The effect of polymer length in liquid-liquid phase separation

Gilberto Valdes-Garcia[1,3], **Kasun Gamage**[2,3], **Casey Smith**[2], **Karina Martirosova**[2], **Michael Feig**[1], **Lisa J. Lapidus**[1,2,4,*]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[2]Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA

[3]These authors contributed equally

[4]Lead contact

SUMMARY

Understanding the thermodynamics that drive liquid-liquid phase separation (LLPS) is quite important given the number of diverse biomolecular systems undergoing this phenomenon. Many studies have focused on condensates of long polymers, but very few systems of short-polymer condensates have been observed and studied. Here, we study a short-polymer system of various lengths of poly-adenine RNA and peptides formed by the RGRGG sequence repeats to understand the underlying thermodynamics of LLPS. Using the recently developed COCOMO coarse-grained (CG) model, we predicted condensates for lengths as short as 5–10 residues, which was then confirmed by experiment, making this one of the smallest LLPS systems yet observed. A free-energy model reveals that the length dependence of condensation is driven primarily by entropy of confinement. The simplicity of this system will provide the basis for understanding more biologically realistic systems.

Graphical Abstract

The thermodynamics driving liquid-liquid phase separation in short polymers are poorly understood. Valdes-Garcia et al. use a coarse-grained model to show that, when keeping the volume fraction of polymers constant, shorter, more numerous polymers will cost too much entropy to condense compared with fewer, longer polymers.

INTRODUCTION

Liquid-liquid phase separation (LLPS) is an increasingly predominant phenomenon found in many types of cells and under many conditions.[1,2] Despite the diversity of the formation conditions and their contents, there seem to be commonalities of the resulting condensates that can be described by physics-based models.[3] While condensates have been observed for folded proteins and RNAs,[4] the majority of systems studied are of disordered proteins and nucleic acids. Polymeric models describing intra- and intermolecular interactions balanced against entropy have been developed to describe such observations. The simplest model has been proposed by Flory[5] and Huggins[6] to describe the free energy change of mixing a homopolymer with a solvent. Choi et al.[7] have expanded this theory into the stickers-and-spacers model to include specific interactions in heteropolymers while still accounting for the entropy of polymers. Similarly, Banani et al.[8] have described condensation in terms of valency of client binding to scaffold molecules.

Experimentally, most model systems use proteins and/or RNAs that are dozens to hundreds of residues long.[9–13] Alshareedah et al.[14] worked with sequences similar to those

used here but at lengths of 500 bases and 50 amino acids. Bai et al.[10] demonstrated condensation with a 21-base oligonucleotide and a 30-residue peptide in the presence of inert crowders. Lim et al.[15] observed condensates in 23-residue peptides rich in histidine and tyrosine. Akahoshi et al.[16] observed condensation with poly(A)$_{15}$ single-stranded DNA (ssDNA) and K_7X_3 peptides, where X represents several different residues. Tang et al.[17] conducted an exhaustive computational survey of all dipeptide sequences and experimentally demonstrated liquid droplets of QW. Finally, recent work by Fisher and Elbaum-Garfinkle[18] has shown condensation with uradine diphosphate (UDP) or uradine triphosphate (UTP) and polyR$_{10}$, but not with UDP and polyK$_{10}$, indicating that arginine has stronger interactions with RNA than lysine despite the same charge. These differences have been described computationally as well, using all-atom simulations for polyR$_5$/polyK$_5$-polyuracyl$_5$ mixtures at various concentrations.[19] However, the length dependence at the residue level beyond this precision was not explored. Therefore, a natural question is as follows: how long must such polymers be to observe condensation?

We recently developed a coarse-grained (CG) computational model that reduces each amino acid or base to a single bead (see supplemental experimental procedures and Table S1 for details).[20] The interactions between beads were systematically parameterized to match experimental observations of polymer characteristics and LLPS for many different systems. The model was also developed to accurately account for concentration dependence of LLPS, which allowed quantitative prediction of the length dependence under real experimental conditions. The accuracy of this model compared with others in the literature was achieved with just a few specific parameters: the stiffness of the angular harmonic potential and separate parameters for the strength of cation-π interactions within a protein and between protein and nucleic acids.

Using this model, we predict that short polymers undergo LLPS at moderate (~1 mg/mL) concentrations and systematically investigate LLPS of RNAs and peptides of various lengths computationally and experimentally. The composition and the volume fraction of the polymers are kept constant so that the only variable that changes is the number of covalent bonds in the system. Condensation of molecules as short as 5–10 nt or amino acids is observed, making this one of the smallest LLPS systems yet. A simple energetic model is developed based on the computational results, and we find that the primary driver of the length dependence is the confinement entropy; more short polymers need to be confined compared with long polymers. At the shortest lengths, the enthalpy does not balance the entropy, making the free energy of condensation positive.

RESULTS

CG model of disordered polymers

Using COCOMO, our recently developed CG model,[20] simulations were performed for various lengths of adenine polymers (poly(A)$_N$) and different repetitions, M, of the [RGRGG]$_M$ peptide for a fixed volume fraction of 0.13% of each polymer (Table S2). Snapshots shown at the end of the simulations indicate system-dependent condensation (Figure S1). According to these initial simulations, there are minimum peptide and nucleic acid polymer lengths before condensation is observed. The minimum peptide length required

for phase separation depended on the length of poly(A)$_N$ and vice versa. For example, [RGRGG]$_4$ was the minimum length to form clusters with poly(A)$_{20}$, whereas poly(A)$_{10}$ required at least the length of [RGRGG]$_{10}$ to form clusters (Figure S1). Moreover, we found that, even for RNA as long as 300 bases, no clusters were observed for [RGRGG]$_1$ or [RGRGG]$_2$, indicating that there is a minimal peptide length required for phase separation. No condensates were observed for poly(A)$_5$ for peptides as long as [RGRGG]$_{15}$, also suggesting a minimal RNA length, but the peptide length was not extended beyond 75 amino acids. These simulations are generally in agreement with previous computational works on RNA[21–23] and DNA[24] showing that LLPS is enhanced with longer nucleic acids lengths.

Experimental observation of condensates

To confirm the predictions from the CG model, LLPS was tested for different mixes of poly(A)$_N$ (N = 5, 10, and 20) with [RGRGG]$_M$ (M = 1, 2, 3, 4, 6, 8, and 10). For all experiments, the total concentrations of RNA and peptide were maintained at 1 mg/mL, and the ratio of positive to negative charge was 0.75. Phase separation was observed either by fluorescence imaging, using Cy3-labeled poly(A)$_N$ or Cy5-labeled [RGRGG]$_1$, or by differential interference contrast (DIC) imaging. We observed that the Cy3 in the labeled RNA induces phase separation by itself, likely because of dye hydrophobicity and stacking interactions in oligomers.[25] To find out the threshold at which Cy3 starts inducing phase separation in our systems, the concentration of poly(A)$_{10}$-Cy3 was varied from 0–100 μM in a mixture of poly(A)$_{10}$ and [RGRGG]$_2$. The results showed that no condensates formed up to 20 μM, but condensates were observed at 30 mM and above (Figure S2), indicating that the threshold lies between 20 and 30 μM. Therefore, Cy3-labeled RNA was kept at a very low concentration (5 μM) compared with unlabeled poly(A) to prevent it from inducing phase separation.

Figure 1 (top) shows confocal microscopy images for all combinations of [RGRGG]$_{1,2,3,4,6,8,10}$ and poly(A)$_{5,10,20}$ and long-chain poly(A) (poly(A)$_{>600}$). Surprisingly, the minimum peptide length for poly(A)$_{20}$ to phase separate was [RGRGG]$_2$, and poly(A)$_{10}$ separated with [RGRGG]$_3$, shorter than predicted from the initial simulations. Additionally, poly(A)$_5$ phase separated with [RGRGG]$_4$, whereas the COCOMO model initially did not predict phase separation at any peptide length (Figure S1). For each polymer, increasing the length by no more than 5 residues was sufficient to induce phase separation. It was observed by naked eye that all samples with condensation turned cloudy when the peptide was mixed with RNA, but [RGRGG]$_4$ and [RGRGG]$_6$ with poly(A)$_5$ were less cloudy compared with all of the other mixtures, suggesting that condensation for these mixtures is less stable than for longer lengths. The results also show that no phase separation occurred for the shortest peptide, [RGRGG]$_1$, even with poly(A)$_{>600}$, indicating that [RGRGG]$_1$ does not interact sufficiently favorably with the RNA polymer to form condensates.

To further confirm that condensates below the diffraction limit of the microscope were not present, dynamic light scattering of several samples were measured. Figure S3 shows that the particle sizes for mixtures with no observed condensates are smaller than 10 nm, and the particles sizes for mixtures with condensates are larger than 500 nm. The one exception to these observations is poly(A)$_5$ with [RGRGG]$_3$, where particle sizes of ~150 nm were

observed. These complexes of ~4.3 million molecules may be only marginally stable or may be less liquid-like than longer-length mixtures and unable to grow.

The images in Figure 1 shows that there were no substantial differences in droplet size right after mixing for any of the combinations. To confirm this, confocal images were analyzed to determine the area of each condensate. Figure 2 shows histograms of various combinations of RNA and peptide, which were fit to a Poisson distribution. Within the uncertainty of such distribution, there were no substantial differences at the earliest time between different combinations of RNA and peptide. The average condensate size increases with time, but differences in growth do not appear to depend on the lengths of the components. This suggests that the composition of the condensed phase, especially the residue densities, does not change much with polymer length, as discussed below. Nucleation appeared to occur within the experimental time of mixing the components and creating the first confocal image, but it is possible that the rate of nucleation does depend on peptide or RNA length. While there was qualitatively good agreement between initial computer predictions and experiments, we considered how initial simulations using the COCOMO model may be improved to match experiments quantitatively. The model contains various parameters for different types of interactions, which have been calibrated using protein/RNA polymer properties data and condensation measurements from the literature. A straightforward modification would have been to increase the strength of the RNA-peptide cation-π interactions because it makes phase separation more likely and would decrease the length threshold, similar to experiments. However, we also noted that the density in the condensed phase was extremely high,[26,27] suggesting that there would be little to no water within them under experimental conditions[28,29] (Figure S4). Because the experimentally observed condensates showed spherical droplets exhibiting liquid-like growth over time, such high densities may be unexpected. Increasing the size of the beads, described via the σ_i parameter, by 20% decreased the density commensurately (Figure S4). At the same time, increasing the size of the beads decreased the minimum lengths for condensation, in quantitative agreement with experiments (Figure 1). This modification to the model is referred to as COCOMO 1.2σ and was employed in the rest of the simulations in this work (see Table S3 for composition details).

To address the liquid nature of the condensates obtained in the simulations, we analyzed the diffusion inside the clusters and the exchange of peptides/RNA chains with outside the cluster (Table S4). Our results showed that diffusion is retarded inside the clusters, but liquid-like behavior is retained. The residence time of polymers in the condensates range from tens of nanosecond to microseconds. In general, residence times inside the cluster increase with the length of the polymer in the system. Video S1 shows clear polymer diffusion inside the cluster, and Video S2 shows small condensates merging to form a larger one during the simulations. Because of the CG nature of our model, the diffusion rates have only qualitative meaning because diffusion is likely too fast because hydrodynamic effects are not considered, and molecular friction is likely underestimated.

We also investigated whether the system size of the simulations influences the observed results by repeating some of the simulations with a box size twice as big in each dimension, 200 nm. The larger systems sampled RNA-protein mixtures over the entire phase space,

including conditions closer to the phase separation boundary. In all cases, the increased size of the system did not change the results (Figure S5). Larger condensates were obtained, as expected, because of the higher number of molecules available.

To understand the thermodynamics of condensation, isothermal titration calorimetry (ITC) was performed. For mixtures without condensation, the heat change per mole of injectant is relatively flat, while for mixtures with condensation, a strong cooperative transition is observed, and the peptide:RNA mixing ratio increases with peptide length, as expected (Figure S6). Here we estimated the enthalpy of condensation as the change in heat released before and after condensation, yielding enthalpies around −60 to −90 kJ/mol of peptide for all analyzed mixtures. However, for a more detailed thermodynamic analysis of the ITC curves, a specific binding model that applies well to the condensation of disordered polymers will be required.[30] We therefore conclude that the attractive interactions between RNA and peptide are the main drivers of condensation. However, these results do not explicitly explain the length dependence of condensation.

The experiments and simulations indicate minimum peptide and RNA lengths for condensation. However, we found that shorter peptides (i.e., [RGRGG]$_1$) may participate in condensates when they are formed by longer peptides. This led us to speculate that shorter polymers may be able to compensate when the concentration of a longer peptide was too low to observe condensation. To demonstrate this phenomenon, we reduced the concentration of [RGRGG]$_2$ in a mixture with poly(A)$_{20}$ until condensation was lost, between 0.4 and 0.5 mg/mL. Then [RGRGG]$_1$ was added until phase separation was recovered, between 0.6 and 0.7 mg/mL. This observation was confirmed by the CG model; the threshold for [RGRGG]$_2$ condensation was ~0.55 mg/mL, and for [RGRGG]$_1$, phase separation recovery was ~1.8 mg/mL (Figures 3 and S7). We also showed that [RGRGG]$_1$ alone, even at high concentrations, cannot induce phase separation (Figure S8). See Table S5 for composition details of these simulations.

Energetic model of condensation

The phase boundary for condensation is anticorrelated in length of the two components (Figure 1); as poly(A)$_N$ increases in length, the minimum peptide length required decreases while keeping the concentration of residues (adenine, arginine, and glycine) the same. This suggests that the length-dependent condensation in these systems is primarily determined by entropy, as has been mentioned for other protein-RNA phase-separating systems,[21] and is supported by results from the CG model simulations. We also observed that the radial distribution functions (RDFs) between residues in the condensates are not substantially changed by length (Figure S9). In other words, condensates of longer polymers are not more tightly bound than condensates of shorter ones.[31,32]

To quantitatively understand the polymer length dependence of phase separation in the peptide-RNA mixtures considered here, we built a thermodynamic model using details obtained from the CG simulations (see supplemental experimental procedures for details). The model focuses on estimating the stability of the condensates from the free energy based on enthalpy-entropy decomposition. The model holds the monomer (i.e., Amino

acid and nucleotide) concentration constant and does not consider the conditions for phase coexistence required for condensation to be observed.

To find the enthalpy, h, for each monomer in the condensate based on pairwise interactions (i.e., adenine-adenine, adenine-arginine, adenine-glycine, arginine-arginine, arginine-glycine, and glycine-glycine), the RDF from the CG simulations is convolved with the CG potential, integrated, and then multiplied by the number of monomers in the condensate. Relevant here is the change in enthalpy from the disperse phase to the condensed phase. The enthalpy in the disperse phase is estimated in the same manner from bulk-phase RDFs (see supplemental information) and the total number of monomers in the system. The enthalpy of phase separation is then the difference between the enthalpy of the condensate and the enthalpy of the disperse phase. The enthalpy of the disperse phase is about 10% of the enthalpy of the condensate (Figure S10). The enthalpy of the dilute phase in coexistence with the condensate is neglected because it is very small compared with the other enthalpy contributions. When expressed in units of kilojoule per mole monomers that are present in the respective phases, the enthalpies are nearly constant as a function of peptide length and vary relatively little as a function of RNA length (Figure S10).

To find the entropy, s, per polymer in the condensate, different possible sources of entropy were considered. Condensation certainly involves the loss of translational freedom. In addition, differences in conformational entropy, differences in mixing entropy (the peptide/RNA composition is different in the condensates than in the disperse phase), and counterion condensation may also be contributing factors.

For changes in conformational entropy, we examined the radii of gyration for each polymer combination in the condensed and disperse (before condensation is observed) phases. Figure S11 shows that they are indistinguishable for the measured lengths. Only for very long RNA and peptides are there more notable differences in radii of gyration, indicating that they become more extended within the condensate. More extended conformational ensembles in the condensed phase have been related to enhanced phase separation with disordered proteins by reducing steric hindrance and therefore maximizing intermolecular contacts.[33] Additionally, Figure S12 shows that the probability of intrachain distances between residues 1 and 5 in peptides and RNA is unchanged with polymer length and between the disperse and condensed phases. We therefore conclude that the conformational entropy of individual polymers is unchanged between phases.

Based on the simulations, condensates have different compositions of peptides vs. RNA as in the initial disperse systems (Table S6). Consequently, there is a contribution to the change in mixing entropy, but the amount is small, less than 0.2 kJ/mol for $-T\,S$ at 300 K (Figure S13). Therefore, changes in mixing entropy were neglected in the subsequent analysis.

Counterions that may associate with polymers in the disperse phase and release upon condensation were not explicitly considered in COCOMO (or COCOMO 1.2σ) because of its implicit nature. The electrostatic term in the potential has been adjusted to account for charge shielding effects, but the entropic gain of releasing counterions from the polymers during condensation is not considered directly. An initial assumption may be that counterion

association is independent of the polymer lengths as long as the total amount of peptide or RNA is held constant, as in the work presented here. To explore counterion association in more detail, we carried out explicit solvent all-atom simulations of peptides and RNA in the presence of sodium phosphate at pH 8.0 and experimental polymer concentrations of about 1 g/L (Figure S14). The analysis suggests that there may be increased phosphate association with the shortest peptide (RGRGG$_1$), in part because of end effects with the charged N terminus but similar ion association per peptide repeat with longer peptides (RGRGG$_2$ vs. RGRGG$_4$). Na$^+$ association with RNA does appear to increase with RNA length, but the number of ions bound to a single RNA molecule is small (~1) compared with the more significant number of phosphates associated with a peptide (~7 for RGRGG$_4$). This suggests that, while counterion release may be a significant driving force for condensation, it may not be a major factor in explaining the observed polymer length dependence of condensation, which is the focus of the present work.

Therefore, we focus our subsequent analysis on confinement of individual polymers within the condensate as the remaining source of entropy. The change in translational entropy is estimated from the ratio of the accessible volume in the condensate to the volume of the box in the simulations (Equation S10). The volume accessible to a polymer in the condensate was estimated from the molecular volume of the polymer with the argument that the free space inside the dense condensates is too fragmented for a given polymer to fit elsewhere and that the accessible volume is therefore just the volume already occupied by the polymer. It is important to note that, while enthalpy contributions are estimated per monomer, entropy is calculated per polymer because it is the translational freedom of each polymer that is being restricted. The additional restriction of rotational degrees of freedom was not considered in the analysis. Figure 4 shows the resulting enthalpy, entropy ($-T \, S$ at 300 K) and total free energy change between the disperse and condensed phase for each polymer mixture of poly(A)$_{5,10,20}$ and [RGRGG]$_{1,2,3,4,6,8,10}$. With increasing length, DG becomes more negative at the simulated phase boundary. Energies in Figure 4 are given in units of kilojoule per mole of total amino acids. The normalization is therefore independent of peptide and RNA length. If the enthalpy values in Figure 4 are multiplied by the number of amino acids in different size peptides, then the resulting enthalpies are about -100 to -200 kJ/(mol total peptide), which can be compared with the ITC results and is found to be in qualitative agreement.

With the normalization in Figure 4, it appears that the magnitude of the enthalpy decreases significantly for shorter peptides and shorter RNA, different from the enthalpies in Figure S10, which are normalized with respect to the number of monomers inside the condensate and disperse phases. At the same time, entropy ($-T \, S$) appears to increase more gradually for shorter peptides and RNA. The reason is that, with shorter peptides or RNA, the condensates are smaller (Figure S15), containing fewer monomers of the shorter polymer species (peptides or RNA) (Figure S16). This results in smaller total condensate enthalpies, which becomes apparent when enthalpies are normalized against the fixed total number of amino acids in all of the systems but not when enthalpies are normalized against the (decreasing) number of monomers in the condensate. Fewer peptides (or RNA) participating in condensation when the polymers become shorter presumably compensates for the increase in entropy because more molecules are needed to be confined to the condensate

to maintain the same monomer density inside the condensate. As a consequence, the fraction of peptides in the condensate (Figure S17) and the ratio of peptides to RNA (Table S6) change with the shortest peptides and RNA, at least within the simulation. Note that there is little evidence of smaller condensates or altered compositions in the experiment. The theoretical energies based on RDFs for longer RNA and peptides do not reproduce this effect because they assume constant condensate sizes and monomer densities inside the condensates although the total free energy closely matches the actual simulation results. For even shorter polymers, condensation entropy increases rapidly to the point where it cannot be compensated anymore by favorable enthalpy, thereby inhibiting condensation (Figure 4). Fundamentally, this is a consequence of entropy scaling with the number of molecules that increases for shorter peptides, whereas enthalpy scales with the number of (overall constant) monomers.

DISCUSSION

We have systematically shown the effect of length on condensation of positively charged peptides and negatively charged RNA. We observed condensation for combinations of relatively short peptides and RNA, but experiments and simulations show that there does exist a lower limit in terms of RNA or peptide lengths for the systems studied here. The essential driving force for condensation is electrostatic attraction between RNA and arginine residues, counteracted by the entropic cost of condensation. The key reason for the observed length dependence is that the enthalpy of condensation scales with the number of charged units, whereas entropy scales with the number of polymers. We expect that, with peptides with lower charge density (e.g., arginine residues spaced more widely), longer peptides and/or longer RNA would be required for phase separation to be observed. On the other hand, systems with higher charge density have been proven to enable phase separation even with single nucleotides, such as UDP (−3 charge) or UTP (−4 charge), and $polyR_{10}$, as observed in other work.[18] Our study furthermore demonstrates that still shorter polymers may participate in condensates as clients or as partial drivers of condensation together with longer polymers.

The remarkable agreement between experiment and CG simulations suggests that the computational approach could be extended to other sequences and systems as well as other dimensions in phase space, such as concentration or peptide content. The parameterization of the CG model[20] explored a wide range of sequences and concentrations and accounts; for example, for the higher condensation propensity of arginine with nucleotides than lysine, as observed by Fisher and Elbaum-Garfinkle.[18] However, the relative simplicity of the CG model neglects counterion effects that are known to be important factors during condensation[34,35] and does not consider partial secondary structures that are present in many intrinsically disordered proteins (IDPs). Extending the model to explore the importance of these factors will be the subject of future studies.

LLPS has practical applications for inducing high-concentration phases of certain biomolecules. The work here illustrates a quantitative framework for predicting the system components necessary for LLPS. On the other hand, this work demonstrates that a wide range of peptides and RNA can lead to LLPS. In the biological context, this means that

many biomolecules may drive and/or participate in condensate formation in a dynamic manner as cellular concentrations of peptides and RNA fluctuate.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact—Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Lisa Lapidus (lapidus@msu.edu).

Materials availability—This study did not generate any new unique reagents. All RNA and peptide sequences were commercially synthesized as described below.

Data and code availability—All experimental data will be shared by the lead contact upon request. A Jupyter notebook illustrating how to run the model via OpenMM along with sample analysis is available on Github at https://doi.org/10.5281/zenodo.7818659. A Jupyter notebook implementing the energetic model is available at https://doi.org/10.5281/zenodo.7818637. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental methods

LLPS was studied for short proteins and RNA. The peptides $[RGRGG]_{1,2,3,4,6,8,10}$ and Cy5-labeled $[RGRGG]_1$ were obtained from Bio-Synthesis. RNA poly(A)$_{5,10,20}$ and Cy3-labeled poly(A)$_{10,20}$ were obtained from Horizon Discovery. These constructs were used without modification and dissolved in 20 mM sodium phosphate buffer at pH 8.0. Mixtures of unlabeled RNA and peptide were created at concentrations of 1 mg/mL, except where noted, along with 5μM fluorescently labeled samples.

A Nikon A1 Rsi confocal laser-scanning microscope configured on an automated Nikon Eclipse Ti inverted microscope (Nikon Instruments) equipped with a 100× Plan Apo total internal reflection fluorescence (TIRF) oil objective (NA 1.45) was used to capture the confocal images at 100× objective magnification and photomultiplier tube (PMT) detector set to 31 HV. The Cy3 and Cy5 fluorescence were excited using a diode laser at 561 nm and 647 nm and recorded through 595/50 nm and 700/75 nm band-pass emission filters, respectively. Image acquisition was performed using the Nikon NIS Elements software (v.5.21.03). Transmitted light images were recorded using DIC optics at 561 nm.

Confocal images were generally obtained within 3 min of mixing for all combinations of RNA and peptide. To study the growth of condensates with time, 10 images were obtained during a time interval of ~2 min for each of the following: right after mixing and after 10 and 20 min of mixing. Size distribution analysis of condensates was performed using ImageJ software.

Bright-field microscopy was used to identify the effect of Cy3 on condensate formation. Images were obtained using an AmScope compound microscope equipped with a 10× objective (NA 0.25).

Dynamic light scattering (DLS) measurements were carried out using a Zetasizer Ultra Red Advanced Series instrument. Measurements were taken at 25°C using the 173° backscatter detector at a wavelength of 633 nm.

ITC was carried out using a Micro-Cal VP-ITC system at 25°C by injecting a ~250 μM solution of the peptide into either ~17 μM or ~31 μM solutions of the RNA. 30 injections were performed at 10 μL each. These concentrations are much lower than those used in the rest of the measurements in this work. Bright-field microscopy was used to confirm the presence of condensates, as indicated in Figure S6.

CG simulations

Initial observations of length-dependent cluster formation were done with simulations using the COCOMO model[20] (Figure S1). To improve agreement with experimental data and reduce unexpectedly high polymer densities in the condensates, the short-range interaction parameter σ was increased 1.2 times for all residues and nucleotides. This modified version of the model is referred to as COCOMO 1.2σ. The model is described in more detail in the supplemental experimental procedures.

Energetic analysis of peptide-RNA phase separation

Following an analytical treatment introduced by Dutagaci et al.,[4] the residue-level enthalpies and polymer-level entropies are calculated from the CG simulations. Details are given in the supplemental experimental procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

1. Zhang H, Ji X, Li P, Liu C, Lou J, Wang Z,Wen W, Xiao Y, Zhang M, and Zhu X (2020). Liquid-liquid phase separation in biology: mechanisms, physiological functions and human diseases. Sci. China Life Sci. 63, 953–985. 10.1007/s11427-020-1702-x. [PubMed: 32548680]

2. Hyman AA, Weber CA, and Jülicher F (2014). Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. 30, 39–58. 10.1146/annurev-cellbio-100913-013325. [PubMed: 25288112]

3. Banani SF, Lee HO, Hyman AA, and Rosen MK (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 18, 285–298. 10.1038/nrm.2017.7. [PubMed: 28225081]

4. Dutagaci B, Nawrocki G, Goodluck J, Ashkarran AA, Hoogstraten CG, Lapidus LJ, and Feig M (2021). Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. Elife 10, e64004. 10.7554/eLife.64004. [PubMed: 33496264]

5. Flory PJ (1941). Thermodynamics of high polymer solutions. J. Chem. Phys. 9, 660. 10.1063/1.1750971.

6. Huggins ML (1941). Solutions of long chain compounds. J. Chem. Phys. 9, 440. 10.1063/1.1750930.

7. Choi J-M, Holehouse AS, and Pappu RV (2020). Physical principles underlying the complex biology of intracellular phase transitions. Annu. Rev. Biophys. 49, 107–133. 10.1146/annurev-biophys-121219-081629. [PubMed: 32004090]

8. Banani SF, Rice AM, Peeples WB, Lin Y, Jain S, Parker R, and Rosen MK (2016). Compositional control of phase-separated cellular bodies. Cell 166, 651–663. 10.1016/j.cell.2016.06.010. [PubMed: 27374333]

9. Bremer A, Farag M, Borcherds WM, Peran I, Martin EW, Pappu RV, and Mittag T (2022). Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. Nat. Chem. 14, 196–207. 10.1038/s41557-021-00840-w. [PubMed: 34931046]

10. Bai Q, Zhang Q, Jing H, Chen J, and Liang D (2021). Liquid-liquid phase separation of peptide/oligonucleotide complexes in crowded macromolecular media. J. Phys. Chem. B 125, 49–57. 10.1021/acs.jpcb.0c09225. [PubMed: 33373232]

11. Ambadipudi S, Biernat J, Riedel D, Mandelkow E, and Zweckstetter M (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. Nat. Commun. 8, 275. 10.1038/s41467-017-00480-0. [PubMed: 28819146]

12. Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CCH, Eckmann CR, Myong S, and Brangwynne CP (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proc. Natl. Acad. Sci. USA 112, 7189–7194. 10.1073/pnas.1504822112. [PubMed: 26015579]

13. Kaur T, Raju M, Alshareedah I, Davis RB, Potoyan DA, and Banerjee PR (2021). Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies. Nat. Commun. 12, 872. 10.1038/s41467-021-21089-4. [PubMed: 33558506]

14. Alshareedah I, Kaur T, Ngo J, Seppala H, Kounatse LAD, Wang W, Moosa MM, and Banerjee PR (2019). Interplay between short-range attraction and long-range repulsion controls reentrant liquid condensation of ribonucleoprotein-RNA complexes. J. Am. Chem. Soc. 141, 14593–14602. 10.1021/jacs.9b03689. [PubMed: 31437398]

15. Lim J, Kumar A, Low K, Verma CS, Mu Y, Miserez A, and Pervushin K (2021). Liquid–liquid phase separation of short histidine- and tyrosine-rich peptides: sequence specificity and molecular topology. J. Phys. Chem. B 125, 6776–6790. 10.1021/acs.jpcb.0c11476. [PubMed: 34106723]

16. Yuto Akahoshi MM, Shinkai Y, Kurita R, and Shiraki K (2023). Shunsuke Tomita Phase-separation propensity of non-ionic amino acids in peptide-based complex coacervation systems. Biomacromolecules 24, 704–713. [PubMed: 36640113]

17. Tang Y, Bera S, Yao Y, Zeng J, Lao Z, Dong X, Gazit E, and Wei G (2021). Prediction and characterization of liquid-liquid phase separation of minimalistic peptides. Cell Reports Physical Science 2, 100579. 10.1016/j.xcrp.2021.100579.

18. Fisher RS, and Elbaum-Garfinkle S (2020). Tunable multiphase dynamics of arginine and lysine liquid condensates. Nat. Commun. 11, 4628. 10.1038/s41467-020-18224-y. [PubMed: 32934220]

19. Paloni M, Bussi G, and Barducci A (2021). Arginine multivalency stabilizes protein/RNA condensates. Protein Sci. 30, 1418–1426. 10.1002/pro.4109. [PubMed: 33982350]

20. Valdes-Garcia G, Heo L, Lapidus LJ, and Feig M (2022). Modeling concentration-dependent phase separation processes involving peptides and RNA via residue-based coarse-graining. Preprint at bioRxiv. 10.1101/2022.08.19.504518.

21. Laghmach R, Alshareedah I, Pham M, Raju M, Banerjee PR, and Potoyan DA (2022). RNA chain length and stoichiometry govern surface tension and stability of protein-RNA condensates. iScience 25, 104105. 10.1016/j.isci.2022.104105. [PubMed: 35378855]

22. Sanchez-Burgos I, Espinosa JR, Joseph JA, and Collepardo-Guevara R (2022). RNA length has a non-trivial effect in the stability of biomolecular condensates formed by RNA-binding proteins. PLoS Comput. Biol. 18, e1009810. 10.1371/journal.pcbi.1009810. [PubMed: 35108264]

23. Tejedor AR, Garaizar A, Ramírez J, and Espinosa JR (2021). 'RNA modulation of transport properties and stability in phase-separated condensates. Biophys. J. 120, 5169–5186. 10.1016/j.bpj.2021.11.003. [PubMed: 34762868]

24. Lebold KM, and Best RB (2022). Tuning Formation of protein–DNA coacervates by sequence and environment. J. Phys. Chem. B 126, 2407–2419. 10.1021/acs.jpcb.2c00424. [PubMed: 35317553]

25. Iqbal A, Arslan S, Okumus B, Wilson TJ, Giraud G, Norman DG, Ha T, and Lilley DMJ (2008). Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. Proc. Natl. Acad. Sci. USA 105, 11176–11181. 10.1073/pnas.0801707105. [PubMed: 18676615]

26. Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J, and Fawzi NL (2019). Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. Nat. Struct. Mol. Biol. 26, 637–648. 10.1038/s41594-019-0250-x. [PubMed: 31270472]

27. Brady JP, Farber PJ, Sekhar A, Lin Y-H, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, and Kay LE (2017). Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. Proc. Natl. Acad. Sci. USA 114, E8194–E8203. 10.1073/pnas.1706197114. [PubMed: 28894006]

28. Ahlers J, Adams EM, Bader V, Pezzotti S, Winklhofer KF, Tatzelt J, and Havenith M (2021). The key role of solvent in condensation: mapping water in liquid-liquid phase-separated FUS. Biophys. J. 120, 1266–1275. 10.1016/j.bpj.2021.01.019. [PubMed: 33515602]

29. Burke KA, Janke AM, Rhine CL, and Fawzi NL (2015). Residue-by-Residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. Mol. Cell 60, 231–241. 10.1016/j.molcel.2015.09.006. [PubMed: 26455390]

30. Kayitmazer AB (2017). Thermodynamics of complex coacervation. Adv. Colloid. Interface Sci. 239, 169–177. 10.1016/j.cis.2016.07.006. [PubMed: 27497750]

31. Sing CE (2017). Development of the modern theory of polymeric complex coacervation. Adv. Colloid. Interface Sci. 239, 2–16. 10.1016/j.cis.2016.04.004. [PubMed: 27161661]

32. Ou Z, and Muthukumar M (2006). Entropy and enthalpy of polyelectrolyte complexation: Langevin dynamics simulations. J. Chem. Phys. 124, 154902. 10.1063/1.2178803. [PubMed: 16674260]

33. Garaizar A, Sanchez-Burgos I, Collepardo-Guevara R, and Espinosa JR (2020). Expansion of intrinsically disordered proteins increases the range of stability of liquid–liquid phase separation. Molecules 25, 4705. [PubMed: 33076213]

34. Khokhlov AR, and Nyrkova IA (1992). Compatibility enhancement and microdomain structuring in weakly charged polyelectrolyte mixtures. Macromolecules 25, 1493–1502. 10.1021/ma00031a021.

35. Salehi A, and Larson RG (2016). A molecular thermodynamic model of complexation in mixtures of oppositely charged polyelectrolytes with explicit account of charge association/dissociation. Macromolecules 49, 9706–9719. 10.1021/acs.macromol.6b01464.
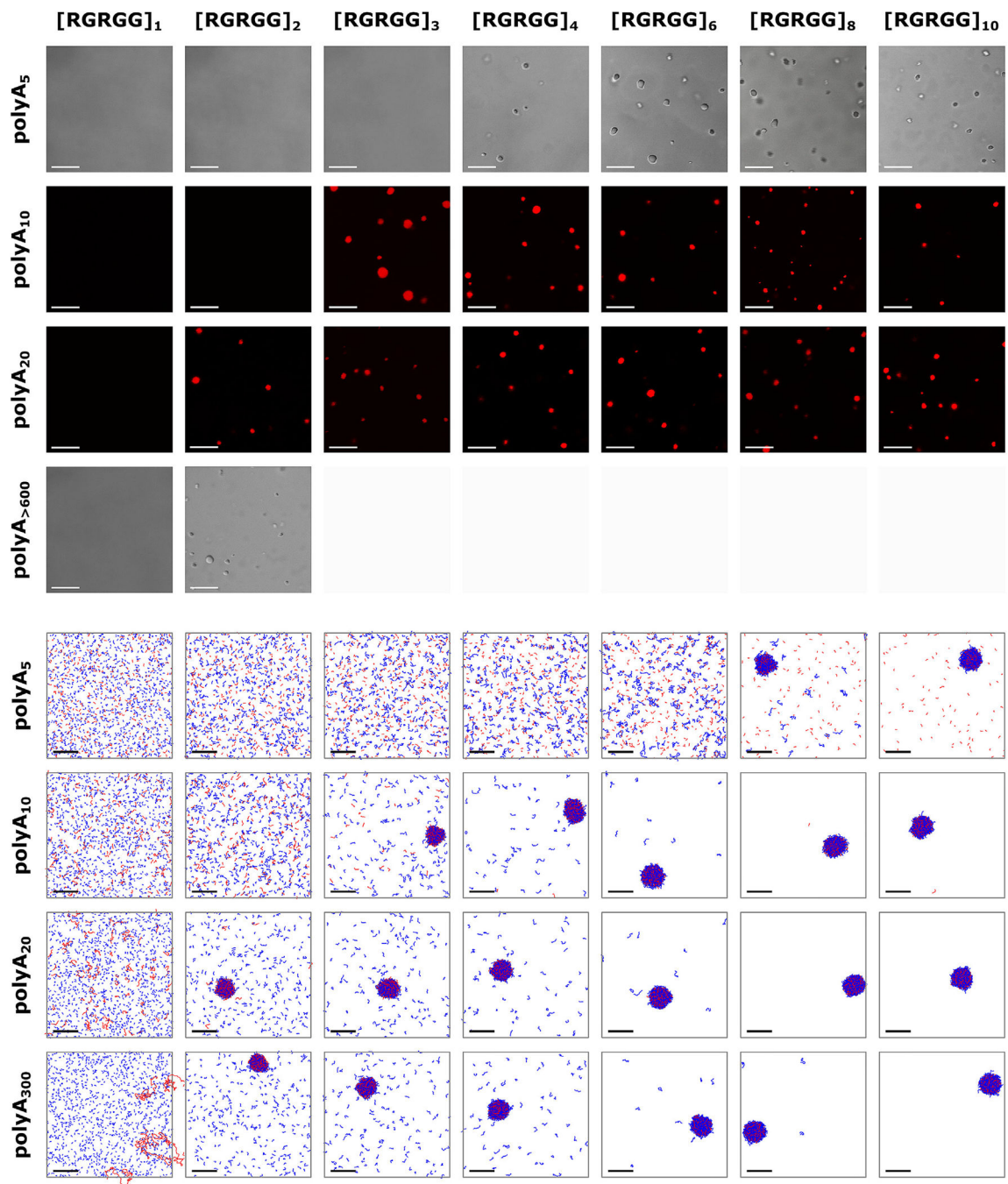
Highlights

This work investigates the length dependence of LLPS using simulation and experiment

Polymers as short as 10–20 monomer subunits condense at a concentration of 1 mg/mL

Confinement entropy is the primary driver of the length dependence of condensation

More entropy is lost to confine more, shorter polymers than fewer, longer polymers

**Figure 1. Condensates of various lengths of RNA and peptide**

Experimental (top) and simulation using COCOMO 1.2σ (bottom) results are shown for different mixtures of RNA and peptides at 1 mg/mL. Images were obtained by confocal and DIC microscopy. 5 µM Cy3-labeled poly(A) was used for fluorescence. The final frames of the simulation trajectories show the central box of the periodic system for each simulated system, with RNA and peptide colored in red and blue, respectively. Scale bars represent 10 µm in experimental panels and 20 nm in simulation panels.
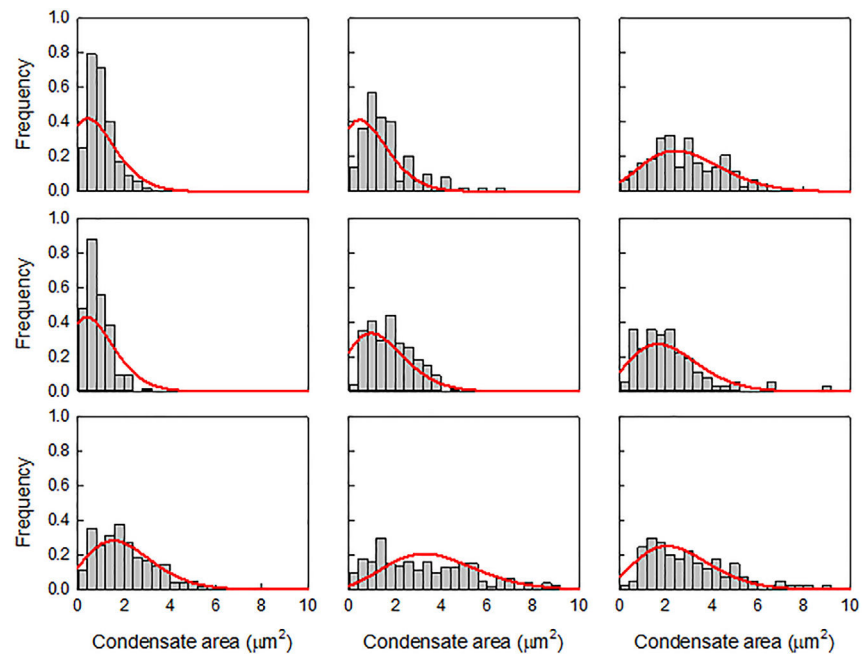
Figure 2. Condensate size distribution over time

Top row: poly(A)$_{10}$ and [RGRGG]$_4$ around (from left) 4:15, 11:05, and 20:56 min after mixing. Middle row: poly(A)$_{10}$ and [RGRGG]$_{10}$ around 3:23, 10:50, and 20:44 min after mixing. Bottom row: poly(A)$_{20}$ and [RGRGG]$_4$ around 3:32, 11:54, and 21:11 min after mixing. The bin width is 0.4 μm^2. The red lines fit a Poisson distribution. Concentrations are the same as in Figure 1.
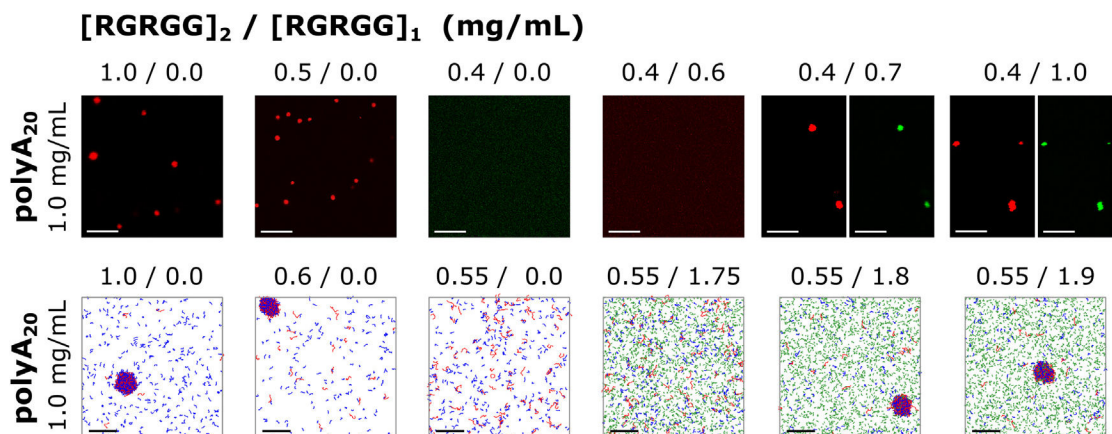
Figure 3. Phase separation recovery with small peptides

Experimental (top) and simulation using COCOMO 1.2σ (bottom) results show the loss of LLPS when lowering the [RGRGG]$_2$ concentration below a certain threshold and its recovery when adding enough [RGRGG]$_1$. Poly(A)$_{20}$ was kept at 1 mg/mL for these assays. The red condensates in the top panels show the fluorescence of [RGRGG]$_2$-Cy3, and the green condensates show the fluorescence of [RGRGG]$_1$-Cy5, indicating coexistence in the condensates. The final frames of the trajectories are shown in the bottom panels for each simulated system, with poly(A)$_{20}$, [RGRGG]$_1$, and [RGRGG]$_2$ colored in red, green, and blue, respectively. Scale bars represent 10 μm in experimental panels and 20 nm in simulation panels.
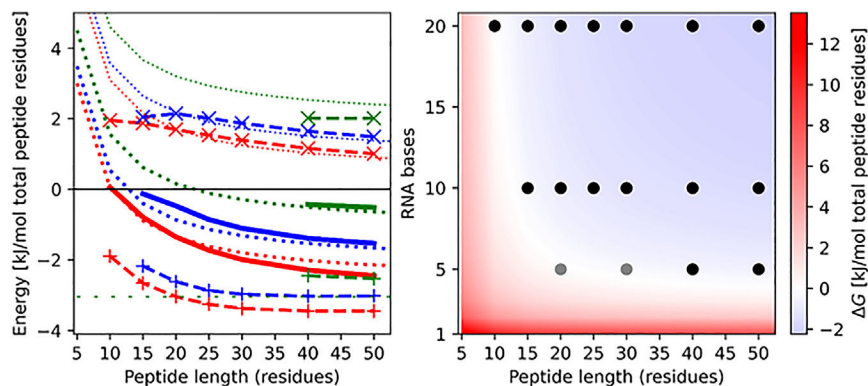
Figure 4. Energetic analysis of peptide-RNA condensates based on enthalpy-entropy composition
Condensate enthalpies according to Equation S3 (dashed lines with +), entropies ($-T\,S$ at 300 K) according to Equation S9 (dashed lines with x), and total free energies according to Equation S2 (solid lines) are shown on the left as a function of peptide length with different RNAs (poly(A)$_{20}$, red; poly(A)$_{10}$, blue; poly(A)$_5$, green). Energies are given in units of kilojoules per mole of the total number of amino acid residues in the peptides. Energies were estimated by averaging over five replicate simulations. The statistical errors of the mean are less than 0.1 kJ/mol and are not shown. Lines with short dashes reflect total free energy estimates using densities and RDFs based on poly(A)$_{20}$/[RGRGG]$_4$. Lines with long dashes and dotted lines show the estimated enthalpic and entropic contribution to the free energy, respectively. The contour plot on the right shows the total free energies as a function of peptide and RNA length obtained with the same parameters. Dots indicate combinations of peptide/RNA for which condensates were observed experimentally (black and gray) and in the simulations (black only).