Supporting Information: Simulation of FUS protein condensates with an adapted coarse-grained model

Zakarya Benayad,[†,‡] Sören von Bülow,[†] Lukas S. Stelzl,[†,§] and Gerhard Hummer[*,†,¶]

†*Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany*

‡*Département de Chimie, École Normale Supérieure, PSL University, 75005 Paris, France*

¶*Institute for Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany*

§*Present address: Institute of Molecular Physiology, Johannes Gutenberg University Mainz, Johann-Joachim Becher-Weg 7, 55128 Mainz, Germany; and Institute of Molecular Biology (IMB), 55128 Mainz, Germany*

E-mail: gerhard.hummer@biophys.mpg.de

Phone: +49 69 6303-2501

Table of Contents

Supporting Text

Surface tension from droplet shape fluctuations

Following Henderson and Lekner[1] we assume that the energetics underlying the thermal fluctuations in the droplet shape is dominated by the changes in surface area. The potential energy U of such surface-shape fluctuations is dominated by the surface tension γ:

$$U = \gamma \, \delta A \tag{S1}$$

where δA is a small change in surface area relative to the sphere. 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, The surface of the droplet can be described as a sum of spherical harmonics Y_l^m:

$$r(\theta, \phi) = R + \zeta(\theta, \phi) \tag{S2}$$

with r the radius as a function of the polar and azimuthal angles θ and ϕ, and

$$\zeta(\theta, \phi) = \sum_{\substack{l,m \\ l \geq 2,\, |m| \leq l}} \zeta_{lm} Y_l^m(\theta, \phi) \tag{S3}$$

R is the average radius and ζ_{lm} is the coefficient of the mode (l, m). For small amplitudes, the potential energy U then becomes:

$$U = \frac{\gamma}{2} \sum_{l,m} (l-1)(l+2) \left|\zeta_{lm}\right|^2 \tag{S4}$$

For small perturbations ($|\zeta|^2 \ll R$), each mode (l, m) is thus effectively harmonic and independent. As a consequence, the equipartition theorem gives a relation between the surface tension and the mean squared amplitudes of the shape fluctuations:

$$\frac{\gamma}{2}(l-1)(l+2) \left\langle \left|\zeta_{lm}\right|^2 \right\rangle = \frac{k_\mathrm{B} T}{2} \tag{S5}$$

Each mode (l, m) gives an independent estimate for the surface tension. Focussing on the low-frequency modes, we approximate the instantaneous droplet shape as an ellipsoid with axes a, b and c. If this ellipsoid is aligned with the Cartesian coordinate system, its spherical coordinate representation is:

$$r^2(\theta, \phi) = a^2 \sin^2\theta \cos^2\phi + b^2 \sin^2\theta \sin^2\phi + c^2 \cos^2\theta \tag{S6}$$

We expand $r(\theta, \phi)$ to second order in $a = R + \delta a$, $b = R + \delta b$, $c = R + \delta c$. By projecting onto spherical harmonics, we find:

$$r = \zeta_{00} Y_0^0 + \zeta_{20} Y_2^0 + \zeta_{22} Y_2^2 + \zeta_{2,-2} Y_2^{-2} \tag{S7}$$

where $Y_l^m = Y_l^m(\theta, \phi)$ and $r = r(\theta, \phi)$. The expansion coefficients are:

$$\zeta_{00} = \sqrt{4\pi} \left(R + \frac{\delta a + \delta b + \delta c}{3} \right) \tag{S8}$$

$$\zeta_{20} = \sqrt{\frac{4\pi}{45}} (2\delta c - \delta a - \delta b) \tag{S9}$$

$$\zeta_{2,\pm 2} = \sqrt{\frac{2\pi}{15}}(\delta a - \delta b) \tag{S10}$$

The condition of constant volume requires that:

$$V = \frac{4\pi}{3}abc = \frac{4\pi}{3}R^3 \tag{S11}$$

To lowest order, we have

$$V = \frac{4\pi}{3}[R^3 + R^2(\delta a + \delta b + \delta c)] \tag{S12}$$

and thus

$$\delta a + \delta b + \delta c = 0 \tag{S13}$$

By imposing this condition on the amplitudes of the modes, we find

$$\zeta_{00} = \sqrt{4\pi}R \tag{S14}$$

$$\zeta_{20} = -\sqrt{\frac{4\pi}{5}}(\delta a + \delta b) \tag{S15}$$

$$\zeta_{2,\pm 2} = \sqrt{\frac{2\pi}{15}}(\delta a - \delta b) \tag{S16}$$

We thus arrive at an expression for the potential energy associated with surface area changes $U = \gamma\, \delta A$ to second order in a spherical harmonics expansion:

$$\begin{aligned} \delta U &= \frac{\gamma}{2}\sum_{l\geq 2}\sum_{m}(l-1)(l+2)|\zeta_{lm}|^2 \\ &\approx 2\gamma\left[|\zeta_{20}|^2 + |\zeta_{2,-2}|^2 + |\zeta_{2,2}|^2\right] \\ &= \frac{\gamma}{2}\frac{16\pi}{15}\left[3\left\langle(\delta a + \delta b)^2\right\rangle + \left\langle(\delta a - \delta b)^2\right\rangle\right] \end{aligned} \tag{S17}$$

This expression for the potential energy combined with the equipartition theorem for the two independent modes, $\delta a + \delta b$ and $\delta a - \delta b$, gives us the two independent expressions for the surface tension, eqs 3 and 4 of the main text.

Alternative derivation of relation between surface tension and shape fluctuations. The preceding expression for the surface tension can also be derived directly from eq S1. We again describe the instantaneous droplet shape by an ellipsoid with axes a, b, and c. We expand the surface area of the ellipsoid to second order in δa, δb and δc. In terms of spherical polar angles θ and ϕ, the first fundamental form defining the surface area element can be written as

$$\begin{aligned} dA &= \sin\theta \, \big(a^2 \, b^2 \, \cos^2\theta + b^2 \, c^2 \, \sin^2\theta \, \cos^2\phi \\ &\quad + a^2 \, c^2 \, \sin^2\theta \, \sin^2\phi \big)^{1/2} \end{aligned} \tag{S18}$$

The area of the ellipsoid then becomes

$$A = \int_0^{\pi} d\theta \int_0^{2\pi} d\phi \, dA \tag{S19}$$

In our statistical mechanical model of droplet shape fluctuations, we (1) impose droplet volume conservation by setting $c = R^3/(ab)$, (2) introduce new variables $a = R + (\delta u + \delta v)/2$ and $b = R + (\delta u - \delta v)/2$, (3) expand the area element dA to second order in δu and δv, and (4) integrate over the polar angles to obtain an expression for the area of the ellipsoid to second order in δu and δv. In this way, we arrive at

$$A = 4\pi R^2 + \frac{8\pi}{5}\delta u^2 + \frac{8\pi}{15}\delta v^2 \tag{S20}$$

Substituting the area change $\delta A = A - 4\pi R^2$ into eq S1 for the corresponding potential energy, we find that the fluctuations in $\delta u = \delta a + \delta b$ and $\delta v = \delta a - \delta b$ are thus harmonic and uncoupled. The equipartition theorem combined with the surface energy eq S1 then gives us the expressions for the surface tension in eqs 3 and 4 of the main text.

Droplet shape

We described the droplet shape in terms of a general ellipsoid with axis lengths a, b, c. We estimated a, b and c from principal component analysis (PCA) of the mass distribution. Let $\boldsymbol{r}_{\mathrm{CMS}}$ be the center of mass of the atoms i with positions $\boldsymbol{r}_i$ and masses m_i in the droplet (excluding water, ions and other small molecules in the solvent):

$$\boldsymbol{r}_{\mathrm{CMS}} = \frac{\sum_i \boldsymbol{r}_i m_i}{\sum_i m_i} \tag{S21}$$

For simplicity, we gave all protein beads in MARTINI an equal mass weight $m_i = 1$. We then calculated a 3×3 mass-weighted covariance matrix of the Cartesian positions $\boldsymbol{r}_i = (r_i^1, r_i^2, r_i^3)^T$ with elements

$$C_{\alpha,\beta} = \frac{m_i(r_i^\alpha - r_{CMS}^\alpha)(r_i^\beta - r_{CMS}^\beta)}{\sum_i m_i} \tag{S22}$$

The eigenvalues λ_1, λ_2, and λ_3 of the matrix $\boldsymbol{C}$ are proportional to the squared ellipsoidal axes: $\lambda_1 = \nu a^2$, $\lambda_2 = \nu b^2$, $\lambda_3 = \nu c^2$. By imposing the condition of constant volume $R^3 = abc$, we eliminate ν:

$$a = \frac{R\lambda_1^{1/3}}{(\lambda_2\lambda_3)^{1/6}} \tag{S23}$$

$$b = \frac{R\lambda_2^{1/3}}{(\lambda_1\lambda_3)^{1/6}} \tag{S24}$$

$$c = \frac{R\lambda_3^{1/3}}{(\lambda_1\lambda_2)^{1/6}} \tag{S25}$$

where R is the average radius of the droplet, obtained from a sigmoidal fit of the density profiles (eq 2 of the main text). If the instantaneous droplet shape has ellipsoidal axes $a_1(t)$, $a_2(t)$, $a_3(t)$ at time t along an MD trajectory, the instantaneous axis fluctuations are then $\delta a_i(t) = a_i(t) - R$. From the deviations of the instantaneous axes from the mean droplet

radius averaged over the three independent combinations of axes, we estimate the variances

$$\left\langle (\delta a \pm \delta b)^2 \right\rangle = \frac{1}{3} \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left\langle (\delta a_i \pm \delta a_j)^2 \right\rangle \tag{S26}$$

Substituting these averages into eqs 3 and of the main text, we arrive at the following two independent estimates of the surface tension

$$\gamma_{20} = \frac{15 k_{\mathrm{B}} T}{16\pi \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left\langle (\delta a_i + \delta a_j)^2 \right\rangle} \tag{S27}$$

$$\gamma_{22} = \frac{45 k_{\mathrm{B}} T}{16\pi \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left\langle (\delta a_i - \delta a_j)^2 \right\rangle} \tag{S28}$$

We expect that the surface-tension estimates are consistent, $\gamma = \gamma_{20} = \gamma_{22}$.

Supporting Tables

Table S1: Amino-acid sequence of the 163-residue FUS low complexity domain.

```
  1-MASNDYTQQA  TQSYGAYPTQ  PGQGYSQQSS  QPYGQQSYSG  YSQSTDTSGY-50
 51-GQSSYSSYGQ  SQNTGYGTQS  TPQGYGSTGG  YGSSQSSQSS  YGQQSSYPGY-100
101-GQQPAPSSTS  GSYGSSSQSS  SYGQPQSGSY  SQQPSYGGQQ  QSYGQQQSYN-150
151-PPQGYGQQNQ  YNS-163
```

Table S2: Simulation characteristics. Listed are the starting configuration, the width of the cubic box, the number of FUS-LCD chains, their overall concentration, and the α interaction rescaling parameters. Slab box size "*" corresponds to $90 \times 20 \times 20\,\mathrm{nm}^3$. A prime in the α column indicates a run length of 12 µs instead of 24 µs.

Starting configuration	Box width (nm)	Number of proteins	Concentration (mg/mL)	Time (µs)	α
Homogeneous	40	134	60	12	0.1, 0.2, 0.3, 0.5, 0.55, 0.6, 0.65, 0.7
		336	150	12	0.1, 0.2, 0.3, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 1
		672	300	12	0.1, 0.2, 0.3, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 1
Preformed droplet	40	134	60	12	0.5, 0.55, 0.65, 0.7, 0.75, 0.8, 0.85
	50	50	11	24	$0.6'$, $0.625'$, $0.65'$, 0.7, $0.75'$
		100	23	24	0.625, $0.65'$, 0.7, $0.75'$
		134	31	24	$0.6'$, 0.625, 0.65, 0.7, $0.75'$
	60	200	26	24	0.625, 0.65, 0.7, 0.75
Single FUS chain	20	1	3.6	6×5	0.6, 0.625, 0.65, 0.7, 0.75
Slab	*	100	79	36	0.6, 0.625, 0.65, 0.7

Table S3: Trajectory segments used for the surface tension calculation (in µs) for simulations with different numbers N of FUS-LCD proteins and different scaling factors α. In these segments, the droplet had relaxed and, at $\alpha = 0.625$, was clearly discernible.

	Shape fluctuations				Interfacial width			
α	$N = 50$	100	134	200	$N = 50$	100	134	200
0.625	–	–	–	–	6-12	6-12, 22-24	6-24	6-24
0.65	10-12	10-12	6-24	6-24	6-12	10.5-12	16-24	18-24
0.7	16-24	16-24	15-24	6-24	16-24	15-24	6-24	6-24
0.75	4-12	7-12	6-12	11-12	8.7-12	10-12	8.7-12	6-12

Supporting movies

Movie S1: Phase separation of FUS at $\alpha = 0.7$ leading to the formation of a spherical droplet starting from a homogeneous solution of $N = 134$ proteins in a $40 \times 40 \times 40\,\mathrm{nm}^3$ box. The system adopts different structures, being first a "tube" spanning the box that then ruptures to form a sphere. Shown are the FUS-LCD chains in different colors, with water and ions omitted for clarity. The length of the shown trajectory is 12 µs.

Movie S2: Water inside phase-separated FUS droplet for $\alpha = 0.7$ with $N = 134$ proteins. In this MD simulation snapshot, FUS-LCD proteins are shown in red, and water beads in light blue. A clip plane for the protein is moved through the droplet, leaving behind the exposed water beads to reveal the hydration gradually, ending with a "droplet of water".

References

(1) Henderson, J.; Lekner, J. Surface Oscillations and the Surface Thickness of Classical and Quantum Droplets. *Mol. Phys.* **1978**, *36*, 781–789.

Supporting Figures

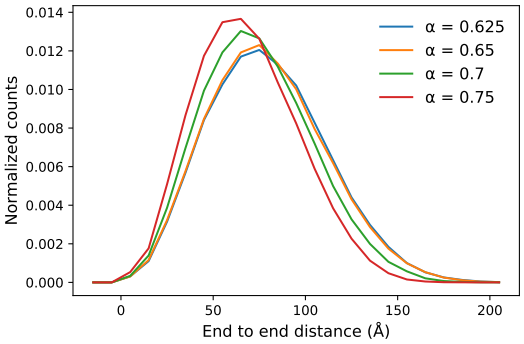

Figure S1: Distribution of end-to-end distances at various values of α in MD simulations with $N = 134$ proteins.

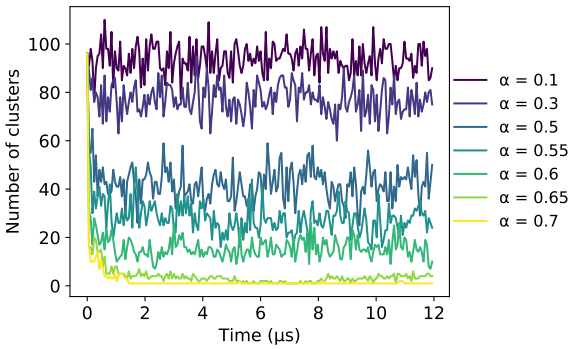

Figure S2: Cluster formation. The number of FUS-LCD clusters is plotted as function of time for different α. The simulations started from a homogeneous solution of 134 proteins in a $40 \times 40 \times 40\,\mathrm{nm}^3$ box, equilibrated with $\alpha = 0$. During the first 2 µs, following the increase of α from zero to the desired value at time zero, larger clusters formed and, as a result, the overall number of clusters decreased.

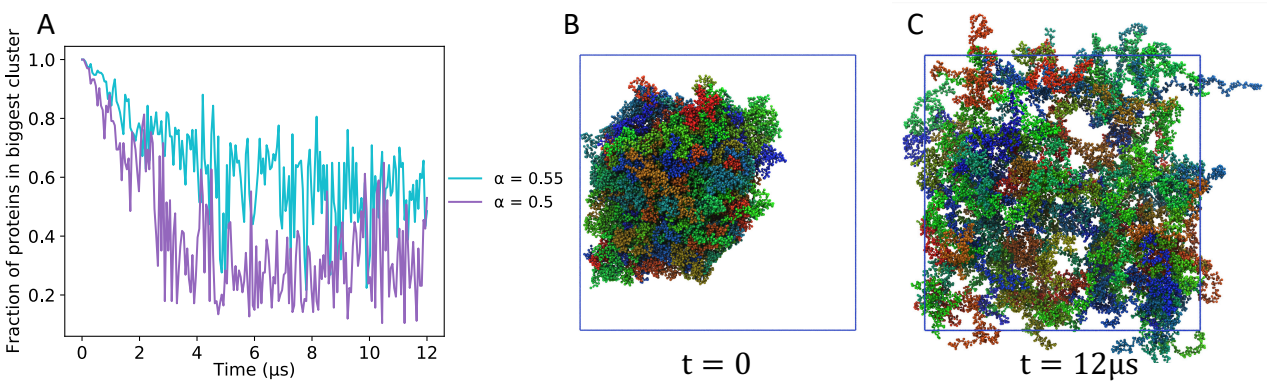

Figure S3: Reversibility of the FUS-LCD phase separation. (A) Fraction of proteins in the largest cluster as a function of time for $\alpha = 0.5$ (magenta) and $\alpha = 0.55$ (light blue). The MD simulations started with a droplet preformed at $\alpha = 0.7$ with $N = 134$ proteins. This droplet dissolved after α was lowered to 0.55 or 0.5. (B) Starting configuration with a preformed droplet. (C) Final configuration at $\alpha = 0.5$ with dispersed FUS LCD.

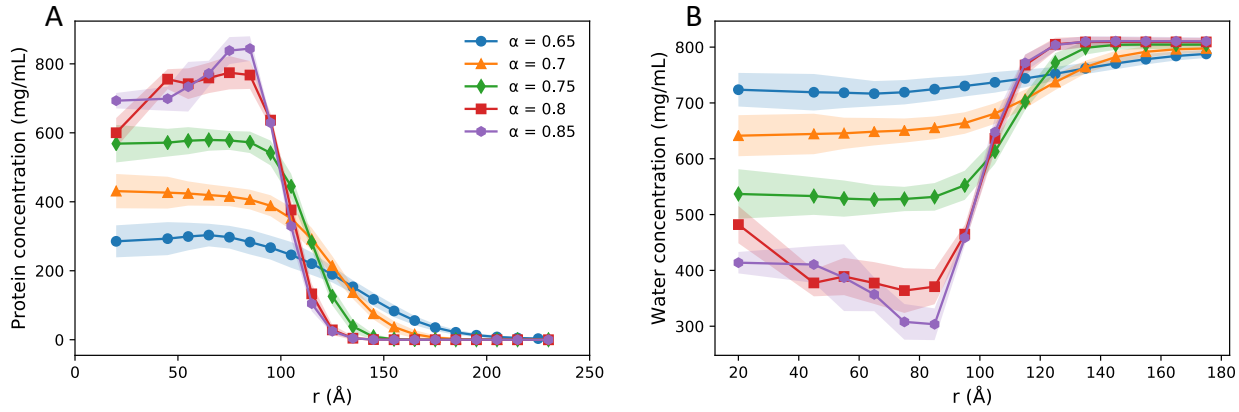

Figure S4: Protein concentration inhomogeneities in FUS-LCD droplets for $\alpha > 0.75$. (A) Radial protein density profiles for α between 0.65 and 0.85. (B) Corresponding water density profiles.

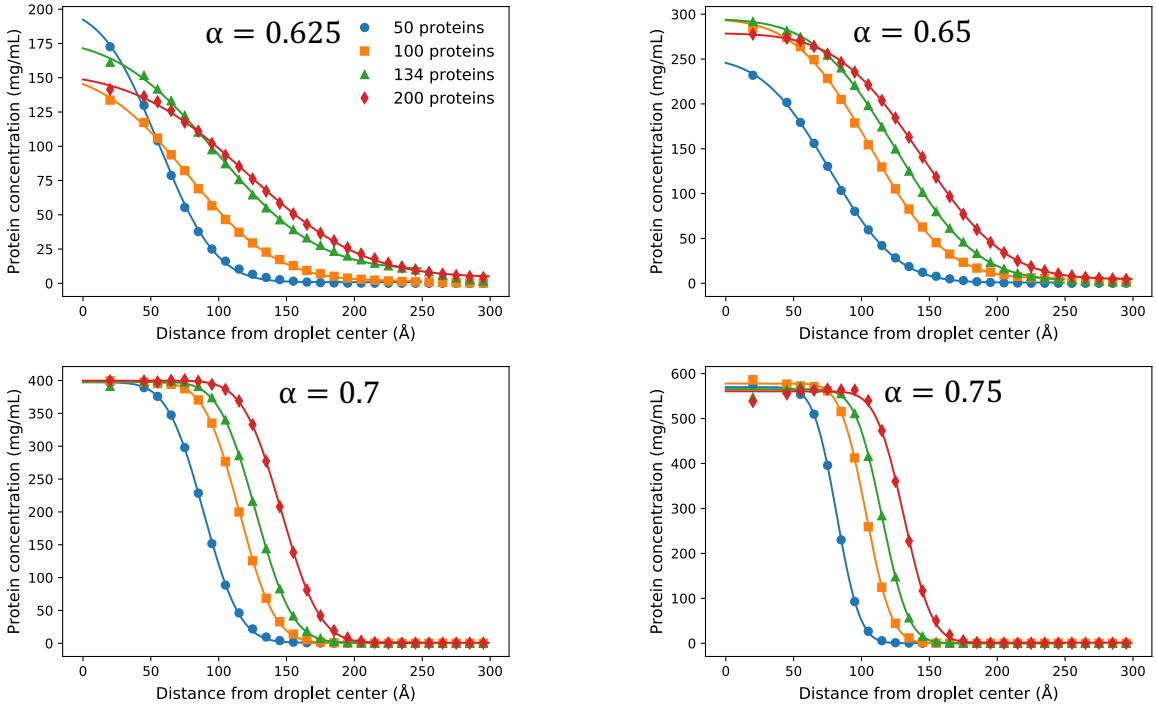

Figure S5: Radial protein concentration profiles as function of distance from droplet center (symbols: simulation results; lines: fit to error function density profile). Results are shown for MD simulations with different total numbers N of proteins in a box of volume $50 \times 50 \times 50\,\mathrm{nm}^3$ for $N = 50, 100, 134$ proteins and $60 \times 60 \times 60\,\mathrm{nm}^3$ for $N = 200$ proteins, and for different values of α.

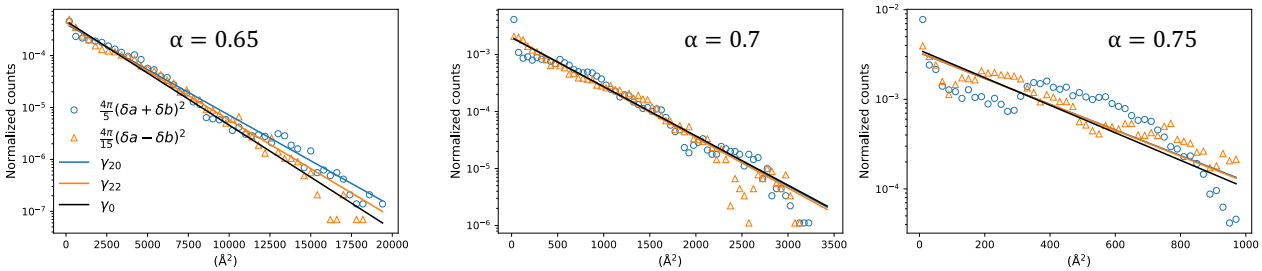

Figure S6: Droplet shape fluctuations. Shown are the normalized histograms of the scaled ellipsoidal mode amplitudes, $|\zeta_{20}|^2$ and $|\zeta_{22}|^2 + |\zeta_{2,-2}|^2$, for a droplet made of 134 proteins at different values of α. The lines are predictions from the capillary wave models, using the different estimates of the surface tension as input, as indicated. Deviations at high values of α (right) are likely the result of slow shape relaxations of the viscous drops on the MD timescale.

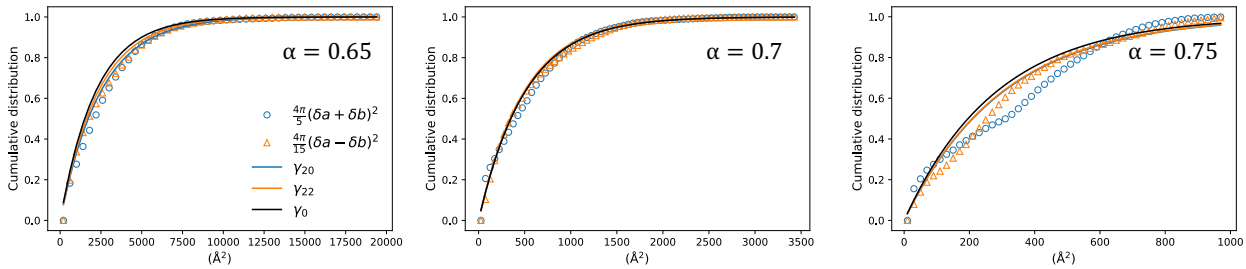

Figure S7: Cumulative distributions of the scaled ellipsoidal modes amplitude $|\zeta_{20}|^2$ and $|\zeta_{22}|^2 + |\zeta_{2,-2}|^2$, for a droplet made of 134 proteins at different α. The lines are predictions from the capillary wave models. See Figure S6 for the corresponding histograms.

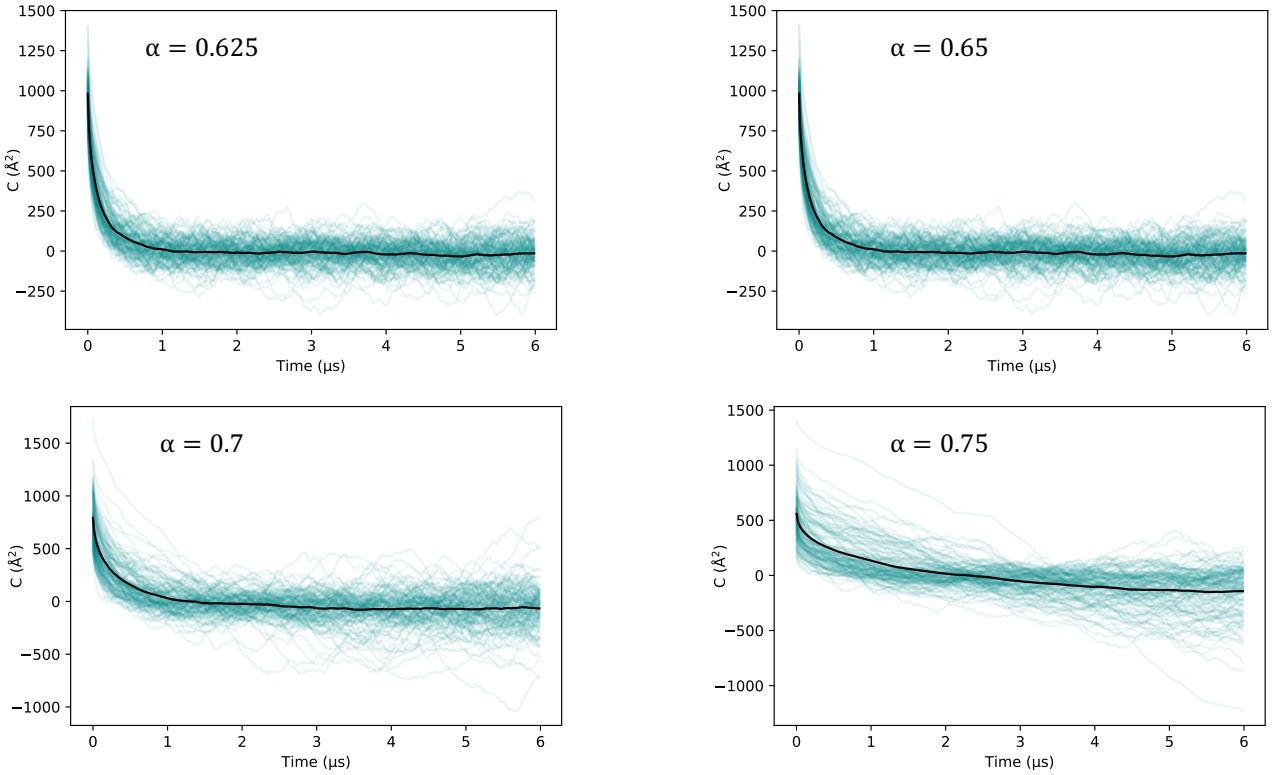

Figure S8: Autocorrelation functions of the FUS-LCD end-to-end distance at different values of α. Thin turquoise lines are the results for individual FUS-LCD chains. The average of the autocorrelation functions across the ensemble of 134 protein chains is shown in black.

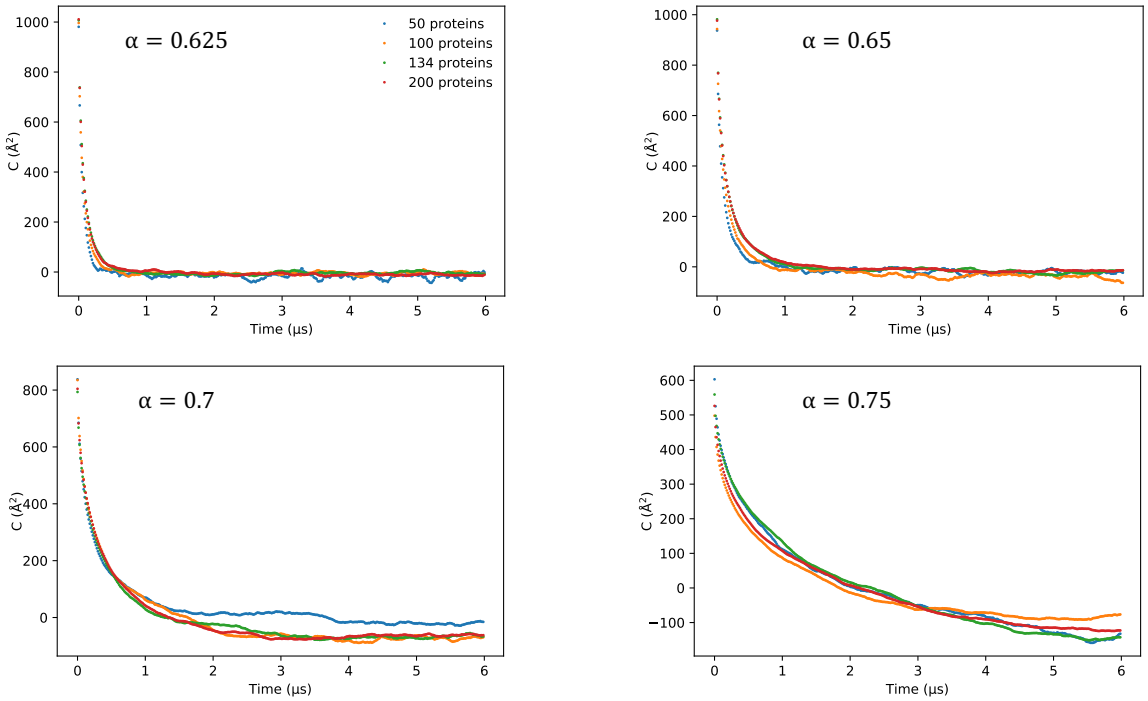

Figure S9: Average of the FUS-LCD end-to-end distance autocorrelation function across the ensemble of chains at various values of α and for different numbers of proteins.

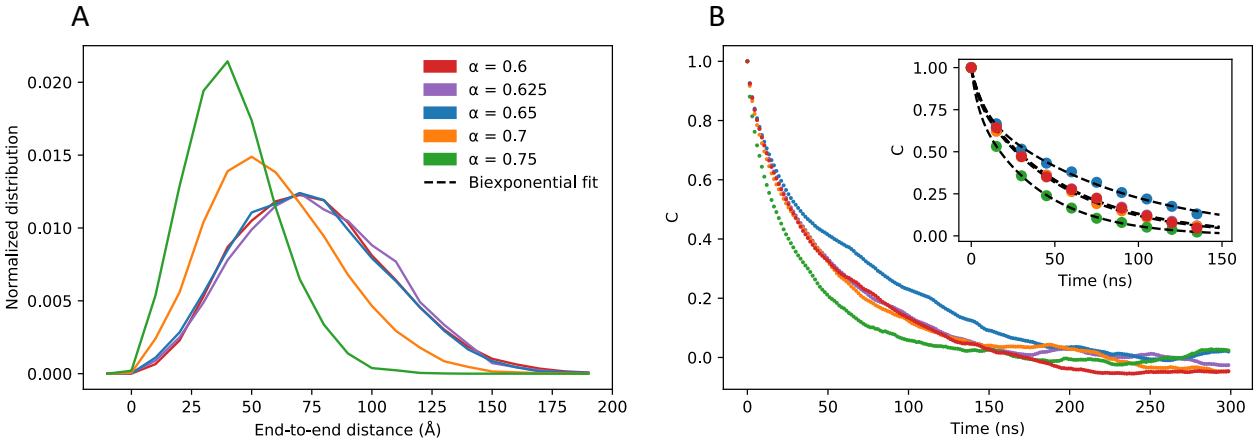

Figure S10: Determination of effective diffusion coefficient for end-to-end distance of isolated FUS-LCD chains. (A) Normalized distributions of the end-to-end distances of an isolated chain free in solution at different values of α. Note that chain compaction sets in for $\alpha \geq 0.7$. (B) Autocorrelation functions of the end-to-end distances of isolated chains, averaged over five independent runs of 6 µs each. The inset shows the biexponential fits from which the relaxation times were computed (crosses in Figure 8 of the main text).

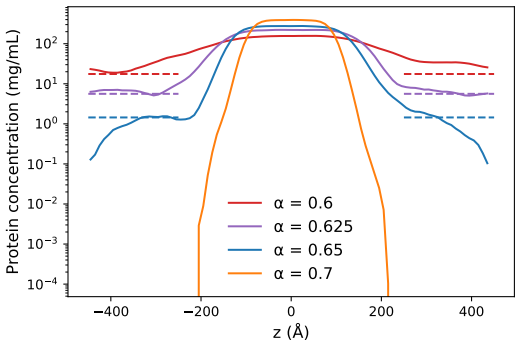

Figure S11: Slab concentration profiles. The dashed lines show the dilute-phase concentrations obtained from clustering. Results are shown for MD simulations with 100 proteins in a box of volume $20 \times 20 \times 90\,\mathrm{nm}^3$ and for different values of α, as indicated.

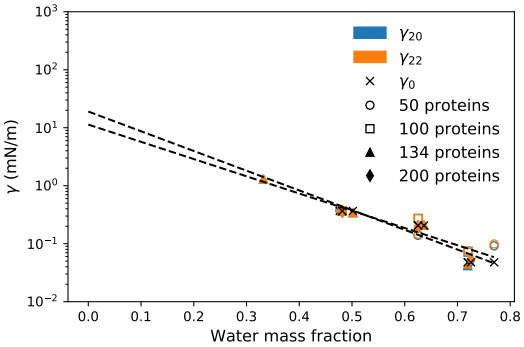

Figure S12: Connection between surface tension and hydration. Results are shown for $\alpha = 0.8, 0.75, 0.7$ and 0.65, which appear as distinct groups of points from left to right. The black dashed lines are exponential fits to γ_{20} and γ_{22}, and to γ_0, in an effort to extrapolate the surface tension to zero water content (i.e., a dry droplet).