Droplet growth, Ostwald's rule, and emergence of order in Fused in Sarcoma

Farkhad Maksudov[1], Mauro L. Mugnai[2], Laura Dominguez[1,3], Dmitrii Makarov[1,4], and D. Thirumalai[*,1]

[1]Department of Chemistry, The University of Texas at Austin, 105 E 24th Street, Stop A5300, Austin, TX 78712, USA.

[2]Institute of Soft Matter Synthesis and Metrology, Georgetown University, Washington DC 20057, USA.

[3]Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, Coyoacán, CDMX, Mexico

[4]Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, United States

Email: dave.thirumalai@gmail.com

Abstract

The low complexity domain of Fused in Sarcoma (FUS-LC consisting of 214 residues) undergoes phase separation, resulting in a dense liquid-like phase that forms early and slowly matures to reach ordered gel-like state on long time scales. Upon maturation, core-1, comprising of the 57 residues (39-95) in the N-terminus become structured, resulting in the formation of a non-polymorphic fibril. The truncated FUS-LC-C (residues 110-214) construct forms a fibril in which core-2 (residues 112-150) adopts a β-sheet structure. Using coarse-grained

monomer SOP-IDP model simulations of FUS-LC, we predict that residues 155-190 in the C-terminal (core-3) form rapidly, followed by core-2, and finally core-1. The time scale of formation of the cores and their stabilities are inversely correlated, as anticipated by the Ostwald's rule of stages. Unbiased multichain simulations show that the chemical potentials in the two phases are equal and the calculated densities of the dense and dilute phases are in agreement with experiments. The dense phase, which forms by a nucleation mechanism, coarsens over time by a process that is reminiscent of Ostwald ripening. AlphaFold predictions of the core-3 structure and the simulations show that β-strand emerges in the core-3 region early during the droplet formation, and drives the initiation of FUS-LC assembly. The techniques introduced here are general and could be used to probe assembly of other IDPs such as TDP-43, which shares many features with FUS-LC.

Introduction

Phase separation in polymers, resulting in the coexistence of dense and dilute phases, is a well known phenomenon in synthetic [1] and biological molecules [2]. Evidence of phase separation in a mixture of glycoproteins and polysaccharides, and gelatin existed nearly a century ago [3], which coined the term "complex coacervation" to describe this phenomenon). The dense phase consisted of polymers and water and the dispersed low density phase was a solution containing one or more polymers. However, only with the demonstration that phase separation is an important mechanism of cellular organization there has been a great interest in its relevance in biology and biophysics [4–7]. Several studies have shown that besides proteins, RNA in complex with proteins and chromatin also phase separate spontaneously both *in vivo* and *in vitro* [8–10]. In many cases, proteins, typically intrinsically disordered proteins (IDPs) or proteins with disordered regions, could phase separate, at least in *in vitro*. Although the vast number of experimental [11], theoretical [12], and computational studies [13–19] have shed light on the molecular basis of phase separation, many unanswered questions remain. Here, we investigate the mechanism of droplet growth during phase separation in Fused in Sarcoma (FUS), an RNA-binding protein, using computer simulations. Our focus is on the low-complexity (LC) N-terminal domain of FUS (referred to as FUS-LC), which is critical for its ability to undergo phase separation and to form fibril that bind to hydrogels [20, 21].

The 214 residue long FUS-LC (the see the sequence in Fig. 1A), which forms non-polymorphic fibrils, has several unusual attributes that are also shared by other IDPs, such as TDP43 and hnRNPA1. A number of experimental studies [21, 22], summarized previously [12], have shown that FUS-LC and related constructs undergo phase separation under nominal conditions (room temperature and neutral pH). (1) Solid state NMR experiments showed that, on very long time scales, FUS-LC forms a non-polymorphic fibril in which residues 39-95 (referred to as core-1) form a cross-β structure with an S-bend topology [20]. In contrast, FUS-LC-C (residues 111-214) also forms a non-polymorphic structure but with a protolfilament consisting of a dimer (core-2) with each monomer resembling a U-bend in which only residues 112-150 are structured. The two ordered cores structures do not coexist

in FUS-LC even though their free energies are similar [23, 24]. (2) A more recent study demonstrated [21] that a C-terminal region, specifically residues 155 to 190 (core-3), plays a crucial role in hydrogel binding, which means that it is structured [25]. Using a combination of tyrosine-to-serine mutations and systematic deletion of residues from the C-terminal ends of FUS-LC, it was found that mutations in the C-terminal region significantly reduced the ability of FUS-LC to bind to preformed hydrogels [21]. Surprisingly, even after deleting a substantial portion of the N-terminal residues (1-61) that includes part of the core-1 residues (39-61), hydrogel binding was not abolished. These findings indicate that FUS self-assembly must involve at least two distinct interaction regions, raising the question of how these regions interact to regulate FUS assembly. (3) Deletion of most of the residues in core-2 retains the ability to bind to the hydrogel. The results, summarized in (2) and (3), suggest that the FUS-LC domain contains at least two distinct regions that could regulate the self-assembly of FUS: an N-terminal core-1 (residues 39-95) and a C-terminal core-3 (residues 155-190). (4) FUS-LC-C (residues 111-214) polymerizes faster than FUS-LC-N (residues 1-110). Conversely, thermal denaturation studies showed that the FUS-LC-N polymers are more stable, requiring higher temperatures for disassembly into soluble monomers, than the FUS-LC-C polymers.

Here, we used coarse-grained simulations based on the Self-Organized Polymer for Intrinsically Disordered Proteins (SOP-IDP) model [26, 27] to first demonstrate that the kinetics and stability of fibril-like structures are inversely correlated, which explains the experimental observation described above (item (1)). Second, multichain simulations show that the FUS-LC droplet forms by a nucleation and growth mechanism that is consistent with Ostwald ripening, in which the concentration of monomers and oligomers decreases at the expense of the growth of the dense droplet. The coexisting concentrations between the dense and dilute phases, which we establish by showing the near-equality of the chemical potentials between the two phases, are in near-quantitative agreement with experiments. Third, the dense phase is initially driven by the formation of core-3. SOP-IDP simulations show that key residues (164-176) adopt the β-strand conformation. The structure of core-3 fibril determined using AlphaFold shows prominent intra- and interchain contacts in this region. The combination of molecular simulations and structures derived from AlphaFold has resolved several puz-

zles in the FUS-LC assembly and sets the stage applications to other low-complexity IDP sequences.

Results

Ostwald's rule of stages: To establish the relationship between the kinetics of formation of the three cores and their stabilites with respect to the random coil (RC), which is the ground state, we used the N^* theory (summarized in the Supplementary Information) posits that within the native basin of attraction of the monomer high free-energy conformations are populated with low probability. In particular, the N^* states resemble structural motifs found in amyloid fibrils [28–30].

The conformations belonging to the N^* state are determined using long equilibrium simulations. We used the structural overlap function, χ (see Eq. 3 in the SI), describing the similarity between the conformation of the monomer and the structure in the fibril to characterize the excited states. For core-1 (core-2), we used fibril structure determined using solid-state NMR [20] (PDB ID: 5W3N; see Fig. 5A) (cryo-EM [23] (PDB ID: 6XFM; see Fig. S15)). We used AlphaFold2 [31] and AlphaFold3 [32] to generate structures for the prediction core-3 fibril (see Supplementary Results). We define the N^* states as fibril-like conformations with structural overlap parameters $\chi_{fib.i} \geq \chi_{c.i}$, where $i = 1, 2, 3$ denotes the core. The procedure for selecting the values of $\chi_{c.i}$ is described in the SI Methods. The N^*-states are populated in the conformational ensemble of monomers at extremely low levels (usually less than $\sim$1%). Populations of core-1 and core-2 calculated using long equilibrium simulations showed [20, 24] that core-1 is only modestly more stable than core-2, which makes the absence of core-2 fibrils in FUS-LC [23] puzzling. Moreover, neither estimates of the time scales of formation of the fibril-like structures of the three cores nor the fate of core-3 has been investigated in experiments.

To establish a relation between the kinetics of formation of the three cores to their stabilities, we calculated the distribution of first passage times, $P(\tau_{FPT}) = \frac{1}{M}\sum_{i=1}^{M}\delta(\tau_{\mathrm{FPT}} - \tau_{1k})$, where τ_{FPT} is the first time at which the transition from the random coil (RC) ground state to the corresponding N^* state occurs in the kth trajectory. We associated τ_{FPT} with the

time at which a structure with $\chi_{fib.i} \geq \chi_{c.i}$ is reached for the first time in various trajectories. Using 200 independent Brownian dynamics simulations (see Methods), each lasting 280 ms, we observed 42 transitions to core-1, 87 transitions to core-2, 116 transitions to the U-bend conformation of core-3, and 124 transitions to the S-bend conformation of core-3. The smallest number of transitions to core-1 suggests that it is kinetically least accessible.

Because the first passage time (FPT) is not reached in all the trajectories, only truncated distributions of FPTs were sampled (Fig. 1B-D) in the simulations. Consequently, only the lower bounds for the true mean FPTs can be estimated. The error bars represent the standard error of the mean. We could conclude based on the calculated MFPTs that time for forming core-3 is the shortest while core-1 forms on the longest time. Because only a small fraction of the trajectories reach the N^* states, there is uncertainty in reaching this conclusion, which we resolve using a method described below.

To assess the statistical significance of the differences in the mean first-passage times (MFPTs) to the various N^* states, we conducted pairwise comparisons between the MFPT values using Welch's unequal variances t-test [33]. The MFPT for core-1 (129.17 ± 11.73 ms) is significantly greater than for core-2 (97.47 ± 8.70 ms), the U-bend conformation of core-3 (94.17 ± 6.59 ms), and the S-bend conformation of core-3 (99.41 ± 6.61 ms), with p-values between 0.02 and 0.04. However, the differences between core-2, core-3 U-bend, and core-3 S-bend are not statistically significant ($p > 0.6$), nor is the difference between the two core-3 conformations. The lack of statistical significance is due to the finite duration of the simulations, which limits the sampling of rare transitions, resulting in underestimation of the MFPTs. To resolve this issue, we apply a correction described in Eq. 2 below.

Although $P(\tau_{FPT})$, (Fig. 1 (B-D)) is broad, the RC $\rightarrow$ N^* transitions occur by two-state kinetics. The logarithm of the FPT distributions, $\ln[\tau_{FPT}]$, which could reveal kinetic traps as distinct modes [34], shows that the distributions are predominantly unimodal (Fig. 1 (E-G)). Moreover, the fraction of trajectories which have not transited from RC $\rightarrow$ N^* transitions as a function of time, computed as $P_{RC}(t) = 1 - \int_0^t P(\tau_{FPT})\, d\tau_{FPT}$ also exhibit approximately exponential behavior (see Fig. S2F). Thus, we model the transitions using two-state kinetics, which allows us to approximate $P(\tau_{FPT})$, by an exponential distribution, $P(\tau_{FPT}) = \lambda_i e^{-\lambda_i \tau_{FPT}}$, where $\lambda_i = 1/\widehat{\langle \tau_{c.i} \rangle}$ (Fig. 1H). We estimated the true MFPTs, $\widehat{\langle \tau_{c.i} \rangle}$

using Eq. 2. The precisely calculated MFPT values are $\widehat{\langle \tau_{c.1} \rangle} = 1554 \pm 16$ ms, $\widehat{\langle \tau_{c.2} \rangle} = 685 \pm 10$ ms, $\widehat{\langle \tau_{c.3U} \rangle} = 426 \pm 7$ ms, and $\widehat{\langle \tau_{c.3S} \rangle} = 404.9 \pm 7.0$ ms (see Fig. 1I). These results unambiguously show that the order of formation of fibril-like structures is, core-3 followed by core-2 and finally core-1, $\widehat{\langle \tau_{c.1} \rangle} > \widehat{\langle \tau_{c.2} \rangle} > \widehat{\langle \tau_{c.3} \rangle}$. We should emphasize that MFPT estimates are only proxies for the times needed to form the fibrils, which occur on timescales of days in FUS-LC [20, 21, 23, 35].

We then assessed whether the N^* states identified in the isolated monomer are structurally distinct from the rest of the conformational ensemble by comparing the residue-wise propensity for β-strand formation in fibril-like conformations—defined by $\chi_{fib.i} > \chi_c$ for each core $i = 1, 2, 3$—denoted as $p_\beta^{c.i}$, to that in the remainder of the ensemble, which predominantly consists of random-coil-like structures. Let p_β^{RC} be the residue-dependent β-strand content in fibril-like N^* states (see Fig. S3). For each core i, the ratio $p_\beta^{c.i}/p_\beta^{RC}$ exceeds unity primarily within the corresponding core region of FUS-LC. This finding provides strong evidence that the N^* states, although sampled by the isolated monomer, capture key structural features of fibril-like conformations. We also calculated the radius of gyration (R_g) and end-to-end distance (R_{ee}) for the N^* states (see Fig. S4). The core-1 fibril-like conformations ($\chi_{fib.1} > \chi_c$) are more compact compared to both the random coil ensemble and the fibril-like conformations associated with core-2 and core-3 (see Fig. S4C). In contrast, core-2 and core-3 fibril-like conformations exhibit R_g and R_{ee} values that are more similar to those of the random coil states.

Having shown that the excited states in the monomer ensemble are fibril-like structures, we calculated the relative stabilities of the three N^* states (see SI Methods) sampled in the equilibrium Langevin simulations of FUS-LC. The free energy difference between the core-1 and core-2 is $\Delta F_{1,2} = -0.93 \pm 0.13$ kcal/mol, which is reasonably close to the value obtained from solubility measurements [23], 0.35 ± 0.19 kcal/mol. The relative stability of core-2 with respect to the U-bend topology of core-3 is $\Delta F_{2,3U} = -7.37 \pm 0.15$ kcal/mol. Among the core-3 variants, the one with the S-bend topology is more stable than the U-bend conformation ($\Delta F_{3S,3U} = -5.59 \pm 0.12$ kcal/mol), while both are less stable than core-2 ($\Delta F_{2,3S} = -1.78 \pm 0.11$ kcal/mol). The stability hierarchy is reflected in the energy differences $\Delta F_{1,3U} = -8.29 \pm 0.22$ kcal/mol and $\Delta F_{1,3S} = -2.70 \pm 0.10$ kcal/mol. Strikingly, the rates

of formation of the three fibril-like structures are inversely correlated with with free energy differences between different cores (see Fig. 1I), which is in accord with the Ostwald's rule of stages, found previously in dipeptide supramolecular polymers [36].

Absence of core-3 accelerates the formation of core-1 and core-2: We repeated the calculations described above for FUS-LC-N (residues 1–163) to illustrate the effect of core-3 on the kinetics of core-1 and core-2 formation. In 200 Brownian trajectories, each lasting 60 ms, there are 16 transitions to core-1 and 100 transitions to core-2. The values of the MFPTs, estimated from the completed transitions are: $\langle \tau_{c.1|s} \rangle = 31.03 \pm 4.6$ ms, $\langle \tau_{c.2|s} \rangle = 27.41 \pm 1.7$ ms (see Fig. S2). The unimodal (Fig. S2) distributions of the FPTs imply $P(\tau_{\rm FPT}) = \lambda_i e^{-\lambda_i \tau_{\rm FPT}}$, allows us to calculate the true MFPTs, $\widehat{\langle \tau_{c.i} \rangle}$, using Eq. 2. We find that $\widehat{\langle \tau_{c.1} \rangle} = 694.97 \pm 0.01$ ms, $\widehat{\langle \tau_{c.2} \rangle} = 85.14 \pm 0.002$ ms (see Fig. 1J). Equilibrium simulations of FUS-LC-N yielded a free energy difference between core-1 and core-2 of $\Delta F_{1,2} = -2.24 \pm 0.06$ kcal/mol. As is the case for FUS-LC, the formation rates of fibril-like structures and the free energy differences between the competing core structures in FUS-LC-N are inversely correlated (see Fig. 1J). Interestingly, core-1 and core-2 form 3-fold and 11-fold slower in FUS-LC than in FUS-LC-N, which lacks core-3 except for 9 residues (full core-3 consists of 36 residues). This finding is sufficient to conclude that core-3 must play a key role in the assembly of FUS-LC (however, see below).

Droplet formation and Ostwald ripening: We initiated five multi-chain simulations from five distinct initial conformations in a homogeneous solution containing 200 randomly distributed FUS-LC chains (refer to Methods; see also a top panel in Fig. 2A; see also Fig. S5). The box size corresponded to an overall protein concentration of 600 μM, which is larger than the typical range (100-250 μM [37, 38]) used in *in vitro* experiments but is sufficiently low to observe both the condensed and dilute phases in equilibrium (see Fig. 2A).

The unbiased simulations show that oligomers of different sizes form spontaneously (see Supplementary Movie 1). We divided the oligomers into three groups: small oligomers (2-9 proteins), large oligomers (10-31 proteins), small droplets (32-70 proteins), and large droplets (¿ 70 proteins). At the start of the simulations ($t < 0.1 \times 10^9 \tau$, where τ is a simulation time step defined in Methods), the FUS-LC monomers quickly coalesce into small oligomers, which rapidly disintegrate into monomers and reassemble repeatedly. At intermediate times, some

of the small oligomers act as nucleation sites, leading to the formation of large oligomers. The initial clustering is in accord with the expectations of the classical nucleation theory (CNT) for the formation of the subcritical nuclei [39]. Eventually, one of the oligomers serves as a nucleation site, leading to the spontaneous formation of a small droplet at $t > 0.5 \times 10^9 \tau$. The transition from a subcritical to a supercritical nucleus is a hallmark of CNT, signifying the instant when the nucleus exceeds the critical radius and initiates the stable growth of a large FUS droplet (see Figs. 2 and S5). The onset time for stable droplet growth varies significantly across different simulation runs (see Fig. 2), suggesting that the droplet assembles by multiple pathways, which is also consistent with CNT predictions.

Most commonly, monomers disintegrate from the oligomers or smaller droplets, thus exposing a higher fraction of sites to the solvent (modeled implicitly here). These are subsequently reincorporated into larger droplets, which is reminiscent of Ostwald ripening [40, 41], a phenomenon in which less stable smaller droplets or oligomers disintegrate into monomers (see Figs. 2C and S9A; see also Supplementary Movie 2) due to their higher surface energy and chemical potential (see Fig. 3C). Such monomers subsequently diffuse and are re-incorporated into larger droplets or oligomers, which are free energetically more favorable because of their lower surface-to-volume ratio and reduced interfacial energy. In our simulations, the higher fraction of solvent-exposed sites in smaller droplets (see Fig. S6) accelerates their dissolution, effectively supplying material for the growth of larger droplets, as anticipated by Ostwald ripening [41, 42]. This process results in the coarsening of the system over time, with larger droplets becoming more prominent while smaller ones diminish. The observed dynamics highlight the interplay between thermodynamic stability and molecular exchange, which governs the temporal evolution of the phase-separated FUS-LC droplets.

Monomer conformations in the dilute and dense phase are similar but not identical: We compared the Small-Angle X-ray Scattering (SAXS) profiles ($I(q)$ with q being the scattering wave vector) calculated for FUS-LC chains within the droplet to $I(q)$ in the dispersed phase (see Fig. 3A). The profiles were virtually indistinguishable from each other, which is consistent with several experimental studies on FUS-LC of varying sequence length. (1) In a recent work [43], NMR signals for FUS NTD (residues 1-267) were indistinguishable between the dilute and condensed phases. Similar conclusions were reached using solid-state

NMR studies of FUS-LC [44]. (2) Solution-state NMR studies of FUS-LC-N (residues 1-163) [45] showed no significant conformational changes upon phase separation on the timescale of the experiment. (3) Double electron-electron resonance (DEER) spectroscopy studies of FUS NTD, consisting of residues (1-267) [46] also showed that the conformational similarity of FUS in dilute and condensed phases. The measured distance distributions between selected pairs of residues within FUS NTD monomers inside and outside the condensed phase were similar [46]. We calculated these distributions between the same set of residues in FUS-LC and found that they are similar both the phases (Fig. 4). The averages and standard deviations of these distances show reasonable agreement with those reported for FUS NTD (see Table S4; see also Fig. 4). (4) The inferred distances between the N-terminal residue and three selected residues in FUS-LC-N monomers both inside and outside the droplet using experiments smFRET [47] showed only a modest increase in the condensed phase. We computed the same distances for FUS-LC and similarly observed a minor increase for monomers inside droplets (see Fig. S11). This increase in distance between the N-terminal and other residues suggests a modest degree of monomer unwinding within the condensed phase, which is also reflected in the distributions of the radius of gyration (R_g) and end-to-end distance (R_{ee}) for FUS-LC proteins inside and outside the droplet (see Fig. S11).

Inherent Structures (IS): The findings described above show that while the conformations of monomeric FUS-LCs remain largely similar in homogeneous and phase-separated environments, subtle structural differences emerge in the latter. To further investigate this, we calculated the residue-dependent β-strand propensity of FUS-LC monomers in both the environments (see SI Methods) (see Fig. 3D). The overall β-strand propensity remains low, suggesting that monomeric FUS-LCs inside droplets are also disordered. However, the monomers within droplets have nearly threefold higher propensity for β-strand formation compared to free monomers (see Fig. 3D). To ascertain that the strand content is enhanced in the droplet, we determined its inherent structure (IS) [48] using a method that was introduced to probe packing in dense liquids. The IS, which locates local minima in the energy landscape, was determined by quenching the conformation of the droplet to a low temperature. We find that the β-strand (see Fig.3E) content is typically greater in the cores than elsewhere. In particular, the greater β-strand content in core-3 shows that there is a higher tendency for

structural ordering in this region even before core-1 forms, which is consistent with experimental observations [21] and Ostwald's rule of stages (see Fig. 1). It is worth bearing in mind that fibrils form only on times scales on the order of days [20, 21, 35], which implies that the droplets in our simulations and experiments [45] are metastable.

Chemical potentials in the two phases are similar: For $t > 3.5 \times 10^9\,\tau$, a stable droplet forms from which only the monomers disassociate on long time scales (see Fig. 2). However, polymer chains chains exhibit continuous dynamic exchange, moving in and out of the droplet (see Fig. 2B; see also Supplementary Movie 3), as required for phase equilibrium. To ensure that we characterize the equilibrium state of the condensate, we discarded the first $3.5 \times 10^9\tau$ time steps in the trajectory. To assess that equilibrium is established, we calculated the chemical potentials in the condensed (μ_c) and dispersed (μ_d) phases using the approach described in SI. We assumed that the polymers that are not in the condensed phase constitute the dispersed phase. In other words, the variations in the oligomer sizes found in the dispersed phase are not treated as separate species. The relative difference, $\Delta\mu = (\langle\mu_c\rangle - \langle\mu_d\rangle)/\langle\mu_c\rangle$, is 0.003 ± 0.03. The large error bars that exceed the mean value are due the finite system size (see Fig. S7).

The similarity between the chemical potential distributions of the two phases, $P_d(\mu)$ and $P_c(\mu)$ (see Fig. 3B), was assessed using the Jensen-Shannon divergence (JSD) (see SI Methods). The calculated divergence, $D_{\mathrm{JS}}(P_d(\mu) \parallel P_c(\mu)) = 0.01$. The small value of the JSD shows that the two distributions are similar. Because their mean values are comparable, we surmised that the observed differences arise from the larger variance in μ for the condensed phase (see Fig. 3B). Given that the system size that can be reliably simulated is small (maximum of 200 chains), the conclusion that the chemical potentials between the coexisting phases are similar is unreasonable.

Concentrations of the coexisting phases: The establishment of the equality of the chemical potentials in the two phases allows us to reliably calculate the coexisting protein concentrations (see SI Methods for details). The average FUS-LC concentration of the condensed phase is $\rho_c = (16.24 \pm 8 \times 10^{-4})$ mM (see, Figs. 2D and 3H; see also Fig. S12B), which is in excellent agreement with a few experiments. Raman imaging experiments were used to estimate the dense phase concentration to be 15.0 mM under standard conditions

(pH 7.6 and 150 mM NaCl) [37]. An independent study using quantitative phase imaging reported that the FUS-LC concentration in the condensed phase to be 337.3 mg/mL (15.7 mM, assuming a molecular weight of 21.5 kDa for FUS-LC) [49], which is comparable to the 20 mM concentration reported for FUS NTD (residues 1-267) [46]. The average FUS-LC concentration in the dispersed phase in the simulations is $\rho_d = (0.28 \pm 2 \times 10^{-4})$ mM (see, Fig. 3H), which is in reasonable agreement with the experimentally reported concentration of ≈ 100 μM [46, 50]. In our simulations we observed that on average there are 114.80 ± 0.05 FUS-LC chains in the condensed phase, while 85.2 ± 0.05 are in the dilute phase (see Figs. 2C and 3G), which agrees with experimental estimates [46]. The elevated overall protein concentration used in the simulations likely results in the larger value ρ_d of the low density phase. The density of the droplet calculated using simulations is about ≈ 58 higher than in the dispersed phase. The corresponding experimental value is ≈ 157. The good agreement between the simulations and several experiments further validates our model.

Mechanism of droplet formation: Given that the chains in the droplet are disordered, it is natural to wonder if specific interactions between the FUS-LC polymers initiate phase separation. To address this issue, we analyzed interchain interactions. The results, which were qualitatively consistent across all the simulation runs, are illustrated in Figs. S8 and S9. At short times, interchain interactions are rare because most proteins exist as free monomers or small oligomers. However, the majority of interchain interactions that form during the early times are mediated by the core-3 region, with frequent interactions between residues 164-176 of one protein and the corresponding residues of another. Subsequently, other non-preferential interactions, which are intermittent and disrupted over time, are observed. The contacts between residues 164-176 consistently form the most populated interchain interactions. Interactions mediated by core-1 are discernible only at later stages of droplet formation, with preferential contacts involving core-1 residues 61-77. Interestingly, amino acid residues in core-2 (residues 112-150) do not engage in preferential interactions at either the beginning of the simulation (Figs. S8 and S9) or not at later stages (Fig. 3I), except when mediated by core-1 or core-3. At equilibrium, interactions between core-2 regions of different chains are the least probable (see Fig. 3I), a finding that supports the observation that core-2 is not required to bind to hydrogels [21].

These findings were derived using simulations conducted at low friction, which could alter the kinetic pathways and the temporal order of events. To mitigate these effects, we performed additional five independent Brownian dynamics simulations for a system of 64 FUS proteins in a solvent with viscosity set to 10^{-3} Pa·s (see Methods). In the first 150 μs of each simulation, transient oligomers containing 4-8 chains form and dissociate (see Fig. S9A), in accord with the results from low-friction simulations (see Fig. 2). Analysis of the interchain interactions during Brownian dynamics further revealed that residues 164-176 primarily mediate contacts between the chains in the oligomers (see Fig. S9). From the temporal progression of the contact maps (see Fig. S9C), we find that condensate formation is initiated from oligomers that act as nucleation sites, predominantly involving core-3. This suggests that core-3 plays a central role in mediating droplet formation (see below). This is further supported by the interaction patterns observed in dimers and small oligomers (fewer than 10 chains; see Fig. S9D): although the core-1 region also participates in interchain interactions, their relative contribution decreases in larger oligomers, where core-3-mediated interactions become dominant.

Core 3 residues drive FUS-LC condensation: It is surprising that the most frequent and the strongest interchain interactions are mediated by the C-terminal tail of FUC-LC, particularly residues ^{164}SSGGGGGGGGGGN^{176} (see Fig. 3F,I). Strikingly, these residues have the highest propensity to form β-strands at the monomer level (see Figs. 3D and S10). The relevance of core-3 in FUS-LC assembly was demonstrated [21] by systematically truncating residues from C terminus, and probing their ability to bind to hydrogel. Truncation beyond residue 180 resulted in a complete loss of the ability of GFP-tagged test proteins to bind to mCherry hydrogels formed by the intact FUS-LC (see Fig. 2A in ref. [21]), thus suggesting that residues that are critical for FUS-LC condensate formation is within the 165-180 range. This finding is supported by our simulations which show that the most probable interactions are mediated by residues 164-176. In addition, our simulations show that residues ^{71}TPQGYGSTGGYGSS^{84} (Fig. 3F,I) also form several interchain interactions within the droplet. Kato and McKnight also systematically created FUS-LC variants by truncating chunks of residues from the N terminus (see Fig. 2B in ref. [21]). The results of these experiments show that the crucial residues for FUS-LC hydrogel binding activity are within

residues 61-77, which also aligns with our simulation results. The early formation of structure in residues (164-176), containing mostly Glycines, the absence of discernible order in core-2, and emergence of order late in core-1 show that core-3 that initiates the assemblly of FUS-LC.

Structural model for core-3: In light of the importance of core-3 in FUS-LC assembly, we used to AlphaFold2 (AF2; [31]) and AlphaFold3 (AF3; [32]) to predict the unknown structure of core-3. To assess the reliability of AF2 and AF3, we first determined the structures of core-1 and core-2 and compared them to the experimentally resolved counterparts (see Figs. 5 and S15 for AF3). AlphaFold3 generates structurally diverse models of the fibrils based on different random seeds [51]. Although the predicted and experimentally derived structures deviate to some extent, the majority of residues that form β-sheet secondary structures in the experimental fibrils are also found in β-sheets in the AF3-predicted models (see Supplementary Results for details). Moreover, the predictions are sufficiently accurate to validate Ostwald's rule of stages. This conclusion is based on a comparison of mean first-passage times (MFPTs) computed using both AF3-predicted and experimental fibril structures (see Supplementary Results for details).

To assess the reliability of the AF3 predictions, we analyzed the quality metrics provided by AlphaFold—pLDDT, iPTM, and pTM [32]. These metrics exhibited substantial variability depending on the chain configuration. By examining their distributions across different chain numbers (see SI for details), we determined the most reliable structural model (Fig. S14). For $FUS_{155-190}$ fibrils, which consist of four subunits (Fig. S15D), we observed higher pLDDT, iPTM, and pTM values compared to other assemblies (see Fig. S14). The selected core-3 fibril model had the highest values across all three quality metrics: iPTM, pTM, and average pLDDT (see Table S3). These values fall within the typical range observed for amyloid fibrils [51]. A similar analysis for $FUS_{141-214}$, which forms core-3 fibrils experimentally [21], revealed that the optimal structural model also consisted of a four-subunit complex (Fig. 5C), with iPTM, pTM, and average pLDDT (see Table S3). Given the experimental confirmation of core-3 fibril formation for this sequence, we adopted the structures derived for $FUS_{141-214}$ in further analysis.

AF3 predictions show that the two outer extended segments of core-3 correspond to

residues 155–163 and 182–190, while the inner segment consists of residues 168–177 (Fig. 5C). Within this architecture, residues 156–163 and 172–190 form β-strands in a serpentine-like β-sheet arrangement, whereas residues 164–166 adopt a β-turn conformation that facilitates chain reversal. Residues 155 and 167–172 remain in a random coil conformation. Figures S16A and S16B display per-atom pLDDT scores for the $FUS_{155-190}$ and $FUS_{141-214}$ fibrils, respectively. The monomeric unit in the $FUS_{155-190}$ fibril adopts a U-bend topology with a "kinked" region spanning residues 176–184 (Fig. S15D). In contrast, the core-3 region (residues 155–190) of the $FUS_{141-214}$ monomer adopts an S-bend topology (Fig. 5C). Interestingly, the "kinked" region (residues 176–184) is conserved across both models.

Given that core-3 forms cross-β polymers [21], we used the predicted structures to determine the residue dependent secondary structural propensity [52]. Both the U-bend and S-bend topologies exhibited a β-sheet secondary structure throughout most residues (see Figs. S16C,D). Additionally, residues 164-176, which showed strong inter-chain interactions in our simulations (Figs. 2 and S9), also displayed prominent intra- and interchain contacts in the predicted core-3 fibrils (see Figs. S16E,F). This result further highlights the importance of core-3 in mediating interactions critical for fibril assembly.

Discussion

The in-depth study of phase separation in the low complexity domain of FUS and its variants has resolved a few long-standing problems related to its assembly. For instance, based on the weak interactions between the polymer chains, one would expect that there are no preferential contacts between the chains within the droplet. Unexpectedly, our simulations reveal that interchain interactions involving residues 164–176 (the center of the core-3) are structured early and initiate the formation of the FUS-LC droplet, which implies that the presence of structure in core-3 propagates to core-1 over time. Importantly, the lack of interchain contacts associated with core-2 (Figs. 3, S8, and S9) supports the experimental finding that its formation is not essential for hydrogel binding [21]. Besides these conclusions, there are several other findings whose implications are explored below.

Nucleation and Ostwald ripening: Droplet formation and growth takes place in two

steps. In the first step, monomers collide to form oligomers that frequently disintegrate until one with a critical size emerges (see Supplementary Movie 1), which subsequently grows to a larger droplet forms. In the second step, excess monomers (and possibly small oligomers) are "devoured" (the word used in [41]) by largest droplet leading to the coarsening of the condensate (see Supplementary Movie 2). The simulations vividly illustrate the two processes. During progressive coarsening larger droplets become increasingly dominant ("winner take all" scenario [53]) as the smaller ones shrink. The observed dynamics underscore the balance between thermodynamic stability and molecular exchange, which dictates the temporal evolution of the phase-separated FUS-LC droplets.

Our study provides a structural perspective on the nucleation and growth of the droplet, bridging the initial nucleation events described by Classical Nucleation Theory (CNT) with the long-time coarsening governed by Ostwald ripening (see Supplementary Movie 1). Because of the structural heterogeneity of monomers and oligomers, nucleation mechanism may be most efficient only when a subset of structures of the heterogeneous conformations are involved. Indeed, the relevance of the heterogeneous structures sampled in the dispersed phase in controlling the droplet formation is supported by theoretical and experimental findings [54, 55]. Despite the complexity, remnants of CNT are evident in the droplet growth. Our simulations show that the dependence of droplet stability on size is in accord with both theoretical frameworks: CNT establishes a critical nucleus size required for stable growth, while Ostwald ripening explains the preferential growth of larger droplets at the expense of smaller ones. The simulations capture this size-dependent behavior, elucidating its role in shaping the overall droplet population dynamics.

Core-3 structure prediction: To determine the kinetics of core-1 and core-2 formation using the spectrum of monomer states, we utilized the known experimental structures for reference [20, 23]. To determine the rate of formation of core-3, we relied on the structures predicted by AlphaFold. Although there is some variability among the structures, core-3 is predicted to have both a U-bend and S-bend topologies, with the same residues positioned at the kink. Using the structural overlap parameter, we calculated the mean first passage times and found them to be similar for both types of structures. This shows the conclusions pertaining to Ostwalds's rule of stages are robust even if alternative conformations predicted

by AlphaFold are used (Fig. S20). We do not claim that any of the predicted structures of core-3 are correct in reality. Indeed, the performance of AlphaFold for IDRs and even for fold-switching proteins has been criticized [51, 56, 57]. Despite these caveats, our results for FUS-LC are in excellent agreement with experiments [21], suggesting that predicted structural models for core-3 are likely to be reasonable.

Ostwald's rule of stages: The detection of the excited N^* fibril-like structures the spectrum of monomers [24] informs our understanding of fibrillar assemblies in IDPs and proteins that have a high propensity to aggregate [22, 30, 58]. The inverse correlation between the stability of the fibril cores and the kinetics of their formation, demonstrated from the knowledge of the N^* states, is accord with the Ostwald rule of stages that was previously shown for dipeptide supramolecular polymers [59] and Aβ peptides [60]. Our simulations establish that determination of the N^* states, perhaps by relaxation dispersion nuclear magnetic resonance spectroscopy [58, 61] or by time resolved solid state NMR [62], would be important in anticipating the mechanism of condensate formation and subsequent transition to fibrils. It is worth noting that the validity of the Ostwald's rule of stages is shown using the structures of the monomer N^* states. In actuality, FUS-LC fibrils form on time scales on the order of days [20, 21, 35]. Nevertheless, we expect that Ostwald's rule would hold as the droplets transition to a solid.

Although the N^* conformations sampled by the isolated monomer are structural proxies for their conformation within core-1, core-2, or core-3 fibrils, our results show that these regions exhibit an increased propensity to form β-strand in the corresponding regions of FUS-LC (see Fig. S3) even at the early stages of assembly. Thus, our findings validate the relevance of N^* states in characterizing fibril formation even though they are sampled in isolated monomers but exhibit structural features that mirror those in the fibrils. This observation accords well with previous studies indicating that such N^* states, though rarely populated under native conditions, are encoded within the monomer's free energy landscape and can serve as precursors to aggregation-prone conformations [30].

Role of core-3: The mutual exclusion in the formation of core-1 and core-3 structures was established using ALS-associated mutations [21]. Our simulations provide indirect evidence for this key finding. It is based on a comparison of the MFPTs for fibril-like conformations

in FUS-LC and FUS-LC-N (see Figs. 1 and S2). The absence of the core-3 facilitates the formation of core-1 and core-2, implying that core-3 either energetically or through steric interactions hinders these transitions. The increased β-strand propensity in the core-1 region of FUS-LC-N compared to FUS-LC further shows that the presence of core-3 suppresses the cores with greater stability (see Fig. S10). Importantly, the relative stability between core-1 and core-2, $\Delta F_{1,2}$, increases from -0.93 ± 0.13 kcal/mol in FUS-LC to -2.24 ± 0.06 kcal/mol in FUS-LC-N. The enhanced stability suggests that core-3 exerts a greater destabilizing effect on core-1 than on core-2, further supporting its role in modulating fibril formation pathways.

Implication for FUS-LC assembly: Our results confirm that the thermodynamically more stable core-1 compared to core-3 ($\Delta F_{1,3S} = -2.70 \pm 0.10$ kcal/mol) forms significantly later than core-3 in FUS-LC (see Fig. 1I). Kinetic studies revealed that core-3 forms within 50 hours of incubation (see Fig. in Ref. [21]). In contrast, core-1 forms only after an incubation period of 5–7 days [20]. The combination of kinetic and thermodynamic findings suggests that core-3 must transition into core-1 over time. Although core-1 and core-3 do not coexist in FUS-LC, our kinetic studies of FUS1-163 shows that the absence of the core-3 region facilitates the formation of core-1. The time to form fibril-like core-1 structure decreases by a factor of $\approx$4 in FUS-LC-N compared to FUS-LC (compare Figs. 1 I and J). This implies that, at the early stages of FUS-LC assembly, core-3 likely hinders core-1 formation. The mutual exclusivity of core-1 and core-3 was previously shown in experiments [21] using ALS-associated mutations.

Our simulations predict that in FUS-LC core-2 fibril-like structures form after core-3 but before core-1. Experiments demonstrated that core-2 is not required for hydrogel binding [21], which accords well with our simulations that show that there are a limited number of interchain contacts between the core-2 regions in FUS-LC monomers within the condensates (see Fig. 3). The scarcity of favorable interactions likely impedes the formation of stable core-2 fibrils. The destabilization of core-2 could also arise from entropic repulsion between the "fuzzy coats" surrounding the ordered regions of the non-polymorphic FUS-LS fibrils [20, 23].

Finally, there are similarities between FUS-LS and the low complexity domain of the TAR DNA-binding protein 43 (TDP- 43). In the latter, core-1 ("aggregation" core) and core-2

(about 10 residues outside of core-1) transition from liquid to solid morphology [63]. The ordered cores do not typically coexist (as in FUS-LC) and form on different time scales [64]. It appears that Ostwald's rule of stages should be manifested in TDP-43 as well. The generality of the results discovered here (Ostwald's rule, nucleation and growth of condensates, maturation of droplets into fibrils an extremely long time scales) for other low complexity sequences remains to be established.

Methods

Transition to fibril-like monomer N^* states: Previously, we showed that the monomer ensemble has fingerprints of sparsely populated high free energy conformations with high degree of similarity to monomers in the fibrils [24]. Such conformations, belonging to the N^* state, may be determined using long equilibrium simulations. The N^* states were identified using the structural overlap parameter (see Eq. (2) in the SI). To calculate the rates of transition to the N^* states, we carried out Brownian dynamics simulations in the high friction regime, by setting the solvent viscosity of 10^{-3} Pa.s, which is roughly the water viscosity. The equation of motion of each bead i, derived by neglecting the inertial term in Eq.2 is,

$$\dot{q}_i = -\frac{1}{\gamma}\frac{\partial U_{SOP-IDP}}{\partial q_i} + \Gamma_i. \tag{1}$$

We integrated Eq.1 using the Euler algorithm in the custom LAMMPS code [65]. In the over-damped limit, the natural unit of time is $\tau_{HF} = \frac{\gamma a^2}{k_B T}$, where $\gamma = 6\pi\eta a$, as described previously [66]. With $a = 2.88$ Å, the average bead radius in our model, we obtain $\tau_{HF} =$ 109.5 ps. To derive reliable estimates of the first passage times (FPTs), we generated $N_{sim} =$ 200 independent trajectories for each polymer at 298 K. The duration of each trajectories is 240 ms, which may not sufficient to observe transitions to the N^* states. To correct for this shortcoming the procedure described below to obtain the true estimate of the transition times to the N^* states.

True Mean FPT: Because the duration of the simulation trajectories is finite the transition

from the disordered ground state to the N^* state cannot be observed in all the trajectories. As a result, the mean first passage time is underestimated. To correct for the underestimate, we used the maximum likelihood estimate (MLE) to calculate the true mean FPT, by assuming that the first passage times FPTs are exponentially distributed. The true mean first passage time, denoted as $\langle\tau\rangle$, is given by,

$$\langle\tau\rangle = \frac{\sum_{k=1}^{N_S} \tau_k + N_E \cdot T}{N_O}. \tag{2}$$

The equation above (Eq. 2), derived elsewhere [67] is similar to the approach in [68], accounts for the contributions from events that occur within the duration (T) of the simulation trajectory as well those that do not on T. In Eq. 2, τ_k is the FPT that is less than T, N_S is the number of trajectories, N_E denotes the number of simulations where the FPT exceeds T (indicating cases where the first passage did not occur within T), and N_O is the total number of observed FPTs that are less than T. The robustness of Eq. 2 was tested across different values of N_S and T. It is worth pointing out that the validity of Eq. 2 was established by explicit comparisons to precisely solvable models [67].

Multi-chain simulations: Each simulation run was initiated from different initial configuration of the system consisting of 200 polymers. The initial configurations were derived from five distinct conformations generated in the FUS-LC monomer simulations. The conformations correspond to a cluster in the ground state (see the Methods in SI). The dimensions of the monomers, R_g and R_{ee} values, were close to the ensemble average. Each of the five monomers was replicated 40 times and arranged in a simulation box using PACKMOL software [69], ensuring that there are no interactions between monomers in the initial configuration. The distance between the monomers in the initial configuration was 10-15 nm, which is larger than the average R_{ee} and two-three times larger than the average R_g of FUS-LC (Fig. S11). We use a temperature (T = 209 K) to facilitate droplet formation under equilibrium conditions.

All the simulations were carried out at the salt concentration of 150 mM NaCl. To calculate the inherent structures [48], we quenched the system to a low temperature, by isolating the droplet and then performed Langevin dynamics at $T = 3$ K until the potential

energy converged. Multi-chain simulations were performed on a GPU using OpenMM [70]. The equations of motion were solved using the leap-frog integration scheme with a time step of $h = 0.05\tau$, where the natural time unit is, $\tau = a_0\sqrt{m_0/\varepsilon_0}$. Here, $m_0 = 1$ Da represents the mass scale, $a_0 = 1$ Å is the characteristic length scale, and $\varepsilon_0 = 1$ kcal/mol sets the energy scale. To accelerate sampling, simulations were conducted in the underdamped regime by reducing the effective viscosity of water by a factor of 100. Each trajectory was propagated for a total of 5×10^9 integration steps.

Clustering: For each snapshot, a monomer was considered part of a cluster (i.e., an oligomer or droplet) if it was in contact with at least two other monomers within that cluster. Two monomers were defined to be in contact if there existed at least one pair of residues, i and j, from the respective monomers such that the distance between their side chains was ≤ 0.8 nm.

Fibril structures: To predict the three-dimensional structures of the fibrils, we employed AlphaFold2 (AF2; [31]) through ColabFold, a Google-hosted AlphaFold collaborative Jupyter notebook that facilitates modeling of protein monomers as well as homo- and hetero-oligomeric complexes [71]. ColabFold was run with default parameters, with the number of subunits varied to assess the structural diversity. For AlphaFold3 (AF3; [32]) predictions, we used the AlphaFold Server [32] under default settings. To evaluate the reliability of each predicted structure, we analyzed multiple AlphaFold metrics: the Predicted local distance difference test (pLDDT), which reflects confidence in structure predictions; the Template Modeling (pTM) score and the interface predicted template modeling (ipTM) score, which estimates the overall quality of the complex and the interfacial interactions within complexes, respectively [31]. The pTM and ipTM scores approximate the Template Modeling (TM) score [72], providing insights into the predicted structural accuracy. We present these metrics for all the AlphaFold models in the SI Figs. S14, S16, S17.

Acknowledgments: We are grateful to Steve McKnight and Balaka Mondal for useful comments and critical reading of the manuscript. LD was partly supported by a Fulbright Fellowship. This work was supported by a grant from the National Science Foundation (CHE 2320256), and a grant from the Welch Foundation (F-0019) administered through the Collie-Welch Regents Chair. This work used Bridges-2 [73] at Pittsburgh Supercomputing

Center through allocations bio220054p and bio240056p from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources that have contributed to the research results reported within this paper. URL: `http://www.tacc.utexas.edu`.

References

1. Flory, P. J. *Principles of polymer chemistry* (Cornell university press, 1953).
2. Overbeek, J. T. & Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *Journal of Cellular and Comparative Physiology* **49,** 7–26 (1957).
3. Bungenberg De Jong, H. G. & Kruyt, H. R. Coacervation (partial miscibility in colloid systems). *Proc. Royal Academy of Amsterdam* **32,** 849 (1929).
4. Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F. & Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324,** 1729–1732 (2009).
5. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176,** 419–434 (2019).
6. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews Molecular cell biology* **18,** 285–298 (2017).
7. Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annual Review of Biophysics* **49,** 107–133 (2020).
8. Rippe, K. Liquid–liquid phase separation in chromatin. *Cold Spring Harbor perspectives in biology* **14,** a040683 (2022).
9. Roden, C. & Gladfelter, A. S. RNA contributions to the form and function of biomolecular condensates. *Nature Reviews Molecular Cell Biology* **22,** 183–195 (2021).
10. Trcek, T., Douglas, T. E., Grosch, M., Yin, Y., Eagle, W. V., Gavis, E. R., Shroff, H., Rothenberg, E. & Lehmann, R. Sequence-independent self-assembly of germ granule mRNAs into homotypic clusters. *Molecular cell* **78,** 941–950 (2020).
11. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22,** 196–213 (2021).

12. Thirumalai, D., Kumar, A., Chakraborty, D., Straub, J. E. & Mugnai, M. L. Conformational fluctuations and phases in fused in sarcoma (FUS) low-complexity domain. *Biopolymers* **115,** e23558 (2024).

13. Shea, J.-E., Best, R. B. & Mittal, J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Current opinion in structural biology* **67,** 219–225 (2021).

14. Ranganathan, S., Dasmeh, P., Furniss, S. & Shakhnovich, E. Phosphorylation sites are evolutionary checkpoints against liquid–solid transition in protein condensates. *Proceedings of the National Academy of Sciences* **120,** e2215828120 (2023).

15. Ranganathan, S. & Shakhnovich, E. The physics of liquid-to-solid transitions in multidomain protein condensates. *Biophysical Journal* **121,** 2751–2766 (2022).

16. 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* **1,** 732–743 (2021).

17. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annual review of physical chemistry* **71,** 53–75 (2020).

18. 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* **118,** e2111696118 (2021).

19. Wang, L., Brasnett, C., Borges-Araújo, L., Souza, P. C. & Marrink, S. J. Martini3-IDP: improved Martini 3 force field for disordered proteins. *Nature Communications* **16,** 1–14 (2025).

20. 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* **171,** 615–627 (2017).

21. Kato, M. & McKnight, S. L. The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores. *Proceedings of the National Academy of Sciences* **118,** e2114412118 (2021).

22. 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* **162,** 1066–1077 (2015).

23. Lee, M., Ghosh, U., Thurber, K. R., Kato, M. & Tycko, R. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. *Nature communications* **11,** 5735 (2020).

24. Kumar, A., Chakraborty, D., Mugnai, M. L., Straub, J. E. & Thirumalai, D. Sequence determines the switch in the fibril forming regions in the low-complexity FUS protein and its variants. *The journal of physical chemistry letters* **12,** 9026–9032 (2021).

25. Kato, M., Zhou, X. & McKnight, S. How do protein domains of low sequence complexity work? *RNA* **28,** 3–15 (2022).

26. Baul, U., Chakraborty, D., Mugnai, M. L., Straub, J. E. & Thirumalai, D. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *The Journal of Physical Chemistry B* **123,** 3462–3474 (2019).

27. Mugnai, M. L., Chakraborty, D., Maksudov, F., Kumar, A., Nguyen, H. T., Zeno, W., Stachowiak, J. C., Straub, J. E. & Thirumalai, D. Sizes, conformational fluctuations, and SAXS profiles for Intrinsically Disordered Proteins. *Prot. Sci.* **34,** e70067 (2025).

28. Straub, J. E. & Thirumalai, D. Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annual review of physical chemistry* **62,** 437–463 (2011).

29. Li, M. S., Co, N. T., Reddy, G., Hu, C.-K., Straub, J. & Thirumalai, D. Factors governing fibrillogenesis of polypeptide chains revealed by lattice models. *Physical review letters* **105,** 218101 (2010).

30. Zhuravlev, P. I., Reddy, G., Straub, J. E. & Thirumalai, D. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *Journal of molecular biology* **426,** 2653–2666 (2014).

31. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

32. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D. & Jumper, J. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature,* 1–3 (2024).

33. Welch, B. L. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* **34,** 28–35 (1947).

34. Woods, E. J. & Wales, D. J. Analysis and interpretation of first passage time distributions featuring rare events. *Physical Chemistry Chemical Physics* **26,** 1640–1657 (2024).

35. Berkeley, R. F., Kashefi, M. & Debelouchina, G. T. Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophysical Journal* **120,** 1276–1287 (2021).

36. Levin, A., Mason, T. O., Adler-Abramovich, L., Buell, A. K., Meisl, G., Galvagnion, C., Bram, Y., Stratford, S. A., Dobson, C. M., Knowles, T. P., *et al.* Ostwald's rule

of stages governs structural transitions and morphology of dipeptide supramolecular polymers. *Nature communications* **5,** 5219 (2014).

37. Yokosawa, K., Kajimoto, S., Shibata, D., Kuroi, K., Konno, T. & Nakabayashi, T. Concentration quantification of the low-complexity domain of fused in sarcoma inside a single droplet and effects of solution parameters. *The Journal of Physical Chemistry Letters* **13,** 5692–5697 (2022).

38. 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. *Molecular cell* **60,** 231–241 (2015).

39. Debenedetti, P. G. Metastable liquids: concepts and principles (1996).

40. Sugimoto, T. General kinetics of Ostwald ripening of precipitates. *Journal of Colloid and Interface Science* **63,** 16–26 (1978).

41. Lifshitz, I. M. & Slyozov, V. V. The kinetics of precipitation from supersaturated solid solutions. *Journal of physics and chemistry of solids* **19,** 35–50 (1961).

42. Wagner, C. Theorie der alterung von niederschlägen durch umlösen (Ostwald-reifung). *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie* **65,** 581–591 (1961).

43. Emmanouilidis, L., Bartalucci, E., Kan, Y., Ijavi, M., Escura Pérez, M., Afanasyev, P., Boehringer, D., Zehnder, J., Parekh, S. H., Bonn, M., Michaels, T. C. T., Wiegand, T. & Allain, F. H.-T. A solid beta-sheet structure is formed at the surface of FUS droplets during aging. *Nature Chemical Biology,* 1–9 (2024).

44. Wilson, C. B., Lee, M., Yau, W.-M. & Tycko, R. Conformations of a low-complexity protein in homogeneous and phase-separated frozen solutions. *Biophysical journal* **123,** 4097–4114 (2024).

45. 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. *Nature structural & molecular biology* **26,** 637–648 (2019).

46. Esteban-Hofer, L., Emmanouilidis, L., Yulikov, M., Allain, F. H.-T. & Jeschke, G. Ensemble structure of the N-terminal domain (1–267) of FUS in a biomolecular condensate. *Biophysical Journal* **123,** 538–554 (2024).

47. Joshi, A., Walimbe, A., Avni, A., Rai, S. K., Arora, L., Sarkar, S. & Mukhopadhyay, S. Single-molecule FRET unmasks structural subpopulations and crucial molecular events during FUS low-complexity domain phase separation. *Nature Communications* **14,** 7331 (2023).

48. Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **225,** 983–989 (1984).

49. McCall, P. M., Kim, K., Fritsch, A. W., Iglesias-Artola, J., Jawerth, L., Wang, J., Ruer, M., Peychl, J., Poznyakovskiy, A., Guck, J., *et al.* Quantitative phase microscopy enables precise and efficient determination of biomolecular condensate composition. *BioRxiv,* 2020–10 (2020).

50. Maltseva, D., Chatterjee, S., Yu, C.-C., Brzezinski, M., Nagata, Y., Gonella, G., Murthy, A. C., Stachowiak, J. C., Fawzi, N. L., Parekh, S. H., *et al.* Fibril formation and ordering of disordered FUS LC driven by hydrophobic interactions. *Nature chemistry* **15,** 1146–1154 (2023).

51. Wojciechowska, A. W., Wojciechowski, J. W. & Kotulska, M. Non-standard proteins in the lenses of AlphaFold3-case study of amyloids. *bioRxiv,* 2024–07 (2024).

52. Law, S. M., Frank, A. T. & Brooks III, C. L. PCASSO: A fast and efficient Cα-based method for accurately assigning protein secondary structure elements. *Journal of computational chemistry* **35,** 1757–1761 (2014).

53. Li, M. S., Klimov, D., Straub, J. & Thirumalai, D. Probing the mechanisms of fibril formation using lattice models. *The Journal of chemical physics* **129** (2008).

54. Bertrand, E., Demongin, C., Dobra, I., Rengifo-Gonzalez, J. C., Singatulina, A. S., Sukhanova, M. V., Lavrik, O. I., Pastré, D. & Hamon, L. FUS fibrillation occurs through a nucleation-based process below the critical concentration required for liquid–liquid phase separation. *Scientific Reports* **13,** 7772 (2023).

55. Oranges, M., Jash, C., Golani, G., Seal, M., Cohen, S. R., Rosenhek-Goldian, I., Bogdanov, A., Safran, S. & Goldfarb, D. Core-shell model of the clusters of CPEB4 isoforms preceding liquid-liquid phase separation. *Biophysical Journal* **123,** 2604–2622 (2024).

56. Ragonis-Bachar, P., Axel, G., Blau, S., Ben-Tal, N., Kolodny, R. & Landau, M. What can AlphaFold do for antimicrobial amyloids? *Proteins: Structure, Function, and Bioinformatics* **92,** 265–281 (2024).

57. Chakravarty, D., Schafer, J. W., Chen, E. A., Thole, J. F., Ronish, L. A., Lee, M. & Porter, L. L. AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nature communications* **15,** 7296 (2024).

58. Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundström, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M. & Kay, L. E. Structure of an intermediate state in protein folding and aggregation. *Science* **336,** 362–366 (2012).

59. Levin, A., Mason, T. O., Adler-Abramovich, L., Buell, A. K., Meisl, G., Galvagnion, C., Bram, Y., Stratford, S. A., Dobson, C. M., Knowles, T. P. & Gazit, E. Ostwald's rule of stages governs structural transitions and morphology of dipeptide supramolecular polymers. *Nature communications* **5,** 5219 (2014).

60. Chakraborty, D., Straub, J. E. & Thirumalai, D. Energy landscapes of Aβ monomers are sculpted in accordance with Ostwald's rule of stages. *Science Advances* **9,** eadd6921 (2023).

61. Ceccon, A., Tugarinov, V., Torricella, F. & Clore, G. M. Quantitative NMR analysis of the kinetics of prenucleation oligomerization and aggregation of pathogenic huntingtin exon-1 protein. *Proceedings of the National Academy of Sciences* **119,** e2207690119 (2022).

62. Jeon, J., Yau, W.-M. & Tycko, R. Early events in amyloid-β self-assembly probed by time-resolved solid state NMR and light scattering. *Nature communications* **14,** 2964 (2023).

63. Fonda, B. D., Jami, K. M., Boulos, N. R. & Murray, D. T. Identification of the rigid core for aged liquid droplets of an RNA-binding protein low complexity domain. *Journal of the American Chemical Society* **143,** 6657–6668 (2021).

64. Jami, K. M., Farb, D. C., Osumi, K. M., Shafer, C. C., Criscione, S. & Murray, D. T. Small heat shock protein HSPB8 interacts with a pre-fibrillar TDP43 low complexity domain species to delay fibril formation. *bioRxiv,* 2025–01 (2025).

65. Shi, G., Liu, L., Hyeon, C. & Thirumalai, D. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature communications* **9,** 3161 (2018).

66. Veitshans, T., Klimov, D. & Thirumalai, D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding and Design* **2,** 1–22 (1997).

67. Shi, G., Shin, S. & Thirumalai, D. Static Three-Dimensional Structures Determine Fast Dynamics Between Distal Loci Pairs in Interphase Chromosomes. *bioRxiv,* 2025–01 (2025).

68. Al-Athari, M. F. M. Estimation of the mean of truncated exponential distribution. *Journal of Mathematics and Statistics* **4,** 284 (2008).

69. Martınez, L., Andrade, R., Birgin, E. G. & Martınez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem* **30,** 2157–2164 (2009).

70. Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **13,** e1005659 (2017).

71. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. ColabFold: making protein folding accessible to all. *Nature methods* **19,** 679–682 (2022).

72. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics* **57,** 702–710 (2004).

73. Brown, S. T., Buitrago, P., Hanna, E., Sanielevici, S., Scibek, R. & Nystrom, N. A. in *Practice and Experience in Advanced Research Computing 2021: Evolution Across All Dimensions* 1–4 (2021).

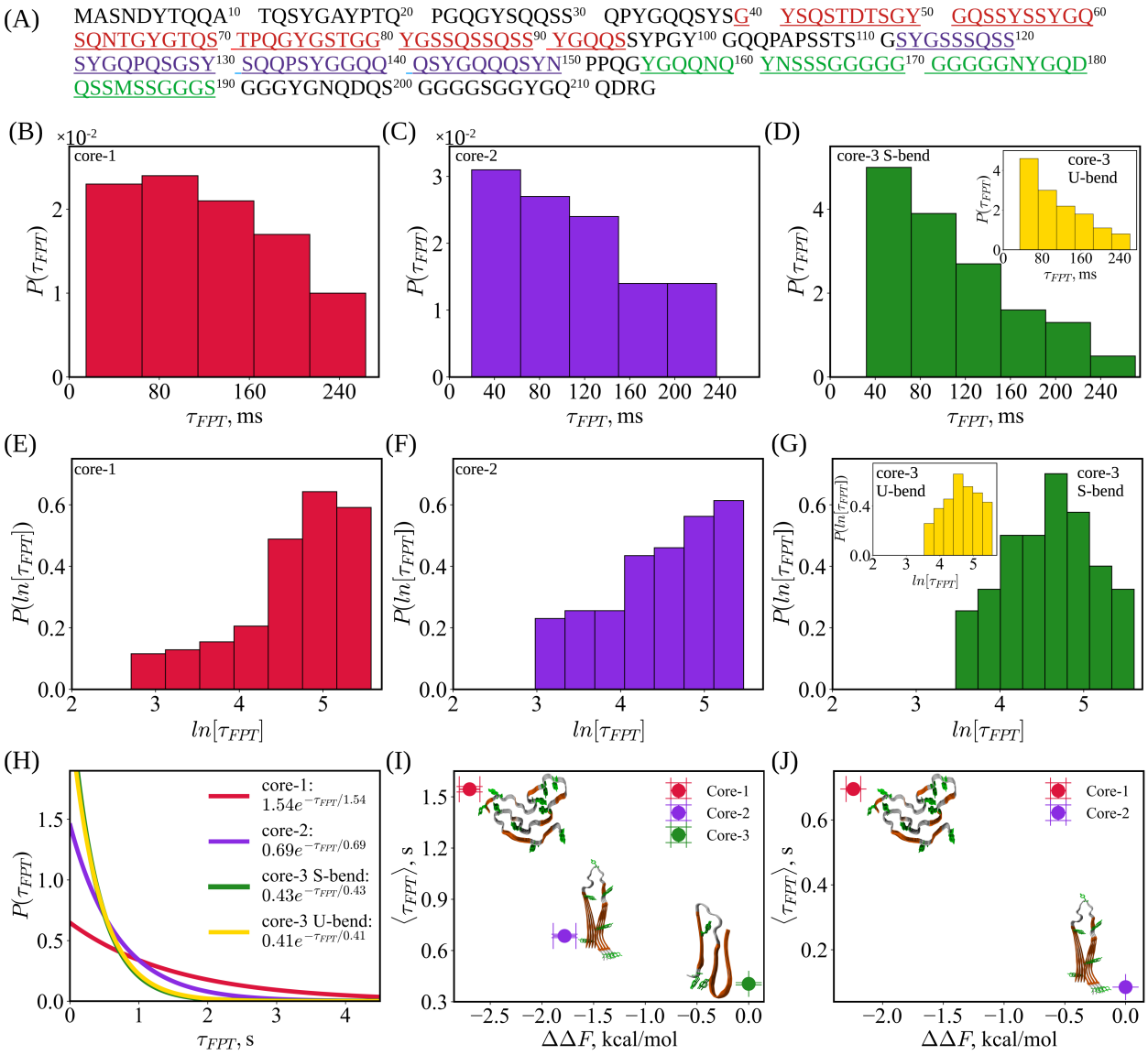

Figure 1: **Ostwald's rule of stages:** (A) The FUS-LC sequence using a one-letter code for the amino acids. Core-1 residues 39-95 are in red, core-2 residues are highlighted in purple, and those forming core-3 are in green. Panels B-D: Distributions of first-passage times (FPTs) for the transition from the disordered ground state to the fibril-like N^* states in FUS-LC (residues 1–214). Shown are the τ_{FPT} distributions sampled from Brownian dynamics simulations (see Methods) for the transition from the free energy ground state of FUS-LC to the core-1 fibril-like structure (panel B), to the core-2 fibril-like structure (panel C), and to the core-3 S-bend and U-bend (inset) fibril-like structures (panel D) predicted by AF3. Panels E-G: Distributions of logarithm of first passage times (FPTs) $ln[\tau_{FPT}]$ for the transition from the disordered ground state to the fibril-like N^* states. (H) Theoretical distributions of FPTs calculated using the correction formula (see Eq. 2). (I): MFPTs as a function of relative stability with respect to core-3 (S-bend) for FUS-LC. The inset shows the experimental structures of core-1 (PDB ID: 5W3N [20]) and core-2 (PDB ID: 6XFM [23]), along with the predicted S-bend structure of core-3. (J) MFPTs as a function of relative stability with respect to core-2 for FUS-LC-N (residues 1–163). The distributions of FPTs for the transition from the disordered ground state to the fibril-like N^* states for FUS-LC-N are show in Fig. S2.

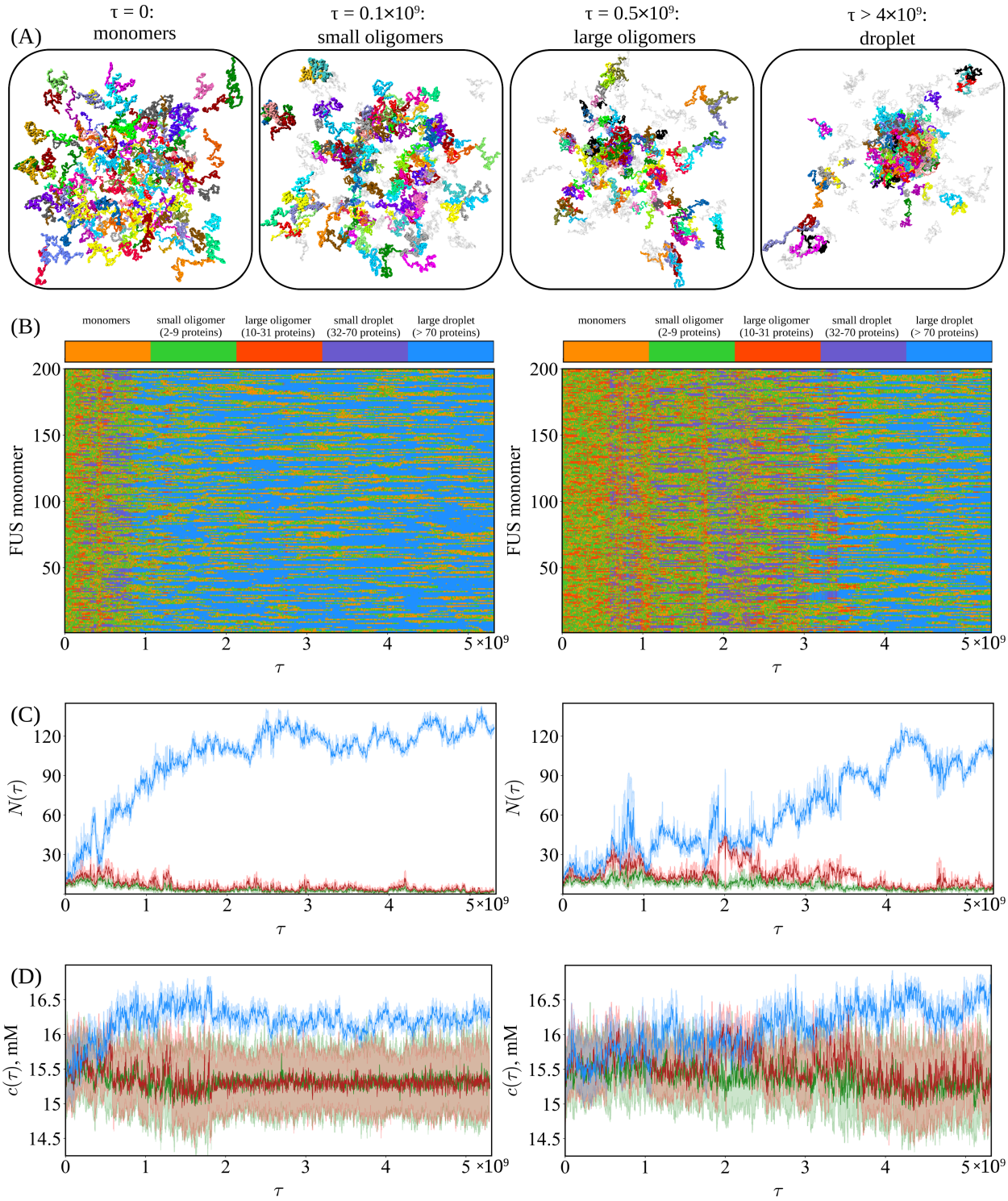
(A)
τ = 0:
monomers
τ = 0.1×10⁹:
small oligomers
τ = 0.5×10⁹:
large oligomers
τ > 4×10⁹:
droplet
(B)
monomers
small oligomer (2-9 proteins)
large oligomer (10-31 proteins)
small droplet (32-70 proteins)
large droplet (> 70 proteins)
FUS monomer
τ
(C)
N(τ)
(D)
c(τ), mM

Figure 2: **Phase separation in FUS-LC.** (A) Representative snapshots illustrating the progression of phase separation in a system of 200 FUS-LC chains. The first panel shows the initial homogeneous configuration in the single dilute phase. Subsequent panels depict the emergence of small oligomeric nuclei, followed by their growth into a larger oligomer. This structure eventually matures into a droplet, representing the equilibrium state characterized by the coexistence of condensed and dilute phases. FUS-LC proteins in the monomeric state are rendered in transparent gray. (B) The temporal progression of individual FUS protein molecules is shown as a sequence of discrete states over the course of the simulation τ is a simulation time step, which is defined in Methods. Conformational states of the proteins are represented through a color-coded scheme. Two independent trajectories are displayed. (C) Temporal evolution of the sizes of the three largest droplets (blue, red, and green) corresponding to the trajectories in panel B. Droplet size is defined as the number of FUS-LC chains contained within each droplet. (D) Protein mass concentration c as a function of simulation time for the droplets shown in panel C.

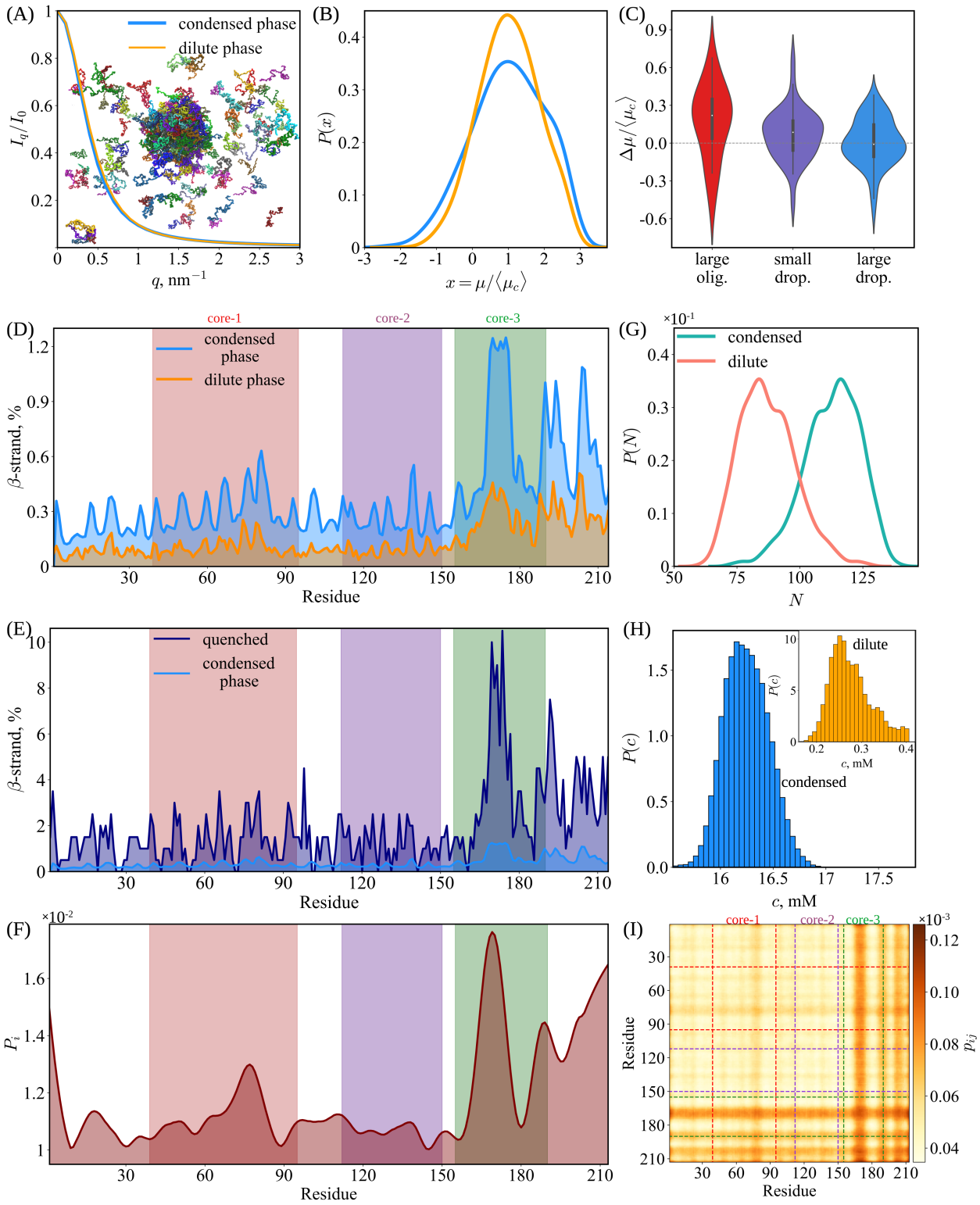
(A)
condensed phase
dilute phase
I_q/I_0
q, nm^{-1}
(B)
$P(x)$
$x = \mu/\langle\mu_c\rangle$
(C)
$\Delta\mu/\langle\mu_c\rangle$
large olig.
small drop.
large drop.
(D)
core-1
core-2
core-3
condensed phase
dilute phase
β-strand, %
Residue
(E)
quenched
condensed phase
β-strand, %
Residue
(F)
$\times 10^{-2}$
P_i
Residue
(G)
$\times 10^{-1}$
condensed
dilute
$P(N)$
N
(H)
dilute
condensed
$P(c)$
c, mM
(I)
core-1
core-2
core-3
$\times 10^{-3}$
Residue
p_{ij}

Figure 3: **Equilibrium characteristics of FUS-LC condensates.** (A) Small-angle X-ray scattering (SAXS) profiles for FUS-LC polymers in the condensed phase (within the phase separated droplets) and the dilute phase (outside droplets). The scattering intensity $I(q)$ as a function of the scattering vector q. The inset presents representative images from a simulation of 200 FUS-LC chains, illustrating the coexistence of dilute and condensed phases. (B) Distributions of the chemical potential in the dispersed and condensed phases, scaled by the average chemical potential in the condensed phase, $\langle \mu_c \rangle$. (C) Distributions of the chemical potential difference between condensed and dilute phases, defined as $\Delta\mu = \langle \mu_c \rangle - \langle \mu_d \rangle$, scaled by $\langle \mu_c \rangle$ for assemblies consisting of different numbers of chains, where $\langle \mu_c \rangle$ and $\langle \mu_d \rangle$ denote the average chemical potentials in the condensed and dilute phases, respectively. (D) Ensemble-averaged residue-wise percentages of β-strand content for FUS-LC monomers in homogeneous and phase-separated environments. Vertical dashed lines are positioned at the residues corresponding boundaries of the core regions: core-1 (residues 39-95), core-2 (residues 112-150), and core-3 (residues 155-190). (E) Amino acid dependent propensity for β-strand formation calculated from the inherent structure of droplet. The propensity values are averaged over all the chains in the condensate. (F) Cumulative probability for each residue of FUS-LC to form inter-chain contacts within the condensate. For a given residue i, P_i is obtained by summing the average probabilities of all its intermolecular contacts with residues j (see panel I). (G) Droplet size distributions in the condensed and dilute phases. (H) Distributions of protein concentration in the condensed and dilute phases. (I) Intermolecular contact probability p_{ij} between residues i and j, calculated over the equilibrium portion of the trajectory ($t > 3 \times 10^9 \tau$).

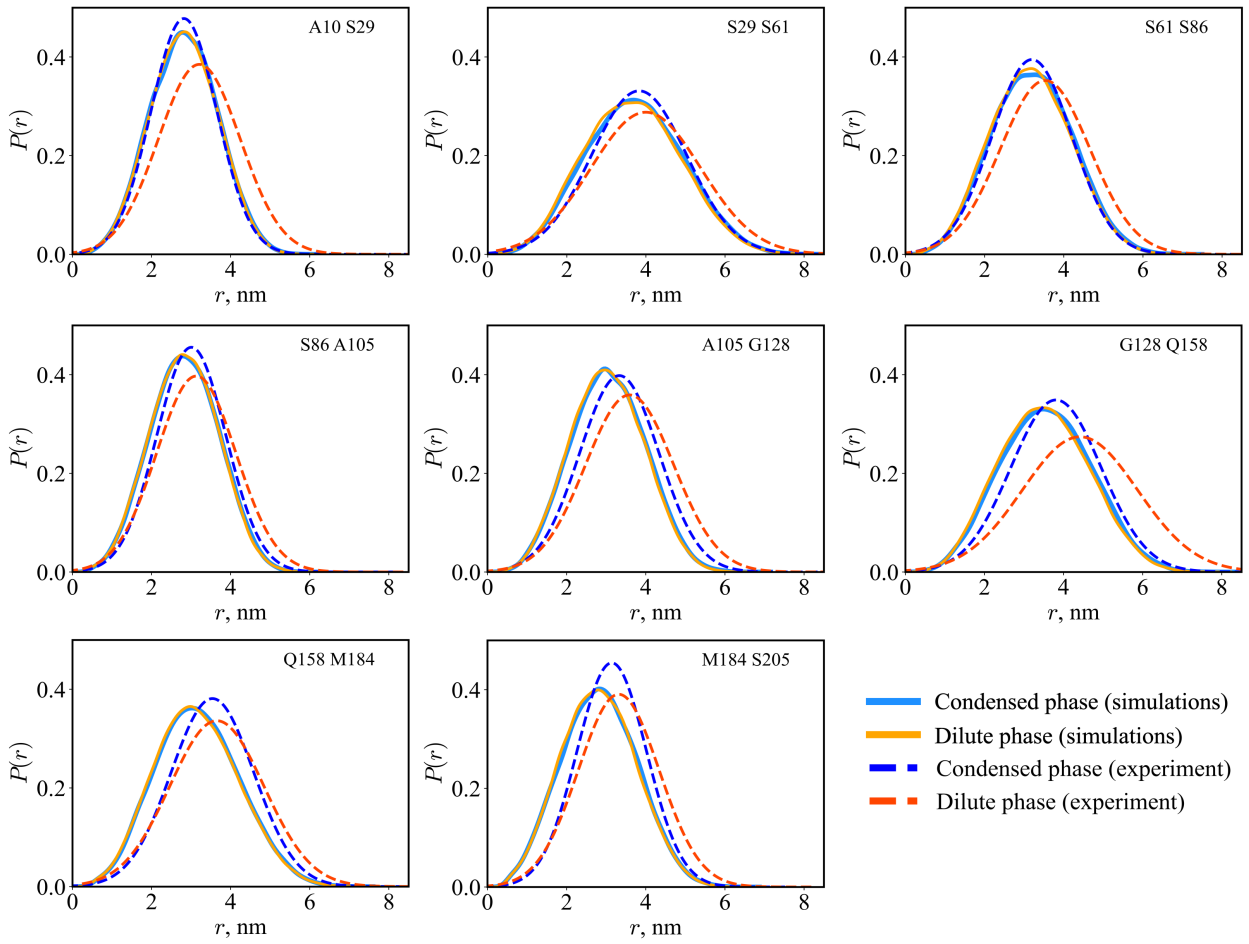

Figure 4: **Conformations of monomers in dilute and dense phase are similar.** Distance distributions between select residues (base on an experimental study [46]) within the same FUS-LC chain for chains in the dilute phase and those within the condensed droplet from simulations. These distributions are compared to experimental distance distributions for $FUS_{1\text{-}267}$, extracted from Fig. 2 in Ref. [46]. The corresponding mean distances and standard deviations are provided in Table S4.

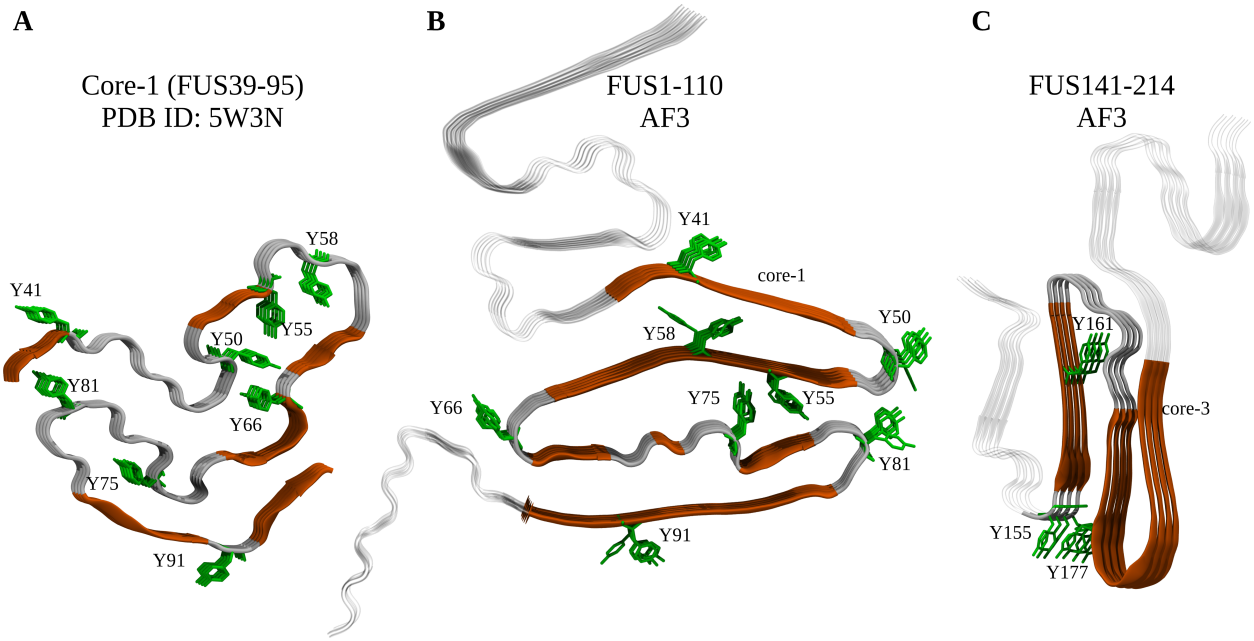

Figure 5: **Structural models of core-1 and core-3 fibrils predicted by AlphaFold3.** (A) Structure of core-1 (residues 39–95) resolved by solid-state NMR [20]. Tyrosine (TYR) side chains are shown as sticks, with β-sheets colored brown and turn/coil regions in gray. (B, C) AlphaFold3 predictions for pentameric assembly of $FUS_{1\text{-}110}$ (B) and tetrameric assembly of $FUS_{141\text{-}214}$ (C). The color scheme is the same as in panel (A). Core-1 residues (39–95) in (B) and core-3 residues (155–190) in (C) are depicted as in panel (A), while the remaining fibril regions are rendered transparent for clarity.