

Modeling Concentration-dependent Phase Separation Processes Involving Peptides and RNA via Residue-Based Coarse-Graining

Gilberto Valdes-Garcia, Lim Heo, Lisa J. Lapidus, and Michael Feig*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 669–678

Read Online

ACCESS |



Metrics & More

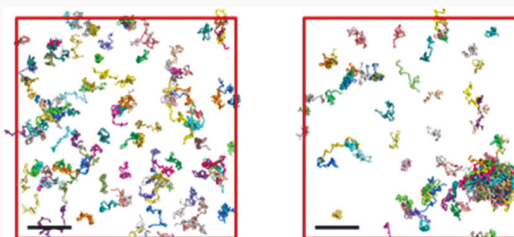


Article Recommendations



Supporting Information

ABSTRACT: Biomolecular condensation, especially liquid–liquid phase separation, is an important physical process with relevance for a number of different aspects of biological functions. Key questions of what drives such condensation, especially in terms of molecular composition, can be addressed via computer simulations, but the development of computationally efficient yet physically realistic models has been challenging. Here, the coarse-grained model COCOMO is introduced that balances the polymer behavior of peptides and RNA chains with their propensity to phase separate as a function of composition and concentration. COCOMO is a residue-based model that combines bonded terms with short- and long-range terms, including a Debye–Hückel solvation term. The model is highly predictive of experimental data on phase-separating model systems. It is also computationally efficient and can reach the spatial and temporal scales on which biomolecular condensation is observed with moderate computational resources.



COCOMO simulates IDP/RNA LLPS

INTRODUCTION

Liquid–liquid phase separation (LLPS) is gaining importance in understanding membrane-less subcellular organization. In the cell, liquid condensation is mediated by polymers, mainly proteins and RNA.¹ The first membrane-less compartment observed was within the nucleus of neuronal cells in the 1830s, later termed the nucleolus.² Today, many other compartments that are not delimited by membranes are also known. Examples include the Cajal bodies,^{3–5} PML bodies,³ and nuclear speckles^{6,7} in the nucleus, as well as the stress granules,^{8,9} P-bodies,¹⁰ and germ granules¹¹ in the cytoplasm. More recent studies indicate that biomolecular condensation may be much more ubiquitous than these well-known cellular components.¹² Despite differences in composition, location, and function, the condensates share similarities in shape, dynamics, and assembly mechanisms.¹³

Many biophysical techniques, including microscopy and structural and compositional analysis, have been applied to study phase separation (PS).¹⁴ On the theoretical side, analytical approaches based on polymer theories have been applied.^{15,16} Finally, computer simulations have been used to explain the interactions that stabilize PS, with the most detailed insight derived from atomistic simulations.^{17–19} However, atomistic simulations are challenged²⁰ by the significant computational resources required to reach the time scales (μ s–ms) and spatial scales (>100 nm) on which LLPS is observed experimentally. Coarse-grained (CG) models are a computationally more efficient alternative^{21,22} and they have been used successfully to study PS via simulation.²³ Earlier studies stem from the colloid field, with more limited applicability to specific biological systems.²⁴ More recently,

biology-focused models at different resolution levels have been developed, ranging from models representing proteins/RNA at the molecule level as single particles²⁵ to patchy particles,²⁶ residue-based models,^{23,27–33} and higher-resolution models with multiple particles per residues.^{34–36}

Residue-based, sequence-dependent models have become very popular for studying PS as they combine computational efficiency with an ability to retain key physicochemical features of specific biological systems. Many rely on a hydrophobicity scale (HPS) using an Ashbaugh–Hatch modified Lennard–Jones potential³⁷ to describe short-range interactions.^{23,27–29,33} In some cases, the HPS are implemented without further optimization,²³ while other models apply machine learning and Bayesian parameter-learning procedures for optimization.^{27,29} Further HPS optimizations have focused on cation– π interactions given its importance in PS.^{31,32} These CG models have been able to reproduce some experimental data reasonably well. One limitation of the existing models, as will be shown below, is that concentration-dependent PS is not reproduced well, which limits their predictive ability since concentration is a key factor governing phase behavior.³⁸ Moreover, most models focus on describing only protein

Received: August 19, 2022

Published: January 6, 2023



systems without a compatible nucleic acid model, preventing studies of increasingly important peptide-RNA condensates.

Here, we propose a new residue-based CG model, termed COCOMO (Concentration-dependent Condensation Model) to describe PS in peptide-only and peptide-RNA systems. Our goal was to develop a simple yet accurate model for describing coacervation in systems containing only peptides or mixtures of peptides and RNA in a concentration-dependent manner. The model was designed to minimize the number of necessary parameters to maintain as much general applicability as possible, and it introduces a term to account for solvation effects at the residue level. We demonstrate that our model can accurately reproduce experimental PS data, including in systems that were not included in the parameterization while maintaining a balanced description of individual polymer properties of peptides and RNA molecules.

METHODS

Coarse-Grained Model. In COCOMO, each residue, either a protein amino acid or an RNA nucleotide, is represented as a single spherical particle. The total interaction energy is given by:

$$U_{\text{total}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{short-range}} + U_{\text{electrostatic}} \quad (1)$$

where U_{bond} is the harmonic potential for chain connectivity:

$$U_{\text{bond}} = \sum_{i=1}^{N-1} \frac{1}{2} k_{\text{bond}} (l_{i,i+1} - l_0)^2 \quad (2)$$

where $l_{i,i+1}$ is the distance between two neighboring residues, k_{bond} is the spring constant, and l_0 the equilibrium bond length. We choose $k_{\text{bond}} = 4184 \text{ kJ/mol}\cdot\text{nm}^2$, which is a softer value than in all-atom potentials; for proteins, $l_0 = 0.38 \text{ nm}$, from the average $\text{C}\alpha$ - $\text{C}\alpha$ distance, and, for nucleotides, $l_0 = 0.5 \text{ nm}$, corresponding to the average distance between backbones for single-stranded nucleic acids.³⁹

U_{angle} is the angle potential between three neighboring particles to account for chain stiffness

$$U_{\text{angle}} = \sum_{i=1}^{N-2} \frac{1}{2} k_{\text{angle}} (\theta_{i,i+1,i+2} - \theta_0)^2 \quad (3)$$

where $\theta_{i,i+1,i+2}$ is the angle between three neighboring beads, with the angle constant, k_{angle} , equal to $4.184 \text{ kJ/mol}\cdot\text{rad}^2$ for proteins and $5.021 \text{ kJ/mol}\cdot\text{rad}^2$ for nucleic acids. The target angle was set to $\theta_0 = 180^\circ$.

Nonbonded pairwise interactions consist of a short-range 10–5 Lennard-Jones potential, $U_{\text{short-range}}$, and a long-range Debye–Hückel potential, $U_{\text{electrostatic}}$, as follows:

$$U_{\text{short-range}} = \sum_{i,j} 4(\epsilon + \epsilon_{\text{cation}-\pi}) \left(\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{10} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^5 \right) \quad (4)$$

$$U_{\text{electrostatic}} = \sum_{i,j} \frac{(A_{i,j} + A_{0,i,j})}{r_{i,j}} \bullet e^{-r_{i,j}/\kappa} \quad (5)$$

where $r_{i,j}$ is the interparticle distance, $\sigma_{i,j}$ is the distance at which the potential is zero, ϵ is the depth of the potential well, $\epsilon_{\text{cation}-\pi}$ is added to augment cation– π interactions, $A_{i,j} = A_i \times A_j$ describes attractive or repulsive long-range interactions, $A_{0,i,j} = A_{0i} + A_{0j}$ reflects the effective repulsion between polar

residues due to solvation effects, and κ is the Debye–Hückel screening length.

Optimized nonbonded parameters values are as follows: $\epsilon_{\text{polarresidues}} = 0.4 \text{ kJ/mol}$ and $\epsilon_{\text{non-polarresidues}} = 0.41 \text{ kJ/mol}$. Arg, Asn, Asp, Cys, Gln, Glu, His, Lys, Ser, and Thr were considered polar residues; Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Tyr, and Val were considered nonpolar. For nucleotides, we used $\epsilon_{\text{nucleotides}} = 0.41 \text{ kJ/mol}$. We further adjusted for cation– π interactions by adding $\epsilon_{\text{R/K-F/Y/W}} = 0.3 \text{ kJ/mol}$ for interactions within proteins and $\epsilon_{\text{R/K-nucleic}} = 0.2 \text{ kJ/mol}$ for protein-RNA cation– π interactions. The effective radii $\sigma_i = 2r_i \times 2^{-1/6}$ were set from the radius, r_i , of a sphere with equivalent volume of a given residue. A_i was calculated from residues charge (q_i) + 1 for Arg/Lys, –1 Asp/Glu and nucleotides, and 0 for the rest according to $\text{sign}(q_i) \sqrt{0.75|q_i|}$ as proposed previously.²⁵ A_0 was set to 0.05 for polar residues and nucleotides, it was set to 0 for nonpolar residues. Finally, $\kappa = 1 \text{ nm}$, except when noted, corresponding to an ionic strength of $\sim 100 \text{ mM}$. Table S1 reports all residue-specific parameters employed in the model.

Molecular Dynamics Simulations. The model was simulated by molecular dynamics simulations using OpenMM 7.7.0.⁴⁰ Langevin dynamics was applied with a friction coefficient of 0.01 ps^{-1} . Initially, a 5,000 step steepest descent minimization was performed followed by 20,000 steps of MD with a time step of 0.01 ps. After that, systems were run for production using a time step of 0.02 ps. Nonbonded interactions were calculated using periodic boundary conditions and truncated at a cutoff distance of 3 nm. We also tested the use of force switching according to Steinbach and Brooks (eq 10 in that work)⁴¹ with $r_{\text{on}} = 2.9 \text{ nm}$ and $r_{\text{off}} = 3.1 \text{ nm}$ and found little difference for selected systems (Figure S13). However, the simulations slowed down by 20%. Therefore, we continued to use truncated potentials in the remainder of this work. The cutoff was chosen long enough for the potentials to approximately reach zero. However, the cutoff distance remains a model parameter and may affect model performance as other recent work has shown. Residues separated by one bond were excluded from nonbonded interactions energy calculation.

Individual protein and RNA chains were simulated in a cubic box with a side length of 300 nm at 298 K. For systems described here, five replicates were run over 500 ns, saving coordinates every 200 ps for each system. This was long enough to establish converged ensembles from which average radii of gyration and persistence length could be extracted. The first 100 ns of each trajectory were excluded from the analysis. An initial random conformation for each chain was obtained using a custom python script. Topology files were generated with the MMTSB Tool Set⁴² and CHARMM v44b2.⁴³ Using a GeForce RTX 2080 Ti GPU card, we could simulate 100 ns of a 100-residue protein in 5 min.

Polymer properties were calculated and averaged from the five replicates. For protein sequences we calculated the radius of gyration using MDTraj library.⁴⁴ For RNA chains we also determined the end-to-end distance and the orientational correlation factor (OCF) as a function of separation along the chain $li - jl$, calculated according to:

$$\text{OCF}(li - jl) = \langle \cos \theta_{i,j} \rangle = \langle \hat{r}_i \bullet \hat{r}_j \rangle \quad (6)$$

where $\hat{r}_i$ and $\hat{r}_j$ are normalized vectors between any $i, i + 1$ and $j, j + 1$ bonded residues in the chain, respectively.

Systems consisting of multiple chains, either proteins or protein-RNA mixtures, were simulated in boxes of side length ranging from 100 to 200 nm to match different concentrations for different systems (see below). Multiple-chain systems were assembled as follows: (i) short simulations were carried out for each of the system components, as described above, (ii) representative conformations were obtained as cluster centers from RMSD-based clustering using the *MiniBatchKMedoids* method implemented in MSMBuild, ⁴⁵ (iii) conformations were randomly picked and placed in the simulation box at random positions and with random orientations, but avoiding any two beads between different molecules to be closer than 5 nm, using a custom python script, until the desired concentration was reached, (iv) topology files were generated for the assembled systems using the MMTSB Tool Set ⁴² and CHARMM v44b2. ⁴³ For all systems, one replicate was run for 20 μ s, with coordinates saved every 500 ps. This was considered long enough to determine the ability to form condensates, as typical times for nucleating condensate formation were on the order of microseconds. In the concentrated systems, we evaluated the ability of our model to reproduce heterotypic and homotypic PS in a concentration-dependent manner. In addition, temperature phase diagrams were constructed for some systems by running simulations at fixed initial concentrations at temperatures ranging from 250 to 310 K. Using a GeForce RTX 2080 Ti GPU card, we could simulate 100 ns of a 30,000-residue system in 15 min.

Clustering analysis on multichain systems was performed via contact-based criteria. Using an in-house Python script, we calculated pairwise distances between residues. Two residues were considered in contact if they were closer than the cutoff distance of 0.9 nm, and two chains were considered part of the same cluster if they have at least one contact between any of their residues. We computed the largest cluster size along simulation time and cluster size distribution in terms of the number of members.

To characterize the protein–protein and protein-RNA aggregates formed in different systems, we calculated the mass concentration radially from the cluster's center of mass (COM) to a distance equal to half of the box side length. The results were fitted to a sigmoid curve:

$$f(x) = \frac{L}{1 + e^{a(x-b)}} + c \quad (7)$$

where L , a , b , and c are fitting parameters. Conveniently, b is the distance where the concentration drops to half the value from cluster COM and allowed us to determine the cluster dimension. We calculated the density of condensed and diluted phases as the average density at distances $\leq b/2$ and $\geq 2b$, respectively.

To characterize the dynamic of the cluster, we determined the translational diffusion, D_{tr} , from the mean square displacement (MSD) of the chains between time t and $(t + \tau)$ for a given lag time, τ . Then, D_{tr} is calculated from the linear fitting MSD(τ) versus τ

$$D_{tr} = \frac{\text{MSD}(\tau)}{6\tau} \quad (8)$$

The analysis was performed over 4 μ s of trajectory. Fitting was made up to $\tau = 10$ ns. We also calculate the cluster

residence time, as the average time the molecules remain in the dense phase over the last 4 μ s of the trajectory.

All figures from simulations were prepared using PyMOL. ⁴⁶ A Jupyter notebook illustrating how to run the model via OpenMM along with sample analysis is available on github at: <https://github.com/feiglab/cocomo>.

Model Parametrization and Test Systems. To parametrize the model, we considered a set of 45 intrinsically disordered or unfolded proteins ranging from 8 to 198 residues in length (Table S2), primarily to reproduce the experimentally measured radii of gyration (Table S3). We note that some of the systems are true disordered peptides, while for others disorder was induced by denaturants. A number of studies measuring intramolecular FRET of Trp–Cys quenching have observed that foldable proteins in high denaturant are significantly more expanded and dynamic than in low or no denaturant, but the effect of denaturant is much less pronounced for intrinsically disordered proteins, which have similar polymer properties to denatured proteins. ^{47–49} Since this model is primarily concerned with modeling the polymeric properties of intrinsically disordered proteins, we consider both types of systems a good parameterization set. A second set of 26 intrinsically disordered proteins not included in the set used for parameterization (Tables S4 and S5) was used for validation. PolyAde-30, polyUra-30, and polyUra-40 were used to parametrize RNA single chain stiffness (Table S6). In addition, parameterization focused on reproducing the concentration-dependent homotypic PS for FUS LCD and LAF-1 RGG peptides (Table S7). The model was then validated for predicting PS in three additional homotypic systems (A1 LCD, hTau40-k18, and Ddx4, Table S7), for heterotypic protein systems (FUS LCD with FUS RGG3 or [RGRGG]₅, Table S8) and protein-RNA systems for which PS has been reported experimentally (Table S9).

To comparatively assess the performance of COCOMO, we ran simulations with different residue-based models published recently. For protein only systems we compared with the models developed by Dignon et al. 2018, ²³ Dannenhoffer et al. 2021, ²⁷ Regy et al. 2021, ²⁸ and the M3 parameters set from Tesei et al. 2021. ²⁹ On the other hand, systems involving also RNA were simulated using the model by Regy et al. 2020. ³³ Simulations using other models were done as described in their original papers, that is all force field parameters and simulation parameters were applied as published.

RESULTS AND DISCUSSION

Model Parametrization. The bonded and nonbonded terms in the COCOMO model were parametrized mainly via iterative parameter scans. The main goal was to maximize the agreement with experimental radii of gyration (R_g) based on χ^2 values for the IDP systems given in Tables S2 and S3, reproduce RNA polymer parameters given in Table S6, and reproduce PS at the concentrations given from experiment for FUS LCD and LAF-1 RGG peptides.

The bonded terms are comprised of bond and angle potentials. The bond length was based on geometry and a k_{bond} value “softer” than those from an all-atom potential was chosen. Our chosen value was in the range of other residue-based models published previously, ^{23,27–29} and it was not optimized further. For the angle potential, we parametrized k_{angle} to reproduce R_g distributions for intrinsically disordered proteins (IDPs) (Figure 1), reproduce R_g and persistence length values for RNA (Figure 2). We note that other recently

proposed residue-based models for IDPs do not have an angle term.^{23,27–29}

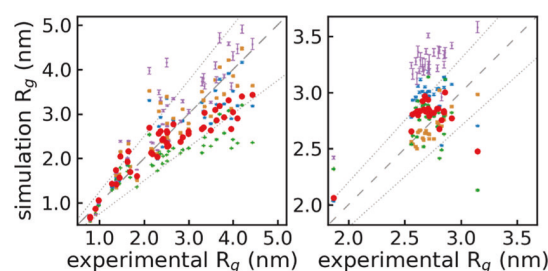


Figure 1. Experimental vs simulated radius of gyration. Scatter plots of the parametrization set with 45 proteins (A) and the validation set consisting of 26 proteins (B) are shown for COCOMO (red circles) and results obtained by us using the models by Regy et al. 2021²⁸ (blue triangle), Tesei et al. 2021²⁹ (tan square), Dignon et al. 2018²³ (green filled plus), and Dannenhoffer et al. 2021²⁷ (purple star). Error bars indicate standard errors from variations between five replicates. Deviations between simulated ensembles and experimental values

according to $\chi^2 = \sum_{i=1}^N \frac{(R_{g,i,\text{exp}} - R_{g,i,\text{sim}})^2}{R_{g,i,\text{exp}}^2}$ were 3.69, 3.39, 3.50, 7.92,

and 6.45 in the parametrization set (A), and 0.40, 0.95, 0.10, 0.81, and 3.1 in the validation set (B), respectively, for COCOMO, and the other models in the order listed above. As a guide to the eye, a dashed line indicates the identity function, and a dotted line shows 25% (A) and 10% (B) deviations from the experimental R_g values.

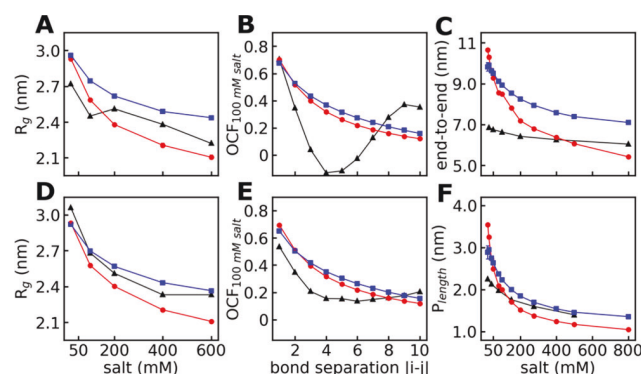


Figure 2. Polymer properties of short RNA sequences. Effect of charge screening on the radius of gyration (R_g) of polyAde-30 (A), and polyUra-30 (D), end-to-end distance of polyUra-40 (C), and persistence length of polyUra-40 (F). The orientational correlation factor (OCF) for polyAde-30 (B), and polyU-30 (E) at 100 mM salt concentration is also shown. Results are shown in red circles, blue squares, and black triangles for COCOMO, the Regy et al. 2020³³ model, and experimental values,^{55,57} respectively. Different salt concentrations were reflected by varying the Debye length, κ , in eq 5, as they can be related according to $\kappa = (De_0k_B T / 2c_i N_A q_e^2)^{1/2}$, where D is the dielectric constant equal to 80, ϵ_0 is the permittivity of free space, k_B is Boltzmann's constant, T is the temperature, c_i is the salt concentration, N_A is Avogadro's constant, and q_e is the charge of an electron.

Nonbonded interactions affect both polymer properties and PS behavior, so both properties were considered together during optimization. For short-range interactions, we initially started with a single value of ϵ for all interactions, but we found slightly better performance when separating ϵ values between polar and nonpolar residues, although the final optimized values that are very similar. It was more important to account

explicitly for cation– π interactions that contribute significantly to PS^{50–54} by increasing ϵ values for protein/protein, $\epsilon_{R/K-F/Y/W}$, and protein/RNA, $\epsilon_{R/K-nucleic}$ interactions involving interactions between basic amino acids and aromatic moieties. Long-range interactions were mostly determined by nominal charges of amino acids and/or RNA bases, but an additional correction term, A_0 , was applied to polar residues to effectively account for solvation effects by creating weak repulsion relative to interactions between hydrophobic residues.

The sensitivity of the model to each of the finally chosen parameters is illustrated with the analysis shown in Figures S1 and S2.

Sampling of Intrinsically Disordered Proteins (IDPs).

The performance of COCOMO on the training (Table S2) and validation (Table S4) sets in terms of reproducing experimental R_g values is shown in Figure 1. For comparison, we also ran individual protein chains with four recently published residue-based coarse-grained models.^{23,27–29} The results of our model are similar to the top performing models among those we tried (Figure 1A) based on χ^2 values of 3.69 for the training set and 0.4 for the validation set. For the majority of proteins, the simulated R_g fall within 25 and 10% of the experimental values in the parametrization and test set, respectively (Figure 1, dotted lines). The agreement between simulation and experiment is generally good, but as with most other models, our model also shows systematic deviations where smaller systems tend to be less compact, whereas larger systems are more compact than suggested by the experimental values.

To test how R_g is affected by the parameters of our model, we systematically varied one parameter after another while keeping all other values at their final optimized values and repeated simulations (Figure S1). Our results show that variations of A_0 , ϵ_{polar} , and $\epsilon_{\text{non-polar}}$ have a strong effect on R_g , k_{angle} has a moderate effect, and $\epsilon_{\text{cation-}\pi}$ and k_{bond} have minimal effects (Figure S1). This analysis shows that slightly more optimal parameters could be found if the goal is only to reproduce the experimental R_g values. However, with those values, the concentration-dependent PS behavior in concentrated systems is not reproduced correctly (see below).

Sampling of RNA. Experimental data on polymer properties is available for polyAde-30, polyUra-30, and polyUra-40. Individual chains of these polynucleotides were simulated using COCOMO and compared with the RNA bead model developed by Regy et al. 2020.³³ The results show good agreement with the experimental measurements of R_g , end-to-end distances, and persistence lengths (P_{length}) (Figure 2A,C,D,F). We also calculated the orientation correlation function (OCF) to quantify the directional persistence of the chain. The correlation decreases with chain distances (Figure 2B,E) similar to experimental values. However, short-range interactions due to base stacking⁵⁵ between bases separated by 3–5 bases that give rise to a dip in the correlation function in the experiment, especially for polyAde-30 are not reproduced because our model does not include beads to represent side chains.

In the simulations a Debye–Hückel term was used to treat electrostatic screening of charge interactions by ions in solution. Therefore, a variation in ionic strength could be modeled by changing the screening length, κ . This treatment of electrostatics has been successful in reproducing the ionic strength dependence of polyelectrolyte macromolecules'

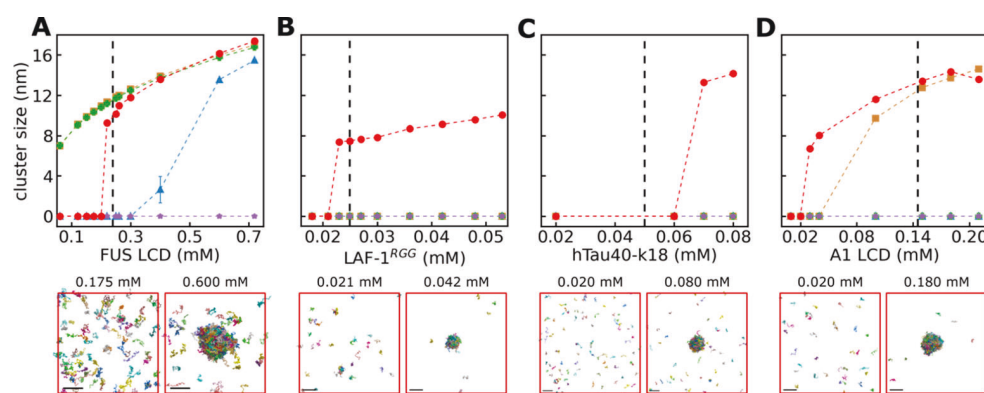


Figure 3. Protein homotypic phase separation. Cluster sizes as a function of concentration and simulation snapshots are shown for FUS LCD (A), LAF-1^{RGG} (B), hTau40-k18 (C), and A1 LCD (D). Cluster sizes were averaged over the last 4 μ s of the simulation. Results with COCOMO (red circles) are compared with results using models from Regy et al. 2021²⁸ (blue triangle), Tesei et al. 2021²⁹ (tan square), Dignon et al. 2018²³ (green filled plus), and Dannenhoffer et al. 2021²⁷ (purple star). Experimental LLPS concentration thresholds^{58–61} are shown as dashed lines. The final frames of simulations using our model (lower panels) are shown for concentrations below (left) and above (right) the experimental LLPS threshold for each system. Coloring is used to indicate different chains. The size bar represents 20 nm.

association in similar models.⁵⁶ We also find here that the experimental trends are well-reproduced with a decrease of R_g and P_{length} values at higher salt concentration as in the experiments (Figure 2A,D,F). Best agreement is observed around 100 mM salt concentration ($\kappa = 1$ nm), which was the condition chosen for our simulations.

Compared to the Regy et al. 2020³³ model, COCOMO agrees similarly well with the experimental data, but while the Regy et al. model generates polymers that are slightly less compact than in experiment, COCOMO generates conformations that are slightly more compact.

The sensitivity of RNA simulations to the value of parameters was also evaluated. As observed for proteins, the R_g and P_{length} values are most sensitive to the choice of A_0 and $\epsilon_{\text{non-polar}}$ values, while the choice of k_{angle} and k_{bond} showed much smaller effects (Figure S2).

Protein Homotypic Phase Separation. The main focus of COCOMO is on modeling PS phenomena with a model that also maintains realistic polymer behavior of individual molecules. We begin by describing the performance of homotypic systems. We parameterized the model using data from the FUS LCD and LAF-1^{RGG} proteins and then tested it with hTau40-k18 and A1 LCD. Detailed information on the simulated homotypic systems is given in Table S7 and results are summarized in Figure 3. We observe PS dependent on the concentration for all systems in good agreement with experimental thresholds (Figure 3, dashed lines).^{58–61} For comparison, we ran the same protein system using previously published models.^{23,27–29} With those other models, the PS behavior was inconsistent among different protein systems. For FUS LCD, with the Tesei et al. 2021²⁹ and Dignon et al. 2018²³ models PS occurred at all of the concentrations tested here, while concentration dependence was seen for the Regy et al. 2021²⁸ model and no PS when simulating with the Dannenhoffer et al. 2021²⁷ model (Figures 3 and S3). For systems containing LAF-1^{RGG} and A1 LCD, no PS was observed at any tested concentration for the four models. Finally, the Tesei et al. 2021²⁹ model was the only one showing PS in A1 LCD simulations. This suggests that the Regy et al. 2021²⁸ and Tesei et al. 2021²⁹ models may be able to at least qualitatively describe concentration-dependent PS,

but with the concentration threshold shifted to higher and lower values, respectively.

Analysis of the trajectories showed that the largest cluster size generally stabilizes by 10 μ s and increasing concentration accelerates the time necessary to form clusters (Figures S4 and S5). In the case where condensates were not observed, it is in principle possible that condensates did not form due to slow nucleation near the condensation concentration threshold. To test for this possibility, we performed additional simulations for the FUS LCD system starting from the final snapshot at 0.22 mM, with a formed condensate, but then increased the box sizes to lower the concentrations to 0.12 to 0.20 mM, where PS was not observed when starting from randomly distributed polymers. The initial condensate melted at all concentrations except for 0.20 mM (Figure S6), indicating that there is only slight hysteresis around the reported concentration thresholds due to slow nucleation kinetics when forming or melting condensates near the critical concentration.

In some cases, all the proteins in the box became part of the condensate by the end of the simulations, while in other cases, there was coexistence between the dilute and condensed phases. For all the systems, the cluster size increased with concentration, mainly due to more material being available to condense while the density of the dense phase remained the same (Figure S7). Condensates of FUS LCD in simulations with the Tesei et al. 2021²⁹ and Dignon et al. 2018 models²³ were larger and denser than the clusters in our model (Figure S7). The smaller condensate size with COCOMO is a result of a fraction of proteins remaining in the dilute phase, whereas the higher density in the other models may suggest overpacking of the chains during condensate formation.

We further tested our model by constructing phase diagrams as a function of temperature. Starting from a box with FUS LCD at 0.26 mM, we simulated the system at temperatures ranging from 260 to 310 K. A temperature-concentration phase diagram was constructed based on the densities in the dilute and condensed phases (Figure 4). We obtained very good agreement with the experimental coexistence densities reported for FUS LCD.⁶² On the other hand, results for Ddx4 indicate that COCOMO may result in overpacking in the dense phase with respect to the experimental values (Figure 4).⁶³ We note that, Regy et al.²⁸ were able to

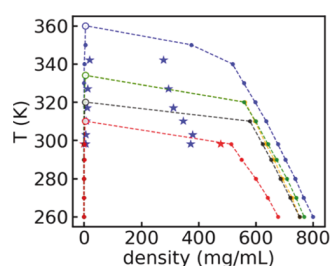


Figure 4. Temperature phase diagram. Simulation densities at different temperature (T) for a system starting at 4.6 mg/mL FUS LCD (red), hTau40-k18 (black), LAF-1^{RGG} (green), A1 LCD (orange), and at 5 mg/mL for Ddx4 (blue). Open circles indicate the lowest temperature where no PS was observed in the set, corresponding to T_c . Note that LAF-1^{RGG} (green) and A1 LCD (orange) traces are almost completely overlapped. The data shown are the averaged values over the last 4 μ s of the trajectory. Experimental coexistence densities for FUS LCD⁶² and Ddx4⁶³ are shown using star symbols.

reproduce experimental densities for this protein, and it will require further investigation to what extent correct packing can be reproduced accurately with a residue-based CG model.

Recently, Tejedor et al.⁶⁴ described a positive correlation between the critical temperature, T_c , above which no LLPS is observed with the experimental saturation concentration for various phase-separating proteins, using the model developed by Dignon et al. (used in this work for comparison as well) with an additional reparameterization to include cation- π interactions.³¹ With COCOMO, we also find such a correlation among the proteins used in the homotypic PS studies (Figure 4). The experimental protein saturation

thresholds reported are 0.013, 0.024, 0.050, and 0.24 mM for A1 LCD, LAF-1^{RGG}, hTau40-k18, and FUS LCD, respectively.^{58–61} The order of the experimental concentration thresholds (FUS LCD > hTau40-k18 > LAF-1^{RGG} > A1 LCD) matched the order of critical temperatures, as evident from the phase diagrams (Figure 4). We note that the phase diagrams and T_c values were very similar between LAF-1^{RGG} and A1 LCD. This may be expected since the experimental saturation concentration values of these two proteins are the lowest and close to each other.

We observed cluster densities between 500 and 800 g/L depending on the system. Water was not included explicitly in the COCOMO model, but from the densities, one can estimate the water fraction of 50–20% since water has a density of 1000 mg/mL. We note that for such systems, a wide range of water content is reported in experimental studies.^{65,66}

Protein Heterotypic Phase Separation. We further evaluated COCOMO with systems containing more than one protein, as those systems can also lead to PS.^{58,67,68} We focused on heterotypic protein PS of FUS LCD at increasing amounts of [RGRGG]₅ and FUS LCD^{RGG3} peptides because these proteins have been well studied both experimentally and computationally.^{17,28,50,58,67} Details about the simulated systems for heterotypic PS can be found in Table S7. Results shown in Figure 5 indicate that COCOMO reproduces concentration-dependent PS upon addition of the peptides. As for homotypic PS, the models from Tesei et al. 2021²⁹ and Dignon et al. 2018²³ also resulted in condensation and are independent of the FUS LCD/peptide ratio. No condensation was observed with the models of Regy et al. 2021²⁸ and Dannenhoffer et al. 2021²⁷ at any concentration (Figures 5 and S8). In all cases where condensates were observed, an increase

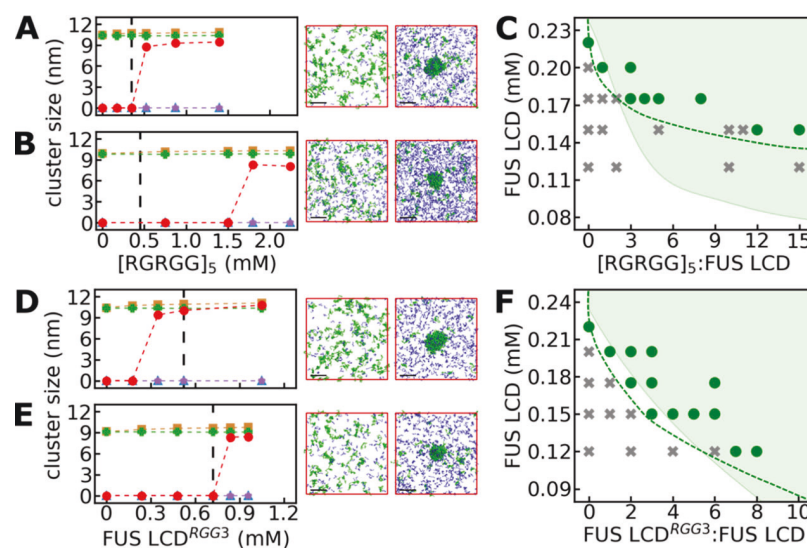


Figure 5. Protein heterotypic phase separation. Results are shown for FUS LCD 0.175 mM (A) and FUS LCD 0.150 mM (B) at increasing concentrations of the [RGRGG]₅ peptide and FUS LCD 0.175 mM (C) and FUS LCD 0.120 mM (D) at increasing concentrations of the FUS LCD^{RGG3} peptide. Results obtained with COCOMO (red circle) are compared with other models: Regy et al. 2021²⁸ (blue triangle), Tesei et al. 2021²⁹ (tan square), Dignon et al. 2018²³ (green filled plus), and Dannenhoffer et al. 2021²⁷ (purple star). Cluster sizes were averaged over the last 4 μ s of the simulations. The experimental LLPS concentration threshold for the addition of the peptide is shown as a dashed line.⁵⁸ Final frames of the simulations with COCOMO are shown for the lowest (left) and highest (right) concentrations added of peptide in each case. FUS LCD and peptide chains were colored in green and blue, respectively. The size bar represents 20 nm. A two-dimensional concentration-dependent phase diagram for FUS LCD as a function of different concentration ratios of [RGRGG]₅ (C) or FUS LCD^{RGG3} (F) is shown based on COCOMO simulations. Markers show the presence (green circles) or absence (gray crosses) of condensates during simulation. The shaded region is the experimental PS regime estimated from experimental phase diagrams.⁵⁸ The dashed line is an aid to the eye for the PS regime boundary in our model.

in peptide concentration was accompanied by faster cluster growth and larger final clusters (Figures S9 and S10).

Multicomponent phase diagrams as a function of concentration also show overall good agreement with experimental data (Figure 5C,F). Qualitatively, the main features of the experimental phase diagrams are reproduced well. Quantitative agreement is also very good for the FUS LCD/FUS LCD^{RGG3} system (Figure 5F), but for the FUS LCD: [RGRGG]₅ system, the model predicts a shift to larger minimum FUS LCD concentrations compared to the experimental data (Figure 5C, shaded area vs dashed line). Finally, the morphology of the coacervates in the simulations is also generally in agreement with the experiments, as they are composed of both FUS LCD and the peptide (Figures 5 and S9). We found that condensates are enriched in FUS LCD over the peptide by about tenfold, but there is no experimental data to validate the model prediction. Otherwise, our results reproduce the experimental observation that peptides enhance PS of FUS LCD.⁵⁸

Protein–RNA Phase Separation. Finally, we turn to protein–RNA condensation. Experimental evidence has shown that RNA can modulate the stability^{50,69,70} and kinetic properties^{19,59} of protein condensates, and protein–RNA condensation is receiving increasing attention. Therefore, another goal of our model was to describe PS in systems including proteins and nucleic acids. COCOMO was successful in capturing the phase behavior of different protein–RNA mixtures for which LLPS has been described experimentally (Figure 6 and Table S9). Condensates were

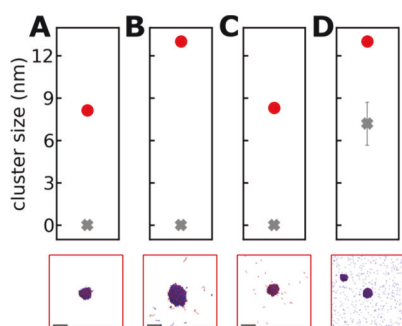


Figure 6. RNA–protein phase separation. Results are shown for using polyAde-21–(RRLR)₆–SSSGSS (A), polyUra-40–FUS LCD^{RGG3} (B), polyUra-10–polyArg-50 (C), and polyAde-500–(RGRGG)₅ (D) simulated at concentrations where PS was observed experimentally^{50,51,58,71} (see details in Table S9). Results with COCOMO (red circle) are compared with simulations using the Regy et al. 2020³³ (gray cross) model. Cluster sizes were averaged over the last 4 μ s of the simulation. The final frames of simulations using COCOMO are shown on the lower panels for each system. RNA and protein chains are colored in red and blue, respectively, and the size bar represents 20 nm.

composed of both proteins and RNA and were observed in all systems that we studied. For comparison, we simulated the same systems using the model by Regy et al. 2020.³³ Using this model, no PS was found except for polyAde-500–(RGRGG)₅, where transient cluster formation was observed without clear condensation into larger clusters (Figures 6, S11 and S12).

We analyzed the diffusion inside the clusters and the time scales on which peptides or RNA exchange with the environment (Table S10). In all cases, diffusion inside the clusters is retarded, as expected, but liquid-like behavior

appears to be retained. Generally, polymers remain in the condensates on μ s time scales before exchanging with the dilute phase, and the systems with longer peptides/RNA displayed slower diffusion and larger residency times inside the condensates. However, we note that due to the CG nature of our model, the resulting dynamics is only qualitatively meaningful and likely too fast since hydrodynamics is omitted and molecular friction is underestimated.

We note that the choice of the value for the $\epsilon_{R/K-nucleic}$ parameter affects the agreement with the experimental data.^{50,51,58,71} Previous studies showed contrasting views on this term, Das et al.³¹ argued that despite augmenting $\epsilon_{R/K-nucleic}$ parameter, their model failed to capture the experimental PS propensity, while other authors⁶⁴ demonstrated that increased cation– π interactions can reproduce the experimental trend of their selected set of proteins. Here, we find that with this interaction added explicitly, we can describe PS between proteins and RNA in agreement with experiments.

CONCLUSIONS

In this work, we present a simple coarse-grained model that can balance the polymer properties of disordered proteins and RNA with their propensity to phase separate. Our model differs from the other four used for comparison^{23,27–29} in various key aspects, such as the presence of an angle term to account for protein chain stiffness. Also, we describe short-range interactions with a 5–10 Lennard-Jones potential,⁷² while the other potentials compared here use an Ashbaugh–Hatch modified Lennard-Jones potential,³⁷ with different optimized scales, to describe the hydrophobicity of protein/RNA residues/nucleotides. For long-range interactions, instead of using a regular Debye–Hückel potential, we used a modification that allows the implicit description of solvation effects along with salt-screened electrostatic interactions. Finally, we added a term for cation– π interactions, as in agreement with previous studies,²⁶ this interaction is an important aspect to consider in modeling LLPS systems.

Our model reproduces experimental R_g values for peptides well and captures the dependency of RNA polymer properties on salt concentration. At the same time, the model can describe homotypic and heterotypic protein PS as well as PS involving protein–RNA mixtures in agreement with various in vitro experimental systems. Other models can also accurately reproduce R_g values for different peptide sequences, and some of those models can also describe PS. However, in those models, concentration-dependent PS involves significantly lower temperatures or higher concentrations than what is reported in the experiments. Our results showed that only the Regy et al. 2021²⁸ and Tesei et al. 2021²⁹ models may describe the concentration-dependent PS and only in some systems. However, it appears that these models continue to be improved.⁷³ Other improvements in angle and dihedral terms have been proposed for the Regy et al. model to more accurately reproduce structural features like helical propensity of IDPs,⁷⁴ but it is unclear how this may affect the prediction of PS. The key strength of the model here is that experimental behavior is reproduced at the level of single polymers as well as for condensation at the same temperature and in a concentration-dependent manner that matches experiments. Therefore, the model is more predictive with respect to when and under what conditions PS would be expected for systems for which experimental data is not available.

Even though the COCOMO CG model presented here performs well when compared to experimental observations, there are also significant limitations. The implicit description of salt effects based on the Debye–Hückel formalism is only approximate and does not capture changes in free energy of the ions themselves as they condense along with the biopolymers. COCOMO does not fully capture the entropic contributions given by counterions a translational entropy,⁷⁵ association and dissociation,⁷⁶ and their partitioning among dilute and condensed phases⁷⁷ that are important in LLPS. In addition, the residue-level approximation neglects any shape-dependent packing interactions during condensation and, more generally, anisotropic or directional interactions, for example, due to aromatic stacking or hydrogen bonding. Although this could be addressed using higher-resolution models,^{78–80} they are computationally more expensive to cover the range of concentrations and scales of LLPS systems that we can describe with COCOMO. The next steps to further improve the model would be to convert the condensate models generated by the COCOMO to higher-resolution representations and test their physical viability via simulation. For example, atomistic simulations of reconstructed condensates may inform how to improve the CG model without increasing the computational complexity.

A physically realistic and computationally efficient CG model that is predictive and can reach the time and spatial scales on which biomolecular condensation is observed experimentally opens up a wide range of applications. Despite much experimental information obtained so far about biomolecular condensates, many questions about how condensation depends on molecular compositions remain unclear. The model here allows such questions to be explored not just for homo- and heterotypic peptide systems but also for peptide-RNA mixtures. We expect that there will be many applications where such simulations can interpret experimental observations and suggest new hypotheses to be tested experimentally. Furthermore, we think that the COCOMO model is predictive for a wide variety of peptide sequences as long as they remain largely disordered. Our model has not been designed at this point to describe systems with more structured peptides or folded proteins. However, a possibility is to follow a similar strategy, as described in the Rizuan et al.⁷⁴ paper, or adding elastic internal restraints to preserve the fold, respectively. This remains an interesting direction for future work.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00856>.

Residue-specific parameters used in COCOMO; sequences of disordered or unfolded proteins and radii of gyration of proteins in the parametrization set; sequences of disordered proteins and radii of gyration in the test set; RNA experimental properties; homotypic and heterotypic protein phase separation systems composition; protein – RNA phase separation systems composition; dynamics of condensed and diluted phases in different systems; protein Rg sensitivity; RNA polymer properties sensitivity; representative snapshots for homotypic protein systems simulated with different models; cluster formation during homotypic phase

separation; cluster size distributions in homotypic protein systems; Reversible phase separation in FUS LCD with COCOMO; density profiles; representative snapshots for heterotypic LLPS in different systems; cluster formation in heterotypic phase separation; cluster size distributions in heterotypic phase separation; representative conformations of different RNA-protein systems; cluster characterization in RNA – protein phase separation; and FUS LCD phase separation using the COCOMO potential using the switching function (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michael Feig – Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0001-9380-6422; Phone: +1-517-432-7439; Email: mfeiglab@gmail.com

Authors

Gilberto Valdes-Garcia – Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States

Lim Heo – Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0002-3153-2363

Lisa J. Lapidus – Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0003-3136-2680

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jctc.2c00856>

Funding

Funding was provided by the National Science Foundation grants MCB 1817307 and by the National Institute of Health (NIGMS) grant R35 GM126948. We used computational resources at the Institute for Cyber-Enabled Research/High Performance Computing Cluster (ICER/HPCC) at Michigan State University.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, 357, No. eaaf4382.
- (2) Pederson, T. The nucleolus. *Cold Spring Harbor Perspect. Biol.* **2011**, 3, a000638.
- (3) Mao, Y. S.; Zhang, B.; Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **2011**, 27, 295–306.
- (4) Gall, J. G. Cajal bodies: The first 100 years. *Annu. Rev. Cell Dev. Biol.* **2000**, 16, 273.
- (5) Gall, J. G. The centennial of the Cajal body. *Nat. Rev. Mol. Cell Biol.* **2003**, 4, 975–980.
- (6) Lamond, A. I.; Spector, D. L. Nuclear speckles: A model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **2003**, 4, 605–612.
- (7) Galganski, L.; Urbanek, M. O.; Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* **2017**, 45, 10350–10368.
- (8) Buchan, J. R.; Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **2009**, 36, 932–941.
- (9) Protter, D. S. W.; Parker, R. Principles and properties of stress granules. *Trends Cell Biol.* **2016**, 26, 668–679.
- (10) Luo, Y.; Na, Z.; Slavoff, S. A. P-Bodies: composition, properties, and functions. *Biochemistry* **2018**, 57, 2424–2431.

- (11) Voronina, E.; Seydoux, G.; Sassone-Corsi, P.; Nagamori, I. RNA granules in germ cells. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a002774.
- (12) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 215–235.
- (13) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285–298.
- (14) Mitrea, D. M.; Chandra, B.; Ferrolino, M. C.; Gibbs, E. B.; Tolbert, M.; White, M. R.; Kriwacki, R. W. Methods for physical characterization of phase-separated bodies and membrane-less organelles. *J. Mol. Biol.* **2018**, *430*, 4773–4805.
- (15) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **2015**, *11*, 899–904.
- (16) Posey, A. E.; Holehouse, A. S.; Pappu, R. V. Phase separation of intrinsically disordered proteins. *Methods Enzymol.* **2018**, *611*, 1–30.
- (17) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular details of protein condensates probed by microsecond long atomistic simulations. *J. Phys. Chem. B* **2020**, *124*, 11671–11679.
- (18) Paloni, M.; Bailly, R.; Ciandrini, L.; Barducci, A. Unraveling molecular interactions in liquid-liquid phase separation of disordered proteins by atomistic simulations. *J. Phys. Chem. B* **2020**, *124*, 9009–9016.
- (19) Wei, M. T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **2017**, *9*, 1118–1125.
- (20) Feig, M.; Chocholoušová, J.; Tanizaki, S. Extending the horizon: Towards the efficient modeling of large biomolecular complexes in atomic detail. *Theor. Chem. Acc.* **2006**, *116*, 194–205.
- (21) Predeus, A. V.; Gul, S.; Gopal, S. M.; Feig, M. Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. *J. Phys. Chem. B* **2012**, *116*, 8610–8620.
- (22) Kar, P.; Feig, M. Recent advances in transferable coarse-grained modeling of proteins. *Adv. Prot. Chem. Struct. Biol.* **2014**, *96*, 143–180.
- (23) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *Plos Comp. Biol.* **2018**, *14*, No. e1005941.
- (24) Mani, E.; Lechner, W.; Kegel, W. K.; Bolhuis, P. G. Equilibrium and non-equilibrium cluster phases in colloids with competing interactions. *Soft Matter* **2014**, *10*, 4479–4486.
- (25) Dutagaci, B.; Nawrocki, G.; Goodluck, J.; Ashkarran, A. A.; Hoogstraten, C. G.; Lapidus, L. J.; Feig, M. Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. *eLife* **2021**, *10*, No. e64004.
- (26) Joseph, J. A.; Espinosa, J. R.; Sanchez-Burgos, I.; Garaizar, A.; Frenkel, D.; Collepardo-Guevara, R. Thermodynamics and kinetics of phase separation of protein-RNA mixtures by a minimal model. *Biophys. J.* **2021**, *120*, 1219–1230.
- (27) Dannenhoffer-Lafage, T.; Best, R. B. A data-driven hydrophobicity scale for predicting liquid-liquid phase separation of proteins. *J. Phys. Chem. B* **2021**, *125*, 4046–4056.
- (28) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Sci.* **2021**, *30*, 1371–1379.
- (29) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2111696118.
- (30) Latham, A. P.; Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *J. Chem. Theory Comput.* **2020**, *16*, 773–781.
- (31) Das, S.; Lin, Y. H.; Vernon, R. M.; Forman-Kay, J. D.; Chan, H. S. Comparative roles of charge, pi, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 28795–28805.
- (32) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat. Comput. Sci.* **2021**, *1*, 732–743.
- (33) Regy, R. M.; Dignon, G. L.; Zheng, W. W.; Kim, Y. C.; Mittal, J. Sequence dependent phase separation of protein-polynucleotide mixtures elucidated using molecular simulations. *Nucleic Acids Res.* **2020**, *48*, 12593–12603.
- (34) Wu, H.; Wolynes, P. G.; Papoian, G. A. A coarse-grained force field for intrinsically disordered proteins. *J. Phys. Chem. B* **2018**, *122*, 11115–11125.
- (35) Holmstrom, E. D.; Liu, Z.; Nettels, D.; Best, R. B.; Schuler, B. Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. *Nat. Commun.* **2019**, *10*, 2453.
- (36) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (37) Ashbaugh, H. S.; Hatch, H. W. Natively unfolded protein stability as a coil-to-globule transition in charge/hydrophobicity space. *J. Am. Chem. Soc.* **2008**, *130*, 9536–9542.
- (38) Farahi, N.; Lazar, T.; Wodak, S. J.; Tompa, P.; Pancsa, R. Integration of data from liquid-liquid phase separation databases highlights concentration and dosage sensitivity of LLPS drivers. *Int. J. Mol. Sci.* **2021**, *22*, 22.
- (39) Ghobadi, A. F.; Jayaraman, A. Effect of backbone chemistry on hybridization thermodynamics of oligonucleic acids: a coarse-grained molecular dynamics simulation study. *Soft Matter* **2016**, *12*, 2276–2287.
- (40) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Comp. Biol.* **2017**, *13*, No. e1005659.
- (41) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667–683.
- (42) Feig, M.; Karanickolas, J.; Brooks, C. L., III MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Modell.* **2004**, *22*, 377–395.
- (43) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (44) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.
- (45) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: modeling conformational dynamics at the picosecond to millisecond scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (46) Schrödinger, L. L. C. The PyMOL Molecular Graphics System, 2015. Version 1.8
- (47) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 16155–16160.
- (48) Waldauer, S. A.; Bakajin, O.; Lapidus, L. J. Extremely slow intramolecular diffusion in unfolded protein L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13713.
- (49) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; et al. Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc.* **2012**, *134*, 12565–12577.

- (50) Alshareedah, I.; Kaur, T.; Ngo, J.; Seppala, H.; Kounatse, L. D.; Wang, W.; Moosa, M. M.; Banerjee, P. R. Interplay between short-range attraction and long-range repulsion controls reentrant liquid condensation of ribonucleoprotein-RNA complexes. *J. Am. Chem. Soc.* **2019**, *141*, 14593–14602.
- (51) Fisher, R. S.; Elbaum-Garfinkle, S. Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nat. Commun.* **2020**, *11*, 4628.
- (52) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **2020**, *71*, 53–75.
- (53) Paloni, M.; Bussi, G.; Barducci, A. Arginine multivalency stabilizes protein/RNA condensates. *Protein Sci.* **2021**, *30*, 1418–1426.
- (54) Ukmar-Godec, T.; Hutten, S.; Grieshop, M. P.; Rezaei-Ghaleh, N.; Cima-Omori, M. S.; Biernat, J.; Mandelkow, E.; Söding, J.; Dormann, D.; Zweckstetter, M. Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* **2019**, *10*, 2909.
- (55) Plumridge, A.; Andresen, K.; Pollack, L. Visualizing disordered single-stranded RNA: connecting sequence, structure, and electrostatics. *J. Am. Chem. Soc.* **2020**, *142*, 109–119.
- (56) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66.
- (57) Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 799–804.
- (58) Kaur, T.; Raju, M.; Alshareedah, I.; Davis, R. B.; Potoyan, D. A.; Banerjee, P. R. Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies. *Nat. Commun.* **2021**, *12*, 872.
- (59) Elbaum-Garfinkle, S.; Kim, Y.; Szczepaniak, K.; Chen, C. C. H.; Eckmann, C. R.; Myong, S.; Brangwynne, C. P. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 7189–7194.
- (60) Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* **2017**, *8*, 275.
- (61) Bremer, A.; Farag, M.; Borchers, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, M. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **2022**, *14*, 196–207.
- (62) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **2019**, *26*, 637–648.
- (63) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y. H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E8194–E8203.
- (64) Tejedor, A. R.; Garaizar, A.; Ramírez, J.; Espinosa, J. R. RNA modulation of transport properties and stability in phase-separated condensates. *Biophys. J.* **2021**, *120*, 5169–5186.
- (65) Ahlers, J.; Adams, E. M.; Bader, V.; Pezzotti, S.; Winkhofer, K. F.; Tatzelt, J.; Havenith, M. The key role of solvent in condensation: Mapping water in liquid-liquid phase-separated FUS. *Biophys. J.* **2021**, *120*, 1266–1275.
- (66) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **2015**, *60*, 231–241.
- (67) Boeynaems, S.; Bogaert, E.; Kovacs, D.; Konijnenberg, A.; Timmerman, E.; Volkov, A.; Guharoy, M.; De Decker, M.; Jaspers, T.; Ryan, V. H.; et al. Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* **2017**, *65*, 1044–1055 e1045.
- (68) Sanders, D. W.; Kedersha, N.; Lee, D. S. W.; Strom, A. R.; Drake, V.; Riback, J. A.; Bracha, D.; Eeftens, J. M.; Iwanicki, A.; Wang, A.; et al. Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* **2020**, *181*, 306–324.
- (69) Schwartz, J. C.; Wang, X.; Podell, E. R.; Cech, T. R. RNA seeds higher-order assembly of FUS protein. *Cell Rep* **2013**, *5*, 918–925.
- (70) Banerjee, P. R.; Milin, A. N.; Moosa, M. M.; Onuchic, P. L.; Deniz, A. A. Reentrant phase transition drives dynamic substructure formation in ribonucleoprotein droplets. *Angew. Chemie Int. Ed.* **2017**, *56*, 11354–11359.
- (71) Bai, Q.; Zhang, Q.; Jing, H.; Chen, J.; Liang, D. Liquid-liquid phase separation of peptide/oligonucleotide complexes in crowded macromolecular media. *J. Phys. Chem. B* **2021**, *125*, 49–57.
- (72) Lennard-Jones, J. E.; Cohesion, J. E. Cohesion. *Proc. Phys. Soc* **1931**, *43*, 461–482.
- (73) Tesei, G.; Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe* **2022**, *2*, 94.
- (74) Rizuan, A.; Jovic, N.; Phan, T. M.; Kim, Y. C.; Mittal, J. Developing bonded potentials for a coarse-grained model of intrinsically disordered proteins. *J. Chem. Inf. Model.* **2022**, *62*, 4474–4485.
- (75) Khokhlov, A.; Nyrkova, I. Compatibility enhancement and microdomain structuring in weakly charged polyelectrolyte mixtures. *Macromolecules* **1992**, *25*, 1493–1502.
- (76) Salehi, A.; Larson, R. G. A molecular thermodynamic model of complexation in mixtures of oppositely charged polyelectrolytes with explicit account of charge association/dissociation. *Macromolecules* **2016**, *49*, 9706–9719.
- (77) Vis, M.; Peters, V. F.; Ern , B. H.; Tromp, R. H. Ion entropy in phase-separated aqueous mixtures of polyelectrolyte and neutral polymer. *Macromolecules* **2015**, *48*, 2819–2828.
- (78) Tsanai, M.; Frederix, P. W.; Schroer, C. F.; Souza, P. C.; Marrink, S. J. Coacervate formation studied by explicit solvent coarse-grain molecular dynamics with the Martini model. *Chem. Sci.* **2021**, *12*, 8521–8530.
- (79) Liu, X.; Chen, J. HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32421–32432.
- (80) Zhang, Y.; Liu, X.; Chen, J. Toward accurate coarse-grained simulations of disordered proteins and their dynamic interactions. *J. Chem. Inf. Model.* **2022**, *62*, 4523–4536.