Evolution of a fuzzy ribonucleoprotein complex in viral assembly

Huaying Zhao[1], Tiansheng Li[2], Sergio A. Hassan[3], Ai Nguyen[1], Siddhartha A.K. Datta[1], Guofeng Zhang[4], Camden Trent[1], Agata M. Czaja[1], Di Wu[5], Maria A. Aronova[6], Kin Kui Lai[7], Grzegorz Piszczek[5], Richard D. Leapman[6], Jonathan W. Yewdell[2], Peter Schuck[1*]

[1] Laboratory of Dynamics of Macromolecular Assembly, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

[2] Cellular Biology Section, Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[3] Bioinformatics and Computational Biosciences Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[4] Electron Microscopy Unit, Trans-NIH Shared Resource on Biomedical Engineering and Physical Science, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

[5] Biophysics Core Facility, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

[6] Laboratory of Cellular Imaging and Macromolecular Biophysics, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

[7] HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[*]Correspondence to Peter Schuck: schuckp@mail.nih.gov

Abstract

SARS-CoV-2 assembly entails condensation of viral RNA by the viral nucleocapsid (N) protein into ribonucleoprotein particles (RNPs). Lacking high-resolution structural information, biochemical and biophysical approaches have revealed their architectural principles, which involve cooperative interactions of several protein-protein and protein-RNA interfaces, initiated through oligomerization of conserved transient helices in the central disordered linker of N. Here we study the impact of defining N-protein mutations in variants of concern on RNP formation, using biophysical tools, a virus-like particle assay, and reverse genetics experiments. We find convergent evolution in independent introduction of amino acid substitutions strengthening existing binding interfaces. Furthermore, N:P13L of Omicron variants creates a self-association interface *de novo*, enhancing RNP assembly and increasing viral fitness. We hypothesize that the formation of polydisperse, fuzzy N-RNA clusters with distributed weak binding interfaces optimizes reversible RNA condensation while allowing for a large sequence space to be explored to support host adaptation and evolution.

Introduction

Intrinsically disordered proteins are ubiquitous and play key role in many dynamic cellular processes, including cellular signaling, transcriptional regulation, as well as spatio-temporal organization(*1*, *2*). Intrinsic disorder is particularly prevalent in proteins of RNA viruses, for reasons deeply connected to evolution in presence of their high mutation rates and quasispecies nature (*3–6*) .

On a physicochemical level, disorder supports multi-functionality through modulation of the conformational ensemble dependent on ligands and post-translational modifications (*7–13*) and the local environment (*14*, *15*), the possibility of transient folding (*16–18*), and the propensity for protein clustering and liquid-liquid phase separation (LLPS) (*12*, *19–23*). Disorder also permits a significant degree of sequence variation, as many functionally important features are not dependent on the specific sequence but are encoded in non-local biophysical properties(*24–26*). Opposite to the achievement of mutational tolerance through high stability, such as exhibited, for example, by thermostable enzymes, mutational tolerance can be achieved in loosely packed and disordered proteins by lowering the potentially deleterious effect of mutations on stability (*27*). In addition, disorder and flexibility magnifies potential impact of mutations on conformation (*25*). Furthermore, the high degree of sequence variability supports the exploration of short linear interaction motifs (SLiMs) in intrinsically disordered regions (IDRs) for interfacing with a large variety of eukaryotic regulatory and signaling processes, thus favoring adaptability and evolvability on the level of the host-virus interface (*28–31*).

In recent years it has become apparent that assemblies of disordered proteins frequently retain significant disorder and conformational flexibility of their interaction partners, generating “fuzzy” complexes (*3*, *32*, *33*). The potential functional advantages of intrinsic disorder in protein interactions include leveraging of weak binding interfaces through allovalency and the adaptability to multiple binding partners (*33*, *34*). In fuzzy complexes the total binding energy is distributed into multiple distinct ultra-weak interaction sites(*34*) . Similar to individual RNA virus proteins with loose or absent structure, maintaining disorder and a spatial distribution of low-energy interactions in the protein complexes may increase the tolerance for mutations and improve evolvability of protein complexes.

The unprecedented worldwide sequencing effort of SARS-CoV-2 genomes during its rapid evolution in humans provides a unique opportunity to examine these concepts. The genomic database now exceeds in size that of any other virus by orders of magnitude(*35*, *36*), providing the basis for phylogenetic analyses and for monitoring the emergence of variants of concern (VOCs) and their geographic spread (*37*). In addition, it exhaustively samples the mutational landscapes of amino acids that can occupy any position of the viral proteins, which reflects their biophysical constraints (*18*, *25*, *38*). For example, the SARS-CoV-2 nucleocapsid (N-)protein – the most abundant viral protein in infected cells and the focus of the present work – has 419 positions, 86% of which can be assumed by on average 4-5 different amino acids, and up to 12 in positions in the three IDRs(*18*) (**Figure 1A**), evidently without fatally compromising dozens of reported N-protein functions (*39*). Observed N-protein mutations can modulate basic biophysical properties including its oligomeric state, thermodynamic stability of its two folded domains, LLPS propensity, charge distributions, and secondary structure content(*18*, *25*), as well as its interactions with RNA (*40*, *41*), kinases (*9*, *42*) and other host proteins through altered SLiMs (*28*, *43–45*).

In the present work we study the eponymous function of N-protein, which is the spatial condensation of the long genomic RNA (gRNA) into ribonucleoprotein particles (RNPs) for viral assembly. N-protein has two folded domains (the nucleic acid binding domain and the C-terminal dimerization domain, NTD and CTD, respectively), and three IDRs that include the linker between NTD and CTD, and the N-terminal and C-terminal extensions, N-arm and C-arm, respectively (**Figure 1A**) (*46*). The IDRs amount to ≈45% , which renders N-protein highly flexible with a radius of gyration fluctuating from 5 nm to >8 nm (*47*). RNPs are ≈15 nm diameter particles composed of ≈10-15 N-proteins that bind stretches of gRNA (*48*, *49*). 30-40 RNPs are distributed like beads on a string in the ≈100 nm virion, preferentially near the membrane (*48*, *49*). RNPs appear highly heterogeneous in electron microscopy (EM) (*10*, *49*), and as of now, a high-resolution structure has not been determined. The Morgan laboratory has described an *in vitro* model for the assembly of RNPs, where N-protein in the presence of stem-loops of the 5'-UTR RNA readily forms polymorphic ribonucleoprotein complexes that match in size, symmetry, and RNA content what would be expected for the assembled RNPs in virions (*10*, *13*). In conjunction with biophysical experiments and point mutations exposing essential binding interfaces, this has allowed us recently to develop a coarse-grained structural model of RNPs (*50*).

To examine how architecture and energetics of RNP assemblies can be impacted by N-protein mutations we study a panel of N-proteins derived from ancestral Wuhan-Hu-1 and different VOCs, including Alpha, Delta, Lambda, and Omicron (see **Table 1**), in biophysical experiments, VLP assays, and mutant virus. Even though it is thought that mutations affecting efficiency of viral assembly are contributing to viral fitness and evolution(*9*, *18*, *51*, *52*) , mutations in N-protein appear to be secondary to defining mutations in the viral spike protein, and sequential VOCs can differ significantly in their accompanying set of obligatory N-protein mutations. Nonetheless, we found SARS-CoV-2 has independently evolved different strategies for enhanced stability of RNPs in convergent evolution. As is characteristic for fuzzy complexes, significant disorder is retained in the RNPs, in addition to considerable polydispersity and the distribution of total binding energy into a series of weak protein interfaces that act in concert. It appears this design has facilitated modulation, where some interfaces have been repeatedly strengthened in different ways, and others have emerged *de novo* as a result of N-protein mutations in VOCs. We believe the latter displays yet another dimension of plasticity of RNA virus proteins that extends to the architectural principles of critical complexes.

Results

Essential protein-protein interfaces in RNP assembly

Based on biophysical experiments of protein-protein and protein-NA interfaces of full-length (FL) N-protein and domain constructs, in combination with the *in vitro* RNP assembly assay(*10*, *13*) we have recently proposed a coarse-grained structural model of RNPs that satisfies all known binding interfaces. These include at least two known sites of protein-protein and two for protein-nucleic acid interactions distributed throughout most of the protein (*50*) (**Figure 1B/C**). N-proteins are constitutive dimers with K_D in the low nM range by virtue of domain-swapped beta hairpins in the CTD (*46*, *53–56*). N dimers exhibit only ultra-weak further self-association with ≈1 mM K_D (*18*, *53*, *56*) (although NA impurities can artificially scaffold N-protein into higher oligomers (*13*, *57*)). However, higher-order assembly is initiated by the occupation of the nucleic acid binding site in the NTD, which allosterically promotes a helical conformation of the leucine-rich sequence (LRS) in the intrinsically disordered linker, allowing LRS helices to form promiscuous coiled-coil oligomers stabilized through hydrophobic interactions with low μM K_D (*58*). As demonstrated with the LRS point mutant N:L222P that abrogates these transient helices, LRS oligomerization forms the basis for oligomerization of N-protein dimers into RNPs, for example, dodecamers as hexamers of dimers. These oligomers are stabilized by multi-site interactions with double-stranded RNA simultaneously in two sites of the CTD dimer, and by binding preferentially single-stranded RNA to each of the NTDs (*14*, *59–61*), with both RNA interactions allowing inter-dimer interactions that crosslink N-protein dimers within the RNP (**Figure 1C**)(*50*). Additional stabilizing protein-protein interfaces may exist in the C-arm (*10*, *62*), although we have been unable to detect self-association of the C-arm by analytical ultracentrifugation up to low mM concentrations (*58*).

As suggested in the cartoon of **Figure 1C**, this supports a three-dimensional arrangement with a central LRS oligomer with symmetry properties and dimensions similar to low resolution EM images of model RNPs (*10*, *13*) and cryo-ET of RNPs in virions (*48*, *49*). It should be noted, however, that the arrangement sketched in **Figure 1C** is not unique and other subunit orientations could be envisioned that satisfy all constraints from experimentally observed binding interfaces. Extending previous ColabFold structural predictions that show multiple N-protein dimers self-assembled *via* the LRS coiled-coils(*58*), we attempted the AF3 modeling of RNPs combining multiple N dimers with SL7 RNA ligands, mimicking our biophysical assembly model. Current AF3 restrictions limit the prediction to pentamers of N-protein dimers with 10 copies of SL7 RNA. While only inconsistent results were obtained – which is not surprising given the large intrinsically disordered regions -- some models did produce an overall RNP organization similar to **Figure 1C**, suggesting such an arrangement is at least sterically reasonable with regard to possible N-protein subunit orientations in an RNP (**Supplementary Figure S1**).

Using the *in vitro* RNP assembly model of N-protein in mixtures with stem-loop RNA SL7 we measured secondary structure content in circular dichroism (CD) experiments (**Figure 1D**). While a significant structural change in the RNA when bound to N-protein can be deduced in the near UV, after correction for RNA contributions, in the far-UV the CD spectrum of the RNP mixtures is very similar to that of N-protein alone. A small gain in helicity can be discerned for RNP assembly mixtures at ≈220 nm, consistent with the expected coil transition in the LRS stabilizing the RNP. Similarly, small increases in

helicity can be discerned for LRS-stabilizing cysteine mutants. It is noteworthy that the CD spectrum of the RNP mixture appears to be dominated by disordered chains. Although their signature minimum ellipticity at ≈200 nm cannot be observed directly in the high salt buffer required for RNP assembly, the data are quantitatively consistent in their strongly decreasing slope from 205 – 210 nm with the previously measured disordered signature of N-protein in low-salt buffer (without RNP formation). This result is consistent with the absence of significant structure formation in the RNP, and with N-proteins IDRs retaining most of their disorder outside the LRS.

As is depicted in the diagram **Figure 1B**, assembly of approximately 6 N-protein dimers and ≈500bp of RNA requires simultaneous binding at multiple protein-protein and protein-RNA interfaces. Many of these interactions are weak, but they are multivalent and act cooperatively (*50*). This feature may aid in dynamic assembly and disassembly, and shape the ensemble of complex states to efficiently populate functioning RNPs. As we describe in the following, mutations of N-protein have led to diverse mechanisms modulating and promoting RNP formation through their effect on protein-protein interfaces relevant to RNP stability.

A novel self-association interface through transient β-sheets enabled by the P13L mutation

In early epidemiological analyses, the N:P13L mutation has been identified as the most important driver for N-protein fitness, and it has become an obligatory mutation of all Omicron variants (*63*, *64*). In our recent survey of biophysical effects of N-protein mutations, we observed a distinct ability of the N-arm mutant peptide $N_{1\text{-}43}$:P13L to form large assemblies at ≈mM concentrations (which is not seen with the ancestral peptide), while FL N:P13L exhibited enhancement of LLPS (*25*), both indicative of weak interactions. Furthermore, after prolonged storage of P13L N-arm peptide solutions at ≈mM concentrations at 4 °C, an increase in the solution viscosity was observed (*25*). In the present work we studied these effects of the P13L mutation in more detail. As shown in **Figure 2A**, negative-stained EM images show the formation of fibrils of N-arm peptides of Omicron variant $N_{1\text{-}43}$:P13L,Δ31-33. Similarly, fibril formation was observed for $N_{1\text{-}43}$:P13L N-arm peptide lacking the deletion Δ31-33, but not in controls with ancestral N-arm, the ancestral N-arm carrying only the deletion $N_{1\text{-}43}$:Δ31-33, or the disordered ancestral C-arm (**Supplementary Figure S2**). Thus, the N-arm mutation P13L is responsible for the formation of fibrils. Some of these N-arm fibrils exhibit a twisted morphology with width of ≈5 nm (**Figure 2A**), in some instances exhibiting patterns of strand breaks. Such fibrils are frequently encountered in proteins that can stack β-sheets, such as in amyloids(*65*).

To test the formation of β-sheet structure in N:P13L we carried out CD experiments of ancestral N-arm $N_{1\text{-}43}$, and mutants carrying the P13L mutation, or Δ31-33, or both in the Omicron variant $N_{o,1\text{-}43}$:P13L,Δ31-33 (**Figure 2B**). While the ancestral N-arm exhibits CD spectra with a minimum at ≈200 nm typical of disordered conformations (black), the Omicron N-arm has a significantly higher structure content (magenta), consistent with β-sheets, as revealed by the strong negative mean residue ellipticity in the 210-220 nm range. Interestingly, raising the concentration from 0.4 mM (dashed) to 1 mM (solid) led to a large shift in the far-UV spectra from negative to positive ellipticity at ≈200 nm, as well as a shift

in the minimum to higher wavelengths, both indicative of a significant reduction of disorder. This is consistent with the stabilization of β-sheets in a strongly cooperative self-association process with an effective K_D in the high μM to low mM range. Dissecting the origin of the increased β-sheet content, it is apparent that the majority of the effect arises from the P13L mutation (red) alone, with a minor contribution by Δ31-33.

Finally, confirming the interpretation of the EM images and the CD data, the structural prediction of N:P13L in ColabFold displayed oligomers with stacking β-sheets of residues 12-18 with typical hydrogen bond patterns (**Figure 2C**), whereas ancestral peptides did not lead to well-organized structures (**Supplementary Figure S3**). Analogous predictions of N-arms with two other frequent mutations, P13S and P13T, did not lead to β-sheet structures as in P13L.

<u>Enhanced oligomerization of the leucine-rich sequence through cysteine mutations</u>

The protein-protein interaction interfaces driving LRS self-association have been studied in great detail in biophysical experiments and MD simulations for the ancestral molecule and several mutations in the mutational landscape (*18*). A key feature is a pattern of hydrophobic residues that can combine to form a transient helix creating a hydrophobic surface stretching from ≈222 to 234 on one side of the helix. The helical propensity has previously also been observed by NMR, and the monomeric helix has an additional role in mediating binding to NSP3 by docking into a hydrophobic binding groove of the Ubl1 domain of NSP3 (*16*).

We have previously discovered the self-association of the LRS after analysis of the mutational landscape (*18*) and reports of a conspicuous defining mutation N:G215C in Delta variants that correlated with its rise in 2021 among clades with identical spike mutations (*18*, *66*, *67*) . As deduced from MD simulations, the cysteine at position 215 is located at the base (N-terminal end) of the transient helix and, through its lower flexibility than the glycine, serves to redirect the adjacent upstream disordered residues such that helices are more prone to form stabilizing coiled-coil interactions (*18*, *58*). As shown experimentally by sedimentation velocity analytical ultracentrifugation (SV-AUC) in reducing conditions, this enhances non-covalent self-association of LRS peptides as well as FL N dimers by 2-3 orders of magnitude (*18*, *58*). Covalent disulfide bonds in the LRS in non-reducing conditions were found to further promote LRS oligomerization. However, there is no conclusive data yet whether covalent bonds in the LRS occur *in vivo*, or any G215C effect is entirely non-covalent due to the significant strengthening of LRS helix oligomerization (see **Discussion**). In any event, the G215C mutation leads to enhanced assembly in a VLP assay (*50*), and as shown in reverse genetics experiments, *in vivo* confers a significant replication advantage and an altered virion morphology (*68*).

Here we studied a mutation of G214, which in the mutational landscape exhibits a similar mutation pattern as G215. In particular, we focus on an independent introduction of a cysteine in the LRS that occurred in the Lambda variant, prevalent in South America in 2020-2021 (*69*), with the defining N-protein mutations P13L, R203K/G204R, and G214C (**Table 1**). The effect of the G214C mutation is unknown. Due to the close proximity of 214 and 215, we asked whether it enhances LRS self-association

similarly to G215C, and to this end first synthetized a LRS peptide comprising $N_{210-246}$:G214C. Unexpectedly, unlike the chemically identical $N_{210-246}$:G215C peptide, it exhibited low solubility. This prohibited its characterization at sufficiently high concentrations for study of self-association by SV-AUC, which required in excess of ≈0.4 mM for $N_{210-246}$:G215C peptides to display oligomers. However, dynamic light scattering (DLS) revealed the presence of $N_{210-246}$:G214C complexes with hydrodynamic radii ranging from 6 to 40 nm (in comparison to 1-2 nm for $N_{210-246}$:G215C(*18*)) in reducing conditions, and slightly larger in non-reducing conditions (**Supplementary Figure S4**). This shows that while G214C also strengthens the LRS self-association interface, it exhibits different properties compared to G215C.

To gain more insight in the different behavior of the cysteine LRS mutants we carried out MD simulations of monomeric and trimeric LRS peptides $N_{210-246}$. As shown in **Figure 3**, both cysteine mutants extend the helix by stabilizing the flexible GG motif into a well-defined α-helix turn. Importantly, however, the sulfhydryl groups in position 214 *vs* 215 assume different orientations relative to the hydrophobic patch serving as the oligomerization interface. Under reducing conditions, this has the potential to also cause variation in physicochemical properties of resulting oligomers, e.g., through repositioning of the adjacent glutamate E216 and resulting changes in the surface electrostatic potential, and to thereby alter the oligomerization scheme (**Supplementary Figure S5**). It can be expected that under oxidative conditions the differences between 214C and 215C are exacerbated due to their different relative orientation of the hydrophobic interface of the helix and the sulfhydryl groups, leading to different oligomeric states and different phase separation properties.

<u>Impact of enhanced self-association interfaces on the oligomerization of full-length N</u>

In principle, since N-protein is a constitutive dimer coupled in the CTD, binding interfaces in the LRS and N-arm may form intra-dimer bridges simply further stabilizing the dimeric state. Indeed, coarse-grained molecular simulations of N-protein dimers have revealed large conformational fluctuations where the LRS, for example, can make rare intra-dimer contacts (*47*). On the other hand, given the large flexibility of the disordered chain, the same interfaces may lead to higher oligomers if they establish one or two inter-dimer bridges across different dimers, generating different classes of tetramers. We would expect a concentration-dependent probability of such inter-dimer contacts for N-protein in solution.

Experimentally, in the absence of NA ligands, ancestral N-protein at low μM concentrations is essentially dimeric with only hints of reversible higher-order oligomers (*53*, *56*). With the strengthened LRS helix stability through the G215C mutation, in reducing conditions, we observe reversible tetramerization with a K_D of 1.0 (0.7 – 1.5) μM (**Figure 4A**), consistent with previous work(*18*). We made analogous observations by SV-AUC for N_λ (containing the G214C mutation) in reducing conditions, demonstrating it similarly can form reversible inter-dimer bonds (**Figure 4A**). Clearly, strengthened LRS helices can alter the N-protein oligomeric state by crosslinking dimers into tetramers.

To study the potential impact of disulfide bonds on N-protein oligomeric state, we prepared oxidized FL N:G215C and N_λ by extensively dialyzing the reduced protein in TCEP-free buffer while purging air

through the dialysate for gentle passive oxygenation. This oxidized sample is referred to as N:G215C* and N_{λ}* . Using a DTNB assay we assessed the percentage of free sulfhydryl groups to be ≈30% for both N:G215C* and N_{λ}*, respectively. Since the LRS mutation provides the sole cysteine in N-protein, we concluded that the majority of N:G215C* and N_{λ}* is disulfide-linked in the LRS, which was confirmed by non-reducing SDS-PAGE (**Supplementary Figure S6**). Due to the non-covalent dimerization in the CTD as well as non-covalent tetramerization *via* LRS interfaces (see above), these samples may be complex mixtures of dimers and/or higher oligomers with different patterns of covalent and non-covalent intra- and inter-dimer LRS interactions. Using SV-AUC we established that close to half of both the N:G215C* and N_{λ}*sample is tetrameric at low μM concentrations (**Figure 4A**). Because oxidation significantly increases the tetramer population, we conclude that one or two covalent inter-dimer bonds can form that enhance N-protein oligomerization.

In contrast to the modulation of the coiled-coil LRS interfaces, the *de novo* creation of the N-arm self-association interface through beta-sheet interactions enabled by P13L cannot be readily observed in FL N-protein at low μM concentrations. Similar to the ancestral LRS interface, it provides only ultra-weak binding energies that require mM concentrations to significantly populate oligomers. This is fully consistent with the previous observation by SV-AUC that neither N_{FL}:P13L,Δ31-33 nor $N_{FL,o}$ with the full set of Omicron mutations show any significant higher-order self-association at low μM concentrations, whereas at high local concentrations – as observed in phase-separated droplets – they can modulate and cooperatively enhance self-association processes (*25*). Another process that increases the local concentration of N-arm chains is the tetramerization of FL N-protein. As described earlier, occupancy of the NA-binding site in the NTD allosterically promotes self-assembly of the LRS into higher oligomers (*53*). We hypothesized that these oligomers may be cooperatively stabilized by additional N-arm interactions in P13L mutants.

To this end, we carried out SV-AUC and DLS experiments of 2 μM ancestral N, and the FL N-arm mutant N:P13L,Δ31-33 in the presence of a short oligonucleotide T_{10} that occupies the NTD binding site but is too short for scaffolding multiple N-protein dimers(*53*) (**Figure 4B**). Low salt buffer (10 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2PO_4 , 1.8 mM KH_2PO_4 , pH 7.4) in this experiment ensures maximal occupancy of the NTD binding site for NA by 10 μM T_{10}, which under these conditions has a K_D below 0.1 μM (*53*). In a control experiment by SV-AUC we measured the affinity of the T_{10} oligonucleotide for ancestral N and N:P13L,Δ31-33 and observed no significant difference in moderate ionic strength (**Supplementary Figure S7**). Ancestral N_{FL} protein in the absence of oligonucleotide sediments at ≈4.0 S, reflecting its dimeric state afforded by the CTD dimerization domain (*53*). When LRS oligomerization is abrogated through introduction of a L222P mutation, binding of T_{10} to the N-protein dimer still induces a conformational change and increases its *s*-value to ≈4.9 S (*58*). With the native LRS in the ancestral N_{FL} this T_{10} -ligated state allows LRS oligomerization, which can be observed through the formation of a reaction boundary in SV-AUC with an *s*-value of ≈5.6 S, reflective of a mixture of dimers and tetramers in rapid exchange relative to the time-scale of sedimentation (*58*, *70*) . As a control, we reproduced this previously reported result (**Figure 4B**, blue). Introduction of the N-arm mutations N:P13L,Δ31-33 causes a further increase of the *s*-value of the reaction boundary to ≈5.8 S, indicating an increase in the tetramer stability (**Figure 4B**, magenta). This stabilization of the tetramer is corroborated independently

by an increase in the average hydrodynamic radius of N:P13L,Δ31-33 in mixture with T_{10} (5.51 nm) relative to ancestral N (5.37 nm) or the LRS mutant N:P13L,Δ31-33,L222P (4.84 nm) (**Figure 4B inset**). Thus, the N-arm mutation clearly strengthens inter-dimer interactions, even though the added binding energy is too weak to produce detectable tetramer populations at micromolar concentrations by itself. In principle, allosteric interactions between the distant disordered N-arm and the LRS in the disordered linker might exist that cause enhanced tetramerization of N:P13L,Δ31-33 with occupied NA-site in the NTD. However, a more parsimonious explanation is that the additional self-association interface in the N-arm created by P13L makes inter-dimer contacts that add to the separate oligomerization of the LRS helices in stabilizing tetramers.

<u>Mutation effects on RNP assembly and stability</u>

As the assembly of RNPs requires the concerted effect of several binding interfaces, we asked whether the enhanced LRS coiled-coil stability and the novel N-arm self-association interface impact the RNP stability. To examine this experimentally we carried out *in vitro* RNP assembly experiments using the assay developed previously by the Morgan laboratory (*10*, *13*). As mentioned above, it is based on the observation that mixtures of N-protein with stem-loop RNA from the viral 5'-UTR at low μM concentrations readily form polymorphic ribonucleoprotein complexes that match in size, symmetry, and RNA content what would be expected for the assembled RNPs in virions (*10*, *13*). We embarked on the experimental roadmap introduced previously(*50*) consisting of SV-AUC experiments that hydrodynamically resolve RNPs in dynamic assembly equilibrium in solution as fast-sedimenting reaction boundaries(*70*, *71*), in combination with complementary mass photometry (MP) experiments that can resolve populations of different protein/RNA complexes and their dissociation products at sub-μM concentrations through interferometric sizing of single-molecule surface adsorption events(*72*).

Sedimentation coefficient distributions of ancestral and different mutant FL N at 3 μM in the presence of 3.4 μM SL7 are shown in **Figure 5A**. For reference, under these conditions the RNPs of the ancestral N-protein form a reaction boundary with a weighted average sedimentation velocity of 19.7 S (black). Under otherwise identical conditions, significantly faster sedimentation can be discerned for N:P13L,Δ31-33 (20.2 S, red). Faster sedimentation can reflect an increase in size of the complexes, and/or increased populations and lifetimes of complexes in dynamic equilibrium, i.e., higher affinity and stability. By contrast, N:R203K/G204R exhibits a slightly lower *s*-value of 19.5 S (blue). Both R203K/G204R and P13L,Δ31-33 combine in the set of defining Omicron mutations, N_o , which produces RNP boundaries at 20.5 S (cyan). It appears the N-arm mutations can more than compensate for the loss of RNP stability through the R203K/G204R mutation in the linker. The introduction of the LRS cysteine in N:G215C is nearly as effective, with an *s*-value of 20.4 S (orange). Finally, the largest increase in sedimentation velocity can be discerned for the combination of P13L/R203K/G204R with the LRS enhancing cysteine G214C in N_λ with 21.0 S (magenta) (see **Table 1**).

When these samples are diluted and equilibrated at tenfold lower concentration, the RNPs are largely dissociated, as may be discerned from the reduced amplitude of the rapidly sedimenting peak and

significant population of species with *s*-values between 7 and 18 S (**Figure 5B**). This highlights the cooperative oligomerization of the N-dimer/2SL7 subunits observed previously(*50*) . Interestingly, the only partial dissociation of N_o , N:G215C, and N_λ RNPs suggests the existence of a subpopulation with higher stability. By contrast, the sedimentation coefficient distribution of RNPs from 0.3 μM N:P13L,Δ31-33 is more similar to that of the ancestral N-protein and N:R203K/G204R, suggesting the augmented RNP population caused by combining these mutations is kinetically not as stable and the contributions from N-arm interfaces are more transient.

Side-by-side to the SV experiments, MP measurements were carried out on the same samples to shed more light on the mass distribution of the RNP particles. For discrimination of individual surface adsorption events on the coverslip, the sample concentrations cannot exceed 0.3 μM protein with 0.34 μM SL7. As shown in **Figure 5C**, this results in a ladder of oligomers of N-dimer/2SL7 subunits ranging from ≈120 kDa (a single N-dimer/2SL7 subunit) to ≈700 kDa (a hexamer of subunits) (*10*, *50*). While this is analogous to the solution data in **Figure 5B**, for the detailed comparison of SV and MP distributions their different signal weights should be considered: where MP counts individual particles producing a number distribution, SV records a signal proportional to mass, and therefore is strongly skewed toward larger particles compared to MP. Also, in a trade-off between resolution and susceptibility to systematic errors, there is a potential for chemical properties to bias the surface adsorption, and for some adventitious experimental variation originating from the glass substrates. Nonetheless, we obtained results qualitatively consistent with the SV-AUC data, where N:P13L is largely dissociated like ancestral N-protein and N:R203K/G204R at 0.3 μM; and at the same time, N_o , N:G215C, and N_λ RNPs retain more of the oligomeric state. Interestingly, N:G215C shows the least fully dissociated dimer and retains more of a pentameric species, which is in contrast to the other cysteine-containing mutant N_λ (**Figure 5C**).

The ladder of oligomeric subunits resolved in MP poses the question whether the largely assembled state at 3 μM is uniform. Under these conditions SV-AUC shows a single RNP reaction boundary, but it is relatively broad, which would be equally consistent with a reaction boundary from a single complex in rapid association/dissociation exchange on the time-scale of sedimentation (< 1000 sec) (*71*), as with a polydisperse mixture of several unresolved large oligomers. In order to gain more insight in the mass distribution under these 3 μM conditions, for some constructs we employed a microfluidic accessory device for the MP instrument allowing rapid dilution with a dead time of < 0.1 sec to minimize RNP dissociation. The resulting data in **Figure 5D** again exhibit a ladder of oligomeric peaks, where the majority of N-protein assembled in a heterogeneous mixture of RNPs between tetramer and hexamer of subunits. Thus, the dissociation products after dilution in **Figure 5C** do not seem to originate from a single assembled complex. Furthermore, we observe characteristic differences in the oligomeric distribution between different constructs. Similar to the equilibrated lower concentration conditions, relative to ancestral N, the hexamer population is augmented for N:P13L,Δ31-33, N:G215C, and N_λ. Notably there is again a prominent pentamer population for N:G215C.

Since the disassembly of RNPs after viral entry is another critical step in the viral life cycle, we aimed to probe the kinetic stability of the RNP complexes. To this end, we applied a modified pipettor-based sample application protocol where rapid dilution of the 3 μM mixture was followed by several

consecutive 20 sec periods of data acquisition. The data from the acquisition immediately following the dilution (acquired between 3 – 23 sec) is shown in **Figure 5E**. It provides mass distributions showing substantial assembly into heterogeneous populations of RNPs. Qualitatively consistent with the 3 μM mixtures in SV-AUC, the N:R203K/G204R has a clear destabilizing effect on the RNP, whereas all mutants with enhanced binding interfaces exhibit higher populations of larger RNPs, with the greatest enhancement for N:P13L,Δ31-33. In order to focus on the kinetics of RNP dissociation we calculated the number average molecular weight of RNPs, which is plotted as a function of decay time in **Figure 5F** and empirically fitted as a single exponential decay attaining the separately measured equilibrium value measured at 0.3 μM (for distributions see **Figure 5C**, for number averages **Table 1**). While N:P13L,Δ31-33 produces the largest increase in RNP molecular weight between 0.3 μM and 3 μM, these RNPs have the shortest lifetime (τ = 44 sec *vs* 66 sec for the ancestral N), suggesting the added H-bond interactions in the N-arm to be rapidly reversible. On the other hand, RNPs of N:G215C have a lower average mass (being dominated by a pentameric oligomer of subunits), but show an increased kinetic stability (τ = 231 sec). RNPs of N_λ carrying both an LRS cysteine and the N-arm mutation are simultaneously of higher average molecular weight and persist most upon dilution (with a best-fit RNP equilibrium level of 669 kDa). (**Figure 5F**).

Finally, we were curious about the impact of disulfide-linked oligomers of the cysteine mutants on their RNP structure and stability, and carried out analogous experiments with the substantially oxidized protein preparations depicted above in **Figure 4A**. As shown in **Figure 6A**, sedimentation coefficient distributions of 3 μM oxidized N:G215C* and oxidized N_λ* (carrying 214C) in assembly mixtures with SL7 show faster reaction boundaries when oxidized as compared to their reduced form, with 20.8 S *vs* 20.4 S for N:G215C, and 21.5 S *vs* 21.0 S for N_λ. Upon dilution the RNPs of oxidized N:G215C* protein dissociate similar as those of reduced N:G215C. For oxidized N_λ RNPs dissociation is less, but still significant. The molecular weight distributions in MP measured in the first 20 sec after rapid dilution of the 3 μM stock (**Figure 6B**) show largely assembled mixtures of tetramers, pentamers, and hexamers of the N-dimer/2SL7 subunits, with a large increase in the population of hexameric RNPs in oxidized *vs* reduced forms. The tail of even higher molecular weight species is enhanced, pointing to subpopulations of heptameric species especially for oxidized N_λ RNPs. Interestingly, the preference of pentameric RNPs formed by reduced N:G215C is absent after oxidation. N_λ* exhibits the largest population of hexamers observed in any sample, in agreement with the highest *s*-value of the reaction boundary in SV. As indicated by the time-course of RNP dissociation after tenfold dilution (**Figure 6C**), oxidation enhances the average molecular weight of the RNPs but not their lifetime. We conclude that disulfide-linked N-protein tetramers can incorporate into and aid in the formation of RNPs, modulating their preferred oligomeric state, and that cumulative effect of the N-arm mutation P13L and the LRS cysteine persists for disulfide-linked protein.

<u>Virus-like particle formation and infectivity</u>

We asked whether the observed modulation of RNP size and stability by N mutations impacts the formation and infectivity of virus-like particles (VLPs) (*50*, *73*). In this assay, producer 293T cells are co-

transfected with plasmids for the four structural proteins S, E, M, and N of SARS-CoV-2, alongside a plasmid containing the viral packaging signal with a luciferase reporter gene. This leads to assembly of VLPs, which are collected from the supernatant and applied to receiver cells transfected with entry factors ACE2 and TMPRSS2. Infected receiver cells then express the luciferase reporter and luminescence is measured, as an indicator for the combined efficiency of protein expression, VLP assembly in the producer cells, and entry into the receiver cells (*73*).

First, a control experiment was carried out with the N:L222P mutant previously shown to abrogate LRS oligomerization and RNP formation. Similar to a second control without N-protein plasmid, N:L222P produced very little luminescence relative to ancestral N (**Figure 7**). A positive control was N:R203M (green), which was previously shown to significantly enhance the signal of the VLP assay(*73*), and did so in the present VLP experiments. Similarly consistent with previous reports(*52*, *73*), R203K/G204R (blue) led to increased VLP signals. Of the mutations related to N-protein binding interfaces examined in the present work, neither P13L (red), nor Δ31-33 (light red), nor G215C (yellow) alone led to significant enhancement; only G214C (orange) produced a small but statistically significant enhancement. However, combination of the latter with P13L in P13L/G214C (brown) increased the gain, and further incorporation of the R203K/G204R double mutation to produce the full set of N_λ mutations (purple) substantially increased the measured luminescence. A similar observation was made for the combination of P13L/Δ31-33/R203K/G204R constituting the full set of N_o mutations (magenta).

<u>P13L mutation in N-protein promotes SARS-CoV-2 replication in cell lines</u>

Finally, we characterized the fitness of P13L mutation by introducing it into a recombinant SARS-CoV-2 reporter virus expressing an mCherry gene fused to N *via* a P2A linker (*74*). We infected Vero-TMPRSS2 and A549-ACE2 cells at an MOI of 0.01 with Wuhan-Hu-1 (ancestral) and mutant viruses, and measured mCherry fluorescence as well as viral release into the supernatant at indicated time points post-infection. This revealed that the P13L mutant exhibited a stronger mCherry signal throughout the course of infection (**Figure 8A,B**) and generated more progeny virus strain in both cell lines (**Figure 8C**). These findings are consistent with P13L providing a fitness advantage over ancestral virus.

Discussion

Rapid evolution of protein binding interfaces has frequently been observed in viral protein complexes, notably in the virus-host interface, including viral surface glycoproteins as well as ribonuclear proteins and non-structural proteins, with fitness advantages being accomplished, for example, through reshaping the binding interfaces, modulating protein structural dynamics, or altering physicochemical properties (*35*, *75–78*). Also, entirely new interactions can arise through the viral mimicry of eukaryotic short linear motifs as a result of frequent mutations in the viral protein IDRs, which can greatly augment the virus-host interface (*28*, *79*). The mutations we have studied here are of a different category,

impacting the interactions among viral proteins that enhance viral multi-protein complexes. Previous examples include the intra-host diversity of polymerase subunit interfaces in H5N1 influenza viruses (*80*) . While such mutations are not directly targeted towards the host, they may still contribute to host adaptation, or be balancing other mutation effects in an epistatic network, conceivably involving modulation of local effective protein concentrations (*81*). Irrespective of their complete context, they can provide valuable insights in viral protein mechanisms.

Specifically, we have described three different mutations of SARS-CoV-2 N-protein that, in convergent evolution, strengthen the formation of RNPs and enhance viral assembly. N:G214C, N:G215C, and N:P13L have been independently introduced (as highlighted in the phylogenetic trees **Supplementary Figure S8-10** generated in Nextstrain(*37*)), and persisted in the defining set of mutations in their respective variants of concern (Lambda, Delta, and Omicron, respectively). We have shown here that N:P13L confers a fitness advantage in cell lines, and similarly, N:G215C was shown by Kubinski *et al.*(*68*) to impart improved viral growth. This correlates well with our results studying their molecular mechanisms. For both of the cysteine mutants, molecular dynamics simulations and biophysical studies show how cysteines augment self-association interfaces by extending and redirecting the transiently formed helical coiled-coils in the intrinsically disordered LRS, which play a central role in the assembly of RNPs. By contrast, for N:P13L, unexpectedly, the evolution of RNP stability goes beyond modulation of a previously existing binding interface, and instead we observe the *de novo* formation of an additional dynamic self-association interface in the distant disordered N-arm through the stacking of transient β-sheets, that, we hypothesize, cooperatively contributes to the stability of RNPs. Even though the solution affinity of the N-arm P13L interface is ultra-weak, based on the RNP volume observed in cryo-EM(*48*) and assuming a dodecameric N cluster, the average local concentration of N-arm chains across the RNP volume is ≈1 mM, such that disordered N-arm peptides could well create populations of N-arm clusters stabilizing RNPs through this interface. (Unfortunately, transient structural changes in N-arm would remain undetectable in the CD spectrum of the RNPs in the *in vitro* assay of **Figure 1D**, due to limitations in the buffer transparency in the far UV, their small relative contributions, and the dominance of N-protein spectra by the strong signature of disordered chains.)

This leads to a picture of SARS-CoV-2 RNPs that is far from structurally well defined, matching the concept of fuzzy complexes (*82*). On a molecular level, large portions of the SARS-CoV-2 N-protein (the N-arm, C-arm, and linker) are intrinsically disordered and highly flexible(*47*, *83*), which persists in the presence of bound NA(*40*, *84*) . It appears that conformational freedom is also retained to a significant degree in the RNPs. This flexibility could be advantageous for accommodating various RNA secondary structures(*10*) , and favorably balance the energetic cost of RNP disassembly that is required immediately after viral entry. Also, this serves to accommodate significant sequence variation supporting, for example, the mimicry and rapid evolution of eukaryotic SLiMs (*28*, *29*, *31*) mentioned above. SARS-CoV-2 RNPs appear highly heterogeneous in EM (*10*, *49*), and this is reflected in the polymorphic oligomeric states of RNP species we observe in SV-AUC and MP, that we believe is driven by promiscuous self-association or clustering of transient LRS helices(*18*). Extending previously described characteristics of fuzziness in protein complexes(*29*, *32*, *85*), plasticity seems to involve even basic architectural principles, considering not only the emergence of new distant stabilizing interfaces

such as described here in the N-arm, but also the possibility of RNP assembly of truncated $N_{210-419}$* lacking one of the major NA binding interfaces (*9*, *51*, *86–88*) (see below).

Unfortunately, this intrinsic heterogeneity poses significant methodological hurdles. Nonetheless, salient structural features and assembly principles may be derived from constraints of known binding interfaces and oligomeric states of the RNP and its subunits, as observed in SV-AUC and MP. While the arrangement sketched in **Figure 1C** satisfies these requirements, alternate less symmetrical configurations can be conceived that seem at least equally likely and may coexist in polydisperse mixtures of RNPs. For example, there is no evidence to exclude the possibility of anti-parallel LRS helices pointing the folded NA-binding domains in different relative orientations, or of mixed co-assemblies with $N_{210-419}$* subunits lacking the NTD (**Supplemental Figure S11**). Uniformity of N-protein/RNA clusters may not be relevant for adequate gRNA condensation.

Beyond the structural model, to study the effect of a larger number of N-protein mutations derived from VOCs side-by-side in the context of virus assembly, we have carried out experiments using a VLP assay(*73*) (**Figure 7**). In these experiments, all four structural proteins are transfected into 293T cells to package a reporter RNA into VLPs and their infection of receiver cells can be compared. It should be recognized that the results do not directly reflect the relative efficiency of RNP assembly only, since protein expression levels, their localization, and their posttranslational modifications are not controlled for. Susceptibility for such factors might be exacerbated with mutations that modulate weak protein interactions. For example, as shown previously(*9*, *50*), a GSK3 inhibitor inhibiting N-protein phosphorylation significantly enhances VLP formation and eliminates the advantage provided for by the N:G215C mutation relative to the ancestral N – presumably due to an increase in assembly-competent, non-phosphorylated N-protein erasing an affinity advantage. A similar process may be underlying the absent or marginal improvement in VLP readout from the cysteine LRS mutants and P13L at the achieved transfection level in the present work, and the enhanced signal from R203K/G204R and R203M (the latter being consistent with previous reports(*45*, *73*)) modulating protein phosphorylation. Nonetheless, mirroring the results of the biophysical *in vitro* experiments, the addition of RNP-stabilizing P13L and G214C mutations on top of R203K/G204R led to a significantly larger VLP signal. Irrespective of the VLP conditions, in subsequent reverse genetics experiments the sole N:P13L mutation was shown to significantly increase viral fitness (**Figure 8**).

Regarding the cysteine mutations that have been repeatedly introduced in the LRS prior to the rise of the Omicron VOCs, it is an open question whether they lead to covalent bonds *in vivo*. While examples of disulfide-linked viral nucleocapsid proteins have been reported (*68*, *89*, *90*), a methodological difficulty in their detection is artifactual disulfide bond formation post-lysis of infected cells (*68*, *89*). Our results clearly show that a major effect of the cysteines enhancing RNP formation derives from extension of the LRS helices, and concomitant redirection of the disordered N-terminal sequence, already without any covalent bonds. While oxidized tetrameric N-proteins of N:G214C and N:G215C can be incorporated into RNPs, the covalent bonds provided only marginally improved RNP stability. Interestingly, the introduction of cysteines imposes preferences of RNP oligomeric states dependent on oxidation state, consistent with our MD simulations highlighting the impact of cysteine orientation of 214C *versus* 215C relative to the hydrophobic surface of the LRS helices. Overall, considering potentially

detrimental structural constraints from covalent bonds on LRS clusters seeding RNPs, energetic penalties on RNP disassembly, as well as the required monomeric state of the LRS helix for interaction with the NSP3 Ubl domain(*16*), at present it is unclear to what extent the formation of disulfide linkages between LRS helices would be beneficial or detrimental in the viral life cycle.

Recent work by the Soranno laboratory has identified an additional function of the disordered N-arm in transiently interacting with the NTD(*83*) and dynamically enhancing the affinity of the NTD for RNA (*40*). Using single-molecule Förster Resonance Energy Transfer (smFRET) a fourfold modulatory effect of the P13L/Δ31-33 mutation on the NTD RNA binding affinity was observed in N-arm-NTD constructs. Through MD simulations the reduced NTD affinity for RNA was attributed to the N-arm Δ31-33 deletion (*40*). Superficially, this may seem in slight conflict with our results of similar T_{10} affinity of FL ancestral N with and without the P13L/Δ31-33 mutation, but results were obtained in different buffer conditions (50 mM TRIS, pH 7.4 in (*40*) *versus* 20 mM HEPES, 150 mM NaCl, pH 7.5 in the present work). In any event, RNA binding of NTD and stabilization of the RNP are different processes; any modulation of N-arm contributions to NTD-RNA interactions through Omicron N-arm mutations Δ31-33 may coexist and be over-compensated for by N-arm self-association interfaces through P13L modulating RNP subunit interactions in the high local N-arm density of the RNP.

The double mutant R203K/G204R arose early in the pandemic and was adopted in several VOCs (including Alpha, Gamma, Lambda, Zeta, and Omicron BA.1) with the triple nucleotide changes G28881A, G28882A, and G28883C (*9*, *87*) (**Supplementary Figure S12**). As mentioned above, on the protein level N:R203K/G204R has been shown to alter phosphorylation (though in different ways in *in vitro* VLP or *in vivo* reverse genetics experiments)(*9*, *42*, *91*) ; and phosphorylation in turn reduces NA binding and promotes viral replication as opposed to assembly functions (*8*, *9*, *13*, *51*). Adding to such a switch, in the present work we observed the loss of RNP stability of N:R203K/G204R relative to the ancestral N, extending the previous observation of reduced LLPS propensity of N:R203K/G204R (*25*). Simultaneously, on the RNA level the N:R203K/G204R mutations also lead to the new formation of a TRS sequence ACGAAC underlying the expression of $N_{210\text{-}419}$* in virus-infected cells (though not expected to occur with N:R203K/G204R in the VLP assay lacking the viral RNA-dependent RNA polymerase). It has been hypothesized that $N_{210\text{-}419}$* confers increased viral fitness through the suppression of the host anti-viral response(*87*, *88*), and that it can assist RNP formation(*9*, *51*). However, the contribution of $N_{210\text{-}419}$* to assembly is still unclear: although it is remarkably capable of forming RNPs *in vitro* and VLP assays (*9*, *51*, *86*), in infected cells and virions $N_{210\text{-}419}$* has been detected only as a minority species(*87*, *88*). Also, the most recent major Omicron XEC variant(*92*) (which had close to 60% global frequency at the beginning of 2025; **Supplementary Figure S12**) exhibits a fourth consecutive nucleotide change G28884C that maintains a similar RG mutation forming R203K/G204P but ablates the canonical TRS sequence, such that continued expression of $N_{210\text{-}419}$* in XEC is in question. We propose that an alternative or additional mechanism to retain viral assembly functions may be presented by the accompanying P13L mutation, which our data suggest can more than restore loss of RNP stability in the combination of RG mutations with P13L. This combination occurs in N_o and all Omicron variants so far, and was even further stabilized with a cysteine in the LRS in N_λ.

In conclusion, it has been proposed that mutations in SARS-CoV-2 N protein that affect viral assembly can impact infectivity and fitness (*9*, *18*, *51*, *52*). We believe the observed modulations of the RNP assembly and stability studied here highlight a key mechanism for this. Although effects on fitness of viruses carrying N mutations are most likely multi-factorial, they have been observed in reverse genetics tissue culture experiments previously for N:R203K/G204R (*42*, *52*, *87*), N:G215C (*68*), and in the present work for N:P13L. On the other hand, the rise of new VOCs was usually dominated by their spike protein mutations (with the exception of 21I replacement by 21J which has identical spike mutations but acquired N:G215C(*18*, *66*, *67*) in the rise of Delta variant), and in many cases N mutations of previously dominant variants were completely replaced by another set of N mutations (dramatically exemplified in the displacement of Delta by Omicron variants). This reinforces the view that these N mutations are secondary to alterations in the immune landscape and transmissibility as the primary driver of evolution(*93*). Nonetheless, the remarkable plasticity of RNPs offers multiple avenues to module stability and to compensate for potentially RNP-destabilizing effects of mutations that are beneficial in other ways. In convergent evolution, this has been a constant theme of N protein mutations throughout the SARS-CoV-2 pandemics up until to date. We hypothesize that the 'fuzziness' and pleomorphic ability of RNP assembly, with its variable distribution of overall binding energy into several different weak or ultra-weak protein interfaces, and the poor structural definition ranging from flexible chain configurations to polydisperse oligomeric states, provides an evolutionary advantage of orchestrated disorder to promote epistatic interactions and facilitate host adaptation.

Methods

Protein expression and purification

Full-length N-protein of the wild type and mutant SARS-CoV-2 were expressed and purified as described previously (*25*, *50*) . Briefly, One Shot BL21(DE3)pLysS E.coli (Thermo Fisher Scientific, Carlsbad, CA) was transformed using a pET29a(+) plasmid vector, which contains a kanamycin-resistant gene and the gene encoding the N-protein of interest preceded by 6xHis tag followed by a Tobacco Etch Virus (TEV) cleavage site at the N-terminus. After cell lysis, the protein was purified by Ni^{2+} affinity chromatography, where on-column unfolding by urea and refolding steps were carried out to remove protein-bound bacterial nucleic acid (*13*). After tag cleavage by TEV protease, 6xHis tag removal was verified through another round of affinity chromatography, and/or *via* mass spectrometry. Cleaved protein was subjected to a final size exclusion chromatography followed by dialysis into working buffer (20 mM HEPES, 75 mM NaCl, pH 7.50, supplemented with 1 mM TCEP for cysteine containing proteins, unless otherwise mentioned). Protein purity was validated by SDS-PAGE and the absence of nucleic acid was confirmed by an absorbance ratio 260/280 of ≈0.50-0.55. Final protein concentration was determined by UV-Vis spectrophotometry or by refractive index detected SV-AUC.

N peptides were purchased from ABI Scientific (Sterling, VA), purified by HPLC, examined by matrix-assisted laser desorption/ionization for purity and identity, and lyophilized. The oligonucleotide T_{10} and stem–loop RNA SL7 were purchased from Integrated DNA Technologies (Skokie, IL) and purified by HPLC and lyophilized by the vendor. After reconstitution, SL7 was subject to thermal denaturation at 95 °C for 2 min followed by gradual cooling to room temperature over 1–2 h. For sequences of oligonucleotides and peptides see **Supplementary Table S1**.

The 5,5'-Dithiobis-(2-Nitrobenzoic Acid) (DTNB) (catalog #22582) and Cysteine-HCl (catalog #44889) were purchased from Thermo Fisher Scientific Inc. (Waltham, MA). Free thiols in the protein samples were quantified using the Ellman's assay by following the standard protocol from the vendor. Briefly, free thiols react with DTNB, resulting in a measurable yellow-colored product, TNB. The quantity of sulfhydryl groups was calculated by the absorbance of the sample using the molar extinction coefficient of TNB at 412 nm (14,150 $M^{-1}cm^{-1}$). The same samples were analyzed by SDS-PAGE without the addition of a reducing agent. The relative intensities of the monomer and dimer bands reflect the amounts of the reduced and oxidized forms, respectively.

Structure prediction and MD simulations

In the studies of LRS peptides, the initial structures of the monomer and oligomers of the ancestral $N_{210-246}$ (**Supplementary Table S1**) were predicted using AlphaFold3 (AF3). The predicted conformations agreed with those reported in (*18*, *58*), with the oligomers showing a parallel, left-handed coiled-coil signature. Point mutations were introduced by replacing the corresponding glycine residue with cysteine. Graphics were created using ChimeraX.

MD simulations were performed for monomers and trimers of the ancestral peptides $N_{210-246}$ and two mutant sequences (G214C and G215C), following the protocol described in (*58*). In each case, the initial

structure was the top-ranked AF3 model. Each simulation was extended for 250 ns after thermal equilibration under experimental conditions (T = 20°C, P = 1 atm, pH 7.5, and 75 mM NaCl) using the isothermal-isobaric ensemble as implemented in NAMD. The all-atom representation of the CHARMM (param36) force field was used. Structural stabilization of the helical regions was observed within a few nanoseconds, and data were analyzed over the last 200 ns of the simulations.

Structure of the N-arm oligomers was predicted using ColabFold (*94*), assembling 20 copies of ancestral N_{10-20}, N_{10-20}:P13L, N_{10-20}:P13S, or N_{10-20}:P13T. Structures were analyzed and displayed using ChimeraX (*95*).

Electron microscopy

Carbon coated 200 mesh copper TEM grids were glow discharged for ≈15 seconds. Next, 4 µL of the sample solution was deposited onto the grids and incubated for 2 minutes. After incubation, excess sample solution was removed by gently touching the edge of the grids with filter paper. A large drop of distilled water was then placed on the grids for 1 minute, followed by removal of the water by contacting the grid edge with filter paper. This rinsing step was repeated three times. The grids were then stained with 5 µL of 1% uranyl acetate (UA) solution for 20 seconds. Any excess UA was removed by touching the edge of the grids with filter paper and then left to air dry. Finally, the grids were examined with a FEI Tecnai12 Transmission Electron Microscope (FEI, Hillsboro, Oregon), operating at 120 keV beam energy. TEM images are captured using a high-speed, high-resolution Gatan Rio 3k x 3k CMOS camera (Gatan, Warrendale, PA).

Sedimentation velocity analytical ultracentrifugation

SV-AUC experiments were conducted in a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter, Indianapolis, IN) using standard protocols as previously described (*96*). AUC cell assemblies filled with samples using 12- or 3-mm charcoal-filled Epon double-sector centerpieces with sapphire windows were loaded into An-50 or An-60 rotors and temperature equilibrated in the rotor chamber at 20 °C for 2-3 hrs. Subsequently, radial scans were acquired with Rayleigh interference optics and absorbance optics at 260 nm and/or 280 nm. Calibration factors for the instrument were determined according to previously published methods(*97*). Sedimentation boundary data were analyzed using sedimentation coefficient distribution *c*(*s*) model in the software SEDFIT (*98*) (https://sedfitsedphat.nibib.nih.gov/software).

Mass photometry

Mass photometry measurements were performed in a TwoMP instrument (Refeyn, UK) following the standard protocol (*72*) unless otherwise mentioned. Samples were loaded in the mini-wells formed by a silicone gasket which was placed on top of a coverslip mounted on the microscope stage. Two configurations of sample loading were used in the current study. For the time-dependent experiments, the samples were first prepared in the Eppendorf tubes prior to MP experiment. Then the working buffer (9 µL) was loaded onto the coverslip for focusing the objective. Subsequently, the sample was added to the buffer droplet, gently mixed and the measurement was initiated immediately as the 1st time point. The subsequent acquisitions for the same sample continued for specific time intervals. For

the samples which were not subject to this dilution/mixing configuration, sample mixtures were equilibrated 2 – 3 hours, focusing was achieved by using the buffer-free option provided by the data acquisition software, and data was collected immediately after sample application. Calibration of the TwoMP instrument was performed using the two calibrants, Beta-Amylase from Sweet Potato (Sigma A8781) and Thyroglobulin from Bovine Thyroid (Sigma T9145) as recommended by the manufacturer. MP data was acquired with AcquireMP software and the analysis was performed with DiscoverMP software (Refeyn, UK).

Rapid mixing experiments were carried out on a OneMP-MassFluidix HC system (Refeyn Ltd., Oxford, UK). A rapid-dilution microfluidic chip (MP-CON-51001, Refeyn Ltd., Oxford, UK) was connected to the buffer, sample, and waste lines. Sample was injected into the chip at a flow rate of 8 µL/min, while the buffer was flowed at 1100 µL/min, and 1 min videos were recorded for data acquisition.

Circular dichroism spectroscopy

CD spectra were acquired in a Chirascan Q100 (Applied Photophysics, UK) at 20 °C. Measurements were performed in 0.1 mm (peptides) or 1 mm (proteins) pathlength cuvettes with 1 nm steps, and a 1 sec integration time per data point. Each spectrum represents the average of three independent scans with background subtraction applied.

Dynamic light scattering

Dynamic light scattering measurements of the samples were performed in a Prometheus Panta (Nanotemper, Germany) instrument at 20°C. The samples were loaded into capillaries (Nanotemper PR-AC002) and autocorrelation functions (ACFs) were acquired using the 405 nm laser at the detection angle of 140°. The ACFs were analyzed with discrete species models and size-distribution models in SEDFIT (*99*).

Virus-like particle assay

The plasmids pLVX-EF1alpha-SARS-CoV-2-E-2xStrep-IRES-Puro (#141385), pLVX-EF1alpha-SARS-CoV-2-M-2xStrep-IRES-Puro (#141386), and pLVX-EF1alpha-SARS-CoV-2-N-2xStrep-IRES-Puro (#141391) were obtained from Addgene. Plasmid pLuc-T20 was a kind gift from Jennifer A. Doudna. The plasmid pIRES2-EGFP (Cat # V011106) was purchased from NovoPro. Plasmids encoding ACE2, pcDNA3.1(+)-SARS-CoV-2 WA1-S and pGAGGS-TMPRSS2 were previously described (*100*, *101*). To construct the plasmid pcDNA3.1(+)-N, the N gene was cloned into pcDNA3.1(+) between the BamHI and NotI sites. Mutations in N were generated by QuikChange™ site-directed mutagenesis, and verified by whole plasmid sequencing. Similarly, to prepare plasmid pIRES2-ME, the M gene was first cloned into pIRES2-EGFP between NheI and BamHI sites, followed by the insertion of the E gene into the resulting plasmid between NotI and a PvuI site introduced after the start codon of EGFP.

The SARS-CoV-2 virus-like particles (VLPs) were prepared as previously described(*50*, *73*) with some modifications. For SC2-VLP production, 0.8×10^6 293T cells per well were plated in a 6-well plate and allowed to grow 24 hours before transfection. Plasmids Cov2-N (1.34), CoV2-M-IRES-E (0.659), CoV-2-Spike (0.0032) and Luc-T20 (2.0) at indicated mass ratios for a total of 4 µg of DNA were diluted in 250

µL Opti-MEM reduced serum medium (Gibco cat# 31985062) at room temperature. 12 µL of TransIT®-293 Transfection Reagent (Mirus Bio, cat# MIR 2704) equilibrated to room temperature was added to each sample, and immediately vortexed. The transfection mixtures were incubated for 25 min at room temperature and then added dropwise to 293T cells after media change with 2 mL of DMEM containing fetal bovine serum and penicillin/streptomycin. Media was changed after 24 h of transfection and at 48 h post-transfection, VLP containing supernatant was collected and filtered using a 0.22 µm syringe filter. Cells in each well were rinsed with 1 mL of PBS and then lysed directly in the wells with 300 µL of NuPAGE™ LDS Sample Buffer (Invitrogen, cat# NP0008) containing HALT protease inhibitors (Thermo Scientific, cat# 87786) and 10 mM DTT.

For the luciferase assay, the wells in 12-well plates were pre-treated with 1 mL poly-D- Lysine (Thermo Fisher Scientific, cat# A3890401) for 30 min, which was removed, washed with PBS once, and finally the plates were airdried. In each well of these pre-treated plates, 1.5×10^5 receiver cells (293T cells transfected with the plasmids encoding ACE2 and pGAGGS-TMPRSS2 at a mass ratio of 1:1 using the TransIT®-293 transfection reagent) were plated. Next day, the media was replaced, and the cells infected with 250 µL of supernatant from the producer cells. After 24 h, the media was removed, and cells were rinsed with PBS and lysed in 150 µL passive lysis buffer (Luciferase Assay System, Promega, cat# E1500) for 15 min at room temperature with gentle rocking. 20 µL of each lysate was transferred to an opaque black 96-well plate in triplicate, and 50 µL of reconstituted luciferase assay buffer was added and mixed with each lysate. Luminescence was measured immediately after mixing using a TECAN plate reader.

Generation of recombinant SARS-CoV-2

The generation and use of recombinant SARS-CoV-2 (rSARS-CoV-2) viruses in tissue culture at biosafety level 3 were approved by NIH Institutional Biosafety (IBC) and the Dual Use Research of Concern Institutional Review Entity (DURC-IRE) Committees (IBC approved case number: RD-22-XI-11).

We generated rSARS-CoV-2 viruses using a bacterial artificial chromosome (BAC)-based SARS-CoV-2 reverse genetics system (*74*), with detailed construction and recovery procedures described in our previous study (*102*). Specifically, we fused the mCherry gene with the N gene via a 2A linker and created an intermediate plasmid, pUC57-NEM, containing a portion of the pBAC-SARS-CoV-2 genome digested with BamHI and SalI restriction enzymes. The P13L mutation was introduced via site-directed mutagenesis on the pUC57-NEM plasmid. PCR amplification was performed to generate fragments containing the NEM genes, which were then assembled with the BamHI/SalI-digested larger fragment using the NEBuilder® HiFi DNA Assembly Master Mix. A furin cleavage site mutation (R685S) served as the backbone for introducing mutations in the N protein.

Plasmids were purified using the QIAGEN Plasmid Maxi Kit. Confluent BHK21-ACE2 cells (2×10^6 cells/well in 6-well plates, in duplicate) were transfected with 2.5 µg/well of pBAC-SARS-CoV-2 using the TransIT-LT1 transfection reagent. After 6 hours, the medium was replaced with fresh DMEM containing 2% FBS. At 48-72 hours post-transfection, mCherry-positive cells exhibiting signs of viral infection were detached, collected with the supernatant, labeled as P0, and stored at −80°C. The P0 viral stock was

centrifuged to remove cell debris and used to infect fresh Vero E6-TMPRSS2 cells for 48-72 hours. The resulting P1 viral stock was collected, and viral titers were determined following next-generation sequencing (NGS) confirmation of the viral genome sequence.

SARS-CoV-2 infection

Vero-TMPRSS2 and A549-hACE2 cell lines were seeded in 12-well plates at 3×10^5 cells/well. The next day, cells were infected at MOI 0.01 with an inoculation period of 1 hour, followed by a medium change to fresh culture medium. Supernatants were collected, and infected cells were fixed with 4% paraformaldehyde (PFA) for 30 minutes at indicated time before removal from the BSL-3 laboratory for fluorescence imaging using Cytation 5 (BioTek). Virus released into the supernatant was titrated using Vero cells via the limiting dilution method.

Acknowledgements

This work was supported by the Intramural Research Programs of NIBIB (ZIA EB000099-02), NHLBI, NCI, and NIAID at the National Institutes of Health. This work utilized the computational resources of the NIH HPC Biowulf cluster for molecular dynamics simulations.

References

1. A. S. Holehouse, B. B. Kragelund, The molecular basis for cellular function of intrinsically disordered protein regions. Nature Research [Preprint] (2024). https://doi.org/10.1038/s41580-023-00673-0.

2. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. [Preprint] (2005). https://doi.org/10.1038/nrm1589.

3. N. Tokuriki, C. J. Oldfield, V. N. Uversky, I. N. Berezovsky, D. S. Tawfik, Do viral proteins possess unique biophysical features? *Trends Biochem Sci* **34**, 53–59 (2009).

4. M. N. Gupta, V. N. Uversky, Protein structure–function continuum model: Emerging nexuses between specificity, evolution, and structure. *Protein Science* **33**, 1–36 (2024).

5. C. J. Brown, A. K. Johnson, A. K. Dunker, G. W. Daughdrill, Evolution and disorder. *Curr Opin Struct Biol* **21**, 441–446 (2011).

6. B. Xue, D. Blocquel, J. Habchi, A. V. Uversky, L. Kurgan, V. N. Uversky, S. Longhi, Structural disorder in viral proteins. American Chemical Society [Preprint] (2014). https://doi.org/10.1021/cr4005692.

7. S. Ranganathan, P. Dasmeh, S. Furniss, E. Shakhnovich, Phosphorylation sites are evolutionary checkpoints against liquid-solid transition in protein condensates. *Proc Natl Acad Sci U S A* **120** (2023).

8. M. Botova, A. R. Camacho-Zarco, J. Tognetti, L. M. Bessa, S. Guseva, E. Mikkola, N. Salvi, D. Maurin, T. Herrmann, M. Blackledge, A specific phosphorylation-dependent conformational switch in SARS-CoV-2 nucleocapsid protein inhibits RNA binding. *Sci Adv* **10**, 1–15 (2024).

9. A. M. Syed, A. Ciling, I. P. Chen, C. R. Carlson, A. N. Adly, H. S. Martin, T. Y. Taha, M. M. Khalid, N. Price, M. Bouhaddou, M. R. Ummadi, J. M. Moen, N. J. Krogan, D. O. Morgan, M. Ott, J. A. Doudna, SARS-CoV-2 evolution balances conflicting roles of N protein phosphorylation. *PLoS Pathog* **20**, e1012741 (2024).

10. C. R. Carlson, A. N. Adly, M. Bi, C. J. Howard, A. Frost, Y. Cheng, D. O. Morgan, Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation. *Journal of Biological Chemistry* **298**, 102560 (2022).

11. A. Jack, L. S. Ferro, M. J. Trnka, E. Wehri, A. Nadgir, X. Nguyenla, D. Fox, K. Costa, S. Stanley, J. Schaletzky, A. Yildiz, SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol* **19**, e3001425 (2021).

12. A. Savastano, A. Ibáñez de Opakua, M. Rankovic, M. Zweckstetter, Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat Commun* **11**, 6041 (2020).

13. C. R. Carlson, J. B. Asfaha, C. M. Ghent, C. J. Howard, N. Hartooni, M. Safari, A. D. Frankel, D. O. Morgan, 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 (2020).

14. C. A. Roden, Y. Dai, C. A. Giannetti, I. Seim, M. Lee, R. Sealfon, G. A. McLaughlin, M. A. Boerneke, C. Iserman, S. A. Wey, J. L. Ekena, O. G. Troyanskaya, K. M. Weeks, L. You, A. Chilkoti, A. S. Gladfelter, Double-stranded RNA drives SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures. *Nucleic Acids Res* **50**, 8168–8192 (2022).

15. I. V. Nesmelova, D. L. Melnikova, V. Ranjan, V. D. Skirda, "Translational diffusion of unfolded and intrinsically disordered proteins" in *Progress in Molecular Biology and Translational Science* (Elsevier B.V., 2019)vol. 166, pp. 85–108.

16. L. M. Bessa, S. Guseva, A. R. Camacho-Zarco, N. Salvi, D. Maurin, L. M. Perez, M. Botova, A. Malki, M. Nanao, M. R. Jensen, R. W. H. Ruigrok, M. Blackledge, The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci Adv* **8** (2022).

17. T. R. Alderson, I. Pritišanac, Đ. Kolarić, A. M. Moses, J. D. Forman-Kay, Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proceedings of the National Academy of Sciences* **120**, 2022.02.18.481080 (2023).

18. H. Zhao, A. Nguyen, D. Wu, Y. Li, S. A. Hassan, J. Chen, H. Shroff, G. Piszczek, P. Schuck, Plasticity in structure and assembly of SARS-CoV-2 nucleocapsid protein. *PNAS Nexus* **1**, pgac049 (2022).

19. S. Alberti, A. S. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

20. S. Ranganathan, E. I. Shakhnovich, Dynamic metastable long-living droplets formed by sticker-spacer proteins. *Elife* **9**, 1–25 (2020).

21. M. Zachrdla, A. Savastano, A. Ibáñez de Opakua, M. Cima-Omori, M. Zweckstetter, Contributions of the N-terminal intrinsically disordered region of the severe acute respiratory syndrome coronavirus 2 nucleocapsid protein to RNA-induced phase separation. *Protein Science* **31**, 2003–2005 (2022).

22. S. Brocca, R. Grandori, S. Longhi, V. Uversky, Liquid–liquid phase separation by intrinsically disordered protein regions of viruses: Roles in viral life cycle and control of virus–host interactions. *Int J Mol Sci* **21**, 1–31 (2020).

23. T. M. Perdikari, A. C. Murthy, V. H. Ryan, S. Watters, M. T. Naik, N. L. Fawzi, SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J* **39**, 1–35 (2020).

24. T. Zarin, B. Strome, G. Peng, I. Pritišanac, J. D. Forman-Kay, A. M. Moses, Identifying molecular features that are associated with biological function of intrinsically disordered protein regions. *Elife* **10**, 1–36 (2021).

25. A. Nguyen, H. Zhao, D. Myagmarsuren, S. Srinivasan, D. Wu, J. Chen, G. Piszczek, P. Schuck, Modulation of biophysical properties of nucleocapsid protein in the mutant spectrum of SARS-CoV-2. *Elife* **13** (2024).

26. J. J. Alston, A. Soranno, A. S. Holehouse, Conserved molecular recognition by an intrinsically disordered region in the absence of sequence conservation. *bioRxiv*, doi: 10.1101/2023.08.06.552128 (2023).

27. A. Aharoni, L. Gaidukov, O. Khersonsky, S. M. Q. Gould, C. Roodveldt, D. S. Tawfik, The "evolvability" of promiscuous protein functions. *Nat Genet* **37**, 73–76 (2005).

28. P. Schuck, H. Zhao, Diversity of short linear interaction motifs in SARS-CoV-2 nucleocapsid protein. *mBio* **14**, e02388-23 (2023).

29. N. Duro, M. Miskei, M. Fuxreiter, Fuzziness endows viral motif-mimicry. *Mol Biosyst* **11**, 2821–2829 (2015).

30. T. Hagai, A. Azia, M. M. Babu, R. Andino, Use of Host-like Peptide Motifs in Viral Proteins Is a Prevalent Strategy in Host-Virus Interactions. *Cell Rep* **7**, 1729–1739 (2014).

31. N. E. Davey, G. Travé, T. J. Gibson, How viruses hijack cell regulation. *Trends Biochem Sci* **36**, 159–169 (2011).

32. P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2–8 (2008).

33. S. Longhi, L. M. Bloyet, S. Gianni, D. Gerlier, How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cellular and Molecular Life Sciences* **74**, 3091–3118 (2017).

34. J. G. Olsen, K. Teilum, B. B. Kragelund, Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. Birkhauser Verlag AG [Preprint] (2017). https://doi.org/10.1007/s00018-017-2560-7.

35. N. D. Rochman, Y. I. Wolf, E. V. Koonin, Molecular adaptations during viral epidemics. *EMBO Rep* **23**, 1–22 (2022).

36. S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33–46 (2017).

37. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

38. J. D. Bloom, R. A. Neher, Fitness effects of mutations to SARS-CoV-2 proteins. *Virus Evol* **9**, 2023.01.30.526314 (2023).

39. W. Wu, Y. Cheng, H. Zhou, C. Sun, S. Zhang, The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virol J* **20**, 6 (2023).

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

41. K. Dhamotharan, S. M. Korn, A. Wacker, M. A. Becker, S. Günther, H. Schwalbe, A. Schlundt, A core network in the SARS-CoV-2 nucleocapsid NTD mediates structural integrity and selective RNA-binding. *Nature Communications* **15**, 1–16 (2024).

42. B. A. Johnson, Y. Zhou, K. G. Lokugamage, M. N. Vu, N. Bopp, P. A. Crocquet-Valdes, B. Kalveram, C. Schindewolf, Y. Liu, D. Scharton, J. A. Plante, X. Xie, P. Aguilar, S. C. Weaver, P.-Y. Shi, D. H. Walker, A. L. Routh, K. S. Plante, V. D. Menachery, Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. *PLoS Pathog* **18**, e1010627 (2022).

43. J. Ren, S. Wang, Z. Zong, T. Pan, S. Liu, W. Mao, H. Huang, X. Yan, B. Yang, X. He, F. Zhou, L. Zhang, TRIM28-mediated nucleocapsid protein SUMOylation enhances SARS-CoV-2 virulence. *Nat Commun* **15** (2024).

44. K. V. Tugaeva, A. A. Sysoev, A. A. Kapitonova, J. L. R. Smith, P. Zhu, R. B. Cooley, A. A. Antson, N. N. Sluchanko, Human 14-3-3 Proteins Site-selectively Bind the Mutational Hotspot Region of SARS-CoV-2 Nucleoprotein Modulating its Phosphoregulation. *J Mol Biol* **435**, 167891 (2023).

45. Y. Li, M. Li, H. Xiao, F. Liao, M. Shen, W. Ge, J. Ou, Y. Liu, L. Chen, Y. Zhao, P. Wan, J. Liu, J. Chen, X. Lan, S. Wu, Q. Ding, G. Li, Q. Zhang, P. Pan, The R203M and D377Y mutations of the nucleocapsid protein promote SARS-CoV-2 infectivity by impairing RIG-I-mediated antiviral signaling. *PLoS Pathog* **21**, 1–28 (2025).

46. C. K. Chang, S. C. Sue, T. H. Yu, C. M. Hsieh, C. K. Tsai, Y. C. Chiang, S. J. Lee, H. H. Hsiao, W. J. Wu, W. L. Chang, C. H. Lin, T. H. Huang, Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci* **13**, 59–72 (2006).

47. B. Różycki, E. Boura, Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data. *Biophys Chem* **288**, 106843 (2022).

48. S. Klein, M. Cortese, S. L. Winter, M. Wachsmuth-Melm, C. J. Neufeldt, B. Cerikan, M. L. Stanifer, S. Boulant, R. Bartenschlager, P. Chlanda, SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885 (2020).

49. H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li, S. Li, Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738.e13 (2020).

50. H. Zhao, A. M. Syed, M. M. Khalid, A. Nguyen, A. Ciling, D. Wu, W. Yau, S. Srinivasan, D. Esposito, J. A. Doudna, G. Piszczek, M. Ott, P. Schuck, Assembly of SARS-CoV-2 nucleocapsid protein with nucleic acid. *Nucleic Acids Res* **52**, 6647–6661 (2024).

51. M. Bouhaddou, A. K. Reuschl, B. J. Polacco, L. G. Thorne, M. R. Ummadi, C. Ye, R. Rosales, A. Pelin, J. Batra, G. M. Jang, J. Xu, J. M. Moen, A. L. Richards, Y. Zhou, B. Harjai, E. Stevenson, A. Rojc, R. Ragazzini, M. V. X. Whelan, W. Furnon, G. De Lorenzo, V. Cowton, A. M. Syed, A. Ciling, N. Deutsch, D. Pirak, G. Dowgier, D. Mesner, J. L. Turner, B. L. McGovern, M. L. Rodriguez, R. Leiva-Rebollo, A. S. Dunham, X. Zhong, M. Eckhardt, A. Fossati, N. F. Liotta, T. Kehrer, A. Cupic, M. Rutkowska, I. Mena, S. Aslam, A. Hoffert, H. Foussard, C. O. Olwal, W. Huang, T. Zwaka, J. Pham, M. Lyons, L. Donohue, A. Griffin, R. Nugent, K. Holden, R. Deans, P. Aviles, J. A. Lopez-Martin, J. M. Jimeno, K. Obernier, J. M. Fabius, M. Soucheray, R. Hüttenhain, I. Jungreis, M. Kellis, I. Echeverria, K. Verba, P. Bonfanti, P. Beltrao, R. Sharan, J. A. Doudna, L. Martinez-Sobrido, A. H. Patel, M. Palmarini, L. Miorin, K. White, D. L. Swaney, A. Garcia-Sastre, C. Jolly, L. Zuliani-Alvarez, G. J. Towers, N. J. Krogan, SARS-CoV-2 variants evolve convergent strategies to remodel the host response. *Cell* **186**, 4597-4614.e26 (2023).

52. H. Wu, N. Xing, K. Meng, B. Fu, W. Xue, P. Dong, W. Tang, Y. Xiao, G. Liu, H. Luo, W. Zhu, X. Lin, G. Meng, Z. Zhu, Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host Microbe* **29**, 1788-1801.e6 (2021).

53. H. Zhao, D. Wu, A. Nguyen, Y. Li, R. C. Adão, E. Valkov, G. H. Patterson, G. Piszczek, P. Schuck, Energetic and structural features of SARS-CoV-2 N-protein co-assemblies with nucleic acids. *iScience* **24**, 102523 (2021).

54. L. Zinzula, J. Basquin, S. Bohn, F. Beck, S. Klumpe, G. Pfeifer, I. Nagy, A. Bracher, F. U. Hartl, W. Baumeister, High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochem Biophys Res Commun* **538**, 54–62 (2021).

55. J. Cubuk, J. Jeremias Incicco, K. B. Hall, A. S. Holehouse, M. D. Stuchell-Brereton, A. Soranno, The dimerization domain of SARS CoV 2 Nucleocapsid protein is partially disordered as a monomer and forms a high affinity dynamic complex. [Preprint] (2024). https://doi.org/10.1101/2024.09.25.614883.

56. I. M. Yu, C. L. T. Gustafson, J. Diao, J. W. Burgner, Z. Li, J. Zhang, J. Chen, Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *Journal of Biological Chemistry* **280**, 23280–23286 (2005).

57. A. Tarczewska, M. Kolonko-Adamska, M. Zarębski, J. Dobrucki, A. Ożyhar, B. Greb-Markiewicz, The method utilized to purify the SARS-CoV-2 N protein can affect its molecular properties. *Int J Biol Macromol* **188**, 391–403 (2021).

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

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

60. S. M. Korn, K. Dhamotharan, C. M. Jeffries, A. Schlundt, The preference signature of the SARS-CoV-2 Nucleocapsid NTD for its 5'-genomic RNA elements. *Nat Commun* **14**, 3331 (2023).

61. C. Iserman, C. A. Roden, M. A. Boerneke, R. S. G. Sealfon, G. A. McLaughlin, I. Jungreis, E. J. Fritch, Y. J. Hou, J. Ekena, C. A. Weidmann, C. L. Theesfeld, M. Kellis, O. G. Troyanskaya, R. S. Baric, T. P. Sheahan, K. M. Weeks, A. S. Gladfelter, Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol Cell* **80**, 1078–1091 (2020).

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

63. F. Obermeyer, M. Jankowiak, N. Barkas, S. F. Schaffner, J. D. Pyle, L. Yurkovetskiy, M. Bosso, D. J. Park, M. Babadi, B. L. MacInnis, J. Luban, P. C. Sabeti, J. E. Lemieux, Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science (1979)* **1208**, 1–14 (2022).

64. A. Oulas, M. Zanti, M. Tomazou, M. Zachariou, G. Minadakis, M. M. Bourdakou, P. Pavlidis, G. M. Spyrou, Generalized linear models provide a measure of virulence for specific mutations in SARS-cov-2 strains. *PLoS One* **16**, 1–17 (2021).

65. A. K. Paravastu, R. D. Leapman, W.-M. Yau, R. Tycko, Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proceedings of the National Academy of Sciences* **105**, 18349–18354 (2008).

66. V. Marchitelli, C. Troise, O. Vesuviano, A. Parisi, I. Zooprofilattico, A. Bianco, I. Zooprofilattico, L. Del Sambro, I. Zooprofilattico, R. Somma, O. Vesuviano, S. Ingegneria, U. Basilicata, G. De Natale, Evidence for the dependence of the SARS-Cov-2 Delta high diffusivity on the associated N : G215C nucleocapsid mutation. *Res Sq*, 1–9 (2021).

67. A. Stern, S. Fleishon, T. Kustin, M. Mandelboim, O. Erster, E. Mendelson, O. Mor, N. S. Zuckerman, D. Bucris, D. Sofer, D. Bar-Ilan, M. Geva, O. Asraf, G. Rechavi, E. Glick-Saar, N. Rainy, C. Weiner, R. Sorek-Abramovich, Y. Yegorov, A. Vishnevsky, P. Benveniste-Lekovitz, A. H. Ramzia, A. B. Chaim, The unique evolutionary dynamics of the SARS-CoV-2 Delta variant Israel Consortium of SARS-CoV-2 sequencing. *medRxiv*, 2021.08.05.21261642 (2021).

68. H. C. Kubinski, H. W. Despres, B. A. Johnson, M. M. Schmidt, L. K. Estes, A. Pekosz, J. W. Crothers, P. Roychoudhury, A. L. Greninger, K. R. Jerome, B. Martorelli Di Genova, D. H. Walker, B. A. Ballif, M. S. Ladinsky, P. J. Bjorkman, V. D. Menachery, E. A. Bruce, Variant mutation in SARS-CoV-2 nucleocapsid enhances viral infection via altered genomic encapsidation. *bioRxiv*, doi: 10.1101/2024.03.08.584120 (2024).

69. P. L. Wink, F. C. Z. Volpato, F. L. Monteiro, J. B. Willig, A. P. Zavascki, A. L. Barth, A. F. Martins, First identification of SARS-CoV-2 lambda (C.37) variant in Southern Brazil. *Infect Control Hosp Epidemiol* **43**, 1996–1997 (2022).

70. P. Schuck, Sedimentation patterns of rapidly reversible protein interactions. *Biophys. J.* **98**, 2005–2013 (2010).

71. P. Schuck, H. Zhao, *Sedimentation Velocity Analytical Ultracentrifugation: Interacting Systems* (CRC Press, Boca Raton, FL, 2017; https://doi.org/10.1201/b21988).

72. D. Wu, G. Piszczek, Standard protocol for mass photometry experiments. *European Biophysics Journal* **50**, 403–409 (2021).

73. A. M. Syed, T. Y. Taha, T. Tabata, I. P. Chen, A. Ciling, M. M. Khalid, B. Sreekumar, P.-Y. Chen, J. M. Hayashi, K. M. Soczek, M. Ott, J. A. Doudna, Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles. *Science (1979)* **374**, 1626–1632 (2021).

74. C. Ye, K. Chiem, J. G. Park, F. Oladunni, R. N. Platt, T. Anderson, F. Almazan, J. C. de la Torre, L. Martinez-Sobrido, Rescue of SARS-CoV-2 from a single bacterial artificial chromosome. *mBio* **11**, 1–10 (2020).

75. D. Evseev, K. E. Magor, Molecular Evolution of the Influenza A Virus Non-structural Protein 1 in Interspecies Transmission and Adaptation. *Front Microbiol* **12** (2021).

76. V. Barozi, A. L. Edkins, Ö. Tastan Bishop, Evolutionary progression of collective mutations in Omicron sub-lineages towards efficient RBD-hACE2: Allosteric communications between and within viral and human proteins. *Comput Struct Biotechnol J* **20**, 4562–4578 (2022).

77. D. Focosi, P. G. Spezia, F. Maggi, Subsequent Waves of Convergent Evolution in SARS-CoV-2 Genes and Proteins. *Vaccines (Basel)* **12**, 887 (2024).

78. C. Planchais, A. Reyes-Ruiz, R. Lacombe, A. Zarantonello, M. Lecerf, M. Revel, L. T. Roumenina, B. P. Atanasov, H. Mouquet, J. D. Dimitrov, Evolutionary trajectory of receptor binding specificity and promiscuity of the spike protein of SARS-CoV-2. *Protein Science* **31**, 1–19 (2022).

79. N. E. Davey, M. S. Cyert, A. M. Moses, Short linear motifs - Ex nihilo evolution of protein regulation Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Communication and Signaling* **13**, 9–11 (2015).

80. M. R. A. Welkers, H. A. Pawestri, J. M. Fonville, O. D. Sampurno, M. Pater, M. Holwerda, A. X. Han, C. A. Russell, R. E. Jeeninga, V. Setiawaty, M. D. de Jong, D. Eggink, Genetic diversity and host adaptation of avian H5N1 influenza viruses during human infection. *Emerg Microbes Infect* **8**, 262–271 (2019).

81. P. Li, B. Xue, N. J. Schnicker, L.-Y. R. Wong, D. K. Meyerholz, S. Perlman, Nsp3-N interactions are critical for SARS-CoV-2 fitness and virulence. *Proceedings of the National Academy of Sciences* **120**, 2017 (2023).

82. H. Wu, M. Fuxreiter, The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. *Cell* **165**, 1055–1066 (2016).

83. J. Cubuk, J. J. Alston, J. J. Incicco, S. Singh, M. D. Stuchell-Brereton, M. D. Ward, M. I. Zimmerman, N. Vithani, D. Griffith, J. A. Wagoner, G. R. Bowman, K. B. Hall, A. Soranno, A. S. Holehouse, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* **12**, 1936 (2021).

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

85. M. Fuxreiter, Fuzziness in protein interactions—a historical perspective. *J Mol Biol* **430**, 2278–2287 (2018).

86. A. N. Adly, M. Bi, C. R. Carlson, A. M. Syed, A. Ciling, J. A. Doudna, Y. Cheng, D. O. Morgan, Assembly of SARS-CoV-2 ribonucleosomes by truncated N∗ variant of the nucleocapsid protein. *Journal of Biological Chemistry* **299**, 105362 (2023).

87. H. V Mears, G. R. Young, T. Sanderson, R. Harvey, J. Barrett-Rodger, R. Penn, V. Cowton, W. Furnon, G. De Lorenzo, M. Crawford, D. M. Snell, A. S. Fowler, A. M. Chakrabarti, S. Hussain, C. Gilbride, E. Emmott, K. Finsterbusch, J. Luptak, T. P. Peacock, J. Nicod, A. H. Patel, M. Palmarini, E. Wall, B. Williams, S. Gandhi, C. Swanton, D. L. V. Bauer, Emergence of SARS-CoV-2 subgenomic RNAs that enhance viral fitness and immune evasion. *PLoS Biol* **23**, 1–29 (2025).

88. R. P. Mulloy, D. Evseev, M. P. Bui-Marinos, N. Sharlin, J. A. Corcoran, A truncated SARS-CoV-2 nucleocapsid protein enhances virus fitness by evading antiviral responses. *bioRxiv*, 1–68 (2025).

89. S. K. Wootton, D. Yoo, Homo-Oligomerization of the Porcine Reproductive and Respiratory Syndrome Virus Nucleocapsid Protein and the Role of Disulfide Linkages. *J Virol* **77**, 4546–4557 (2003).

90. E. N. Prokudina, N. P. Semenova, V. M. Chumakov, I. A. Rudneva, Transient disulfide bonds formation in conformational maturation of influenza virus nucleocapsid protein (NP). *Virus Res* **99**, 169–175 (2004).

91. J. S. Yun, H. Song, N. H. Kim, S. Y. Cha, K. H. Hwang, J. E. Lee, C. H. Jeong, S. H. Song, S. Kim, E. S. Cho, H. S. Kim, J. I. Yook, Glycogen Synthase Kinase-3 Interaction Domain Enhances Phosphorylation of SARS-CoV-2 Nucleocapsid Protein. *Mol Cells* **45**, 911–922 (2022).

92. F. Scarpa, F. Branda, G. Ceccarelli, C. Romano, C. Locci, N. Pascale, I. Azzena, P. L. Fiori, M. Casu, S. Pascarella, M. Quaranta, D. Benvenuto, R. Cauda, M. Ciccozzi, D. Sanna, SARS-CoV-2 XEC: A Genome-Based Survey. *Microorganisms* **13** (2025).

93. P. V Markov, M. Ghafari, M. Beer, K. Lythgoe, P. Simmonds, N. I. Stilianakis, A. Katzourakis, The evolution of SARS-CoV-2. *Nat Rev Microbiol* **21**, 361–379 (2023).

94. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

95. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).

96. P. Schuck, H. Zhao, C. A. Brautigam, R. Ghirlando, *Basic Principles of Analytical Ultracentrifugation* (CRC Press, Boca Raton, FL, 2015; https://doi.org/10.1080/19420862.2020.1810488).

97. R. Ghirlando, A. Balbo, G. Piszczek, P. H. Brown, M. S. Lewis, C. A. Brautigam, P. Schuck, H. Zhao, Improving the thermal, radial, and temporal accuracy of the analytical ultracentrifuge through external references. *Anal. Biochem.* **440**, 81–95 (2013).

98. P. Schuck, *Sedimentation Velocity Analytical Ultracentrifugation: Discrete Species and Size-Distributions of Macromolecules and Particles* (CRC Press, Boca Raton, FL, 2016).

99. E. T. Parker, P. Lollar, Measurement of the translational diffusion coefficient and hydrodynamic radius of proteins by dynamic light scattering. *Bio Protoc* **11**, e4195 (2021).

100. G. Shi, A. D. Kenney, E. Kudryashova, A. Zani, L. Zhang, K. K. Lai, L. Hall-Stoodley, R. T. Robinson, D. S. Kudryashov, A. A. Compton, J. S. Yount, Opposing activities of IFITM proteins in SARS-CoV-2 infection. *EMBO J* **40**, e106501 (2021).

101. G. Shi, A. I. Chiramel, T. Li, K. K. Lai, A. D. Kenney, A. Zani, A. C. Eddy, S. Majdoul, L. Zhang, T. Dempsey, P. A. Beare, S. Kar, J. W. Yewdell, S. M. Best, J. S. Yount, A. A. Compton, Rapalogs

downmodulate intrinsic immunity and promote cell entry of SARS-CoV-2. *Journal of Clinical Investigation* **132**, e160766 (2022).

102. T. Li, I. Kang, Z. Hu, J. Gibbs, C. Ye, I. Kosik, G. Shi, J. Holly, M. Kosikova, A. Compton, L. Martinez-Sobrido, R. F. Johnson, H. Xie, J. W. Yewdell, Syncytia Formation Promotes Virus Resistance to Interferon and Neutralizing Antibodies. *bioRxiv*, doi: 10.1101/2023.12.12.571262 (2023).

Table 1. Overview of N-protein mutant species studied

designation	N-protein mutations	in set of defining VOC mutations [1]	Reaction boundary s-value in RNP assay (S)	MP final Average Mw in 500-1500 kDa range (kDa)	best-fit mass increase from 0.3 to 3 μM (kDa)	best-fit effective RNP life-time (sec)
N (ancestral)	none	Wuhan-Hu-1	19.7	614	63.8	66.3
N:P13L	P13L	λ, o (all)				
N:Δ31-33	Δ31-33	o (all)				
N:P13L/Δ31-33	P13L, Δ31-33	o (all)	20.2	625	118.2	43.6
N:R203K/G204R	R203K, G204R	α, γ, λ, ζ, and o (except XEC)	19.5	596	70.3	54.6
N_o	P13L, Δ31-33, R203K, G204R	o (except XEC)	20.5	617	77.2	58.5
N:G215C	G215C (reduced)	δ (all 21J)	20.4	619	47.8	231
N:G215C*	G215C (oxidized)	δ (all 21J)	20.8	649	56.1	41.8
N_λ	P13L, R203K, G204R, G214C (reduced)	λ	21.0	669	20.0	67.4
N_λ*	P13L, R203K, G204R, G214C (oxidized)	λ	21.5	660	51.6	98.2
N:R203M	R203M	δ, κ				
$N_{210\text{-}419}$*	Δ1-209	o (not XEC)				

[1] Referring to the most common mutations in the VOCs and VOIs, excluding sporadic spontaneous reversions or other variations.

Figures

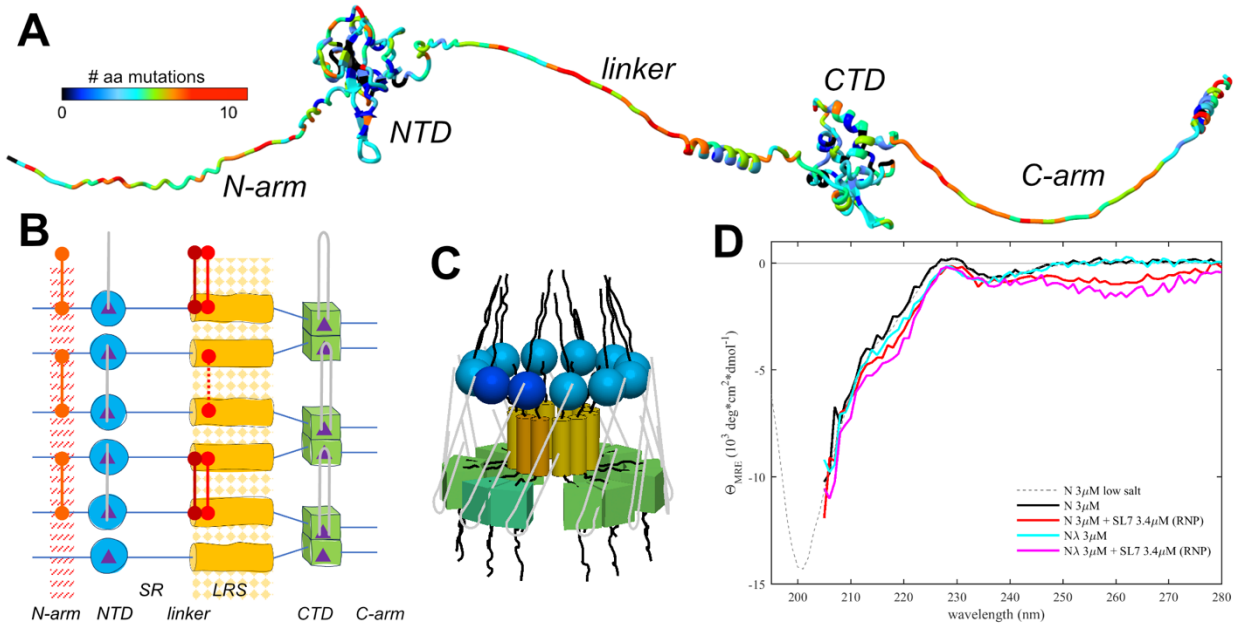

Figure 1. Basic organization of N-protein and RNPs. N-protein (1-419) has two folded domains, NTD (45-180) and CTD (248-363), and three IDRs including the N-arm (1-44), the central linker (181-247) and the C-arm (364-419). (A) Displayed is an AF2 structure where the disordered N-arm, linker, and C-arm are artificially stretched for clarity. The residues are color-coded according to the number of different amino acids that have been observed at this position in the mutational landscape replacing the Wuhan-Hu-1 sequence. The bioinformatic analysis was carried out as described, previously (*58*), updated to August 26, 2024, using a threshold of >5 genomes for each mutation. (B) Schematic of protein-protein and protein/RNA interfaces in RNP assembly. The dimerization domain (CTD) is indicated in green. The linker is subdivided in a serine and arginine-rich region (180-205, SR) and a L-rich region (206-247, LRS). LRS can transiently fold into helices (yellow) that create a hydrophobic patch for promiscuous self-association (indicated as yellow pattern). For clarity the cartoon only shows 3 neighboring N-protein dimers, although higher-order oligomers assemble in RNPs. Nucleic acid binding sites (purple triangles) preferentially bind single-stranded RNA at the NTD (grey lines), and double stranded RNA at the two sites per CTD dimer, with the ability to cross-link neighboring dimers potentially in various configurations. New inter-dimer interactions evolved in VOCs are indicated by red connectors, including the promotion of beta-sheet oligomerization through the N:P13L mutation in the N-arm (as in Omicron and Lambda variants), and the introduction of cysteines at the base of the LRS helices in N:G214C (as in Lambda variants) and N:G215C (as in Delta variants). (C) Three-dimensional cartoon of the circular organization of N-protein domains in RNPs, with one dimer shaded slightly darker to highlight the dimeric building blocks. For clarity, subunit sizes are not drawn to scale. (D) CD spectra of N-protein in the presence of SL7 under near physiological salt conditions leading to majority assembly of RNPs. Spectra are corrected for free SL7 contributions. Shown are spectra of ancestral N-protein alone (black), in the presence of SL7 forming RNPs (red), Nλ alone (cyan) and in the presence of SL7 forming RNPs (magenta). Spectra are truncated at < 205 nm due to limited buffer transparency. For comparison the dotted line shows a previously published CD spectrum of ancestral N-protein in low-salt buffer that permits measurement at shorter wavelengths (*25*).

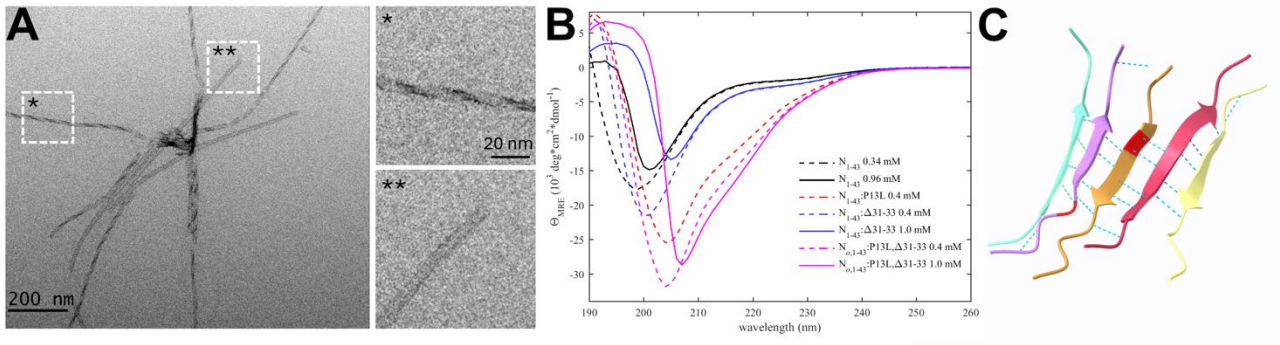

Figure 2. The P13L mutation creates self-association interfaces in the N-arm through stabilization of β-sheets. (A) Electron micrograph of negatively stained omicron N-arm $N_{o,1\text{-}43}$:P13L,Δ31-33 after equilibration at 10 μM in 20 mM HEPES, 150 mM NaCl, pH 7.50. The magnified regions are examples of twisted (*) and straight (**) fibrils. (B) CD spectra of $N_{1\text{-}43}$:P13L (red), $N_{1\text{-}43}$:Δ31-33 (blue), and Omicron $N_{o,1\text{-}43}$:P13L,Δ31-33 (magenta) at 0.4 mM (dashed) and 1.0 mM (solid), in comparison with ancestral N-arm (black). (D) Subset of ColabFold prediction of multimers of $N_{10\text{-}20}$:P13L highlighting hydrogen bonds. The P13L residue is highlighted in red in the middle peptide.

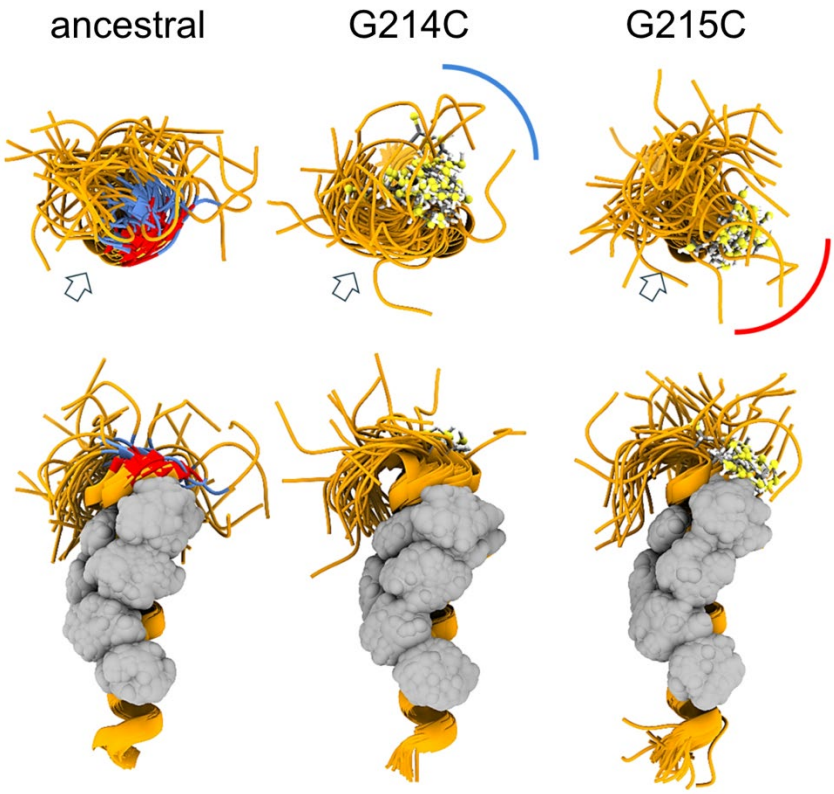

Figure 3. MD simulation of ancestral LRS and comparison with 214C and 215C mutants of Delta and Lambda variants. Snapshots at equal time intervals (4 ns) taken from the 200-ns MD simulations of the ancestral peptide $N_{210-246}$ and single-point G→C mutants (all monomers were oriented by overlying the helical region and displayed in orthosteric view). The upper row shows a view from the N-terminus side (with the helix axes perpendicular to the plane of the figure), while the lower row presents a side view (with axes on the plane). For clarity, the disordered C-terminal segment (residues 236-246) has been removed. The glycine and cysteine residues at positions 214 and 215 (colored blue and red, respectively, in the ancestral peptide and rendered as ball-and-stick in the mutants) restructure the flexible backbone around the GG motif into a well-defined α-helix turn, directing the sulfhydryl group in specific orientations (illustrated schematically by the blue and red arches). These orientations can be quantified relative to the Leu-rich central region (indicated by arrows and rendered as gray van der Waals spheres), which forms the hydrophobic interfaces of the oligomers (*58*). Under reducing conditions, this reorientation of the N-terminus relative to the helix can influence helix binding during the early stages of oligomerization or alter the conformation and physicochemical properties of the resulting oligomers, as illustrated in **Supplementary Figure S5**.

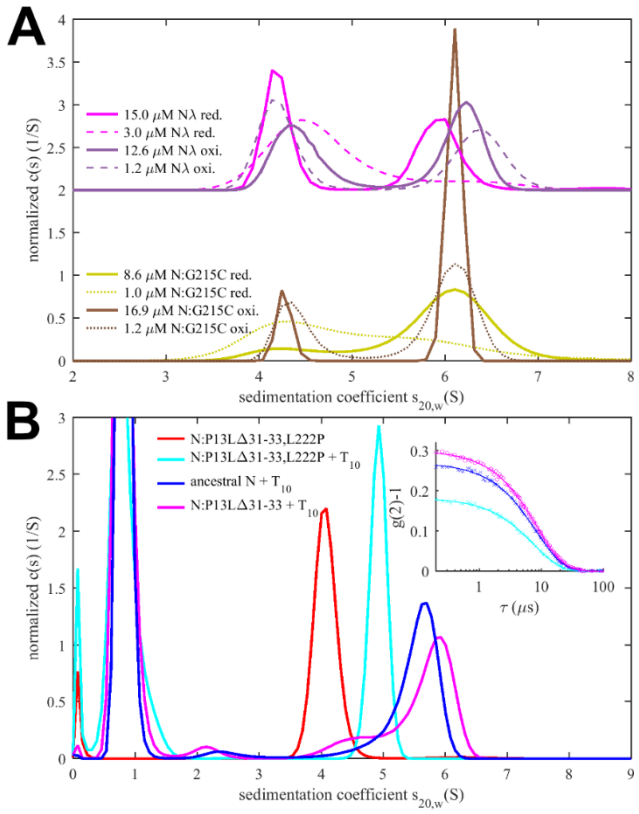

Figure 4. Impact of mutations in self-association interfaces on oligomeric state. (A) Sedimentation coefficient distributions of cysteine mutants N:G215C (reduced in yellow) and N:G215C* (oxidized in brown), as well as reduced N_λ (reduced in magenta) and N_λ* (oxidized in violet), with N_λ data offset by 2. For each sample data were acquired at high (solid lines) and low (dashed lines) concentration. (B) Sedimentation coefficient distributions of 2 μM N:P13L,Δ31-33, ancestral N, and N:P13L,Δ31-33,L222P in the presence of 10 μM T_{10} in low salt buffer. The inset shows DLS autocorrelation data of the same samples (symbols) and single-species fits (lines).

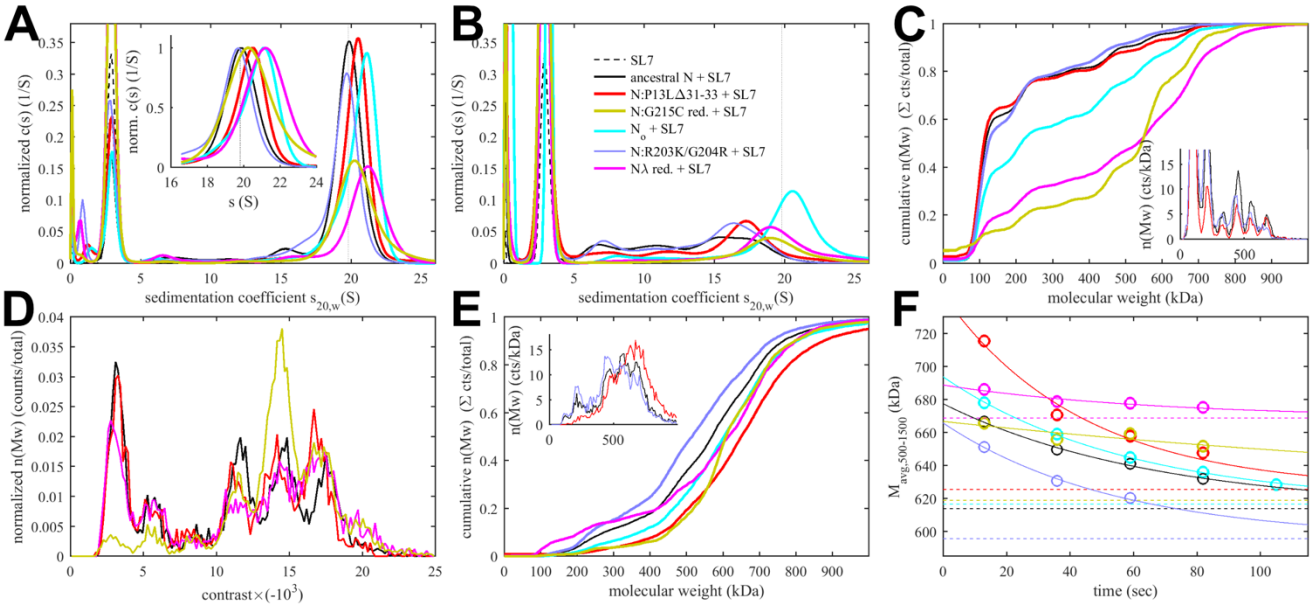

Figure 5. Size distributions and stability of ancestral and mutant RNPs. Show are SV (A,B) and MP data (C-F) for mixtures of N-protein with stem-loop RNA SL7 in molar ratio of 1(N):1.15(SL7) at high concentration of 3 μM (A,D) or low concentration of 0.3 μM (B,C,E,F) in reducing buffer conditions. All panels use the same color scheme for N-protein: ancestral (black), N:P13L,Δ31-33 (red), N:G215C (orange), N_o (cyan), N:R203K/G204R (blue), N_λ (magenta). (A) Sedimentation coefficient distributions of mixtures equilibrated at high concentration, with reaction boundary peaks magnified in the inset. For reference, the ancestral RNP s-value is drawn as dotted vertical line. Absorbance data are recorded at 260 nm and are weighted by SL7 content of sedimenting species. Higher reaction boundary s-values signify greater affinity or lifetime of the mutant RNPs. (B) Sedimentation coefficient distributions of the same samples as in (A), tenfold diluted and equilibrated, highlight dissociation of most RNPs into a range of intermediate size complexes. (C) MP experiments of equilibrated 0.3 μM mixtures. The measured number distributions are presented as cumulative distributions, which display higher percentages of large species as shifts to the right. Most samples are largely dissociated into dimers, with remaining peaks corresponding to populations of dimer to hexamers of $N_2/SL7_2$ subunits (as highlighted in the differential distributions in the inset). As an example for the resolution of distinct species, the inset shows the differential distribution (histogram) for N:R203K/G204R (blue), ancestral N (black), and N:P13L,Δ31-33 (red). (D) Mass distributions acquired in stopped-flow configuration applied to 3 μM mixtures. Larger (negative) contrasts correspond to higher molecular weights, with major peaks corresponding to species containing 1, 4, 5, and 6 $N_2/SL7_2$ subunits. (E) For kinetic experiments, mass distributions were acquired in different time intervals after tenfold dilution of 3 μM mixtures, here showing data collected from 3 sec to 23 sec. (F) Number-average masses of assembled RNPs between 500 and 1500 kDa observed in consecutive 20 sec data acquisition intervals after tenfold dilution of 3 μM mixtures (circles). The dashed horizontal lines are number-averages determined from the equilibrated 0.3 μM mixtures in (E). The solid lines are a best-fit single exponentials constrained to decay to the measured equilibrium values, yielding RNP lifetimes listed in **Table 1**.

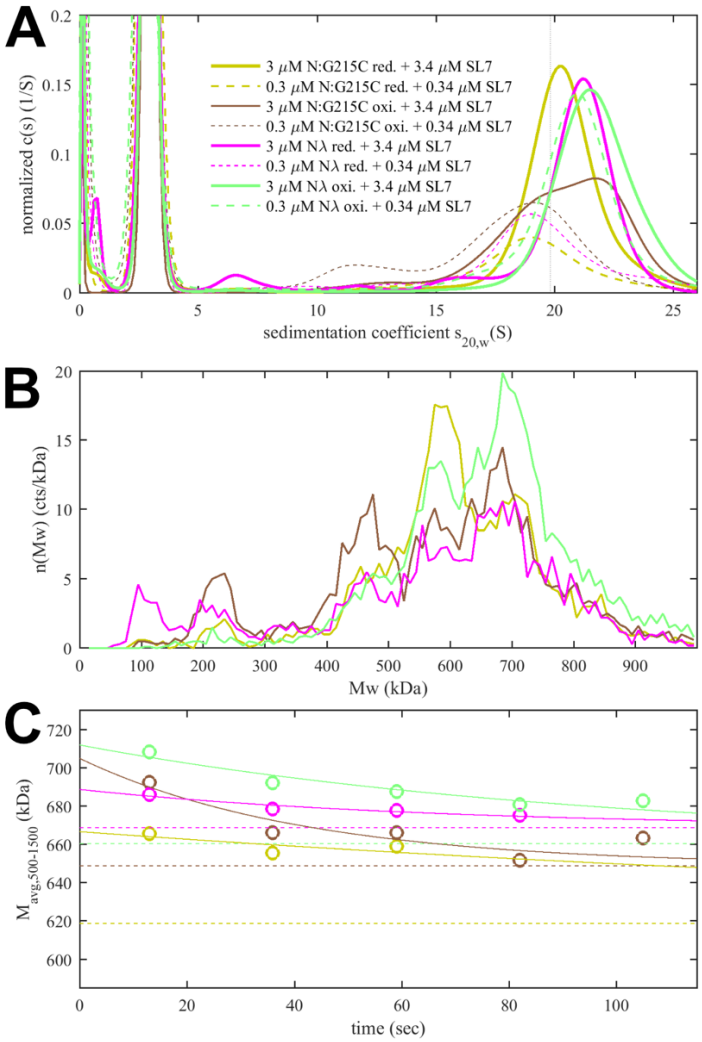

Figure 6. Impact of LRS disulfide bonds on the size and stability of RNPs. N-protein with cysteine mutations in the LRS were oxidized to form disulfide-linked oligomers (as shown in **Figure 4A**) and mixed with stem-loop RNA SL7 in molar ratio of 1(N):1.15(SL7). Shown are data for oxidized N:G215C* (brown) and oxidized N_λ* (green), and for comparison, reduced N:G215C (yellow) and reduced N_λ (magenta). (A) Sedimentation coefficient distributions at 3 μM (solid lines) and 0.3 μM (dashed lines) protein. (B) Molecular weight distributions in MP experiments of the same mixtures rapidly diluted to 0.3 μM protein, acquired from 3 – 23 sec after dilution. (C) Time-course of number-average RNP mass between 500 and 1,500 kDa (circles), determined from rapid dilution experiments in (B) for consecutive 20 sec data acquisition intervals. The solid lines are best-fit single-exponential decays constrained to attain the separately measured equilibrium values at 0.3 μM protein (dashed lines), with lifetimes listed in **Table 1**.

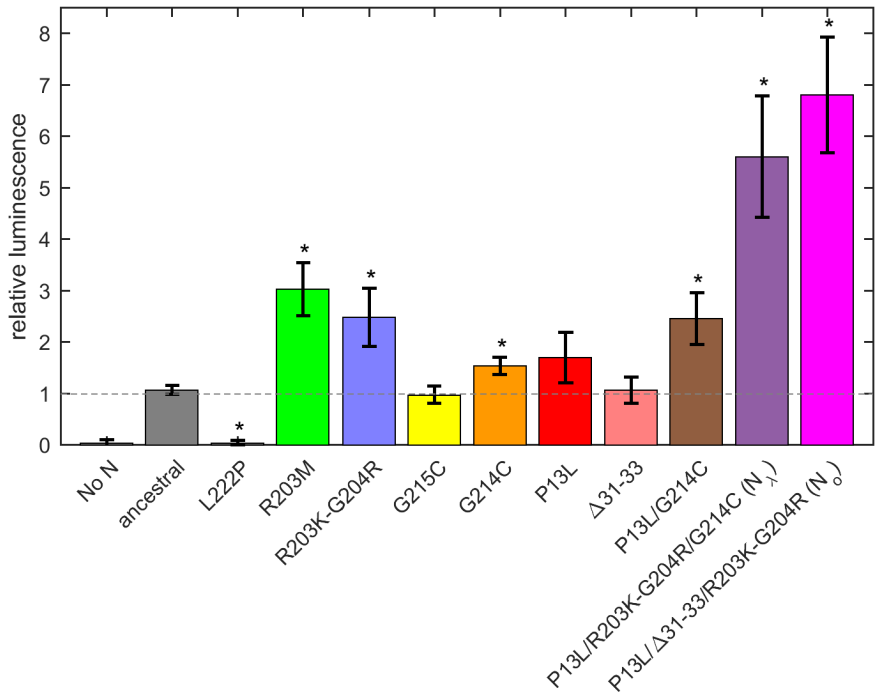

Figure 7. Mutation effect on packaging and cell entry in a VLP assay. Error bars are standard deviations from n = 4. Stars indicate significance (P > 0.95) of a two-sided Kolmogorov-Smirnov test comparing the control ancestral measurements with mutants.

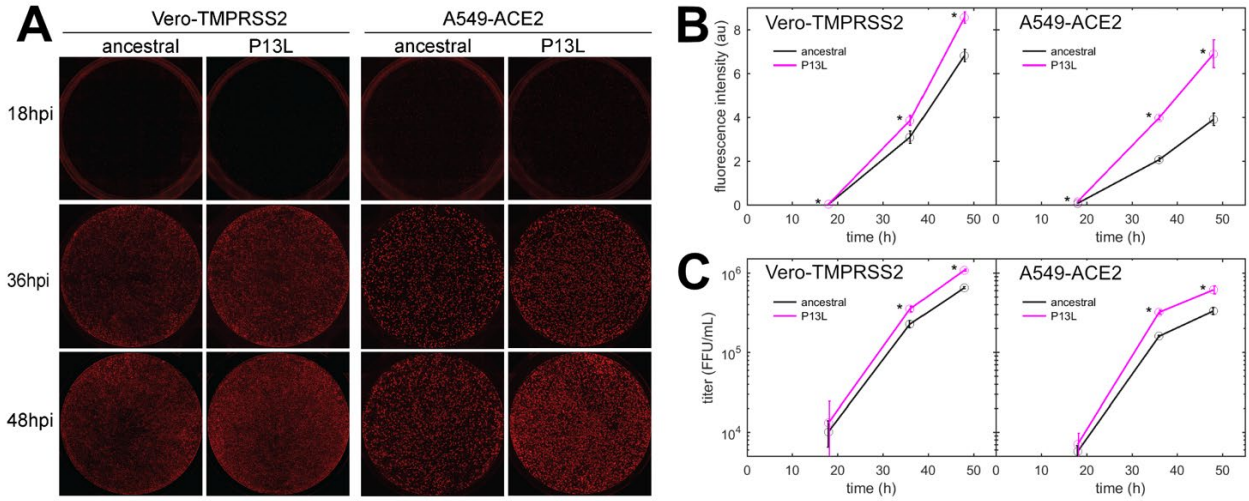

Figure 8. Replication kinetics of recombinant SARS-CoV-2 reporter viruses in cell lines. (A) Representative images of Vero-TMPRSS2 and A549-ACE2 cells infected with SARS-CoV-2 P13L or WT at different time points post-infection. (B) Quantification of fluorescence intensity from P13L and WT virus infections shown in (A). (C) Viral titers in the supernatant from infected cells. Error bars are standard deviations (n = 3), and stars indicate significant differences on a P=0.95 confidence level.