Characterization of SARS-CoV-2 N protein reveals multiple functional consequences of the C-terminal domain

Chao Wu[1], Abraham J. Qavi[1*], Asmaa Hachim[2*], Niloufar Kavian[2,3,4*], Aidan R. Cole[1*], Austin B. Moyle[6*], Nicole D. Wagner[6*], Joyce Sweeney-Gibbons[7*], Henry W. Rohrs[6], Michael L. Gross[6], J. S. Malik Peiris[2,8], Christopher F. Basler[7], Christopher W. Farnsworth[1#], Sophie A. Valkenburg[2#], Gaya K. Amarasinghe[#1], and Daisy W. Leung[1,5#],

[1]Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[2]HKU-Pasteur Research Pole, School of Public Health, The University of Hong Kong, Hong Kong, China

[3]Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Assistance Publique–Hôpitaux de Paris, Hôpital Universitaire Paris Centre, Centre Hospitalier Universitaire Cochin, Service d'Immunologie Biologique, Paris, France.

[4]Institut Cochin, INSERM U1016, Université Paris Descartes, Sorbonne Paris Cité, Paris.

[5]Department of Internal Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[6]Department of Chemistry, Washington University in St. Louis, St. Louis, MO, USA.

[7]Center for Microbial Pathogenesis, Institute for Biomedical Sciences, Georgia State University, Atlanta, GA, USA

[8]Division of Public Health Laboratory Sciences, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

*equal contributions

[#]corresponding authors

Key words: SARS-CoV-2, COVID-19, RNA binding, oligomerization, LLPS, nucleoprotein, serology

Summary

Nucleocapsid protein (N) is the most abundant viral protein encoded by SARS-CoV-2, the causative agent of COVID-19. N plays key roles at different steps in the replication cycle and is used as a serological marker of infection. Here we characterize the biochemical properties of SARS-CoV-2 N. We define the N domains important for oligomerization and RNA binding that are associated with spherical droplet formation and suggest that N accessibility and assembly may be regulated by phosphorylation. We also map the RNA binding interface using hydrogen-deuterium exchange mass spectrometry. Finally, we find that the N protein C-terminal domain is the most immunogenic by sensitivity, based upon antibody binding to COVID-19 patient samples from the US and Hong Kong. Together, these findings uncover domain-specific insights into the significance of SARS-CoV-2 N and highlight the diagnostic value of using N domains as highly specific and sensitive markers of COVID-19.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus and the causative agent of COVID-19. Coronavirus has a single-stranded, positive-sense RNA genome. The genome encodes four major structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N). The N protein is the second most proximal to the 3' end of the genome and is one of the most abundantly expressed viral proteins given the multifunctional roles N has during viral replication and assembly (Fung and Liu, 2019; Kim et al., 2020; McBride et al., 2014; Perlman and Netland, 2009). It is estimated that 1,000 copies of N are incorporated into each virion versus only 100 copies of S (Bar-On et al., 2020). Whereas N exists mostly in a phosphorylated state in the cytoplasm, it is predominantly in a dephosphorylated state in mature virions, suggesting that genome packaging is regulated by the phosphorylation state of N (Wu et al., 2014; Wu et al., 2009).

A critical function of N is to encapsidate the ssRNA viral genome to evade immune detection and to protect the viral RNA from degradation by host factors (Chang et al., 2014; McBride et al., 2014). N has two structural domains (**Figure 1A**): a N-terminal domain (N_{NTD}; amino acid residues 44-176) and a C-terminal domain (N_{CTD}; residues 248-369). N_{NTD} is often referred to as the RNA-binding domain (RBD), although the exact regions involved in RNA binding are not yet well defined (Chang et al., 2014; Grossoehme et al., 2009; Gui et al., 2017; Kang et al., 2020; McBride et al., 2014). The isolated N_{CTD} exists as a dimer in solution and is potentially involved in RNA binding as well (Bouhaddou et al., 2020; Gui et al., 2017; Takeda et al., 2008). A conserved serine/arginine rich-linker region (N_{LKR}) connects the N_{NTD} and the N_{CTD}. Phosphorylation of residues in the LKR is believed to regulate discontinuous transcription, particularly for shorter subgenomic

mRNA closer to the 3' end during early stages of replication (Wu et al., 2014; Wu et al., 2009). The LKR, along with Narm and Carm, have been shown to be intrinsically disordered (Chang et al., 2014; Cubuk et al., 2020). However, the LKR is similarly conserved as the NTD and CTD (**Figure 1B** and **Supp. Figure 1B**), supporting its essential functional role across strains of coronaviruses (Chang et al., 2014), while the Narm and Carm are the least conserved regions.

Given its abundant expression and conservation within the genome, N has been used as an antigen for serology tests (Chew et al., 2020; Tang et al., 2020a, b), such as the widely used Roche Elecsys Anti-SARS-CoV-2 assay and Abbott SARS-CoV-2 IgG assay (Tang et al., 2020a, b). Comparison of the serological response to the entire proteome of the SARS-CoV-2 virus using a luciferase immunoprecipitation assay (LIPS) (Hachim et al., 2020) showed that N-specific antibodies dominated the overall antibody response. Furthermore, the T cell responses directed towards N are highly immunodominant in SARS-CoV and SARS-CoV-2 infection, with N-specific memory T cell responses evident 17 years after the initial SARS-CoV infection (Le Bert et al., 2020). Due to epitope conservation of the N protein, there is some T cell cross-reactivity towards the N of the SARS-CoV-2 (Le Bert et al., 2020). The N protein stability, RNA binding characteristics, abundance, and conservation altogether impact immunogenicity for T and B cell immunities.

Previous studies, including our own, have examined fundamental properties of nucleocapsid or nucleoproteins (N) from RNA viruses, which revealed overlapping and unique functions (Arragain et al., 2019; Ding et al., 2016; Lu et al., 2020; Luo et al., 2020; Raymond et al., 2010; Su et al., 2018; Wan et al., 2017). These include the insights into

RNA binding, oligomerization, as well as potential roles in RNA synthesis and immune evasion. However, such studies for SARS-CoV-2 N protein remain incomplete. To address this gap, here we use a series of biochemical and biophysical assays to probe essential functions of N. Our results reveal that oligomeric N provides a continuous RNA binding platform and that N-RNA has different morphologies, including spherical droplets. Our data also suggest how phosphorylation could modulate these processes, which can be exploited by drugs targeting of these cellular processes (Bouhaddou et al., 2020). Given the recent progress in SARS-CoV-2 spike-based vaccines, knowledge of N will likely provide the basis to differentiate individuals with immunity due to natural infections from those that are immunized. Such tools will play a critical role in managing herd immunity. In addition, our domain-specific insights into the immunogenicity of N provide opportunities to further enhance sensitivity and specificity of serology testing. Therefore, insights gained here regarding N protein function and regulation are critical for improving diagnostic testing for the ongoing COVID-19 pandemic and future outbreaks.

Results

Multiple domains within N protein are important for oligomerization. To characterize how the domains of N contribute to oligomerization, we used dynamic light scattering (DLS) to determine the hydrodynamic properties of isolated N domains. The results reveal that there are two major oligomeric species for wildtype (WT) N (N_{WT}; 46 kDa), one with a hydrodynamic radius (R_h) of 8.9 nm and another of 450 nm (**Figure 1C and 1D**). For comparison, the R_h values for maltose binding protein (44 kDa) and bovine serum albumin (66 kDa) are 2.9 nm and 3.7 nm, respectively. Removal of the Narm and Carm ($N_{NTD\text{-}LKR\text{-}}$

$_{CTD}$) generates two major species, similar to N_{WT}. However, both $N_{NTD\text{-}LKR\text{-}CTD}$ populations display reduced polydispersity (peak width, **Supp. Figure 2A**) in the high order oligomeric species, suggesting that both arms contribute to N oligomerization. Removal of the CTD ($N_{NTD\text{-}LKR}$) results in a single peak representing a dimeric species (R_h = 3.9 nm), but with considerable polydispersity. N_{NTD} and N_{CTD} alone are stable domains; N_{NTD} is a monomer (R_h = 2.3 nm) whereas N_{CTD} is a dimer in solution (R_h = 3.5 nm). Exact mass measurement by denaturing mass spectrometry yields values corresponding to the expected sequence (± 1 Da) and supports the identity of the constructs used here (**Supp. Figure 2B-E**).

N provides an oligomerization platform for high affinity RNA binding. N encapsidates viral genomic ssRNA in a sequence-independent manner (Chang et al., 2014). To gain insight into how each N domain contributes to ssRNA binding, we measured fluorescence polarization (FP) upon binding of a FITC-labeled 20-nt ssRNA (Sequence: UUUCACCUCCCUUUCAGUUU) (**Figure 2A**). We find that N_{WT} binds the 20-nt ssRNA with high affinity (K_D = 0.007 ± 0.001 μM). Removal of the Narm and Carm do not impact ssRNA binding (K_D = 0.006 ± 0.002 and 0.006 ± 0.002 μM for $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ and $N_{NTD\text{-}LKR\text{-}CTD}$, respectively) (**Figure 2B**). The individual domains N_{NTD} and N_{CTD} have low affinity binding (K_D = 20 ± 10 and 13 ± 5 μM, respectively), both of which bind with increased affinity after inclusion of the LKR region (**Figure 2B-2D**). Interestingly, addition of CTD that dimerizes onto NTD-LKR improves the binding affinity to the single digit nM range (**Figure 2C**). The affinity increase does not happen when CTD is provided *in trans* (compare $N_{NTD\text{-}LKR}$ + N_{CTD} with $N_{NTD\text{-}LKR\text{-}CTD}$). This affinity increase also occurs when NTD is added to LKR-CTD *in cis* (**Figure 2D**, compare $N_{LKR\text{-}CTD}$ + N_{NTD} with $N_{NTD\text{-}}$

137 $_{LKR-CTD}$). Similar binding curves and K_D values were obtained when fluorescence
138 anisotropy values were converted from polarization (**Supp. Figure 3A**). Collectively, our
139 data show that NTD, CTD, and LKR all contribute to ssRNA binding, and the presence of
140 three domains in tandem confers N with high affinity binding.

141

142 N also binds dsRNA and is proposed to disrupt dsRNA structures formed by transcription
143 regulatory sequences (TRS) during discontinuous transcription (Grossoehme et al., 2009;
144 Keane et al., 2012; Sola et al., 2015). We next tested N binding to a stem-loop RNA
145 (slRNA) (Sequence: GGAAGAUUAAUAAUUUUCC) (**Figure 2E**). We find that N_{WT} binds
146 with relatively high affinity (K_D = 0.051 ± 0.004 μM) whereas both N_{NTD} and N_{CTD} alone
147 have very weak binding affinities (K_D = 120 ± 80 and 60 ± 40 μM, respectively). The
148 addition of LKR significantly improves binding, consistent with binding results for ssRNA.
149 Overall, binding of N to slRNA appears similar to binding to ssRNA, but at an order of
150 magnitude lower than that of ssRNA (**Supp. Figure 3B**). This may be due, in part, to the
151 energetic penalty of unfolding the stem loop structure. Furthermore, it seems that Narm
152 and Carm contribute more to slRNA binding than ssRNA because the reduction of K_D for
153 slRNA are more pronounced after removing the Narm and Carm (**Supp. Figure 3C**).

154

155 **N-RNA forms spherical droplets, requiring multiple N domains.** Purification of
156 recombinant N protein on size exclusion chromatography revealed three populations,
157 including two RNA-bound states (p1 and p2) and an RNA-free state (p3) despite
158 purification in high salt (500 mM) to eliminate RNA binding (**Figure 3A**). Truncation of the

Narm results in an increase of the RNA-free peak (p3), suggesting that N truncations can alter the structure of N and correspondingly impact N-related properties, such as RNA binding and oligomerization, thereby potentially impacting the physiological function of N. We observe an even greater shift to p3 when both Narm and Carm were removed, suggesting that both arms contribute to RNA binding interactions. To gain additional insight into the two RNA-bound populations p1 and p2, we visualized these samples by using negative-stain electron microscopy (EM) in near-physiological salt concentrations (150 mM). N_{WT} p1 contains N-RNA with a loose-coil appearance (**Figure 3B, top left**), similar to that observed for other RNA-bound nucleocapsids (Bharat et al., 2012; Mavrakis et al., 2002). Other recent studies have also observed that N protein undergoes liquid-liquid phase separation (LLPS) when mixed with RNA(Carlson et al., 2020; Cubuk et al., 2020; Iserman et al., 2020; Jack et al., 2020; Savastano et al., 2020). In agreement with these results, for N_{WT} p2, we mostly observe spheres corresponding to liquid droplets separated from the surrounding buffer (**Figure 3B, top right**).

We next examined the role of Narm and Carm on high-order assembly. $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ behaves similarly to N_{WT} and has loose coils in p1 (**Figure 3B, middle left**) and forms spherical liquid droplets in p2 (**Figure 3B, middle right**). However, examination of p2 from $N_{NTD\text{-}LKR\text{-}CTD}$ (**Figure 3B, bottom right**) revealed a much smaller population of liquid droplets (red arrow) and mostly crystal-like needle aggregates, suggesting that the Carm is important for droplet formation. A transition from spherical liquids to needle-like solids is consistent with the liquid-to-solid transitions observed for other proteins that undergo phase separation(Patel et al., 2015).

Phosphorylation of LKR modulates RNA binding and higher-order assembly. Previous studies have shown that N is dephosphorylated in mature virions of a closely related strain, SARS-CoV, and predominantly phosphorylated during active transcription and replication in the cytoplasm (Wu et al., 2014). For SARS-CoV, glycogen synthase kinase (GSK)-3 was shown to phosphorylate N at Ser177 (corresponding to Ser176 in SARS-CoV-2 N) (Wu et al., 2009). Phosphorylation of Ser177 is preceded by phosphorylation of Ser189 and Ser207 (Ser188 and 206 in SARS-CoV-2 N) by other priming kinases (Wu et al., 2009). Moreover, N protein phosphorylation has been qualitatively shown to modulate both RNA binding and phase separation(Carlson et al., 2020; Lu et al., 2020; Savastano et al., 2020). To test if phosphorylation impacts RNA binding and the phase separation of SARS CoV-2 N-RNA, we generated a set of N phosphomimics by mutating specific serine residues to aspartate residues. Size exclusion chromatography showed that, compared to N_{WT}, $N_{S188D/S206D}$ produced a reduced RNA-free peak (p3) and an increased RNA-bound peak (p1) (**Figure 3C**). Introduction of S176D to generate $N_{S176D/S188D/S206D}$ resulted in an even greater shift in p1 and p3 distributions, showing how phosphorylation can affect N interactions with RNA. The height of p2 remains relatively the same for all preparations. Examination of these protein peaks using electron microscopy revealed that $N_{S188D/S206D}$ displays similar loose coils in p1 (**Figure 3D, top left**) and spherical droplets in p2 (**Figure 3D, top right**) for the RNA-bound species. Similar observations were made for $N_{S176D/S188D/S206D}$ (**Figure 3D**). To describe this interaction further, we measured ssRNA binding to the N phosphomimics (**Figure 3E** and **Supp. Figure 4A**). $N_{S176D/S188D/S206D}$ displays ~5-fold lower binding affinity to ssRNA compared to N_{WT} binding, a result consistent with previous work examining the

impact of LKR phosphorylation on RNA binding(Savastano et al., 2020). We observed a similar trend for the $N_{NTD\text{-}LKR}$ construct. Furthermore, binding to slRNA is also affected by these mutations (**Supp. Figure 4B**). Collectively, our data suggest that phosphorylation of the LKR region modulates N interactions with RNA, causing changes in solution properties. Interestingly, there are 14 serine residues in the LKR, of which 13 are found in SARS-CoV, and an increase in phosphorylation in this region may further enhance these changes for RNA interaction and subsequent viral replication and immunogenicity.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) locates the RNA binding interface of NTD-LKR. We performed HDX-MS to locate regions that become protected upon RNA binding. Using $N_{NTD\text{-}LKR\text{-}CTD}$, we observed phase-separation and aggregation upon RNA binding, causing a 100-fold mass spectrometric signal loss for the bound state possibly attributable to phase separation or aggregation inducing poor digestion. However, this problem does not pertain to $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ where digestion yielded 152 peptides after sequential FXIII and peptic digestion and provided 93.3% sequence coverage (**Supp. Figure 5A**).

HDX analysis of $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ shows clear protection in four distinct regions of upon RNA binding (aa residues 41-63, 105-108, 146-171, and 213-230) (**Figure 4A** and **4B**). Residues 133-143 are not perturbed by RNA binding, but peptides covering 146-171 show clear protection. The largest differences in HDX are observed where 50-80% of the residues of unbound peptides undergo a burst phase of HDX in the first 10 s (146-156, 163-171, and 213-230 aa), i.e., the peptides cover regions of little hydrogen bonding in the unbound state. When bound to RNA, the fraction of residues participating in the burst

phase decreases, resulting in observed protection. Then, HDX either converges over time (146-156, 163-171, and 219-223 aa) consistent with protein conformation or RNA binding dynamics, or the HDX never converges in the timescale of the experiment (222-230 aa), consistent with relatively static binding. Interestingly, peptides covering 103-108 aa and 156-159 aa undergo very little HDX throughout the experiment, consistent with either hydrogen bonding of secondary and tertiary structure or a hydrophobic pocket. Of note, HDX decreases for the bound state in these peptides only after 1 h. The low initial HDX limits the dynamic range of binding-induced protection from HDX, but statistically significant protection is still observed.

Further, HDX analysis revealed that the protected regions (**Figure 4C**) overlap well with a basic patch groove in the N_{NTD} structure (**Figure 4D**). In addition, a region (213-230 aa) within the LKR domain, after the structured domain and SR-motif, shows statistically significant HDX protection. Interestingly, we did not detect HDX protection in the SR-motif, which was proposed to bind RNA. This may be due to the Ser-to-Asp mutations introduced into this region, changing the RNA binding patterns. Altogether, HDX results along with our biochemical data define an RNA binding interface within the NTD-LKR region.

N_{CTD} is a sensitive serological marker. Current sero-diagnostic assays to identify COVID-19 positive individuals are based on the detection of antibodies against N due to its abundant expression and corresponding high immune response (Chew et al., 2020; Tang et al., 2020a, b). However, these N-directed serological assays are highly variable and their sensitivity depend on the sampling time-points, ranging from 0% to 93.75% (Liu

et al., 2020; Tang et al., 2020a, b), suggesting that serological markers for SARS-CoV-2 infection can be further improved.

Given that the domains of N impact the various essential properties of N, we next assessed which domains contribute to the observed N immunodominance in RT-PCR confirmed COVID-19 patients. Plasma samples were collected from two cohorts, one in St. Louis, USA (n = 45) and one in Hong Kong (n = 23), at different time points of infection (Supplementary Table 1). Using these samples, we performed enzyme-linked immunosorbent assays (ELISAs) to detect IgG present in COVID-19 patient plasma to different N domains. First, we confirmed that purified N_{WT} is a sensitive serological marker to differentiate between COVID-19 positive and negative individuals (**Figure 5A**). As shown in **Figure 5B**, antibodies against all five N constructs were detected in the COVID-19 cohort ($p < 0.0001$ versus negative controls for all). A cut-off based on the mean of the negatives plus three standard deviations allowed us to assess the performance of each N construct at detecting IgG antibodies in COVID-19 positive individuals (**Figure 5C**). We find that $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ shows the lowest sensitivity (41.2%), whereas the truncated $N_{NTD\text{-}LKR\text{-}CTD}$ can detect more COVID-19 positive individuals (70.6%). Furthermore, N_{CTD} shows the highest combination of sensitivity (75%) and specificity (96.4%) over the other N constructs tested. This is demonstrated by the lowest cut-off score for the N_{CTD} for negative control samples, despite a comparable level of amino acid sequence conservation of the N_{CTD} (29-41%) to the N_{NTD} (32-48%) and N_{LKR} (28-42%) domains with common cold corona viruses (Figure 1B).

We next compared the immunogenicity of $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ to N_{CTD} and N_{NTD} on an IgG heatmap during natural infection to an independent panel of 67 COVID-19 samples from Hong Kong. The magnitude of the IgG response to the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ tends to follow the same trend as that of N_{CTD} (**Figure 5D**). When we assessed the kinetics of N_{CTD} and the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ responses, we find that the magnitude of the N_{CTD} IgG detection tends to reach a similar level to that of $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ at convalescent time-points (after day 14) (**Figure 5E**). The ELISA ratio of $N_{CTD}/N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ demonstrates this finding, pointing to a maturation of the humoral immune response towards the N_{CTD} with time after infection (**Figure 5F**, $p < 0.0001$ for acute versus convalescent time-points). The early dominance of the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ IgG response may reflect the recruitment of a cross-reactive pre-existing N-specific response. This response becomes more specific with time for the N_{CTD} domain as a *de novo* antibody response is made.

Given that some RNA viral N proteins are known immune antagonists, including prior studies of SARS-CoV, we hypothesized that N from SARS-CoV-2 may also suppress the type-I interferon (IFN) signaling pathway (Messaoudi et al., 2015). Using an IFN-β promoter assay, we showed that N has a role in suppressing IFN signaling pathway when stimulated by Sendai virus (SeV) infection (**Figure 5G**). N_{WT} can inhibit IFN-β promoter activity, although not as well as Měnglà virus (MLAV) VP35, a potent inhibitor of IFN signaling (Williams et al., 2020). Both $N_{Narm\text{-}NTD\text{-}LKR}$ and N_{NTD} show modest inhibition at the highest concentration tested. However, $N_{CTD\text{-}Carm}$ shows similar levels of inhibition as N_{WT}, and N_{CTD} displays the highest inhibition even at lower protein concentrations. In

296 summary, N is a potent inhibitor of IFN signaling and the CTD domain appears to be the
297 region critical for mediating this function.

298

299 **Discussion**

300 SARS-CoV-2 N protein is a core viral protein produced by the subgenomic RNA,
301 positioned proximal to the 3' end of genome, display high transcription levels, and is in
302 high abundance in virions. N is prone to forming higher-order oligomers that have a role
303 in binding to the RNA genome of SARS-CoV-2 and nucleocapsid formation. N is also
304 regulated by post-translational modifications, including phosphorylation, which changes
305 the physiochemical properties of N and likely directs its multiple roles at different stages
306 of the viral replication cycle.

307

308 Here we used a series of N constructs to dissect how each domain contributes to
309 oligomerization and RNA binding, and how phosphorylation can modulate these
310 properties. N oligomerization and RNA binding are likely linked; N oligomerization
311 provides a platform for high affinity RNA interactions whereas genomic RNA serves as a
312 string connecting N proteins. The modular domains of N provide multiple regulatory layers
313 for genomic access. For example, phosphorylation of N in the LKR region reduces RNA
314 binding to N and modify distribution of N-RNA species.

315

We also gained insight into the antigenicity of individual domains of N and its potential utility in serological studies. We data revealed that N_{CTD} acts more specifically in detecting infection of SARS-CoV-2, the causative agent of COVID-19, from plasma in comparison to N_{WT}. Consistent with our observation, the N_{CTD} region is also predicted to encompass major antigenic sites of N (Bussmann et al., 2006; Liang et al., 2005). Owing to the relative conservation of N within coronaviruses, it is crucial to understand how N-directed antibodies generated by different coronaviruses are cross-reactive with those that are derived when exposed to SARS-CoV-2. The common cold coronavirus protective immunity is short-lasting (Edridge et al., 2020); reinfection with the same seasonal coronavirus occurred frequently at 12 months after the initial infection using an ELISA approach against the N C-terminal region (coronavirus NL63, 215-377 aa; 229E, 213-389 aa; OC43, 328-428 aa; HKU1, 326-441 aa), which is similar to the N_{CTD} region we used (SARS-CoV-2, 248-369 aa). Interestingly, 2 out of 10 individuals assessed in this study of longitudinal donors unexposed to SARS-CoV-2 by Edridge *et al.*, produced broadly reactive antibodies towards SARS-CoV-2 N_{WT}. The possibility of broadly reactive antibodies in unexposed individuals highlights the need for domain specific serology, such as our use of the N_{CTD} for increased sensitivity to discriminate COVID-19 cases, while reducing the false-positive rate from cross reactive antibodies generated by infections of the common cold coronaviruses.

In conclusion, we describe our efforts to characterize domain specific insights into essential biochemical and serological properties of SARS-CoV-2 N. Our results advance the understanding of viral replication processes and highlight the diagnostic value of using

N domains as a highly specific and sensitive markers of COVID-19. While our study highlights the important functions associated with N_{CTD}, much remains to be characterized, including the mechanistic link between the high immunogenicity of N_{CTD} and the physical properties such as RNA binding and oligomerization.

Methods

Patients and sample collection. Our study enrolled a total of 67 patients with RT-PCR confirmed COVID-19 infection: with 45 patients from St. Louis, MO, USA, and 23 patients from Hong Kong, PRC. The negative samples (n = 28) used in this study were from St. Louis, USA, and were obtained from patients following the start of the pandemic. Plasma samples were obtained from patients at Barnes-Jewish Hospital (St. Louis, MO, USA) and the Hong Kong Island West Cluster of Hospitals (Hong Kong, PRC). Both hospital systems are urban, tertiary-care, academic medical centers. Positive and negative patients from all cohorts were confirmed using standard of care, RT-PCR based methods. The collection of patient plasma was approved by the Human Research Protection Office at Washington University in St. Louis (IRB reference number 202007097) and the Institutional Review Board of The Hong Kong University and the Hong Kong Island West Cluster of Hospitals (IRB reference number UW20-169). Plasma samples were collected from heparinized blood. Sample day was defined as days post-symptom onset.

Enzyme-linked immunosorbent assay (ELISA). ELISA assays were performed with Nucleoprotein (N) proteins made in house, as described below. Briefly, recombinant N proteins were coated on 96 well flatbottom immunosorbent plates (Nunc Immuno MaxiSorp) at a concentration of 500 ng/mL, in 100 µL coating buffer (PBS with 53% Na_2CO_3 and 42% $NaHCO_3$, pH 9.6) at 4°C overnight. An additional plate coated with a non‐specific protein (blocking buffer, PBS with 5% fetal bovine serum (FBS)) was used to measure the background binding of each plasma sample. Following FBS blocking and thorough washing, diluted plasma samples (1:100) were bound for 2 hours, further washed and then detected by an anti‐human IgG secondary antibody labelled with HRP (Invitrogen), and absorbance detected at 450nm on a spectrophotometer (Wallac).

Protein Expression and Purification. SARS-CoV-2 N constructs were expressed as His-tag fusion proteins in BL21 (DE3) *E. coli* cells (Novagen). At OD_{600} of 0.6-0.7, recombinant protein expression was induced with 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for 12-14 h at 18˚C. Cells were harvested and resuspended in lysis buffer containing 20 mM Tris (pH 7.5), 1 M NaCl, 20 mM imidazole, 5 mM 2-mecaptoethanol (BME). Cells were lysed using an EmulsiFlex-C5 homogenizer (Avestin) and lysates were clarified by centrifugation at 30,000 x *g* at 4 ˚C for 40 min. N proteins were purified using affinity tag and gel filtration columns. Purity of N proteins were determined by Coomassie staining of SDS-PAGE.

Negative Staining EM. 2 µL of N sample at a concentration of 1 mg/mL was applied to a glow-discharged copper grid (Ted Pella), washed twice with water before staining with

2% uranyl acetate for 30 s and air dried. Grids were imaged using a JEOL JEM-1400plus Transmission Electron Microscope operating at 120 kV and recorded with an AMT XR111 high-speed 4k x 2k pixel phosphor-scintillated 12-bit CCD camera.

Dynamic Light Scattering (DLS). DLS experiments were performed on a DynaPro-PlateReader II (Wyatt Technologies Corporation). Measurements of N samples in triplicates (1 mg/mL) were obtained at 25 ºC and analyzed using Dynamics software (Wyatt).

Fluorescence Polarization Assay (FPA). FPA experiments were performed on a Cytation5 plate reader (BioTek) operating on Gen5 software. Excitation and emission wavelengths were set to 485 and 528 nm, respectively, with a bandpass of 20 nm. Read height and G factor were set to 8.5 mm and 1.26, respectively using the autogain function. For RNA binding experiments, fluorescein isothiocyanate (FITC) labelled 20 nt-ssRNA or 19 nt slRNA at a final concentration of 1 nM was loaded on N samples (in 20 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM TCEP, 5% glycerol) at concentrations ranging from 0.4 nM to 10 μM in a 96-well plate. After 10 min of incubation, fluorescence polarization signals were read. The fluorescence polarization values were then plotted against N concentrations to fit the dissociation constant, K_D, using ORIGIN software. For anisotropy plots, anisotropy values were converted from polarization according to previous research (Kozlov et al., 2012).

LC-MS Analysis. Unless otherwise indicated, all chemical reagents were sourced from Millipore Sigma and used without further purification. For LC-MS analyses, 30 pmol of

protein in 50 µL of 1:1 solvent mixture of acetonitrile:water with 0.1% formic acid (CovaChem) was loaded onto a C8 trap (ZORBAX Eclipse XDB C8 column, 2.1 x 15 mm, Agilent), desalted for 3 min by using water/0.1% formic acid at a flow rate of 100 µL/min, and eluted using an 14 minute gradient from 0 to 80% acetonitrile/0.1% formic acid at a flow rate of 100 µL/min. Samples were analyzed using a MaXis 4G Q-TOF (Bruker Daltonics). The mass spectrum was extracted guided by the elution peak and submitted to PMI Intact Mass and searched for M values ranging from 5-50 kDa.

HDX-MS. $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ was incubated with a 20-nt ssRNA at a 1:1 ratio. After incubation, 2 µL of 50 µM protein/protein-RNA in PBS (pH 7.4) was diluted 10-fold (v/v) with labeling buffer (PBS in D_2O, pD 7.0) (D_2O from Cambridge Isotope Laboratories), incubated for 10, 30, 300, and 3600 s on ice, quenched by using a 60% dilution with 3 M urea, PBS (pH 2.5), and flash frozen for later LC-MS analysis. A 0 s control was prepared with PBS in H_2O. Prior to incubation, each 50 µL of 2 µM sample was thawed for 1 min at 37 °C before injection into a custom-built liquid chromatography (LC) apparatus for LC-MS analysis. The labelled protein passed through two in-house packed protease columns (2 mm x 20 mm), coupled so that the first using protease from *Aspergillus saitoi* type XIII (FXIII) and the second porcine pepsin (0.1% formic acid, flow rate 200 µL/min); the resulting peptides were trapped on a ZORBAX Eclipse XDB C8 column (2.1 mm x 15 mm, Agilent), desalted for 3 min, and then separated on a Hypersil Gold C18 column (2.1 x 50 mm, Thermo Fisher) with a 10.5 min linear gradient from 4 – 40% acetonitrile/0.1% formic acid (flow rate 100 µL/min). All valves, tubes, and columns (except for the protease columns, which lose activity at low temperature) were submerged in ice during the

experiment to minimize back exchange. Peptides were eluted into a Bruker Maxis HM Q-TOF MS for mass analysis. Experiments were in duplicate unless otherwise indicated. The HDX data processing was performed by using HDExaminer (version 2.5.1, Sierra Analytics, Inc.).

IFN-β promoter reporter gene assay. HEK-293T cells (5 x 10^4) were co-transfected using Lipofectamine 2000 with 25 ng of an IFN-β promoter-firefly luciferase reporter plasmid, 25 ng of pRL-TK *Renilla* luciferase reporter plasmid, and 125, 12.5, and 1.25 ng of the indicated viral protein expression plasmid. Twenty-four hours post-transfection, cells were mock-treated or SeV (15 hemagglutination units / ml) infected. Eighteen hours post-treatment or post-infection, cells were lysed and analyzed for luciferase activity using a Dual-Luciferase reporter assay system (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase activity. Assays were performed in triplicate; error bars indicate the standard error of the mean (SEM) for the triplicate. Viral protein expression was confirmed by Western blot analysis.

Acknowledgements

We thank Dr. N. Krogan (UCSF) for sharing SARS-CoV-2 plasmids, Drs. A. Holehouse and A. Soranno (Washington University School of Medicine) for providing critical feedback, and R. Ridings for coordinating studies between WUSM and external groups. Research was supported by Fast Grant ##2161 (Emergent Ventures) to G.K.A. and NIH grants (P01AI120943, R01AI123926 to C.F.B, G.K.A. and C.F.B; R01AI107056 to D.W.L.; P41GM103422 and R24GM136766 to M.L.G.; P01AI120943, R01AI143292 and R01AI148663 C.F.B). S.A.V was supported by COVID190115 and COVID190126

454 Health and Medical Research Fund, Food and Health Bureau, Hong Kong and NIH/NIAID
455 CEIRS contract HHSN272201400006C. A.J.Q. is supported by a NIH T32 training grant
456 (T32CA009547).

457

458 **Author contributions:** CW, GKA, and DWL conceived the overall project. All authors
459 were integral to the design and execution of the study. CW, AQ, GKA, and DWL wrote
460 the initial draft with significant input from all authors.

461

462 **Figures**

463 **Figure 1. Characterization of N oligomerization with dynamic light scattering. A**.
464 Domain architecture of N. N has two structural domains: NTD and CTD. The sequence
465 between NTD and CTD is a linker region (LKR) containing a serine-arginine rich motif.
466 The Narm, LKR, and Carm are predicted to be disordered based upon sequence analysis.
467 **B.** Sequence identity between N of SARS-CoV-2 and common cold coronaviruses and
468 other epidemic severe coronaviruses MERS-CoV and SARS-CoV. **C.** Measurements of
469 N oligomerization with dynamic light scattering. Measured hydrodynamic radii, R_h, are
470 reported in D. **D.** Table summarizes the DLS data for all constructs. Numbers are reported
471 as average and standard deviation of three experiments.

472

473 **Figure 2. SARS-CoV-2 nucleocapsid protein binds ssRNA with high affinity. A.**
474 Principles of fluorescence polarization assay to measure RNA binding. Increasing

concentrations of N was titrated into 1nM of labeled RNA. Protein binding to labeled RNA leads to slower tumbling of labeled RNA, resulting in increased fluorescence polarization. **B.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black square), 16 ± 12 μM (N_{NTD}, red circle), 13 ± 5 μM (N_{CTD}, blue up triangle), 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), and 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, green diamond). **C.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.006 ± 0.002 nM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), 13 ± 5 μM (N_{CTD}, blue circle), 0.50 ± 0.08 μM ($N_{NTD\text{-}LKR}$, orange star), and 0.44 ± 0.04 μM ($N_{NTD\text{-}LKR}$ + N_{CTD}, purple pentagon). **D.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), 16 ± 12 μM (N_{NTD}, red circle), 0.35 ± 0.04 μM ($N_{LKR\text{-}CTD}$, orange up triangle), and 0.72 ± 0.09 μM (N_{NTD} + $N_{LKR\text{-}CTD}$, purple down triangle). **E.** Fluorescence polarization binding curves of N constructs to a slRNA. The fitted K_D values are 0.051 ± 0.004 μM (N_{WT}, black square), 124 ± 84 μM (N_{NTD}, red circle), 65 ± 44 μM (N_{CTD}, blue up triangle), 2.5 ± 0.5 μM ($N_{NTD\text{-}LKR}$, magenta down triangle), 0.22 ± 0.02 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, green diamond), and 0.10 ± 0.01 μM ($N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, navy left triangle). **F**. Table summarizes K_D values (μM) for key constructs binding to ssRNA and slRNA. Numbers are reported as average and standard deviation of two experiments.

Figure 3. N-RNA phase separates and phosphorylation modulates N-RNA. A. Size exclusion chromatography of N constructs (N_{WT}, black; $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, red; $N_{NTD\text{-}LKR\text{-}CTD}$, blue) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol from a 2 L pellet.

Samples from peak 1 (p1) and p2 contain RNA whereas p3 are RNA-free based upon absorbance from the 260/280 ratio. **B.** Negative stain electron microscopy (EM) image of p1 and p2 for N_{WT}, $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, and $N_{NTD\text{-}LKR\text{-}CTD}$ in 150 mM NaCl. Samples were diluted into 150 mM NaCl before negative-staining fixation by uranyl acetate. **C.** Size exclusion chromatography of N constructs (N_{WT}, black; $N_{S188D/S206D}$, blue; $N_{S176D/S188D/S206D}$, red) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol. **D**. Negative stain electron microscope image of $N_{S188D/S206D}$ and $N_{S176D/S188D/S206D}$ in 150 mM NaCl. **E.** Fluorescence polarization binding curves of N mutants to a 20-nt ssRNA. The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black square), 0.015 ± 0.002 μM ($N_{S176D/S188D/S206D}$, black circle), 0.505 ± 0.075 μM ($N_{NTD\text{-}LKR}$, orange up triangle), and 1.1 ± 0.2 μM ($N_{NTD\text{-}LKR\ S176D/S188D/S206D}$, orange down triangle). Numbers are reported as average and standard deviation of two experiments.

Figure 4. HDX-MS mapping of RNA binding to $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$. A. Woods' plot showing cumulative differential HDX and validating differences using global significance limits. The horizontal bars depict the cumulative HDX differences between the RNA-bound and unbound $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$. Standard deviations are shown for each peptide. Peptides showing statistically significant differences are differentiated by global significance limit (*, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$). The blue shade of the peptide bar indicates differing statistical significance (light blue, medium blue, and navy, respectively); gray peptide bars depict peptides where statistically significant differences in HDX were not observed. Vertical bars show previously reported binding sites (residues reported for RNA-binding CoV2 N-protein (Dinesh et al., 2020; Ye et al., 2020), AMP-binding HCoV-

OC43 (Lin et al., 2014; Ye et al., 2020), and for both are shown in red, yellow, and orange, respectively). Secondary structure (PDB 6M3M) is shown above. **B**. Representative kinetic plots showing peptide level HDX as a function of exchange time (unbound, black; bound to RNA, red). **C**. Sites of protection measured by HDX mapped on the N_{NTD} structure (PDB 6M3M). Statistically significant HDX protection, regions of no difference in HDX, and regions where lacking proteolytic coverage results in no data are shown in teal, light gray, and dark gray, respectively. Those residues unresolved in the structure are shown as a dashed line, with the exception of those reporting a statistically significant difference in teal. **D**. Electrostatic potential calculated with APBS mapped on to the N_{NTD} structure (PDB 6M3M) shows a major positive charge groove. Red and blue represent negative and positive electrostatic potential. The color scale is in kT/e units.

Figure 5. The CTD of N is a highly sensitive serological marker. A. ELISA data of N_{WT} screened against plasma of COVID-19 positive and negative individuals from a combined Hong Kong, PRC and St. Louis, MO, USA cohort. Black solid line indicates the mean OD_{450} value for each population. **** $p < 0.0001$. **B.** ELISAs with the various N constructs for patient IgG. ELISAs were performed on plasma samples from COVID-19 patients (n = 68) and negative controls (n = 28). The cut-off is represented by the dotted line and calculated as the mean + 3 standard deviations of the negative population. Mean values ± standard deviation of COVID-19 and negative groups are shown. **C.** Sensitivity and specificity for each of the N domains calculated from the ELISA results. **D.** Heat-map of ELISA results for $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, N_{CTD}, and N_{NTD} constructs from COVID-19 samples (n = 67). Each column represents an individual sample. **E.** Maturation of the N_{CTD} and

$N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ IgG response over time (n = 67). **F.** Ratio of OD_{450} for N_{CTD} and $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ for acute and convalescent time-points. Mean values ± standard deviation of acute and convalescent COVID-19 samples are shown. Experiments were repeated twice. Statistical significance was calculated by unpaired Student's t-test, ****p < 0.0001. **G**. Inhibition of SeV-induced IFNβ promoter activation by N constructs. Fold changes are relative to vector-only (V) transfections without SeV infection. MLAV VP35 served as a positive control for inhibition. Three transfection concentrations were used: 1.25, 12.5, and 125 ng/well. Statistical significance was determined by performing a one-way ANOVA followed with Tukey multiple comparison as compared to Sendai virus-infected control; **** p < 0.0001, *** p < 0.0002, ** p < 0.0021, * p < 0.0332.

Supplementary Figure Legends

Supplementary Figure 1, related to Figure 1. Multiple sequence alignment of coronavirus nucleocapsids. A. Multiple sequence alignment of coronavirus nucleocapsids. Sequences were aligned using Clustal Omega. Accession numbers used are 229E (APT69891.1), NL63 (YP_003771.1), HK1 (AAT98585.1), OC43 (AAR01019.1), MERS (AKL80590.1), SARS (AAP30037.1), SARS2 (YP_009724397.2). Alignments were analyzed using ESPript3. The three serines (176, 188, and 206) are labeled with red arrows. **B**. Sequence identity between SARS-CoV-2 N and that of common cold coronaviruses and MERS and SARS. FL, full length. All units are in %. Percent identity matrixes for corresponding domains of N are generated using Clustal2.1.

566 **Supplementary Figure 2, related to Figure 1. Denaturing mass spectra of N protein**
567 **truncations N_{NTD}, N_{CTD}, $N_{NTD-LKR}$, and $N_{NTD-LKR-CTD}$. A.** DLS polydispersity table for N
568 constructs. Higher values indicate broader size distributions. Numbers are reported as
569 average and standard deviation of three experiments. Deconvolution yields experimental
570 M values of 16,881 Da, 16,078 Da, 24,155 Da, and 37,829 Da for (**B**) N_{NTD}, (**C**) N_{CTD}, (**D**)
571 $N_{NTD-LKR}$, and (**E**) $N_{NTD-LKR-CTD}$ respectively, matching theoretical values within 1 Da, based
572 on protein sequence. Deconvoluted mass spectra (right) and adduct series corresponding
573 to pervasive trifluoroacetic adducts (delta mass 114 Da, circle) and a-N-gluconoylation
574 (delta mass 178 Da, star). TFA adducts are introduced by the ion pairing reagent in
575 solvent, while α-N-gluconoylation is a common modification occurring on His-tagged
576 proteins. Native spray of NTD (not pictured) yielded no peaks with delta mass 114 Da,
577 but retained a single delta mass 178 Da, confirming transient TFA adducts are an artifact
578 of the denaturing experiment, but the α-N-gluconoylation of the His-tag is covalent.

579

580 **Supplementary Figure 3, related to Figure 2. Nucleocapsid binds stem-loop RNA**
581 **with reduced affinity. A.** Fluorescence anisotropy binding curves of N constructs to a
582 20-nt ssRNA. Anisotropy values were converted from polarization according to previous
583 research (Kozlov et al., 2012). The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black
584 square), 0.006 ± 0.002 μM ($N_{NTD-LKR-CTD}$, magenta circle), 14 ± 5 μM (N_{CTD}, blue up
585 triangle) and 18 ± 14 μM (N_{NTD}, red down triangle). These values are very close to those
586 of polarization. In this system, binding monitored by anisotropy is similar to that of
587 polarization. **B.** Fitted K_D values for N constructs binding to ssRNA (black) and slRNA
588 (grey). **C.** Ratio of K_D of slRNA over that of ssRNA for N constructs. The reduced binding

589 to slRNA is around 5-fold for most N constructs. The reduction is higher for those of N_{NTD-}
590 $_{LKR-CTD-Carm}$ and $N_{NTD-LKR-CTD}$, suggesting Narm and Carm are more involved in slRNA
591 binding.

592

593 **Supplementary Figure 4, related to Figure 3. Phosphorylation mimetics of N reduce**
594 **RNA binding. A.** Fluorescence polarization binding curves of N constructs to a 20-nt
595 ssRNA. The fitted K_D values are 0.007 ± 0.001 μM for N_{WT}, 0.015 ± 0.002 μM for
596 $N_{S188D/S206D}$, and 0.023 ± 0.006 μM for $N_{S176D/S188D/S188D}$. **B.** Fluorescence polarization
597 binding curves of N constructs to a 19-nt slRNA. The fitted K_D values are 1.3 ± 0.3 μM
598 ($N_{LKR-CTD}$, black square), 3.0 ± 0.5 μM ($N_{NTD-LKR}$, red circle), and 2.9 ± 1.4 μM ($N_{NTD-LKR}$
599 $_{S176D/S188D/S206D}$, blue up triangle).

600

601 **Supplementary Figure 5, related to Figure 4. Sequence coverage of $N_{NTD-LKR}$**
602 **$_{S176D/S188D/S206D}$ in HDX-MS and HDX of the unbound state. A.** Protein coverage map of
603 unbound state $N_{NTD-LKR\ S176D/S188D/S206D}$ HDX yielding 152 peptides with 93.3% sequence
604 coverage. Peptide bars are colored according to their average %HDX relative to the color
605 bar, where cooler colors depict low average %HDX and warmer colors depict high
606 average %HDX. The secondary structure reported by PDB 6M3M is shown above the
607 sequence. Overall, the HDX of the unbound state is largely consistent with the reported
608 secondary structure and a well-ordered tertiary structure; regions outside of the reported
609 structure undergo relatively rapid HDX, consistent with a lack of backbone hydrogen
610 bonding. Interestingly, despite a lack of reported secondary structure in the region of 155-

160, relatively low HDX was observed, consistent with either hydrogen bonding of secondary/tertiary structure or a hydrophobic pocket. SR-motif in LKR are boxed in red. **B**. All kinetic plots used in the peptide-level difference plot in **Figure 5A** show peptide level HDX as a function of exchange time (unbound, black; bound to RNA, red).

References

Arragain, B., Reguera, J., Desfosses, A., Gutsche, I., Schoehn, G., and Malet, H. (2019). High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms. eLife *8*, e43075.

Bar-On, Y.M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the numbers. eLife *9*, e57309.

Bharat, T.A.M., Noda, T., Riches, J.D., Kraehling, V., Kolesnikova, L., Becker, S., Kawaoka, Y., and Briggs, J.A.G. (2012). Structural dissection of Ebola virus and its assembly determinants using cryo-electron tomography. Proceedings of the National Academy of Sciences *109*, 4275-4280.

Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V.V., Correa Marrero, M., Polacco, B.J., Melnyk, J.E., Ulferts, S., Kaake, R.M.*, et al.* (2020). The Global Phosphorylation Landscape of SARS-CoV-2 Infection. Cell *182*, 685-712.e619.

Bussmann, B.M., Reiche, S., Jacob, L.H., Braun, J.M., and Jassoy, C. (2006). Antigenic and cellular localisation analysis of the severe acute respiratory syndrome coronavirus nucleocapsid protein using monoclonal antibodies. Virus Res *122*, 119-126.

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. Molecular Cell.

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 Research *103*, 39-50.

Chew, K.L., Tan, S.S., Saw, S., Pajarillaga, A., Zaine, S., Khoo, C., Wang, W., Tambyah, P., Jureen, R., and Sethi, S.K. (2020). Clinical evaluation of serological IgG antibody

response on the Abbott Architect for established SARS-CoV-2 infection. Clinical Microbiology and Infection *26*, 1256.e1259-1256.e1211.

Cubuk, J., Alston, J., Incicco, J., Singh, S., Stuchell-Brereton, M., Ward, M., Zimmerman, M., Vithani, N., Griffith, D., Wagoner, J.*, et al.* (2020). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA (bioRxiv).

Dinesh, D.C., Chalupska, D., Silhan, J., Veverka, V., and Boura, E. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. bioRxiv, 2020.2004.2002.022194.

Ding, B., Qin, Y., and Chen, M. (2016). Nucleocapsid proteins: roles beyond viral RNA packaging. Wiley Interdiscip Rev RNA *7*, 213-226.

Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., Bakker, M., Klein, M., Loens, K., Jebbink, M.F., Matser, A., Kinsella, C.M., Rueda, P.*, et al.* (2020). Seasonal coronavirus protective immunity is short-lasting. Nature Medicine.

Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. Annual Review of Microbiology *73*, 529-557.

Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., Leibowitz, J.L., and Giedroc, D.P. (2009). Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. Journal of Molecular Biology *394*, 544-557.

Gui, M., Liu, X., Guo, D., Zhang, Z., Yin, C.-C., Chen, Y., and Xiang, Y. (2017). Electron microscopy studies of the coronavirus ribonucleoprotein complex. Protein & Cell *8*, 219-224.

668 Hachim, A., Kavian, N., Cohen, C.A., Chin, A.W.H., Chu, D.K.W., Mok, C.K.P., Tsang,
669 O.T.Y., Yeung, Y.C., Perera, R.A.P.M., Poon, L.L.M.*, et al.* (2020). ORF8 and ORF3b
670 antibodies are accurate serological markers of early and late SARS-CoV-2 infection.
671 Nature Immunology *21*, 1293-1301.
672 Iserman, C., Roden, C., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., Park, C.,
673 Boppana, A., Fritch, E., Hou, Y.J.*, et al.* (2020). Specific viral RNA drives the SARS CoV-
674 2 nucleocapsid to phase separate. bioRxiv, 2020.2006.2011.147199.
675 Jack, A., Ferro, L.S., Trnka, M.J., Wehri, E., Nadgir, A., Costa, K., Schaletzky, J., and
676 Yildiz, A. (2020). SARS CoV-2 nucleocapsid protein forms condensates with viral
677 genomic RNA. bioRxiv, 2020.2009.2014.295824.
678 Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z.,
679 Chen, Q.*, et al.* (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA
680 binding domain reveals potential unique drug targeting sites. Acta Pharmaceutica Sinica
681 B *10*, 1228-1238.
682 Keane, S.C., Liu, P., Leibowitz, J.L., and Giedroc, D.P. (2012). Functional Transcriptional
683 Regulatory Sequence (TRS) RNA Binding and Helix Destabilizing Determinants of Murine
684 Hepatitis Virus (MHV) Nucleocapsid (N) Protein. Journal of Biological Chemistry *287*,
685 7063-7073.
686 Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The
687 Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914-921.e910.
688 Kozlov, A.G., Galletto, R., and Lohman, T.M. (2012). SSB–DNA Binding Monitored by
689 Fluorescence Intensity and Anisotropy. In Single-Stranded DNA Binding Proteins:
690 Methods and Protocols, J.L. Keck, ed. (Totowa, NJ: Humana Press), pp. 55-83.

691 Le Bert, N., Tan, A.T., Kunasegaran, K., Tham, C.Y.L., Hafezi, M., Chia, A., Chng, M.H.Y.,
692 Lin, M., Tan, N., Linster, M.*, et al.* (2020). SARS-CoV-2-specific T cell immunity in cases
693 of COVID-19 and SARS, and uninfected controls. Nature *584*, 457-462.
694 Liang, Y., Wan, Y., Qiu, L.-w., Zhou, J., Ni, B., Guo, B., Zou, Q., Zou, L., Zhou, W., Jia,
695 Z.*, et al.* (2005). Comprehensive Antibody Epitope Mapping of the Nucleocapsid Protein
696 of Severe Acute Respiratory Syndrome (SARS) Coronavirus: Insight into the Humoral
697 Immunity of SARS. Clinical Chemistry *51*, 1382-1396.
698 Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S., and Hou, M.-H. (2014).
699 Structural Basis for the Identification of the N-Terminal Domain of Coronavirus
700 Nucleocapsid Protein as an Antiviral Target. Journal of Medicinal Chemistry *57*, 2247-
701 2257.
702 Liu, W., Liu, L., Kou, G., Zheng, Y., Ding, Y., Ni, W., Wang, Q., Tan, L., Wu, W., Tang, S.*,*
703 *et al.* (2020). Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked
704 Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. Journal of Clinical
705 Microbiology *58*, e00461-00420.
706 Lu, S., Ye, Q., Singh, D., Villa, E., Cleveland, D.W., and Corbett, K.D. (2020). The SARS-
707 CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA
708 and the membrane-associated M protein. bioRxiv, 2020.2007.2030.228023.
709 Luo, M., Terrell, J.R., and Mcmanus, S.A. (2020). Nucleocapsid Structure of Negative
710 Strand RNA Virus. Viruses *12*, 835.
711 Mavrakis, M., Kolesnikova, L., Schoehn, G., Becker, S., and Ruigrok, R.W.H. (2002).
712 Morphology of Marburg Virus NP–RNA. Virology *296*, 300-307.

McBride, R., Zyl, M.v., and Fielding, B. (2014). The Coronavirus Nucleocapsid Is a Multifunctional Protein. Viruses *6*, 2991-3018.

Messaoudi, I., Amarasinghe, G.K., and Basler, C.F. (2015). Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. Nature Reviews Microbiology *13*, 663-676.

Patel, A., Lee, Hyun O., Jawerth, L., Maharana, S., Jahnel, M., Hein, Marco Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, Titus M.*, et al.* (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Perlman, S., and Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. Nature Reviews Microbiology *7*, 439-450.

Raymond, D.D., Piper, M.E., Gerrard, S.R., and Smith, J.L. (2010). Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proceedings of the National Academy of Sciences *107*, 11769-11774.

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. Nature Communications *11*, 6041.

Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annual Review of Virology *2*, 265-288.

Su, Z., Wu, C., Shi, L., Luthra, P., Pintilie, G.D., Johnson, B., Porter, J.R., Ge, P., Chen, M., Liu, G.*, et al.* (2018). Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. Cell *172*, 966-978.e912.

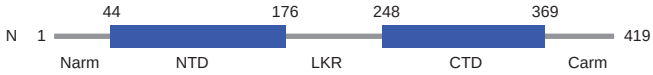
Takeda, M., Chang, C.-k., Ikeya, T., Güntert, P., Chang, Y.-h., Hsu, Y.-l., Huang, T.-h., and Kainosho, M. (2008). Solution Structure of the C-terminal Dimerization Domain of

736 SARS Coronavirus Nucleocapsid Protein Solved by the SAIL-NMR Method. Journal of
737 Molecular Biology *380*, 608-622.
738 Tang, M.S., Hock, K.G., Logsdon, N.M., Hayes, J.E., Gronowski, A.M., Anderson, N.W.,
739 and Farnsworth, C.W. (2020a). Clinical Performance of the Roche SARS-CoV-2
740 Serologic Assay. Clinical Chemistry *66*, 1107-1109.
741 Tang, M.S., Hock, K.G., Logsdon, N.M., Hayes, J.E., Gronowski, A.M., Anderson, N.W.,
742 and Farnsworth, C.W. (2020b). Clinical Performance of Two SARS-CoV-2 Serologic
743 Assays. Clinical Chemistry *66*, 1055-1062.
744 Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., and Briggs, J.A.G.
745 (2017). Structure and assembly of the Ebola virus nucleocapsid. Nature *551*, 394-397.
746 Williams, C.G., Gibbons, J.S., Keiffer, T.R., Luthra, P., Edwards, M.R., and Basler, C.F.
747 (2020). Impact of Měnglà Virus Proteins on Human and Bat Innate Immune Pathways.
748 Journal of Virology *94*, e00191-00120.
749 Wu, C.-H., Chen, P.-J., and Yeh, S.-H. (2014). Nucleocapsid Phosphorylation and RNA
750 Helicase DDX1 Recruitment Enables Coronavirus Transition from Discontinuous to
751 Continuous Transcription. Cell Host & Microbe *16*, 462-472.
752 Wu, C.-H., Yeh, S.-H., Tsay, Y.-G., Shieh, Y.-H., Kao, C.-L., Chen, Y.-S., Wang, S.-H.,
753 Kuo, T.-J., Chen, D.-S., and Chen, P.-J. (2009). Glycogen Synthase Kinase-3 Regulates
754 the Phosphorylation of Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid
755 Protein and Viral Replication. Journal of Biological Chemistry *284*, 5229-5239.
756 Ye, Q., West, A.M.V., Silletti, S., and Corbett, K.D. (2020). Architecture and self-assembly
757 of the SARS-CoV-2 nucleocapsid protein. Protein Science *29*, 1890-1901.

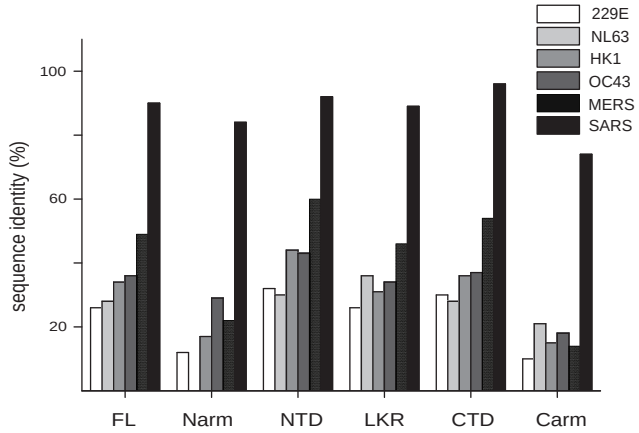
758

Figure 1. Characterization of N oligomerization with dynamic light scattering.

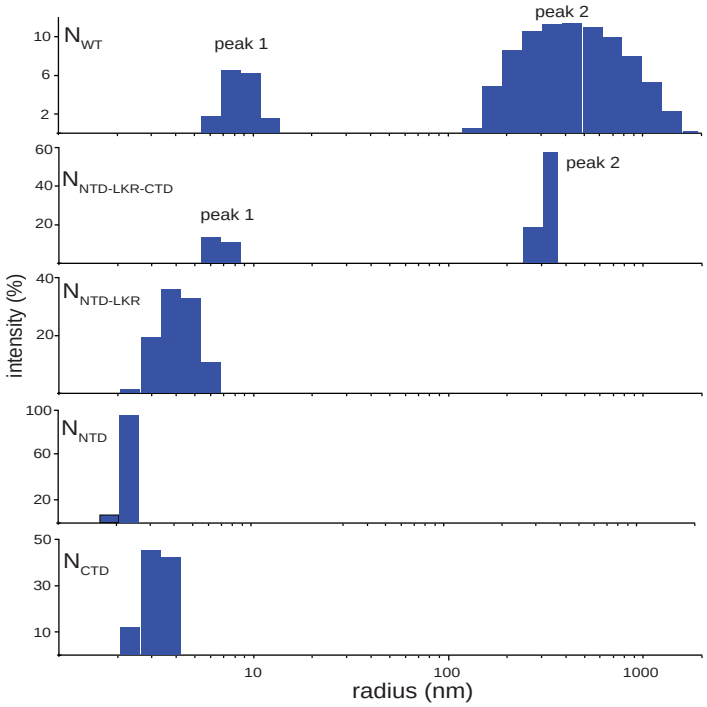
A



B



C

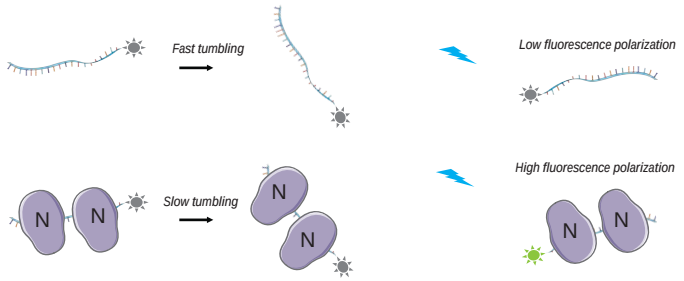


D

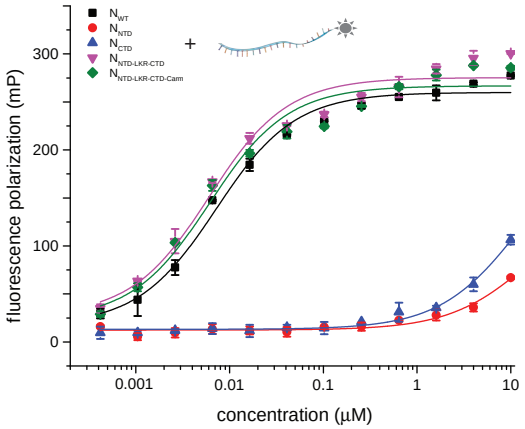
	MW (kDa)	Rh (nm)	
		peak 1	peak 2
N_{WT}	46	8.9 ± 0.2	450 ± 100
$N_{NTD-LKR-CTD}$	35	6.5 ± 0.4	360 ± 20
$N_{NTD-LKR}$	22	3.9 ± 0.1	
N_{NTD}	14	2.3 ± 0.1	
N_{CTD}	14	3.5 ± 0.1	

Figure 2. Nucleocapsid binds RNA with high affinity.

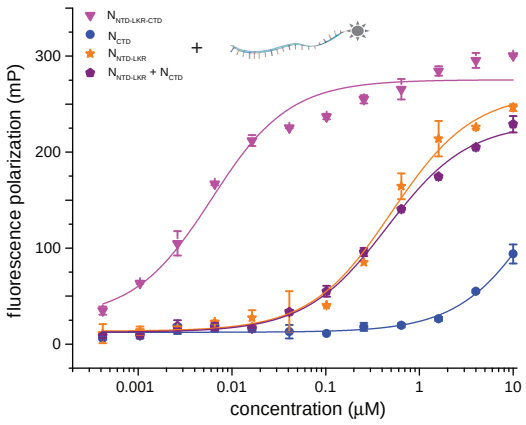
A



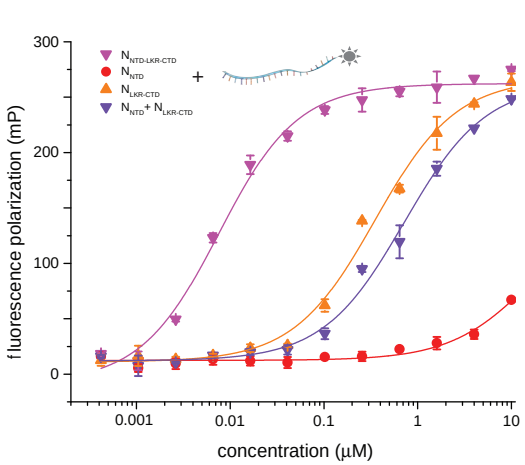
B



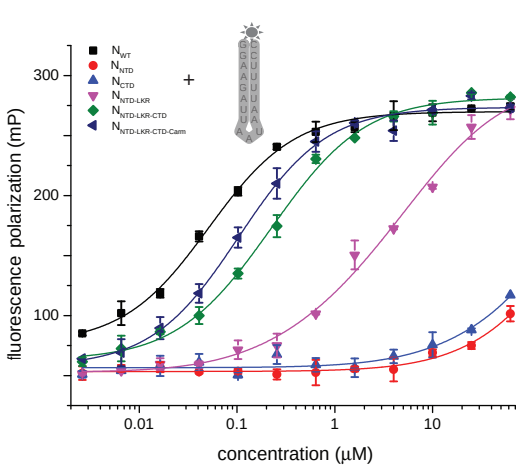
C



D



E



F

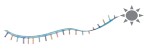
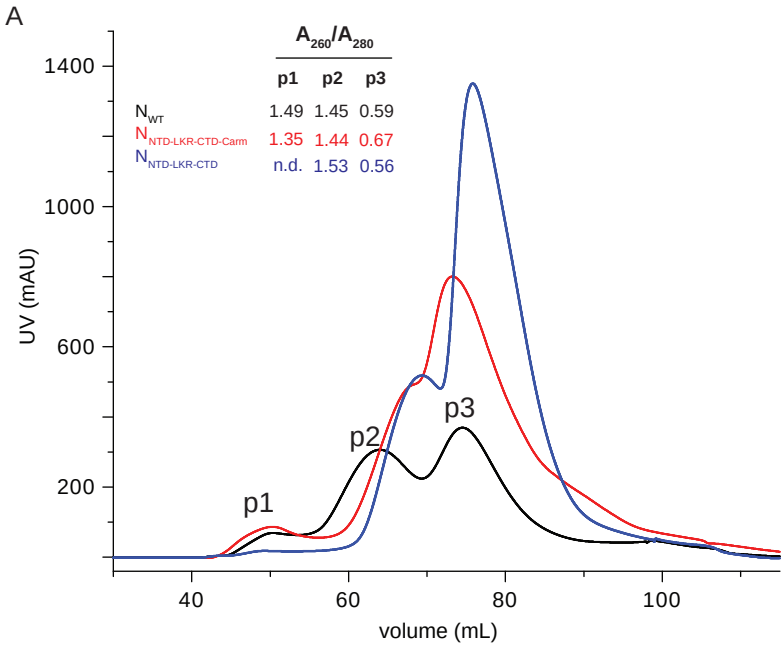
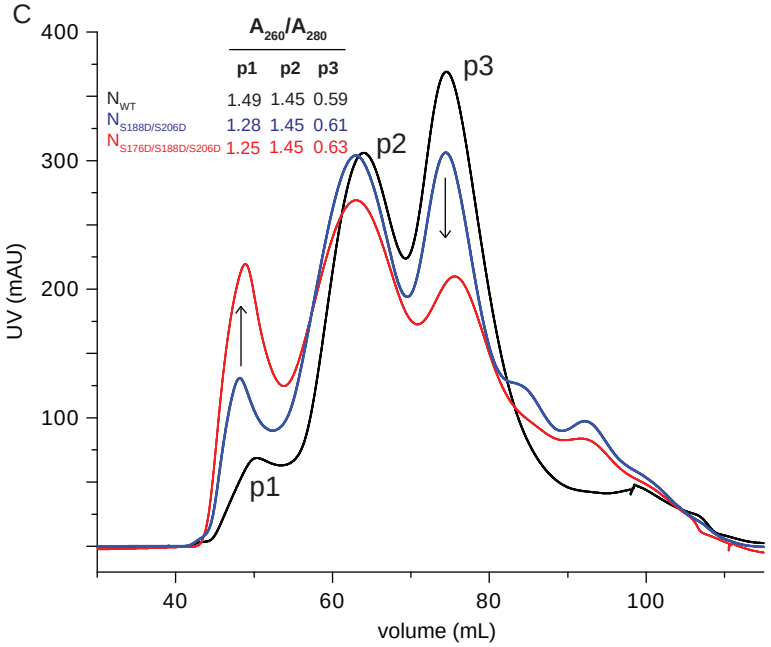
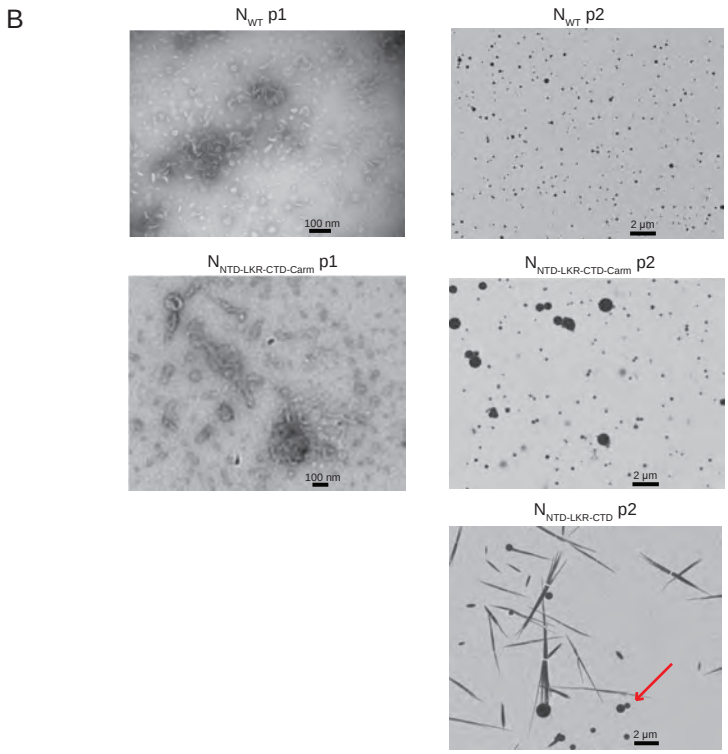
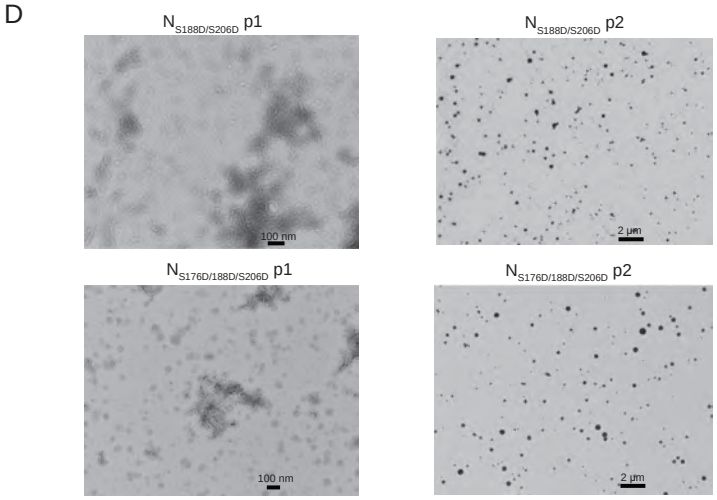
	N_{WT}	$N_{NTD\text{-}LKR\text{-}CTD\text{-}Carn}$	$N_{NTD\text{-}LKR\text{-}CTD}$	$N_{NTD\text{-}LKR}$	N_{NTD}	N_{CTD}
	0.007 ± 0.001	0.006 ± 0.002	0.006 ± 0.002	0.50 ± 0.08	20 ± 10	13 ± 5
	0.051 ± 0.004	0.10 ± 0.01	0.22 ± 0.02	2.5 ± 0.5	120 ± 80	60 ± 40

Figure 3. N-RNA phase separates and phosphorylation modulates N-RNA









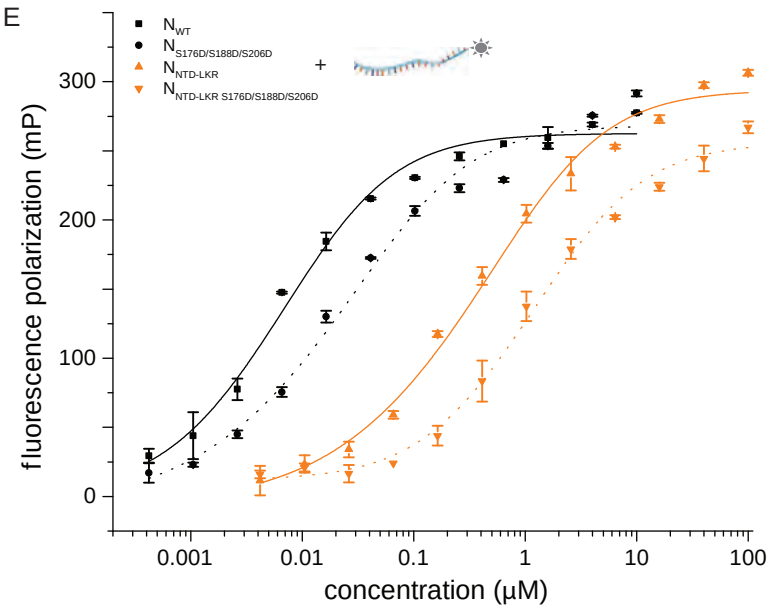


Figure 4. HDX-MS mapping of RNA binding interface within NTD1/CP1276/S188D/S206D*

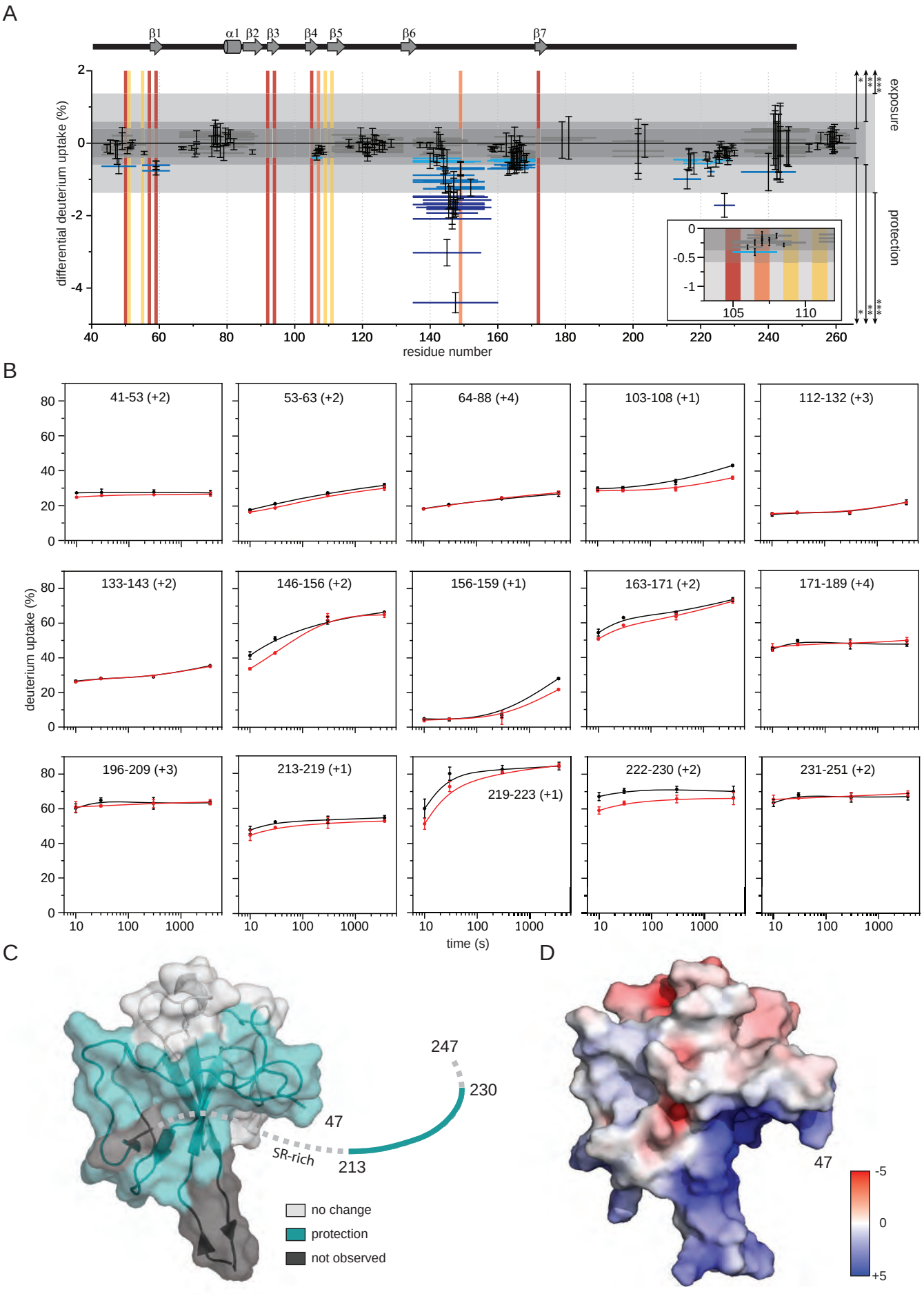
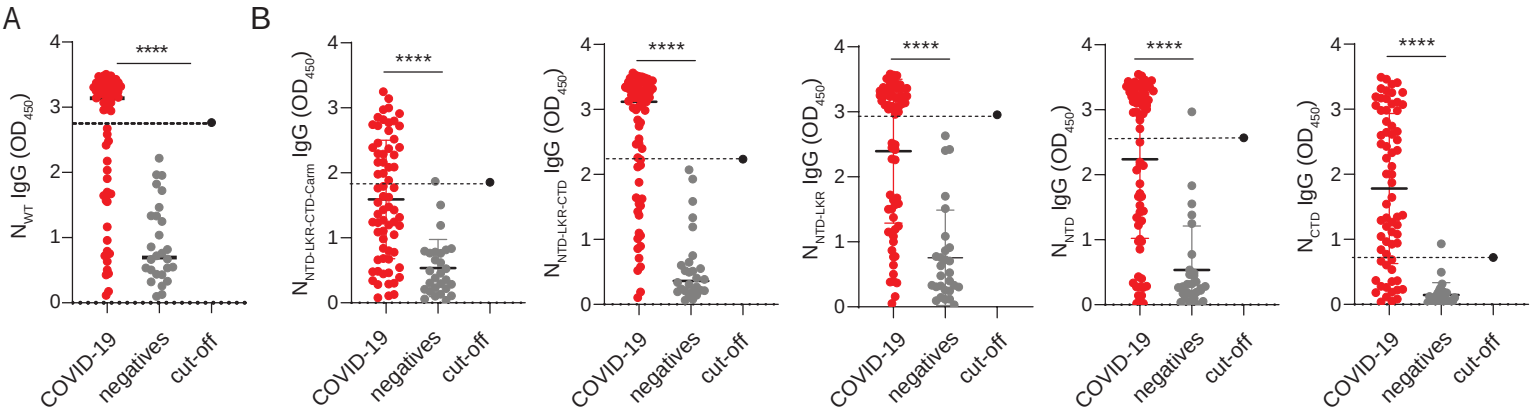
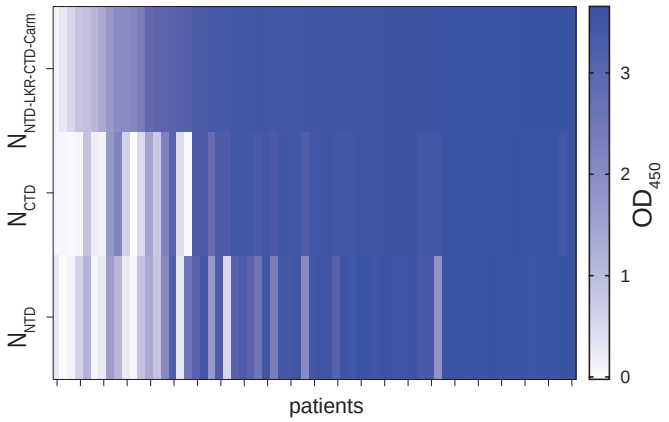


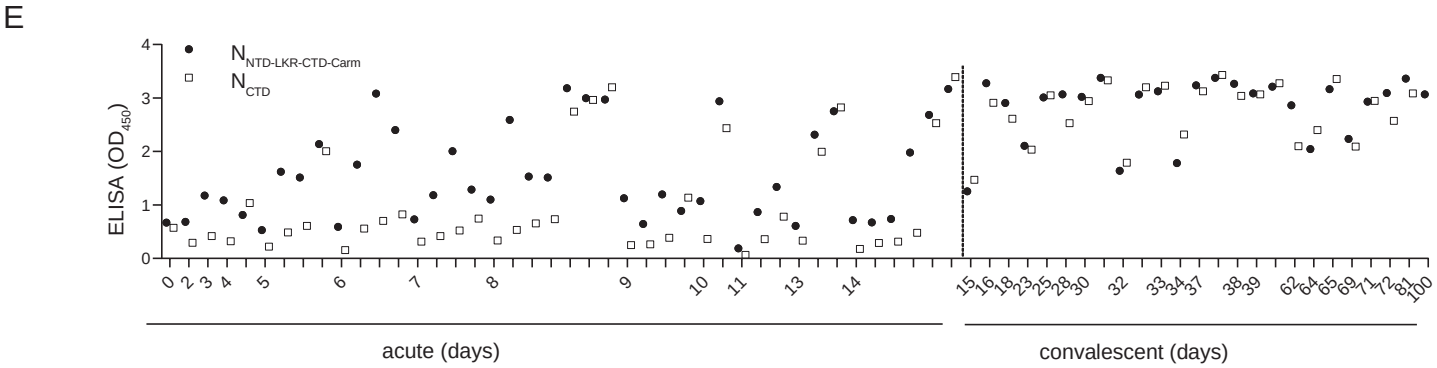
Figure 5. The CTD of N is a highly sensitive serological marker

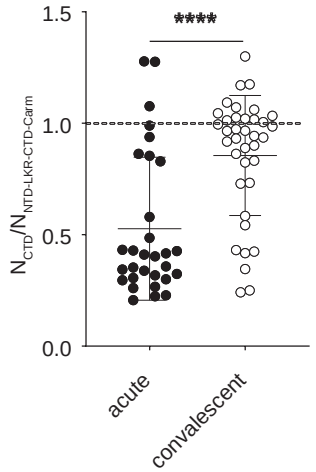


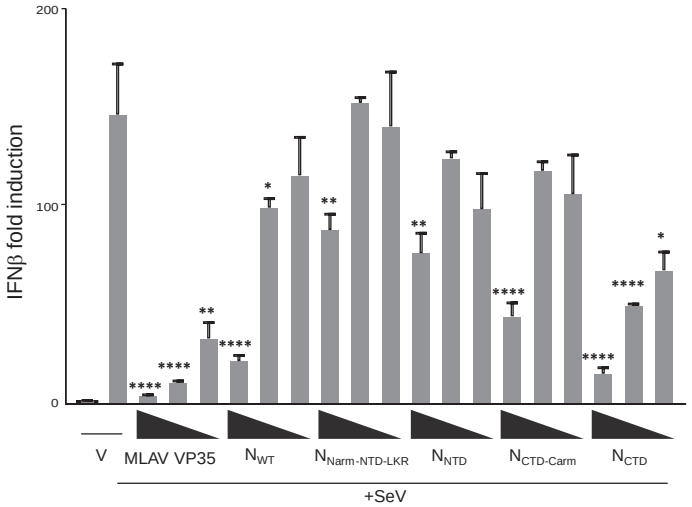
C

	sensitivity (%)	specificity (%)
N_{WT}	66.2	100
$N_{NTD-LKR-CTD-Carm}$	41.2	96.4
$N_{NTD-LKR-CTD}$	70.6	100
$N_{NTD-LKR}$	54.4	100
N_{NTD}	54.4	96.4
N_{CTD}	75.0	96.4



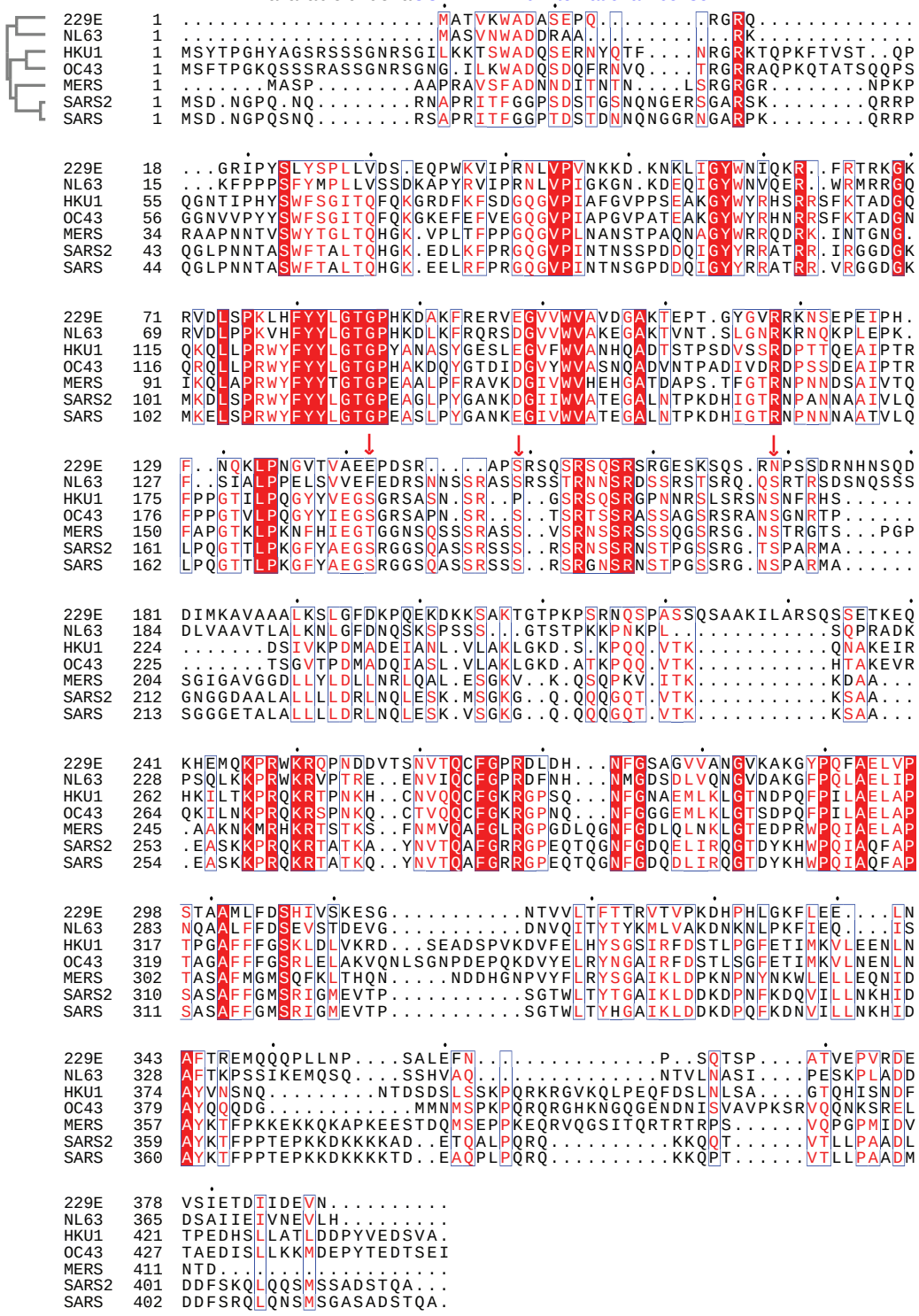






Supplementary Figure 1. related to figure 1. Multiple sequence alignment of coronavirus nucleocapsids.

A



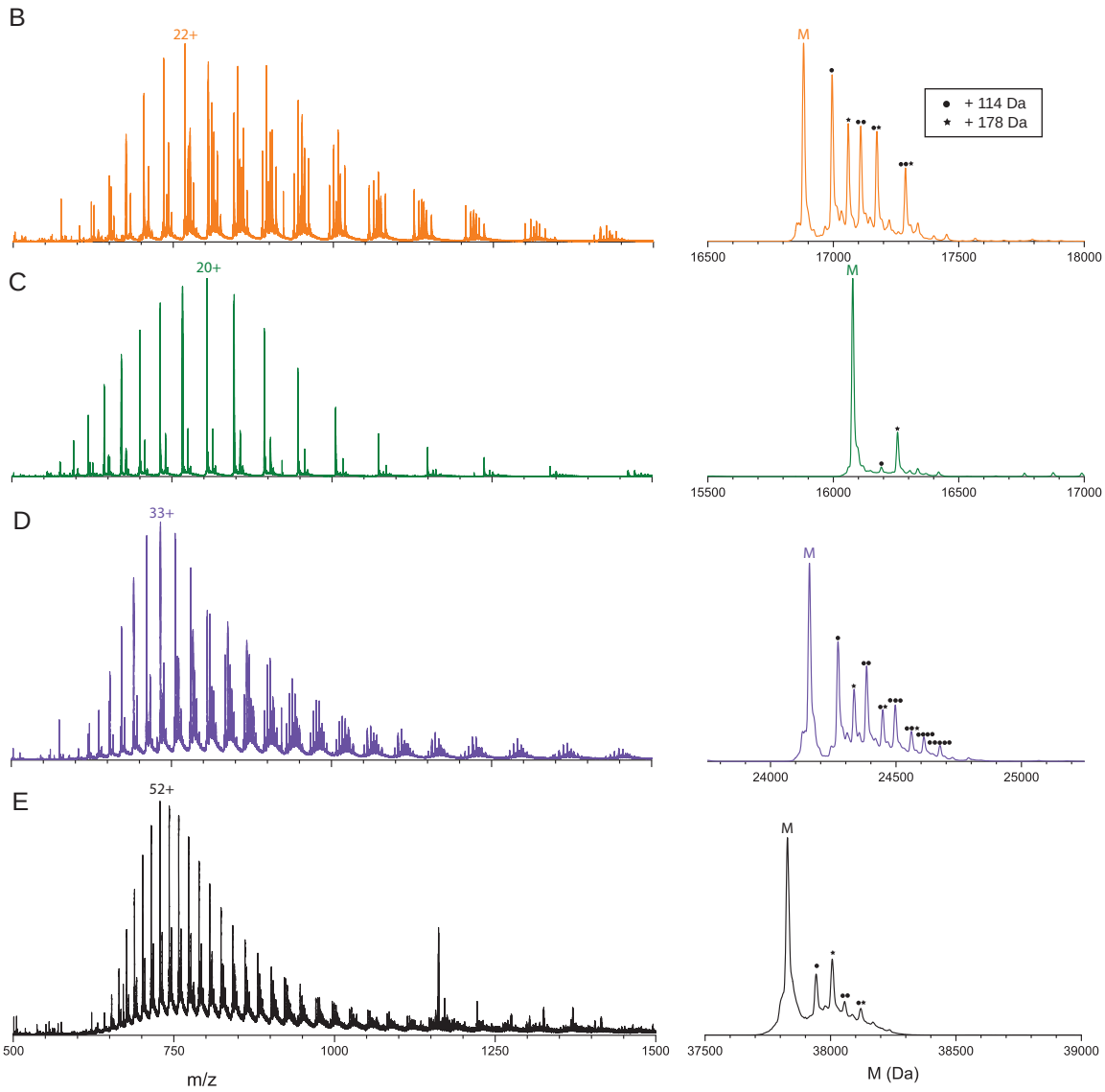
B

	229E	NL63	HK1	OC43	MERS	SARS
FL	26	28	34	36	49	90
Narm	12	0	17	29	22	84
NTD	32	30	44	43	60	92
LKR	26	36	31	34	46	89
CTD	30	28	36	37	54	96
Carm	10	21	15	18	14	74

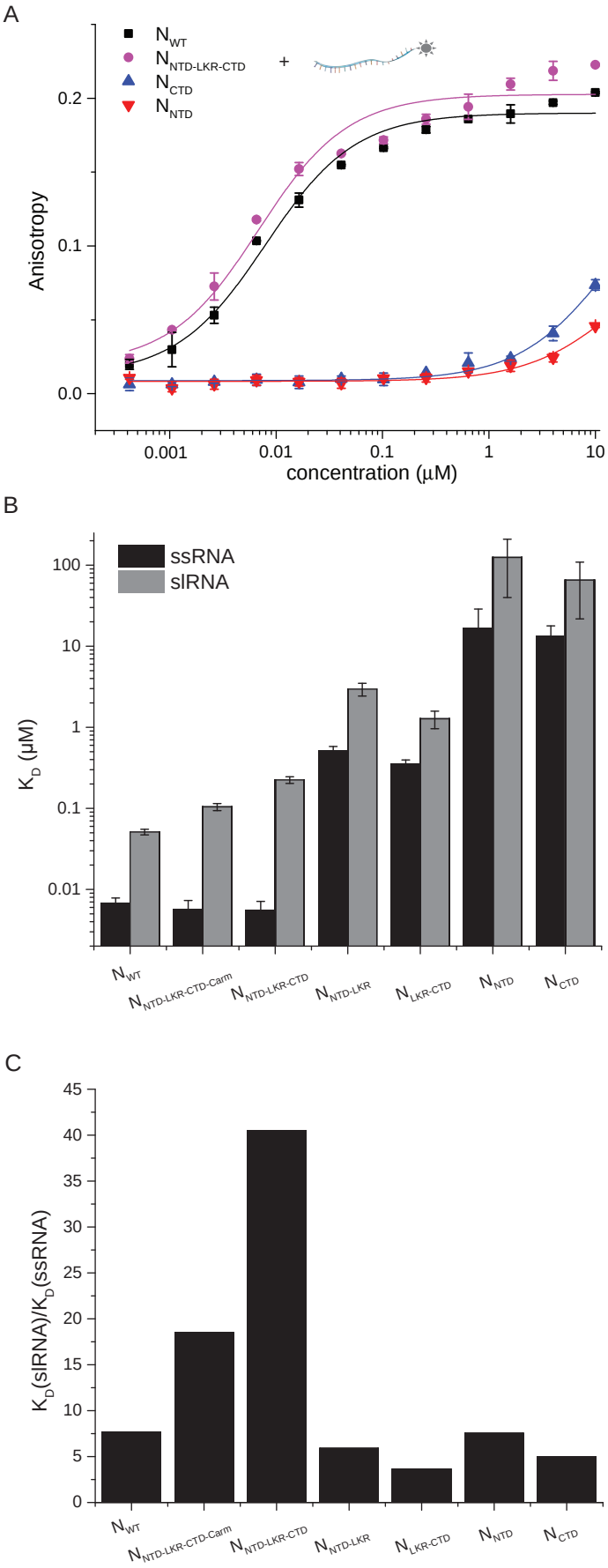
Supplementary Figure 2, related to Figure 1. Denaturing mass spectra of N protein truncations.

A

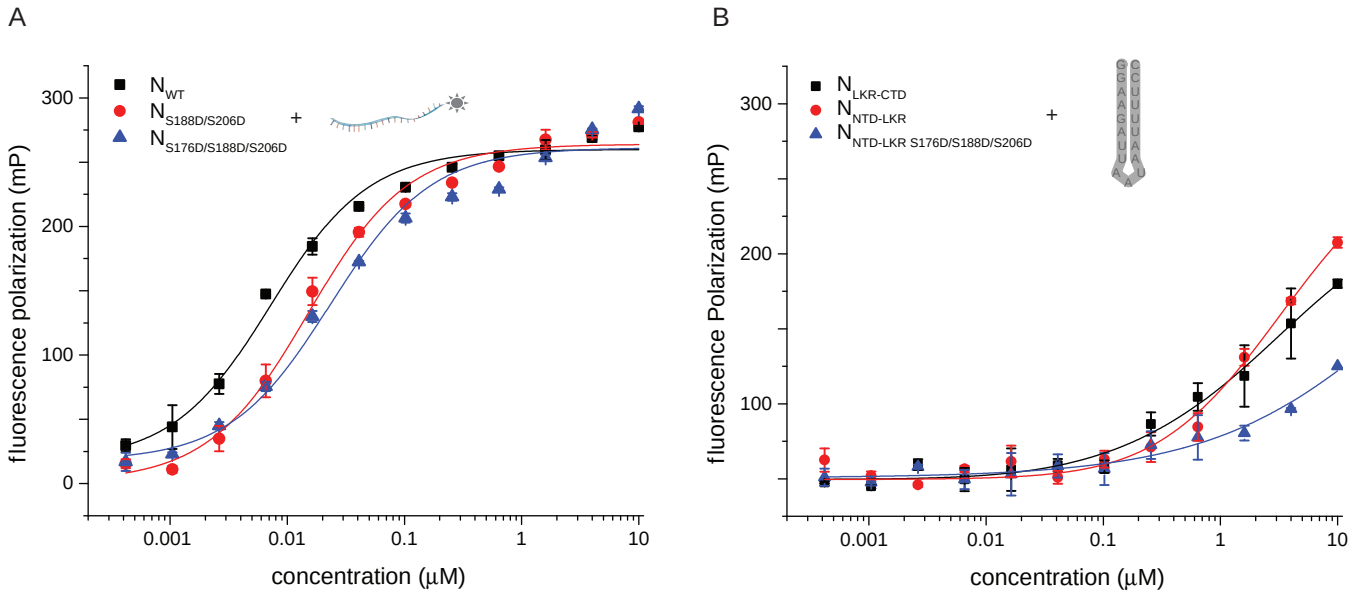
	MW (kDa)	polydispersity (%)	
		peak 1	peak 2
N_{WT}	46	20 ± 1	40 ± 20
$N_{NTD-LKR-CTD}$	35	17 ± 7	30 ± 20
$N_{NTD-LKR}$	22	23 ± 3	
N_{NTD}	14	7 ± 2	
N_{CTD}	14	15 ± 1	



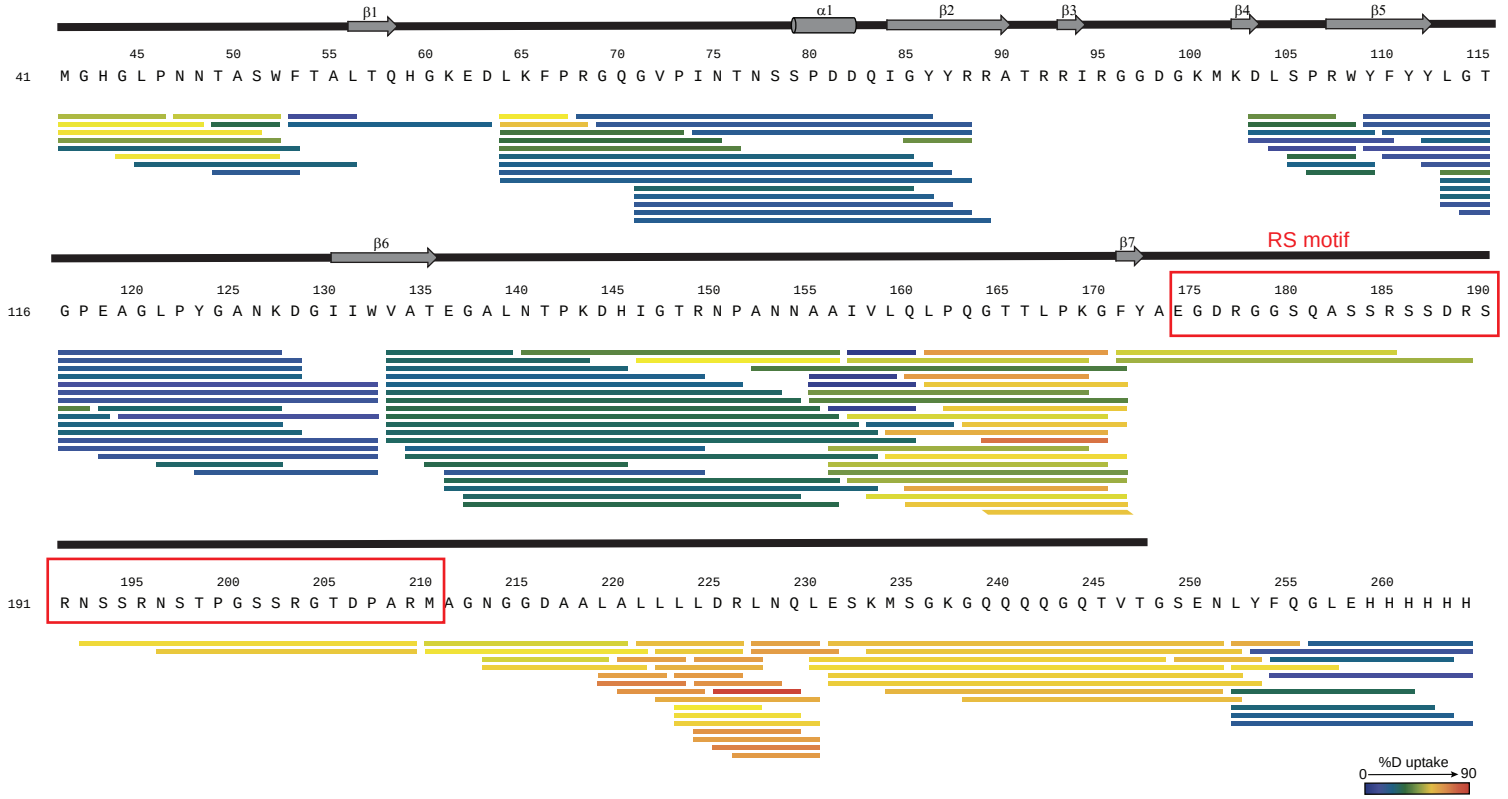
Supplementary Figure 3, related to Figure 3. Nucleocapsid binds stem-loop RNA with reduced affinity.



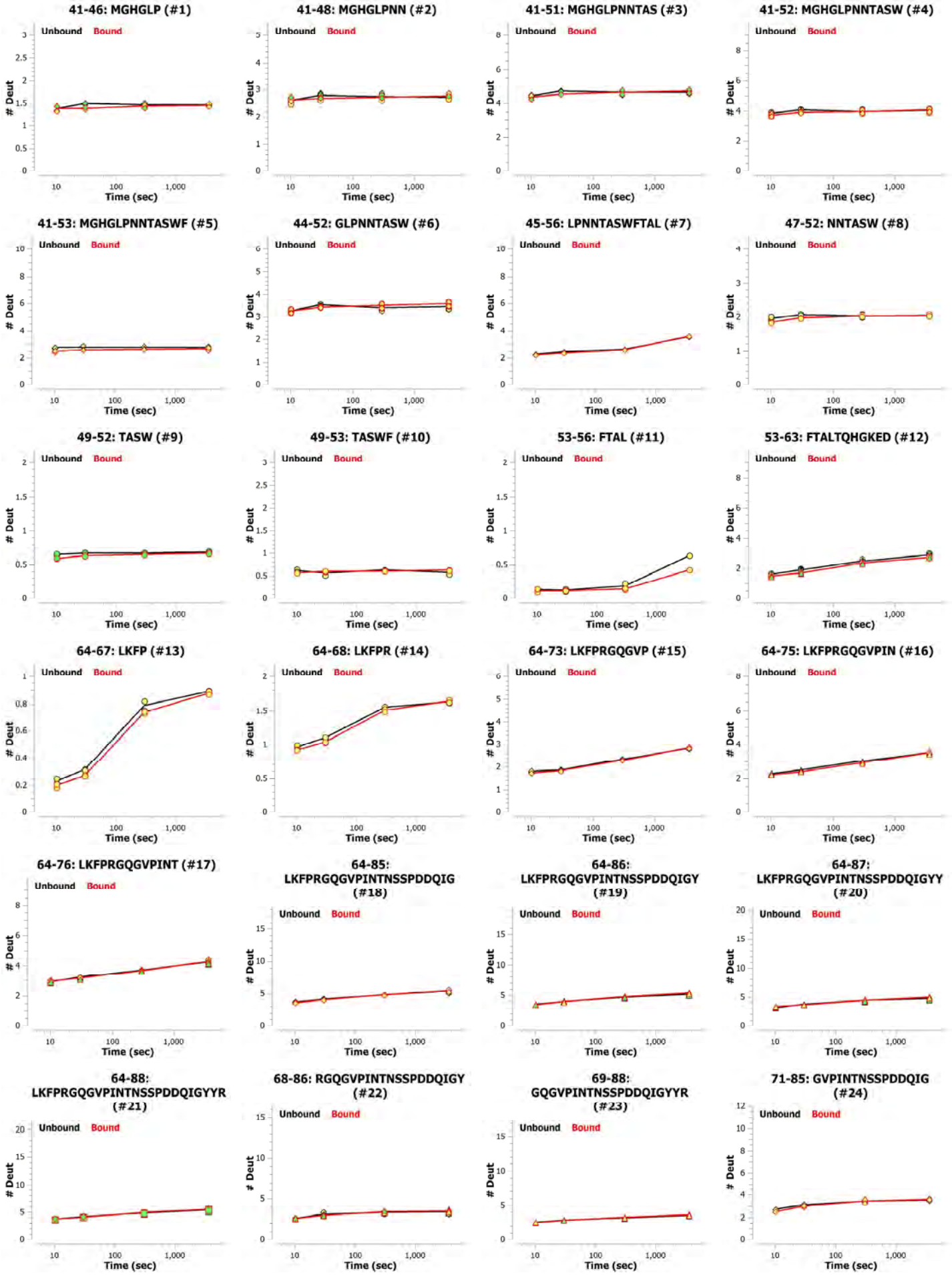
Supplementary Figure 4, related to Figure 4. Phosphorylation mimetics of N reduces RNA binding.



Supplementary Figure 5A, related to Figure 4. Sequence coverage of NTD-LKR S176D/S188D/S206D in HDX-MS.

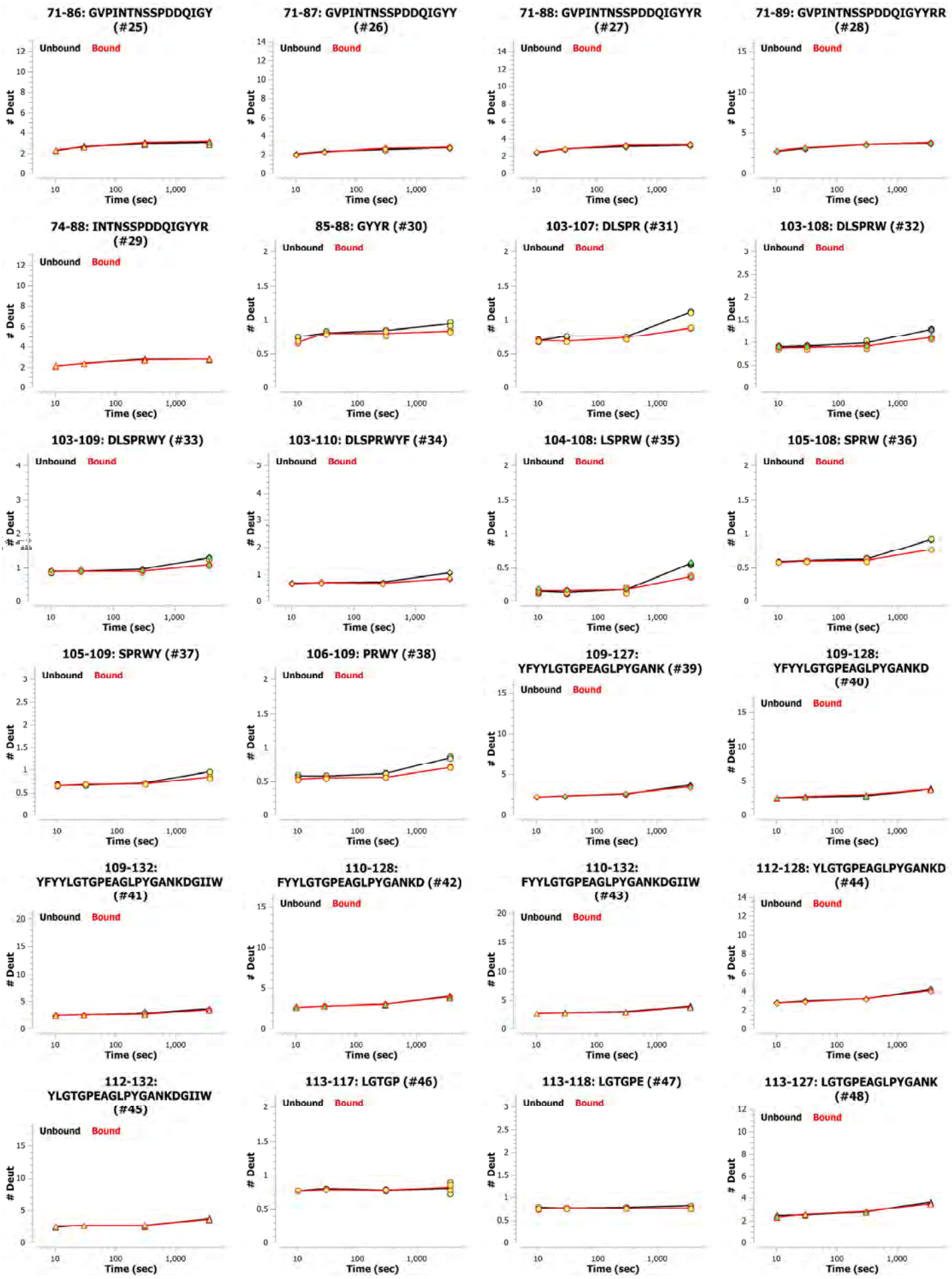


Supplementary Figure 5B-1, related to Figure 4. Sequence coverage of N-NTD-LKR S176D/S188D/S206D in HDX-MS



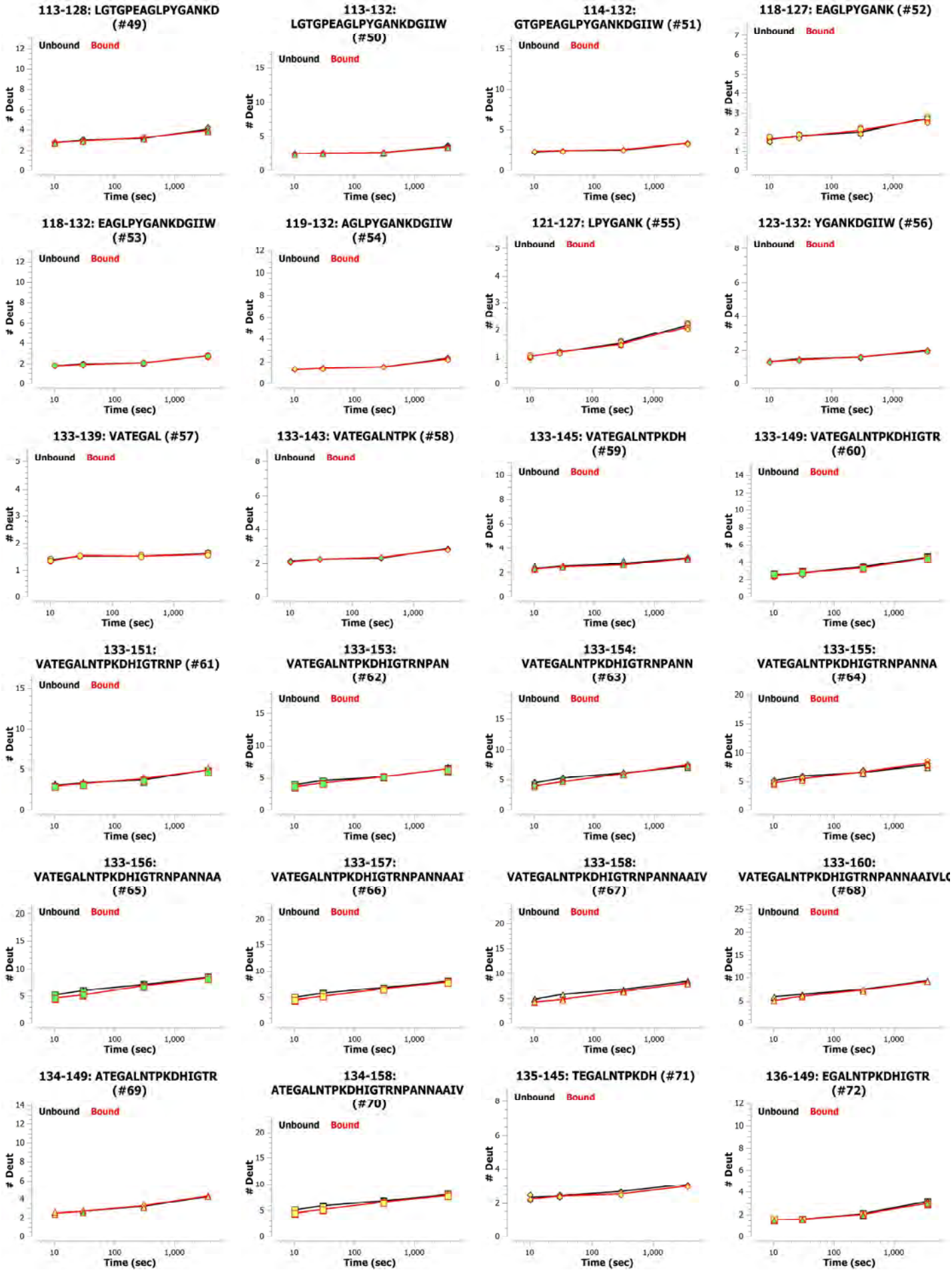
Supplementary Figure 5B-2, related to Figure 5. Sequence coverage of N in HDX-MS

NTD-LKR S176D/S188D/S206D



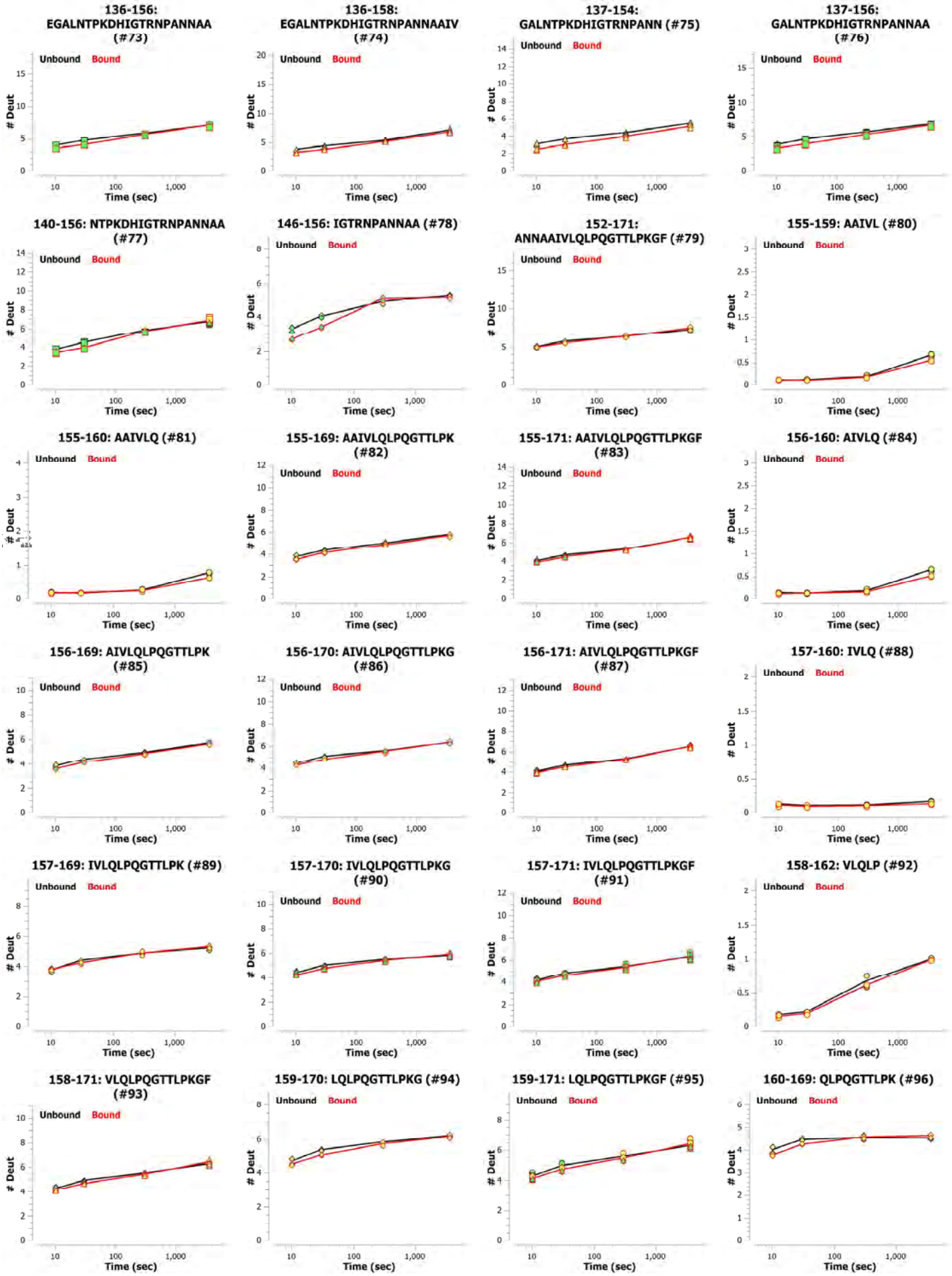
Supplementary Figure 5B-3, related to Figure 5. Sequence coverage of N- in HDX-MS

NTD-LKR S176D/S188D/S206D



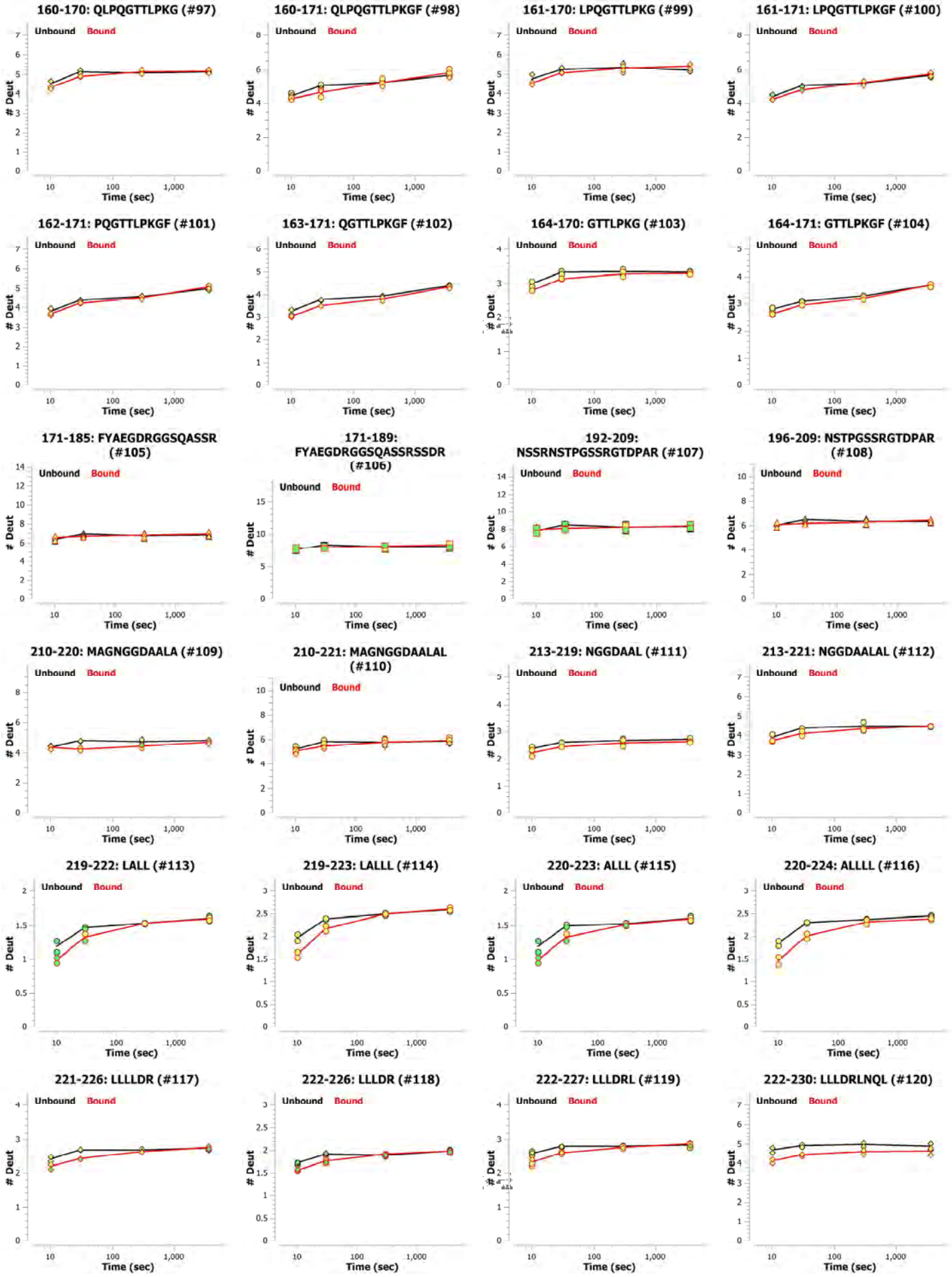
Supplementary Figure 5B-4, related to Figure 5. Sequence coverage of N- in HDX-MS

NTD-LKR S176D/S188D/S206D



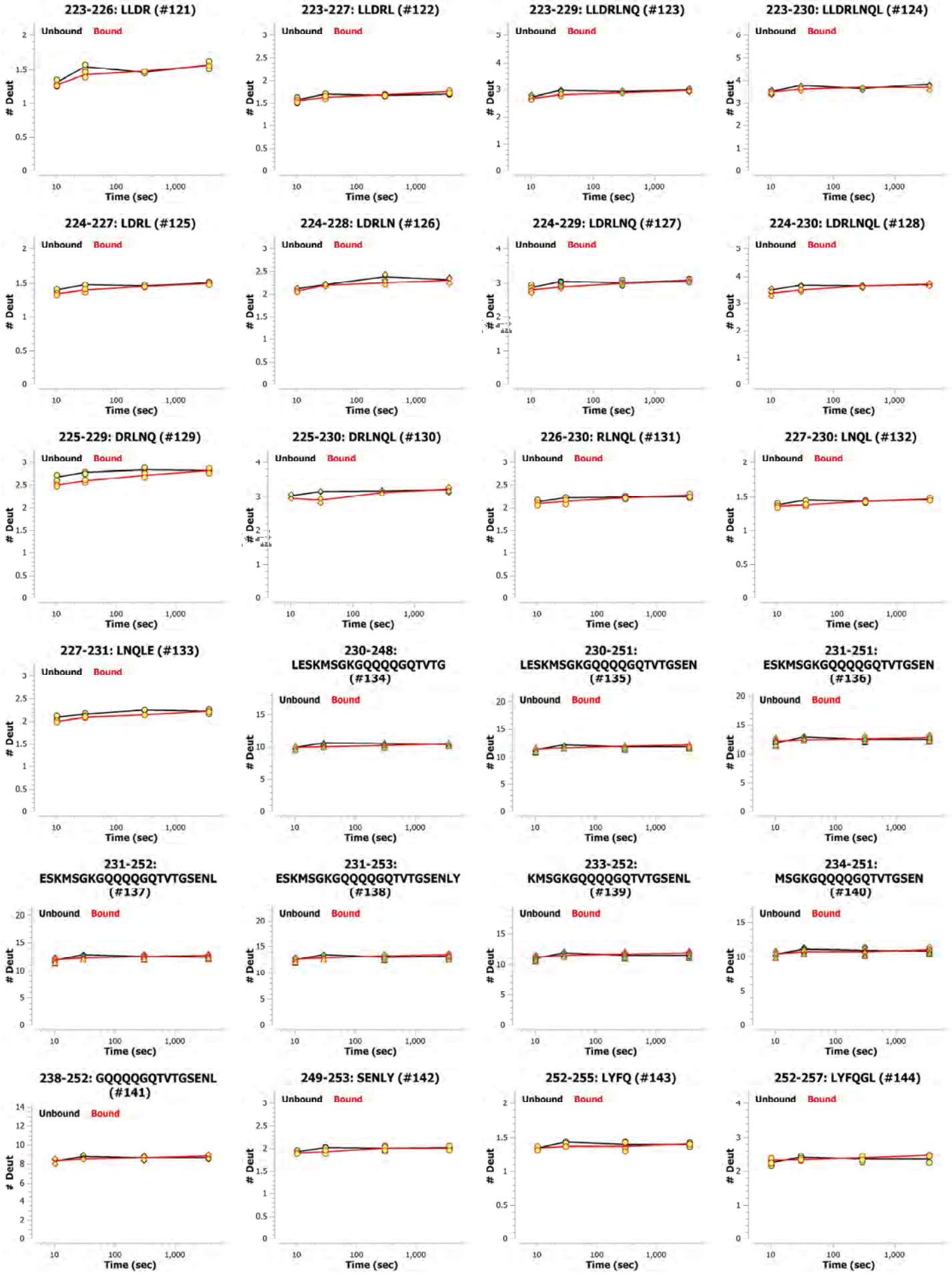
Supplementary Figure 5B-5, related to Figure 5. Sequence coverage of N- in HDX-MS

NTD-LKR S176D/S188D/S206D



Supplementary Figure 5B-6, related to Figure 5. Sequence coverage of N- in HDX-MS

NTD-LKR S176D/S188D/S206D



Supplementary Figure 5B-7, related to Figure 5. Sequence coverage of N- in HDX-MS NTD-LKR S176D/S188D/S206D