Tangled balls of RNA

1. **RNA phase transitions in repeat expansion disorders**
   Authors: A. Jain and R. D. Vale
   Nature, 2017. doi: 10.1038/nature22386.

2. **Condensates in RNA repeat sequences are heterogeneously organized and exhibit reptation dynamics**
   Authors: H. T. Nguyen, N. Hori, and D. Thirumalai
   Nature Chemistry, 2022. doi: 10.1038/s41557-022-00934-z

Recommended with a Commentary by Omar A. Saleh, University of California, Santa Barbara

The field of molecular biology is undergoing a close examination of the potential functional role of 'membraneless organelles', or biomolecular condensates, within cells. These are spontaneously-formed, dense, and amorphous groupings of biomolecules (proteins and/or RNA) held together by various types of attractive binding interactions, and free to exchange components with the surrounding cellular medium due to the lack of an encapsulating membrane. Studies of these assemblies have shown that, at least in the *in vitro* context, many are well-described by the equilibrium thermodynamics of phase separation. For example, this is tested by showing that condensation occurs only beyond a minimum concentration of the constituent biomolecules, or that there is a regime in which the condensates coexist with a dilute biomolecular solution.

A notable addition to this field was the first of the two recommended works, the 2017 paper by Jain and Vale [1], that investigated the mechanistic basis of so-called 'nucleotide repeat expansion disorders.' Such disorders are malfunctions of the cellular machinery that lead to the presence in the genomic DNA of multiple repeats of a short sequence, such as the trinucleotide repeat $(CAG)_n$. The presence of these repeats are associated with various human diseases [2], though only if there are enough repeats (roughly $n \gtrsim 30$)[3]. Jain and Vale argued that biomolecular phase separation might lie at the root of the dysfunction caused by these repeats. Specifically, transcription causes the repeat sequences to be copied into RNA molecules, which can strongly (though not perfectly) self-bind due to the nearly 'palindromic' nature of the sequence– that is, through the Watson-Crick-Franklin rules of nucleic-acid basepairing, a sequence CAG can form a relatively strong bond with a second CAG sequence through basepairing between the two C/G pairs. Notably, the paired sequences can both be located on the same molecule (i.e. if the RNA folds onto itself), or on different molecules. Thus, if enough such RNAs are present, there is the possibility that

multi-valent, inter-strand basepairing will lead to a condensed network of associated RNA strands. Jain and Vale showed that condensates of RNA do indeed form for a few relevant sequences, both in purified form (*in vitro*) and within the cell, provided they are long enough (with the critical repeat number close to the ≈ 30 associated with disease), and with an observed critical concentration, thus pointing towards the potential relevance of phase separation.

The significance of the Jain/Vale hypothesis for human disease is complex and not settled [2, 4]. That said, the RNA repeat system is quite compelling as a physical model for condensate formation due to its simplicity. First, condensation occurs with only a single biomolecular component. Further, that component is a nucleic acid, whose homogeneous electrostatic charge and limited sequence heterogeneity makes its physical characteristics notably more uniform than the proteins that constitute many other condensates. Finally, the dominant monomer binding mechanism is the sequence-specific basepairing mechanism, rather than more non-specific mechanisms such as hydrophobic or electrostatic interactions. The relative simplicity of RNA-repeat condensates should make them amenable to more detailed, quantitative studies, and indeed the second recommended paper is a careful physical investigation from Nguyen, Hori, and Thirumalai [5], who carried out extensive coarse-grained simulations of the condensation behaviors of CAG-repeat RNAs, and were able to replicate the existence of a critical length, and critical concentration, for phase separation.

A primary physical question about RNA-repeat condensates is how they are able to condense in the first place– one might expect these sorts of RNA molecules to form stable, single-RNA folded structures (e.g. hairpins) in which the C/G bases all form intramolecular pairs, and are thus rendered non-reactive and unavailable to form the intermolecular bonds needed for condensation. The situation is somewhat akin to a long piece of adhesive tape: as anyone who has struggled with wrapping presents can attest, tape can easily fold onto itself into a 'passivated' hairpin that is no longer able to stick to anything else. Nguyen *et al.* showed that such hairpins do indeed form, but frequently in a slightly misaligned form which leaves a tail of unpaired bases (imagine folding the tape at a position slightly off center, leaving a sticky tab). This tail serves as a starting point for intermolecular interactions that ease the kinetic transition of an RNA molecule into the condensate. The equilibrium energetics of this transition are also intriguing: The authors found that intramolecular basepairing is nearly entirely replaced by intermolecular basepairing when an RNA enters the condensate from the dilute phase, but the data indicate the fraction of C/G bases formed is quite similar, $\approx 85 - 90\%$, in both states. This implies that basepairing, while obligatory, is not the primary driving force for condensation. Instead, the key would appear to either be entropic, due to the wide array of structures available to the RNAs within the condensate relative to the relatively constrained ensemble of the single-chain folded state, or associated with the bending penalty incurred upon folding.

A second crucial insight from Nguyen *et al.* concerned the polymeric nature of the condensate. *A priori*, one might imagine that the propensity of these RNA molecules to fold onto themselves would make them somewhat globular, and only capable to binding to other RNAs through whatever unpaired segments remain, leading to a more 'colloidal' dense phase of particles, each with an unshared pervaded volume. Instead, though, the simulation results show that the RNAs significantly expand upon condensing, taking on configurations well-modeled as ideal chains (i.e. with Flory exponent $\nu \approx 1/2$) that are highly entangled with

(and bound to) their neighbors. As with polymer melts, the ideality of the ensemble arises because, in the dense condensate, intra- and inter-chain monomer interactions are identical, leading to no excess intra-chain repulsion or attraction that would, respectively, swell or condense each RNA. The relevance of polymeric behavior extends to dynamic properties within the condensate, where individual bases are seen to display reptation-like dynamics (mean-squared displacements growing as $t^{1/4}$), indicative that each RNA is constrained to move along a 1-D tube created by entanglements with the other chains. Of course, unlike reptation in a melt, here the entanglements are frequently basepaired interactions, so diffusion is likely activated rather than purely Brownian.

Ultimately, even a putatively simple biomolecular condensate is still quite an involved system, and Nguyen *et al.* offered hints of other complexities. For example, nucleic acid basepairing leads to double-helices that are rod-like; accordingly, the authors showed that, in moderately-sized condensates, there is a nematic order apparently associated with such rods. This order decreases (though doesn't disappear) in larger droplets. One wonders what other physical features might be affected by such rods, and how the highly-entangled nature of double-helical binding might affect reptation dynamics. Other quantitative questions are briefly touched on– for example, certain data imply that, in a $(CAG)_{47}$ condensate, each RNA contacts roughly 10-15 other chains at a time (with each contact being transient), though neither the dependency of this valency on chain length, nor the number of basepairs per contact are noted. And, as discussed by the authors, electrostatic effects are significant: as homogeneously negatively-charged polymers, RNA necessarily requires charge neutralization by positively charged salt ions to form a dense phase, with Jain and Vale finding a complex dependence of condensate stability on the amount of divalent and monovalent salt. However, the type of simulation performed by Nguyen *et al.* didn't allow full exploration of this aspect. Finally, Jain and Vale reported a liquid-to-gel transition of the experimental condensates over time, an aspect of the system's behavior for which there are only hints in the limited time-window of the simulation. Nonetheless, the existence of various loose ends only highlights the pioneering nature of the article by Nguyen *et al.*, which represents a significant advance in the understanding of a physically compelling, and potentially medically-relevant, condensate system.

References

[1] A. Jain and R. D. Vale, Nature 546 (7657), 243-247 (2017).

[2] I. Malik et al., Nat. Rev. Mol. Cell Biol. 22 (9), 589-607 (2021).

[3] D.-Y. Lee and C. T. McMurray, Curr. Opin. Genet. Dev. 26, 131-140 (2014).

[4] D. W. Sanders and C. P. Brangwynne, Nature 546 (7657), 215-216 (2017).

[5] H. T. Nguyen et al., Nat. Chem. 14 (7), 775-785 (2022).