bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dual RNA modulation of protein mobility and stability within phase-separated condensates

Andrés R. Tejedor,[1] Adiran Garaizar,[2] Jorge Ramírez,[1, a)] and Jorge R. Espinosa[2, b)]

[1)] *Department of Chemical Engineering, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006, Madrid, Spain*

[2)] *Maxwell Centre, Cavendish Laboratory, Department of physics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

(Dated: 20 April 2021)

One of the key mechanisms employed by cells to control their spatiotemporal organization is the formation and dissolution of phase-separated condensates. Such balance between condensate assembly and disassembly can be critically regulated by the presence of RNA. In this work, we use a novel sequence-dependent coarse-grained model for proteins and RNA to unravel the impact of poly-uridine RNA in modulating the protein mobility and stability within different biomolecular condensates. We explore the behavior of FUS, hnRNPA1 and TDP-43 proteins along with that of their corresponding prion-like domains and RNA-recognition motifs, from absence to moderately high RNA concentration. By characterising the phase diagram, key molecular interactions, surface tension and transport properties of the condensates, we report a dual RNA-induced behavior: On the one hand, poly-uridine enhances phase separation at low concentration, whilst at high concentration it inhibits the ability of proteins to self-assemble. On the other hand, as a consequence of such stability modulation, the viscoelastic liquid properties of the condensates are significantly enhanced at moderately high RNA concentration as long as the length of poly-uridine strands is comparable or moderately shorter than those of the proteins, whereas protein self-diffusion barely depends on poly-uridine length. On the whole, our work rationalizes the different routes by which RNA can regulate phase separation and condensate dynamics, as well as the subsequent aberrant rigidification implicated in the emergence of various neuropathologies and age-related diseases.

Keywords: Intrinsically disordered proteins, liquid-liquid phase separation, RNA, biomolecular condensates

I. INTRODUCTION

Liquid-liquid phase separation (LLPS) is one of the key processes employed by cells to control the spatiotemporal organization of their many components[1–6]. This phenomenon – displayed by a large variety of biomolecules such as multivalent proteins and nucleic acids[6–11] – is involved in wide-ranging aspects of the cell function such as membraneless compartmentalisation [6,12–16], signaling[2,17], genome silencing[18–20], formation of super-enhancers[21], helping cells to sense and react to environmental changes[22], or buffering cellular noise[23] among many others[24–27]. The spontaneous demixing of the cell components into different coexisting liquid compartments occurs both inside the cytoplasm (e.g., P granules[1] and RNA granules/bodies[28,29]) and in the cell nucleus (e.g., Cajal bodies[30], nucleoli[31], nuclear speckles[32,33] or heterochromatin domains[19,20]), and enables the coordinated control of thousands of simultaneous chemical reactions that are required to maintain biological activity[34]. Besides these diverse functionalities, membraneless organelles have also been observed to exert mechanical forces to induce chromatin reorganization[35,36] or to act as molecular sensors of intracellular and extracellular exchanges[22]. Still novel biological roles, such as the propensity of condensates to buffer protein concentrations against gene expression noise continue being discovered[23,37].

The biomolecular building blocks behind LLPS in the cell are usually proteins with intrinsically disordered regions (IDRs) or proteins with globular domains connected by flexible linkers that can establish multiple homotypic self-interactions or heterotypic interactions with cognate biomolecules (e.g. a different IDR, RNA or DNA) over their interactions with the solvent[9,11]. Several DNA and RNA-binding proteins such as FUS[38–40], hnRNPA1[15,16], TDP-43[41,42], LAF-1[43] or HP1[19,20] have been observed to undergo phase separation both *in vivo* and *in vitro*. These proteins, besides their intrinsically disordered regions, frequently present additional specific domains with high physico-chemical affinity for RNA (termed as RNA recognition motifs (RRMs))[44] or DNA[45]. In particular, the intermolecular binding between IDRs and RNA (either via specific RNA-RRMs interactions or non-selective electrostatic or π-π interactions) have been found to be critical in regulating LLPS[43,46–49]. In that respect, determining the precise consequences of RNA in LLPS is a matter of huge interest[5,43,48–52].

In vitro experimental evidences show how protein aggregation is enhanced upon addition of RNA at low concentration, whilst it inhibits phase separation at high concentration[47,50,53,54]. Such re-entrant behavior

[a)] Electronic mail: jorge.ramirez@upm.es

[b)] Electronic mail: jr752@cam.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is in agreement with the hypothesis that solid-like aggregates are more readily formed in the cytoplasm than in the cell nucleus, where the abundance of RNA is higher[50]. Moreover, besides modulating the stability of the condensates, RNA can affect their kinetic properties. A viscosity reduction of LAF-1 droplets (a key protein in P granules formation) after addition of short RNA strands has been observed without significantly affecting droplet stability[43]. On the contrary, the inclusion of long RNA chains inside the condensates can also enhance notably their viscosity at certain given concentrations[46,55]. This RNA-induced modulation of droplet viscoelasticity (and also recently observed by DNA[56]) is crucial in the regulation/dysregulation of the liquid-like behavior of RNA-binding proteins (RBPs) such as FUS[38–40], hnRNPA1[15,16], TDP-43[41,42,57], TAF-15[50,58] or EWSR1 among others[48,50,58]. The resulting rigidification of these condensates can lead to the formation of pathological solid aggregates which are behind the onset of several neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS), frontotemporal dementia or Alzheimer[15,59–63]. Because of that, a huge effort is being devoted in understanding the underlying molecular factors involved in RNA-induced regulation of condensates stability and viscoelasticity[6,8,12,51,52,64,65].

Recent experimental advances in single-molecule Forster resonance energy transfer (smFRET) have enabled the direct observation of the structural and dynamic protein behavior in diluted conditions[66–68], however, the thermodynamic and kinetic aspects inside the condensates are still hardly accessible[69]. Notably, particle tracking microrheology techniques have been successfully used to provide data about the mean squared displacement (MSD) of marked beads inside droplets, and then, via that MSD, condensate viscosity has been estimated[43,46,55,70]. Nevertheless, other fundamental magnitudes such as the protein mean squared displacement, end-to-end distance relaxation times, protein radius of gyration or droplet surface tensions are extremely challenging to obtain[51]. Moreover, direct measurements of the molecular contacts that promote phase separation are of great relevance, and rarely, this information can be unequivocally extracted[39,58,71]. The mutation and/or phosphorylation of specific residues along sequences can help in deciphering which contacts might be key in sustaining LLPS[72,73], but still, higher level of mechanistic and molecular resolution is needed.

In that respect, computer simulations emerges as a great tool to enlighten this blind spot[6,74–76]. The most recent advances in computational science have allowed to carry out impressive calculations mimicking *in vivo* conditions[77]. Atomistic Molecular Dynamics (MD) simulations have been also successfully proved in characterizing the conformational ensembles of single proteins and protein complexes[76,78,79], pinpointing the link between chemical modifications, sequence mutations, and the modulation of protein–protein and protein-DNA interactions[80–82], and guiding the development of chemically accurate coarse-grained models for LLPS[83,84]. Simultaneously, a huge effort in developing different levels of coarse-grained (CG) potentials is being devoted, including mean field models[85–89], lattice-based simulations[90–93], minimal models[94–98], and sequence-dependent models[83,99–101]. By retaining the specific physico-chemical features of proteins, DNA and RNA, while averaging out others for computational efficiency, CG models have been widely used to elucidate key factors behind LLPS and their dependency on the protein length[102,103], amino acid sequence[83,99,100,104,105], multivalency[90,106–111], conformational flexibility[112,113] and multicomponent composition[114–118]. Nonetheless, regarding the role of RNA in LLPS and the molecular driving forces contributing to the formation of RNA-RBP condensates, further work is required[119]. On the one hand, atomistic MD simulations have provided binding free energies of specific protein/RNA complexes, but are limited to very few protein replicas[120,121]. On the other hand, coarse-grained models have been recently proposed to elucidate the effect of RNA on phase separation of small prion-like domains such as those of FUS[122], protamine[123] or LAF-1[116]. Remarkably, the work by Regy *et al*[116] presents a detailed parametrization of a CG model for RNA within the framework of the HPS protein potential[83], opening up new possibilities to link the molecular mechanisms of RNA-RBP condensates to their macroscopic phase behavior.

The present work aims to narrow down this gap by shedding light on the re-entrant behavior of RNA in LLPS and on its implications on condensate transport properties. By employing a novel high-resolution CG model for RNA and IDPs[83,116,124], we explore the ability of different RNA-binding proteins which undergo LLPS (FUS, hnRNPA1, TDP-43 as well as their corresponding prion-like domains and RNA-recognition motifs) to phase separate in presence of poly-uridine (poly-U) RNA from low to moderately high concentration. First, we validate the model predictions against the relative experimental protein saturation concentration of these condensates in absence of RNA at physiological salt concentration, finding a remarkable qualitative agreement. We also analyse the key molecular interactions and primary protein domains promoting aggregation, as well as the impact of heterogeneous sequence connectivities in the structure and surface tension of the condensates. Secondly, we investigate how poly-U RNA regulates the coexistence line of the condensates as a function of RNA concentration for a constant poly-U strand length, as well as for different strand lengths at a constant protein/poly-U concentration. Besides evidencing RNA re-entrant behavior of enhanced LLPS at low poly-U concentration and RNA-inhibited LLPS at moderately high poly-U saturation, we find a critical poly-U length below which, LLPS is no longer promoted even at low concentrations. Finally, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

characterize the transport properties (i.e., protein mobility and viscosity) of the condensates as a function of poly-U saturation, showing that from moderate RNA concentration, protein mobility is considerably enhanced. However, condensate viscosity is only reduced without significantly altering their stability when strands of a minimal critical RNA length are introduced ($\sim$ 100 nucleotides). When longer strands are added, viscosity moderately increases. Taken together, our work provides a framework to rationalise the ubiquitous dual effect of RNA in the stability and kinetics of RNA-RBP condensates.

II. RESULTS

A. Sequence-dependent model validation

Biomolecular condensates are stabilized by chemically diverse weak protein–protein interactions, which are determined by the specific nature (e.g., hydrophobicity, aromaticity, and charge) of the encoded protein amino acids[58,108]. Here, to capture such sequence specificity, we employ a novel reparametrization[124] of the high-resolution HPS model from Mittal group[100] which accounts for sequence-dependent hydrophobic and cation–π interactions by means of short-range pairwise potentials, and for electrostatic interactions through Yukawa long-range potentials. Bonded interactions between subsequent amino acids are restrained by a harmonic potential, and non-bonded hydrophobic interactions are modelled via an Ashbaugh-Hatch potential. Additionally, cation–π and electrostatic interactions are described by Lennard-Jones and Yukawa/Debye-Hückel potential terms respectively. The low salt physiological concentration regime ($\sim$ 150 mM) of the implicit solvent is controlled by the screening length of the Yukawa/Debye-Hückel potential. Given that the original HPS model[100] has been shown to underestimate LLPS-stabilising cation–π interactions[82], we employ the recently proposed reparametrization by Das *et al.*[124]. Additionally, to account for the 'buried' amino acids contained in the protein globular domains, we scale down those interactions with respect to the original set of HPS parameters by a 30% as proposed in Ref[82,125]. All the details regarding the model parameters and simulation setups are provided in the Supplementary Information (SI).

To further validate the model[100,124], we elucidate the relative ability to phase-separate of several archetypal RNA/DNA-binding proteins that are known to undergo LLPS either *in vivo* and *in vitro*. These proteins are: FUS[38–40], hnRNPA1[15,16] and the TAR DNA-binding protein 43 (TDP-43)[41,42] (Fig. 1A). We evaluate the phase diagram for either the full protein sequences and for some of their specific domains such as the RNA recognition motifs (RRMs) or the prion-like domains (PLDs). More precisely, we focus on the following sequences: FUS (full-sequence), FUS-PLD, hnRNPA1 (isoform A1-B, and hereafter named as hnRNPA1), hnRNPA1-PLD, hnRNPA1-RRM, TDP-43 (full-sequence), and TDP-43-PLD (sequences are provided in the SI). For TDP-43, we distinguish between two different variants, one including the α–helix structured domain in the C-tail intrinsically disordered region (h-TDP-43), and another in which the whole PLD region remains fully disordered (wt-TDP-43). Despite h-TDP-43 and wt-TDP-43 only differing in less than 10 % of their sequence structural conformation[126,127], the presence of the $\alpha-$helical structured domain has been shown to moderately affect the protein ability to phase-separate[128]. We also study the low complexity domain (LCD) of the isoform A1-A of hnRNPA1[15] (termed as hnRNPA1-A-LCD), since it has been shown to be a key part of the hnRNPA1 sequence in promoting LLPS in absence of RNA[15].

By means of Direct Coexistence simulations (DC)[129–131] in combination with the critical exponent law and the law of rectilinear diameters[132], we compute the phase diagram (Fig. 1B) of all the aforementioned proteins (hnRNPA1-PLD and hnRNPA1-RRM are shown in Fig. S2 of the SI; see SI of Ref.[98] for details on extracting coexisting densities from DC simulations). The predicted renormalized critical points (T/T_c' where T_c' refers to the highest critical temperature of the set) against the experimental saturation concentration of the proteins to undergo LLPS in FUS[58,82], FUS-PLD[58], hnRNPA1[15], hnRNPA1-A-LCD[15], wt-TDP-43[42,82], h-TDP-43[42,82] and TDP-43-PLD[133] are shown in Fig. 1C (please note that the experimental saturation concentration reported by Molliex *et al.*[15] corresponds to the isoform A1-A, but the difference in the critical concentration between the two isoforms is assumed to be minor). We find a positive correlation between the critical point in our simulations and the saturation concentration of the proteins in experiments (both at physiological salt concentration). Such impressive qualitative agreement (a coarse-grained model with explicit solvent is not expected to quantitatively capture the actual T_c) demonstrates that the cation-π reparametrization proposed by Das *et al.*[124] on top of Mittal's group model[100] is able to successfully describe the relative ability of these proteins to self-assemble into phase-separated condensates. Furthermore, we observe a non-negligible difference between the phase diagram of the α–helical structured TDP-43 and that of the wt-TDP-43, showing the latter a moderately lower critical temperature, as reported in Ref.[128]. Notably, both prion-like domains of FUS and TDP-43 exhibit a significant lower ability to phase separate than their full counterparts as experimentally found[58]. On the contrary, hnRNPA1-A-LCD exhibits a similar critical temperature to that of hnRNPA1 full-sequence. To rationalise these observations, in the following section, we perform a detailed molecular and structural characterisation of the condensates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

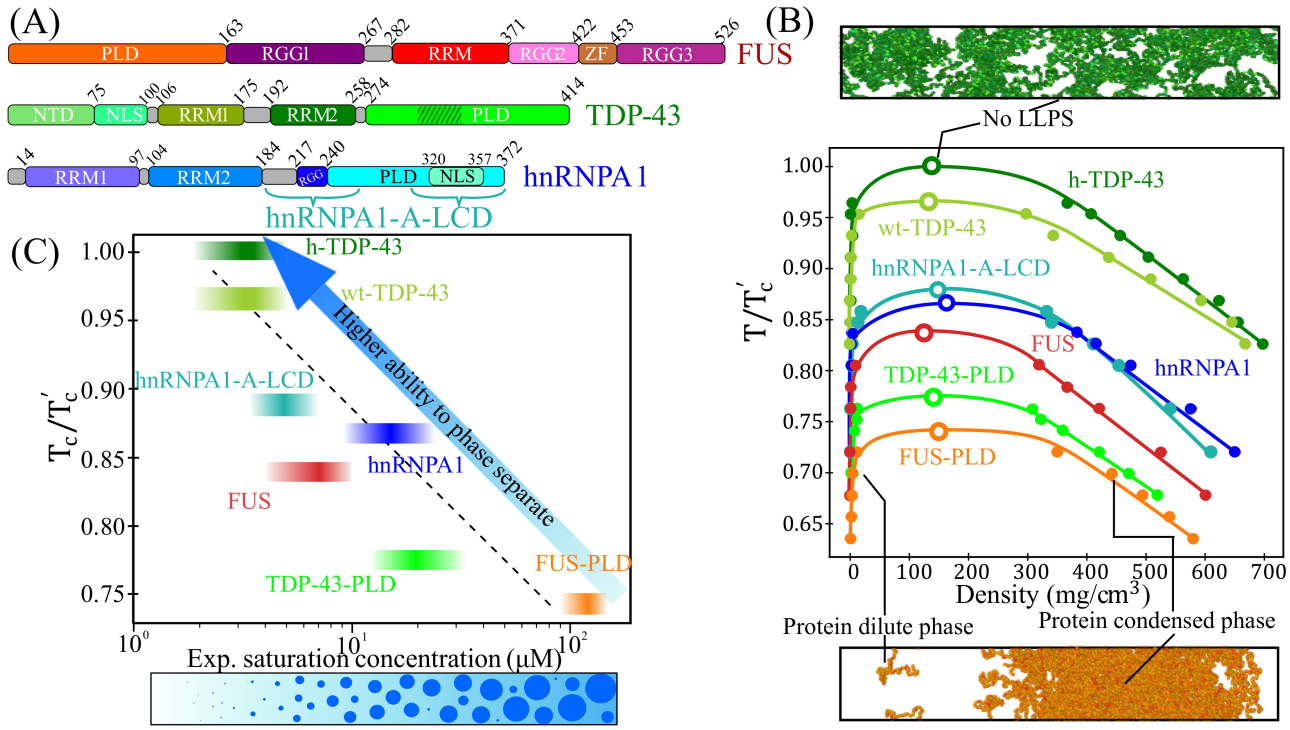

FIG. 1. Experimental validation of the sequence-dependent model. (A) Different sequence domains of the three studied proteins: Prion-like domain (PLD), Arginine-Glycine-Glycine rich domain (RGG), RNA-recognition motifs (RRM), zinc finger (ZF), N-tail domain (NTD) and nuclear localization sequence (NLS). Dashed lines in TDP-43 PLD indicate the position of the α–helical domain. Braces in hnRNPA1 sequence indicates the two LCD segments composing the sequence of the isoform A1-A of hnRNPA1 (hnRNPA1-A-LCD), corresponding to the residues 186-251 + 304-372 (see section SII for the sequence) (B) Phase diagram in the (T/T_c')-density plane for FUS (red), FUS-PLD (orange), hnRNPA1-A-LCD (turquoise), hnRNPA1 (blue), h-TDP-43 (dark green), wt-TDP-43 (lime green) and TDP-43-PLD (light green). Filled circles indicate the coexistence densities obtained through DC simulations, and empty circles the estimated critical points via the critical exponent and rectilinear diameter laws[132] (Eqs. (S6) and (S7) in the Supplementary Information). T_c' accounts for the highest critical temperature of the protein set (h-TDP-43), which is $T_c' = 472K$. Top: DC simulation of h-TDP-43 above the critical point where LLPS is no longer observed. Bottom: Direct Coexistence simulation of FUS-PLD at T/T_c'=0.70 exhibiting two coexisting phases. (C) Experimental saturation concentration of the proteins to undergo phase separation *versus* the renormalized critical temperatures shown in (B). The experimental saturation concentration[134] at physiological salt conditions for FUS[47,58,82] (including FUS-PLD), hnRNPA1[15], hnRNPA1-A-LCD[135,136], TDP-43[42,82] and TDP-43-PLD[133] are depicted by intervals to consider concentration uncertainty. The height of the intervals accounts for the computational uncertainty in T/T_c'. The dashed black line is a linear fit to the displayed data (considering the mean concentration of the interval) and the blue arrow indicates higher ability to phase separate. At the bottom, a schematic cartoon summarizing the expected phase behavior while increasing protein concentration is included. Note that temperatures in this model are unrealistic and only describe the relative ability of the different proteins to phase-separate, thus, temperature is only meaningful when is renormalized.

B. Structural and interfacial properties of the condensates without RNA

The specific composition and patterning of the amino acids along the sequence has a huge impact on the protein macroscopic phase behavior[83,100,127]. Moreover, beyond sequence, their conformational ensemble plays a crucial role not only in their ability to phase-separate, but also in the condensate structure[113,127,128,137,138]. A close example of this is TDP-43, in which a subtle conformational difference on its C-terminal intrinsically disordered domain produces a moderate change on its phase diagram (See Figure 1B). To further characterize the molecular, structural and interfacial properties of the previous protein condensates, we now perform a comprehensive full analysis of their surface tension, LLPS-stabilising most-frequent contacts, protein conformational ensembles in and out of the droplet, and condensate structure.

In Fig. 2A we plot the surface tension (γ) between the condensate (protein-rich) and protein-poor liquid phases as a function of temperature (renormalized by the highest critical temperature of the protein set, T_c' of h-TDP-43). An advantage of computer simulations is that γ between two fluid phases (or between a fluid and a vapour one) can be easily computed, as explained in SIV, compared to more challenging approaches (i.e.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

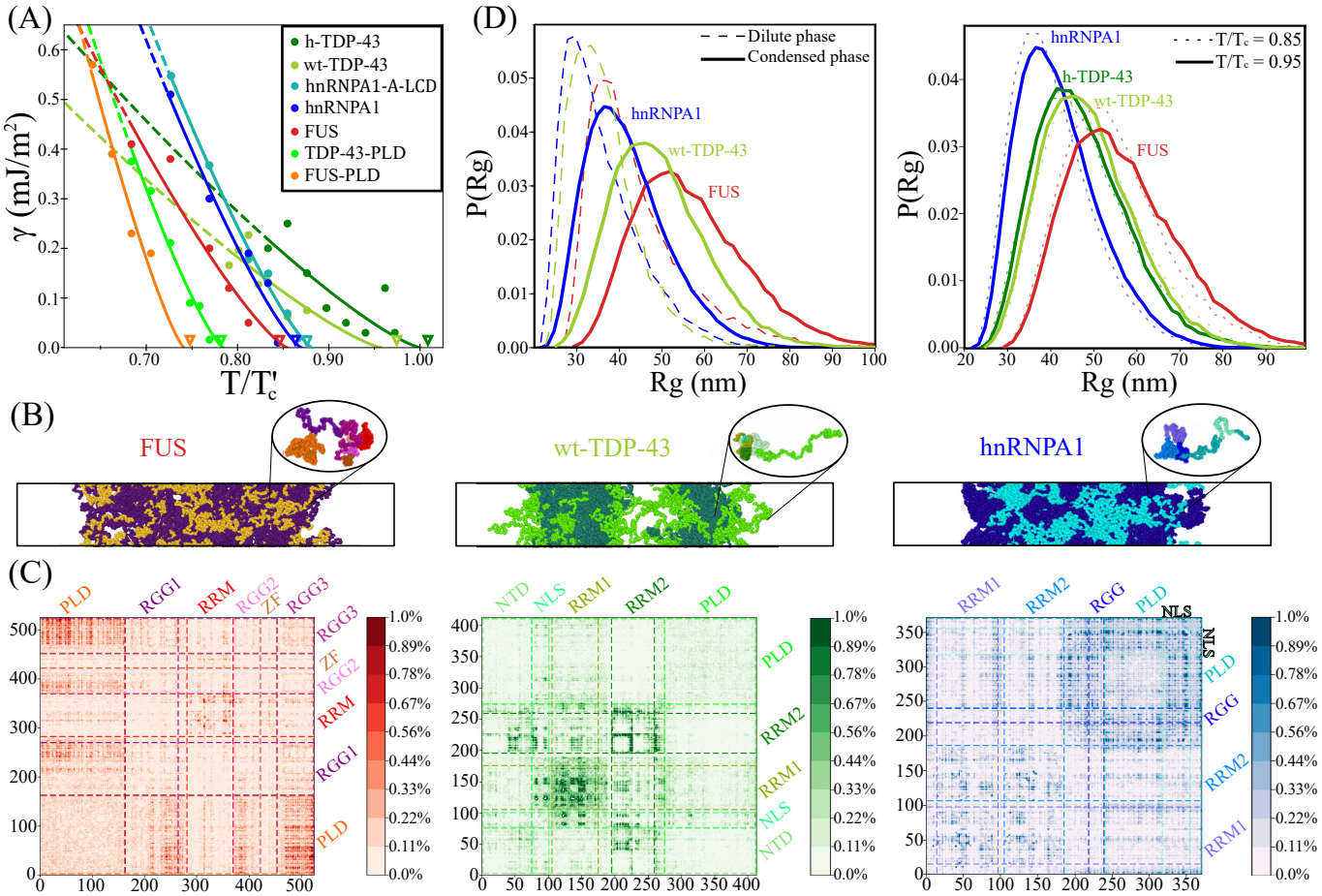

FIG. 2. Molecular, structural and interfacial properties of different RNA-binding protein condensates in absence of RNA. (A) Condensate surface tension (γ) of FUS, FUS-PLD, hnRNPA1, hnRNPA1-A-LCD, wt-TDP-43, h-TDP-43 and TDP-43-PLD as a function of temperature (renormalized by the highest critical temperature of the protein set, $T_c' = 472K$ for h-TPD-43). Filled circles indicate the obtained γ from DC simulations (see section SIV in the Supp. Info. for further details on the calculations) and solid curves the $\gamma \propto (T_c - T)^{1.26}$ fit to our data[132] (dashed curves depict the predicted surface tension at low T extrapolated from the fit). Empty triangles represent the obtained (renormalized) critical temperatures of each sequence using the law of rectilinear diameters and critical exponents as in Fig. 1B. (B) Snapshots of Direct Coexistence simulations of the three full sequences at $T = 0.9T_c$ (meaning T_c the critical temperature of each protein): FUS (left, $T = 360K$), wt-TDP-43 (center, $T = 410K$) and hnRNPA1 (right, $T = 390K$). FUS, wt-TDP-43 and hnRNPA1 prion-like domains are highlighted in orange, bright green and cyan respectively, while the rest of their sequences in purple, dark green and dark blue respectively. The structure of the condensates clearly show the contrast between homogeneously distributed PLD domains as in FUS, clustered PLD domains as in hnRNPA1, or interfacially exposed PLD domains as in wt-TDP-43 condensates. (C) Frequency amino acid contact maps of FUS (left), wt-TDP-43 (center) and hnRNPA1 (right) droplets at $T = 0.9T_c$. Scale bars indicate the averaged percentage of amino acid contact pairs per protein (See section SVI in the Supp. Info. for further details). Dashed lines depict the limits of the different protein domains as indicated in Fig. 1A. (D) Protein radius of gyration distribution function of the three sequences at $T/T_c = 0.9$ and at the bulk equilibrium coexisting density of the diluted (dashed curves) and the condensed phase (continuous curves). (E) Protein radius of gyration distribution function within the condensates at moderate ($T/T_c = 0.85$) and high temperature ($T/T_c = 0.95$).

based on the tie-line width of the phase diagrams) as required in experimental setups[51,139]. We find that the conformational difference in the 40-residue helical region of the TDP-43-PLD terminal domain has significant consequences on the droplet surface tension. For the whole range of studied temperatures, wt-TDP-43 shows smaller γ than h-TDP-43. At the same temperature, the presence of the helical structure in h-TDP-43 promotes a more compact assembly of proteins in the condensed phase, increasing the surface tension. Additionally, TDP-43-PLD droplets present much smaller γ than those of any of its two full-sequence variants at moderate temperatures, explaining why TDP-43-PLD domains are markedly exposed towards the interface in wt-TDP-43 condensates (Fig. 2B). Similarly, the surface tension of FUS-PLD droplets is lower than that of FUS (full-sequence). However, interestingly, γ for hnRNPA1 and hnRNPA1-A-LCD droplets is remarkably similar (as

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

their phase diagrams, see Fig. S2), confirming the significant importance of the hnRNPA1-A-LCD sequence in contributing to phase separation[15]. Our results clearly evidence a direct correlation between droplet surface tension and condensate stability. Proteins with higher γ can demix until higher temperatures or at lower protein concentration.

Next, we focus on the structural organization of the different protein condensates. A significant contrasting behavior between both FUS and hnRNPA1 droplets and those of TDP-43 (both variants) is observed. While both FUS and hnRNPA1 exhibit an homogeneous density droplet distribution with their PLDs indistinctly located along the condensate (although in hnRNPA1 more clustered), TDP-43 shows a highly marked heterogeneous distribution exposing its prion-like domains towards the droplet boundaries (Fig. 2B), evidencing that their PLD interactions barely favor aggregation[127,128]. This condensate arrangement allows the minimisation of the droplet surface tension and the simultaneous maximisation of its enthalpic gain (in absolute value) through a higher density of LLPS-stabilising contacts at the droplet core. In the case of wt-TDP-43, such structural heterogeneity is that pronounced, that condensates split into smaller nearly-interacting liquid droplets as shown in Fig. 2B (center). Conversely, the α-helix structure of h-TDP-43 notably favors the interaction between helical domains, and hence, between the rest of the intrinsically disordered neighbour regions significantly enhancing PLD connectivity, and thus, reducing droplet heterogeneity as experimentally suggested[127]. Moreover, our simulations show that that the structured α–helical domain considerably reduces the local density fluctuations of the droplet and further stabilises the condensate (Fig. 1B).

To rationalise the driving molecular forces behind these structural differences, we compute 1) the amino acid contact map frequency of the proteins within the condensates (Fig. 2C and S6) and 2) the most persistent residue-residue pair interactions along the aggregated proteins (Fig. S7-9). We develop a smart cut-off analysis of each specific residue-residue interaction (adapted to the range of the HPS potential[100], see section SVI in the Supp. Info. for further details) to elucidate the key molecular interactions promoting LLPS according to our model[100,124].

In FUS condensates, the most repeated contacts are G-G, R-Y and G-Y (Fig. S7A) highlighting how hydrophobic, cation–π, and more modestly electrostatic interactions contribute in stabilising the droplets. Since Glycine (G) represents nearly the 30% of the residues along FUS sequence, the frequency of G-G is the highest despite not being one of the strongest pair of amino acid interactions[82]. However, when normalizing the computed number of contacts by the amino acid abundance, we find that the cation–π interaction R-Y becomes the most relevant one inducing LLPS[72,73] according to this force field (see Fig. S7B). Furthermore, when analysing the FUS contact map (Fig. 2C), we observe that its prion-like domain, despite showing much lower ability to phase-separate on its own than the full protein, markedly interacts with the three RGG domains. The top contacts of the PLD alone are very different from those of the full-sequence FUS (Fig. S7A), resulting in much worse phase-separation capabilities for the PLD (Fig. 1B). Regarding the RNA-recognition motifs, we find significant LLPS-stabilising interactions among different RRM FUS domains.

While in FUS condensates the PLD plays a crucial role in LLPS[39,100], the aggregation of TDP-43 (wild-type) is mainly sustained by contacts between RRMs, either with themselves or with other protein regions such as the NTD or the NLS, but mostly dominated by RRM1-RRM1 interactions (Fig. 2C). Nonetheless, the wt-TDP-43 PLD region is still the second protein domain establishing more contacts in total after the RRM1 segment, but mostly because of its length. The three most predominant contacts in wt-TDP-43 (according to our model[100,124]) are K-F, K-E and K-D (Fig. S8A), clearly denoting the key role of cation–π and electrostatic interactions in driving condensation. However, when the structured helical region is present (h-TDP-43), R-F contacts sensibly increase becoming the third most dominant interaction. Interestingly, the renormalization of contacts by amino acid abundance in TDP-43 barely modifies the list of most frequent interactions, probably due to the very homogeneous distribution of amino acids along its sequence (Fig. S8C) when compared to that of FUS. However, similarly to FUS, TDP-43-PLD shows a completely different list of most repeated interactions compared to the full protein (Fig. S8A), which may significantly contribute to reduce its critical temperature (Fig. 1B).

In hnRNPA1, the most frequent contacts are G-G, G-S, and G-R (Fig. S9A), but since glycine is the most abundant amino acid ($\sim$ 25%) followed by serine ($\sim$ 15%), the normalized contacts by amino acid abundance show that R-Y, R-F and K-Y are dominant interactions, again highlighting the importance of cation–π interactions in hnRNPA1 LLPS. The list of top interactions of hnRNPA1-PLD, even after normalization, is very similar to that of hnRNPA1 (Fig. S9A-B), which explains why the phase diagrams of both sequences are hardly distinguishable (Fig. S2A). Surprisingly, the list of most frequent interactions of hnRNPA1-A-LCD is also remarkably similar to that of hnRNPA1 full-sequence (Fig. S9A). In fact, the detailed contact map of hnRNPA1-A-LCD corresponds to the region of hnRNPA1 that contains more LLPS-stabilising interactions (dashed lines in Fig. S6). Thus, the ability of hnRNPA1 to phase separate alone can mainly be captured by the present protein interactions in hnRNPA1-A-LCD (see Fig. S2A).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Finally, we investigate the protein conformational ensemble within the condensates and the diluted phase by computing the radius of gyration distribution function $P(R_g)$. Our simulations reveal that in all cases, when proteins transition from the diluted to the condensed phase, their conformations adopt larger radii of gyration (Figure 2D). Also, the width of $P(R_g)$ considerably increases indicating the more versatile conformations that proteins can exhibit within the condensate. This structural behavior allows proteins to maximize their number of intermolecular contacts, and thus, the droplet connectivity as recently shown in Ref.[113]. Phase-separation driven expansion for proteins undergoing homotypic LLPS has been observed for tau-IDP[140] using steady-state fluorescence measurements of pyrene and fluorescein-labeled tau-K18 proteins, a protein associated with Alzheimer disease[59]. Even if modest, phase-separation induced expansion enables IDRs to establish a surplus of enthalpy-maximizing (more energetically favourable) inter-protein contacts in the condensed phase, compared to those that they would adopt if they remained unchanged or underwent collapse. On the other hand, very recently, NMR and EPR spectroscopies have shown that the N-Terminal domain of FUS is compacted when entering in the condensed phase under agarose hydrogel conditions[141]. However, due to the employed different experimental matrix composition, our model predictions cannot be directly related to these striking observations. Now, when regarding to the protein conformational ensemble within the condensates along temperature, we note a mild change in the hnRNPA1, FUS and TDP-43 conformations as we approach to the critical T (Figure 2D), in contrast to those measured in the diluted phase as a function of T (Fig. S5 and Ref.[83]). Moreover, when comparing both TDP-43 variants $P(R_g)$ distributions, we find almost identical protein ensembles, exhibiting the wild-type variant slightly more open conformations. Nonetheless, that small surplus of extended conformations which can enable higher amount of intermolecular contacts[113] is not enough to enhance LLPS as through the $\alpha-\alpha$ helical interactions present in h-TDP-43[127].

C. RNA-induced reentrant behavior in phase separation

RNA has been recently shown to critically regulate both the phase behavior of different RNA-binding proteins[43,47,49,50,54], and most importantly the emergence of aberrant liquid-to-solid pathological phase transitions[48,59]. In this section, we explore the impact of poly-U RNA in LLPS of RBPs from a molecular and a physico-chemical perspective. By means of the novel coarse-grained model of RNA recently proposed by Regy *et al*[116] and Direct Coexistence simulations[129–131], we characterize the condensate stability of different RNA-binding proteins (and domains) from low to moderately high poly-U concentration regimes. We choose poly-U RNA for simplicity[82] and to follow previous landmark works on RNA-RBP phase separation[43,54].

First, we mix poly-U RNA strands of 250 nucleotides with the proteins studied above. Remarkably, not all proteins were able to favorably interact with poly-U in our simulations. We find that FUS-PLD and TDP-43 (including both variants) do not associate with poly-U even at very low RNA concentration (i.e., $\sim$ 0.05 mg poly-U/mg protein). We further test the affinity of wt-TDP-43 with poly-U strands by performing a separate analysis of each of its major protein sequence domains (PLD, RRM1 and RRM2). None of these domains exhibited a conclusive interaction with poly-U at temperatures moderately below the critical one. That is not entirely surprising since: 1) Several experimental studies have shown that TDP-43-RRM1 only presents strong affinity for RNA strands enriched in UG nucleotides[142–144], and 2) TDP-43-RNA heterotypic interactions are mainly driven by the RRM1, whereas the RRM2 plays a supporting role[142]. Furthermore, in the employed model, the interactions between poly-U and TDP-43 are mainly electrostatic, and therefore, other factors such as RNA secondary and tertiary structures that might sensibly promote RRM binding to different RNA sequences are not explicitly considered[145]. On the contrary, the non-interacting behavior between FUS-PLD and poly-U strands was completely expected since the FUS PLD sequence does not present neither RNA-binding domains nor positively charged domains, thus, precluding their association.

We now evaluate the phase diagram of all proteins (or protein domains) that favorably interact with poly-U, these are: FUS, hnRNPA1, hnRNPA1-PLD, hnRNPA1-A-LCD and hnRNPA1-RRMs. In all systems except for hnRNPA1-PLD, the resulting phase behavior is similar to that shown in Fig. 3A-B for FUS (note that hnRNPA1-PLD/poly-U condensates shows a very mild LLPS enhancement at low poly-U concentration, Fig. S3 and Table S3, so hereafter, the results are just discussed for hnRNPA1-A-LCD, FUS, hnRNPA1 and hnRNPA1-RRMs). At low poly-U/protein ratios, the stability of the condensates moderately increases ($\sim$ 2% higher critical temperature), while at high concentration, the critical point decreases below the critical temperature without RNA (Fig. 3D). This re-entrant behavior has been experimentally observed for synthetic peptides such as RP_3 and SR_8 in poly-U mixtures[54] and for RNA-binding proteins such as FUS[47,50,53,54] or LAF-1[43]. Please note that, although a 2% shift in the critical temperature might seem insignificant, the actual increment in temperature according to the force field[100,116,124] may be as large as 10K, which represents a huge temperature rise when referred to the physiological cell environment. To accurately determine the specific RNA-induced temperature raise, atomistic simulations

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

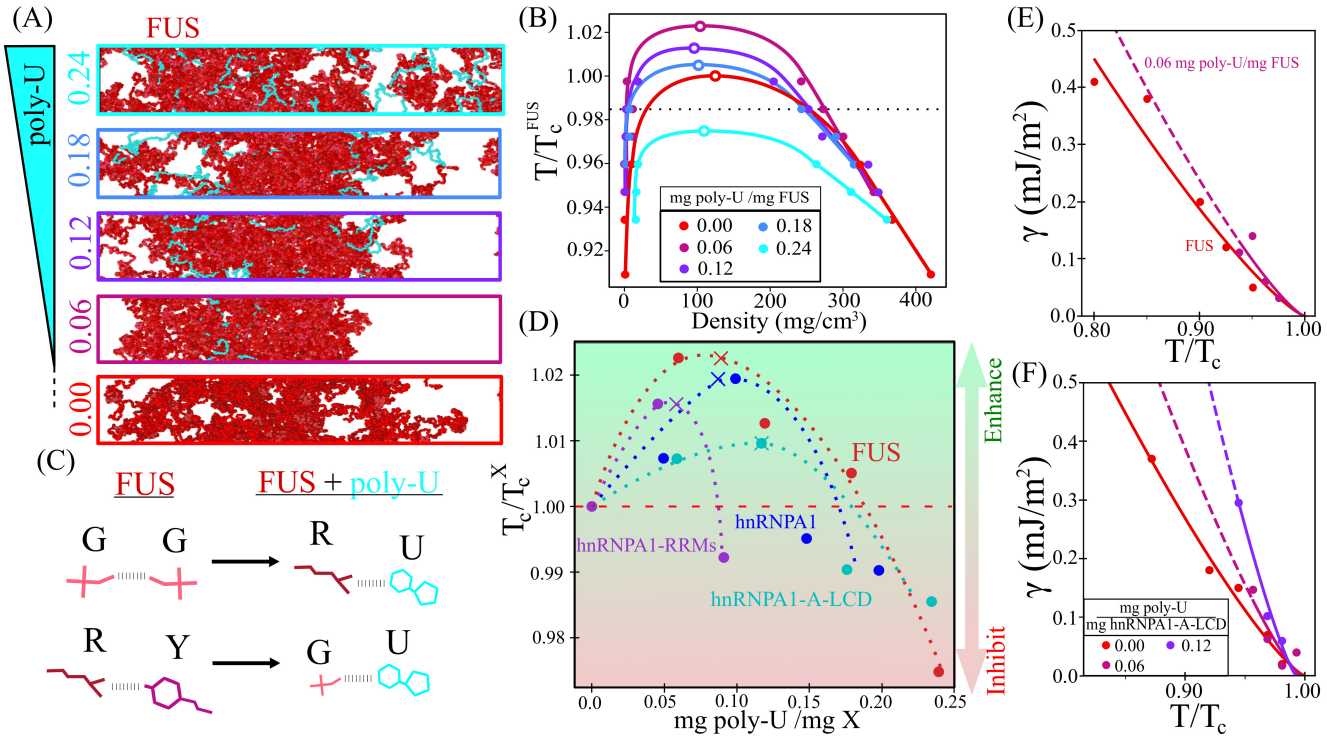

FIG. 3. RNA-induced reentrant behavior in RBP phase separation. (A) Snapshots of Direct Coexistence simulations of FUS (red) and poly-U (cyan) at temperature ($T/T_c^{FUS} \approx 0.99$, where T_c^{FUS} refers to the critical temperature of FUS in absence of poly-U) with increasing poly-U/FUS mass ratios as indicated at the left side of the simulation boxes. (B) Phase diagrams in the temperature-density plane for five different poly-U/FUS mass ratios as indicated in the legend. Empty circles represent the estimated critical point and filled circles the obtained coexisting densities from DC simulations. The horizontal dotted line depicts the temperature at which the DC snapshots shown in (A) were taken. (C) Schematic representation of the two most common contacts of FUS condensates sustaining LLPS with and without poly-U (see further details on section SVI). (D) Re-entrant behavior of several RNA-binding proteins as a function of poly-U/protein mass fraction. Filled circles depict the critical temperature (renormalized by that in absence of poly-U) of the different protein mixtures. Cross symbols indicate the poly-U/protein mass fraction at which condensates possess neutral electrostatic charge, and the horizontal red dashed line shows the limit above which phase separation is enhanced by poly-U (T_c^X is the critical temperature of the pure protein). (E) FUS droplet surface tension (γ) as a function of temperature (renormalized by T_c of the pure component) with (purple) and without (red) poly-U. Filled circles account for the obtained γ values from DC simulations, and solid lines account for the fit given by the following expression[132] $\gamma \propto (T - T_c)^{1.26}$, which can be conveniently extrapolated to moderate lower temperatures (dashed curves). (F) Same as in (E) but for hnRNPA1-A-LCD condensates with and without poly-U at two different mass fractions as indicated in the legend.

would be needed[146], although that is far beyond the current computational capability. Nonetheless, just the fact that a coarse-grained model successfully captures the experimental re-entrant behavior observed in some RBP-RNA condensates is outstanding[116]. For the studied proteins, FUS (red) exhibits the highest variation in critical temperature at either low and high RNA concentration (Fig. 3D). Interestingly, hnRNPA1 (blue) shows an intermediate behavior between that of its A-LCD (cyan) and RRM (purple) domains. The maximum critical temperature in hnRNPA1-RRM is reached at the lowest RNA concentration of the set and it sharply decays after the maximum. Contrarily, hnRNPA1-A-LCD suffers only a moderate increment of the critical temperature, but its re-entrant behavior is smoother and appears at much greater concentration (2 times higher) than that of hnRNPA1-RRM. Overall, hnRNPA1 condensates present higher RNA-induced stabilization in the low RNA regime than that of its PLD and RRMs. Moreover, it is worth mentioning that in all sequences, the larger enhancement of LLPS is reached at a poly-U concentration close to the electroneutrality point (depicted by crosses in Fig. 3D), which emphasizes the major importance of electrostatic nucleotide-amino acid interactions in RNA-RBPs phase separation[54,116].

To characterize the RNA-RBP condensates from a microscopic perspective, we analyze the key molecular contacts enabling phase separation. We find that, near the optimum poly-U/protein concentration, the most frequent contacts promoting LLPS in poly-U/FUS condensates are now R-U and G-U (Fig. 3C and S11A).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

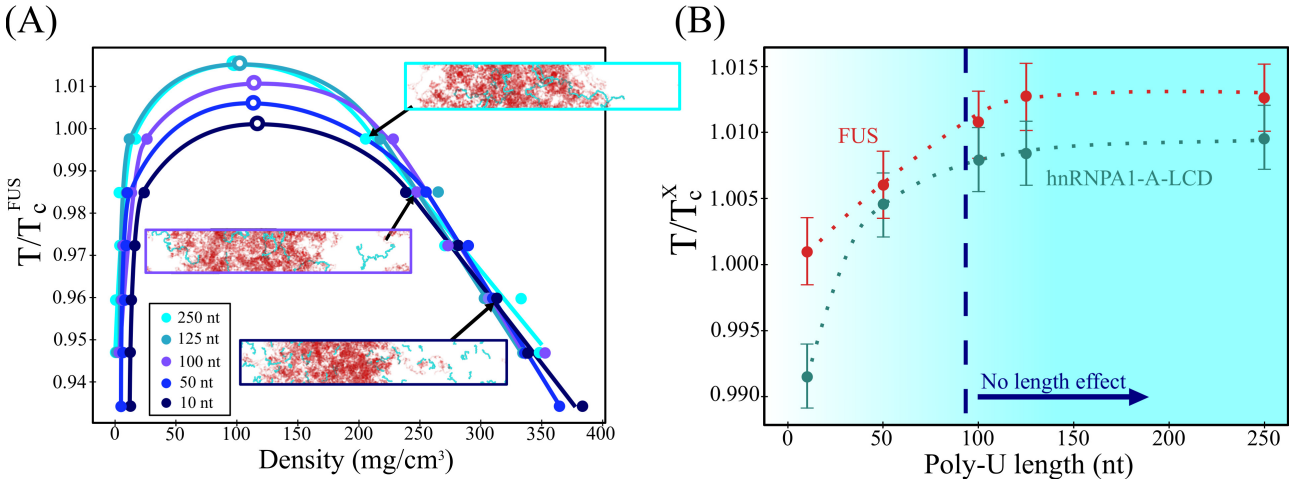

FIG. 4. Condensate stability dependence on poly-U RNA length. (A) Phase diagrams in the temperature-density plane for poly-U/FUS mixtures of different poly-U strand lengths (as indicated in the legend) at a constant concentration of 0.119 mg poly-U/mg FUS. Temperature is normalized by the critical one of FUS (T_c^{FUS}) without poly-U. DC snapshots of three representative cases of poly-U/FUS mixtures at the temperature indicated by the arrow and with poly-U lengths as depicted by the box side color (see legend) are also included. (B) Renormalized critical temperature of poly-U/FUS (red) and poly-U/hnRNPA1-A-LCD (green) condensates as a function of poly-U length for a constant concentration of 0.119 mg poly-U/mg FUS and 0.117 mg poly-U/mg hnRNPA1-A-LCD respectively. Temperature is normalized by the corresponding critical temperature (T_c^X) of each protein in absence of poly-U. The vertical dashed line indicates the minimum RNA length required to maximise droplet stability at this given concentration.

This clearly demonstrates how poly-U (at low fraction) plays a major role in sustaining the condensates, given that the two most frequent contacts are now shifted from G-G and R-Y, to the electrostatic cation-anion R-U interaction, and the G-U interaction. In terms of sequence domains, the RGG regions of FUS are those presenting more contacts with the poly-U strands, explaining why G-U becomes one of the most dominant molecular contacts by proximity (Fig. S10B). On the other hand, in our simulations, the RNA-recognition motif and the PLD region are the domains of FUS that present less favorable interactions with poly-U. Whether that is caused by model deficiencies (lack of secondary- or tertiary-driven interactions), and/or due to the fact that poly-U strands are not specifically recognized by the RRM or ZF domains needs to be further tested[120,147]. Regarding poly-U/hnRNPA1 droplets, our simulations reveal that G-G remains as the dominant amino acid pair interaction (although it substantially decreases by a factor of two), and R-U and G-U become the next two following most frequent contacts (further details in section SVI). However, the behavior of poly-U/hnRNPA1-A-LCD condensates is radically different, despite its phase diagram being altered by poly-U addition, the most frequent contacts remain similar to those in absence of RNA with a very modest excess contribution in R-U interactions (Fig. S11). On the contrary, when just considering the RRM1-RMM2 hnRNPA1 domains (purple curve in Fig. 3D), even at the lowest RNA-protein ratio where the droplet stability attains its maximum value, R-U and K-U emerge as some of the most frequent contacts in spite of the very modest poly-U concentration (Fig. S11). Finally, if we examine the contact map between poly-U and different hnRNPA1 (full-sequence) domains, we strikingly observe that the PLD comprises the highest amount of interactions with poly-U strands. However, that fact is explained through the longer length of the PLD with respect to the two RNA-recognition motifs. Yet, the strongest electrostatic interactions (mainly R-U and K-U) between hnRNPA1 and poly-U are those held through the two RRM domains (Fig. S10B).

We also determine the surface tension (γ) of the condensates in presence of poly-U as a function of temperature (Fig. 3E-F). Either for FUS (E) and hnRNPA1-A-LCD (F) condensates, we observe that poly-U at low concentration significantly increases the droplet surface tension besides further stabilizing the droplets as shown in Fig. 3D. Our simulations suggest that the molecular origin behind such surface tension increase comes from the reallocation of the positively charged residues (R, H and K) within the bulk condensate to maximize the molecular connectivity with poly-U, rather than remaining more exposed to the interface as in the pure component, and therefore, contributing to minimise the droplet surface tension due to their higher hydrophobicity. On the contrary, at moderately high poly-U ratios, the surface tension seems to decrease, although the scattering of our data does not allow us to conclude whether a non-monotonic behavior in γ may also exist.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To further elucidate the role of RNA-regulated RBPs condensate stability, we now focus on the effect of poly-U length in LLPS. A landmark study by Maharana et al.[50] showed that smaller RNAs were more potent than larger ones in solubilizing FUS condensates. On the other hand, Zacco et al.[57] found that longer RNA repeats presented weaker dissociation constants with N-RRM1-2 domains of TDP-43 that 3-fold shorter RNA strands. Given the critical role that RNA performs on the behavior of many different RBP organelles[15,43,55], we investigate the role of RNA length by selecting the poly-U/protein mass ratio which maximises droplet stability ($\sim$ 0.12 mg RNA/mg protein) for FUS and hnRNPA1-A-LCD sequences (Fig. 3D), and by introducing poly-U strands of different lengths (i.e, 10, 50, 100, 125 and 250 nucleotides) at that given concentration. Our simulations reveal that very short poly-U strands ($\sim$ 10 nt) do not enhance phase separation in FUS and hnRNPA1-A-LCD droplets (Fig. 4A-B). In fact, 10-nt poly-U strands in hnRNPA1-A-LCD droplets inhibit LLPS even at low concentration. On the other hand, we strikingly observe that RNA strands longer than $\sim$100 uridines (hereafter called minimal critical length) promote a similar droplet stabilization independently of their length (Fig. 4B). This observed minimal critical length could be likely modulated by some protein/RNA specific features such as RNA sequence, secondary structure interactions, protein charge distribution or RRM patterning effects[48]. A fully characterisation (and rationalization) of the critical aspects controlling RBP-RNA aggregation, such as the RNA length dependence studied here, may provide highly valuable insights for designing therapeutic RNA strategies to combat neurodegenerative diseases whose development is linked to aberrant accumulation and solidification of RBP condensates[60,136].

D. RNA modulates transport properties of RBP condensates

Besides controlling condensate stability, RNA has been proved to play a critical role on regulating the dynamics of many membraneless organelles[15,48,50]. A seminal study of Zhang *et al.*[55] showed that the RNA-binding protein Whi3 phase separates into liquid-like droplets whose biophysical properties can be tuned by changing the concentration of the mRNA binding partner, showing that larger RNA content increases Whi3 droplet viscosity. On the other hand, RNA has been also observed to provoke the opposite effect in LAF-1 condensates when short strands (50 nt) were introduced[43]. Nonetheless, when long RNAs were used (up to 3,000 nt), LAF-1 condensates presented significantly higher viscosity[46]. Moreover, beyond length, RNA sequence can be also an important factor in modulating droplet dynamics[117]. However, a full understanding of the precise effect of RNA in different RBP condensates still requires further work[52]. Here we aim to provide new molecular insights on this discussion by measuring the viscosity and the protein mobility of several RBP condensates as a function of poly-U concentration and for different poly-U lengths.

In vitro, viscosity (η) is usually obtained by bead-tracking within droplets using microrheology techniques[29,43,148,149], so that the trajectory can be registered and the mean squared displacement (MSD) of the beads calculated, and thus, their diffusion coefficient. Then, droplet viscosity is inferred from the diffusion coefficient by using the Stokes-Einstein relation[150]. However, in computer simulations we can measure both observables independently. The linear viscoelasticity of a material can be straightforwardly computed by integrating in time the relaxation modulus $G(t)$ of the system[151,152] (see Section SVII), whereas the diffusion coefficient can be extracted from the MSD of the proteins. Moreover, the direct calculation of $G(t)$ provides useful information about the underlying relaxation mechanisms of the proteins (see Fig. 5A for FUS condensates with and without poly-U), either at short times (white region) where the relaxation modes depend mostly on short-range and intramolecular interactions, or at long time-scales (beige region) where $G(t)$ is dominated by intermolecular forces and the overall conformational relaxation of the proteins.

We characterize the condensate dynamics of FUS, hnRNPA1 and hnRNPA1-A-LCD as a function of poly-U concentration at constant temperature (just below the critical T of each protein in absence of poly-U, $T/T_c \sim 0.98$) and at the corresponding bulk droplet equilibrium density corresponding to each poly-U concentration at that temperature. First, we introduce poly-U strands of 125 nucleotides. As shown in Fig. 4B, the phase diagram for a given concentration is not expected to change either by using strands of 125 or 250 nucleotides. For both FUS and hnRNPA1-A-LCD, we observe a mild non-monotonic behavior with a maximum in viscosity at low poly-U ratios (filled circles in Fig. 5B), which might be directly related to the maximum in droplet stability shown in Fig. 3D, or due to a coincidental scattering of our measurements. Nonetheless, at moderate poly-U mass ratios (i.e., > 0.15 mg poly-U/ mg protein), the viscosity of the condensates is clearly $\sim 30\%$ lower than that without poly-U. On the other hand, a monotonic decreasing trend in viscosity was detected for hnRNPA1 condensates, where almost a $\sim 50\%$ drop in η is found at moderately high poly-U mass fractions. Even though the observed maximum in viscosity could be easily related to the re-entrant behavior depicted in Fig. 3D, further work needs to be devoted to clarify whether this is a real feature of the model, and ultimately of these RBP-RNA condensates. Furthermore, we investigate how poly-U strands of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

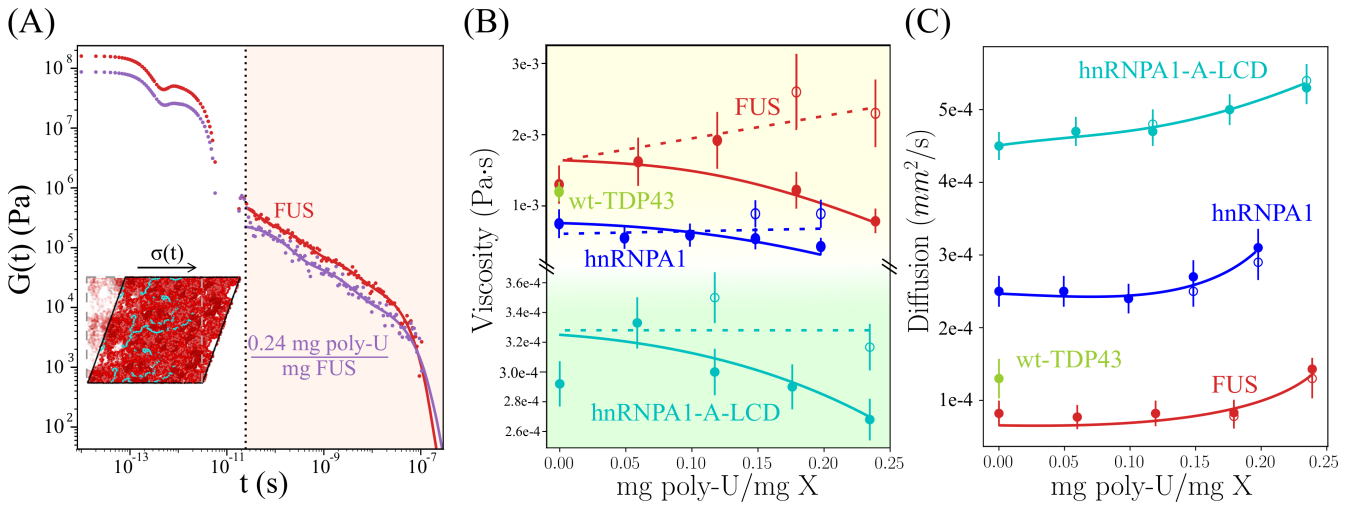

FIG. 5. RNA critically regulates the dynamical properties of RBP condensates. (A) Shear stress relaxation modulus of FUS condensates in absence (red) *versus* presence (purple) of poly-U strands of 125 nucleotides at 0.24 mg poly-U/mg FUS mass fraction, $T/T_c = 0.97$ (where T_c is the critical temperature of FUS pure condensates) and the corresponding equilibrium bulk density of each droplet at such conditions. The vertical dotted line separates the fast-decay relaxation mode regime (white) and the slow-decay relaxation mode one (beige). A snapshot illustrating a shear stress relaxation experiment over a poly-U/FUS condensate simulation box is included. (B) Viscosity of FUS (at $T/T_c = 0.97$), hnRNPA1 (at $T/T_c = 0.985$) and hnRNPA1-A-LCD (at $T/T_c = 0.98$) condensates as a function of poly-U/protein mass ratio. An estimate of wt-TDP-43 viscosity in absence of poly-U at $T/T_c = 0.97$ is also included. Filled circles depict viscosities when poly-U strands of 125 nt were used and empty circles when strands 250 nucleotides were added. Continuous and dashed lines are included as a guide for the eye for strands of 125 and 250 nt respectively. Note that T_c refers to the pure component critical temperature of each protein. Distinct background highlights the discontinuity of the y-axis employed for better visualization of the data. (C) Protein diffusion coefficient (filled circles) as a function of poly-U(125nt)/protein(X) mass ratio. Empty circles account for the protein diffusion coefficient when poly-U strands of 250 nt were added. The same system conditions described in (B) are applied on these calculations. Continuous curves are included as a guide to the eye.

250 nucleotides can regulate droplet viscosity at the same concentrations. While poly-U 125nt strands significantly reduce viscosity at moderate ratios, poly-U 250nt strands barely varies condensate viscosity at the same concentrations, except for FUS, where a moderate viscosity increase was detected (empty symbols in Fig. 5B). These observations are in full agreement with those reported for LAF-1 condensates in presence of short[43] and long[46] RNA strands. Longer RNA chains, even at low to moderate concentrations, could increase droplet viscosity even more due to their own longer relaxation time. We also note that our $G(t)$ values for FUS do not quantitatively match with those of Ref.[70]. That is somewhat expected since our coarse-grained model has been parametrized to describe the radius of gyration[100] and most frequent molecular contacts[124] between proteins rather than dynamic properties such as transport properties within the condensates. Nonetheless, the observed behavior with RNA and between different RBP condensates is expected to qualitatively hold despite the different model approximations (i.e., implicit solvent and amino acids/nucleotides represented by spherical beads).

Finally, we measure the protein diffusion within the condensates for all previous poly-U concentrations and strand lengths (125 and 250 nt). In all cases, we find a positive correlation between viscosity and protein mobility, being the latter considerably higher at moderate poly-U/protein ratios (Fig. 5C). Strikingly, protein diffusion hardly depends on poly-U strand length (empty symbols) as viscosity does (Fig. 5B). That can be rationalised by considering that the shear stress relaxation modulus of the condensates crucially depends on the RNA strand length (longer RNAs imply longer $G(t)$ relaxation decay), while the protein diffusion coefficient does not. The later only depends on droplet density (and temperature), and as shown in Fig. 4 A, condensates densities remain similar either when using strands of 125 and 250 nucleotides. Our simulations suggest that the condensate dynamics dependence on RNA concentration is intimately related to the droplet density decrease as a function of poly-U abundance, as shown in Fig. 3B. Interestingly, we also note that FUS, despite having the lower critical temperature to phase separate, and thus weaker LLPS-stabilising interactions than wt-TDP-43 and hnRNPA1 (Fig. 1B), displays the lowest protein diffusion of the set in absence of poly-U. Such intriguing fact, which might be related to patterning sequence effects[99] or just protein length[103], highlights how beyond stability, condensate dynamics also entail intricate processes that need to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be further investigated. In fact, methods promoting LLPS at lower protein concentration or enhancing protein mobility, such as by short RNA inclusion, could play therapeutic roles on preventing the emergence of pathological solid-like aggregates (by decreasing droplet density and viscosity) related to some neurodegenerative disorders as amyotrophic lateral sclerosis or multisystem proteinopathy[15,47,59].

III. CONCLUSIONS

Here, we investigate the dual effect of poly-U RNA in controlling the stability and dynamics of RNA-binding protein condensates. By means of Molecular Dynamics simulations of a novel sequence dependent coarse-grained model for proteins and RNA[100,116,124], we explore the underlying molecular and thermodynamic driving forces enabling liquid-liquid phase separation of FUS, hnRNPA1 and TDP-43 along their corresponding prion-like and RRM domains, in presence *versus* absence of poly-U strands. First, we validate the model by computing the phase diagram of the previous proteins (with no poly-U) and comparing their relative ability to phase separate in simulations (by estimating their critical temperature) with the experimental protein saturation concentration needed to observe LLPS. We find a remarkable qualitative agreement between the model predicted behavior and the experimental trend at physiological salt concentration. Moreover, for these condensates we elucidate the surface tension with the surrounding protein diluted phase, the key molecular contacts sustaining LLPS along the different sequences, the most frequent protein domain interactions, and the protein conformational distribution along the two phases.

We observe that highly inhomogeneous sequence contact maps, as that of wt-TDP-43, can lead to the emergence of largely heterogeneous droplets with reduced surface tensions compared to their homogeneous counterparts. The exposure of the PLD region of wt-TDP-43 to the droplet interface considerably contributes in lowering its γ, and thus favoring multidroplet appearance[153,154]. However, such condensate heterogeneity is significantly relieved when α–α helical PLD interactions are present, as recently hypothesized by Wang *et al.*[127]. Additionally, the analysis of the intermolecular contact maps within droplets reveals the major importance of certain sequence domain interactions in LLPS, such as hnRNPA1 PLD-PLD interactions, or FUS PLD-RGG interactions. Furthermore, amino acid contacts such as G-G, R-Y, G-S, G-Y, K-F or K-Y have been shown (Fig. S7-S9) to play a leading role in LLPS, highlighting the relevance of cation–π and electrostatic forces, besides hydrophobicity, in the physiological salt regime[58]. Also, the conformational protein ensemble inside the condensates has been demonstrated to be almost independent of temperature, in contrast to those measured in the diluted phase[83]. However, along the diluted-to-condensed transition, a significant variation to more extended conformational ensembles (to maximize protein molecular connectivity) is observed[113].

Moreover, our simulations with poly-U RNA show how the formation of protein condensates is clearly enhanced at low poly-U concentration[47], whereas inhibited at high poly-U-protein ratios[43,50,54]. The poly-U concentration which promotes the highest increase in droplet stability is near the electroneutral charge point of poly-U/FUS and poly-U/hnRNPA1 mixtures (and also for both hnRNPA1 RRMs and A-LCD regions separately) in agreement with findings for LAF-1 PLD condensates[116]. We reveal that such boost in droplet stability is related to an increase of the condensate surface tension at low poly-U ratios. In contrast, neither of the two studied TDP-43 variants, nor their RRMs together or individually, exhibited significantly LLPS enhancement through poly-U addition. Besides, we demonstrate that beyond a certain strand length of ~ 100 nucleotides, the stability of the droplets for a given poly-U concentration reaches a maximum value, whereas below that critical strand length, the effect is more moderate, and for very short lengths (< 10 nt), it can even hinder phase separation[50]. Overall, our results evidence how RBP condensate phase behavior can be drastically modulated by varying RNA concentration and/or RNA length.

Finally, we focus on the transport properties of the RBP condensates as a function of poly-U concentration and length. On the one hand, our simulations demonstrate that at low RNA/protein ratios, droplet viscosity barely changes with respect to that of pure RBP condensates, and it may even show a mild maximum correlated with the highest protein stability. On the other hand, at moderately high poly-U concentration, two different behaviors are observed. When poly-U strands of 125 nt are introduced, viscosity decreases[43], whereas when strands of 250 nucleotides are added, droplet viscosity slightly increases with RNA concentration[46,55]. These are very striking findings considering that droplet stability (phase diagram) for a given poly-U/protein ratio scarcely changes when using these two lengths, since both lengths (125 and 250 nt) are above the minimal poly-U critical length shown in Fig. 4B. However, protein diffusion within the condensates hardly depends on the poly-U length (it mainly depends on droplet density), and it increases with poly-U concentration. In that respect, the addition of moderately short RNA strands (i.e., two or three times shorter than the protein length) could help in promoting condensate dynamics without significantly destabilizing phase separation. Our results suggest that the observed enhanced droplet dynamic behavior at moderate concentrations is mediated by a condensate density reduction upon poly-U

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

addition. Taken together, our observations shed light on the crucial role of RNA (concentration and length) on the formation and phase behavior of RNA-protein complexes[52,123,155]. Moreover, the present work provides a novel estimation of the transport properties of protein condensates by means of computer simulations which could pave the way for future studies characterizing the protein mobility in other relevant systems. Expanding our understanding on LLPS and the role of RNA in this process may drive solutions to precisely modulate aberrant liquid-to-solid transitions in the cell.

IV. ACKNOWLEDGEMENTS

We acknowledge Dr R. Collepardo-Guevara and Dr J. A. Joseph for very useful discussions. This project has received funding from the Oppenheimer Research Fellowship of the University of Cambridge. A. T. is funded by Universidad Politécnica de Madrid (ESTANCIAS-PIF-20-TYOSR8-13-4N0WPQ), A. G. is funded by an EPSRC studentship (EP/N509620/1) and an Winton scholarship. J. R. acknowledges funding from the Spanish Ministry of Economy and Competitivity (PID2019-105898GA-C22). J. R. E. also acknowledges funding from the Roger Ekins Research Fellowship of Emmanuel College. This work has been performed using resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (http://www.hpc.cam.ac.uk) funded by EPSRC Tier-2 capital grant EP/P020259/1.

V. REFERENCES

[1] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Science **324**, 1729 (2009).
[2] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, *et al.*, Nature **483**, 336 (2012).
[3] A. A. Hyman, C. A. Weber, and F. Jülicher, Annual review of cell and developmental biology **30**, 39 (2014).
[4] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Nature reviews Molecular cell biology **18**, 285 (2017).
[5] Y. Shin and C. P. Brangwynne, Science **357** (2017).
[6] S. Alberti, A. Gladfelter, and T. Mittag, Cell **176**, 419 (2019).
[7] T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, *et al.*, Molecular cell **57**, 936 (2015).
[8] G. J. Narlikar, Journal of Biosciences **45**, 5 (2020).
[9] C. P. Brangwynne, P. Tompa, and R. V. Pappu, Nature Physics **11**, 899 (2015).
[10] C. S. Leblond, H. M. Kaneb, P. A. Dion, and G. A. Rouleau, Experimental neurology **262**, 91 (2014).
[11] D. S. Protter, B. S. Rao, B. Van Treeck, Y. Lin, L. Mizoue, M. K. Rosen, and R. Parker, Cell reports **22**, 1401 (2018).
[12] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, *et al.*, Trends in cell biology **28**, 420 (2018).
[13] H. Zhou, Z. Song, S. Zhong, L. Zuo, Z. Qi, L. J. Qu, and L. Lai, Angewandte Chemie - International Edition **58**, 4858 (2019).
[14] S. Saha, C. A. Weber, M. Nousch, O. Adame-Arana, C. Hoege, M. Y. Hein, E. Osborne-Nishimura, J. Mahamid, M. Jahnel, L. Jawerth, A. Pozniakovski, C. R. Eckmann, F. Jülicher, and A. A. Hyman, Cell **166**, 1572 (2016).
[15] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, Cell **163**, 123 (2015).
[16] X. Gui, F. Luo, Y. Li, H. Zhou, Z. Qin, Z. Liu, J. Gu, M. Xie, K. Zhao, B. Dai, *et al.*, Nature communications **10**, 1 (2019).
[17] X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, and R. D. Vale, Science **352**, 595 (2016).
[18] A. G. Larson and G. J. Narlikar, Biochemistry **57**, 2540 (2018).
[19] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, Nature **547**, 236 (2017).
[20] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, Nature **547**, 241 (2017).
[21] B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young, Science **361** (2018).
[22] H. Yoo, C. Triandafillou, and D. A. Drummond, Journal of Biological Chemistry **294**, 7151 (2019).
[23] A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner, Science **367**, 464 (2020).
[24] J. Sheu-Gruttadauria and I. J. MacRae, Cell **173**, 946 (2018).
[25] T. M. Franzmann and S. Alberti, Journal of Biological Chemistry **294**, 7128 (2019).
[26] S. Kroschwald, M. C. Munder, S. Maharana, T. M. Franzmann, D. Richter, M. Ruer, A. A. Hyman, and S. Alberti, Cell reports **23**, 3327 (2018).
[27] J. J. Bouchard, J. H. Otero, D. C. Scott, E. Szulc, E. W. Martin, N. Sabri, D. Granata, M. R. Marzahn, K. Lindorff-Larsen, X. Salvatella, *et al.*, Molecular cell **72**, 19 (2018).
[28] J. J. Moser and M. J. Fritzler, The international journal of biochemistry & cell biology **42**, 828 (2010).
[29] M. Feric and C. P. Brangwynne, Nature cell biology **15**, 1253 (2013).
[30] J. G. Gall, Nature reviews Molecular cell biology **4**, 975 (2003).
[31] C. P. Brangwynne, T. J. Mitchison, and A. A. Hyman, Proceedings of the National Academy of Sciences **108**, 4334 (2011).
[32] H. Lu, D. Yu, A. S. Hansen, S. Ganguly, R. Liu, A. Heckert, X. Darzacq, and Q. Zhou, Nature **558**, 318 (2018).
[33] Y. Chen and A. S. Belmont, Current opinion in genetics & development **55**, 91 (2019).
[34] P. A. Chong and J. D. Forman-Kay, Current opinion in structural biology **41**, 180 (2016).
[35] T. J. Welsh, Y. Shen, A. Levin, and T. P. Knowles, Cell **175**, 1457 (2018).
[36] Y. Shin, Y.-C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, and C. P. Brangwynne, Cell **175**, 1481 (2018).
[37] J. A. Riback and C. P. Brangwynne, Science **367**, 364 (2020).
[38] S. Qamar, G. Wang, S. J. Randle, F. S. Ruggeri, J. A. Varela, J. Q. Lin, E. C. Phillips, A. Miyashita, D. Williams, F. Ströhl, *et al.*, Cell **173**, 720 (2018).
[39] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi, Nature structural & molecular biology **26**, 637 (2019).
[40] S. N. Rhoads, Z. T. Monahan, D. S. Yee, and F. P. Shewmaker, International journal of molecular sciences **19**, 886 (2018).
[41] H.-R. Li, W.-C. Chiang, P.-C. Chou, W.-J. Wang, and J.-r. Huang, Journal of Biological Chemistry **293**, 6090 (2018).
[42] L. McGurk, E. Gomes, L. Guo, J. Mojsilovic-Petrovic, V. Tran, R. G. Kalb, J. Shorter, and N. M. Bonini, Molecular cell **71**, 703 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[43]S. Elbaum-Garfinkle, Y. Kim, K. Szczepaniak, C. C.-H. Chen, C. R. Eckmann, S. Myong, and C. P. Brangwynne, Proceedings of the National Academy of Sciences **112**, 7189 (2015).

[44]A. Cléry, M. Blatter, and F. H. Allain, Current opinion in structural biology **18**, 290 (2008).

[45]S. Sanulli, M. Trnka, V. Dharmarajan, R. Tibble, B. Pascal, A. Burlingame, P. Griffin, J. Gross, and G. Narlikar, Nature **575**, 390 (2019).

[46]M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, Nature Chemistry **9**, 1118 (2017).

[47]K. A. Burke, A. M. Janke, C. L. Rhine, and N. L. Fawzi, Molecular cell **60**, 231 (2015).

[48]L. Guo and J. Shorter, Molecular cell **60**, 189 (2015).

[49]I. Alshareedah, T. Kaur, J. Ngo, H. Seppala, L.-A. D. Kounatse, W. Wang, M. M. Moosa, and P. R. Banerjee, Journal of the American Chemical Society **141**, 14593 (2019).

[50]S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillén-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y.-T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman, and S. Alberti, Science **360**, 918 (2018).

[51]I. Alshareedah, G. M. Thurston, and P. R. Banerjee, Biophysical Journal (2021).

[52]M. Yu and E. A. Lemke, Biophysical Journal (2021).

[53]J. C. Schwartz, X. Wang, E. R. Podell, and T. R. Cech, Cell reports **5**, 918 (2013).

[54]P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, and A. A. Deniz, Angewandte Chemie **129**, 11512 (2017).

[55]H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, and A. S. Gladfelter, Molecular cell **60**, 220 (2015).

[56]F. Muzzopappa, M. Hertzog, and F. Erdel, Biophysical Journal (2021).

[57]E. Zacco, R. Graña-Montes, S. R. Martin, N. S. de Groot, C. Alfano, G. G. Tartaglia, and A. Pastore, Journal of molecular biology **431**, 1671 (2019).

[58]J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, *et al.*, Cell **174**, 688 (2018).

[59]A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, *et al.*, Cell **162**, 1066 (2015).

[60]Y. R. Li, O. D. King, J. Shorter, and A. D. Gitler, Journal of cell biology **201**, 361 (2013).

[61]T. Afroz, E.-M. Hock, P. Ernst, C. Foglieni, M. Jambeau, L. A. Gilhespy, F. Laferriere, Z. Maniecka, A. Plückthun, P. Mittl, *et al.*, Nature communications **8**, 1 (2017).

[62]Y. Lu, L. Lim, and J. Song, Scientific Reports **7**, 1 (2017).

[63]S. Wegmann, B. Eftekharzadeh, K. Tepper, K. M. Zoltowska, R. E. Bennett, S. Dujardin, P. R. Laskowski, D. MacKenzie, T. Kamath, C. Commins, *et al.*, The EMBO journal **37**, e98049 (2018).

[64]J.-M. Choi, A. S. Holehouse, and R. V. Pappu, Annual Review of Biophysics **49**, 107 (2020).

[65]B. Portz, B. L. Lee, and J. Shorter, Trends in Biochemical Sciences (2021).

[66]J. J. Sakon and K. R. Weninger, Nature methods **7**, 203 (2010).

[67]I. König, A. Zarrine-Afsar, M. Aznauryan, A. Soranno, B. Wunderlich, F. Dingfelder, J. C. Stüber, A. Plückthun, D. Nettels, and B. Schuler, nAture methods **12**, 773 (2015).

[68]S. J. LeBlanc, P. Kulkarni, and K. R. Weninger, Biomolecules **8**, 140 (2018).

[69]J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, and T. R. Sosnick, Science **358**, 238 (2017).

[70]L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, *et al.*, Science **370**, 1317 (2020).

[71]V. H. Ryan, G. L. Dignon, G. H. Zerze, C. V. Chabata, R. Silva, A. E. Conicella, J. Amaya, K. A. Burke, J. Mittal, and N. L. Fawzi, Molecular cell **69**, 465 (2018).

[72]Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, *et al.*, The EMBO journal **36**, 2951 (2017).

[73]M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, *et al.*, Cell **149**, 753 (2012).

[74]G. L. Dignon, R. B. Best, and J. Mittal, Annual Review of Physical Chemistry **71**, 53 (2020).

[75]S. Roberts, V. Miao, S. Costa, J. Simon, G. Kelly, T. Shah, S. Zauscher, and A. Chilkoti, Nature communications **11** (2020).

[76]M. Paloni, R. Bailly, L. Ciandrini, and A. Barducci, The Journal of Physical Chemistry B **124**, 9009 (2020).

[77]M. Feig and Y. Sugita, Annual review of cell and developmental biology **35**, 191 (2019).

[78]D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, Proc. Natl. Acad. Sci. U.S.A **82**, 36 (2006).

[79]K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011).

[80]R. Collepardo-Guevara, G. Portella, M. Vendruscolo, D. Frenkel, T. Schlick, and M. Orozco, Journal of the American Chemical Society **137**, 10205 (2015).

[81]D. A. Potoyan and G. A. Papoian, Proceedings of the National Academy of Sciences **109**, 17857 (2012).

[82]G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M. A. Czekalska, *et al.*, Nature Communications **12**, 1 (2021).

[83]G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proceedings of the National Academy of Sciences **115**, 9929 (2018).

[84]B. S. Schuster, G. L. Dignon, W. S. Tang, F. M. Kelley, A. K. Ranganath, C. N. Jahnke, A. G. Simpkins, R. M. Regy, D. A. Hammer, M. C. Good, and J. Mittal, Proceedings of the National Academy of Sciences **117**, 11421 (2020).

[85]J. Lee, Y. O. Popov, and G. H. Fredrickson, The Journal of chemical physics **128**, 224908 (2008).

[86]Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, and H. S. Chan, New Journal of Physics **19**, 115003 (2017).

[87]J. McCarty, K. T. Delaney, S. P. Danielsen, G. H. Fredrickson, and J.-E. Shea, The journal of physical chemistry letters **10**, 1644 (2019).

[88]J.-M. Choi, A. A. Hyman, and R. V. Pappu, Physical Review E **102** (2020).

[89]C. A. Weber, D. Zwicker, F. Jülicher, and C. F. Lee, Reports on Progress in Physics **82**, 064601 (2019).

[90]T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, elife **6**, e30294 (2017).

[91]S. Das, A. Eisen, Y.-H. Lin, and H. S. Chan, The Journal of Physical Chemistry B **122**, 5418 (2018).

[92]T. S. Harmon, A. S. Holehouse, and R. V. Pappu, New Journal of Physics **20**, 045002 (2018).

[93]J.-M. Choi, F. Dar, and R. V. Pappu, PLOS Computational Biology **15**, e1007028 (2019).

[94]H. Liu, S. K. Kumar, and F. Sciortino, The Journal of chemical physics **127**, 084902 (2007).

[95]J. Russo, P. Tartaglia, and F. Sciortino, The Journal of chemical physics **131**, 014504 (2009).

[96]V. Nguemaha and H.-X. Zhou, Scientific reports **8**, 1 (2018).

[97]H.-Y. Chou and A. Aksimentiev, The Journal of Physical Chemistry Letters **11**, 4923 (2020).

[98]J. A. Joseph, J. R. Espinosa, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, Biophysical Journal **120**, 1 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434111; this version posted April 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[99] A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, The Journal of Chemical Physics **152**, 075101 (2020).

[100] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLoS computational biology **14**, e1005941 (2018).

[101] T. M. Perdikari, N. Jovic, G. L. Dignon, Y. C. Kim, N. L. Fawzi, and J. Mittal, Biophysical Journal (2021), 10.1016/j.bpj.2021.01.034.

[102] F. J. Blas, L. G. MacDowell, E. de Miguel, and G. Jackson, The Journal of chemical physics **129**, 144703 (2008).

[103] K. S. Silmore, M. P. Howard, and A. Z. Panagiotopoulos, Molecular Physics **115**, 320 (2017).

[104] R. K. Das and R. V. Pappu, Proceedings of the National Academy of Sciences **110**, 13392 (2013).

[105] M. K. Hazra and Y. Levy, Physical Chemistry Chemical Physics **22**, 19368 (2020).

[106] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, Physical review letters **97**, 168301 (2006).

[107] S. Banjade, Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu, and M. K. Rosen, Proceedings of the National Academy of Sciences **112**, E6426 (2015).

[108] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, Science **367**, 694 (2020).

[109] K. M. Ruff, F. Dar, and R. V. Pappu, bioRxiv (2020).

[110] J. R. Espinosa, A. Garaizar, C. Vega, D. Frenkel, and R. Collepardo-Guevara, J. Chem. Phys **150**, 224510 (2019).

[111] J. Russo, J. Tavares, P. Teixeira, M. T. da Gama, and F. Sciortino, The Journal of chemical physics **135**, 034501 (2011).

[112] F. J. Blas, A. Galindo, and C. Vega, Molecular Physics **101**, 449 (2003).

[113] A. Garaizar, I. Sanchez-Burgos, R. Collepardo-Guevara, and J. R. Espinosa, Molecules **25**, 4705 (2020).

[114] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, Proceedings of the National Academy of Sciences **117**, 13238 (2020).

[115] F. Dar and R. V. Pappu, Biophysical Journal **118**, 213a (2020).

[116] R. M. Regy, G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, Nucleic Acids Research **48**, 12593 (2020).

[117] S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V. Pappu, *et al.*, Proceedings of the National Academy of Sciences **116**, 7889 (2019).

[118] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, Biomolecules **11**, 278 (2021).

[119] R. Damman, S. Schütz, Y. Luo, M. Weingarth, R. Sprangers, and M. Baldus, Nature communications **10**, 1 (2019).

[120] J. Šponer, M. Krepl, P. Banáš, P. Kührová, M. Zgarbová, P. Jurečka, M. Havrila, and M. Otyepka, Wiley Interdisciplinary Reviews: RNA **8**, e1405 (2017).

[121] C. M. Reyes and P. A. Kollman, Journal of molecular biology **297**, 1145 (2000).

[122] T. Kaur, M. Raju, I. Alshareedah, R. B. Davis, D. A. Potoyan, and P. R. Banerjee, bioRxiv (2020).

[123] I. Alshareedah, M. M. Moosa, M. Raju, D. A. Potoyan, and P. R. Banerjee, Proceedings of the National Academy of Sciences **117**, 15650 (2020).

[124] S. Das, Y.-H. Lin, R. M. Vernon, J. D. Forman-Kay, and H. S. Chan, Proceedings of the National Academy of Sciences **117**, 28795 (2020).

[125] T. J. Welsh, G. Krainer, J. R. Espinosa, J. A. Joseph, A. Sridhar, M. Jahnel, W. E. Arter, K. L. Saar, S. Alberti, R. Collepardo-Guevara, and T. P. Knowles, bioRxiv (2020), 10.1101/2020.04.20.047910.

[126] L. Lim, Y. Wei, Y. Lu, and J. Song, PLOS Biology **14**, 1 (2016).

[127] A. Wang, A. E. Conicella, H. B. Schmidt, E. W. Martin, S. N. Rhoads, A. N. Reeb, A. Nourse, D. Ramirez Montero, V. H. Ryan, R. Rohatgi, *et al.*, The EMBO journal **37**, e97452 (2018).

[128] A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, M. Alexandra, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal, and N. L. Fawzi, Proceedings of the National Academy of Sciences **117**, 5883 (2020).

[129] A. J. Ladd and L. V. Woodcock, Chemical Physics Letters **51**, 155 (1977).

[130] R. García Fernández, J. L. Abascal, and C. Vega, The Journal of chemical physics **124**, 144506 (2006).

[131] J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, The Journal of chemical physics **139**, 144502 (2013).

[132] J. S. Rowlinson and B. Widom, *Molecular theory of capillarity* (Courier Corporation, 2013).

[133] W. M. Babinchak, R. Haider, B. K. Dumm, P. Sarkar, K. Surewicz, J.-K. Choi, and W. K. Surewicz, Journal of Biological Chemistry **294**, 6306 (2019).

[134] N. Farahi, T. Lazar, S. J. Wodak, P. Tompa, and R. Pancsa, bioRxiv (2021).

[135] Q. Li, X. Peng, Y. Li, W. Tang, J. Zhu, J. Huang, Y. Qi, and Z. Zhang, Nucleic acids research **48**, D320 (2020).

[136] Y. Lin, D. S. Protter, M. K. Rosen, and R. Parker, Molecular Cell **60**, 208 (2015).

[137] U. Choi, H. Sanabria, T. Smirnova, M. Bowen, and K. Weninger, Biomolecules **9**, 114 (2019).

[138] U. B. Choi, J. J. McCann, K. R. Weninger, and M. E. Bowen, *Structure*, Structure **19**, 566 (2011).

[139] M. Vis, E. M. Blokhuis, B. H. Ern, R. H. Tromp, and H. N. Lekkerkerker, The Journal of Physical Chemistry B **122**, 3354 (2017).

[140] A. Majumdar, P. Dogra, S. Maity, and S. Mukhopadhyay, The Journal of Physical Chemistry Letters **10**, 3929 (2019).

[141] L. Emmanouilidis, L. Esteban-Hofer, F. Damberger, and *et al.*, Nat Chem Biol (2021).

[142] P.-H. Kuo, C.-H. Chiang, Y.-T. Wang, L. G. Doudeva, and H. S. Yuan, Nucleic acids research **42**, 4712 (2014).

[143] P. J. Lukavsky, D. Daujotyte, J. R. Tollervey, J. Ule, C. Stuani, E. Buratti, F. E. Baralle, F. F. Damberger, and F. H. Allain, Nature structural & molecular biology **20**, 1443 (2013).

[144] E. Zacco, S. R. Martin, R. Thorogate, and A. Pastore, Frontiers in molecular neuroscience **11**, 372 (2018).

[145] A. Bhardwaj, M. P. Myers, E. Buratti, and F. E. Baralle, Nucleic acids research **41**, 5062 (2013).

[146] S. Piana, P. Robustelli, D. Tan, S. Chen, and D. E. Shaw, Journal of chemical theory and computation **16**, 2494 (2020).

[147] M. Krepl, A. Cléry, M. Blatter, F. H. Allain, and J. Sponer, Nucleic acids research **44**, 6452 (2016).

[148] T. G. Mason and D. A. Weitz, Physical review letters **74**, 1250 (1995).

[149] R. S. Fisher and S. Elbaum-Garfinkle, Nature communications **11**, 1 (2020).

[150] D. Lavalette, C. Tétreau, M. Tourbez, and Y. Blouquit, Biophysical journal **76**, 2744 (1999).

[151] M. Rubinstein, R. H. Colby, *et al.*, *Polymer physics*, Vol. 23 (Oxford university press New York, 2003).

[152] J. Ramírez, S. K. Sukumaran, B. Vorselaars, and A. E. Likhtman, The Journal of chemical physics **133**, 154103 (2010).

[153] M. G.-J. Navarro, S. Kashida, R. Chouaib, S. Souquere, G. Pierron, D. Weil, and Z. Gueroui, Nature communications **10**, 1 (2019).

[154] T. Lu and E. Spruijt, Journal of the American Chemical Society **142**, 2905 (2020).

[155] H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, and A. S. Gladfelter, Molecular Cell **60**, 220 (2015).