Condensates in RNA Repeat Sequences are Heterogeneously Organized and Exhibit Reptation-like Dynamics

Hung T. Nguyen[1], **Naoto Hori**[2], **and D. Thirumalai**[1,*]

[1]Department of Chemistry, The University of Texas at Austin, Austin, TX, US, 78712
[2]School of Pharmacy, University of Nottingham, Nottingham, UK
[*]dave.thirumalai@gmail.com

ABSTRACT

Although it is known that RNA undergoes liquid–liquid phase separation (LLPS), the interplay between the molecular driving forces and the emergent features of the condensates, such as their morphologies and dynamical properties, are not well understood. We develop a coarse-grained model to investigate the phase separation of trinucleotide repeat RNAs, which are implicated in neurological disorders such as Huntington disease and amyotrophic lateral sclerosis. After establishing that the simulations reproduce key experimental findings (length and concentration dependence of the phase transition in $(CAG)_n$ repeats), we show that once recruited inside the liquid droplets, the monomers transition from hairpin-like structures to extended states. Interactions between the monomers in the condensates result in the formation of an intricate and dense intermolecular network, which severely restrains the fluctuation and mobility of the RNAs inside large droplets. In the largest densely packed high viscosity droplets, the mobility of RNA chains is best characterized by reptation, reminiscent of the dynamics in polymer melts.

Introduction

The discovery that germline P granules form liquid droplets resembling compartments without membranes[1] has resulted in a revolution, providing an impetus for a deeper understanding of how the cytoplasm and nucleus are organized. In turn, this finding has inspired a large number of studies in a variety of unrelated systems that undergo liquid–liquid phase separation (LLPS), postulated to be the mechanism by which such compartments form.[2–14] A vast majority of such studies have focused on liquid organelles composed of intrinsically disordered proteins/regions (IDP/IDR) or IDPs interacting with RNA molecules that could engage in multivalent interactions.[5] These findings are typically explained using analogies to sol-gel phase transitions known in synthetic polymers, although there are differences due to the heteropolymer nature of biological sequences.[4,15–18] However, the molecular details of how proteins and nucleic acids, with diverse sequences, coalesce and undergo LLPS to form condensates, and potentially mature into gel-like or even ordered states, are still elusive. Furthermore, the conformations and dynamics of the molecules inside the condensates are essentially unknown either in experiments or in simulations.

That interactions between RNA and IDP, such as FUS containing RNA binding C-terminal domain and low complexity N-terminal IDP domain, promote phase separation is well documented.[19–21] It is also known that RNA plays an important role in driving the formation, stability and morphology of the biomolecular condensates.[6,12,13,21–29] Recent experiments have also shown that RNA alone is sufficient to self-organize into condensates and gel-like states *in vitro*.[26,30,31] For example, poly-rU with a small addition of short cationic polypeptides can drive reversible LLPS.[26,30,32] Even in germ

granules, containing proteins, there is evidence for well organized RNA clusters stabilized by homotypic interactions.[33,34] Despite the importance of RNA–RNA interactions in driving LLPS, not much is known quantitatively about the organization and dynamics of RNA condensates.

Jain and Vale (JV)[31] showed that the trinucleotide repeat sequences $(CAG)_n$, $(CUG)_n$ and repeats of the hexanucleotide (G_4C_2), linked to several neurological and neuromuscular disorders such as Huntington disease, muscular dystrophy and amyotrophic lateral sclerosis (ALS),[35–38] form condensates. The main findings of the JV study are: (i) The low complexity sequences, $(CAG)_n$ and $(CUG)_n$ as well as polymers made of repeats of G_4C_2 form droplets, which over time, transform to a gel-like state at high RNA concentrations. (ii) Phase separation occurs only if the number of repeats exceeds a critical value.[31,39] (iii) *In vivo,* the gel-like state is abolished (except in the repeats of G_4C_2), and only a liquid-like state is found. Presumably, this is due to active forces generated by an energy source (ATP binding or hydrolysis, for example).

Here, we develop a minimal coarse-grained model to investigate the molecular mechanism of LLPS in RNA repeat sequences. We incorporated only features that are deemed essential for intra- and intermolecular interactions of RNA. In creating the model, we adopted a top-down approach with the goal that it be sufficiently simple so that multi-chain simulations are feasible. The resulting model has only one parameter that sets the energy scale for base pair interactions, which was chosen to reproduce the known structures of a short $(CAG)_2$ duplex.[40] We then performed multi-chain simulations to decipher the LLPS mechanism in the sequence studied by JV *without adjusting any parameter* to match experiments. Our simulations recapitulate the length and concentration dependence of the phase separation, in agreement with the *in vitro* experiments. We show that intermolecular base pair interactions drive phase separation. Unexpectedly, we find that once RNA molecules are recruited in the droplets, they undergo large conformational change, from a hairpin-like conformation in isolation to a stretched state in order to form an extensive network of intermolecular interactions. This soft network, in turn, constrains the RNA conformational fluctuation and mobility. The RNA chains in the high density viscous droplets are conformationally and dynamically heterogeneous. In the largest densely packed high viscosity condensates the RNA chains move predominantly by "slithering" along their contour lengths. Such motions are reminiscent of reptation in polymer melts envisioned 50 years ago by de Gennes.[41] Our work provides important microscopic details of the mechanism of LLPS in the RNA repeat sequences, which currently are not easily accessible in experiments.

Results

LLPS depends on RNA concentration and size

Fig. 1**a** illustrates the transition from monomers to condensates as a function of time, τ, for the $(CAG)_{47}$ system. At $\tau = 0$, the CAG chains exist as monomers. At intermediate times, oligomers form, which subsequently fuse together resulting in large droplets as time progresses (see also Movie S1 for a vivid illustration of the phase separation process). The fraction of RNA chains in the oligomers increases, reaching a peak value at intermediate times, and diminishes at long times. The increase in the droplet size coincides with a decrease in the number of RNA chains in the oligomers, suggesting that the small clusters must fuse to form large droplets.

We then examined the effect of chain length n on the phase separation in $(CAG)_n$. Fig. 1**b** shows that reduction of n from 47 to 30 has little effect on the phase transition behavior. However, if n is smaller than a critical number n^*, the propensity to form liquid droplets ceases entirely. In the $(CAG)_n$ system, n^* is somewhere between 20-30. The value of n^* (or equivalently the critical droplet size) may be estimated using classical nucleation theory that balances the free energy cost of creating a spherical droplet and the

gain in free energy due to favorable interactions between the RNA chains.[42]

Next, we investigated the effect of RNA concentration, $C_{RNA} = \frac{N}{V}$, (N is the number of RNA molecules and V is the volume of the simulation box) on the condensate formation. We reduced (increased) C_{RNA} by increasing (decreasing) V, while keeping the number of chains fixed, $N = 64$. Fig. 1**c** shows that as C_{RNA} for $(CAG)_{47}$ decreases, there is a decrease in the droplet size. At 200 μM, almost all the molecules phase separate and enter the high density phase, leaving very few RNA chains as monomers in the solution. The two largest droplets (blue and green curves) each with ~ 20-30 monomers frequently interact with each other. When C_{RNA} decreases to 50 and 100 μM, only small droplets (< 20 monomers) form. The fluctuations in the size of these droplets are also much less pronounced, indicating that they are less likely to interact with each other and exchange monomers/oligomers.

We calculated the RNA concentrations of the two coexisting low and high density phases using a procedure outlined in the SI. The RNA concentration inside the droplets is enhanced by $\sim$ 50-200 fold compared to the initial concentrations (shown in Fig. 1**d**), in agreement with the experimental value (163-fold).[31] Regardless of the initial C_{RNA}, the concentrations of the two phases are relatively unchanged. Once the initial concentration of $(CAG)_{47}$ reaches 20 μM, which is lower than the concentration of the aqueous (or dispersed) phase, we do not observe stable droplets. At low concentrations, the RNA molecules mostly exist either as monomers or oligomers, and the system as a whole is a single liquid phase. Thus, for a fixed n that is larger than n^*, there is a threshold value of C_{RNA}, which has to be exceeded in order to form stable condensates. From the simulations for different n and C_{RNA}, we determined the putative phase diagram for $(CAG)_n$ (Fig. 1**e**). It shows that for $n \leq 20$, RNA molecules exist in a single phase. For $n > 30$, we predict a coexistence of the high and low density phases, resulting from LLPS of the CAG repeat system.

RNA structure changes dramatically in the condensates

We then characterized the RNA conformations inside the condensates containing multiple chains. The conformational changes of $(CAG)_n$ relative to the monomers is dramatic. (i) The $R(s)$ curve as a function of the sequence distance $s = |i - j|$, shown in Fig. 2**a**, increases at large s, implying that the two ends are not in proximity, as in the isolated chain (see Fig. S2). (ii) Interestingly, $R(s)$ versus s is independent of the droplet sizes. (Oligomers: 2-4 monomers, medium droplets: 5-10 monomers, large droplets: > 10 RNA molecules.) (iii) The bond orientational correlation, $\cos\theta(s) = \left\langle \vec{b}_i . \vec{b}_{i+s} \right\rangle / l_b^2$ (l_b is the bond length), of individual chains lacks periodicity inside the droplets, which is prominent in the monomers and oligomers (Fig. 2**b**). (iv) The end-to-end distance (R_{ee}) distribution for chains inside the droplets shift to higher values (Fig. 2**c**), signaling a disruption of the hairpin structures adopted by the isolated chains. Snapshots from the simulations (Fig. 2**e**) show that the RNA polymers populate extended conformations.

The form factors (Eq. 5 in the SI) in Fig. 2**d** show that the chains inside the droplets are markedly different from the monomeric RNAs. For the chains in the condensates, at small q $\left(q \ll \frac{2\pi}{R_g}\right)$, $S_c(q) \approx N \exp\left(-\frac{1}{3}q^2 \left\langle R_g^2 \right\rangle\right) \approx N\left(1 - \frac{1}{3}q^2 \left\langle R_g^2 \right\rangle\right)$ in the Guinier regime. At higher q, there is a crossover to a power law, $S(q) \sim q^{-1/\nu}$, with $\nu \approx 0.5$, suggesting that the RNAs may be characterized as ideal chains similar in polymer melts. This is because from the perspective of a single chain inside condensates, it is irrelevant whether the interactions arise intramolecularly or from other chains in the droplet. A similar reasoning led to the "Flory ideality hypothesis" proposed in the context of polymer melts (see Fig. 3 for the analyses of intra- and intermolecular interactions). At small spatial length scale, or high q, $S(q) \sim q^{-1}$ is recovered.

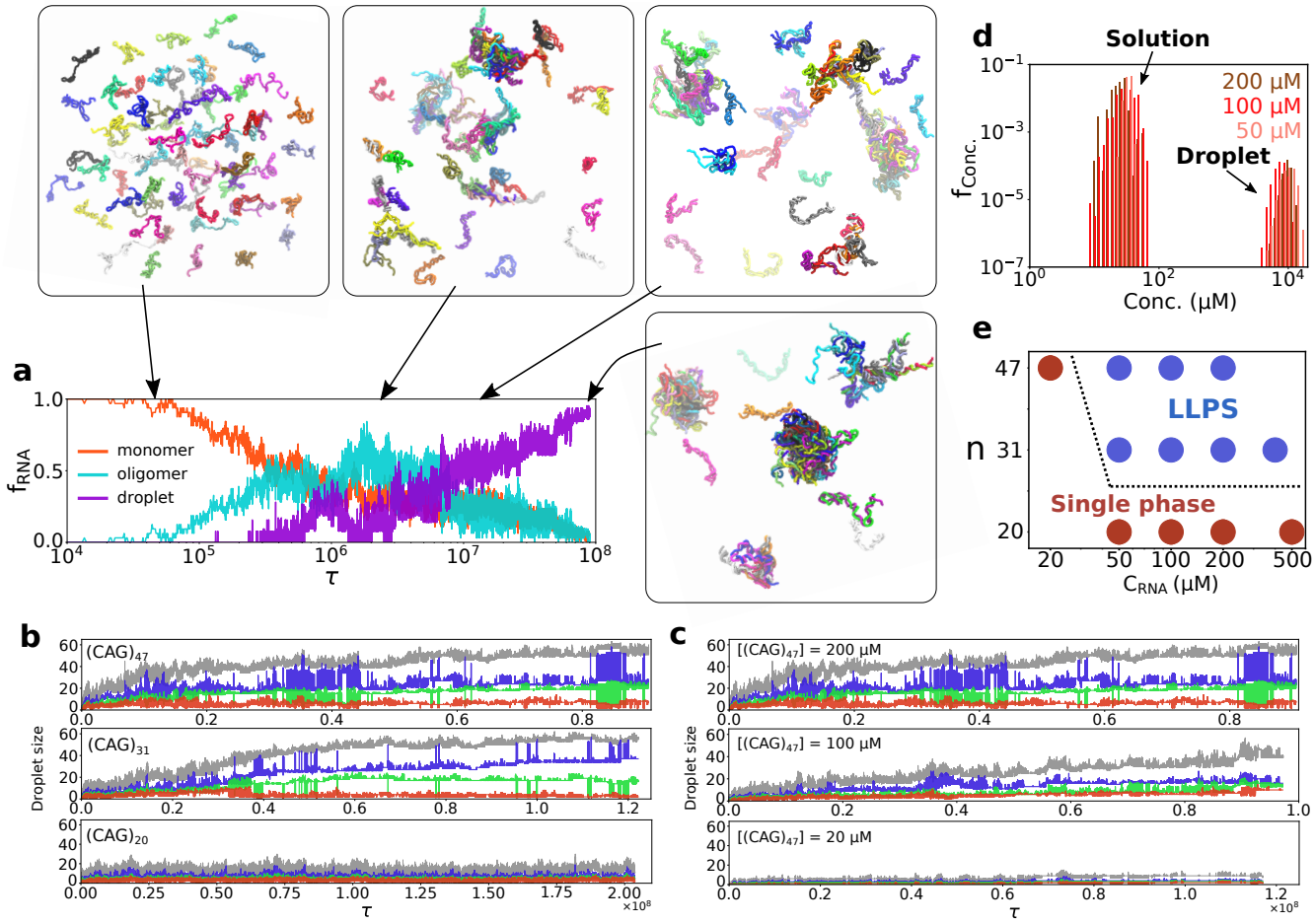

Figure 1. Liquid–liquid phase separation of CAG repeat RNAs. **a**, Fraction of $(CAG)_{47}$ in monomers, oligomers (dimers, trimers and tetramers) and droplets as a function of time. At $\tau = 0$, the RNA chains exist as monomers, whose structural characteristics are described in the Supplementary Information (SI). Subsequently, oligomers form, which fuse together to form large droplets. Snapshots of $(CAG)_{47}$ pictorially illustrate phase separation. **b**, Evolution of the three largest droplet sizes (blue, green and orange, respectively) for different n. The grey curve is the sum of the other three. $(CAG)_{47}$ and $(CAG)_{31}$ form several large droplets while $(CAG)_{20}$ shows little sign of phase separation. **c**, Concentration dependence of $(CAG)_{47}$ phase separation. From top to bottom, the RNA concentration decreases from 200 μM to 20 μM. At 200 μM, very few monomers exist in the solution, as they all interact with each other to form stable droplets. At 20 μM, stable condensate is not observed. **d**, Concentrations of the two phases showing the coexistence of low and high density phases of $(CAG)_{47}$. The concentrations of the two (coexisting) phases are almost invariant, regardless of the initial concentrations, and are separated by almost 3 orders of magnitude. **e**, Putative phase diagram of $(CAG)_n$. For $n = 20$, there is only one phase even at elevated RNA concentrations.

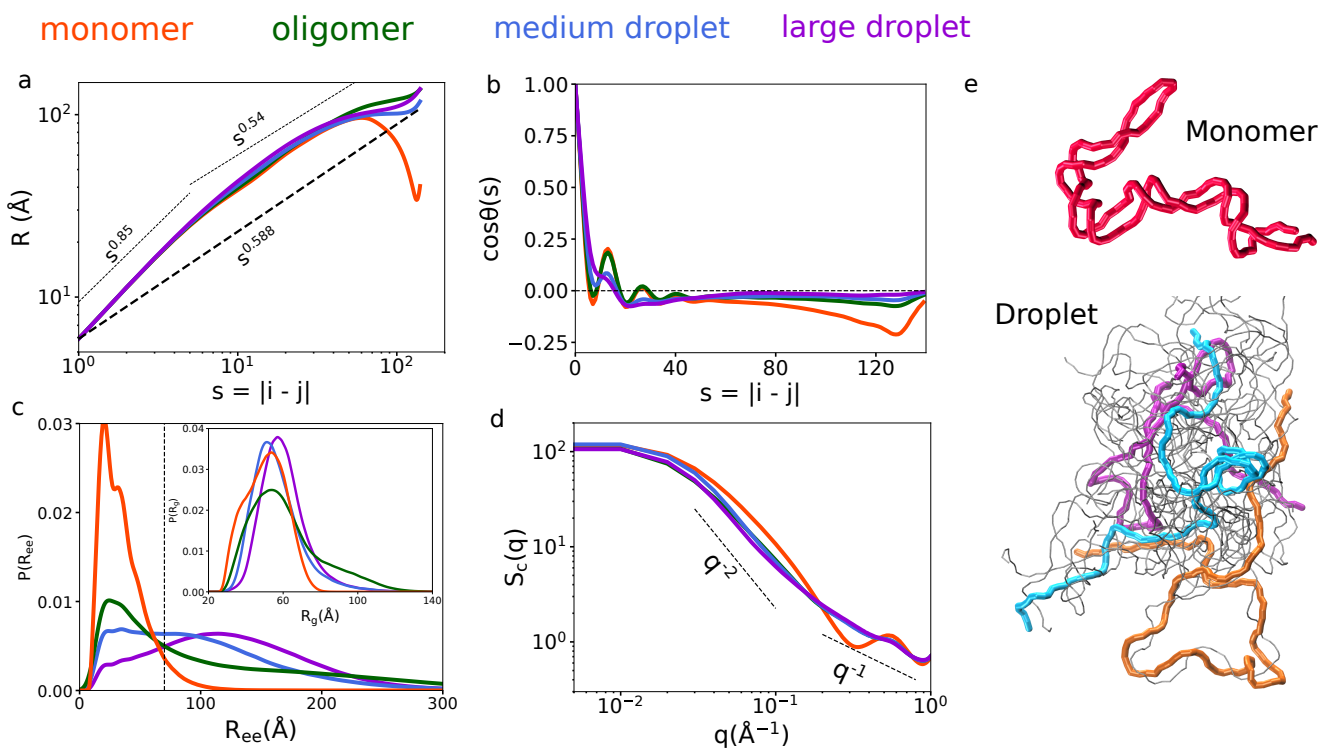

Figure 2. Comparison of the structures of $(CAG)_{47}$ inside the condensates and in isolation. The colors indicate if the molecule is a monomer, or in different sized droplets. **a**, $R(s)$ vs. $s = |i - j|$. The dashed line shows $R(s) \propto s^{0.588}$ for a self-avoiding polymer. **b**, Bond orientational correlation function, $\cos\theta(s)$, as a function of s. Note the absence of periodicity in the large droplet (purple), which is prominent in the oligomers. **c**, Distributions of the end-to-end distance, R_{ee}. The vertical dashed line is the average value for an ideal chain. The R_{ee} values inside the condensates are significantly larger compared to monomers, and the distribution could be fit using a broad Gaussian (see the inset in Fig. 4**a** for the fit). Distributions of the radius of gyration (R_g) are in the inset. **d**, Form factors of RNA chains show that the RNA molecules inside the droplets are similar to an ideal polymer. **e**, Snapshots highlight the conformational differences between isolated chains (top) and chains inside the condensates (bottom).

Intermolecular base-pairing drives RNA condensate formation

We calculated the fraction of formed base pairs (bps), $f_{bp} = \frac{N_{bp}}{N_{CAG}}$ where N_{bp} is the number of bps and N_{CAG} is the number of CAG units in the simulations, to decipher the molecular details of the transition from low order oligomers to condensates. The results were decomposed into intra- and intermolecular bps (shown in Fig. 3**a**). Initially, the chains mostly adopt hairpin-like structures by forming exclusively intramolecular bps. Once small and medium sized droplets form, intermolecular interactions increase at the expense of intramolecular interactions. In order for RNA chains to form interchain bps, the C and G nucleotides from two chains have to satisfy both the distance and orientation criteria, thus requiring population of extended conformations (Fig. 2). In this process, the self-interactions are replaced by the intermolecular interactions without having to compensate for bending on short length scale, s.

We find that intermolecular base pair interaction is the driving force in the coalescence of the RNA repeats.[18,31,43,44] For short chains, ($n = 20$) or at low C_{RNA}, the number of intermolecular interactions is small (Fig. S7). Thus, the majority of the molecules remain as monomers or form relatively small droplets with less than 8 chains. At high concentrations or for longer chains ($n = 47$ and $n = 31$), increase in the number of intermolecular bps leads to phase separation, resulting in the formation of two discrete phases with vastly different RNA densities. The number of bps in the high density phase is much higher than the low density phase (shown in Fig. 3**b**).

Intra- and intermolecular interactions depend on the droplet size (Fig. 3**c**). Unlike the self-interactions in the isolated RNA chain, which feature extensive WC base pairing along the diagonal (Fig. S2), those inside the droplets are more diverse. The propensity to form interactions along the chains increases, beyond those that form along the diagonal. The total number of self-interactions of RNAs inside the droplets is smaller than in the monomers (Fig. 3**a**) due to the formation of extended conformations. Because the RNA conformations are significantly stretched inside the condensates, the interactions between different chains ought to increase. Intermolecular interactions between RNA molecules in droplets are also position-independent, reflecting the repeat nature of the sequence. However, for oligomers, specific 5' to 3' interactions still dominate, similar to what is found in a hybridized duplex.

RNA conformations in the droplets are highly heterogeneous

Due to extensive intermolecular base pair formation inside the droplets, we expect the RNA mobility to be significantly hindered. Each RNA could be kinetically trapped, and only sample conformations around a local minimum in the free energy landscape, much like in a glass or a jammed system.[45] Fig. 4 reveals that this is indeed the case. The distributions of R_{ee} (end-to-end distance) and R_g (radius of gyration) of the individual chains have large dispersion exceeding their usual values, although the histogram averaged over all the chains show behavior resembling that expected for ideal chains (red curves). Therefore, averaging over the ensemble of structures and over the number of RNA monomers in the droplet conceals the conformational heterogeneity. Some of the chains sample conformations with R_{ee} exceeding 200 Å, which is much greater than the dimensions expected for a random coil. As a consequence, there is a great degree of heterogeneity in the conformations that are sampled in the droplets. The unusual conformations that are accessed could arise because of the droplet is arrested in a non-ergodic phase with long overall relaxation times.

Condensates are dynamic

A hallmark of phase separation is that the droplets (or foci) interact with one another, which could result in two (or more) droplets fusing to form a larger droplet by the Ostwald ripening mechanism. The reverse process in which a large droplet could disintegrate into smaller ones could also occur. Our simulations capture both these processes in the phase separation of $(CAG)_n$. Fig. 5**b** shows the time evolution of

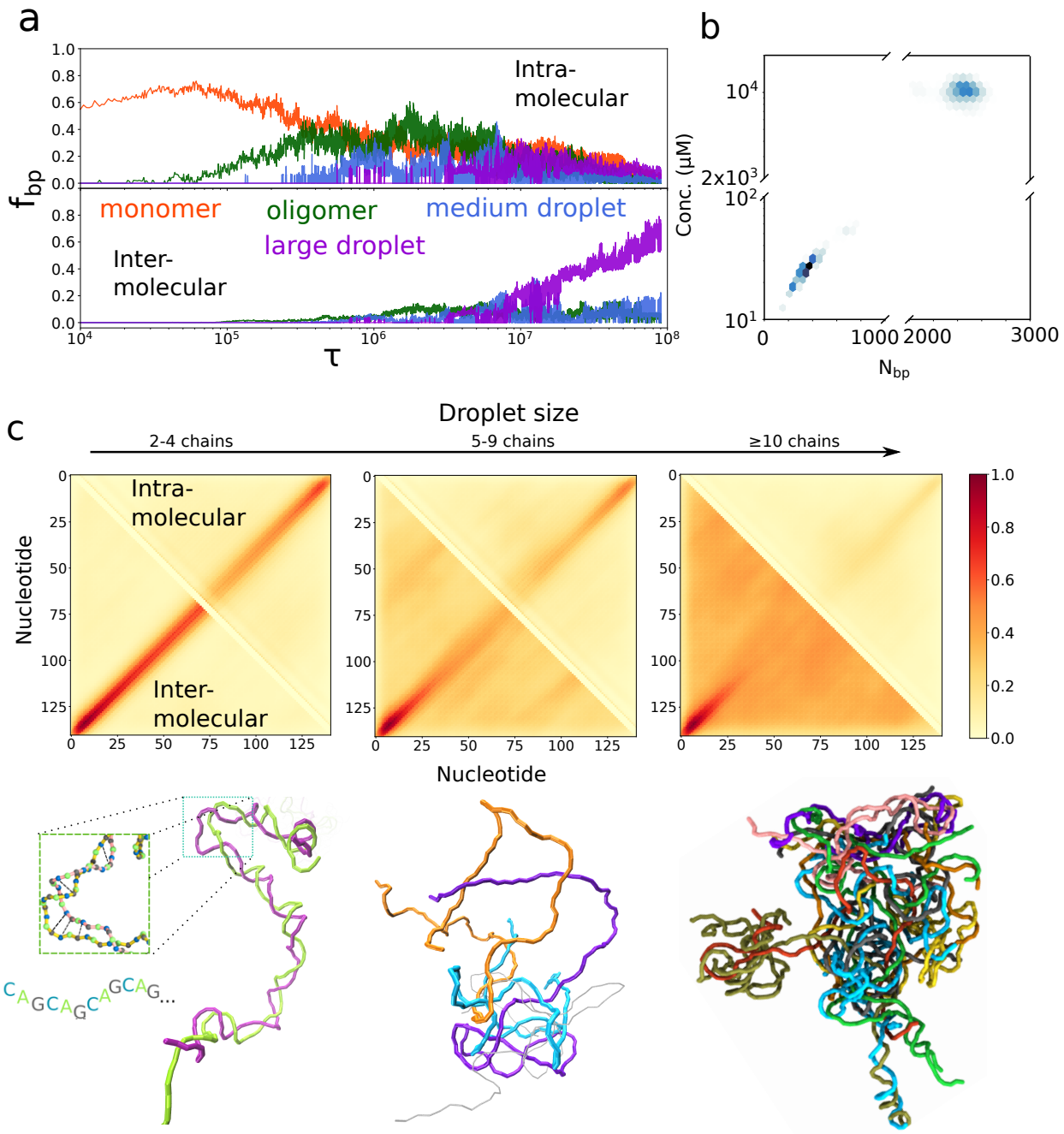

Figure 3. Intermolecular hydrogen bonding drives LLPS in RNA. **a**, Fraction of base pairs, f_{bp}, between G and C nucleotides in $(CAG)_{47}$ at 200 μM (results for others are in Fig. S7). Decomposition into intra- and intermolecular contributions (top and bottom panels, respectively). **b**, Concentration vs. the number of base pair formed in the two coexisting phases. **c**, Distance-based probability of contacts for intra- (top) and intermolecular (bottom) interactions (color scale shown on the right). Left to right correspond to oligomers, medium and large droplets, respectively. In oligomers, intermolecular interactions between the RNAs generate structures similar to RNA duplexes, with signals along the diagonal in the contact map. In larger droplets, the specific interactions decrease, and non-specific interactions predominate. Representative snapshots for different droplets with different sizes are at the bottom.

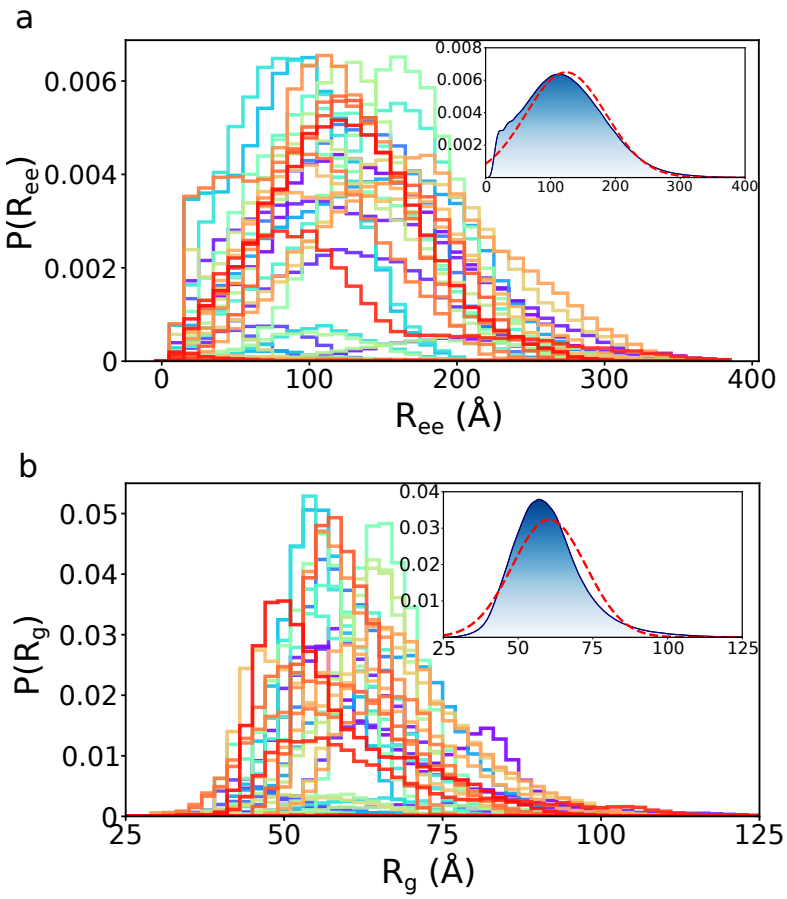

Figure 4. Structural heterogeneity of the RNA chains in large and dense $(CAG)_{47}$ droplets. Distributions of R_{ee} **(a)** and R_g **(b)**, respectively, for each individual RNA chains. The insets shows the average distributions, and the red dashed curves are the Gaussian fit to the data, indicating the behavior of chains inside large droplets is similar to ideal chains. Averaging over the monomers completely masks the remarkable heterogeneity of the monomer structures in the droplets.

each droplet size in the system, illustrating that the condensates formed in the simulations are dynamic, undergoing continuous growth, fusion and fission. The small droplets frequently interact and fuse with one another to form a larger droplet (snapshots shown in Fig. 5**a**). However, not all fusion events are successful. There are instances of failed events in which the two droplets come together, interact for a while but subsequently dissociate (Fig. 5**a**, see also Movie S2). Such fusion-fission processes occur often, suggesting that the droplets are dynamic.

The dynamic nature of the droplets is also highlighted by the lack of persistent internal order, measured by the nematic order parameter (Fig. 5**c**). For small droplets, the internal RNA molecules need to position themselves in a preferred orientation, which would result in a loss of entropy. An extreme example is a dimer, in which the two strands are parallel to maximize the interaction energy (illustrated in Fig. 3**c**). However, in large droplets the RNA could form bp interactions with other chains without sacrificing orientational entropy, which results in a decrease in the nematic order parameter.

Fig. 5**a** shows exquisitely the dynamic nature of the condensate formation. Each row corresponds to a single chain, with the color denoting the droplet to which it is recruited. Each color represents a unique droplet in Fig. 5**b**. At the early stage ($\tau < 2x10^7$), most RNA molecules, at one instance or another, interact with almost all the droplets. Their residence times within a single droplet is relatively short. The monomers exchange with the bulk and are subsequently recruited by another (or the same) droplet. In contrast, at later stages, the chains are mostly restricted inside large droplets (orange and blue droplets in Fig. 5**a**). Despite the restricted movement, the interactions between the RNA chains inside large droplets are nevertheless transient, as illustrated in Fig. 5**d**, which shows the contact lifetime of one molecule with all other chains inside a single droplet. Although some of the contacts are stable most of them are intermittent and are disrupted over time. The lifetime of contacts decay exponentially (Fig. 5**e**). This is because the chains shift their internal positions frequently, and therefore interact with different chains at various times (illustrated in Fig. 5**f** and Movie S2). The contacts with long lifetimes are between chains in the core of the droplet, while those with short lifetimes are formed by the newly recruited chains, usually located at the droplet periphery. Due to the absence of droplet fusion and lack of monomer exchange as time progresses, we surmise that the large droplets in our simulations could be near the onset of coarsening into a rigid gel-like state, which is in accord with *in vitro* experiments.[31]

Monomer dynamics inside the condensates is sluggish: Evidence of reptation

We quantified the dynamics of RNA movement using the mean squared displacement (MSD) of the RNA center-of-mass (Fig. 6**a**). The average MSDs for all the RNA chains increase linearly with time regardless of C_{RNA}, implying that the overall motion is diffusive. However, the dynamics is highly heterogeneous upon undergoing LLPS. The RNA chains in the aqueous phase undergo normal diffusion, but the movement of the polymers in the droplets are significantly affected, especially in large droplets. The MSD exponents are less than unity, and decrease as the droplet size increases (Fig. 6**b**), which is suggestive of sub-diffusive behavior. Indeed, the slowing down of the RNA dynamics inside large droplets could signal the onset of formation of a gel-like state. Given sufficient time, the large droplets could potentially undergo subsequent maturation through a sol-gel transition.[20,45–54]

To probe the jamming dynamics of RNAs inside large droplets, we calculated the MSDs for each nucleotide in the RNA chain (Fig. 6**c**). The average values (the solid line in Fig. 6**c**) show that the MSD scales as $\Delta(\tau) \propto \tau^{1/4}$ at long times. The $\tau^{1/4}$ behavior, which was first observed in polymer melts, is suggestive of reptation-like movement.[41,55,56] The motion of these nucleotides is physically constrained along the contour length of the RNA, reminiscent of the reptation mechanism in which the chain movement is thought to occur in a low (one) dimensional "tube" generated by topological constraints imposed by other chains (see Movie S3 for illustration). In order to exhibit reptation-like dynamics, the RNAs should

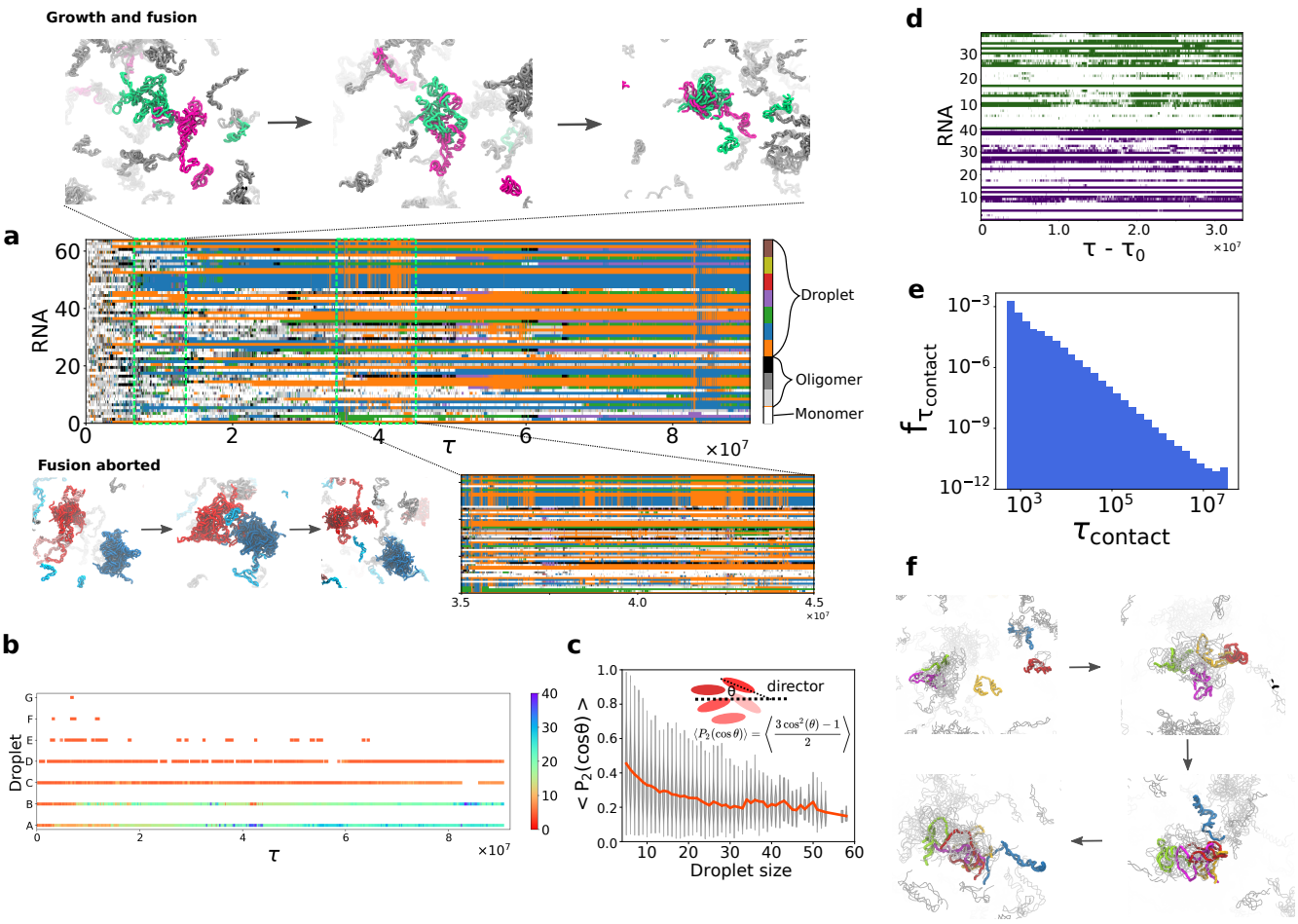

Figure 5. Condensate dynamics. **a**, Each line represents time evolution of a single chain. Colors encode the states of RNA. White is a monomer, oligomers are from grey to black. Seven droplets with different colors contain several RNA chains. Droplet growth and fusion shown on top. Only a few chains are colored, the rest are shown as grey molecules. Aborted fusion event shown on the bottom when the two droplets are in proximity and subsequently dissociate. See also Movie S2. **b**, Evolution of droplets with the color indicating the droplet size (scale on the right), with each row representing a unique droplet, corresponding to one color in **a**. **c**, RNAs in small droplets have high nematic order, $< P_2 \cos(\theta) >$, which decreases as the droplet size increases (the orange curve shows average values). The grey lines show that there are substantial order parameter fluctuations. **d**, Contacts between one particular RNA with other chains inside the same droplet as a function of time. Color code: white - no contact; color - contact. Shown are two plots for two different chains in the same droplet. **e**, Frequency of contact lifetime between chains inside large droplets. The majority of contacts is transient, and the distribution of $\tau_{contact}$ follows an exponential decay. **f**, Illustration of the monomer recruitment and internal rearrangement of RNA. Only a few chains are colored. Chains outside droplets are predominantly in hairpin conformations. Once recruited, they populate more extended states. See also Movie S2.

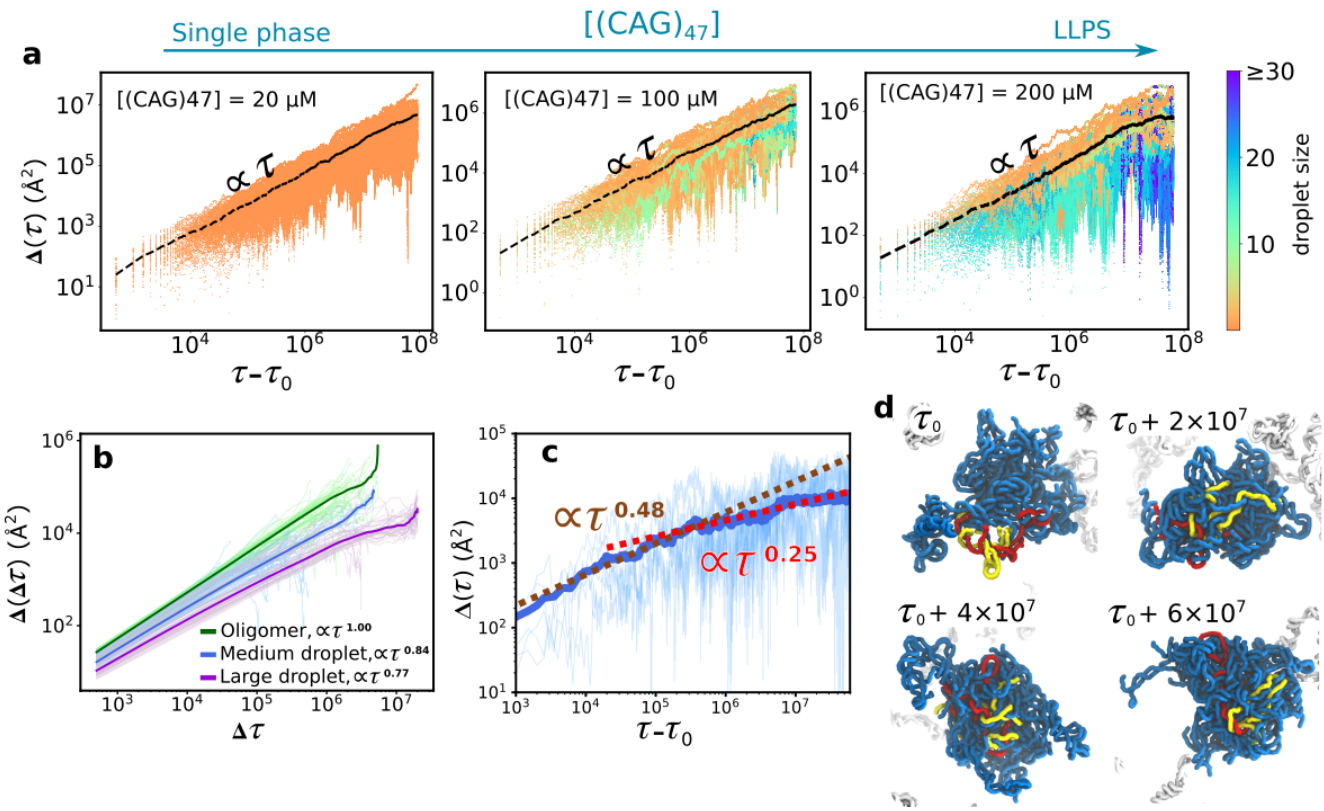

Figure 6. Dynamics of RNA chains in $(CAG)_{47}$ condensates is heterogeneous. **a**, Mean squared displacement (MSD) of the RNA center-of-mass at different concentrations, increasing from left to right, shown in a log scale. Results are shown for individual RNA molecules with color denoting the droplet size. Averaged values are plotted as black lines. The majority of fast moving RNA chains is in oligomeric and monomeric states, while sluggish movement arises from RNA chains inside large droplets. **b**, Time-averaged MSD calculated for RNA chains inside oligomers, medium, and large droplets. The RNA movement in oligomers is diffusive, while for medium and large droplets, they undergo subdiffusive motion. Results for individual RNA molecules are plotted in pale color lines, and the average is shown as solid lines. **c**, Nucleotide MSD in a large viscous droplet. Solid line is the averaged value. In the intermediate regime ($\tau - \tau_o < 10^5$, where τ_o is the initial time when the droplet forms), the MSD exponent is ≈ 0.5, indicating that the chains are jammed. There is a change in slope at $\tau - \tau_o \approx 10^5$, where $\Delta \propto \tau^{1/4}$. This suggests that the movement of connected nucleotides is reminiscent of reptation, which implies that chain dynamics is mainly restricted along the contour length of the RNA due to entanglement. The reptation dynamics is most clearly illustrated in Movie S3. **d**, Representative snapshots show movement of individual chains inside large droplets.

form an intricate and entangled network of interactions, which arises due to intermolecular base pair interactions. Interestingly, the formation of such a network has recently been demonstrated in AU-rich mRNAs.[26,57] The network of intermolecular interactions between the RNAs, therefore, plays a crucial role in dictating the slow heterogeneous dynamics of RNAs inside the droplets.

Discussion and Conclusion

We developed a minimal coarse-grained model that captures many of the observed features in the LLPS of repeat RNA sequences. In accord with experiments, we find that long $(CAG)_n$ repeats ($n = 47, 31$) phase separate at relatively low concentrations, whereas short chains ($n = 20$) remain in a single liquid phase even at elevated RNA concentrations. It is worth noting that there is only a single parameter in the model that was determined using the structures of $(CAG)_2$ duplex. In probing condensate formation in CAG repeats, the model is essentially parameter-free, and hence is transferable to probe other RNA sequences as well.

Recruitment of monomers to condensates involves unwinding of hairpin-like structure: Due to the high GC content of the sequence, the monomeric RNAs adopt an ensemble of hairpin-like structures with a small end-to-end distance (Fig. S2), as is the case in a number of RNA molecules.[40,58,59] In sharp contrast, RNA chains inside the condensates are extended with large 5'-3' distances. In order to form intermolecular interactions with other chains, the RNA monomer has to unwind, exposing the base to form WC base pairs (or $\pi - \pi$ stacking and non-canonical base pair depending on the sequence[23,26,32]). We expect this finding to be a general feature of not only repeat RNAs, but also in any biomolecule that undergoes one-component phase separation, as was shown previously in oligomer formation in a fragment of Aβ peptides.[60]

Topological constraints, conformational and dynamical heterogeneity: As a result of a large number of intermolecular interactions between RNA chains within the condensates, a given RNA molecule is topologically constrained by the neighboring chains. The local environment of individual RNA chains differs greatly, which results in wide variability in the individual distributions of various shape parameters, even though the average distribution could be explained using polymer theory. Because of such topological constraints, the dynamics of RNA chains is also strongly affected. The mobility of the RNA chains is greatly diminished even though diffusion of the system as a whole is liquid-like.[61,62] In large droplets in which long RNA polymers are tightly packed with elevated density and viscosity, the movement of RNA molecules is reminiscent of reptation in polymer melts. RNA chains slither along their contour length due to translational inhibition in other dimensions. Strong entanglement in large droplets could further facilitate the maturation process, eventually driving the system to a gel-like state at long times under *in vitro* conditions. Under cellular conditions, it is likely that active processes (ATP binding and/or hydrolysis) regulate the formation of gel-like states by remodeling the RNA structures.[63–66] For instance, the recruitment of additional protein clients[28] or the presence of ATP-consuming enzymes (for example, helicases or RNA chaperones) in the nucleoplasm that reorganize the RNA base pairings[31,66] could play a role in maintaining the condensates in a liquid-like state.

Final remarks: We have used molecular simulations to probe condensate formation in repeat RNA sequences, and have uncovered general concepts that are difficult to obtain using experiments alone. Our study establishes that in the process of self-assembly of low complexity RNA sequences into droplets, RNA chain makes a transition from hairpin-like structures to extended conformations in order to maximize the number of intermolecular base pair interactions. The droplets are stabilized by a network of soft intermolecular interactions, which repeatedly break and reform as the chains move within the confines of the droplets. The dense network, which is most prominent in large high density droplets, restricts the

mobility of the chain mostly along the contour. As a result, motion occurs by a process that is similar to snake-like movement envisioned in the context of polymer melts. Taken together, our results show that the structural heterogeneity and the pinning of RNA chains by their neighbors makes the dynamics sufficiently sluggish that the droplet may be in a non-ergodic phase.

The entanglement of RNA chains arises in part because the CXG sequence could engage in both intra- and intermolecular WC base pair formation. Because such specific interactions are not possible in poly-rU sequences, which also form droplets (stabilized possibly by $\pi - \pi$ interactions), we believe all that the only requirement for droplet formation in low complexity RNA sequences is the presence of many weak inter-chain interactions. This would suggest that the length dependence for condensate formation is likely to be different in poly-rU sequences than in CAG or CUG polymers. Finally, it is worth emphasizing that our computational framework is well suited to examine the mechanism of self-assembly in other RNA sequences as well.

References

1. Brangwynne, C. P. *et al.* Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **324**, 1729–1732 (2009).
2. Hyman, A. A., Weber, C. A. & Julicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
3. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
4. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
5. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
6. Langdon, E. M. & Gladfelter, A. S. A New Lens for RNA Localization: Liquid-Liquid Phase Separation. *Annu. Rev. Microbiol.* **72**, 255–271 (2018).
7. Berry, J., Brangwynne, C. P. & Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. *Rep. Prog. Phys.* **81**, 046601 (2018).
8. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).
9. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
10. Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
11. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
12. Rhine, K., Vidaurre, V. & Myong, S. RNA Droplets. *Annu. Rev. Biophys.* **49**, 247–265 (2020).
13. Roden, C. & Gladfelter, A. S. RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* 1–13 (2020).
14. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.* (2020).

15. Stockmayer, W. H. Theory of Molecular Size Distribution and Gel Formation in Branched-Chain Polymers. *J. Chem. Phys.* **11**, 45–55 (1943).

16. Flory, P. J. *Statistical Mechanics of Chain Molecules* (NY, Interscience Publishers, 1969).

17. Semenov, A. N. & Rubinstein, M. Thermoreversible Gelation in Solutions of Associative Polymers. 1. Statics. *Macromolecules* **31**, 1373–1385 (1998).

18. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).

19. Han, T. W. *et al.* Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies. *Cell* **149**, 768–779 (2012).

20. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).

21. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).

22. Schwartz, J. C., Wang, X., Podell, E. R. & Cech, T. R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Rep.* **5**, 918–925 (2013).

23. Treeck, B. V. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci.* 201800038 (2018).

24. Treeck, B. V. & Parker, R. Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell* **174**, 791–802 (2018).

25. Langdon, E. M. *et al.* mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).

26. Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci.* **116**, 7889–7898 (2019).

27. Tauber, D., Tauber, G. & Parker, R. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem. Sci.* (2020).

28. Guillen-Boixet, J. *et al.* RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346–361.e17 (2020).

29. Sanders, D. W. *et al.* Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306–324.e28 (2020).

30. Aumiller, W. M., Pir Cakmak, F., Davis, B. W. & Keating, C. D. RNA-Based Coacervates as a Model for Membraneless Organelles: Formation, Properties, and Interfacial Liposome Assembly. *Langmuir* **32**, 10042–10053 (2016).

31. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).

32. Aumiller, W. M. & Keating, C. D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat. Chem.* **8**, 129–137 (2016).

33. Trcek, T. *et al.* Drosophila germ granules are structured and contain homotypic mRNA clusters. *Nat. Comm.* **6**, 7962 (2015).

34. Trcek, T. *et al.* Sequence-Independent Self-Assembly of Germ Granule mRNAs into Homotypic Clusters. *Mol. Cell* **78**, 941–950.e12 (2020).

35. Gatchel, J. R. & Zoghbi, H. Y. Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nat. Rev. Genet.* **6**, 743–755 (2005).

36. La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11**, 247–258 (2010).

37. McMurray, C. T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11**, 786–799 (2010). Number: 11 Publisher: Nature Publishing Group.

38. Krzyzosiak, W. J. *et al.* Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucl. Acids Res.* **40**, 11–26 (2012).

39. Lee, D.-Y. & McMurray, C. T. Trinucleotide expansion in disease: why is there a length threshold? *Curr. Opin. Genet. Dev.* **26**, 131–140 (2014).

40. Kiliszek, A., Kierzek, R., Krzyzosiak, W. J. & Rypniewski, W. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucl. Acids Res.* **38**, 8370–8376 (2010).

41. de Gennes, P. G. Reptation of a Polymer Chain in the Presence of Fixed Obstacles. *J. Chem. Phys.* **55**, 572–579 (1971).

42. Oxtoby, D. W. Homogeneous nucleation: theory and experiment. *J. Phys.: Condens. Matter* **4**, 7627–7650 (1992).

43. de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K. & Krzyzosiak, W. J. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucl. Acids Res.* **39**, 3852–3863 (2011).

44. Ciesiolka, A., Jazurek, M., Drazkowska, K. & Krzyzosiak, W. J. Structural Characteristics of Simple RNA Repeats Associated with Disease and their Deleterious Protein Interactions. *Front. Cell. Neurosci.* **11** (2017).

45. Jawerth, L. *et al.* Protein condensates as aging Maxwell fluids. *Science* **370**, 1317–1323 (2020).

46. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).

47. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).

48. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015).

49. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615–627.e16 (2017).

50. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688–699.e16 (2018).

51. Franzmann, T. M. *et al.* Phase separation of a yeast prion protein promotes cellular fitness. *Science* **359** (2018).

52. Wegmann, S. *et al.* Tau protein liquid–liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, e98049 (2018).

53. Ray, S. *et al.* a-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nat. Chem.* 1–12 (2020).

54. Pytowski, L., Lee, C. F., Foley, A. C., Vaux, D. J. & Jean, L. Liquid–liquid phase separation of type II diabetes-associated IAPP initiates hydrogelation and aggregation. *Proc. Natl. Acad. Sci.* **117**, 12050–12061 (2020).

55. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990).

56. Hsu, H.-P. & Kremer, K. Static and dynamic properties of large polymer melts in equilibrium. *J. Chem. Phys.* **144**, 154907 (2016).

57. Ma, W., Zhen, G., Xie, W. & Mayr, C. Unstructured mRNAs form multivalent RNA-RNA interactions to generate TIS granule networks. *bioRxiv* 2020.02.14.949503 (2020).

58. Marquis Gacy, A., Goellner, G., Juranić, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533–540 (1995).

59. Lai, W.-J. C. *et al.* mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat. Comm.* **9**, 4328 (2018).

60. Nguyen, P. H., Li, M. S., Stock, G., Straub, J. E. & Thirumalai, D. Monomer adds to preformed structured oligomers of Aβ-peptides by a two-stage dock–lock mechanism. *Proc. Natl. Acad. Sci.* **104**, 111–116 (2007).

61. Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci.* **112**, 7189–7194 (2015).

62. Moon, S. L. *et al.* Multicolour single-molecule tracking of mRNA interactions with RNP granules. *Nat. Cell Biol.* **21**, 162–168 (2019).

63. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).

64. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).

65. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353** (2016).

66. Tauber, D. *et al.* Modulation of RNA Condensation by the DEAD-Box Protein eIF4A. *Cell* **180**, 411–426.e16 (2020).

Acknowledgements

We are grateful to Mauro Mugnai, Sumit Sinha and Ryota Takaki for stimulating discussions and critical reading of the manuscript. We thank Anne Bowen for helping with the movies and figures. This work was supported by National Science Foundation Grant (CHE 19-00093) and the Welch Foundation Grant (F-0019) through the Collie–Welch chair. We are thankful to the Texas Advanced Computing Center for providing computational resources.

Author contributions statement

H.T.N. and D.T. conceived and designed research, H.T.N. conducted research, H.T.N., N.H. and D.T. analysed the results and wrote the manuscript.