

# Supplementary Information for

## **LLPS REDIFINE allows the biophysical characterization of multicomponent condensates without tags or labels**

Mihajlo Novakovic<sup>1,\*</sup>, Nina Han<sup>1,2</sup>, Nina C. Kathe<sup>1,2</sup>, Yinan Ni<sup>1,2</sup>, Leonidas Emmanouilidis<sup>1,\*</sup>,  
Frédéric H.-T. Allain<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich, Switzerland

<sup>2</sup>These authors contributed equally

\*Corresponding authors. Email: mihajlo.novakovic@bc.biol.ethz.ch  
leonidas@bc.biol.ethz.ch, allain@bc.biol.ethz.ch

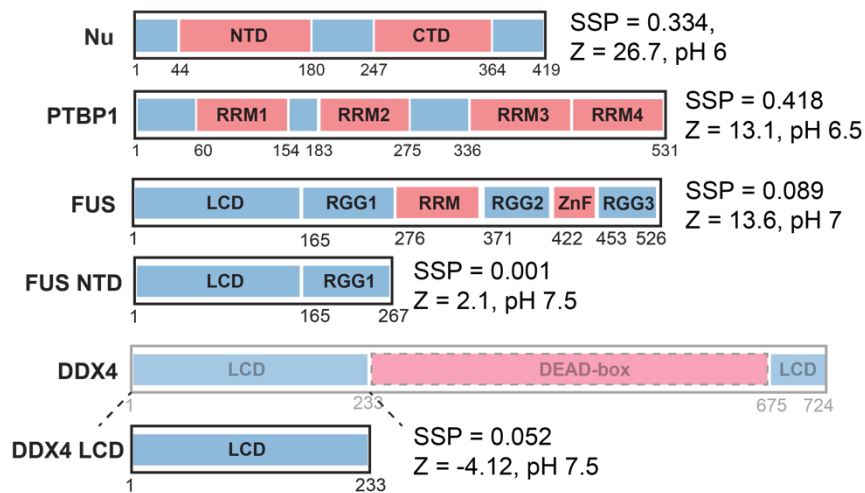
This Supplementary Information PDF contains:

Supplementary Figures

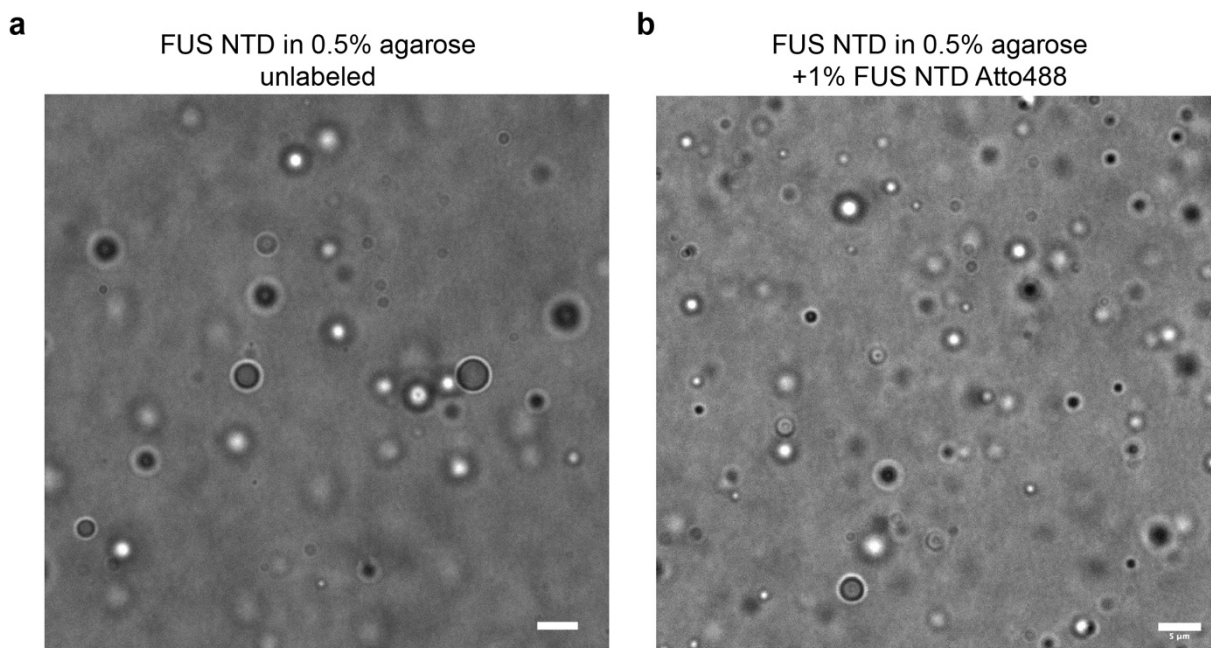
Supplementary Tables

Supplementary Notes

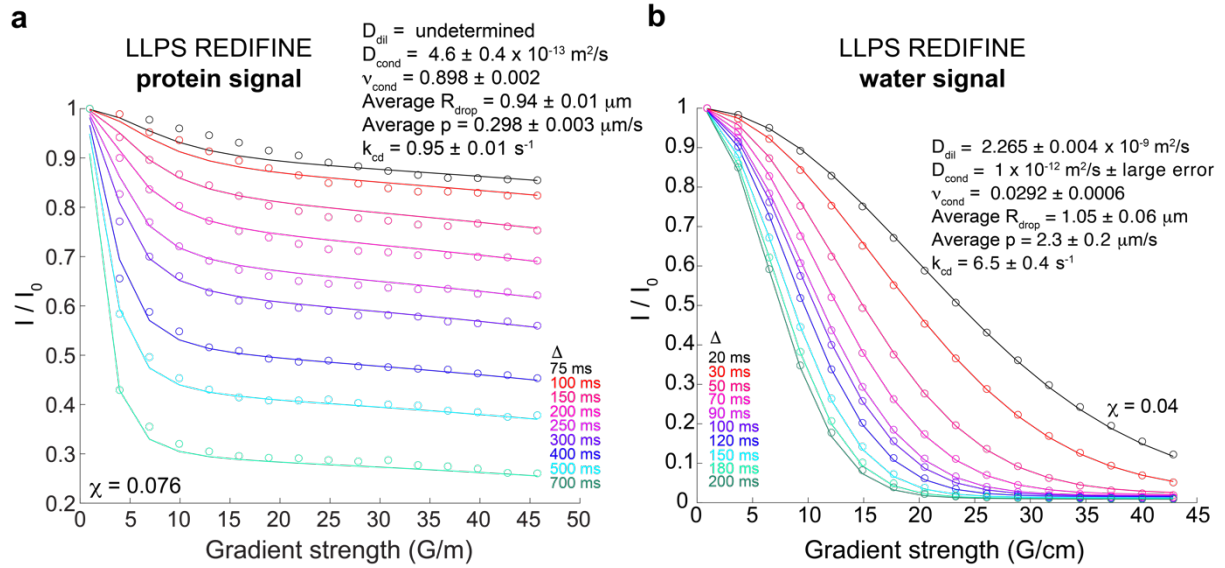
Supplementary References



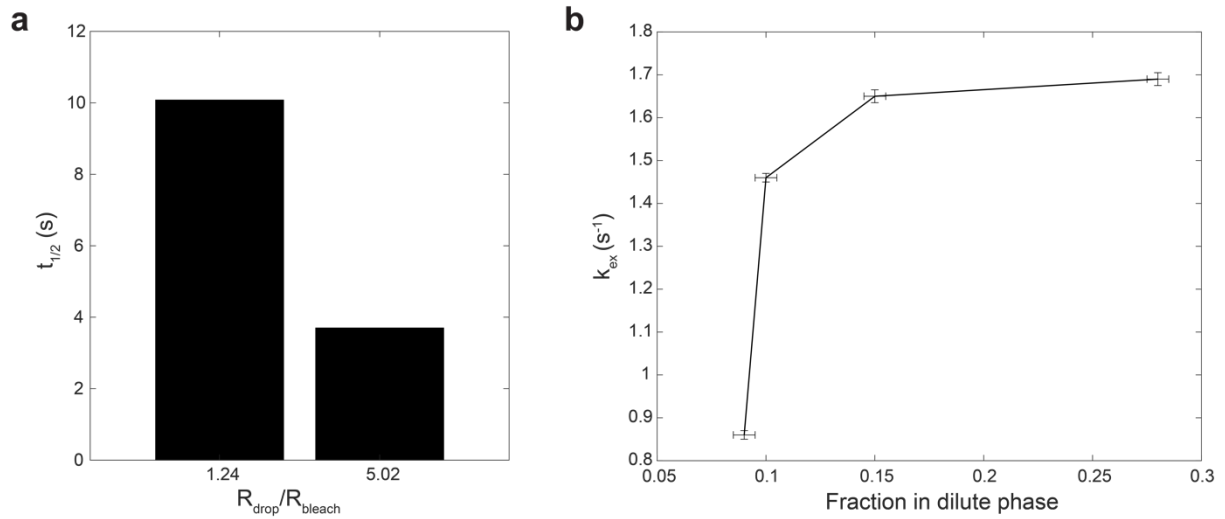
**Supplementary Fig. 1: Domain schemes of proteins utilized in this study.** **a** Primary sequence of proteins used in this study. Nucleocapsid protein consists of structured NTD and CTD domains flanked with 3 IDRs. PTBP1 is mostly structured containing 4 RRM domains, though still containing long disordered regions. A large part of FUS protein is also disordered except RRM motif and partially folded zinc-finger. DDX4 N terminal region that was used in this study is also completely disordered (low complexity domain - LCD). Secondary structure prediction factor SSP and protein net charge Z at the pH used in this study are shown next to each protein sequence. Note the large structural differences between Nu or PTBP1 and FUS, which all classify as RNA binding proteins.



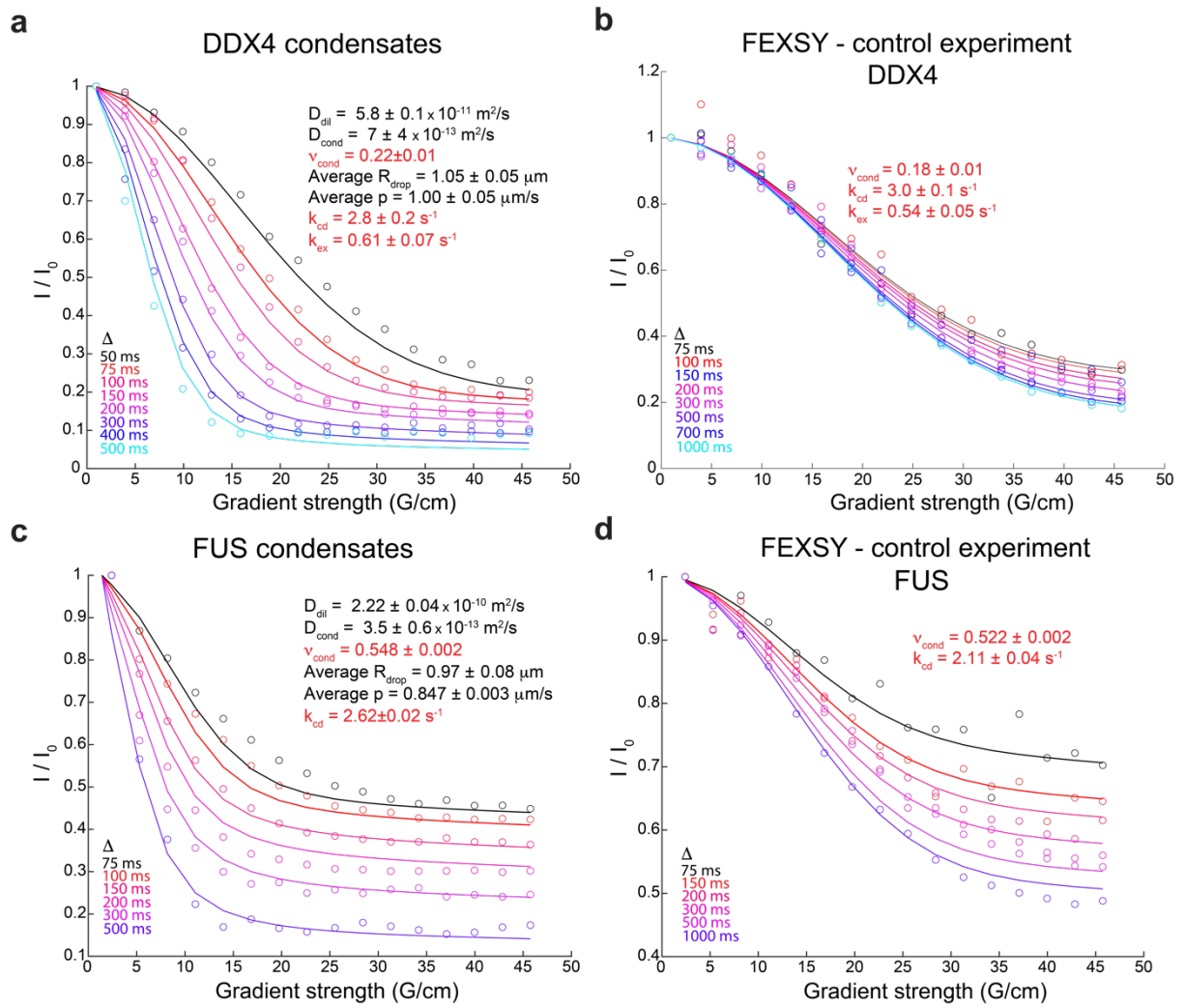
**Supplementary Fig. 2: Bright field images of FUS NTD in 0.5% agarose hydrogel.** **a** Bright field microscopy image acquired on 100  $\mu$ M unlabeled FUS NTD vs. **b** FUS NTD spiked with 1% protein labeled with Atto488 dye. The scale bar is 5  $\mu$ m.



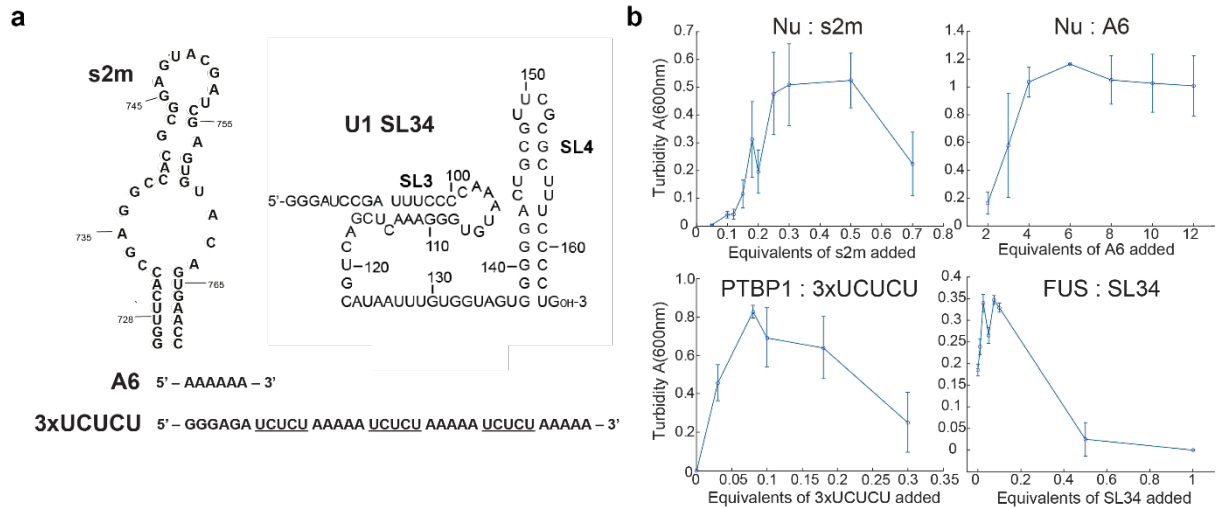
**Supplementary Fig. 3: LLPS REDIFINE characterization of FUS FL condensates.** **a** REDIFINE data set and ensuing fitting parameters determined for FUS FL condensates. **b** Analogous results acquired by integrating water signal in the same FUS FL sample. Experiments in **a** are acquired using 10 ms gradient length while 2 ms was used in in **b**. Large uncertainty of the fitted values is present for  $D_{dil}$  of protein and  $D_{cond}$  of water due to their small corresponding populations. The final values for minimization function  $\chi$  indicating the goodness of the fit are shown in both panels. The uncertainties for each parameter are calculated using covariance matrix and report on the ambiguity of determined values.



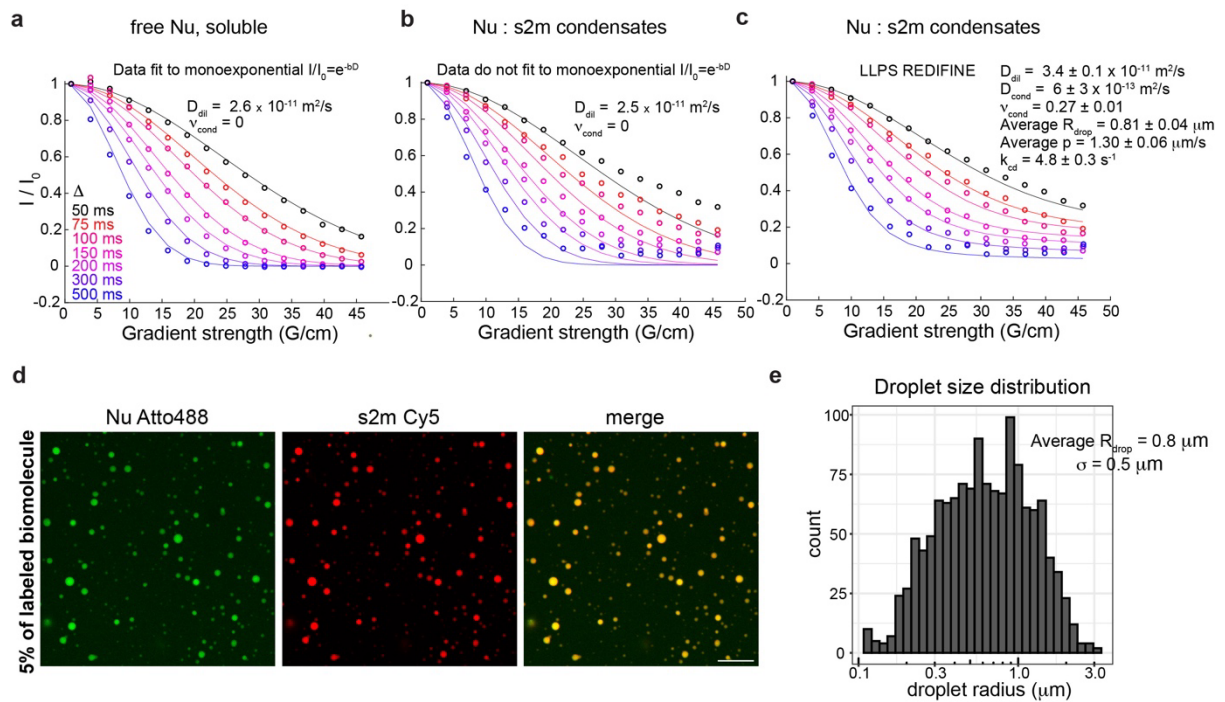
**Supplementary Fig. 4: FRAP recovery and chemical exchange depend on the partitioning coefficient between two phases.** **a** FRAP recovery rates of 1% mCherry-labeled FUS FL are influenced by the ratio of droplet radius and bleached area. Smaller droplets recover considerably slower than larger droplets suggesting that mCherry-labeled FUS FL accumulate mostly inside the droplets. When the whole droplet is bleached there is not enough molecules in dilute phase to recover the bleached droplet. **b** Along the same lines, chemical exchange determined by REDIFINE (Supplementary Table 1) is affected by the fraction of FUS FL protein residing in the dilute phase. We can see that chemical exchange rates considerably drop for FUS FL fractions below 0.1. This confirms the scenario where there is not enough labeled FUS molecules in the dilute phase making FRAP recovery rates apparently slower. The uncertainties for each parameter are calculated using covariance matrix and report on the ambiguity of determined values.



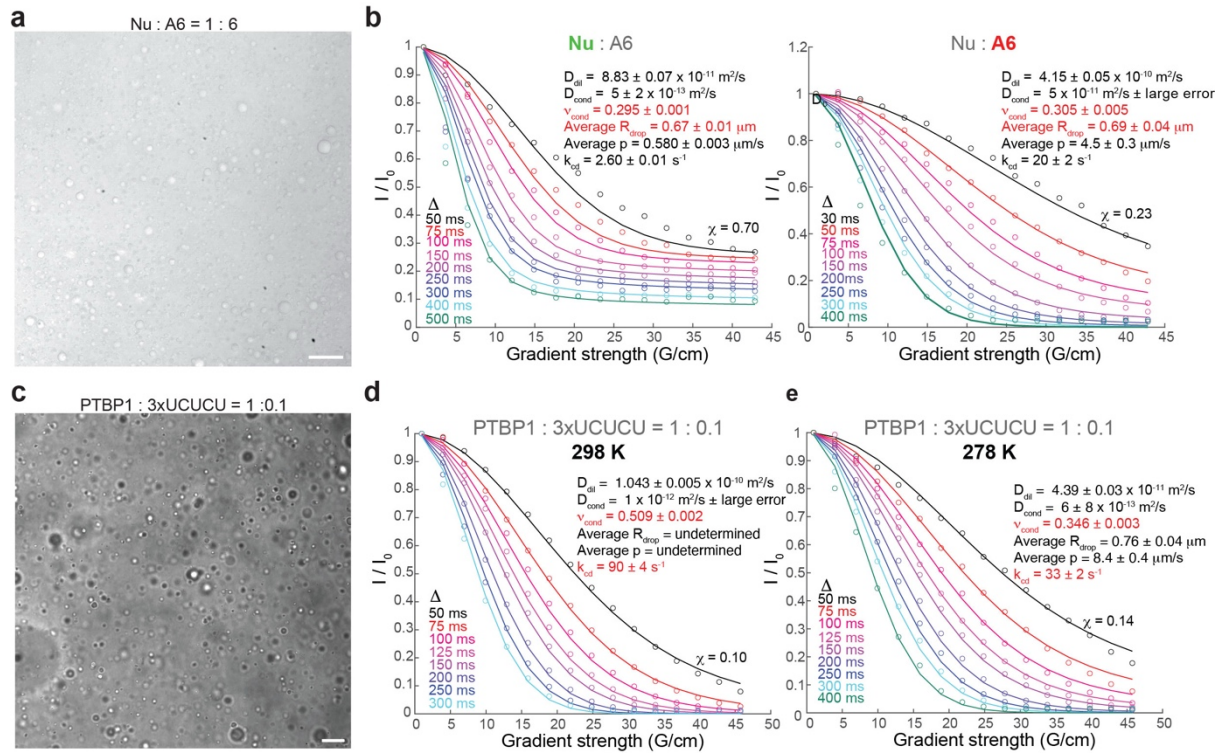
**Supplementary Fig. 5: Validation of LLPS REDIFINE by the complementary FEXSY experiment adapted to condensates.** **a** LLPS REDIFINE acquired on a fully unstructured DDX4 low-complexity domain. **b** Complementary FEXSY experiment adapted to biomolecular condensates and acquired on the same disordered N-terminal domain of DDX4 validating the chemical exchange between the two phases as determined by LLPS REDIFINE. **c,d** Similar analysis was carried out on FUS FL sample that was incubated for 2 months (the sample was still not fibrillar). Very similar exchange rate and populations are determined by both methods. Note that FEXSY experiment can provide exchange information between the two phases, but it is limited to very slow chemical exchange and significant populations of both phases. By utilizing two diffusion encoding blocks and requiring presaturation as a water suppression scheme, the FEXSY experiment is very insensitive and its application is limited by the poor SNR in NMR measurements. Experiments on DDX4 and FUS took almost 5-fold longer acquisition time compared to LLPS REDIFINE. Both proteins were prepared at the concentration of 150  $\mu\text{M}$ . The uncertainties for each parameter are calculated using covariance matrix and report on the ambiguity of determined values.



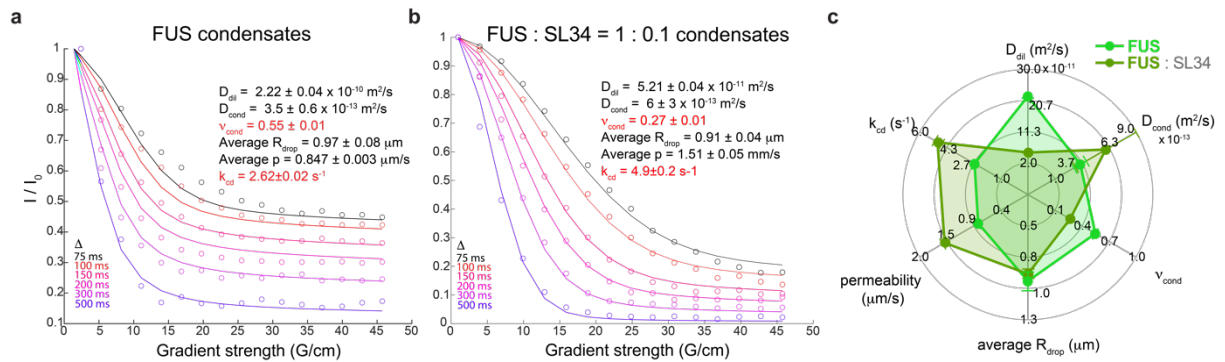
**Supplementary Fig. 6: RNA induces/modulates the phase separation of RBPs.** **a** The secondary structure of s2m and U1 SL34 as well as the primary sequence of single-stranded A6 and 3xUCUCU used for phase separation. **b** Turbidity assays illustrating the phase separation diagrams of Nu, PTB and FUS proteins with respect to their corresponding RNA concentrations. Note that FUS is phase separated even in the absence of RNA. Error bars are the standard deviation of at least 3 technical replicates.



**Supplementary Fig. 7: LLPS REDIFINE provides the condensate properties even when condensed phase plateau is not observed in PGSTE curves.** **a-c** LLPS REDIFINE acquired on dispersed Nu and biphasic Nu condensates. While curves of free Nu can be modeled by mono-exponential decay, the data acquired on Nu condensates can be explained only by LLPS REDIFINE method that enabled determination of the parameters of Nu:s2m condensate. **d** Nucleocapsid colocalize with s2m in the condensed droplets as visualized using fluorescently labeled protein and RNA. Scale bar is 10 μm. **e** Analysis of the size distribution show the mean size of 0.8 μm consistent with LLPS REDIFINE measurements. FRAP experiments on Atto488-labeled Nu could not be performed due to very low bleaching efficiency. The uncertainties for each parameter in REDIFINE are calculated using covariance matrix.



**Supplementary Fig. 8: RNA-induced formation of liquid droplets.** **a** Brightfield microscopy images of Nu:A6 prepared at 110  $\mu\text{M}$  protein concentration in 0.5% agarose. Both protein and RNA peaks are visible in NMR spectrum and are analyzed separately by LLPS REDIFINE. **b** Restricted diffusion of protein and RNA molecules reports the same size of the droplets with significantly different interface permeabilities and diffusion rates for protein and RNA. **c** PTBP1:3xUCUCU condensed samples prepared at 150  $\mu\text{M}$  protein concentration in 0.5 % agarose hydrogel **d** LLPS REDIFINE on condensed PTBP1:3xUCUCU sample acquired at 298 K and **e** 278K. Unambiguous parameters were determined only at 278K when exchange between phases is significantly slower. The final values for minimization function  $\chi$  indicating the goodness of the fit are shown in each panel. The uncertainties for each parameter are calculated using covariance matrix and report on the ambiguity of determined values. Scale bar in **a** is 25  $\mu\text{m}$  and in **c** 10  $\mu\text{m}$ .



**Supplementary Fig. 9: FUS condensates mature differently in presence of specific SL34 RNA.** **a** FUS condensate prepared at 150  $\mu\text{M}$  FUS concentration recorded after two months of incubation at room temperature. **b** FUS condensate prepared at 150  $\mu\text{M}$  FUS and 15  $\mu\text{M}$  RNA concentrations prepared and recorded at the same times as in **a**. In the presence of SL34 RNA, there is less FUS in the droplet phase and it is more dynamic (faster  $D_{\text{cond}}$  and  $k_{\text{cd}}$ ). **c** Summary of the differences induced by the incubation in presence of RNA.

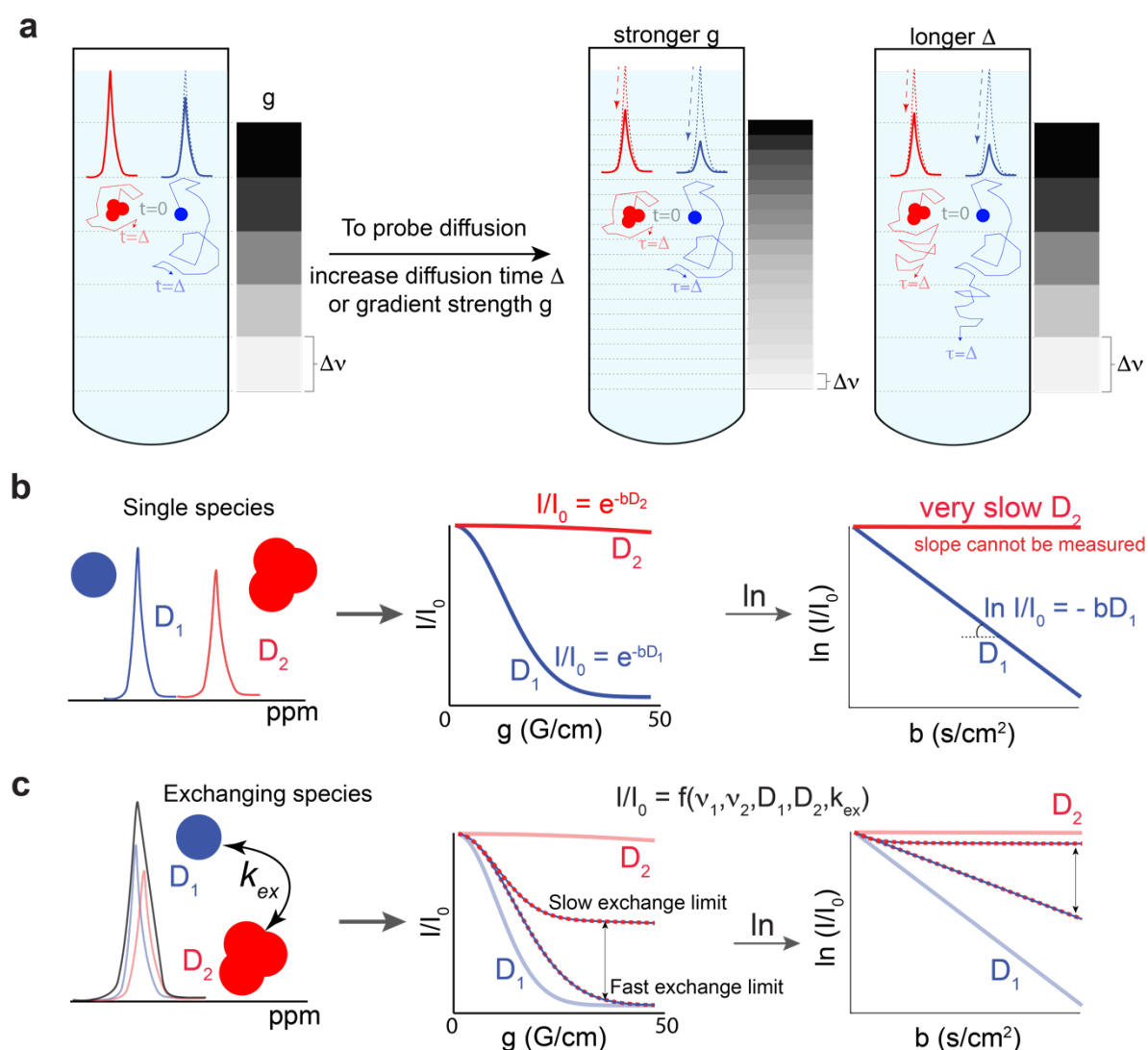
**Supplementary Table 1: LLPS REDIFINE results for the condensates examined in this study.**

	Condensate	$D_{dil}$ (m <sup>2</sup> /s)	$D_{cond}$ (m <sup>2</sup> /s)	Population $\nu$	$R_{drop}$ (μm)	Permeability $p$ (μm/s)	$k_{cd}$ (s <sup>-1</sup> )	$k_{ex}$ (s <sup>-1</sup> )
Structured proteins	Nu : ATP	6.4±0.1 ×10 <sup>-11</sup>	6×10 <sup>-13</sup>	0.36±0.01	0.43±0.02	1.62±0.08	11.3±0.6	4±0.3
	Nu : GTP	9.6±0.2 ×10 <sup>-11</sup>	6±9×10 <sup>-13</sup>	0.32±0.01	0.72±0.03	1.47±0.05	6.2±0.3	2.0±0.1
	Nu : GTP	5.65±0.02 ×10 <sup>-10</sup>	3×10 <sup>-11</sup>	/	/	/	69±3	/
	Nu : A6	8.83±0.07 ×10 <sup>-11</sup>	5±2×10 <sup>-13</sup>	0.29±0.01	0.67±0.01	0.58±0.01	2.6±0.1	0.77±0.01
	Nu : A6	4.15±0.05 ×10 <sup>-10</sup>	5×10 <sup>-11</sup>	0.30±0.01	0.69±0.04	4.5±0.3	20±2	6.1±0.06
	Nu : s2m	3.4±0.1 ×10 <sup>-11</sup>	6±3×10 <sup>-13</sup>	0.27±0.01	0.81±0.04	1.30±0.06	4.8±0.3	1.30±0.07
	Nu : polyA s2m, matured	1.33±0.03 ×10 <sup>-10</sup>	/	0.22±0.01	0.56±0.04	2.6±0.2	14±1	3.1±0.3
	PTBP1 : 3xUCUCU, 278 K	4.39±0.03 ×10 <sup>-11</sup>	6±8×10 <sup>-13</sup>	0.35±0.01	0.76±0.04	8.4±0.4	33±2	11.4±0.7
Disordered proteins	DDX4	5.8±0.1 ×10 <sup>-11</sup>	7±4×10 <sup>-13</sup>	0.22±0.01	1.05±0.05	1.00±0.05	2.8±0.2	0.62±0.05
	FUS, 66 μM	1.2±0.1 ×10 <sup>-10</sup>	5.8±0.5 ×10 <sup>-13</sup>	0.72±0.01	0.97±0.01	0.76±0.01	2.35±0.02	1.69±0.03
	FUS, 106 μM	2.2±0.1 ×10 <sup>-10</sup>	6.4±0.4 ×10 <sup>-13</sup>	0.85±0.01	1.01±0.01	0.65±0.01	1.95±0.02	1.65±0.03
	FUS, 160 μM	/	4.6±0.3 ×10 <sup>-13</sup>	0.90±0.01	1.16±0.01	0.63±0.01	1.62±0.02	1.46±0.02
	FUS, 200 μM	/	4.6±0.4 ×10 <sup>-13</sup>	0.90±0.01	0.94±0.01	0.30±0.01	0.95±0.01	0.86±0.02
	FUS, 200 μM water	2.26±0.01 ×10 <sup>-9</sup>	1×10 <sup>-12</sup>	0.029±0.001	1.05±0.06	2.3±0.2	6.5±0.4	0.19±0.02
	FUS, 150 μM, 2 months old	2.22±0.04 ×10 <sup>-10</sup>	3.5±0.6 ×10 <sup>-13</sup>	0.55±0.01	0.97±0.08	0.85±0.01	2.62±0.02	1.43±0.02
	FUS : SL34, 150 μM, 2 months old	5.21±0.04 ×10 <sup>-11</sup>	6±3×10 <sup>-13</sup>	0.27±0.01	0.91±0.04	1.51±0.05	4.9±0.2	1.32±0.06
	FUS, 150 μM, 6 months old	8.38±0.08 ×10 <sup>-11</sup>	3.9±0.8 ×10 <sup>-13</sup>	0.27±0.01	0.92±0.06	0.83±0.01	2.72±0.02	0.74±0.01
	FUS NTD, 600 μM	8.0±0.2 ×10 <sup>-11</sup>	8.9±0.7 ×10 <sup>-13</sup>	0.75±0.01	1.21±0.03	0.40±0.01	1.00±0.03	0.75±0.03
	FUS NTD, 200 μM	8.5±0.4 ×10 <sup>-11</sup>	1.6±0.2 ×10 <sup>-12</sup>	0.65±0.01	1.19±0.05	0.27±0.04	0.67±0.08	0.44±0.05
	FUS NTD, 200 μM	7.9±0.1 ×10 <sup>-11</sup>	1.4±0.2 ×10 <sup>-12</sup>	0.66±0.01	1.23±0.03	0.25±0.01	0.62±0.02	0.41±0.02
	FUS NTD, 200 μM water	2.23±0.01 ×10 <sup>-9</sup>	7×10 <sup>-13</sup>	0.026±0.001	1.07±0.06	2.4±0.2	6.7±0.6	0.18±0.02

# 1. NMR diffusion experiments with complex biomolecular systems

In the traditional NMR diffusion experiment, a signal of an analyte is mapped as a function of the gradient amplitude applied in the pulsed gradient spin echo (PGSE) experiment.<sup>1,2</sup> The use of gradients can spatially label molecules depending on their position in the NMR tube. Their translational diffusion will lead to a signal attenuation that depends on experimental parameters and diffusion rates (Methods, Supplementary Fig. 10a) and it follows a simple mono-exponential decay (Supplementary Fig. 10b, Methods - Eq. 1) providing a quantitative measure of the displacement. Often, this decay is visualized on a semi-logarithmic plot (Supplementary Fig. 10b, right), where a slope of the line simply represents the diffusion coefficient (Methods - Eq. 3). Note that due to technical limitations of the applied gradient strength, very slow diffusion rates such as those associated with molecular complexes and biomolecules in a viscous environment ( $D < 5 \times 10^{-12} \text{ m}^2/\text{s}$ ) cannot be measured on conventional NMR probes and appear as flat non-decaying curves (Supplementary Fig. 10b).

Generally, proteins and large RNAs inside the viscous droplets undergo very slow diffusion. However, when multiple exchanging species are present in the sample (or molecules coexisting in multiple states) with partially or fully overlapping NMR resonances, the PGSE curve exhibits a complex decay affected by both species (Supplementary Fig. 10c). The chemical exchange imparts the information about the apparently invisible diffusion coefficient  $D_2$  onto



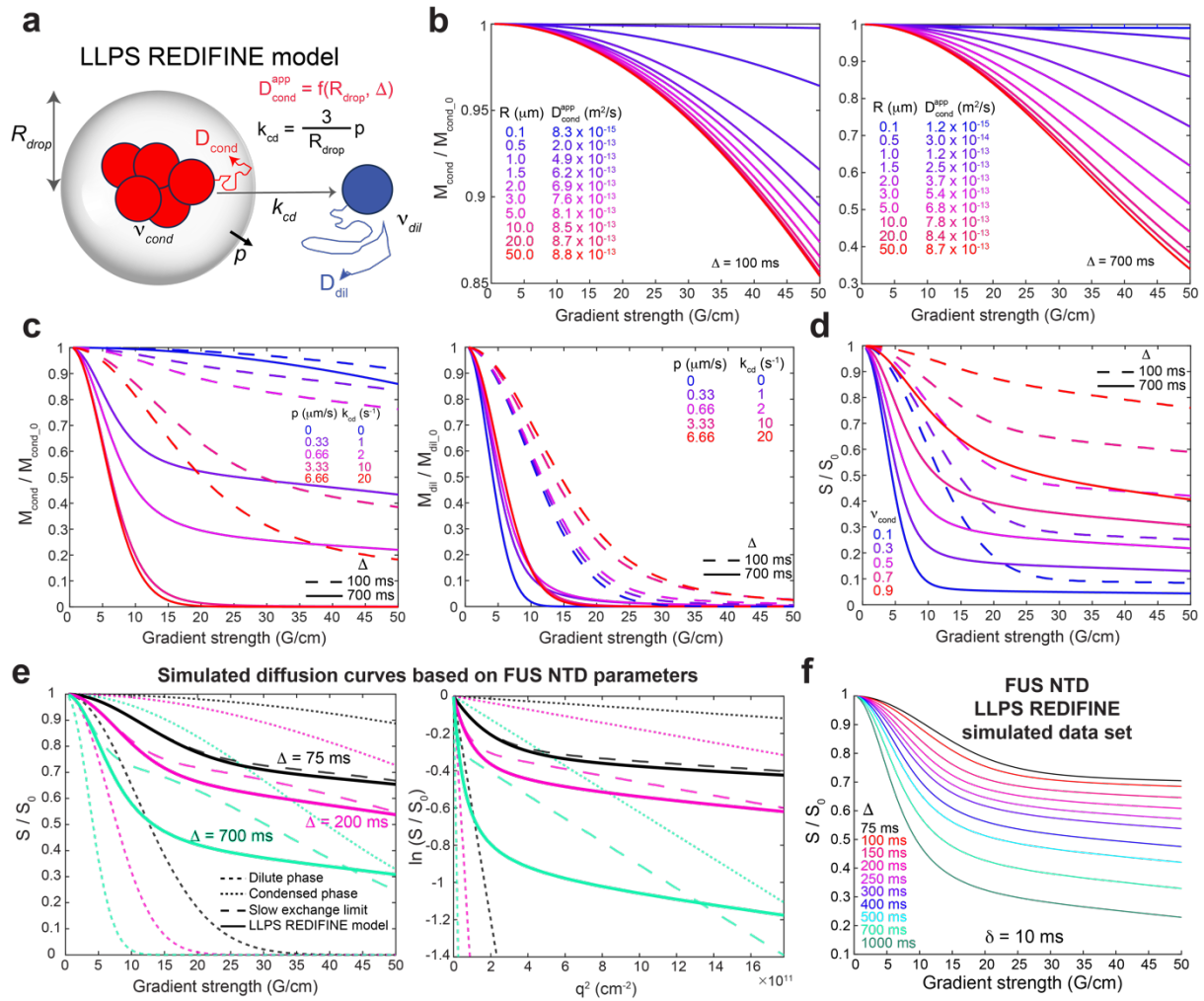
**Supplementary Fig. 10: Principles of diffusion NMR experiments and the effect of exchange in complex systems.** **a** Application of a gradient along  $z$ -axis can effectively “divide” an NMR sample in slices along the same axis. Translational diffusion of a molecule to other neighboring slices will manifest as the signal decay. To quantify diffusion coefficient, traditionally one need to acquire multiple spectra for increasing gradient strength (which makes thinner slices) until all molecules left their initial slice. Note that also increasing diffusion time can yield the same effects since a molecule is allowed longer time to diffuse. **b** Simulated attenuation of the PGSE signal due to a diffusion of single species in solution decays mono-exponentially and appears as a line on semi-logarithmic plot. Very slow diffusion coefficients cannot be probed at most conventional instruments and appear flat in PGSE experiments. **c** The presence of multiple species exchanging with each other encodes an exchange-averaged diffusion information on the NMR diffusion decay. The observed diffusion curve will strongly depend on the exchange rate. Note the large deviation of the resulting diffusion curve from both dilute and condensed phase diffusion. Diffusion coefficients are randomly chosen to be  $D_1 = 9 \times 10^{-11} \text{ m}^2/\text{s}$  and  $D_2 = 3 \times 10^{-13} \text{ m}^2/\text{s}$ , illustrating the typical 2 orders of magnitude diffusivity difference in biomolecular condensates.

the observed diffusion curve, resulting in a trajectory appearing between the slow and fast exchange limits. On the semi-logarithmic plot (Supplementary Fig. 10c, right) this exchange-averaged diffusion decay (defined with an apparent diffusion coefficient) deviates from the diffusion of both dilute and condensed populations not allowing diffusion constants to be determined.

## 2. The framework of LLPS REDIFINE

The foundation of LLPS REDIFINE is illustrated by the model in Supplementary Fig. 11a. The model of chemical exchange equilibrium between different-diffusing species shown in Fig. 1c can be incorporated within the framework of the McConnell equations<sup>3,4</sup> (Methods). Due to restricted diffusion, we defined an apparent diffusion coefficient in a condensed phase that depends on droplet size and the choice of diffusion time (Methods). The apparent diffusion constants and their effect on diffusion in the condensed phase are shown in Supplementary Fig. 11b. For small droplet radiuses and long diffusion times, the apparent diffusion coefficient can be almost 3 orders of magnitude slower than the intrinsic diffusion coefficient while approaching its values for large radiuses emulating almost continuous condensed phase. The influence of permeability and resulting chemical exchange on the PGSE curves of molecules in condensed and dilute phases is presented in Supplementary Fig. 11c. Although the slow diffusion of protein in the condensed phase cannot be accurately measured by NMR PGSE experiment on conventional probes, the chemical exchange effectively mixes the information about the restricted diffusion with the NMR PGSE visible diffusion of the molecules in dilute phase. This leads to apparently faster decays allowing us to extract the information about condensed phase. What is more, NMR diffusion curves are strongly affected by the population of the protein in the condensed phase (Supplementary Fig. 11d). Using our model, we could simulate the behavior of the FUS NTD protein data shown in Fig. 1b to the finest details. Supplementary Fig. 11e shows the importance of including droplet size and permeability. Note the difference between considering two diffusion coefficients in slow exchange (typical for FUS) and the LLPS REDIFINE model. Using the former, diffusion coefficients of molecules in condensed and dilute phases and their populations could not even be approximated. Our simulations show that the PGSE curves are also significantly influenced by the choice of diffusion encoding time. From simulation results, we concluded that multiple diffusion decay curves spanning from short (30-50 ms) to long diffusion times (700-1000 ms), each carrying unique information about the chemical exchange and restricted diffusion (Supplementary Fig.

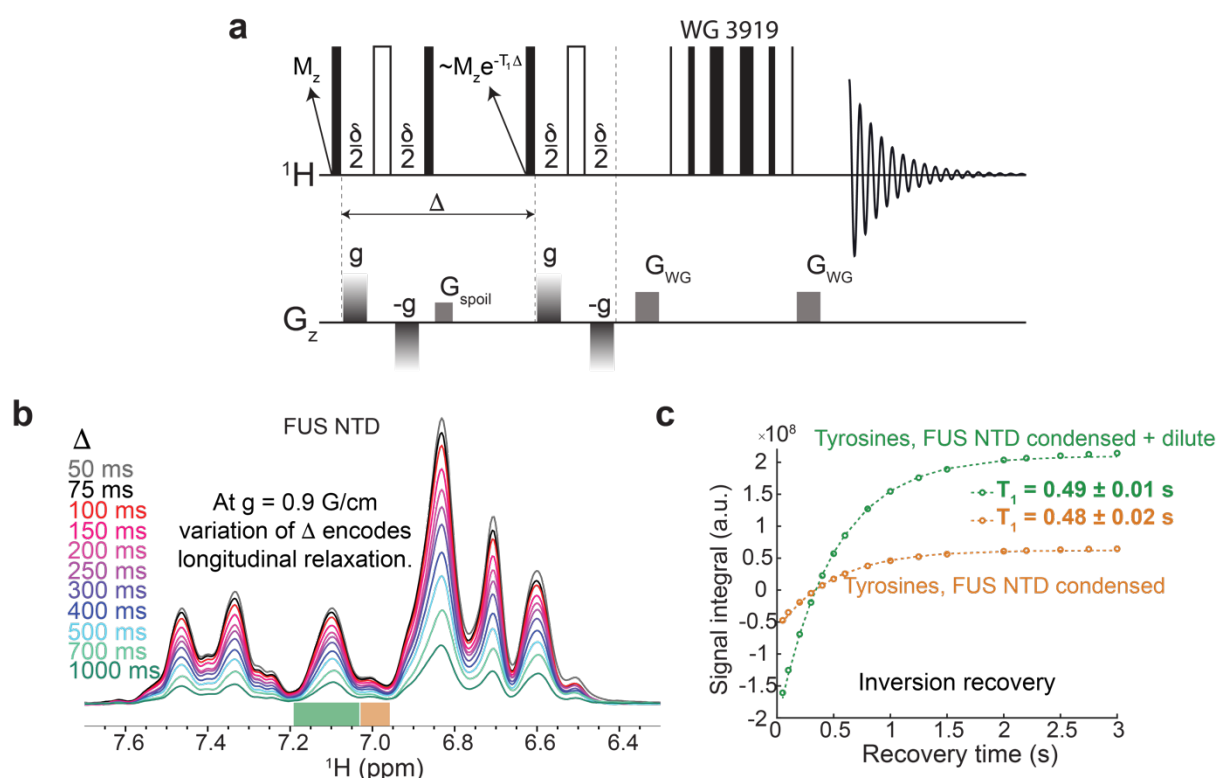
11f), are required to unambiguously describe all the parameters of biphasic condensates presented in the model. It is this rich set of diffusion information that allows LLPS REDIFINE to quantitatively interpret the experimental data acquired on biomolecular condensates.



**Supplementary Fig. 11: Translating the restricted diffusion model with exchange into LLPS REDIFINE – simulations.** **a** The illustration of the REDIFINE model considering the effects of multiple parameters including corresponding phases' diffusion coefficients and populations, permeability and droplet size on the diffusion behavior in condensates. **b** Influence of droplet size on apparent diffusion in condensed phase simulated for two diffusion times, 100 and 700 ms. To isolate only the effect of droplet size, permeability is put close to zero (0.000333 μm/s). **c** Effect of permeability and resulting chemical exchange on the diffusion of molecules in condensed and dilute phase. Condensed phase decay becomes apparently faster with increasing exchange while opposite is true for the dilute phase. This is effective mixing of the information between phases.  $R_{drop}$  is chosen to be 1 μm. **d** Influence of populations on the observed diffusion curves.  $R_{drop}$  and  $p$  were 1 μm and 0.33 μm/s respectively. **e** Reproducing the diffusion curves observed for FUS NTD in Fig. 1b. Note that individual diffusion coefficients and populations cannot be approximated from the diffusion data without implementing LLPS REDIFINE model. **f** Full REDIFINE data sets including multiple diffusion curves acquired for different diffusion times. The data set with at least 6 curves spanning from short (50-75 ms) to long diffusion times (500-1000 ms) is needed to uniquely define the 5 parameters of the system ( $D_{dil}$ ,  $D_{cond}$ ,  $v_{drop}$ ,  $R_{drop}$ , and  $p$ ). The diffusion coefficients used in this simulations are chosen to match the ones of FUS NTD,  $D_{dil} = 8 \times 10^{-11}$  m<sup>2</sup>/s and  $D_{cond} = 8.9 \times 10^{-13}$  m<sup>2</sup>/s.

### 3. LLPS REDIFINE – experimental considerations

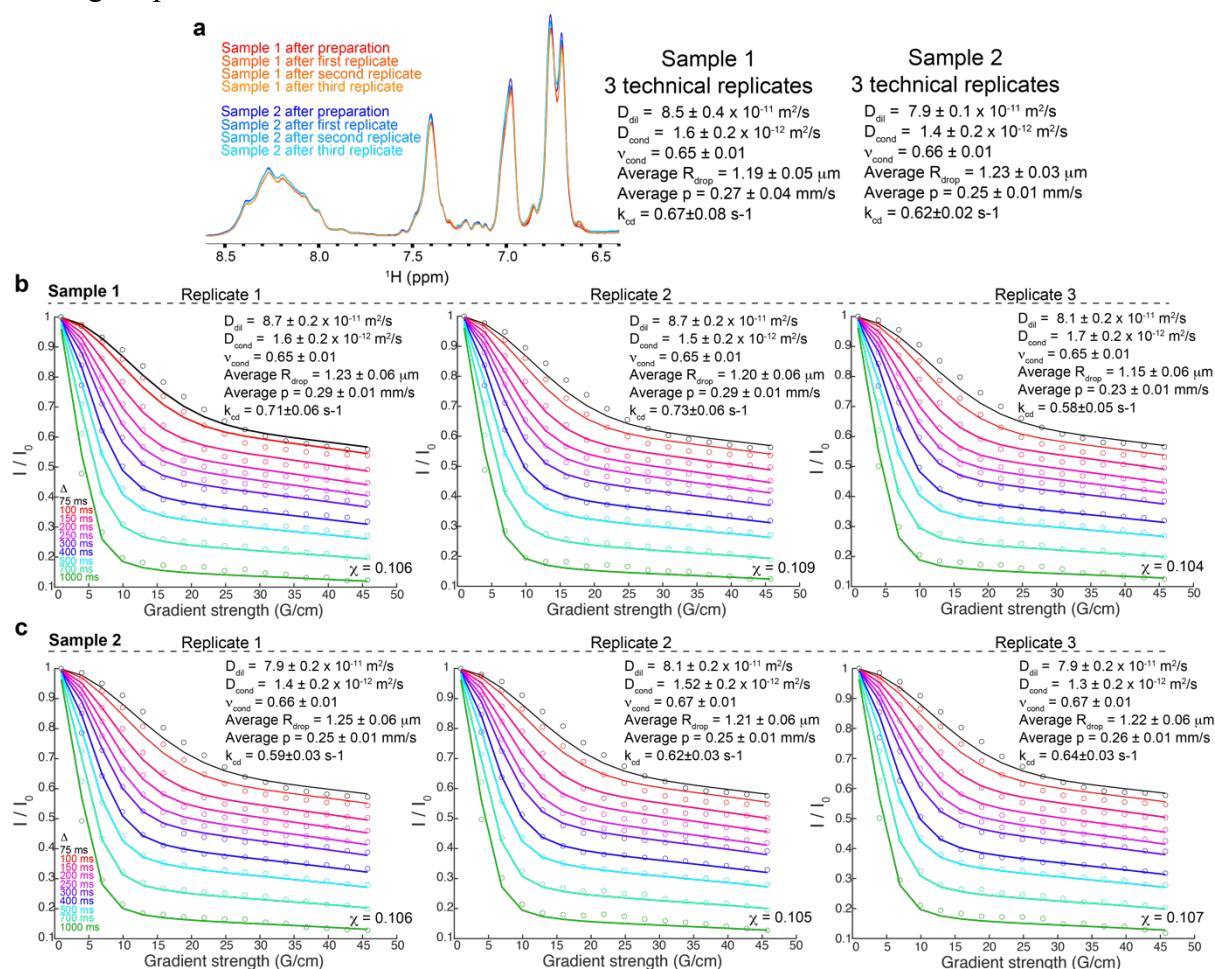
To acquire the experimental data sets, we used a pseudo-3D stimulated echo NMR experiment with bipolar gradients (PGSTE) where both the gradient strength and the diffusion delay are incremented as shown in Supplementary Fig. 12a. The theory shown in methods section for PGSE experiments apply here as well. Stimulated spin-echo probes diffusion while magnetization is stored on the  $z$ -axis, allowing prolonged diffusion encoding times as it relies on long longitudinal relaxation. Therefore, LLPS REDIFINE provides also a  $T_1$  weighting, however as longitudinal relaxation of observable non-labile protons from flexible regions in condensed and dilute phase is nearly identical (Supplementary Fig. 12b,c), relaxation can be disregarded from the model. This is done by normalizing the data sets for different diffusion times to unity. During spin-echo the magnetization is also held transverse during the period of gradient application, and the final signal will also be somewhat affected by the transverse relaxation rate. However, in all experiments, this duration is held constant throughout a dataset so its effect on the signal attenuation is uniform and can be excluded.



**Supplementary Fig. 12: Stimulated spin echo with bipolar gradients is used to acquire LLPS REDIFINE data.** **a** Stimulated gradient spin echo (PGSTE) is a preferred experiment to acquire NMR diffusion curves because it relies on slower  $T_1$  relaxation allowing us to probe longer diffusion encoding time. **b** Signal decay of proton signals experienced by biphasic FUS NTD as a result of varying diffusion time has a  $T_1$ -weighting. Resonance highlighted in green represents tyrosines overlapping signals from protein in dilute and condensed phase in this FUS NTD biphasic sample. Resonance highlighted in orange contains signal solely from condensed protein in the droplets stuck to the wall of NMR tube (assigned based on systematic chemical shift from the main resonance).<sup>17</sup> **c** Independent inversion recovery experimental results shows that the longitudinal relaxation constants of non-exchangeable protons are the same for the protein in condensed and dilute phases based on excellent monoexponential fitting and identical relaxation values.

## 4. LLPS REDIFINE – Biological and technical replicates

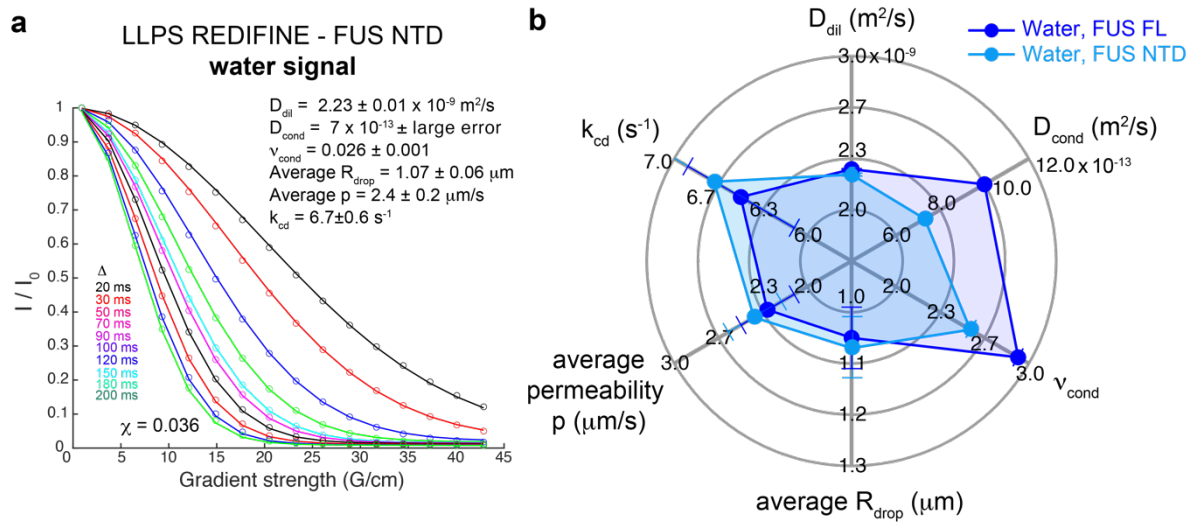
We have prepared two additional FUS NTD samples (using different batch of protein compared to Fig. 1 in the main text, and prepared on different days) and repeated the REDIFINE acquisition 3 times for each sample (technical replicates). Supplementary Fig. 13 summarizes the REDIFINE fitting results. Supplementary Fig. 13a illustrates that both samples have almost indiscernible 1D spectra and that they don't change during the time course of our measurements. For reasons unrelated to the present study FUS NTD sample (600  $\mu$ M in Fig. 1) was  $^{15}\text{N}/^{13}\text{C}$  labeled, hence the different  $^1\text{H}$  spectra due to the presence of heteronuclear couplings compared to Supplementary Fig. 13a. In the same panel, one can find the average values with standard deviation for three replicates for both biological replicates. The REDIFINE parameters are very reproducible and there are also no dramatic inter-sample variabilities. Supplementary Fig. 13b,c shows individual fitting results among samples and technical replicates. It is important to note that these data match well with the results presented in Fig. 1 of the main text, obtained using a different batch of FUS NTD protein and prepared at a higher protein concentration.



**Supplementary Fig. 13: REDIFINE reproducibility.** **a** 1D spectra of two independent FUS NTD samples at different time points during measurement. There is no significant change through time as well as variability between the two samples. Inset reports the average fitting values with standard deviation for two samples, based on three technical replicates. **b,c** Individual fitting for 3 experimental data sets for samples 1 and 2, reporting fitting values and the errors reporting on the correlation between parameters based on covariance matrix. FUS NTD was prepared at the concentration of 200  $\mu$ M.

## 5. REDIFINE applied to water signal in FUS NTD sample

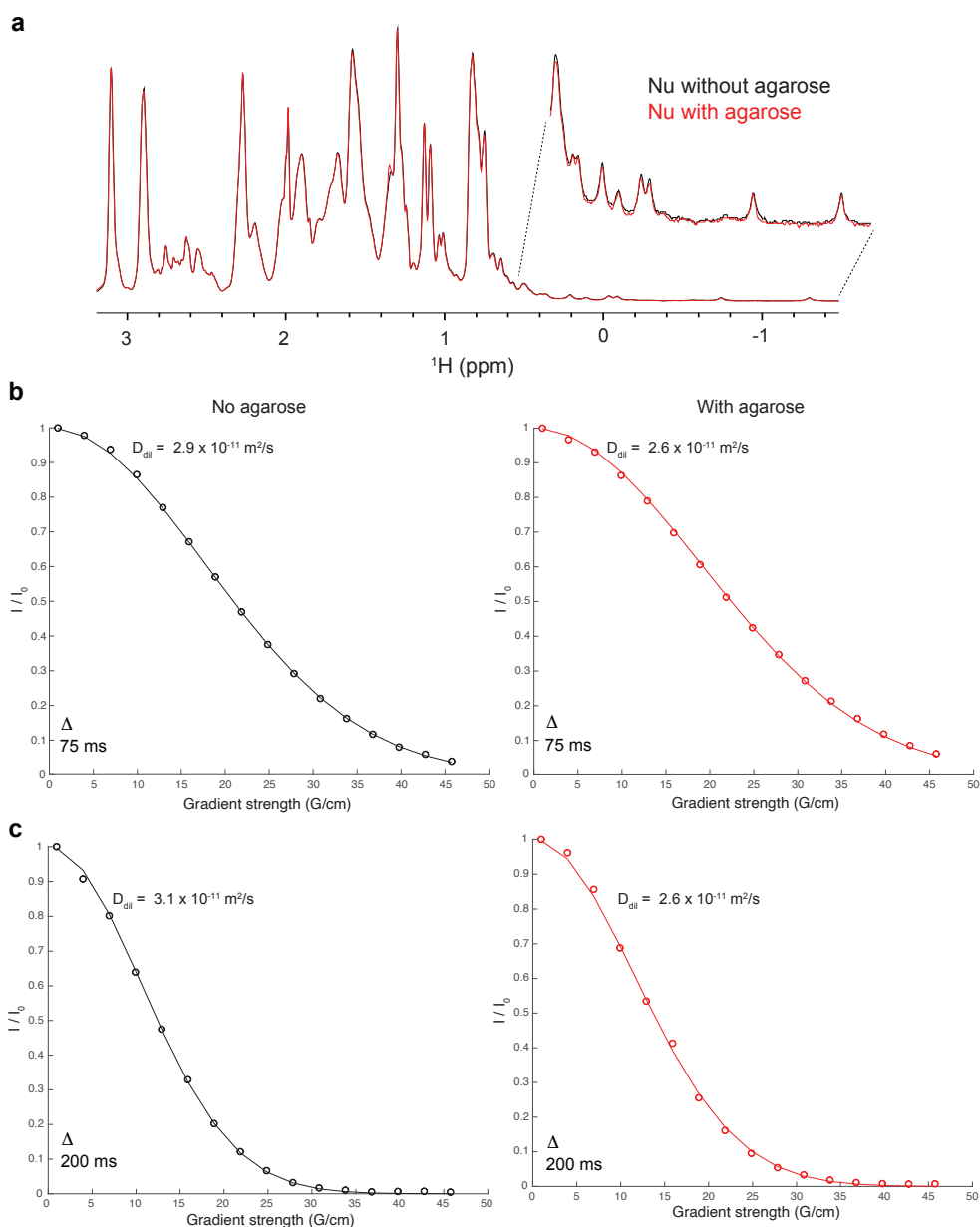
Supplementary Fig. 14a illustrates very good REDIFINE fit on water diffusion and extracted parameters on the structured water in the condensed phase. Combining with protein fit shown in Supplementary Fig. 13, we could calculate an average concentration of FUS NTD in droplets to be around 5.3 mM for this particular sample of truncated FUS construct. Supplementary Fig. 14b compares water-detected data for FUS FL (from Fig. 2 and Supplementary Fig. 3) with truncated FUS NTD. While water reports nearly identical average droplet size for two condensates and has identical diffusion in dilute phase, we can see that there is significantly less water in FUS NTD droplets (~16% less), which is expected as FUS FL has more polar pockets where water can be bound. Interacting less with FUS NTD, water exchanges somewhat faster with the bulk water compared to water in FUS FL.



**Supplementary Fig. 14. REDIFINE on water signal – application on FUS NTD.** **a** Excellent fit of water detected data to REDIFINE model and ensuing fitting parameters **b** Comparison of water REDIFINE acquired on FL FUS and truncated FUS NTD illustrating the similarities and differences between the two condensed samples. As the error for  $D_{cond}$  is large, we will not try to interpret the differences.

## 6. Protein diffusion in presence and absence of agarose

The key role of agarose in our experiments is that it mimics a stabilizing medium, preventing droplets from merging and sedimenting. 0.5% agarose gel is coarse and it barely affects the diffusion of small molecules such as water (we measure  $2.265 \times 10^{-9} \text{ m}^2/\text{s}$  in agarose vs literature values of  $2.299 \times 10^{-9} \text{ m}^2/\text{s}$  of bare water at 298 K<sup>5</sup>). Furthermore, we performed an extensive analysis to show that agarose does not affect the structure and behavior of the biomolecules that we are studying. Supplementary Fig. 15a shows that the NMR spectra, depicting the structure of Nucleocapsid protein (Nu) in phosphate buffer without and with 0.5% agarose, are virtually identical. This confirms that Nucleocapsid, representing a highly complex

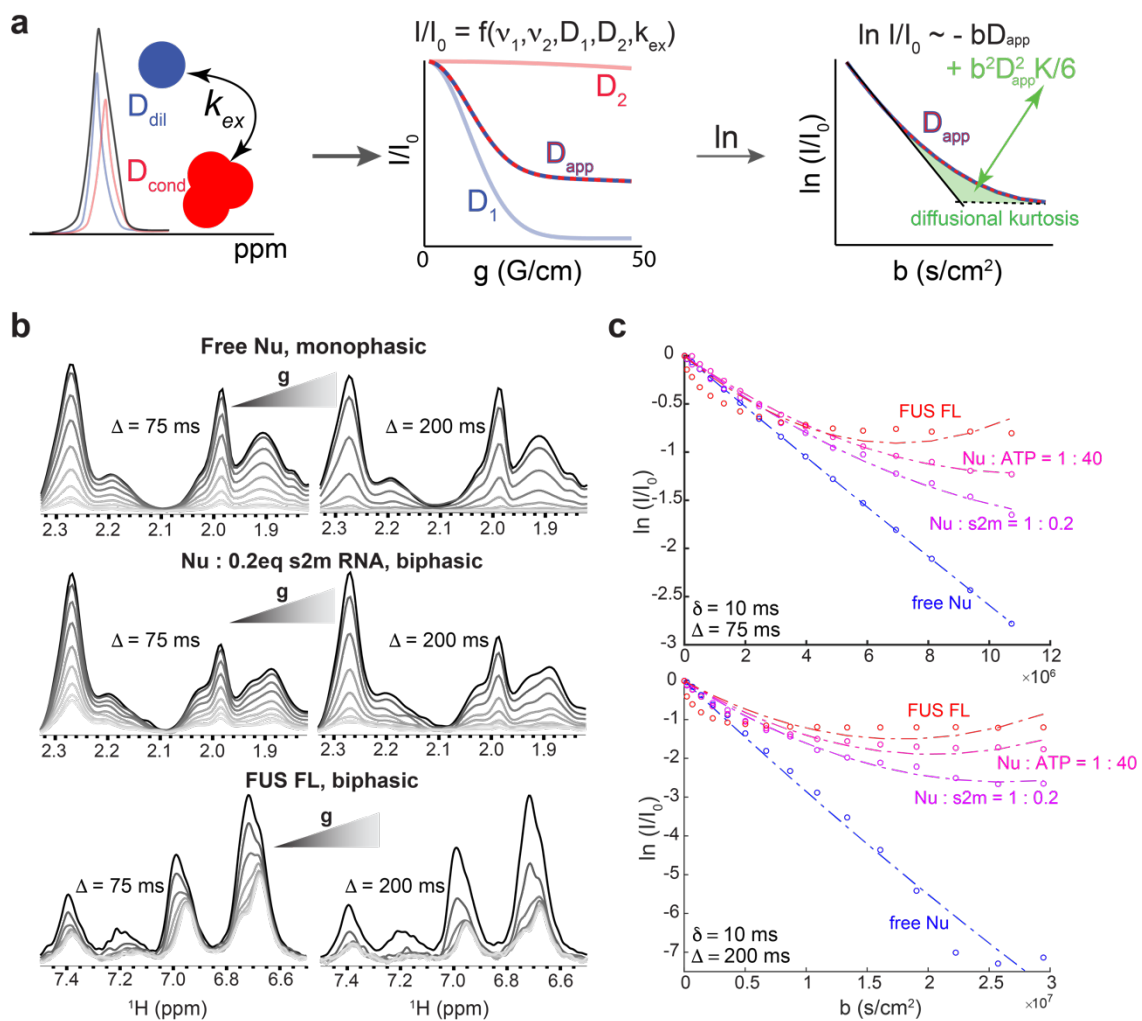


**Supplementary Fig. 15. Agarose is not affecting protein biomolecules.** **a** 1D NMR spectra comparing Nucleocapsid protein (Nu) in absence and presence of 0.5% agarose gel. The nearly identical spectra imply that Nu is not interacting with agarose. **b,c** Diffusion of Nu protein in presence and absence of agarose at two diffusion times (75 ms and 200 ms). These plots show that Nu protein diffuses 10-15% slower in agarose compared to the buffer solution.

protein consisting of structured and disordered domains (92 kDa), is not interacting with agarose. Supplementary Fig. 15b shows that the diffusion coefficients measured for the Nucleocapsid protein without and with agarose are also very similar (ca. 10% difference). As only monoexponential diffusion dynamics is detected even using longer diffusion times (200 ms) as shown in Supplementary Fig. 15c, we can further confirm that Nu protein is not interacting with agarose.

## 7. Diffusional kurtosis effect

The presence of chemical exchange between dilute and condensed phase leads to complex diffusion decays. On the semi-logarithmic plot (Supplementary Fig. 16a) the exchange-averaged diffusion decay (defined with an apparent diffusion coefficient) manifests as a deviation from the linear trend, often described as a kurtosis effect<sup>6</sup>. Kurtosis, denoted by the dimensionless parameter  $K$ , is a statistical metric<sup>6</sup> for quantifying the shape of a probability distribution. By definition, a Gaussian distribution has  $K$  equal to 0. Distributions that are more “peaked” typically have a positive kurtosis ( $K > 0$ ). Positive kurtosis reflects various diffusion environments experienced by diffusing molecules as they encounter barriers,<sup>7</sup> move between compartments, and undergo chemical exchange.<sup>8,9 10,11</sup> It can also be a result of intermolecular



**Supplementary Fig. 16: Biphasic condensates undergo non-gaussian diffusion described with the kurtosis effect.** **a** Chemical exchange between dilute and condensed phase encodes an exchange-averaged diffusion information on the NMR diffusion decay. The observed diffusion curve will strongly depend on the exchange rate (droplet size and permeability). Note the large deviation from the linearity when the curve is shown as a  $\ln \frac{I}{I_0}$  dependence of the diffusion factor  $b$ , described by kurtosis effect (Methods). **b** NMR signal decays of dispersed Nu, and biphasic Nu : RNA and FUS samples with respect to increasing gradient strength in PGSE experiment. Note the different decay trends when diffusion time  $\Delta$  is increased from 75 ms to 200 ms. **c** All the condensate samples show significant diffusional kurtosis effect (deviation from linearity) compared to the free Nucleocapsid sample that experience pure mono-exponential decay. The kurtosis effect varied from 0.06 for free Nu to 1.66 for FUS condensate at 75 ms diffusion time. It is noteworthy that even Nu condensates with different nucleic acids show distinct kurtosis values ( $K = 0.87$  for Nu:s2m and  $K = 1.23$  with Nu:ATP). Both diffusion coefficients and kurtosis factors were notably different when using 200 ms diffusion time.

interactions present in the system. A molecule diffusing with positive kurtosis would typically not travel as far in a given time interval as one that followed Gaussian statistics. In case of diffusional kurtosis, as it is shown in Supplementary Fig. 16a, the PGSE decay deviates from linearity on a semi-logarithmic plot. A traditional way to account for kurtosis is to perform a one-step Taylor series expansion of the exponent in the standard diffusion signal equation (Eq. 1), yielding

$$\frac{I}{I_0} = e^{-b \cdot D_{app} + b^2 D_{app}^2 K / 6} \quad (S1)$$

As in the case of diffusional kurtosis, diffusion coefficient becomes a function of experimental parameters, obstacle size or interaction affinity, apparent diffusion coefficient  $D_{app}$  must be introduced.<sup>12–16</sup>

It has been previously shown in MRI that diffusional kurtosis of water is very sensitive to the tissue type and even could be used to assess structure abnormalities in brain associated to neuropathologies.<sup>8,9</sup> Supplementary Fig. 16b shows the behavior of the NMR signals of different condensates (stabilized in agarose hydrogel)<sup>17</sup> subjected to diffusion NMR experiments. While signals stemming from free Nu in dispersed sample decay to zero at high gradient strengths, the peaks detected in the biphasic samples of Nu (in complex with RNA) and FUS (free) remain detectable even at high gradient values. For FUS protein, this signal plateau was correlated in our previous publication with the fraction of slowly diffusing protein in the condensed phase.<sup>17</sup> Similar behavior was initially anticipated for the biphasic Nu sample, however we noticed that the Nu condensate signals were experiencing a much steeper diffusion decay without an obvious plateau. As multi-component systems (protein, RNA and two phases), all the condensed samples show a significant kurtosis effect that intriguingly depends on the choice of diffusion time (Supplementary Fig. 16c). This implies the presence of restricted diffusion for the biomolecules and chemical exchange between the phases.

## 8. Prospects and limitations of LLPS REDIFINE

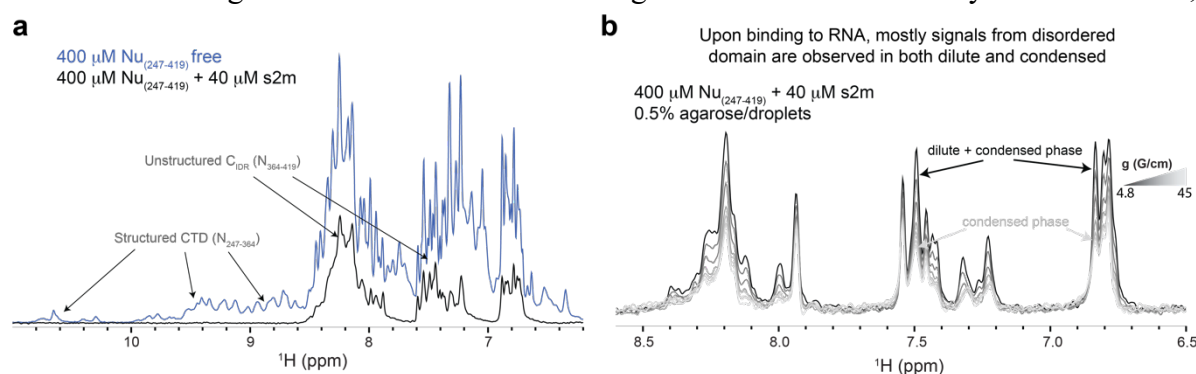
LLPS REDIFINE method proved applicable for the most samples examined in this study. This includes 17 different samples prepared using 5 different proteins having a wide range of biophysical properties (Supplementary Table 1). Due to very fast interphase exchange, LLPS REDIFINE failed to determine droplet parameters for PTBP1:3xUCUCU condensate samples. Chemical exchange of more than  $30 \text{ s}^{-1}$  appears fast on the diffusion time scale averaging out

the effects of restricted diffusion in the condensed phase and preventing unambiguous fitting of the LLPS REDIFINE datasets. Concluding from the simulation and experimental data, ideal chemical exchange for LLPS REDIFINE should fall between  $0.5 \text{ s}^{-1}$  and  $30 \text{ s}^{-1}$ . While the equal population of the species in the two phases gives LLPS REDIFINE the best sensitivity to distinguish accurately condensate parameters, this is often not the case if one of the phases is populated less than 10 %. Then not all parameters can be unambiguously determined. However, the performance of LLPS REDIFINE will also depend on the signal-to-noise ratio (SNR) of signals detected in NMR. This is why very good SNR ratio when detecting water allowed LLPS REDIFINE to determine parameters with good accuracy based on only 3% water in the condensed phase.

To avoid the interference from chemical exchange that labile protons in protein undergo with bulk water,<sup>10,11</sup> only non-exchanging NMR peaks should be integrated in LLPS REDIFINE. Throughout this study, we usually integrated aromatic protons (FUS NTD, full-length FUS, A6 RNA) or aliphatic protons (DDX4, Nucleocapsid, PTBP1).

It appeared that the two-pool model is sufficient to explain most of the data that we acquired. For protein-RNA condensates, we tried to extend the number of species in the model, for example, to take into account protein in free and bound state in both phases. It is generally straightforward to implement this in Eq. 7, however, with more pools corresponding system of coupled differential equations cannot be solved analytically and has to be fitted numerically which increases the calculation time and complexity of the multiparametric fitting procedure. As this hasn't improved the fitting of the data acquired on Nu:A6 condensates (Fig. 3b) at shorter diffusion times we decided to fit all the protein-RNA condensates data with the two-pool approximation (one species in dilute and one in condensed phase).

With structured protein:RNA condensates, another issue can be the broadening of the signal upon binding. While the peaks originating from the folded domains that directly bind to RNA will be most likely broadened beyond detection, signals from flexible, disordered regions remain NMR observable (Supplementary Fig. 17a). As illustrated for Nucleocapsid construct 247-419 consisting of a structured RNA binding domain and intrinsically disordered tail,

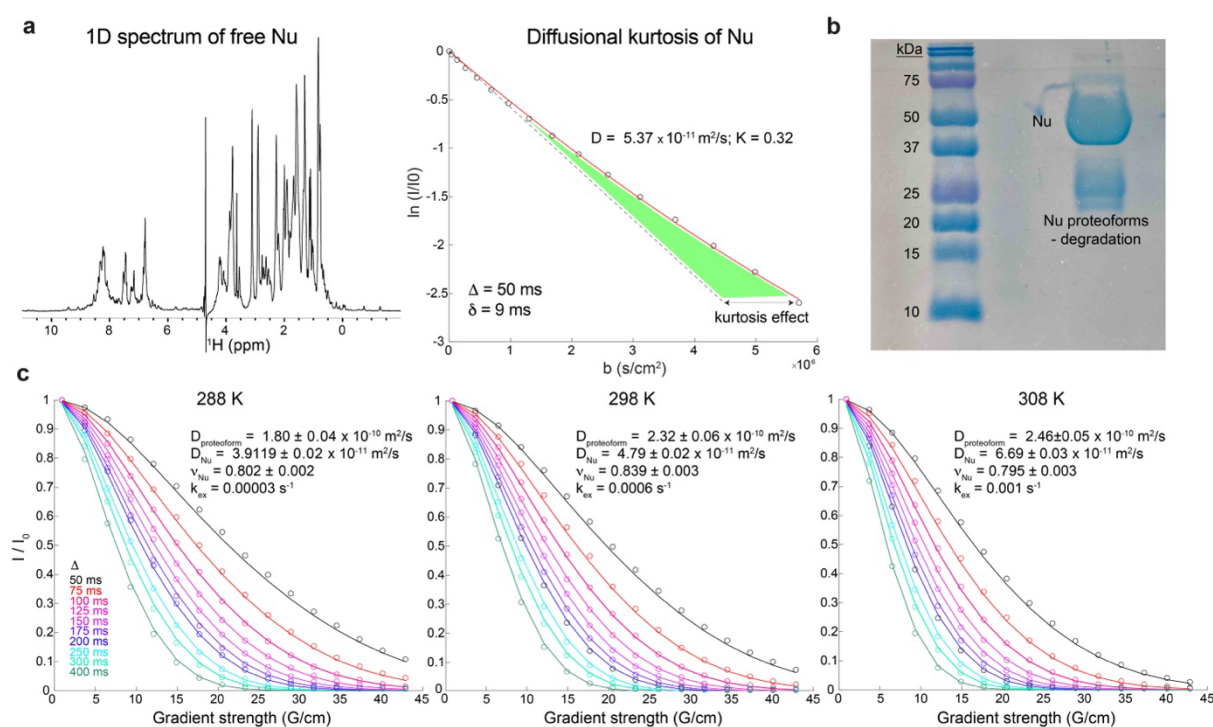


**Supplementary Fig. 17: IDRs remain NMR observable in the condensed phase.** **a** Even though signal of structured domains can broaden beyond detection because of slower tumbling upon protein binding and/or increased viscosity in condensed phase, IDR regions remain flexible. 1D spectra show the comparison of Nu<sub>(247-419)</sub> signal before and after binding to s2m RNA. The signals stemming from disordered regions remain NMR observable upon binding. **b** 1D DOSY spectra acquired for multiple gradient strengths on Nu<sub>(247-419)</sub> condensate sample formed with s2m RNA. The last spectrum acquired with 45 G/cm gradient strength shown in black stems from the protein in condensed phase. Note that the IDR peaks have very similar linewidth in both condensed and dilute phases illustrating that longer disordered regions remain flexible even in the droplets.

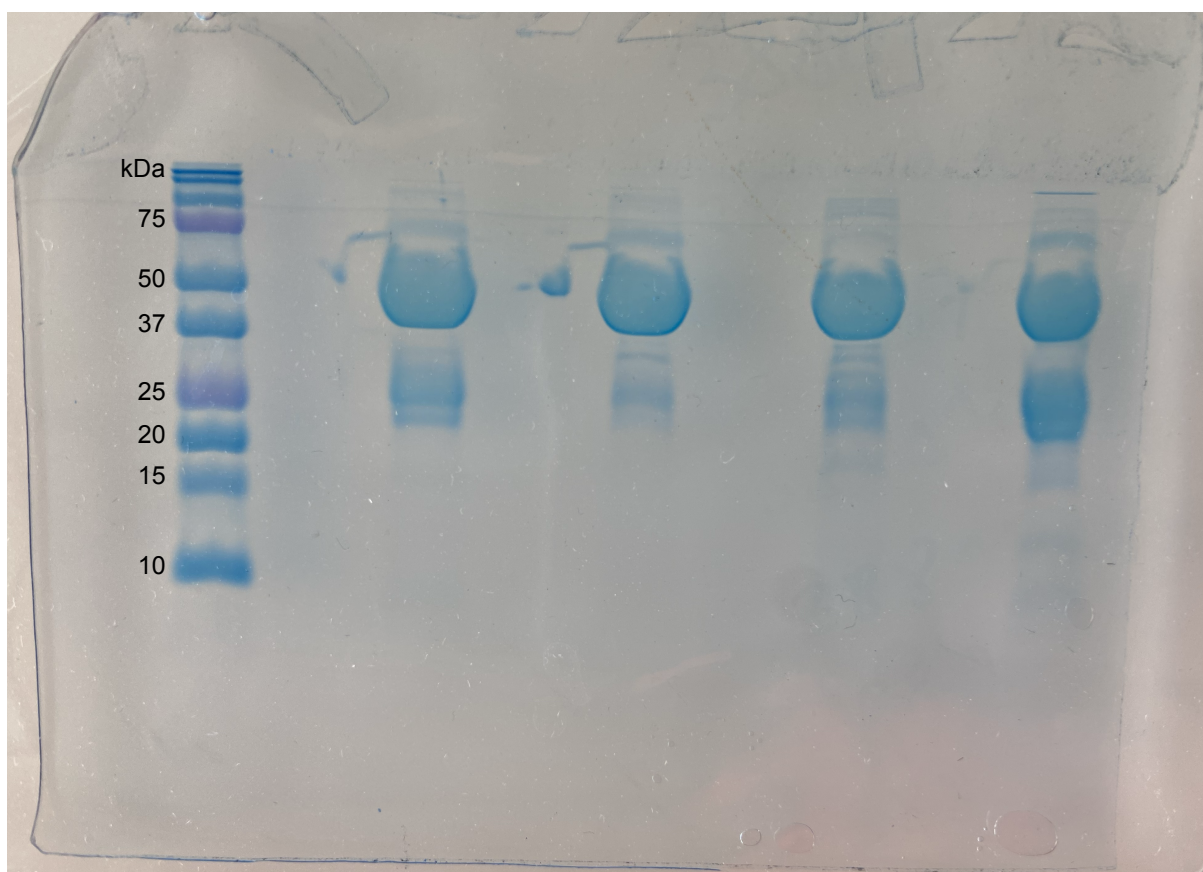
signals from IDR have very similar linewidths in dilute and condensed phase (Supplementary Fig. 17b). This is also the case for full-length Nucleocapsid, PTBP1 and FUS samples. Undetectable NMR signal of protein or RNA upon condensate formation (as for s2m RNA upon binding to Nucleocapsid protein) limits the applicability of LLPS REDIFINE.

Regarding soluble complexes, the methodology is most sensitive when diffusion coefficients of the exchanging species are as different as possible. Furthermore, as in the case of condensates, the chemical exchange ( $k_{\text{off}}$  rate) needs to be in an intermediate regime on the diffusion time scale (residence time 30-1000 ms). This is often the case with RNA binding proteins and their complexes with RNA.

Even the presence of two non-exchanging species could lead to a kurtosis effect which was the case with a 2-week-old Nucleocapsid sample containing degradation proteoforms (Supplementary Fig. 18a,b). Having overlapping NMR signals and showing diffusional kurtosis served as a negative control for our method. REDIFINE methodology resolved that the species don't exchange with each other and determined the population and diffusion coefficients of the individual components. Even at three different temperatures altering diffusion behavior and kinetics, REDIFINE detected no chemical exchange between two species and kept the populations constant (Supplementary Fig. 18c). The only parameters that changed with temperature were diffusion coefficients which was an expected behavior.



**Supplementary Fig. 18: LLPS REDIFINE can also detect protein degradation.** **a**  $^1\text{H}$  spectrum of a free Nucleocapsid protein after a couple of weeks at the room temperature showing at first sight no signs of degradation. However, diffusional kurtosis is observed suggesting the presence of more than one species in solution. **b** On the SDS-PAGE gel it is clear that sample contains also another Nucleocapsid proteoform of around 25 kDa in molecular weight. **c** LLPS REDIFINE data acquired on this degraded Nucleocapsid sample. Two species attributed to free Nucleocapsid and a smaller proteoform are identified. More important, no exchange between two species is detected which is consistent with the existence of two independent proteins in solution. This is confirmed by recording the experiments at 288, 298 and 308 K. All fitting results show constant populations and no chemical exchange between the species.



**Supplementary Fig. 19: Uncropped scan of a gel shown in Supplementary Fig. 18b**

## References

1. James, T. L. & McDonald, G. G. Measurement of the self-diffusion coefficient of each component in a complex system using pulsed-gradient fourier transform NMR. *Journal of Magnetic Resonance (1969)* **11**, 58–61 (1973).
2. Johnson, C. S. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog Nucl Magn Reson Spectrosc* **34**, 203–256 (1999).
3. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J Chem Phys* **28**, 430–431 (1958).
4. Johnson, C. S. Effects of Chemical Exchange in Diffusion-Ordered 2D NMR Spectra. *J Magn Reson A* **102**, 214–218 (1993).
5. Mills, R. Self-diffusion in normal and heavy water in the range 1–45.deg. *J Phys Chem* **77**, 685–688 (1973).
6. Balanda, K. P. & Macgillivray, H. L. Kurtosis: A critical review. *American Statistician* **42**, 111–119 (1988).
7. Tanner, J. E. & Stejskal, E. O. Restricted self-diffusion of protons in colloidal systems by the pulsed-gradient, spin-echo method. *J Chem Phys* **49**, 1768–1777 (1968).
8. Jensen, J. H. & Helpert, J. A. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed* **23**, 698–710 (2010).
9. Jensen, J. H., Helpert, J. A., Ramani, A., Lu, H. & Kaczynski, K. Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med* **53**, 1432–1440 (2005).
10. Liu, M., Toms, H. C., Hawkes, G. E., Nicholson, J. K. & Lindon, J. C. Determination of the relative NH proton lifetimes of the peptide analogue viomycin in aqueous solution by NMR-based diffusion measurement. *J Biomol NMR* **13**, 25–30 (1999).
11. Thureau, P., Ancian, B., Viel, S. & Thévand, A. Determining chemical exchange rates of the uracil labile protons by NMR diffusion experiments. *Chemical Communications* 200–202 (2006) doi:10.1039/b513580j.
12. Price, W. S., Barzykin, A. V., Hayamizu, K. & Tachiya, M. A model for diffusive transport through a spherical interface probed by pulsed-field gradient NMR. *Biophys J* **74**, 2259–2271 (1998).
13. Pfeuffer, J., Flögel, U., Dreher, W. & Leibfritz, D. Restricted diffusion and exchange of intracellular water: Theoretical modelling and diffusion time dependence of <sup>1</sup>H NMR measurements on perfused glial cells. *NMR Biomed* **11**, 19–31 (1998).
14. Balinov, B., Jonsson, B., Linse, P. & Soderman, O. The NMR Self-Diffusion Method Applied to Restricted Diffusion. Simulation of Echo Attenuation from Molecules in Spheres and between Planes. *J Magn Reson A* **104**, 17–25 (1993).
15. Stanisz, G. J., Szafer, A., Wright, G. A. & Henkelman, R. M. An analytical model of restricted diffusion in bovine optic nerve. *Magn Reson Med* **37**, 103–111 (1997).
16. Karunanithy, G. *et al.* INDIANA: An in-cell diffusion method to characterize the size, abundance and permeability of cells. *Journal of Magnetic Resonance* **302**, 1–13 (2019).
17. Emmanouilidis, L. *et al.* NMR and EPR reveal a compaction of the RNA-binding protein FUS upon droplet formation. *Nat Chem Biol* **17**, 608–614 (2021).