

# Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model

Zakarya Benayad, Sören von Bülow, Lukas S. Stelzl, and Gerhard Hummer\*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 525–537



Read Online

ACCESS |



Metrics & More

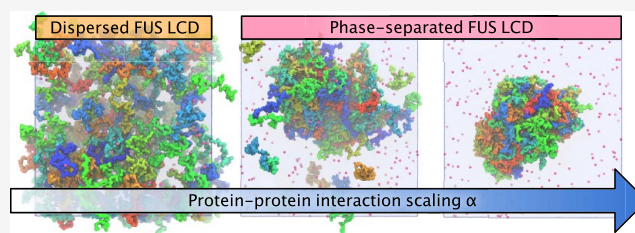


Article Recommendations



Supporting Information

**ABSTRACT:** Disordered proteins and nucleic acids can condense into droplets that resemble the membraneless organelles observed in living cells. MD simulations offer a unique tool to characterize the molecular interactions governing the formation of these biomolecular condensates, their physicochemical properties, and the factors controlling their composition and size. However, biopolymer condensation depends sensitively on the balance between different energetic and entropic contributions. Here, we develop a general strategy to fine-tune the potential energy function for molecular dynamics simulations of biopolymer phase separation. We rebalance protein–protein interactions against solvation and entropic contributions to match the excess free energy of transferring proteins between dilute solution and condensate. We illustrate this formalism by simulating liquid droplet formation of the FUS low-complexity domain (LCD) with a rebalanced MARTINI model. By scaling the strength of the nonbonded interactions in the coarse-grained MARTINI potential energy function, we map out a phase diagram in the plane of protein concentration and interaction strength. Above a critical scaling factor of  $\alpha_c \approx 0.6$ , FUS-LCD condensation is observed, where  $\alpha = 1$  and  $0$  correspond to full and repulsive interactions in the MARTINI model. For a scaling factor  $\alpha = 0.65$ , we recover experimental densities of the dilute and dense phases, and thus the excess protein transfer free energy into the droplet and the saturation concentration where FUS-LCD condenses. In the region of phase separation, we simulate FUS-LCD droplets of four different sizes in stable equilibrium with the dilute phase and slabs of condensed FUS-LCD for tens of microseconds, and over one millisecond in aggregate. We determine surface tensions in the range of 0.01–0.4 mN/m from the fluctuations of the droplet shape and from the capillary-wave-like broadening of the interface between the two phases. From the dynamics of the protein end-to-end distance, we estimate shear viscosities from 0.001 to 0.02 Pa s for the FUS-LCD droplets with scaling factors  $\alpha$  in the range of 0.625–0.75, where we observe liquid droplets. Significant hydration of the interior of the droplets keeps the proteins mobile and the droplets fluid.



## INTRODUCTION

Intracellular compartmentalization into organellar structures is crucial for the organization of cellular biochemistry in time and space. Cell nuclei and mitochondria are examples of subcellular compartments bounded by a lipid membrane, whereas stress granules, processing bodies, nucleoli, or Cajal bodies are membraneless.<sup>1–6</sup> Disordered proteins and nucleic acids can cluster to form biomolecular condensates with liquid-like properties.<sup>1,2,7,8</sup> Such condensates formed by liquid–liquid phase separation (LLPS) in vitro mimic the membraneless organelles in cells.<sup>1,2</sup>

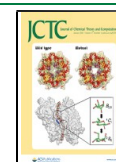
Individually weak but multivalent interactions between intrinsically disordered proteins (IDPs), in some cases amplified by condensing factors, are major drivers of LLPS.<sup>4,8</sup> The biomolecular condensates produced by LLPS behave as liquid droplets immersed in dilute solution.<sup>7</sup> Their liquid-like interior facilitates the rapid diffusion of reactants within the condensates and their exchange with the outside.<sup>9</sup> Their dynamic nature also makes biomolecular condensates a promising template for novel biomimetic materials.<sup>10,11</sup> The rational design of new materials will benefit from predictive

models that relate the static and dynamic materials properties of biomolecular condensates to the protein and nucleic acid sequences<sup>10</sup> and the molecular interactions they encode.

The RNA-binding protein fused in sarcoma (FUS) contains a low-complexity domain (LCD) and is implicated in the formation of membraneless organelles.<sup>12</sup> The FUS-LCD has served as paradigm to understand biomolecular condensates.<sup>13,14</sup> Enriched in the nucleus, FUS participates in transcription, DNA repair, and RNA biogenesis.<sup>15</sup> It has also been shown that disease-associated mutations lead to the liquid-to-solid transition of FUS droplets through the formation of fibrous amyloid-like assemblies.<sup>9,16</sup>

Received: October 9, 2020

Published: December 14, 2020



Theory and simulations<sup>17–25</sup> contribute to the emerging understanding of LLPS. Molecular dynamics (MD) simulations at atomic resolution could, in principle, monitor molecular interactions with a high accuracy. While high computational costs still preclude their widespread use for simulating LLPS, they promise considerable insights into structures and interactions of biomolecules in (subsystems of) phase-separated condensates<sup>26–29</sup> and how these give rise to materials properties.<sup>30</sup> Soft-matter approaches<sup>31</sup> provide access to the large lengthscales and long timescales relevant for phase separation, but may not resolve the detailed effects of protein chemistry and may not be transferable. Overcoming these two issues would require transferable models.<sup>20,21</sup> Coarse-grained simulations using a reduced representation<sup>32</sup> enable direct simulations of phase behavior<sup>21,24,33,34</sup> and can give a good description of effective molecular interactions, which is important for transferability. Dignon et al.<sup>21</sup> studied the phase separation of the low-complexity domain of FUS using a coarse-grained protein model with an implicit solvent. They built the phase diagram in the temperature–concentration space and demonstrated the importance of specific molecular interactions between the IDPs. The construction of system-specific coarse-grained force fields with machine learning offers another alternative to study the phase behavior of a protein of interest.<sup>35</sup>

The parameterization of coarse-grained simulation models for quantitative studies of phase separation is challenging. In the case of proteins in solution, intraprotein, protein–protein, protein–solvent, and solvent–solvent interactions have to be balanced with each other and with configurational and solvation entropy contributions associated with the degrees of freedom integrated out in the coarse-grained representation. Systematic errors in representing molecular interactions are often extensive, i.e., their contributions to the excess free energy difference per molecule between the phases tend to grow linearly with chain length. As a consequence, even small systematic imbalances become amplified. Incidentally, similar challenges<sup>36–38</sup> are faced in simulations of protein self-assembly using highly optimized all-atom force fields. These classical force fields have been parameterized primarily using quantum mechanical data and, in this sense, are also coarse-grained. Indeed, tuning the helix-coil equilibrium of proteins<sup>36</sup> against experimental data turned out to be a critical improvement for simulations of de novo protein folding.<sup>39</sup> For disordered proteins, rebalancing of nonbonded and solvent interactions addressed the issue of compactness in disordered proteins.<sup>37,38</sup>

Here, we adopt this strategy to develop and implement a general approach to fine-tune MD simulation models for studies of biopolymer phase separation. In a first step, we adjust the strength of the nonbonded interactions between the droplet-forming biopolymers to reproduce the experimental excess transfer free energy between the dilute and dense (droplet) phase. Reproducing the experimental densities of both the dilute and the dense phase gives us a first validation test, and an indication that the model may provide a reasonable description also of the molecular structures in the two phases. In a second step, we use the rebalanced model to calculate structural, thermodynamic and dynamic materials properties and to compare them to experiments where possible. We demonstrate this general procedure in simulations of the phase behavior of the disordered protein FUS-LCD with

the MARTINI model,<sup>40</sup> a widely used coarse-grained force field.

We first map the phase diagram in the plane spanned by the protein concentration  $c$  and a parameter  $\alpha$  introduced previously to scale the protein–protein Lennard-Jones (LJ) interaction energy.<sup>41,42</sup> Having identified the region where the FUS-LCD undergoes spontaneous phase separation, we simulate stable FUS droplets at different protein–protein interaction strengths  $\alpha$ . From these simulations, we construct the coexistence line in the  $\alpha$ – $c$  plane as the relationship between  $\alpha$  and the protein concentrations  $c_{\text{dense}}$  and  $c_{\text{dilute}}$  in the droplet and in the coexisting dilute phase, respectively. Using the coexistence data, we determine the value of  $\alpha$  for which the measured excess transfer free energy as well as the two densities are reproduced. We then quantify three biophysical properties: the hydration level of the droplets, their surface tension and their shear viscosity. The strong dependence of these structural, thermodynamic and dynamic properties on  $\alpha$  demonstrates the importance of force field rebalancing to describe biologically relevant behavior.

Our approach of rebalancing MD simulation force fields using experimental information on LLPS is not limited to a particular type of coarse-graining,<sup>32</sup> but is generally applicable. Instead of or in addition to scaling protein–protein interactions, also other aspects of the potential energy surface can be adjusted. Force field rebalancing to match phase boundaries and excess transfer free energies should improve the reliability in molecular simulations of biopolymer phase separation.

## ■ SIMULATION METHODS

**Protein–Protein Interaction Described by Scaled MARTINI 2.2 Force Field.** A variant of the MARTINI 2.2 force field<sup>43</sup> was used for all MD simulations, in which we rescaled the protein–protein LJ interactions following Stark et al.<sup>41</sup> They introduced a parameter  $\alpha$  to scale the protein–protein LJ pair interaction well depth,  $\epsilon_{\alpha} = \epsilon_0 + \alpha(\epsilon_{\text{original}} - \epsilon_0)$ . A value of  $\alpha = 0$  corresponds to a repulsion-dominated interaction in the MARTINI model,<sup>43</sup>  $\epsilon_0 = 2$  kJ/mol, and a value of  $\alpha = 1$  recovers the full interaction in the MARTINI force field,  $\epsilon_1 = \epsilon_{\text{original}}$ . Adopting a common rebalancing strategy for nonspecific protein–protein interactions,<sup>44,45</sup> Stark et al.<sup>41</sup> adjusted  $\alpha$  to match experimentally determined protein–protein virial coefficients, which also gave promising results for carbohydrates.<sup>46</sup> A similar rebalancing of water–protein interactions, starting from the MARTINI 3 force field, was found to describe the structure and dynamics of proteins with ordered and disordered domains well.<sup>47,48</sup> In the spirit of the MARTINI force field development,<sup>40</sup> here we attempt to match the experimentally determined excess transfer free energy for a protein between the dilute phase and the dense phase of condensed droplets. In accordance with the method of Stark et al.,<sup>41</sup> interactions involving the beads P4, Qa, and Qd were not scaled, because they also describe the water and ion beads in the system.

**Simulations to Map the Phase Boundaries in the  $\alpha$ – $c$  Plane.** In a first exploratory step, we qualitatively mapped the phase boundaries in the plane spanned by the protein concentration  $c$  and the interaction strength  $\alpha$  using MD simulations. We initiated these simulations from homogeneous solutions of protein chains in a cubic box with periodic boundary conditions. All simulations were performed with the GROMACS<sup>49,50</sup> v2018.6 software. In the simulations, we

varied the protein concentration and the interaction strength  $\alpha$ . The FUS-LCD is 163 amino acids long. Its amino acid sequence is listed in Table S1. We built an initial atomistic model of the FUS-LCD in an extended conformation using the AmberTools Leap program,<sup>51</sup> which we then coarse-grained using the martinize.py code.<sup>43,52</sup> We built initial simulation systems by randomly inserting copies of the coarse-grained model into the simulation box and adding water beads. Sodium ions were added to ensure electro-neutrality. We replaced 10% of the water beads with “antifreeze” beads to prevent nonphysical solvent freezing. In an initial equilibration, we used repulsion-dominated protein–protein interactions ( $\alpha = 0$ ) to disperse the proteins in the box. We performed a first MD simulation of 2 ns in the NVT ensemble, using the velocity scaling thermostat<sup>53</sup> to establish a temperature of 300 K. A second equilibration step was conducted for 20 ns in the NPT ensemble at a pressure of 1 bar, maintained by an isotropic Berendsen barostat.<sup>54</sup> Production simulations were then started from the final conformation with dispersed FUS-LCD. In the production simulations,  $\alpha$  was set to the desired value between zero and one. At each  $\alpha$  value, we performed simulations of 12  $\mu$ s (see Table S2), using a time step of 30 fs. Temperature and pressure were maintained at 300 K and 1 bar, respectively, by a velocity scaling thermostat<sup>53</sup> and a Parrinello–Rahman barostat.<sup>55</sup> The “New-RF” MD settings as described in ref 56 were used with short-range LJ and electrostatics cutoffs of 1.1 nm. The trajectories were visually inspected using the VMD<sup>57</sup> v1.9.3 software.

**Cluster Formation.** We examined the MD trajectories for possible phase separation by monitoring the formation of protein clusters. Protein clusters were identified on the basis of a pair-distance criterion. In a given simulation structure, two proteins were considered to be in the same cluster if any pair of beads was within a cutoff distance of 5 Å. We monitored the FUS-LCD condensation process by following the number and size of protein clusters along the trajectory.

**Simulation of Droplets.** FUS-LCD droplets were simulated over several microseconds by starting from a preformed droplet. Droplet formation kinetics directly from homogeneous solutions can be severely slowed down by periodic boundaries, which favor percolation in sufficiently dense systems. To avoid percolation, we reduced the protein concentration by increasing the box size and adding solvent. To simulate droplets in equilibrium with the dilute phase over several microseconds, we hence started from a droplet formed at  $\alpha = 0.7$  (see Table S2 for details). We then extended the trajectory at different  $\alpha$  values. The production phase of the droplet simulations started after an equilibration at the target value of  $\alpha$  lasting several microseconds. Table S3 lists the trajectory ranges used for surface tension calculations. For  $\alpha = 0.625$  near the critical value, we restricted the surface tension calculation to the segments of the trajectory where the droplets were clearly discernible.

**Calculation of Droplet Density.** We computed the radial density profile of MARTINI beads with respect to the droplet center of mass. We converted the number densities into mass densities using a molar mass of 17.168 kg/mol for the simulated FUS-LCD. We used the clustering algorithm described above, defined the largest cluster as “droplet” and calculated the radial density profile with respect to its center of mass. The same method was used to compute radial water density profiles.

**Matching of the Excess Transfer Free Energy.** The excess transfer free energy  $\Delta G_{\text{trans}}$  to bring one protein chain from the dilute phase to the dense droplet phase was computed as follows

$$\Delta G_{\text{trans}} = k_{\text{B}}T \ln \left( \frac{c_{\text{dilute}}}{c_{\text{dense}}} \right) \quad (1)$$

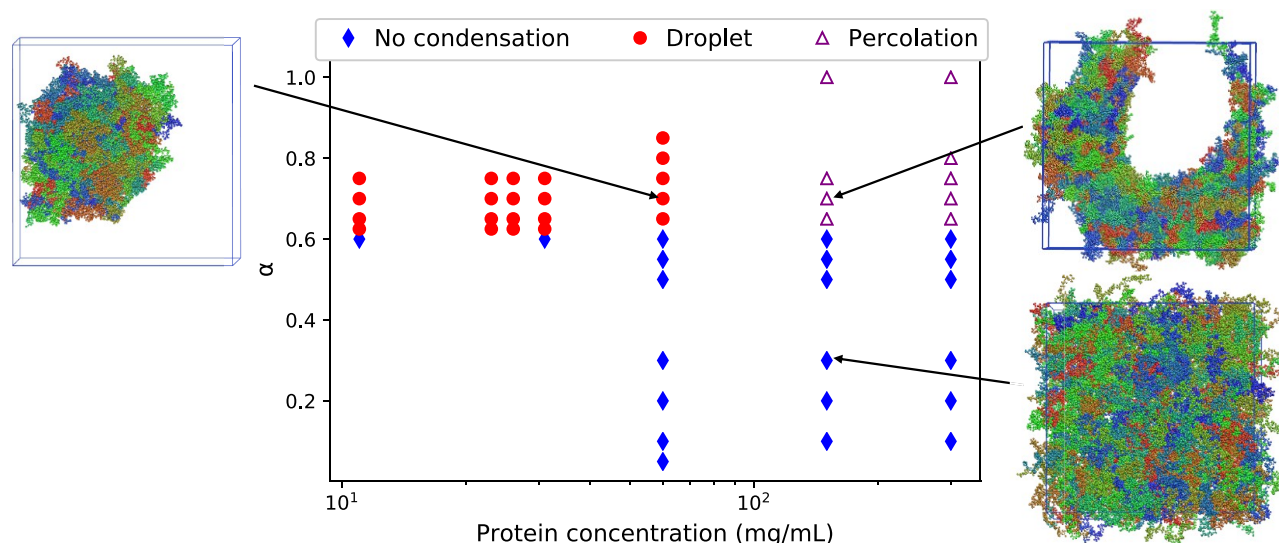
where  $k_{\text{B}}$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $c_{\text{dilute}}$  and  $c_{\text{dense}}$  are the concentrations (i.e., mass densities) in the dilute and dense phases, respectively. Here, we neglected surface effects linked to the droplet curvature. The concentration of FUS in the droplet was calculated using the method described in the previous section. For the concentration of the dilute phase, we divided the number of proteins not part of the largest cluster by the box volume minus the volume of the droplet approximated as a sphere. The droplet radius was obtained from a sigmoidal fit of the protein mass density profiles  $c(r)$

$$c(r) = B - A \operatorname{erf} \left( \frac{r - R}{\sqrt{2} W} \right) \quad (2)$$

where  $r$  is the distance from the droplet center of mass; erf is the error function; and  $A$ ,  $B$ ,  $R$ , and  $W$  are fitted parameters. Note that the derivative of this density profile is a Gaussian distribution with standard deviation  $W$ .<sup>58</sup> For sufficiently large droplets with a defined density plateau at their center,  $B + A = c_{\text{dense}}$  and  $B - A = c_{\text{dilute}}$  correspond to the protein concentrations in the dense and dilute phases. From these limiting densities, we obtained excess transfer free energies for several  $\alpha$  values that we compared to experimental results for the FUS-LCD found in the literature.<sup>13,14,27,59</sup>

**Surface Effects Linked to the Droplet Curvature.** We computed the excess transfer free energy for a flat surface using the slab simulation method.<sup>60,61</sup> A dense phase forming a continuous slab under periodic boundary conditions was simulated in equilibrium with a dilute phase in an elongated box, with slab surfaces normal to the largest dimension (parallel to the  $z$ -axis). We identified the dense phase as the largest cluster, and considered all other proteins as part of the dilute phase. The concentration of the dilute phase was determined by dividing the number of proteins in the dilute phase by the volume remaining after subtraction of the slab volume, approximated as a rectangular parallelepiped of dimension  $L_x \times L_y \times 2z_0$ .  $L_x$  and  $L_y$  are the box dimensions in the  $x$  and  $y$  directions, respectively, and  $\pm z_0$  are the midpoints of the mass density profiles centered on the origin and fitted to the sigmoidal function eq 2. The excess transfer free energy was computed according to eq 1.

**Determination of the Surface Tension.** The surface tension was estimated using two methods, both based on the theory of Henderson and Lekner<sup>62</sup> for the thermal fluctuations of the shape of a spherical droplet composed of an incompressible fluid. In this model, thermally activated capillary waves roughen the interface and result in fluctuations of the droplet shape. These thermal fluctuations are associated with small changes  $\delta A$  in the surface area of the droplet relative to the sphere. The potential energy  $U \approx \gamma \delta A$  of surface-shape fluctuations is assumed to be dominated by the surface tension  $\gamma$ . As derived in the Supporting Information, the fluctuations in the shape of the droplet at lowest order in a spherical harmonics expansion give us two independent estimates of the surface tension



**Figure 1.** Phase diagram of FUS-LCD. Symbols indicate the morphologies observed in MD simulations at different values of the protein concentration and the scaling parameter  $\alpha$  (blue filled diamonds: dispersed proteins without condensation; red filled circles: droplets; open triangles: dense protein aggregates percolated across simulation box boundaries in an inverted topology). Insets show representative structures, connected by arrows to the respective point in the  $\alpha$ - $c$  plane. Shown are the FUS-LCD chains in different colors, with water and ions omitted for clarity. The simulation results correspond to the “homogeneous” starting configuration in Table S2 for concentrations of 150 mg/mL or higher and for 60 mg/mL at  $\alpha \leq 0.7$ , and to the “preformed droplet” simulations for the rest (see Table S2).

$$\gamma_{20} = \frac{5k_B T}{16\pi(\delta a + \delta b)^2} \quad (3)$$

and

$$\gamma_{22} = \frac{15k_B T}{16\pi(\delta a - \delta b)^2} \quad (4)$$

where  $\delta a = a - R$  and  $\delta b = b - R$  are the differences in lengths of any pair of principal axes of a general ellipsoid describing the instantaneous shape of the droplet with respect to the average radius of the droplet,  $R$ . The average  $\langle \cdot \rangle$  is thus both over the droplet shapes along the MD trajectory and over the three distinct combinations of principal axes. We estimated the three axis lengths  $a$ ,  $b$ , and  $c$  from a principal component analysis (PCA) of the mass distribution, as described in the Supporting Information.

We obtained a third estimate of the surface tension from the width of the interface between the droplet and the dilute phase, again relying on the theory of Henderson and Lekner.<sup>62</sup> Capillary wave theory predicts that the squared interface thickness  $W^2$  of an incompressible droplet grows as the logarithm of the droplet radius, with a prefactor proportional to the reciprocal of the surface tension

$$W^2 = W_0^2 + \frac{k_B T}{2\pi\gamma} \ln\left(\frac{R}{B_0}\right) \quad (5)$$

where  $W$  is the interface width,  $W_0$  is an intrinsic width,  $R$  is the droplet radius, and  $B_0$  is a short-wavelength cutoff. Mittal and Hummer<sup>58</sup> used this relation to extract an interfacial tension for the water–vapor interface around a hydrophobic solute.<sup>58</sup> Following their procedure, we simulated droplets of different diameters by varying the number of proteins. We extracted the interfacial width  $W$  of the droplets from fits of eq 2 to the radial concentration profiles. The surface tension was then determined from the slope of a linear fit to  $W^2$  as a

function of  $\ln R$  according to eq 5. Standard errors of  $W^2$  and  $\ln R$  were estimated from block averaging.

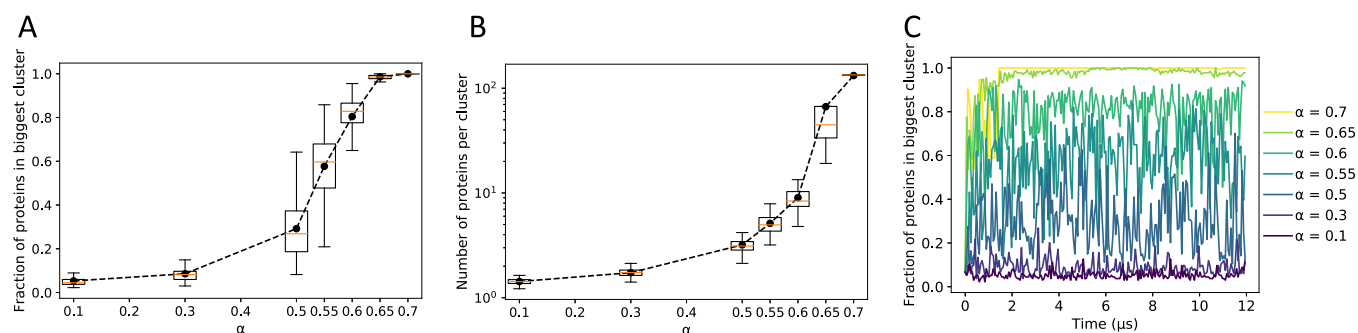
**Estimation of the Shear Viscosity.** To estimate the shear viscosity  $\eta$ , we computed the normalized autocorrelation function of the end-to-end distance of the protein chains. We estimated the autocorrelation time  $\tau$  of the end-to-end distance as the amplitude-weighted sum of the two relaxation times in a biexponential fit to the autocorrelation function averaged over all proteins in the system. We approximated the effective diffusion coefficient  $D$  of the end-to-end motion as the ratio of the variance of the end-to-end distance  $r$  and the autocorrelation time  $\tau$

$$D \approx \frac{\text{var}(r)}{\tau} \quad (6)$$

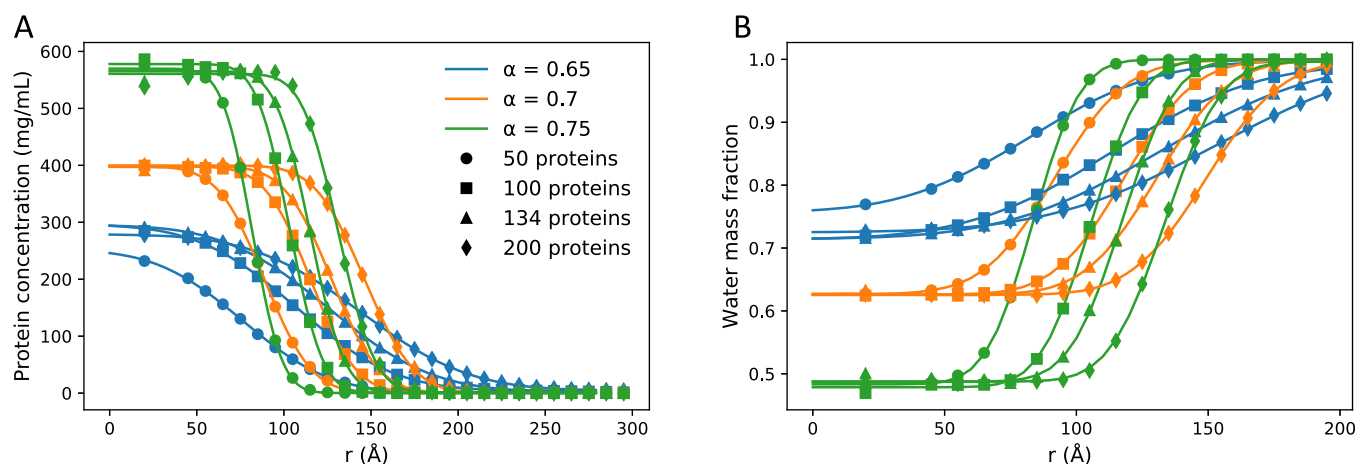
This expression becomes exact in the harmonic limit<sup>63</sup> and accounts for possible effects of  $\alpha$  on the compactness<sup>24</sup> of the FUS-LCD chains within the dense phase (Figure S1). For reference, we performed the same analysis for isolated FUS-LCD chains in aqueous solution, giving us diffusion coefficients  $D_0$ . We then assumed that the effective diffusion coefficients for the end-to-end distance relaxation scale as  $1/\eta$ . Accordingly, we estimated the shear viscosity as

$$\eta(\alpha) \approx \frac{D_0}{D(\alpha)} \eta_w \quad (7)$$

As a reference, we used the effective end-to-end diffusion coefficient  $D_0$  of an isolated FUS-LCD chain with  $\alpha = 0.6$ , where we expect the contributions of attractive nonbonded interactions to the friction to be small<sup>64</sup> such that solvent contributions dominate. For the viscosity of MARTINI water, we used  $\eta_w = 1.0 \times 10^{-3}$  Pa s.<sup>65</sup> With proteins frequently transferring between the droplet and the dilute phase, our analysis does not distinguish if a particular protein is inside or outside the droplet. However, only at values of  $\alpha$  close to 0.6, where the droplet dissolves, contributions from proteins in the dilute phase become significant. In the most relevant regime of



**Figure 2.** Cluster analysis of protein condensation as a function of the scaling parameter  $\alpha$ . (A) Box-whisker plot of the fraction of proteins contained in the biggest cluster in a simulation (mean: filled circle; median: horizontal line; box: interquartile range; error bars: range). (B) Box-whisker plot of the number of proteins per cluster. (C) Fraction of proteins in the largest cluster as a function of time in MD simulations at different values of  $\alpha$ , as indicated in the legend.



**Figure 3.** Radial protein and water concentration profiles in FUS-LCD droplets. (A) Protein mass density as a function of radial distance  $r$  from the droplet center. (B) Relative mass fraction of water as a function of  $r$ . Radial density profiles from MD simulations (symbols) are shown for different values of  $\alpha$  and different droplet sizes, as indicated. Lines are fits to the error function profile eq 2.

$\alpha \geq 0.65$ , nearly all proteins are within the droplet and eq 7 thus reports on the viscosity within the droplet.

## RESULTS

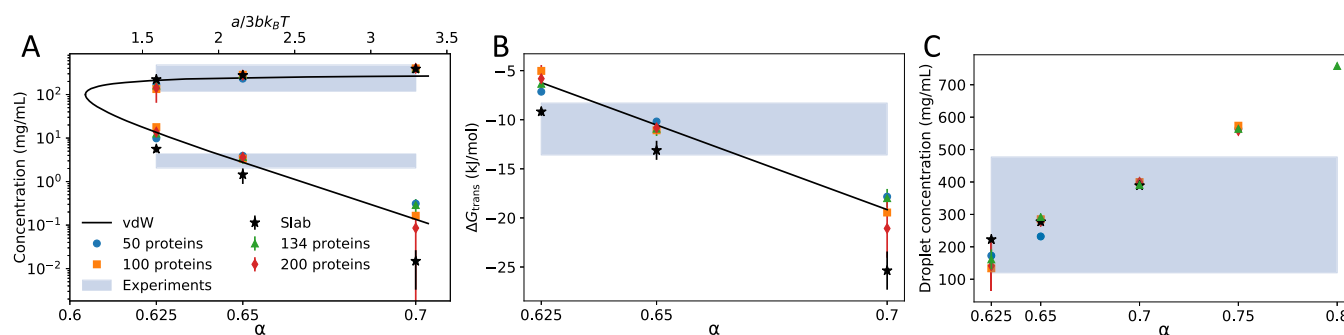
**Phase Separation of FUS-LCD.** We simulated ensembles of FUS proteins at different protein numbers, box sizes, and interaction scaling factors  $\alpha$ , with homogeneous solutions or preformed droplets as starting configurations, as detailed in Table S2. Above a critical interaction strength,  $\alpha > 0.6$ , we observed phase separation at all protein concentrations simulated, 11–300 mg/mL (Figure 1). In simulations starting from configurations with dispersed proteins at low average protein concentration, phase separation led to the formation of a roughly spherical droplet consistent with experimental studies of FUS.<sup>9,66,67</sup> For  $\alpha > 0.6$ , FUS-LCD also formed condensates in MD simulations at high concentration of initially dispersed proteins. However, the topology of the resulting dense phase was inverted. For  $\alpha > 0.6$  and concentrations  $c > 100$  mg/mL, continuous protein condensates formed that percolated across the periodic boundaries of the simulation box (see the inset at the top right of Figure 1).

**Cluster Formation.** We monitored the formation of FUS-LCD clusters to characterize the process of phase separation and to identify droplets. As we increased the interaction scaling factor  $\alpha$  beyond 0.6 at fixed protein concentration and

temperature, we found the system to phase separate through the formation of clusters. Supporting Movie S1 shows the formation of a droplet starting from a homogeneous solution of 134 proteins at  $\alpha = 0.7$ . The proteins first condense into small clusters, which then merge to form a droplet. For  $0.6 < \alpha \leq 0.7$ , FUS-LCD condensation leads to the coexistence of a dense phase and a dilute phase.

**Critical Interaction Strength.** To quantify the critical interaction strength, we measured the time-averaged number and size of clusters at different values of  $\alpha$  (Figure 2). The cluster analysis was performed on simulations starting from a homogeneous solution of 134 proteins in a  $40 \times 40 \times 40$  nm<sup>3</sup> box at different  $\alpha$  values (Table S2, Figure S2). At low  $\alpha$  values ( $\alpha < 0.6$ ), the systems did not undergo phase separation. We found that the proteins remained dispersed, forming only small clusters that each contained only a small fraction of the proteins (Figure S2B). As  $\alpha$  increases beyond 0.6, the number of clusters drops while the average cluster size grows sharply (Figure S2A,B). Phase separation led to the formation of one big cluster containing most of the proteins (>95% for  $\alpha \geq 0.65$ ) in coexistence with a dilute phase. The variations of these quantities with  $\alpha$  indicate  $\alpha_c \approx 0.6$  as the critical interaction strength for our FUS-LCD model at ambient temperature and pressure, above which phase separation occurs.

**Reversibility of Phase Separation.** We confirmed that droplet formation is reversible. We initiated simulations with a



**Figure 4.** Phase diagram and excess transfer free energy for FUS-LCD. (A) Concentration of the coexisting dense and dilute phases from MD simulations (symbols) for different values of  $\alpha$  and droplet sizes, as indicated. The star markers show the densities obtained from the slab simulation method.<sup>60,61</sup> Bars for the largest droplets and the slab systems indicate standard errors. The black line shows the coexistence curve of van der Waals mean field theory with  $a/(3bk_B T)$ , a linear function of  $\alpha$ , as indicated on the top axis. The suitably adjusted critical density  $1/(3b)$  corresponds to a critical protein concentration of 100 mg/mL. Blue shading indicates the range of saturation concentrations reported for FUS-LCD<sup>13,14,59</sup> and for the dense-phase concentration.<sup>13,27</sup> (B) Excess transfer free energy of FUS-LCD between the dilute phase and the dense phase (i.e., the droplet) calculated from eq 1. The black line is a fit to a linear function in  $\alpha$ . Blue shading indicates the range of excess transfer free energies calculated for the densities estimated from experiments, as indicated by shading in (A). (C) Concentration of the dense FUS-LCD phase as a function of the energy scaling parameter  $\alpha$ . The blue area indicates the range of FUS-LCD concentrations in condensates reported from experiment.<sup>13,14,59</sup>

droplet that had been formed at  $\alpha = 0.7$ . After decreasing  $\alpha$  to 0.55 and 0.5, we found the droplets to dissolve (Figure S3). This dissolution of preformed droplets at  $\alpha < 0.6$  is consistent with the phase behavior seen in simulations started with dispersed proteins. Droplet formation is thus reversible on the MD timescale.

**Density Profiles.** For  $\alpha \geq 0.65$ , phase separation led to the formation of a single droplet that remained stable over times  $>10 \mu\text{s}$ . We computed the protein density profiles inside the droplets at different  $\alpha$  values (Figures 3A and S4) from which we extracted the dependence of the dense-phase concentration on  $\alpha$  (Figure 4A,C). We found that the protein density inside the droplets for a given  $\alpha$  value is independent of the overall protein concentration in the system, consistent with the condensate being a distinct phase. The relevant quantities are hence  $\alpha$  and the number of proteins  $N$ , which define the droplet density and size, respectively. The dense-phase concentrations for  $0.65 \leq \alpha \leq 0.7$  are in the range of 300–400 mg/mL (Figure 4C), in line with dense-phase concentrations of IDPs reported from experiments.<sup>27</sup> For  $\alpha \geq 0.8$ , we observed density inhomogeneities inside the “droplets” (Figure S4) that persisted on the simulation timescale. These irregular structures indicate that the strong protein–protein interactions for  $\alpha \geq 0.8$  slowed down the relaxation process beyond the timescale of our MD simulations. In the following, we focus our attention on the  $\alpha$  region where phase separation leads to the formation of stable droplets with liquid-like properties.

**Excess Transfer Free Energy between Dilute and Dense Phases of FUS-LCD Condensate.** The MARTINI force field<sup>40</sup> has been parameterized extensively on the basis of excess transfer free energies (e.g., the water/oil partition coefficient). Adopting this general parameterization approach, we calculated the excess transfer free energy for FUS LCDs between coexisting dilute and dense phases. We used the cluster algorithm and the droplet density profiles to compute the protein concentrations of the two phases at coexistence. From the logarithm of their ratio, according to eq 1, we then obtained the excess transfer free energy.

Figure 4 shows the resulting densities and excess transfer free energies for different values of  $\alpha$ , as obtained from

simulations with different numbers of proteins and, thus, with different droplet sizes. We find that the density of the dilute phase changes by several orders of magnitude in the range of  $0.625 \leq \alpha \leq 0.7$ . By contrast, the protein concentration in the dense phase varies only by a factor of 3 in this  $\alpha$  range.

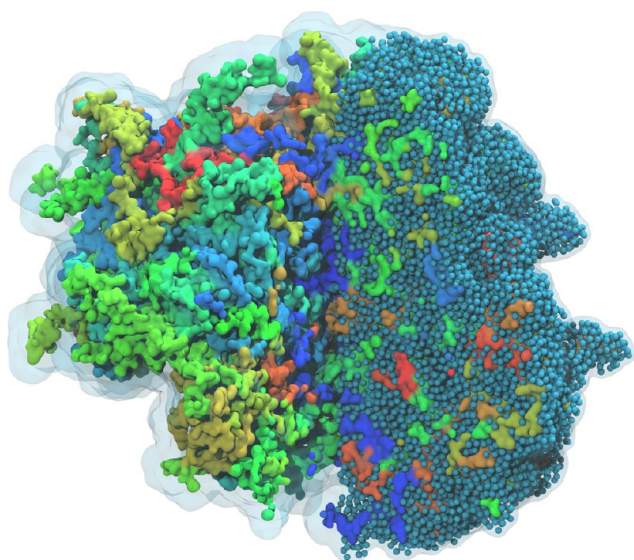
For a value of  $\alpha = 0.65$ , the densities of both the dense and the dilute phase, and so the excess transfer free energy, are consistent with experimental values.<sup>13,14,27,59</sup> We therefore suggest  $\alpha = 0.65$  as a suitable scaling factor to describe the thermodynamics of FUS-LCD phase separation at ambient conditions.

#### Phase Diagram and Effective van der Waals Fluid.

The coexistence behavior in the  $\alpha$ – $c$  plane appears to be captured by van der Waals mean field theory,<sup>68</sup> at least at a semiquantitative level. As shown in Figure 4, a linear relation between the van der Waals interaction parameter  $a$  and  $\alpha$  and a simple scaling of the protein concentration  $c$  bring the observed densities at coexistence into good correspondence with the coexistence curve of van der Waals theory in the plane spanned by the interaction strength  $a$  and density  $\rho$ . Here, we fixed the temperature  $T$  and the volume parameter  $b$  in the van der Waals equation of state,  $p = k_B T / (v - b) - a/v^2$ , with  $p$  the pressure and  $v = 1/\rho$  the volume per particle. This correspondence is consistent with the estimate  $\alpha_c \approx 0.6$  of the critical value of the scaling parameter, below which no condensation occurs. A quantitative analysis of this correspondence would require careful finite-size scaling, which is beyond the scope of this work.

**Droplet Composition.** The presence of explicit solvent in the MARTINI model enabled us to characterize the hydration level of the droplet interior for different  $\alpha$  values (Figures 3B and 5). For  $\alpha = 0.65$ , water accounts for 70% of the droplet mass. At increasing  $\alpha$ , the hydration level decreases. In effect, tightening the protein–protein interactions within the droplet squeezes out the water. Supporting Movie S2 visualizes the water inside the droplet for  $\alpha = 0.7$ .

**Surface Tension of FUS-LCD Droplets from Capillary Wave Theory.** We found that the surface tension of the FUS-LCD droplets calculated from the interfacial width increases with  $\alpha$  (Figure 6B). As shown in Figure 6A, the squared width of the droplet interface grows linearly with the logarithm of the



**Figure 5.** Hydration of the interior of FUS-LCD droplet for  $\alpha = 0.7$  with  $N = 134$  proteins. In this MD simulation snapshot, FUS-LCD proteins in the left half of the droplet are shown in colored surface representation, with a transparent surface indicating water. To indicate the extent of hydration, water beads within the FUS-LCD droplet, or at its surface, are shown as blue spheres in the right half of the droplet. The front right wedge has been removed to provide a clearer view of the droplet interior.

droplet radius, as predicted by capillary wave theory, eq 5. In Figure S5, we show the error function fits to the interfacial density profiles, from which the droplet interfacial widths and radii were extracted. Figure 6B shows the surface tension values calculated from the interfacial widths according to eq 5. At  $\alpha = 0.625$ , where the critical  $\alpha$  is approached, the surface tension is about 25 times lower than at  $\alpha = 0.75$ .

The surface tensions calculated from droplet shape fluctuations and from the interfacial width are consistent. Droplet shape fluctuations are illustrated in Figure 7C–F. For  $0.65 \leq \alpha \leq 0.70$  and droplets with  $>100$  proteins, where shape fluctuations could be resolved and adequately sampled, we obtained excellent agreement with the estimates from the interfacial width. For  $\alpha = 0.625$ , poorly defined droplet shapes

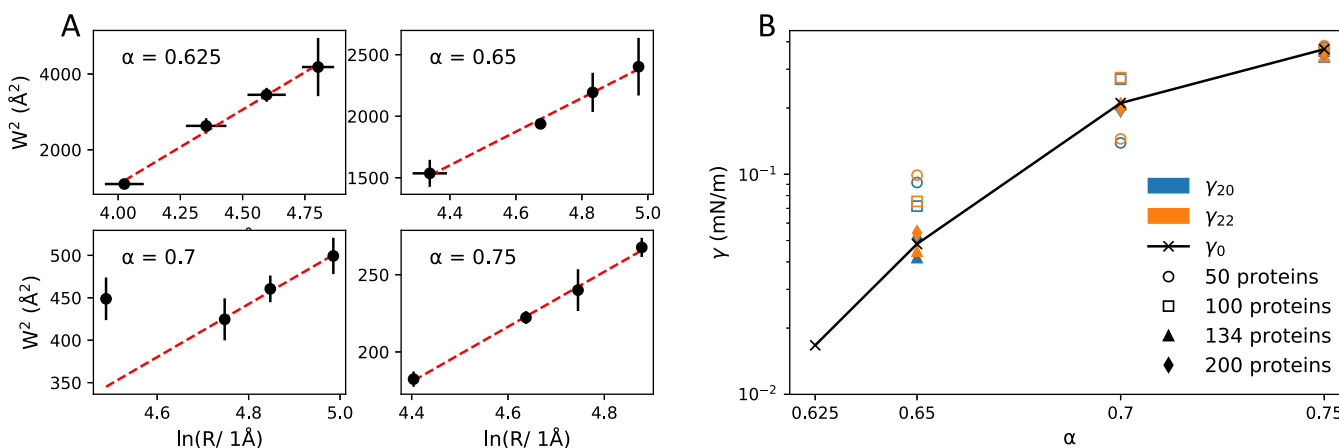
did not allow us to extract reliable ellipsoid axes. For  $\alpha = 0.65$ , where the FUS-LCD excess transfer free energy is matched, the surface tension is consistently about  $\gamma \approx 0.05$  mN/m.

As a further test of capillary wave theory, we compare in Figures S6 and S7 the distributions of the squared amplitudes of the ellipsoidal droplet shape fluctuations to the predicted exponential distributions: we have good correspondence, in line with the agreement in the surface tension values from droplet shapes and interface widths.

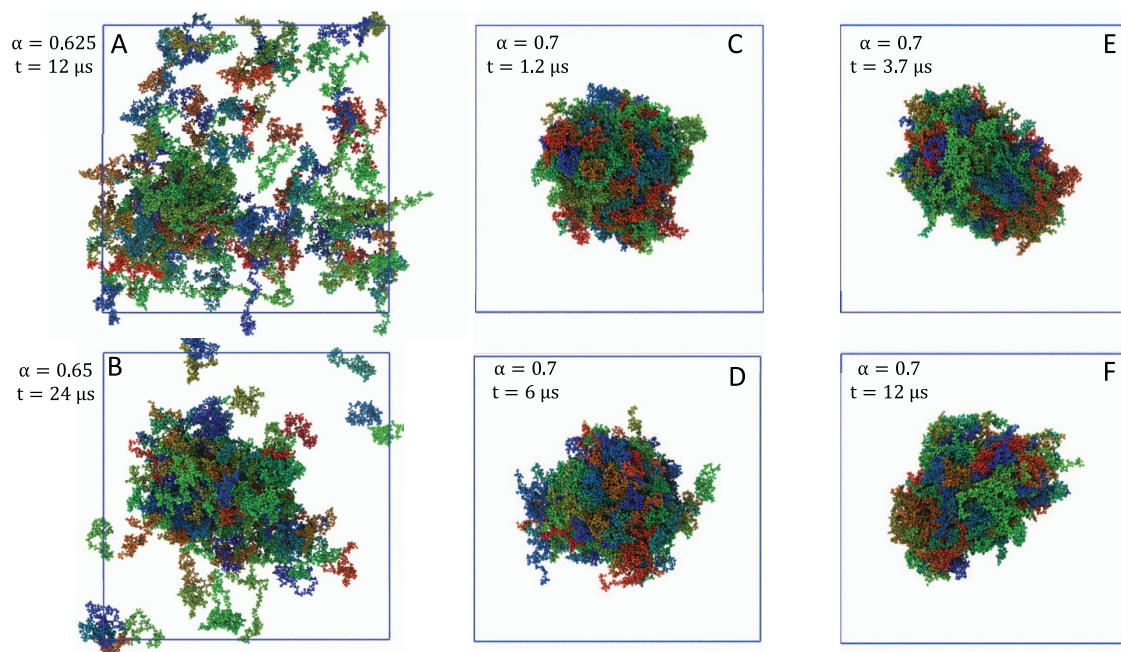
**Droplet Shear Viscosity from FUS-LCD End-to-End Distance Relaxation.** We estimated the shear viscosity of the protein droplets from the standard deviation of the protein end-to-end distance and its relaxation time according to eq 7. We observed that the relaxation time  $\tau$  depends exponentially on  $\alpha$ , increasing by about a factor of 10 between  $\alpha = 0.6$  (where droplets start to form) and  $\alpha = 0.75$  (Figure 8A). By contrast, the relaxation times of isolated FUS-LCD chains free in solution are nearly independent of  $\alpha$ . From the ratio of the standard deviation and the relaxation time (Figures S8 and S9), we obtained the end-to-end diffusion coefficient using eq 6. Then, we used eq 7 to calculate the viscosity, assuming that the diffusion coefficient scales as  $1/\eta$ , with a reference effective end-to-end diffusion coefficient  $D_0$  that we had estimated from MD simulations of a free FUS-LCD chain at  $\alpha = 0.6$  (Figure S10). The calculated viscosities are shown in Figure 8B. We find that  $\eta$  increases exponentially with  $\alpha$  from about 0.001 to 0.02 Pa s. For  $\alpha = 0.65$ , where our rebalanced model matches the experimental excess transfer free energy, we estimate a shear viscosity for FUS-LCD droplets in the range of 0.002–0.004 Pa s.

## DISCUSSION

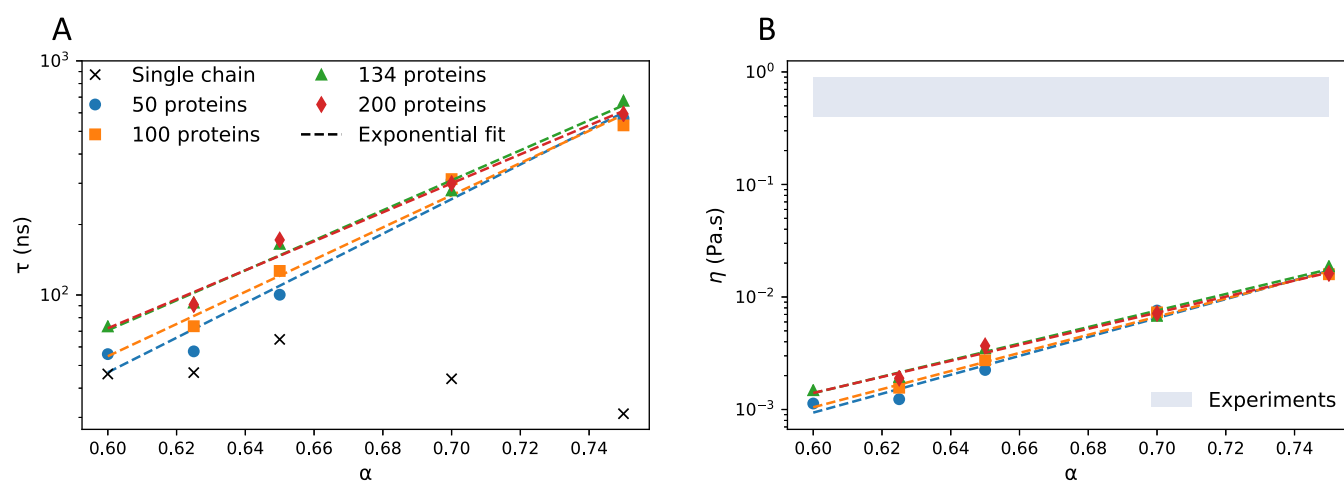
**Phase Separation.** We performed MD simulations of concentrated solutions of FUS-LCD using a modified MARTINI coarse-grained model. We scaled the strength of the protein–protein interactions using the  $\alpha$  parameter introduced by Stark et al.<sup>41</sup> For  $\alpha > 0.6$ , we observed spontaneous phase separation of FUS-LCD solutions into a dense phase and a dilute phase. For sufficiently large and dilute systems, the dense regions coalesced into a spherical droplet. In the dilute phase, the FUS-LCD chains remained dispersed.



**Figure 6.** Surface tension of FUS-LCD droplets. (A) Squared width of droplet interface as a function of the logarithm of the droplet radius (symbols). Dashed lines are linear fits according to eq 5. For  $\alpha = 0.7$ , the smallest droplet was considered an outlier and was thus not included in the fit. Standard errors, as indicated by error bars, were estimated by block averaging. Values of  $\alpha$  are indicated. (B) Surface tension as a function of  $\alpha$  from eqs 3 ( $\gamma_{20}$ ) and 4 ( $\gamma_{22}$ ; filled and open symbols; droplet size is indicated) and from eq 5 ( $\gamma_0$ ; crosses connected by lines).



**Figure 7.** Snapshots of the MD simulation systems at different values of  $\alpha$  and different time points: (A)  $\alpha = 0.625$ , (B)  $\alpha = 0.65$ , (C–F)  $\alpha = 0.7$  with time points indicated. The blue squares indicate the simulation box size.



**Figure 8.** Shear viscosity estimated from the dynamics of the FUS-LCD end-to-end distance. (A) Relaxation time  $\tau$  of the end-to-end distance as a function of  $\alpha$  for droplets of different sizes (filled symbols; as indicated). The crosses show  $\tau$  for an isolated FUS-LCD protein free in aqueous solution. (B) Shear viscosity  $\eta$  of the droplets as a function of  $\alpha$ . The blue shaded region indicates the range of experimental estimates reported for FUS-LCD.<sup>13,67</sup> Dashed lines in (A) and (B) are fits to exponential functions in  $\alpha$ .

The thermodynamics of phase separation is highly sensitive to the strength of the protein–protein interactions. In a narrow window of  $\alpha$  between 0.625 and 0.7, the concentration of the dilute phase varies by about a factor 100 (Figure 4A). In the same window, the dense-phase concentration varies by a factor of 3 (Figure 4C). As a result, the excess free energy for transferring FUS-LCD chain between the two phases varies by about 15 kJ/mol (Figure 4B) if  $\alpha$  is changed by 0.1. These sensitivities to seemingly tiny changes in the interaction strength make it clear that force fields have to be rebalanced to describe biopolymer LLPS in a quantitative manner.

The strong dependence on the energy scaling parameter  $\alpha$  is consistent with the predictions of the van der Waals mean field theory of phase transitions. In Figure 4B, we mapped the van der Waals model to the FUS-LCD data using a linear relation between the respective attraction parameters  $a$  and  $\alpha$ . The

coexistence line of the van der Waals model captures the observed phase behavior remarkably well, giving us a rough estimate of the critical value  $\alpha_c \approx 0.6$ . Indeed, below  $\alpha = 0.6$ , we did not observe phase transition. On the contrary, droplets preformed at higher values of  $\alpha$  dissolved when  $\alpha$  was reduced below 0.6, showing that condensation is reversible on the timescale of the MD simulations (Figure S3).

For  $0.65 \leq \alpha \leq 0.75$ , we verified that the initial protein concentration does not influence the protein density in the dense phase, as expected for a phase in thermodynamic equilibrium. Hence, for the simulation of droplets of different sizes, the relevant quantities are the initial number of proteins, the box size, and  $\alpha$ . For a given  $\alpha$  value, the protein concentrations inside droplets of different sizes were remarkably close (Figure 3), again indicating that thermodynamic equilibrium has been established in the simulations.



However, in simulations of systems with initially dispersed proteins at high total concentrations (i.e., close to the dense-phase concentration along the coexistence curve), protein condensates with inverted topologies formed. In these simulations, the dense protein phase was continuous in the periodic system, causing strong finite-size effects (Figure 1). Interestingly, for  $\alpha = 0.65$ , the protein concentration of >400 mg/mL in this percolated dense phase is substantially higher than the  $\approx 300$  mg/mL inside the spherical droplets and the periodic slab at the same value of  $\alpha$ . Further study of these dense inverted “phases” could be relevant for atomistic simulations of bulk condensates, where box sizes tend to be smaller.

**Optimal Interaction Strength  $\alpha$ .** In the spirit of the MARTINI model,<sup>40</sup> we tuned the  $\alpha$  parameter to reproduce the excess free energy of transferring an FUS-LCD chain from the dilute phase into the dense phase according to eq 1. For FUS-LCD in aqueous solution at ambient conditions, a value of  $\alpha = 0.65$  reproduces the excess transfer free energy and the densities of the dilute and dense phases as estimated from experimental measurements. However, the reported saturation densities depend on temperature and ionic strength,<sup>13</sup> and the range in Figure 4A for the dilute phase may have to be extended higher. Nevertheless, we consider the observed consistency of transfer free energy and the estimated densities as a first validation test.

The dense-phase concentration increases with  $\alpha$  and ranges from 150 to 760 mg/mL for  $0.625 \leq \alpha \leq 0.8$ . Experimental studies reported condensed phase concentrations between 120<sup>13</sup> and 477 mg/mL.<sup>27</sup> These two values bracket the concentration we obtained for  $\alpha = 0.65$ , where our model matches the excess transfer free energy. This result makes us optimistic about the capacity of our model to reproduce also other relevant properties of FUS-LCD condensates.

**Effects of Droplet Curvature.** In matching  $\alpha$  to experiment, we did not take into account surface effects linked to the droplet curvature, which we expect to give rise to a slight increase in the density of the dilute phase. As a rough estimate for  $\alpha = 0.65$ , we use the Kelvin equation and the ideal gas law to estimate the relative increase in the density of the dilute phase above a droplet of radius  $R$

$$\frac{\rho}{\rho_0} \approx \frac{p}{p_0} = \exp\left(\frac{2\gamma}{\rho_m R k_B T}\right) \quad (8)$$

For a density of  $\rho_m \approx 0.009 \text{ nm}^{-3}$  of FUS-LCD in the dense phase, a surface tension of  $\gamma \approx 0.05 \text{ mN/m}$ , and a radius of  $R \approx 13 \text{ nm}$ , the expected increase in the density of the dilute phase is about 20% and thus comparable to the uncertainties in our estimates of the density of the dilute phase. Conversely, following Powles et al.,<sup>69</sup> we solved eq 8 for the surface tension  $\gamma$ , with  $\rho$  the density of the dilute phase above the droplets,  $\rho_0$  the density of the dilute phase above the slab (Figure S11),  $\rho_m$  the density of the droplets, and  $R$  the radius of the droplets. For  $\alpha = 0.65$ , we obtained surface tensions of 0.13–0.27 mN/m, about 2–6 times higher than our estimates from droplet shape fluctuations. In light of the simplifications in eq 8, where we assumed the dilute phase to be ideal, and the uncertainties from finite-size effects and slow convergence of the dilute-phase concentration, we find this semiquantitative agreement in the calculated surface tensions to be reassuring. We refer to ref 17 for a detailed discussion of curvature effects.

**Slow Relaxation at High  $\alpha$ .** We calculated the protein mass density inside the droplets and monitored for possible inhomogeneities. For  $0.65 \leq \alpha \leq 0.75$ , the dense phase appeared homogeneous, consistent with a liquid-like state of these droplets. By contrast, for  $\alpha \geq 0.8$ , significant inhomogeneities emerged. At these high values of  $\alpha$ , FUS-LCD condensed and then got trapped in structures resembling amorphous aggregates subject to slow relaxation on the MD timescale. Extrapolating to such high values of  $\alpha$ , the end-to-end distance relaxation occurs on timescales of several microseconds (Figure 8A). Such slow chain reconfiguration and the associated high viscosity of the dense phase prevent the relaxation on an MD timescale for  $\alpha \geq 0.8$ . Patel et al.<sup>9</sup> reported that physiologically competent FUS droplets have to be liquid, but that aging can lead to aggregation and transition to a solid-like state. By performing long MD simulations of FUS-LCD condensates for  $\alpha \geq 0.8$  (or gradually increasing  $\alpha$ ), it might be possible to mimic aspects of droplet aging and monitor changes in the apparent materials properties.

**Hydration of Droplet Interior.** Thanks to the explicit treatment of the solvent by the MARTINI model, we could quantify the droplet hydration levels. The computed water mass fraction profiles inside the droplets showed that water accounts for  $\approx 70\%$  of the droplet mass at  $\alpha = 0.65$  (Figures 3B and 5). As  $\alpha$  is increased, the water content of the droplets decreases, reaching  $\approx 50\%$  for  $\alpha = 0.75$ . In vivo, water is critical to maintain the fluidity of the droplet, the diffusivity of molecules within, and their biochemical reactivity.<sup>70</sup> Murthy et al.<sup>27</sup> report a water content of  $\approx 65\%$  by volume for FUS-LCD droplets. The close correspondence with the calculated water content again suggests  $\alpha = 0.65$  as a reasonable rescaling parameter. A combined coarse-grained and all-atom MD study of FUS-LCD condensates recently reported a water concentration of  $\approx 600 \text{ mg/mL}$  in the protein-rich phase,<sup>29</sup> about 15% lower than what we found for  $\alpha = 0.65$  here (Figure S4B).

**Surface Tension.** We characterized the droplet surface tension in terms of droplet shape fluctuations and interfacial widths. The three estimates obtained in this way for a given value of  $\alpha$  are in excellent correspondence (Figure 6). We found that the surface tension increases strongly with  $\alpha$  over a narrow range. Indeed, as the critical value  $\alpha_c \approx 0.6$  is approached, the surface tension drops to zero, consistent with theoretical expectations (Figure 6). As  $\alpha$  is increased, strengthened protein–protein interactions increase the cohesive forces and, in turn, the surface tension.

For  $0.625 \leq \alpha \leq 0.75$ , we obtained surface tensions ranging from 0.015 to 0.38 mN/m. Taylor et al.<sup>71</sup> reported surface tensions of 0.004 and 0.68 mN/m for the IDPs Whi3 and LAF1, respectively, in the range of our estimates for FUS-LCD.

For  $\alpha = 0.65$ , where the densities of the dilute and dense phases of FUS-LCD estimated from experiments are reproduced, the surface tension is about 0.05 mN/m. For reference, the water–vapor surface tension at ambient conditions is about 1500 times larger. Even the liquid–vapor surface tension of an LJ fluid at a corresponding state is more than 100 times larger. At a reduced temperature of 0.85, where the excess transfer free energy of the LJ fluid matches that of our FUS-LCD model at  $\alpha = 0.65$ , a reduced surface tension of about 0.837 has been reported,<sup>72</sup> corresponding to  $\gamma = 6.7 \text{ mN/m}$ . Here, we used effective LJ parameters  $\sigma = 0.78 \text{ nm}$  and  $\epsilon = 2.9 \text{ kJ/mol}$  for which the densities at coexistence<sup>72</sup> match the amino-acid densities of FUS-LCD (Figure 3). The wide gap between the surface tensions of FUS-LCD and the

corresponding LJ fluid indicates that for interfacial properties the polymeric nature of the droplets and the immersion in an aqueous solvent cannot be neglected.

In an effort to rationalize the low surface tensions obtained here and reported from experiment, we note that the increase in the surface tension goes along with a decrease in the droplet water content (Figure S12). For large  $\alpha$ , we expect the surface tension to saturate at values for compact and dry protein globules (estimated<sup>73</sup> at  $\approx 50$  mN/m). The extrapolation of the exponential fits of  $\gamma$  to zero water content gives a surface tension of 10–20 mN/m for a compact protein (Figure S12). So, at least roughly, we are indeed interpolating between a compact phase (water content zero) and a dissolved phase ( $\gamma = 0$ ; water content 100%). Note that the exponential dependence on  $\alpha$  has to break down as the dissolved phase and the critical  $\alpha$  are approached, where  $\gamma = 0$ .

**Shear Viscosity.** We also estimated the shear viscosity of the FUS-LCD droplets. We found the viscosity to increase exponentially with  $\alpha$  from 0.001 to 0.02 Pa s for  $0.6 < \alpha < 0.8$ . For  $\alpha = 0.65$ , we found  $\eta \approx 0.004$  Pa s. For reference, Murakami et al.<sup>67</sup> reported a viscosity of 0.4 Pa s, and Burke et al.<sup>13</sup> reported a viscosity of  $\approx 0.9$  Pa s. These values are larger than our estimate by about a factor of 100. By performing further comparisons for other proteins undergoing LLPS, it will be interesting to see if the substantially lower viscosity of the MD model is inherent to coarse-graining the potential energy surface, which tends to smoothen the strong distance and orientation dependence of atomic interactions.

**Transferability.** We emphasize that the optimal value of the scaling parameter determined here,  $\alpha = 0.65$ , applies only to FUS-LCD under the thermodynamic conditions of the present study. For other phase separating systems, the optimal values of  $\alpha$  may be different. Nevertheless, one may hope that the value of  $\alpha = 0.65$  obtained here for FUS-LCD serves as a rough estimate also for other proteins, which would give the rebalancing procedure a degree of transferability. Conversely, explorations of the connection between the optimal values of the scaling parameter  $\alpha$  and the nature of the proteins should give insight into the molecular properties that favor phase separation.

**Interaction Rebalancing in LLPS Modeling.** Our rebalancing approach is general and applicable to different types of coarse-grained simulation models of phase separation. There has been much progress in devising coarse-grained<sup>74–76</sup> and implicit solvent models<sup>77</sup> of disordered proteins and we envisage that our general approach will make it possible to leverage these developments in coarse-grained modeling of disordered biomolecules for simulations of LLPS and biomolecular condensates. We note that our tuning of the relative strengths of protein–protein and protein–solvent interactions is similar to corrections implemented for highly optimized atomistic protein force fields.<sup>36–38,78</sup> Further fine tuning may be required for quantitative atomistic simulations of LLPS.

The strong dependence of the droplet biophysical properties on  $\alpha$  underlines how molecular interactions have an impact on physical properties on larger length scales, as it has been stressed before.<sup>79</sup> In experiment, salt or condensing factors are added to modulate the interaction strength. At least at a qualitative level, changes of  $\alpha$  in the simulations may thus mimic these changes in solvent conditions of the experiments. As we start to use MD simulations to study phenomena of increased complexity and on a larger scale, rebalancing may

thus be both a necessity and an opportunity: a necessity to avoid the large errors resulting from systematic imbalances, and an opportunity to probe the phase behavior over a wide range with only minimal modifications to the system.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01064>.

Detailed derivation of the relations between surface tension and droplet shape fluctuations, and between surface tension and droplet interfacial width; amino acid sequence of FUS-LCD; table of simulations; droplet density inhomogeneities; protein cluster formation; reversibility of phase separation; radial protein density profiles; droplet shape fluctuations; cumulative distributions of the amplitudes of shape fluctuations; autocorrelation functions of end-to-end distance of FUS-LCD chains; average end-to-end distance autocorrelation functions; distributions of end-to-end distances; determination of diffusion coefficient for end-to-end distance; and connection between surface tension and hydration (PDF)

Phase separation of FUS-LCD at  $\alpha = 0.7$  and water inside phase-separated FUS-LCD droplet (Movies S1 and S2) (MP4) (MP4)

## ■ AUTHOR INFORMATION

### Corresponding Author

Gerhard Hummer – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; Institute for Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0001-7768-746X](https://orcid.org/0000-0001-7768-746X); Phone: +49 69 6303-2501; Email: [gerhard.hummer@biophys.mpg.de](mailto:gerhard.hummer@biophys.mpg.de)

### Authors

Zakarya Benayad – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; Département de Chimie, École Normale Supérieure, PSL University, 75005 Paris, France; [orcid.org/0000-0002-0578-8066](https://orcid.org/0000-0002-0578-8066)

Sören von Bülow – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0001-7125-5973](https://orcid.org/0000-0001-7125-5973)

Lukas S. Stelzl – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0002-5348-0277](https://orcid.org/0000-0002-5348-0277)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01064>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge financial support from the German Research Foundation (CRC 902: Molecular Principles of RNA Based Regulation), the Human Frontier Science Program RGP0026/2017 (S.v.B. and G.H.), and the Max Planck Society. They thank Drs. Mateusz Sikora and Jürgen Köfinger as well as Lisa M. Pietrek for insightful and helpful discussions.

## REFERENCES

- (1) 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.
- (2) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176*, 419–434.
- (3) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **2020**, *49*, 107–133.
- (4) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 39–58.
- (5) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **2018**, *28*, 420–435.
- (6) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science* **2017**, *357*, No. eaaf4382.
- (7) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoegge, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324*, 1729–1732.
- (8) Li, P. L.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q. X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483*, 336–340.
- (9) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; et al. A Liquid-To-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162*, 1066–1077.
- (10) Quiroz, F. G.; Chilkoti, A. Sequence Heuristics to Encode Phase Behaviour in Intrinsically Disordered Protein Polymers. *Nat. Mater.* **2015**, *14*, 1164–1171.
- (11) Simon, J. R.; Carroll, N. J.; Rubinstein, M.; Chilkoti, A.; López, G. P. Programming Molecular Self-Assembly of Intrinsically Disordered Proteins Containing Sequences of Low Complexity. *Nat. Chem.* **2017**, 509–515.
- (12) Kato, M.; Han, T.; Xie, S.; Shi, K.; Du, X.; Wu, L.; Mirzaei, H.; Goldsmith, E.; Longgood, J.; Pei, J.; Grishin, N.; Frantz, D.; Schneider, J.; Chen, S.; Li, L.; Sawaya, M.; Eisenberg, D.; Tycko, R.; McKnight, S. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **2012**, *149*, 753–767.
- (13) 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.
- (14) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovskiy, A.; Drechsel, D.; et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-Like RNA Binding Proteins. *Cell* **2018**, *174*, 688–699.
- (15) Yang, L.; Embree, L. J.; Hickstein, D. D. TLS-ERG Leukemia Fusion Protein Inhibits RNA Splicing Mediated by Serine-Arginine Proteins. *Mol. Cell Biol.* **2000**, *20*, 3345–3354.
- (16) Polymenidou, M.; Lagier-Tourenne, C.; Hutt, K. R.; Bennett, C. F.; Cleveland, D. W.; Yeo, G. W. Misregulated RNA Processing in Amyotrophic Lateral Sclerosis. *Brain Res.* **2012**, *1462*, 3–15.
- (17) Weber, C. A.; Zwicker, D.; Jülicher, F.; Lee, C. F. Physics of Active Emulsions. *Rep. Prog. Phys.* **2019**, *82*, No. 064601.
- (18) Jacobs, W. M.; Frenkel, D. Phase Transitions in Biological Systems With Many Components. *Biophys. J.* **2017**, *112*, 683–691.
- (19) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018**, *57*, 2499–2508.
- (20) Espinosa, J. R.; Joseph, J. A.; Sanchez-Burgos, I.; Garaizar, A.; Frenkel, D.; Collepardo-Guevara, R. Liquid Network Connectivity Regulates the Stability and Composition of Biomolecular Condensates with Many Components. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 13238–13247.
- (21) 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 Comput. Biol.* **2018**, *14*, No. e1005941.
- (22) Boeynaems, S.; Holehouse, A. S.; Weinhardt, V.; Kovacs, D.; Van Lindt, J.; Larabell, C.; Van Den Bosch, L.; Das, R.; Tompa, P. S.; Pappu, R. V.; Gitler, A. D. Spontaneous Driving Forces Give Rise to Protein-RNA Condensates With Coexisting Phases and Complex Material Properties. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 7889–7898.
- (23) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-Like Domains. *Science* **2020**, *367*, 694–699.
- (24) Chou, H.-Y.; Aksimentiev, A. Single-Protein Collapse Determines Phase Equilibria of a Biological Condensate. *J. Phys. Chem. Lett.* **2020**, *11*, 4923–4929.
- (25) Berry, J.; Weber, S. C.; Vaidya, N.; Haataja, M.; Brangwynne, C. P. RNA Transcription Modulates Phase Transition-Driven Nuclear Body Assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E5237–E5245.
- (26) Pietrek, L. M.; Stelzl, L. S.; Hummer, G. Hierarchical Ensembles of Intrinsically Disordered Proteins at Atomic Resolution in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 725–737.
- (27) 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.
- (28) Paloni, M.; Bailly, R.; Ciandrini, L.; Barducci, A. Unraveling Molecular Interactions in a Phase-Separating Protein by Atomistic Simulations *bioRxiv* 2020. <https://doi.org/10.1101/2020.05.16.099051>.
- (29) Zheng, W.; Dignon, G. L.; 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 *bioRxiv* 2020. <https://doi.org/10.1101/2020.08.05.237008>.
- (30) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic Cluster Formation Determines Viscosity and Diffusion in Dense Protein Solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 9843–9852.
- (31) Dignon, G. L.; Zheng, W.; Mittal, J. Simulation Methods for Liquid–liquid Phase Separation of Disordered Proteins. *Curr. Opin. Chem. Eng.* **2019**, *23*, 92–98.
- (32) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42*, 73–93.
- (33) Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **2017**, *6*, No. e30294.
- (34) Das, S.; Eisen, A.; Lin, Y.-H.; Chan, H. S. A Lattice Model of Charge-Pattern-Dependent Polyampholyte Phase Separation. *J. Phys. Chem. B* **2018**, *122*, 5418–5431.
- (35) Ruff, K. M.; Harmon, T. S.; Pappu, R. V. CAMELOT: A Machine Learning Approach for Coarse-Grained Simulations of Aggregation of Block-Copolymeric Protein Sequences. *J. Chem. Phys.* **2015**, *143*, No. 243123.
- (36) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (37) Best, R. B.; Zheng, W. W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (38) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

- (39) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- (40) 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.
- (41) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward Optimized Potential Functions for Protein-Protein Interactions in Aqueous Solutions: Osmotic Second Virial Coefficient Calculations Using the Martini Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 4176–4185.
- (42) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS One* **2017**, *12*, No. e0187936.
- (43) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (44) Kim, Y. C.; Hummer, G. Coarse-grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. *J. Mol. Biol.* **2008**, *375*, 1416–1433.
- (45) Jost Lopez, A.; Quoika, P. K.; Linke, M.; Hummer, G.; Köfinger, J. Quantifying Protein-Protein Interactions in Molecular Simulations. *J. Phys. Chem. B* **2020**, *124*, 4673–4685.
- (46) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P.; Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J. Chem. Theory Comput.* **2017**, *13*, 5039–5053.
- (47) Larsen, A. H.; Wang, Y.; Bottaro, S.; Grudin, S.; Arleth, L.; Lindorff-Larsen, K. Combining Molecular Dynamics Simulations with Small-Angle X-Ray and Neutron Scattering Data to Study Multi-Domain Proteins in Solution. *PLoS Comput. Biol.* **2020**, *16*, No. e1007870.
- (48) Milkovic, N. M.; Thomasen, F. E.; Cuneo, M. J.; Grace, C. R.; Martin, E. W.; Nourse, A.; Lindorff-Larsen, K.; Mittag, T. Interplay of Folded Domains and the Disordered Low-Complexity Domain in Mediating hnRNP1 Phase Separation *bioRxiv* 2020. <https://doi.org/10.1101/2020.05.15.096966>.
- (49) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (50) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (51) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T., III; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M. et al. *AMBER 2018*; University of California: San Francisco, CA, 2018. © 2019 by the authors.
- (52) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (53) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
- (54) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (55) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (56) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini Straight: Boosting Performance Using a Shorter Cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1–7.
- (57) Humphrey, W.; Dalke, A.; Schulten, K.; et al. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (58) Mittal, J.; Hummer, G. Static and Dynamic Correlations in Water at Hydrophobic Interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20130–20135.
- (59) Murray, D. T.; Kato, M.; Lin, Y.; Thurber, K. R.; Hung, I.; McKnight, S. L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171*, 615–627.
- (60) Blas, F. J.; MacDowell, L. G.; de Miguel, E.; Jackson, G. Vapor-Liquid Interfacial Properties of Fully Flexible Lennard-Jones Chains. *J. Chem. Phys.* **2008**, *129*, No. 144703.
- (61) Silmore, K. S.; Howard, M. P.; Panagiotopoulos, A. Z. Vapor-Liquid Phase Equilibrium and Surface Tension of Fully Flexible Lennard-Jones Chains. *Mol. Phys.* **2017**, *115*, 320–327.
- (62) Henderson, J.; Lekner, J. Surface Oscillations and the Surface Thickness of Classical and Quantum Droplets. *Mol. Phys.* **1978**, *36*, 781–789.
- (63) Hummer, G. Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005**, *7*, 34.
- (64) Best, R. B.; Hummer, G. Diffusive Model of Protein Folding Dynamics with Kramers Turnover in Rate. *Phys. Rev. Lett.* **2006**, *96*, No. 228104.
- (65) Vögele, M.; Köfinger, J.; Hummer, G. Hydrodynamics of Diffusion in Lipid Membrane Simulations. *Phys. Rev. Lett.* **2018**, *120*, No. 268104.
- (66) Weber, S. C.; Brangwynne, C. P. Getting RNA and Protein in Phase. *Cell* **2012**, *149*, 1188–1191.
- (67) Murakami, T.; Qamar, S.; Lin, J. Q.; Schierle, G. S. K.; Rees, E.; Miyashita, A.; Costa, A. R.; Dodd, R. B.; Chan, F. T.; Michel, C. H.; Kronenberg-Versteeg, D.; Li, Y.; Yang, S.-P.; Wakutani, Y.; Meadows, W.; Ferry, R. R.; Dong, L.; Tartaglia, G. G.; Favrin, G.; Lin, W.-L.; Dickson, D. W.; Zhen, M.; Ron, D.; Schmitt-Ulms, G.; Fraser, P. E.; Schneider, N. A.; Holt, C.; Vendruscolo, M.; Kaminski, C. F.; George-Hyslop, P. S. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **2015**, *88*, 678–690.
- (68) Lekner, J. Parametric Solution of the van der Waals Liquid-Vapor Coexistence Curve. *Am. J. Phys.* **1982**, *50*, 161–163.
- (69) Powles, J. G.; Fowler, R. F.; Evans, W. A. B. A New Method for Computing Surface-Tension Using a Drop of Liquid. *Chem. Phys. Lett.* **1983**, *96*, 289–292.
- (70) Takada, S. Coarse-Grained Molecular Simulations of Large Biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130–137.
- (71) Taylor, N.; Elbaum-Garfinkle, S.; Vaidya, N.; Zhang, H.; Stone, H. A.; Brangwynne, C. P. Biophysical Characterization of Organelle-Based RNA/Protein Liquid Phases Using Microfluidics. *Soft Matter* **2016**, *12*, 9142–9150.
- (72) Errington, J. R. Evaluating Surface Tension Using Grand-Canonical Transition-Matrix Monte Carlo Simulation and Finite-Size Scaling. *Phys. Rev. E* **2003**, *67*, No. 012102.
- (73) Absolom, D.; Van Oss, C.; Zingg, W.; Neumann, A. Determination of Surface Tensions of Proteins II. Surface Tension of Serum Albumin, Altered at the Protein-Air Interface. *Biochim. Biophys. Acta, Protein Struct.* **1981**, *670*, 74–78.
- (74) Mioduszewski, L.; Różycki, B.; Cieplak, M. Pseudo-Improper-Dihedral Model for Intrinsically Disordered Proteins. *J. Chem. Theory Comput.* **2020**, *16*, 4726–4733.
- (75) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2018**, *122*, 11115–11125.
- (76) Baul, U.; Chakraborty, D.; Mugnai, M. L.; Straub, J. E.; Thirumalai, D. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2019**, *123*, 3462–3474.
- (77) Vitalis, A.; Pappu, R. V. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.* **2009**, *30*, 673–699.
- (78) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758–E4766.
- (79) Brangwynne, C. P.; Mitchison, T. J.; Hyman, A. A. Active Liquid-Like Behavior of Nucleoli Determines Their Size and Shape in

Xenopus Laevis Oocytes. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4334–4339.