bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nucleation of Biomolecular Condensates from Finite-Sized Simulations

Lunna Li,[†,¶] Matteo Paloni,[‡,¶] Aaron R. Finney,[†] Alessandro Barducci,[*,‡] and Matteo Salvalaglio[*,†]

†*Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, UK.*
‡*Centre de Biologie Structurale (CBS), INSERM, CNRS, Université de Montpellier, Montpellier, France.*
¶*Contributed equally to the manuscript as co-first authors.*

E-mail: alessandro.barducci@cbs.cnrs.fr; m.salvalaglio@ucl.ac.uk

Abstract

The nucleation of protein condensates is a concentration-driven process of assembly. When condensation is modelled in the canonical ensemble - where the total number of peptides is constant - its dynamics and thermodynamics are influenced by finite-size effects, resulting in the formation of a single, stable condensate droplet. Here, we take advantage of a general theoretical description of the thermodynamics of condensate droplet formation in the canonical ensemble to obtain information on the thermodynamics and kinetics of nucleation in the macroscopic limit. We apply our approach to two phase-separating systems with different physicochemical characteristics: NDDX4 and FUS-LC. From the properties of the steady-state condensate droplet obtained in finite-size, coarse-grained molecular simulations of these two peptides, we estimate the macroscopic equilibrium density of the dilute protein solution and the surface tension of the condensates. Analyses of nucleation free energy barriers reveal that NDDX4 dilute phases are kinetically unstable over a wide range of concentrations. FUS-LC condensation, on the other hand, is consistently associated with activated nucleation events. Differences in the behaviour of the two systems can be explained by different contributions to bulk and surface free energies of the emerging phases, which correlate with single-chain hydrophilicity and the dynamics of monomer exchange between the condensate and lean solution. The approach presented here is general and provides a straightforward and efficient route to obtain emergent, ensemble properties of protein condensates from finite-sized nucleation simulations.

Biomolecular compartments that are not bound by membranes have attracted a lot of attention in the last decade because of their important role in cellular organization.[1,2] The assembly of these membraneless organelles (MLOs) is driven by the formation of dynamical multivalent interactions between proteins and/or nucleic acids.[1,2] Notably, many proteins involved in forming such compartments are either intrinsically disordered or have highly-flexible domains: typical examples include proteins from the DEAD-box[3–5] and FET[6–8] families. The interactions between biomacromolecules in the formation of MLOs was described according to the stickers-and-spacers framework, derived for associative polymers.[7,9] In this framework, polymer chains are characterized by multivalent domains or motifs, named stickers, that govern intermolecular interactions, interspersed by spacer domains that influence the material properties of the condensates.[7] Experiments of simplified systems with one to a few of these disordered protein regions have shown that they can lead to the formation of assemblies via a process of liquid-liquid phase separation (LLPS).[3,4,6] Still, the phenomenon in cells could be more complex, involving different molecular mechanisms.[10–12]

Molecular simulations of simplified systems have provided important insights into the mechanisms and molecular drivers for the formation of biomolecular assemblies. Explicit-solvent molecular dynamics (MD) simulations offer a detailed picture of the structure of the intermolecular interactions and of the relationship between local structure and phase separation propensity.[13–16] Still, system sizes and time scales that can be investigated using atomistic MD are severely limited. To alleviate these difficulties, several Coarse-Grained (CG) models with a one-bead-per-residue resolution were proposed[17–21] and successfully applied to study phase separating systems, establish coexistence conditions, and the effect of mutations and post-translational modifications on phase separation.[17,21,22] Unfortunately, even CG simulations suffer from size limitations and particular strategies, such as the slab method,[17] have to be adopted for minimizing finite-size effects in the study of phase-separation processes. Indeed, in finite-sized systems, the free-energy change associated with the assembly of a condensate droplet is a function of both the concentration and the total volume of the system.[23–27] This dependence emerges from the fact that, in small volumes,[28] the chemical

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

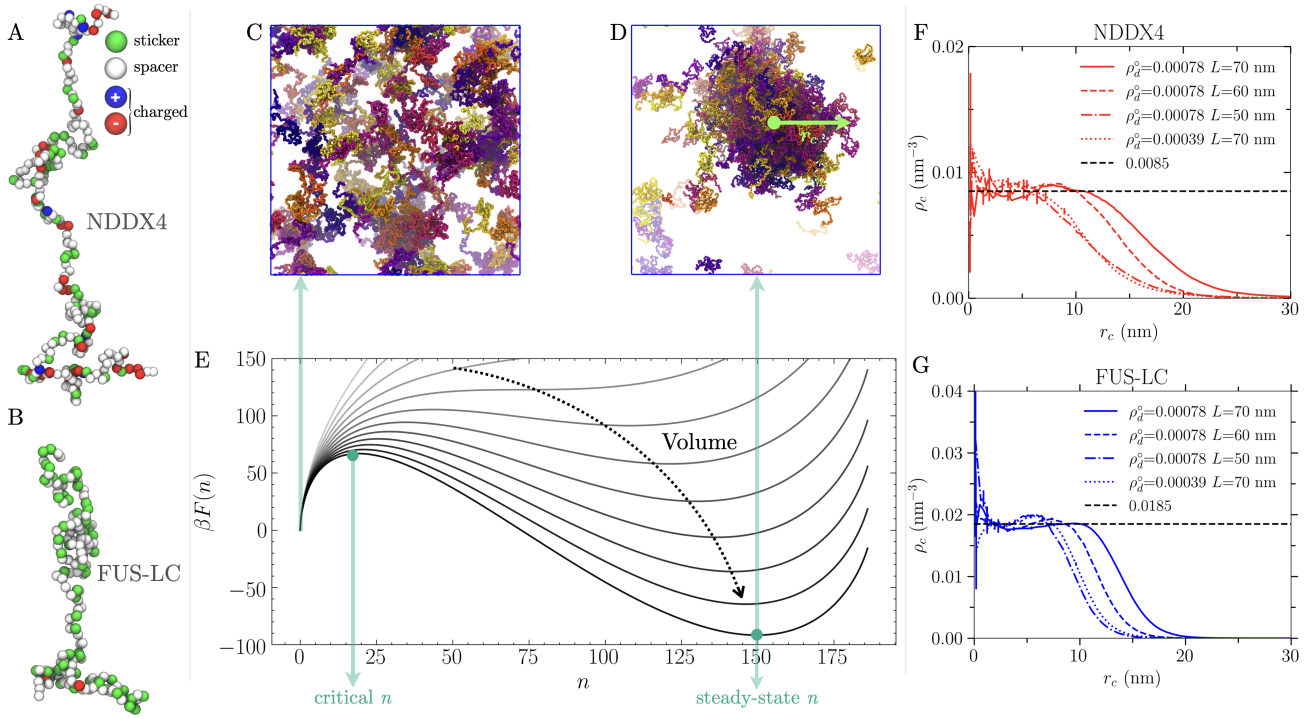


Figure 1: Coarse-grained modelling of the nucleation of bio-molecular condensates in the NVT ensemble. **A-B)** Coarse-grained model of the NDDX4 and FUS-LC chains. **C)** Example of supersaturated homogeneous dilute phase configuration. **D)** Example of a steady-state configuration containing a stable, condensed phase droplet. **E** Example free energy profiles associated with the nucleation of condensed phase droplets, obtained by keeping supersaturation constant while increasing the total volume. The variable n represents the number of chains in the dense phase; the origin corresponds to a homogeneous dilute phase; the local maximum at small n corresponds to the critical nucleus size; and the local minimum at large n corresponds to the steady-state droplet size. **F, G)** Steady-state droplet radial density profiles obtained from simulations performed in different box sizes for NDDX4 (**F**) and FUS-LC (**G**).

potential of the environment surrounding a condensate droplet depends on its size, leading to qualitative and quantitative differences compared to its macroscopic counterpart.[24,27,29] An elegant approach to account for finite-size effects is the Modified Liquid Droplet (MLD) model,[23,24] which provides an expression for the nucleation free energy $F(n)$ in the canonical ensemble under the same set of assumptions typically adopted by classical nucleation theory (CNT). Expressed as a function of the density of the dilute (ρ_d) and condensed (ρ_c) phases, the MLD $F(n)$ reads:

$$\beta F(n) = -n \ln \frac{\rho_d^n}{\rho_d^*} + \beta\sigma A(n) + n(1 - \rho_d^*/\rho_c) + N \ln \frac{\rho_d^n}{\rho_d^\circ} \quad (1)$$

where $\beta = 1/kT$, k is the Boltzmann constant, T is the temperature, N is the total number of chains contained in a simulation box of volume V, n is the number of chains in the condensate droplet, σ is the planar surface tension of the condensed phase, $A(n)$ is the surface area of a droplet of condensed phase formed by n chains, ρ_c is the equilibrium molar density of the condensed phase, ρ_d^* is the equilibrium molar density of the dilute phase, ρ_d° is the density of the dilute phase at n=0, and ρ_d^n is the density of the dilute phase in a system where n chains form a condensed phase droplet at constant N and V:

$$\rho_d^n = \frac{N - n}{V - n/\rho_c}. \quad (2)$$

Examples of $F(n)$ profiles for systems at of increasing size are represented in Fig. 1E, where the stationary points correspond to the critical nucleus size n^* and to the size of the self-limiting steady-state droplet n_{ss}, the values of which both increase in magnitude as the volume and the number of molecules in the system increases. For any concentration, one can always identify a threshold volume below which condensation is inhibited by finite-size effects since $F(n)$, becomes a monotonically increasing function of droplet size, n.[23–27]

Here we adopt this theoretical framework to fully account for the finite-size effects described above to characterize the thermodynamics and kinetics of biomolecular condensation in MD simulations with tractable system sizes. We demonstrate the potential of our approach by focusing on two intrinsically-disordered protein domains: FUS-LC and NDDX4, which are popular model systems for investigating biomolecular condensates.[4,7] While both proteins undergo LLPS at ambient temperature, they display markedly distinct physicochemical characteristics. Notably, FUS-LC is enriched

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

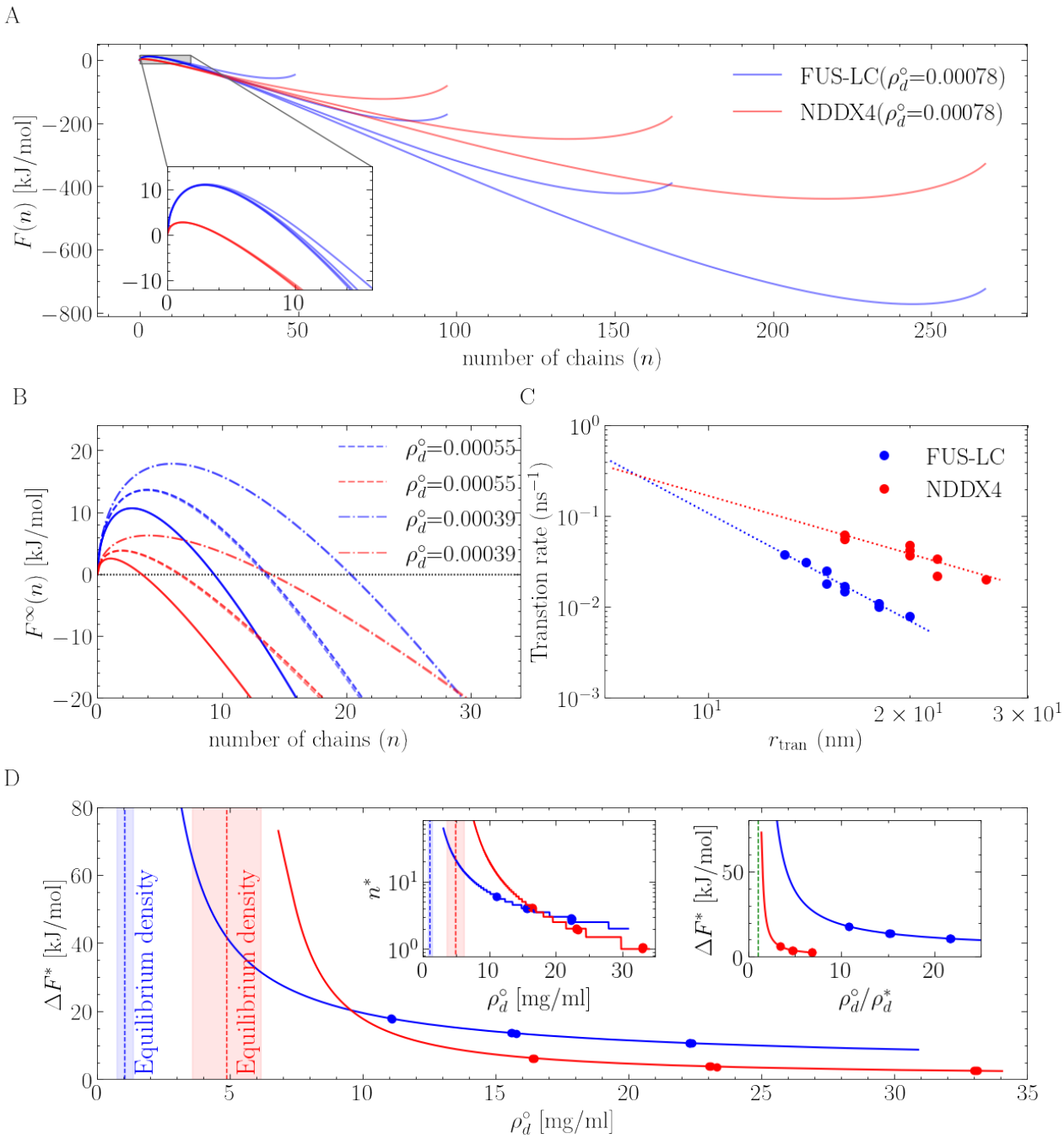


Figure 2: Protein condensate nucleation simulations for NDDX4 (red) and FUS-LC (blue). **A)** Free energy profiles associated with the nucleation of condensed phase droplets. The variable n represents the number of chains in the condensed phase. **B)** Macroscopic free energy profiles, with the effect of the artificial confinement associated with NVT simulations removed. For the same peptide, curves of increasing barriers represent reducing bulk densities/supersaturation. **C)** The rate of condensed-to-dilute transition for droplets of different sizes and the exponential fit. The rate is calculated as the reverse of the mean-first-passage-time of the condensed-to-dilute transition from a Markov-state model (see Figs. S7 and S8). **D)** Estimates of free energy barriers and critical nucleus sizes (left inset) at different bulk densities/supersaturation. The equilibrium density of the dilute phase is reported as a dashed line, with the shaded area representing the 95% confidence interval computed from a bootstrap analysis. The right inset of **D)** is a representation of **D)** plotted in supersaturation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in aromatic and/or polar residues Gln and Ser[7] whereas the NDDX4 sequence is relatively abundant in charged residues organized in patches of opposite signs, significantly contributing to its condensation behavior.[4]

We model these proteins using a sequence-specific CG model with a one-bead-per-residue resolution. Consecutive amino acids are connected using a harmonic potential with an equilibrium distance of 0.38 nm and a spring constant of 1×10^3 kJ/mol/nm^2. Non-bonded amino acids interact through electrostatic interactions and short-range contact potentials. Electrostatic interactions between charged amino acids are described with a Debye-Hückel potential with a screening length of 1 nm, corresponding approximately to an ionic strength of 100 mM. Short-range non-bonded interactions are modelled using a Lennard-Jones potential with residue-dependent σ and ϵ (see Tab. S2) inspired by the stickers-spacers framework.[7,30] In this respect, we take advantage of mutagenesis studies that suggested a key role of amino-acids with large-sized aromatic or planar side chains in driving the phase separation of our model systems.[4,7,31] Thus we define Arg, Phe, Tyr, Trp and Gln residues as Stickers (St) and all the others as Spacers (Sp), and we set the LJ potential as $\epsilon_{St-St} = 1.5\epsilon_{St-Sp} = 3\epsilon_{Sp-Sp}$. The absolute energy scale of these short-range interactions is the only free parameter of our model and we tuned it to reproduce the experimental densities of NDDX4 and FUS-LC proteins.[7,31,32] Functional forms of the interactions and parameters of the CG model are available in SI.

Using this CG potential, we performed a large set of NVT simulations of NDDX4 or FUS-LC to probe condensation from homogeneous solutions at diverse densities and sizes of the systems in standard cubic boxes (see Table 1). For all the systems, we performed for 1 μs to equilibrate the systems, followed by an additional 1 μs for analyses of the resulting steady state. All the simulations were conducted at a temperature of 300 K and an ionic strength of 100 mM in GROMACS 2019.4 (see the SI for further details). We simulated supersaturation regimes sufficient to allow condensed phase formation within a reasonable simulation time, while avoiding extremely high concentrations that may result in large condensates spanning across simulation box periodic boundaries.

In the majority of simulations, small droplet condensates of stationary size n_{ss} - corresponding to local minima in Fig. 1E - were observed. We identified droplets according to a two-dimensional geometric criterion, considering the protein inter-chain contacts and distances between the centres of mass (COM) of individual chains and the droplet COM (see Supporting Information). Analysis of protein radial number density profiles from the COM of steady-state droplets (see Figs. 1 F, G) indicate that densities in the core are independent of the size and overall concentration of the simulated system. The mean densities in this region provide a robust estimate of the equilibrium dense-phase densities for NDDX4 (0.0085 nm^{-3}) and FUS-LC (0.0185 nm^{-3}), which are in good agreement with values obtained by slab coexistence simulations and experimental results. Conversely, the direct estimate of dilute-phase density from finite-volume nucleation simulations can suffer from artefacts that result in errors that can scale as $V^{1/4}$.[33]

We, therefore, rely on the MLD framework that, for finite-sized systems, provides the following Gibbs-Thomson/Kelvin equation for the equilibrium density of the dilute phase:[23]

$$\rho_d^{n_{ss}} = \rho_d^* \exp \frac{2\beta\sigma}{\rho_c r_{ss}} \tag{3}$$

where $\rho_d^{n_{ss}} = \frac{(N-n_{ss})}{(V-n_{ss}/\rho_c)}$, and r_{ss} is the radius of the steady-state droplet. Under the assumption of spherical droplets, one can explicitly introduce the dependence of the coexistence pressure and the droplet radius r_{ss} on the number of chains in the droplet and reformulate Eq. 3 as:

$$\rho_d^* \exp\left[\frac{2\beta\sigma}{\rho_c\left(\frac{3n_{ss}}{4\pi\rho_c}\right)^{1/3}}\right] - \frac{(N-n_{ss})}{(V-n_{ss}/\rho_c)} = 0 \tag{4}$$

. This equation provides the opportunity to fully characterize the assembly process by determining all the key thermodynamic quantities from our finite-size simulations. Indeed, in Eq. 4 N, V and T are defined by the simulation setup while ρ_c and n_{ss} can be directly obtained from the analysis of the droplet density profiles (see Figs. 1F, G; and additional details reported in SI). Most importantly, the dilute phase density (ρ_d^*) and the surface tension (σ) can be computed by a global fitting of Eq.4 on the data from our set of simulations performed at different values of N and V. Using this strategy, we estimated ρ_d^* to be 1.03 ± 0.32 mg/ml for FUS and 4.87 ± 1.44 mg/ml for NDDX4 (error bars indicate 95% confidence intervals from bootstrapped results). Both densities are in excellent agreement with estimates obtained using slab coexistence simulations based on the same CG model: 1.42 ± 0.30 mg/ml and 4.78 ± 0.88 mg/ml for FUS-LC and NDDX4, respectively. Both the slab and nucleation results are comparable to the experimental dilute-phase values of 2 mg/ml[7] and 7 mg/ml.[31] The surface tension σ is estimated to be 0.37 ± 0.11 mN/m, for FUS-LC and 0.101 ± 0.06 mN/m for NDDX4, reflecting the higher hydrophilic character of NDDX4 compared with FUS-LC. Notably, our FUS-LC surface tension estimate agrees with calculations performed by Benayad et al.[34] for an explicit solvent FUS-LC coarse grained model, which identified the surface tension in the 0.01–0.4 mN/m range from fluctuations of the droplet shape and from the broadening of the interface between phases.

To validate the equilibrium density and surface tension parameters obtained from fitting simulation data with Eq. 4, we compare the position of the minima in the function $F(n)$ (Eq. 1), parameterised with ρ_d^* and σ

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Modified-liquid-droplet nucleation simulations

System	L(nm)	N	ρ_d°(nm^{-3})	n_{ss}
NDDX4-1	40	25	0.00039	×
NDDX4-2	40	35	0.00055	△
NDDX4-3	40	50	0.00078	△
NDDX4-4	50	49	0.00039	△
NDDX4-5	50	69	0.00055	45(2)
NDDX4-6	50	98	0.00078	76(2)
NDDX4-7	60	84	0.00039	42(2)
NDDX4-8	60	118	0.00055	84(3)
NDDX4-9	60	169	0.00078	139(2)
NDDX4-10	70	133	0.00039	71(2)
NDDX4-11	70	187	0.00055	130(3)
NDDX4-12	70	268	0.00078	214(2)
FUS-LC-1	40	25	0.00039	×
FUS-LC-2	40	35	0.00055	△
FUS-LC-3	40	50	0.00078	42(0.5)
FUS-LC-4	50	49	0.00039	△
FUS-LC-5	50	69	0.00055	55(1)
FUS-LC-6	50	98	0.00078	86(1)
FUS-LC-7	60	84	0.00039	63(1)
FUS-LC-8	60	118	0.00055	101(1)
FUS-LC-9	60	169	0.00078	153(1)
FUS-LC-10	70	133	0.00039	103(2)
FUS-LC-11	70	187	0.00055	158(1)
FUS-LC-12	70	268	0.00078	243(2)

n_{ss} is the average steady-state cluster size with block average error in the bracket; △ indicates fluctuating cluster (unstable primary droplet/presence of multiple droplets) or that the box sizes are too small to distinguish dilute/dense phase; × indicates no phase separation.

obtained from fitting, with the steady-state droplet size measured in simulations (see Fig. S5). The excellent agreement shown by the parity line demonstrates the method's power for universal calculations of thermodynamic properties such as surface tension and equilibrium vapour pressure using finite-sized nucleation simulations.

Importantly, the determination of ρ_d^* and σ (as well as ρ_c) enables the calculation of free energy profiles in the limit of an infinitely large simulation box. This is achieved by evaluating Eq. 1 in the limit where $N >> n$, and $V >> nv_\ell$. In this limit, $\rho_d^n \rightarrow \rho_d^\circ$, and Eq. 1 reduces to the typical CNT expression for reversible work of formation of a condensate droplet from a supersaturated dilute phase at constant density ρ_d°:

$$\beta F^\infty(n) = -n\ln\frac{\rho_d^\circ}{\rho_d^*} + \beta\sigma A(n) + n(1-\rho_d^*/\rho_c) \qquad (5)$$

Using this approach, we can thus predict nucleation free energy barriers and critical nucleus sizes under conditions that should be optimal for comparison with experiments without the need for computationally-demanding schemes able to mimic open-boundary/infinite-reservoir macroscopic conditions.[35,36]

For NDDX4 and FUS-LC, these are demonstrated in Fig. 2B at the simulated densities. From the nucleation free energy profiles, we can compare nucleation kinetics in the limit of a macroscopic open system. For instance, the nucleation barriers for FUS-LC are consistently higher and are associated with larger critical nuclei cf. NDDX4 (see Fig. 2B) under the conditions studied. We note, however, that using a CNT-based model for nucleation with thermodynamic parameters evaluated using our computational approach indicates a cross-over in the nucleation free energy barrier (and thus in the nucleation rates) for densities lower than those explicitly simulated. This can be seen in Fig. 2D, where the nucleation free energy barrier for FUS-LC becomes lower than that of NDDX4 below 10 mg/mL. The critical nuclei sizes also display a crossover, observed at higher density values of approximately 16 mg/mL. The difference in the crossover density between free energy barrier height and critical nucleus size reflects the differences in the balance between surface and bulk free energy terms for NDDX4 and FUS-LC.

Differences in the physicochemical character of NDDX4 and FUS-LC are further reflected by their different solubility (captured by ρ_d^*), which induces different supersaturation levels at the same bulk densities. Approaching the binodal line (the green dashed line in the right inset of Fig. 2D, corresponding to $\rho_d/\rho_d^* = 1$), both the critical nucleus size and ΔF^* diverge. At these conditions, directly observing nucleation events is extremely unlikely, even in the limit of very large simulations.[29] As such, complete information on the nucleation behavior approaching the binodal can only be inferred from theory. Fig. 2 demonstrates how our methodology provides a direct route to extract the macroscopic thermodynamic parameters that enable such characterisation from simple finite-sized simulations.

Nucleation simulations also provide an extensive sampling of condensed phase droplets in dynamic equilibrium with the dilute phase. We use the data provided by the finite-sized nucleation simulations to analyse the dynamics of single-chain exchange between the condensed and dilute phases using Markov state models (MSMs). The details of the MSM construction and analysis are reported in Supporting Information. Briefly, in the MSMs, the dilute and condensed states are identified based on the inter-chain contacts and on the distance between the COM of the individual chains and the droplet (see Figs. S7, S8). As a measure of the exchange dynamics, we compute the rate of escape of a single chain from the condensed to the dilute phase as a function of the droplet size (see Table S1, Fig. S8). Fig. 2C shows an approximately exponential decay in the condensed-to-dilute transition rate as a function of droplet size, with FUS-LC demonstrating a steeper decay than NDDX4. Moreover, irrespective of the droplet size, it is slower to transfer a FUS-LC chain across the phase boundary compared with NDDX4. The estimated rate of dilute-to-condensed transitions appears instead largely uncorrelated with respect to the droplet

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

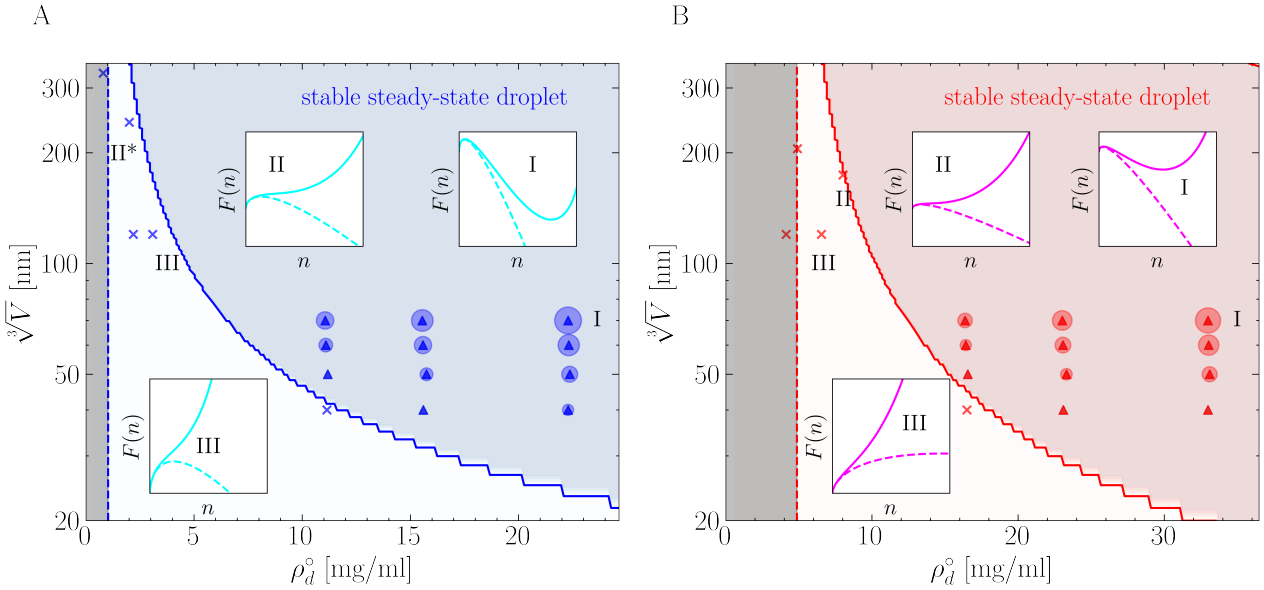


Figure 3: Stability of FUS-LC **A)** and NDDX4 **B)** dense-phase droplets in confined volumes as a function of the total peptide density, predicted using parameters fitted from simulation data from Tables 1 and 2. The grey shaded region represents the region $\rho_d^\circ < \rho_d^*$ where phase separation is thermodynamically unstable, with the dashed lines representing the predicted equilibrium vapour density ρ_d^*. Nucleation is thermodynamically favoured above ρ_d^*. The white region represents the *nominally supersaturated* region of parameter space; however, finite-size effects mean that here, $F(n)$ is a monotonically increasing function, and condensation is inhibited by the effective confinement in the canonical ensemble (see inset III). Simulations performed in these conditions (indicated with $\times$) cannot show nucleation or relax to a homogeneous vapour phase when initialised from a droplet. The blue/red shaded region instead represents the ensemble of conditions in which droplets are thermodynamically stable, characterized by the free energy profile of state II. All simulations performed in this region show phase separation. The simulations used to fit the thermodynamic parameters are represented with circles with sizes proportional to the volume of the corresponding steady-state droplets. The blue/red solid line between the white and the blue/red shaded regions is characterized by the free energy profile of boundary state II. All of the free energy profiles in the insets correspond to simulation data points with the same state labels, except for FUS II, which refers to a slightly higher ρ_d° than the actual data point at II*. Inset states of the same numbering are plotted on the same scale for FUS-LC and NDDX4.

size, but, as expected for a diffusion-dominated process, it fluctuates around the same average for both NDDX4 and FUS-LC. The ratio of the condensed-to-dilute and dilute-to-condensed transition rates is approximately linear with respect to the ratio of the number of peptides in the two phases (see Fig. S6). The faster escape dynamics of single chains from NDDX4 condensate droplets impact fluctuations in their overall size and shape. NDDX4 generally shows slightly larger size fluctuations than FUS-LC (see Table 1), possibly due to a combined effect of longer chain length, different hydrophobicity and faster escape dynamics. In addition, while both NDDX4 and FUS-LC condensate droplets can be effectively approximated as spherical in the theoretical analysis and interpretation of the simulation results, we note that NDDX4 faster exchange dynamics, higher hydrophilicity and more gentle radial density gradients (see Fig. 1, panels F, G) lead to larger deviations from a perfectly spherical shape (see Table S1).

Using the results obtained from nucleation simulations, we can rationalize the effect of finite size on the thermodynamics of phase separation by inserting σ and ρ_d^* into Eq. 1, thus mapping the qualitative nucleation behavior as a function of the total peptide density and system size. Following this strategy, we produce domain diagrams in Fig. 3 that indicate the presence or absence of phase separation for both NDDX4 and FUS-LC. In the blue/red shaded region, exemplified by state I, the finite-sized thermodynamics admit the existence of steady state droplets corresponding to local minima in the nucleation free energy profiles. Increasing the system size and density in this region results in larger droplets, as demonstrated by the size of the circles used to represent our simulation results. In the white region featuring state III, confinement induces a monotonically increasing free energy curve,[24,27] and nucleation will never occur regardless of the simulation time. The blue/red solid line represents the transition boundary, where the free energy curve has a single stationary point corresponding to a flex. This condition is closely approximated by the free energy profile of state II. Simulations initiated on the boundary line from a pre-formed condensate droplet of size close to the sta-

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: Dissolution/nucleation simulations at larger system sizes and lower densities

System	L(nm)	N	ρ_d°(nm^{-3})	Initial condition	Result	Simulation length (μs)
NDDX4-13	120	169	0.00010	NDDX4-9	dissolved	2
NDDX4-14	120	268	0.000155	NDDX4-12	dissolved	4
NDDX4-15	174	1000	0.00019	homogeneous	no nucleation	5
NDDX4-16	205	1000	0.00012	homogeneous	no nucleation	5
FUS-LC-13	120	133	0.00008	FUS-10	dissolved	1
FUS-LC-14	120	187	0.00011	FUS-11	dissolved	6
FUS-LC-15	242	1000	0.00007	homogeneous	no nucleation	5
FUS-LC-16	329	1000	0.00003	homogeneous	no nucleation	5

Initial condition: the final configuration of the system in Table 1 in the increased simulation box, or randomly-distributed homogeneous one-phase initial state. Equilibration time is approximately 1 μs.

tionary point in the free energy profile will experience negligible driving forces to either grow or dissipate.[37] At large system sizes this effect manifests itself as a very slow evolution of the droplet towards the equilibrium state, corresponding to a homogeneous phase with density ρ_d°. Instead, at small system sizes, where far fewer chains are present, the shallow free energy gradient results in large fluctuations of the droplet size. This behavior is confirmed by simulations performed in regimes of volume and peptide density closely approximating these conditions (see Table 2).

Understanding and applying general thermodynamics concepts associated with the precipitation of a dense liquid-like condensate to finite-sized canonical simulations of condensate nucleation is a powerful approach to consistently interpreting the finite-size dependent results of nucleation simulations. This approach yields consistent estimates of thermodynamic properties of protein condensates such as the equilibrium density of the dilute phase in contact with condensates and the surface tension at the dense-dilute phase interface. In turn, information on the nucleation kinetics, such as the critical nucleus size and the free energy nucleation barrier can be obtained. Furthermore, the effect of system size on these thermodynamic and kinetic properties can be clearly demonstrated, which allows a prediction of the phase behavior according to the conditions in microscopic, closed systems.

The method presented in this work demonstrates excellent potential for applications to study phase separation in a wide range of bio-, organic and inorganic systems. On the one hand, our approach provides size-independent thermodynamic parameters from finite-sized simulations. Here we have quantitatively characterized the LLPS of two coarse-grained models of the phase-separating peptides NDDX4 and FUS-LC, establishing thermodynamic properties from their collective behavior under constraints. On the other hand, our approach enables the effect of confinement on phase separations initiated by nucleation to be quantified. This is particularly relevant in the context of biological systems, where phase separations take place in micrometer-scale isolated cellular compartments, as well as in the rational development of technological applications of LLPS for material synthesis in micro/nanofluidic devices.[37,38]

Acknowledgement MS and LL acknowledge the Leverhulme Trust for funding (Project RPG-2019-235), MS and ARF acknowledge funding from the EPSRC Programme Grant Crystallization in the Real World (Grant EP/R018820/1). The authors acknowledge the use of the UCL High-Performance Computing Facilities and associated support services in the completion of this work.

Supporting Information Available

Additional details and figures on the clustering algorithms used to identify the number of chains in the dense phase, calculation of the droplet density profiles, Markov State Models for estimating the exchange kinetics, and force-field parameters are available as Supplementary materials.

References

(1) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.

(2) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. **2017**, *18*, 285–298, Number: 5 Publisher: Nature Publishing Group.

(3) Elbaum-Garfinkle, S.; Kim, Y.; Szczepaniak, K.; Chen, C. C.-H.; Eckmann, C. R.; Myong, S.; Brangwynne, C. P. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences* **2015**, *112*, 7189–7194.

(4) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell* **2015**, *57*, 936–947.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(5) Weis, K. Dead or alive: DEAD-box ATPases as regulators of ribonucleoprotein complex condensation. *Biological Chemistry* **2021**, *402*, 653–661.

(6) 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. **2015**, *60*, 231–241, Publisher: Elsevier.

(7) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688–699.e16.

(8) Borcherds, W.; Bremer, A.; Borgia, M. B.; Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Current opinion in structural biology* **2021**, *67*, 41–50.

(9) 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*, e30294.

(10) A, P.; Weber, S. C. Evidence for and against Liquid-Liquid Phase Separation in the Nucleus. *Non-Coding RNA* **2019**, *5*.

(11) Musacchio, A. On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal* **2022**, *41*, e109952.

(12) Mittag, T.; Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. **2022**, *82*, 2201–2214.

(13) Paloni, M.; Bailly, R.; Ciandrini, L.; Barducci, A. Unraveling Molecular Interactions in Liquid–Liquid Phase Separation of Disordered Proteins by Atomistic Simulations. **2020**, *124*, 9009–9016.

(14) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. **2020**, *124*, 11671–11679.

(15) Conicella, A. E.; Dignon, G. L.; Zerze, G. H.; Schmidt, H. B.; D'Ordine, A. M.; Kim, Y. C.; Rohatgi, R.; Ayala, Y. M.; Mittal, J.; Fawzi, N. L. TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences* **2020**, *117*, 5883–5894.

(16) Paloni, M.; Bussi, G.; Barducci, A. Arginine multivalency stabilizes protein/RNA condensates. *Protein Science* **2021**, *30*, 1418–1426.

(17) 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 Computational Biology* **2018**, *14*, 1–23.

(18) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **2021**, *30*, 1371–1379.

(19) Dannenhoffer-Lafage, T.; Best, R. B. A Data-Driven Hydrophobicity Scale for Predicting Liquid–Liquid Phase Separation of Proteins. *The Journal of Physical Chemistry B* **2021**, *125*, 4046–4056, Publisher: American Chemical Society.

(20) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proceedings of the National Academy of Sciences* **2021**, *118*.

(21) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **2021**, *1*.

(22) Perdikari, T. M.; Jovic, N.; Dignon, G. L.; Kim, Y. C.; Fawzi, N. L.; Mittal, J. A predictive coarse-grained model for position-specific effects of post-translational modifications. *Biophysical Journal* **2021**, *120*, 1187–1197, Publisher: Elsevier.

(23) Reguera, D.; Bowles, R.; Djikaev, Y.; Reiss, H. Phase transitions in systems small enough to be clusters. *The Journal of chemical physics* **2003**, *118*, 340–353.

(24) Wedekind, J.; Reguera, D.; Strey, R. Finite-size effects in simulations of nucleation. *The Journal of chemical physics* **2006**, *125*, 214505.

(25) Grossier, R.; Veesler, S. Reaching one single and stable critical cluster through finite-sized systems. *Crystal Growth and Design* **2009**, *9*, 1917–1922.

(26) Agarwal, V.; Peters, B. Solute precipitate nucleation: A review of theory and simulation advances. *Advances in Chemical Physics: Volume 155* **2014**, 97–160.

(27) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proceedings of the National Academy of Sciences* **2015**, *112*, E6–E14.

(28) Zimmermann, N. E.; Vorselaars, B.; Quigley, D.; Peters, B. Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates. *Journal of the American Chemical Society* **2015**, *137*, 13352–13361.

(29) Salvalaglio, M.; Tiwary, P.; Maggioni, G. M.; Mazzotti, M.; Parrinello, M. Overcoming time scale and finite size limitations to compute nucleation rates from small scale well tempered metadynamics simulations. *The Journal of chemical physics* **2016**, *145*, 211925.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(30) Martin, E. W.; Holehous, 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.

(31) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **2017**, *114*, E8194–E8203.

(32) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular interactions underlying liquidliquid phase separation of the FUS low-complexity domain. *Nature Structural and Molecular Biology* **2019**, *26*, 637–648.

(33) Nilsson, D.; Irbäck, A. Finite-size shifts in simulated protein droplet phase diagrams. *The Journal of Chemical Physics* **2021**, *154*, 235101.

(34) Benayad, Z.; von Bülow, S.; Stelzl, L. S.; Hummer, G. Simulation of FUS protein condensates with an adapted coarse-grained model. *Journal of chemical theory and computation* **2020**, *17*, 525–537.

(35) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *The Journal of chemical physics* **2015**, *142*, 144113.

(36) Liu, C.; Wood, G. P.; Santiso, E. E. Modelling nucleation from solution with the string method in the osmotic ensemble. *Molecular Physics* **2018**, *116*, 2998–3007.

(37) Villois, A.; Capasso Palmiero, U.; Mathur, P.; Perone, G.; Schneider, T.; Li, L.; Salvalaglio, M.; deMello, A.; Stavrakis, S.; Arosio, P. Droplet Microfluidics for the Label-Free Extraction of Complete Phase Diagrams and Kinetics of Liquid–Liquid Phase Separation in Finite Volumes. *Small* 2202606.

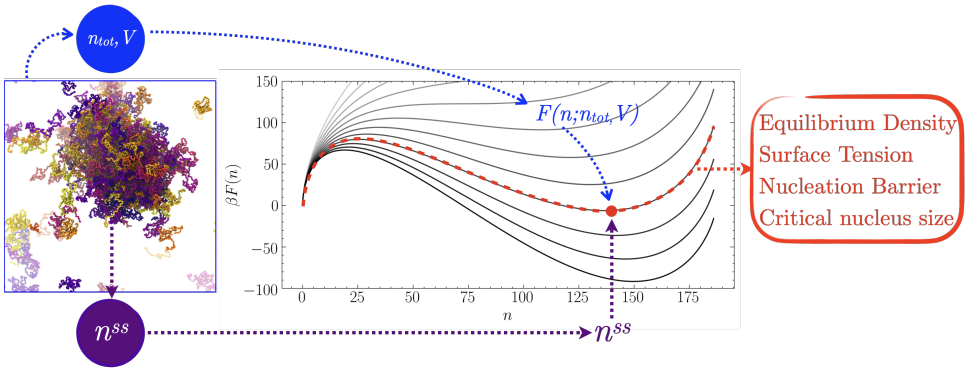
(38) Patel, M.; Shimizu, S.; Bates, M.; Fernandez-Nieves, A.; Guldin, S. Long term phase separation dynamics in liquid crystal-enriched microdroplets obtained from binary fluid mixtures. **2022**,

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For Table of Contents Use Only

Nucleation of Biomolecular Condensates from Finite-Sized Simulations

Lunna Li, Matteo Paloni, Aaron Finney, Alessandro Barducci, Matteo Salvalaglio



bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Material: Nucleation of Biomolecular Condensates from Finite-Sized Simulations

Lunna Li, Matteo Paloni, Aaron Finney, Alessandro Barducci, Matteo Salvalaglio

Peptide Sequence

NDDX4:
MGDEDWEAEINPHMSSYVPIFEKDRYSGENGDNFNRTPASSSEMDDGPSRRDHFMKSGFA
SGRNFGNRDAGECNKRDNTSTMGGFGVGKSFGNRGFSNSRFEDGDSSGFWRESSNDCEDN
PTRNRGFSKRGGYRDGNNSEASGPYRRGGRGSFRGCRGGFGLGSPNNDLDPDECMQRTGG
LFGSRRPVLSGTGNGDTSQSRSGSGSERGGYKGLNEEVITGSGKNSWKSEAEGGES

FUS-LC:
MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQ
SQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSS
SYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNS

Clustering algorithm

We use PLUMED 2.5.2 [1–3] to analyse the steady state condensed phase of NDDX4 and FUS-LC. We adopt a segment-based method in which each peptide chain is represented by different segments. For a NDDX4 molecule of 236 residues, each segment is defined to contain 20 CG beads (n_{beads}=20), with the last segment containing 36 residues. A similar scheme is applied to FUS-LC, which gives a total of 8 segments per FUS-LC chain, with the last segment containing 23 residues. The COM of each segment is then used to construct a radial distribution function (RDF) for the system. Based on the RDF profiles, a threshold distance R_0 can be selected to compute the coordination number (inter-chain/intra-chain contact) for the NDDX4/FUS-LC chains using the COORDINDATION function in PLUMED:

$$\sum_{i\in A}\sum_{j\in B} s_{ij} \tag{6}$$

where s_{ij} corresponds to the following switching function:

$$s_{ij} = \frac{1-(\frac{r_{ij}}{r_0})^6}{1-(\frac{r_{ij}}{r_0})^{12}} \tag{7}$$

The orange dashed line in Fig. S1 indicates that R_0=3.0 nm can be a reasonable choice for NDDX4 and FUS-LC for constructing inter-/intra- chain contacts. Next, the inter- and intra-chain contacts of every peptide chain in the system and over the full simulation length are used to build a 2-dimensional free energy surface (2D-FES). Figure S2 shows that only the inter-chain contact can be used to distinguish different peptide states, as indicated by the dashed blue line. However, it is not sufficient to use one-dimensional (1D) property. As the system size and overall peptide density get smaller, with reduced supersaturation and enhanced finite-size effect, the threshold inter-chain contact tends to shift towards lower values, making it challenging to adapt a universal 1D criteria for distinguish the condensed/dilute phase. Therefore, we also track the distance between the COM of each chain and the condensed phase droplet (COM-distance), to help determining if a peptide chain belong to the dilute or condensed phase. Figures S3 and S4 shows the 2D-FES for cluster analysis of NDDX4 and FUS-LC.

A transition saddle region can be observed in all cases, and it is located at different inter-chain contact/COM-distance values for different system size/peptide density combinations. The orange and cyan dashed lines in Figs. S3, S4 represent the threshold inter-chain contact and COM-distance r_{tran}. For every simulation frame, we assign all the chains with inter-chain contact above the threshold, and with COM-distance below the threshold, as condensed phase chains; all the rest of the chains in the system are considered to be in the dilute phase. Chains that show lower COM-distance but also lower inter-chain contacts represent molecules that are very close to the primary droplet but with much reduced inter-chain interactions; chains with larger COM-distance but moderate inter-chain contact are small clusters of chains in the dilute phase, or a few interface chains that are on the edge of droplet as the condensed phase deviates from a spherical droplet during fluctuation. None are included in the condensed phase counts. For both NDDX4 and FUS-LC, the small clusters of chains prevail in the dilute phase, but any specific cluster typically

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dissipates within short time, in contrast to the stable condensed phase droplet. Similar phenomena are generally discussed for nucleating events ranging from inorganic compounds[4,5] to biological molecules.[6]

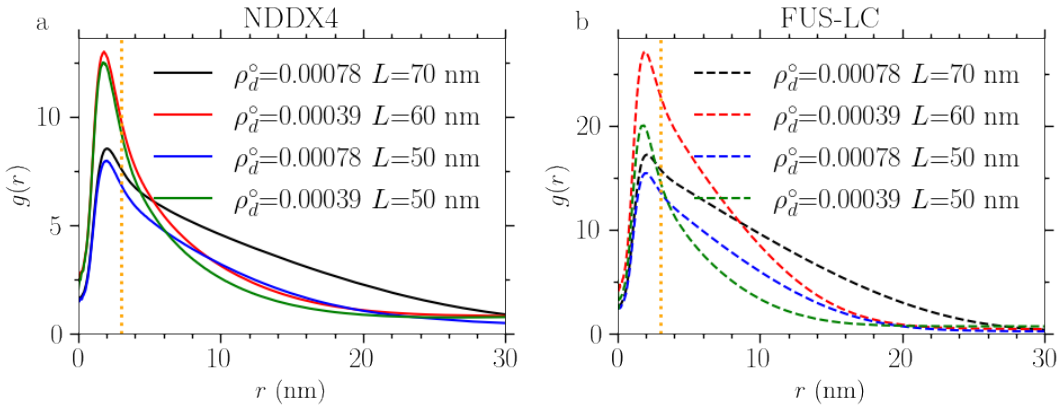


Figure S1: RDF for the segment-COM of NDDX4 and FUS-LC.

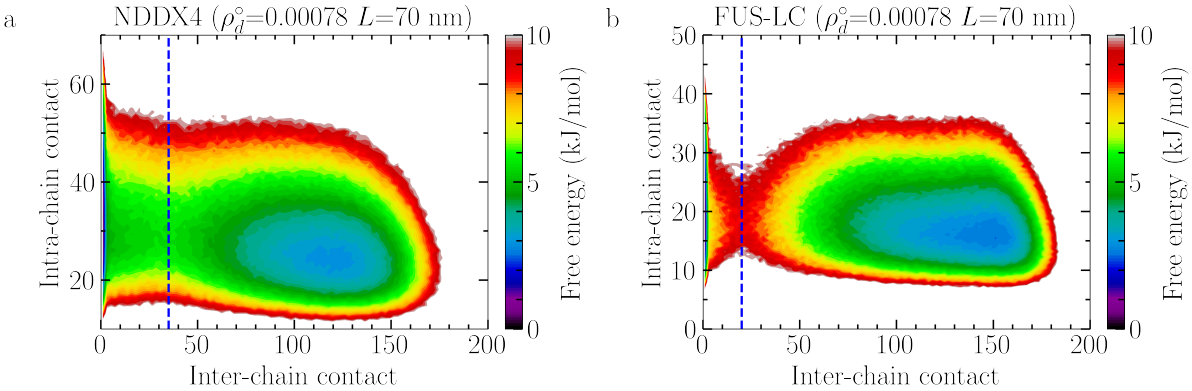


Figure S2: 2D-FES of inter- and intra-chain contacts for NDDX4 and FUS-LC.

Comparison of Figs. S3 and S4 indicates that NDDX4 demonstrates a broader pathway for transition between the dilute and condensed phase. NDDX4 typically shows a larger transition region than FUS-LC on a similar free energy scale. For the limiting case of $\rho_d^\circ = 0.00078$ nm^{-3} and $L = 40$ nm, the NDDX4 droplet is already becoming very fluctuating and much less well-defined due to the combined effect of finite size and reduced supersaturation, while the effect is not yet severe for FUS-LC at the same bulk density/system size condition. For NDDX4 at $\rho_d^\circ = 0.00078$ nm^{-3} and $L = 40$ nm, the threshold COM-distance is $r_{\rm tran}$=16 nm, comparable to the box size $L/2 = 20$ nm, making it challenging to define the condensed/dilute phase boundary.

Table S1 shows the list of threshold COM-distance $r_{\rm tran}$ and droplet radius $r_{\rm ss}$ calculated via $\left(\frac{3n_{ss}v_\ell}{4\pi}\right)^{1/3}$ for NDDX4 and FUS-LC. Droplet size represented by $r_{\rm tran}$ is consistently larger than $r_{\rm ss}$, because $r_{\rm tran}$ also takes into account chains close to the transition saddle region; but $r_{\rm tran}$ and $r_{\rm ss}$ are linearly correlated with each other (see Fig. S5 top inset), so they can be both used to depict droplet size.

As discussed in the main text, the predicted σ of FUS-LC has smaller errors than NDDX4. The slightly improved accuracy may be reflected in the relative shape anisotropy κ^2 of the condensed phase droplets between the two peptides. κ^2 can be used to measures the deviation from a CNT-assumed perfect sphere. Table S1 shows that overall FUS-LC droplet are more spherical than those of NDDX4. As mentioned in the main text, the different sphericity could be associated with the contrasting nature of two peptides, i.e., different hydrophilicity/hydrophobicity. In addition, increased finite-size effect and reduced supersaturation can observe larger deviations from spherical droplet.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1: Droplet size for different system size/bulk combinations

System	L (nm)	ρ_d° (nm^{-3})	r_{tran} (nm)	$r_{ss} = \left(\frac{3n_{ss}v_\ell}{4\pi}\right)^{1/3}$ (nm)	κ^2
FUS-LC-5	50	0.00055	14	8.9	0.08
FUS-LC-6	50	0.00078	15	10.4	0.06
FUS-LC-7	60	0.00039	15	9.3	0.07
FUS-LC-8	60	0.00055	16	10.9	0.06
FUS-LC-9	60	0.00078	18	12.5	0.04
FUS-LC-10	70	0.00039	16	11	0.05
FUS-LC-11	70	0.00078	18	12.7	0.07
FUS-LC-12	70	0.00078	20	14.6	0.03
NDDX4-5	50	0.00055	16	10.8	0.11
NDDX4-6	50	0.00078	20	12.9	0.15
NDDX4-7	60	0.00039	16	10.6	0.12
NDDX4-9	60	0.00055	20	13.3	0.08
NDDX4-9	60	0.00078	22	15.7	0.06
NDDX4-10	70	0.00039	20	12.6	0.08
NDDX4-11	70	0.00055	22	15.4	0.11
NDDX4-12	70	0.00078	26	18.2	0.05

κ^2 is the relative shape anisotropy[7] computed via MDTraj 1.9.4[8] using COM of chains belong to the condensed phase droplet.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

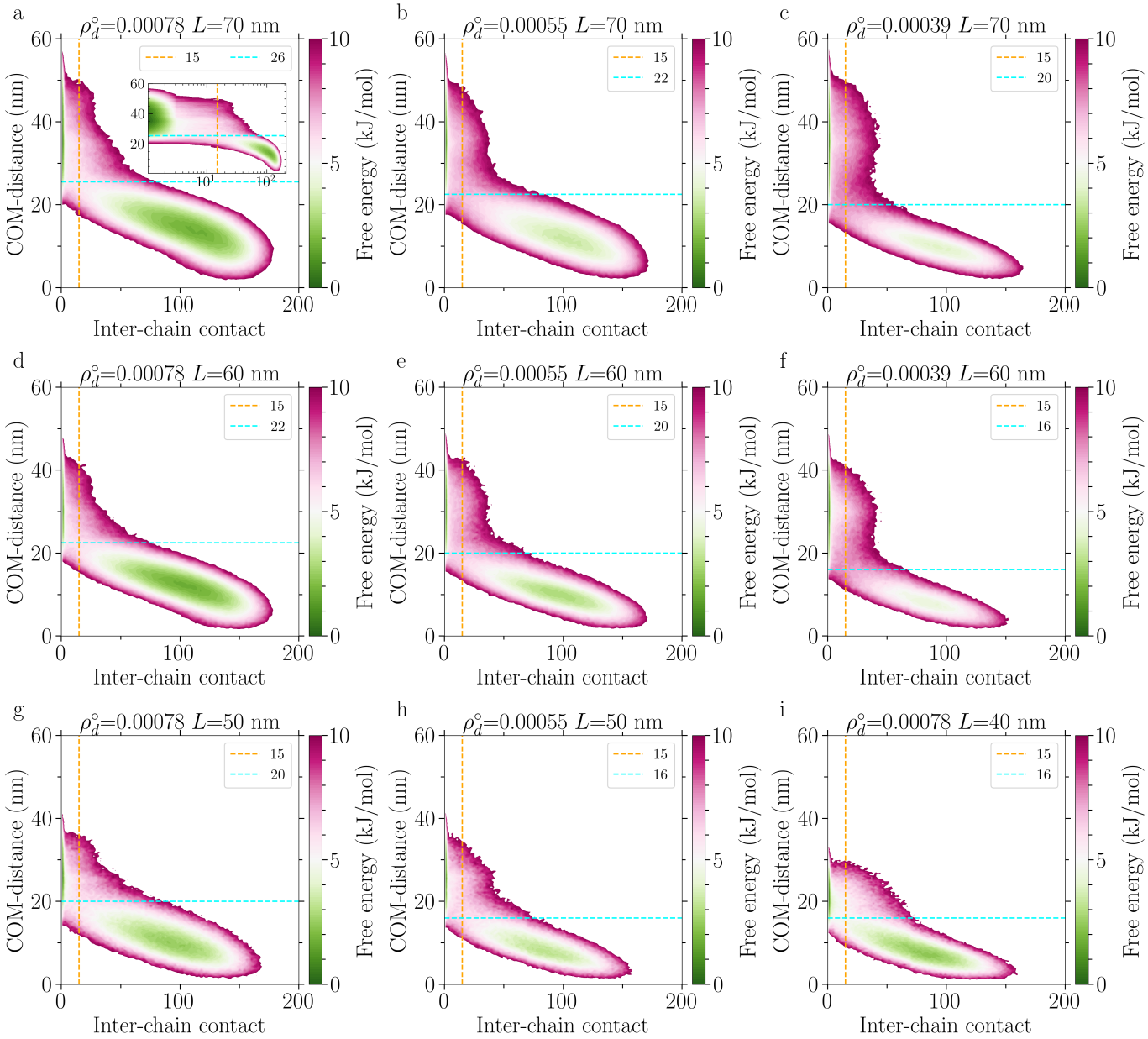


Figure S3: 2D-FES of inter-chain contact and COM-distance for NDDX4

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

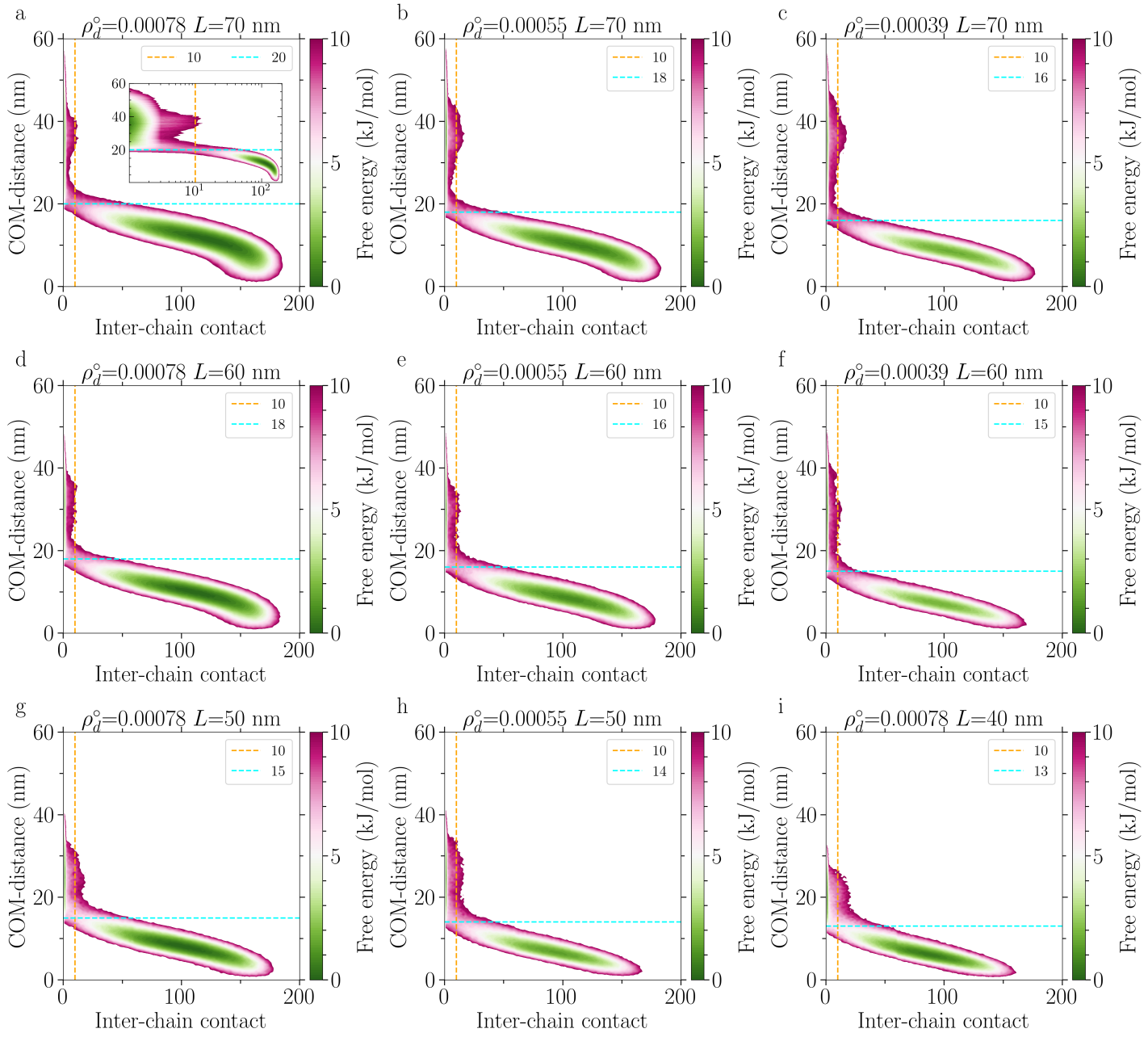


Figure S4: 2D-FES of inter-chain contact and COM-distance for FUS-LC

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

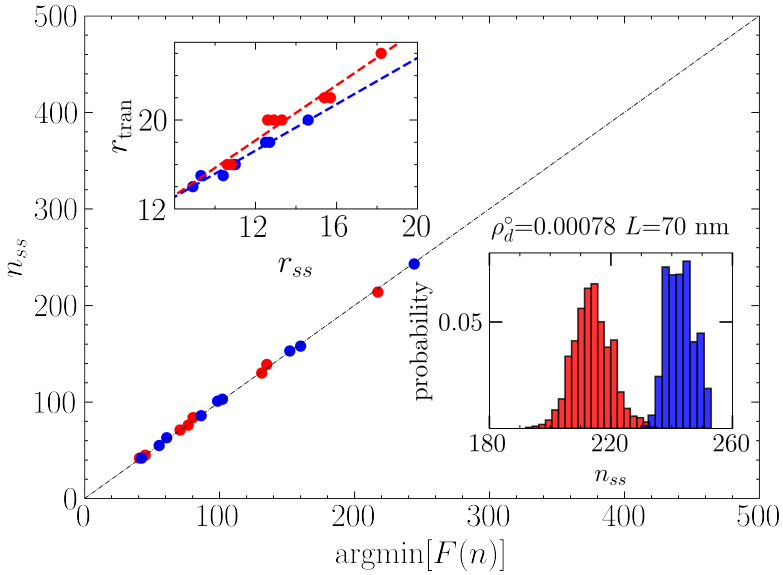


Figure S5: Comparison between the size of the steady-state droplet size measured from simulations and used for performing a global fit of σ and p_d^* via Eq. 4 with the position of the local minimum of the free energy profile expressed by Eq. 1. Top inset: comparison of the steady-state droplet radius r_{ss} with the transition radius $r_{\rm tran}$ and the least squares fitted line. Bottom inset: clustersize distribution for NDDX4 and FUS-LC.

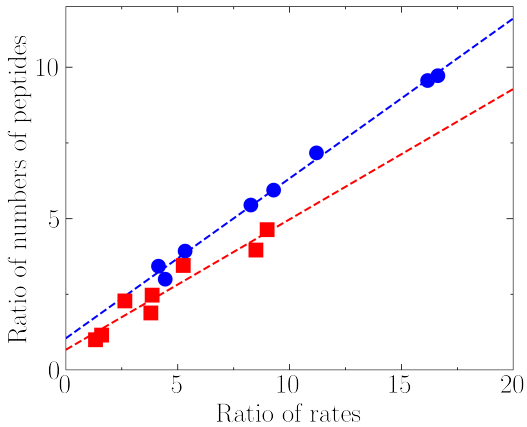


Figure S6: Ratio of the rates of condensed-to-dilute and dilute-to-condensed exchange versus the ratio of the number of peptides in the steady-state condensed and dilute phases for NDDX4 (red) and FUS-LC (blue).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Calculation of steady state droplet density profiles

We use the COM of each chain to build the density profiles for FUS-LC and NDDX4 using an in-house analysis package. The COM trajectories in the steady state are used to construct the largest cluster, representing the liquid droplet, based on a minimum distance criterion. Here, the radial distribution functions for FUS-LC and NDDX4 are first computed and the position of the first maximum informed the truncation distance below which two COMs are considered in direct contact. From these contacts, all of the chains in the liquid droplet are identified and the volume number density of the COMs are computed as a function of the radial distance from the centre of mass of the droplet. Density profiles from this analysis are summarized in Figs. 1 F and G. The small number of chains in the core of the droplets result in significant noise in the computed densities in this region, as highlighted by the uncertainty bars in the figures, computed as the standard deviations from block averaging the full trajectory at every r_c windows.

Markov-state model

We conduct a 2-dimensional Markov-state model (MSM) analysis[9] using PyEMMA[10] for every system size/bulk pepetide density combinations for NDDX4 and FUS-LC, in order to probe the steady state dynamics of chain exchange between the dilute and condensed phases (see Figs. S7 and S8). Overall, the transition regions obtained from the MSM analysis are similar to those observed from 2D-FES (Figs. S3 and S4). We record the reverse of mean-first-passage-time of condensed-to-dilute transition as the condensed-to-dilute rate for plotting Fig. 2C.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

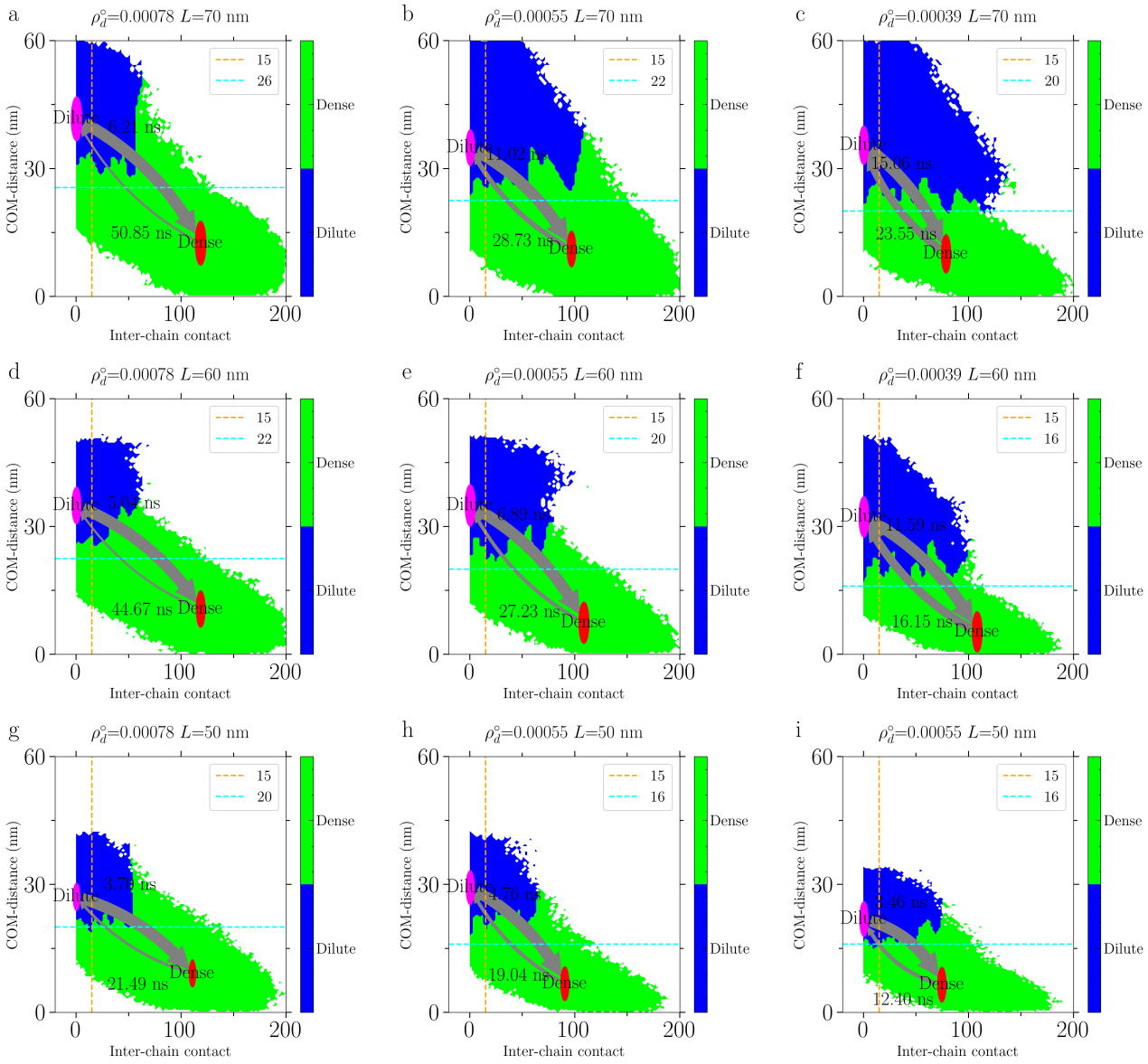


Figure S7: 2D MSM analysis for NDDX4

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

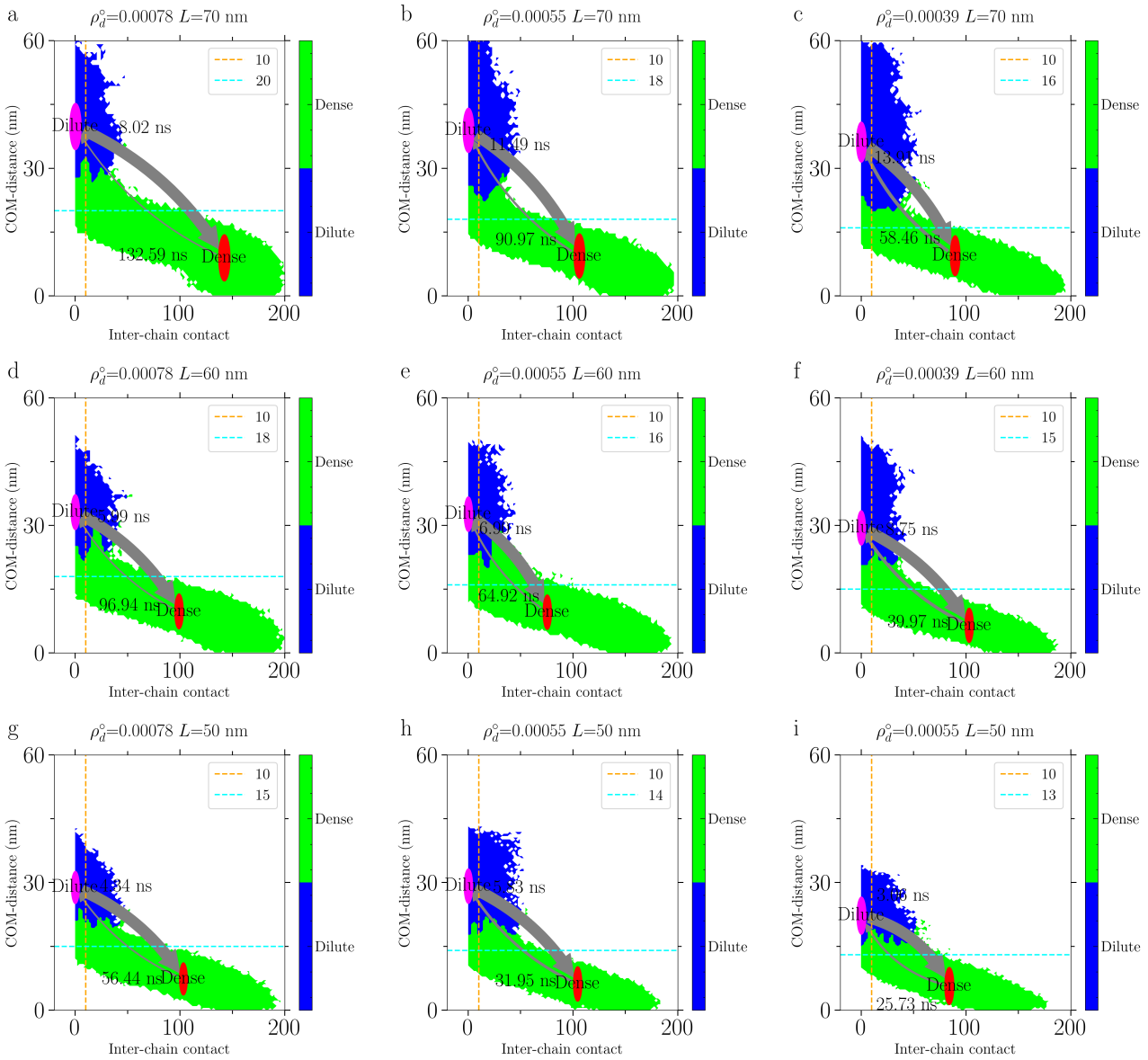


Figure S8: 2D MSM analysis for FUS-LC

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Force-field parameters

The potential energy between two amino acids i and j at distance r is given by:

$$V_{ij} = V_{b,ij} + V_{elec,ij} + V_{LJ,ij} \tag{8}$$

where $V_{b,ij}$ is a harmonic potential between bonded amino acids, with an equilibrium distance of 0.38 nm and a spring constant of 10 kJ/mol/nm^2.

$V_{elec,ij}$ is the Coulombic interaction with a Debye-Hückel screening:

$$V_{elec,ij} = \frac{q_i q_j}{4\pi D r} exp(-r/\kappa) \tag{9}$$

with κ equal to 1 nm.

$V_{LJ,ij}$ is the Lennard-Jones short-range potential:

$$V_{LJ,ij} = 4\lambda_{ij}\epsilon((\frac{\sigma_{ij}}{r})^{12} - (\frac{\sigma_{ij}}{r})^{6}) \tag{10}$$

where $\sigma_{i,j}$ and $\lambda_{i,j}$ are computed with the arithmetic combination from the parameters in Table S2, and ϵ is equal to 0.648 kJ/mol (see Coexistence Simulations section in Supplementary Material)

Table S2 shows our CG parameters for running simulations.

Table S2: Parameters of the CG residues

Residue	m_i[Da]	q_i[e]	σ_i[nm]	λ_i[-]
A	71	0.00	0.504	1/3
C	103	0.00	0.548	1/3
D	115	-1.00	0.558	1/3
E	129	-1.00	0.592	1/3
F	147	0.00	0.636	1
G	57	0.00	0.45	1/3
H	137	0.50	0.608	1/3
I	113	0.00	0.618	1/3
K	128	1.00	0.636	1/3
L	113	0.00	0.618	1/3
M	131	0.00	0.618	1/3
N	114	0.00	0.568	1/3
P	97	0.00	0.556	1/3
Q	128	0.00	0.602	1
R	156	1.00	0.656	1
S	87	0.00	0.518	1/3
T	101	0.00	0.562	1/3
V	99	0.00	0.586	1/3
W	186	0.00	0.678	1
Y	163	0.00	0.646	1

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Effect of parameters for cluster analysis

We have also applied a different segment scheme and R_0 value for cluster analysis, to demonstrate that adjusting the algorithm parameter would not affect the results and hence the predicted thermodynamics and kinetics. The systems tested are FUS-LC and NDDX4 at ρ_d°=0.00078 nm^{-3} and L=70 nm. Figures S9 compares the clustersize distributions estimated using different R_0 and n_{beads}. The results conclude that the parameters used in cluster analysis do not affect the results.

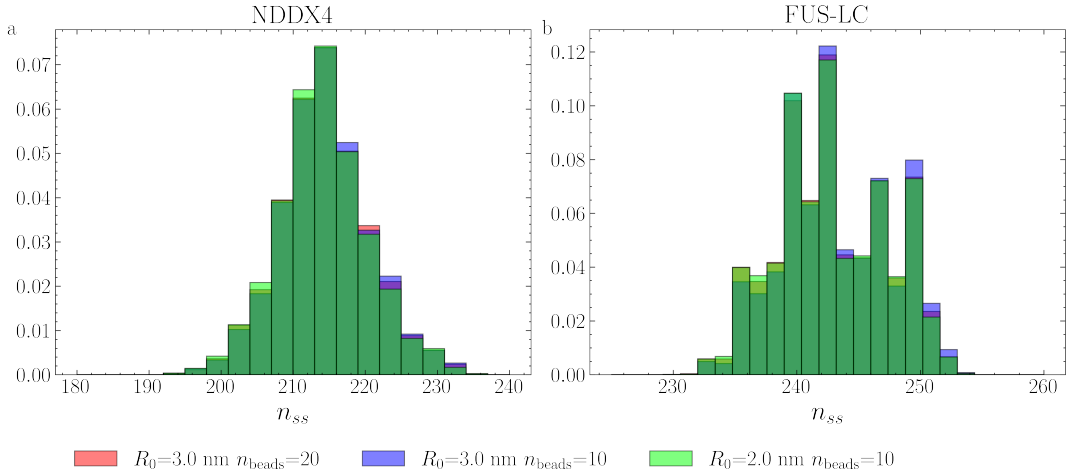


Figure S9: Clustersize distributions estimated using different R_0 and n_{beads} for NDDX4 and FUS-LC segment-COM at ρ_d°=0.00078 and L=70 nm.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Peptide inter-chain interaction

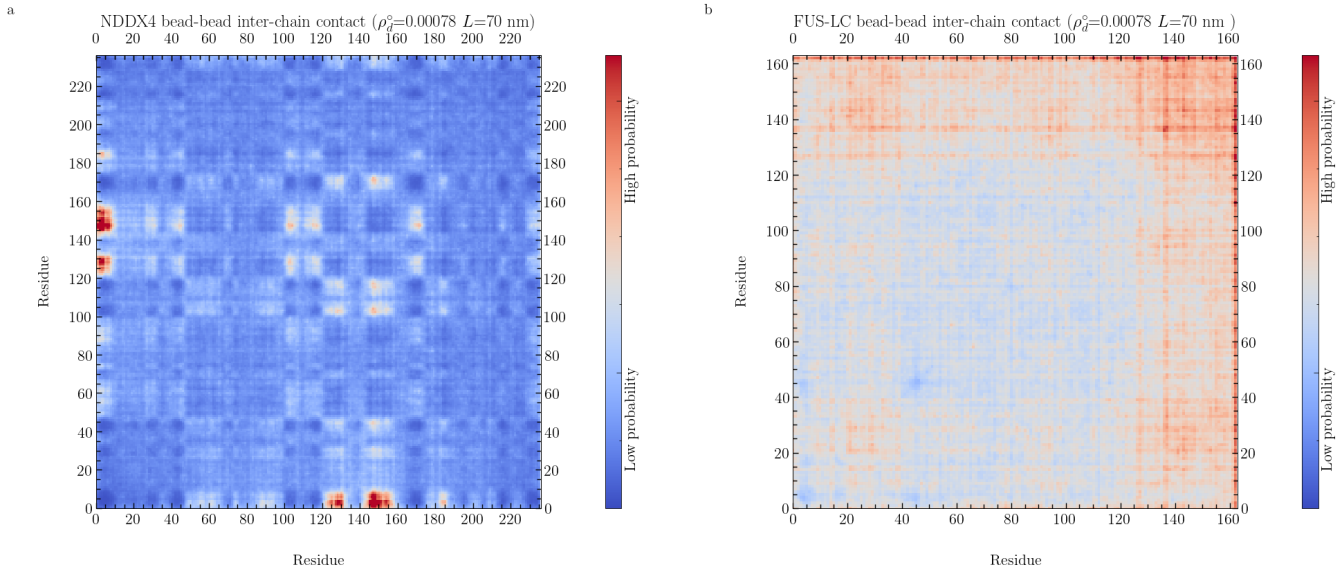


Figure S10: Bead-bead inter-chain contacts for NDDX4 and FUS-LC.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Coexistence Simulations

We performed phase coexistence simulations[? ?] with the following protocol for both systems. 100 copies of the protein molecules were inserted in random positions and orientations in a $20 \times 20 \times 40$ nm^3 ($x \times y \times z$) box using the gmx insert-molecules tool, before extending the simulation box in the z direction to a length of 200nm, keeping the protein center of mass in the center of the simulation box. Finally, we performed production simulations in the NVT ensemble at a temperature of 300K using a leap-frog stochastic dynamics integrator with a timestep of 0.01 ps and a relaxation time of 25 ps. We evaluated the protein concentration along the z axis of the box with the gmx density tool in Gromacs 2018.3,[11] centering the dense liquid protein phase (periodic in x and y) at $z = 0$ at each frame. The dense phase concentration was determined from the values of density around $z = 0$, while dilute phase concentrations were evaluated by averaging concentration values at 60nm$\leq|z|\leq$100nm. We investigated three values of ε, namely 0.627, 0.648, and 0.670 kJ/mol, and chose a value of 0.648 kJ/mol as the one that best reproduced the equilibrium densities of both NDDX4 and FUS-LC (see Fig S11).

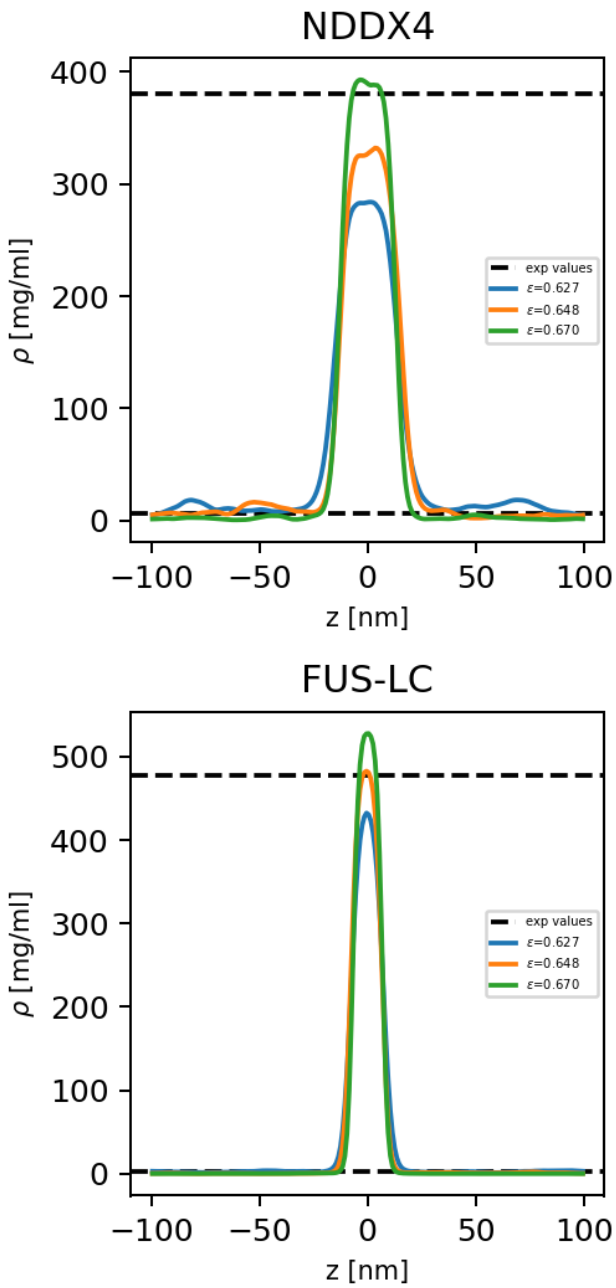


Figure S11: Density profiles from slab simulations of NDDX4 and FUS-LC for different values of the energy scale parameter ε

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.11.515961; this version posted November 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

(1) Colón-Ramos, D. A.; La Riviere, P.; Shroff, H.; Oldenbourg, R. Transforming the development and dissemination of cutting-edge microscopy and computation. *Nat Methods* **2019**, *16*, 667–669.

(2) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604–613.

(3) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.

(4) R. Finney, A.; J. McPherson, I.; R. Unwin, P.; Salvalaglio, M. Electrochemistry, ion adsorption and dynamics in the double layer: a study of NaCl(aq) on graphite. *Chemical Science* **2021**, *12*, 11166–11180.

(5) R. Finney, A.; Salvalaglio, M. Multiple pathways in NaCl homogeneous crystal nucleation. *Faraday Discussions* **2022**,

(6) Vekilov, P. G.; Vorontsova, M. A. Nucleation precursors in protein crystallization. *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 271–282.

(7) Vymětal, J.; Vondrášek, J. Gyration- and Inertia-Tensor-Based Collective Coordinates for Metadynamics. Application on the Conformational Behavior of Polyalanine Peptides and Trp-Cage Folding. *The Journal of Physical Chemistry A* **2011**, *115*, 11455–11465.

(8) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(9) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. Estimation and uncertainty of reversible Markov models. *The Journal of Chemical Physics* **2015**, *143*, 174101.

(10) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.

(11) 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*, 19–25, ADS Bibcode: 2015SoftX...1...19A.