REVIEW

Conformational fluctuations and phases in fused in sarcoma (FUS) low-complexity domain

D. Thirumalai [1,2] ⓘ | Abhinaw Kumar [1] ⓘ | Debayan Chakraborty [1] ⓘ |
John E. Straub [3] ⓘ | Mauro L. Mugnai [4] ⓘ

[1]Department of Chemistry, The University of Texas at Austin, Austin, Texas, USA

[2]Department of Physics, The University of Texas at Austin, Austin, Texas, USA

[3]Department of Chemistry, Boston University, Boston, Massachusetts, USA

[4]Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA

Correspondence
D. Thirumalai, Department of Chemistry, The University of Texas at Austin, Austin, TX 78712, USA.
Email: dave.thirumalai@gmail.com

Abstract

The well-known phenomenon of phase separation in synthetic polymers and proteins has become a major topic in biophysics because it has been invoked as a mechanism of compartment formation in cells, without the need for membranes. Most of the coacervates (or condensates) are composed of Intrinsically Disordered Proteins (IDPs) or regions that are structureless, often in interaction with RNA and DNA. One of the more intriguing IDPs is the 526-residue RNA-binding protein, Fused in Sarcoma (FUS), whose monomer conformations and condensates exhibit unusual behavior that are sensitive to solution conditions. By focussing principally on the N-terminus low-complexity domain (FUS-LC comprising residues 1–214) and other truncations, we rationalize the findings of solid-state NMR experiments, which show that FUS-LC adopts a non-polymorphic fibril structure (core-1) involving residues 39–95, flanked by fuzzy coats on both the N- and C-terminal ends. An alternate structure (core-2), whose free energy is comparable to core-1, emerges only in the truncated construct (residues 110–214). Both core-1 and core-2 fibrils are stabilized by a Tyrosine ladder as well as hydrophilic interactions. The morphologies (gels, fibrils, and glass-like) adopted by FUS seem to vary greatly, depending on the experimental conditions. The effect of phosphorylation is site-specific. Simulations show that phosphorylation of residues within the fibril has a greater destabilization effect than residues that are outside the fibril region, which accords well with experiments. Many of the peculiarities associated with FUS may also be shared by other IDPs, such as TDP43 and hnRNPA2. We outline a number of problems for which there is no clear molecular explanation.

KEYWORDS
fibril morphology depends on length, FUS excited states, phosphorylation is site-specific

1 | INTRODUCTION

Phase separation in polymers dispersed in a solvent has been known for a long time.[1] If the interaction between the solvent and the monomers is not favorable, phase separation occurs, producing a polymer rich phase that coexists with a polymer poor phase. If this process occurs at equilibrium, it can be described using the venerable mean field Flory-Huggins (FH) theory,[1–3] which becomes increasingly accurate as the degree of polymerization increases, because the mixing entropy due to polymers is $\propto \frac{1}{n}$ (n is the degree of polymerization). Extension of the FH theory to polyelectrolyte solutions was given by Overbeek and Voorn,[4] who set out to explain the remarkable observations by Bugenberg de Jong, showing complex coacervation (a term coined in 1929[5]) in polyelectrolytes. The dense phase in a two-component
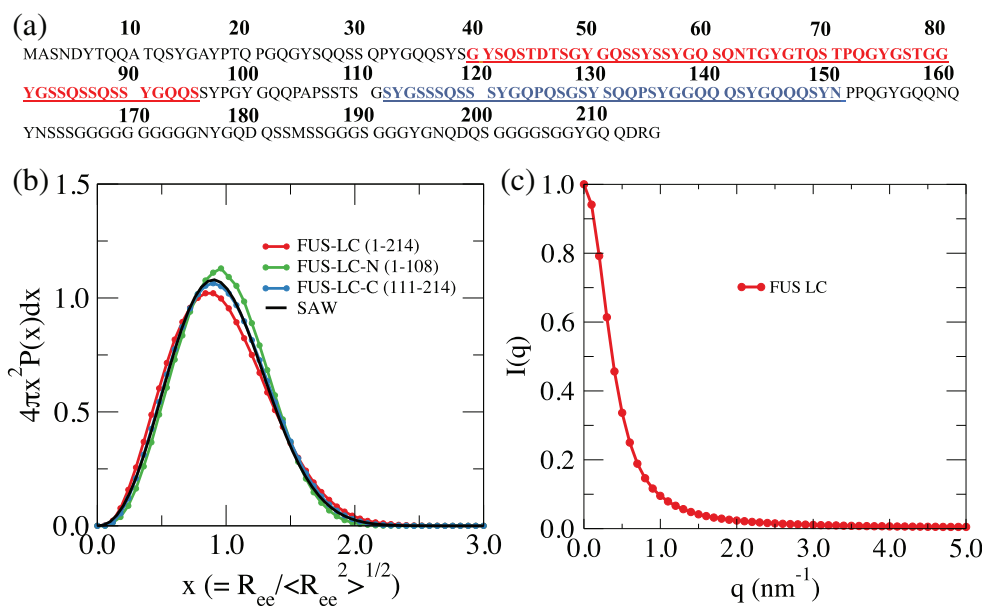
FIGURE 1 (a) Sequence of the Fused in Sarcoma (FUS) Low-Complexity (LC) domain. (b) Distributions of the scaled end-end distance for different FUS-LC variants exhibit the universal Flory random coil behavior. The abbreviation SAW stands for Self-Avoiding walk, an alternate phrase used to describe random coils in the polymer literature. (c) The scattering profile for FUS-LC is reminiscent of random coils.

polyelectrolyte system (gum Arabic, a complex mixture of glycoproteins and polysaccharides, and gelatin) consisted of the polymers with water, and the dilute phase was a solution containing one or both polymers. Bugenberg de Jong pointed out that biological membranes and the complex coacervates may share common characteristics. Indeed, Overbeek, and Voorn suggested potential connections to complex coacervation in biology in the concluding section of their seminal paper.[4] However, the demonstration that condensate (or equivalently coacervate) formation involving biomolecules could produce organelles without membranes by a phase separation mechanism, has given a new perspective on how the cytosol is organized.[6-9] A key paper[6] that appeared in 2009 has set in motion a growing enterprise that has certainly generated considerable interest.

Many of the proteins that produce cellular condensates are intrinsically disordered or contain regions that do not adopt discernible structures, and are often in complex with RNA. Because condensate (we will use this modern terminology as opposed to coacervates as it seems to be the accepted description in biophysics and biology) formation leads to high local density of biomolecules, it could profoundly affect a number of functions, as discussed recently by Alberti and Hyman (see Figure 1 in reference [8]). A number of other articles have described the functional implications of condensate formation.[10,11] These and scores of others, have provided great impetus in developing theories and computational models to describe many aspects of phase separation as well as the mechanical and relaxation properties of condensates. It is hard not to notice that a number of researchers are dedicated in elucidating the principles of condensate formation from molecular level to mesoscopic scales.

In this perspective, we focus only on the RNA-binding protein, Fused in Sarcoma (FUS), which has become a poster child for IDRs

(intrinsically disordered protein that also has ordered regions). The N-terminal region of the 526-residue FUS contains the low-complexity domain, FUS-LC, (residues 1–214) which is rich in QGSY (Q = Glutamine, G = Glycine, S=Serine, and Y = Tyrosine) repeats. Residues beyond 214 contain the RNA-binding domain, whose role in creating various morphologies of the condensates are not fully understood. Here, we focus on the effects of sequence and length on the conformations and phases of FUS-LC and several truncated variants. We first describe the properties of the monomer ensembles of several constructs of FUS-LC, with particular focus on the sparsely populated excited states, referred to as N∗. Interestingly, the N∗ states reveal characteristics that are found in the fibrils of FUS-LC[12] and the truncated N-terminal variant.[13] We find that effects of phosphorylation are site specific,[14,15] as noted in the experiments.[12] Phosphorylation of residues within the fibril region (residues 39–95) in FUS-LC has a greater impact on the stability of the fibril than phosphorylation of residues outside the core. Our results for the FUS monomer have important implications in the early events of condensate formation, as well as the aging of condensates into solid-like aggregates. Finally, we also raise a few puzzles that require additional scrutiny through experimentation and well-designed computations.

2 | RESULTS

2.1 | Sequence analysis

The 214-residue FUS-LC is rich in QGSY. Out of the 214 residues, 174 are from only 4 amino acids (the numbers are: Glutamine, Q = 43, Serine, S = 51, Tyrosine, Y = 27, and Glycine, G = 53). There are only a few charged residues (6 in total). Thus, electrostatic

interactions are negligible, except when certain residues (S and or Y) are phosphorylated or mutated (G156E).[16,17] Because the 1–214 region is abundant in these four residues, with no special pattern along the sequence, FUS-LC is a low-complexity sequence, which may be quantified using sequence or word entropy. FUS-LC is often likened to prion-domains found in yeast prion proteins, because it is enriched in uncharged polar amino acids, as well as glycine.[18,19] The basis for picturing FUS-LC as a prion-like domain (PLD) comes from sequence analysis[20] and experiments on Sup35 and Ure2p from *Saccharomyces Cerevisae*. Among the three major domains of Sup35 (yeast prion), residues 1–123 comprising the N-domain, also referred to as the prion domain, is rich in amino acids N and Q with very few charged residues. The N-domain alone could form self-propagating (by the Griffith mechanism[21]), inheritable amyloids.[22,23] Thus, from the perspective of sequence and the ability to self-propagate, the Sup35 prion domain is different from FUS-LC. Despite the differences in sequence composition between FUS-LC and Sup35 (we do not include mammalian prions whose ground state forms an ordered structure in the C-terminus), the FUS-LC has been christened as a PLD[16]—a terminology that we adopt here.

2.2 | The ground states of FUS-LC and FUS-LC-C are random coils

Like many other IDPs, the ground state of FUS-LC and a number of variants can be described as polymers in good solvents (Flory random coils), which from a polymer theory perspective is surprising but it is what it is. Because these sequences contain hydrophobic and polar residues one would expect that there ought to be detectable deviations from random coil behavior. That this is not so, at least in global properties (radius of gyration for example) is somewhat surprising. These findings are consistent with the narrow chemical shift dispersions of monomers in solution[24] in a truncated variant (residues 1–163). In particular, the C_α and C_β chemical shifts do not deviate significantly from assignments based on what is expected for known random coil structures. Moreover, it was concluded that the chemical shifts in the FUS-LC domain do not change significantly even in the full length FUS that includes the RNA-binding domain.[24] The finding that FUS-LC is disordered is in accord with the random coil behavior observed for many other IDPs.[25] In a number of IDPs, the radius of gyration (R_g), as well as the hydrodynamic radius (R_h) follow Flory's scaling law, $R_g \approx a_D N_T^\nu$ and $R_h \approx N_T^\nu$, where the exponent $\nu \approx 0.58$, and N_T is the protein length.[25]

In order to ascertain if FUS-LC and its variants exhibit the characteristics of a random coil, we generated a monomer conformation ensemble (MCE) of FUS-LC and its truncated variants, FUS-LC-C (residues 110–214) and FUS-LC-N (residues 1–108) using simulations based on the SOP-IDP model.[14] Analyses of the MCE shows that both these constructs are devoid of secondary structures, but exhibit the properties of a homopolymer in a good solvent. For a Flory random coil, the distribution of $x = R_{ee}/\langle R_{ee}^2 \rangle^{1/2}$ should exhibit the following universal behavior,

$$P(x) \sim Cx^g \exp(-Bx^\delta), \quad (1)$$

where B and C are constants, the correlation hole exponent, $g \approx 0.28$, and $\delta = 1/(\nu - 1)$. The distributions $4\pi x^2 P(x)$ for FUS-LC, FUS-LC-N, and FUS-LC-C obtained from simulations show little deviation from Equation 1 (Figure 1b), thus establishing that the ground states of the three constructs behave as polymers in a good solvent. Moreover, the calculated structure factor, $I(q)$, has the characteristic behavior of random coils (Figure 1c).[26] It would be interesting to measure $I(q)$ in order to validate our prediction. Thus, we surmise that the ground states of the three constructs behave as polymers in a good solvent, which also supports the conclusions reached for the truncated (residues 1–163) FUS-LC.[24]

2.3 | Length of FUS-LC determines fibril morphology

In the amyloid and the phase separation fields, there have been attempts to discern the propensity of proteins and IDPs (or IDRs) to self-associate purely on the basis of sequence (sequence gazing),[27,28] an exercise that is not only difficult but also may be incorrect because environmental factors (pH, temperature, and presence of crowding agents) modify the energy landscape substantially.[29,30] The difficulty may be illustrated by contrasting the changes in the fibril structures by systematically chopping the sequences in FUS-LC. For example, solid-state NMR experiments have shown that the prion domain in Sup35 adopts parallel β-sheet structures that lie perpendicular to the fibril axis,[31] much like many other amyloid fibrils (e.g., $A\beta_{42}$[32] and mammalian prions see e.g., References [33,34]). In addition, the structure of the N-domain is not significantly altered in the construct containing both the N-domain and the highly charged M-domain (residues 124–253).[31]

In sharp contrast, the formation of ordered fibril structures in FUS-LC depends on the length of the sequence. The 214-residue forms a non-polymorphic fibril (core-1 spanning residues 39–95) containing a S-bend structure.[12] The remaining residues are disordered, and flank the two sides forming a fuzzy-coat or brush-like structure. In FUS-LC-C (residues 110–214) residues 112–150 also form a non-polymorphic structure but with an entirely different morphology. The protofilament is a dimer with each monomer resembling a U-bend.[13] The dramatic length dependent variation in the fibril structures of FUS-LC is puzzling, and lacks a theoretical explanation, although entropic repulsion between the fuzzy coats is likely at play.[13]

2.4 | Excited (N^*) states of monomer ensembles have fibril-like structures

We have shown in a number of publications that amyloid-like structures[35–37] are the excited states within the MCE. The excited states (N^* states) are sparsely populated (as low as 2% or less), and at present can only be detected using relaxation dispersion NMR experiments[38,39] or potentially using single molecule pulling experiments,
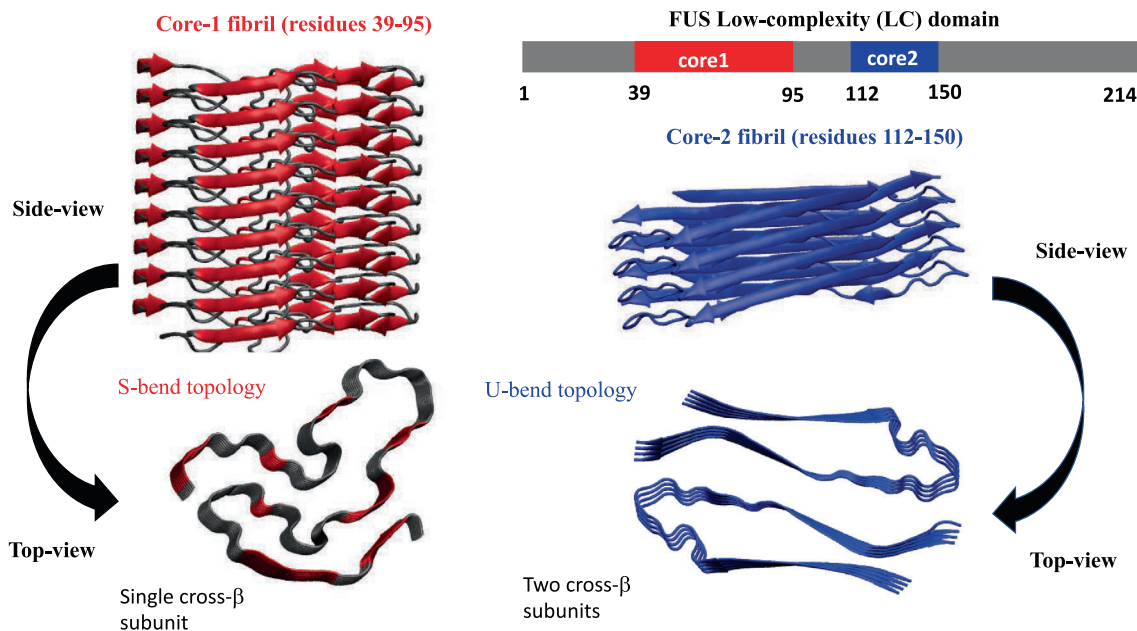
4 of 11

Core-1 fibril (residues 39-95)

FUS Low-complexity (LC) domain

core1 core2

1   39   95   112   150   214

Core-2 fibril (residues 112-150)

Side-view

S-bend topology

Top-view

Single cross-β subunit

Side-view

U-bend topology

Top-view

Two cross-β subunits

FIGURE 2 Length-dependent structural polymorphism in FUS-LC. For the full-length FUS-LC, a S-bend fibril is formed by residues 39 to 95.[12] On the other hand, a truncated variant, FUS-LC-C (residues 110–124) forms a U-bend fibril,[13] with each protofilament consisting of two interacting subunits. In this structure, fibril core is formed by residues 112–150.

as proposed using computations.[40] We showed using simulations[14] that the N* states, bearing striking resemblances to the S-bend (core-1) and the U-bend (core-2) fibril structures, are populated as excited states within the MCE. We summarize the key results from simulations using a simple model before embarking on the difficult multi-chain problem.

Based on a number of studies on amyloid formation in Aβ peptides,[35,37] we theorized that the entire free energy landscape, especially the lowly populated excited N* states bearing a strong resemblance to the ordered structures in the fibrils, controls the emergent structural organization and dynamics of the early events in the FUS-LC assembly. In order to validate this proposal in FUS-LC, let us define the structural overlap parameter,[41]

$$\chi(t) = \frac{1}{2\left(N^2 - 5N + 6\right)} \sum_{ij}^{M} H\left(\Delta - |r_{ij}^t - r_{ij}^{fib}|\right). \quad (2)$$

In Equation 2, the number of residues in the fibril core is N, $H(x)$ is the step function, and r_{ij}^t and r_{ij}^{fib} are the pair distances for the simulated conformations and r_{ij}^{fib} is the corresponding value in the experimental fibril structure determined for FUS-LC using solid-state NMR[12] or cryo-EM.[13] We set $\Delta = 2$Å, which is a stringent criterion for establishing the similarity between the simulated and experimental fibril-like structures. The total number of pair distances is $M = 2\left(N^2 - 5N + 6\right)$.[14] Division by M ensures that $\chi = 1$ if a conformation coincides with the experimental fibril structure. Thus, as defined in Equation 2, the structural overlap satisfies the bound $0 \le \chi \le 1$.

We calculated χ_1 for core-1 (residues 39–95) and χ_2 for the 112–150 region using millions of conformations in the MCE. In doing so, we only considered the simulation coordinates for the ordered

regions. The χ distribution, $P(\chi)$, FUS-LC shows that the population of core-1 is marginally higher than core-2 (see Figure 2 in reference [14]). Because one might assume that the formation of the more complex core-1 structure (with the S-bend) is less likely than the simpler U-bend in core-2, a natural expectation would be that core-2 should have a higher population than core-1. This is essentially an entropic argument but stability is determined by the free energy, which favors core-1, as we describe below. Our assertion is based on a kinetic argument. It assumes that the fluctuations needed to nucleate complex structures from disordered phases are slow, which follows from the Ostwald's rule of stages (see below).[42]

If χ exceeds a threshold value then the structures resemble the monomer unit from the fibril. The N* states for core-1 and core-2 superimposed on the fibril-like structures illustrate that this is indeed the case (Figure 3), thus validating the proposal that sparsely populated states in the MCE are indicators of the propensity of disordered domains of FUS to form fibrils. The overall topology of the N* states for both core-1 and core-2 are similar to the experimental structures (Figure 3), which is gratifying because the parameters in the energy function were not adjusted to obtain agreement. It is worth noting that the structure of the N* state for a truncated variant of FUS-LC (residues 1–163) is topologically similar to the core-1 fibril structure. Furthermore, the core-1 structure is preserved in FUS-LC-N (residues 1–108), an observation that is in accord with the findings of Tycko and coworkers.[13] Taken together, these findings show that as long as the sequence corresponding to core-1 is present, the fibril with S-bend structure is likely to be more probable under nominal experimental conditions, which we explain using free energy estimates. In sharp contrast to ssNMR studies,[12,13] which show the formation of ordered core-1 (FUS-LC) and core-2 (FUS-LC-C), other NMR
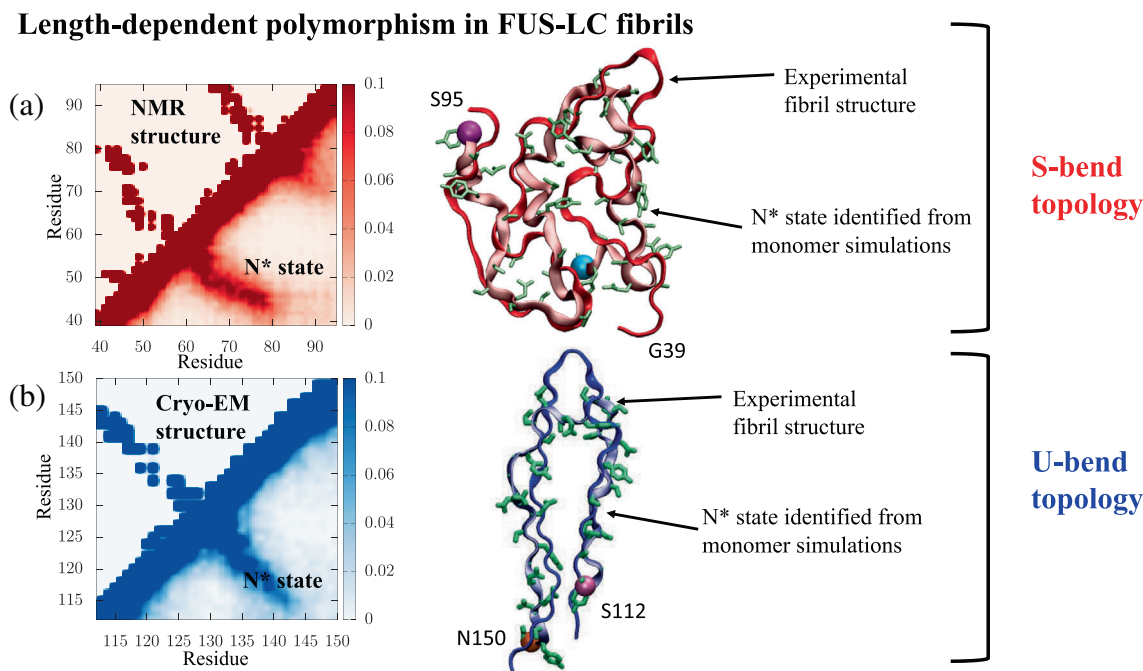
Length-dependent polymorphism in FUS-LC fibrils

FIGURE 3 Contact maps for the N^* ensemble identified from simulations corresponding to the core-1 fibril topology (a) and the core-2 fibril topology (b). The upper triangles in (a) and (b) were computed from the contacts within the monomer units in the experimental fibril structures. For illustration, representative all-atom snapshots from the N^* ensembles are also shown. It is evident that the N^* states bear striking resemblance to the experimental fibril structures.

experiments were used to suggest that the droplets in the construct containing residues 1–163 are mobile exhibiting liquid-like behavior.[24,43,44] The experimental conditions in all the quoted experiments are similar. Our computations predict that core-1 should form in the 1–163 FUS construct.[14] A possible reason for the discrepancy is that the liquid-like droplets have not aged long enough for fibril formation. It is important to resolve this controversy. Early studies[16] suggested that the full length construct, containing the RNA-binding domain, does aggregate. The structure of the aggregated region, which is envisioned to be solid-like, is unknown. The formation of the solid-like component is accelerated by disease causing mutation, G156E.[16]

2.5 | Free energy of core-1 is lower than core-2

The relative free energies of the core-1 and core-2 structures, calculated using the conformations generated in the monomer simulations, rationalizes the absence of core-2 in FUS-LC. First, we enumerated the number of conformations that belong to core-1 and core-2 using χ as the order parameter. From our calculations[14] we find that the free energy difference between core-1 and core-2 using $\Delta G = G_1 - G_2 = -k_B T \times \ln(P_{core-1}/P_{core-2}) = -0.593 \times \ln(8165 \pm 103/666 \pm 29) = -1.47 \pm 0.005$ kcal/mol. The estimates from solubility measurements for ΔG is even less ($\approx 0.35 \pm 0.19$ kcal/mol). It is interesting that $\Delta G \approx 2.5 k_B T$ is sufficient to completely (at least below detectable levels in experiments) eliminate core-2 formation in FUS-LC. From the calculated energy and free

energy difference, the entropy change maybe readily estimated. We find that $T\Delta S \approx -1.08 \pm 0.38$ kcal/mol and the entropy difference ΔS is $S_1 - S_2 = -3.62 \pm 1.27$ cal/mol/K. Thus, the formation of the S-bend is entropically disfavored relative to the U-bend (the former has a more complex topology than the latter). The unfavorable entropy is compensated by the more favorable energy, leading to the prediction that core-1 is preferred in terms of free energy over core-2.

2.6 | Ostwald's rule of stages

A puzzling feature in the fibril morphology of FUS-LC is that the core-2 structure, detectable only in the absence of the sequence giving rise to core-1, (FUS-LC-C residues 110-214), is completely different from core-1. A very small sub-population of excited conformations adopts the U-bend-like structure (Figure 3b) that is found in experiments.[13] More recently, Kato and McKnight have reported that in the construct that does not contain the residues corresponding to core-1 and core-2 (labeled CTC[45]) yet another structure (core-3) emerges. The structure of core-3 is currently unknown. Kinetic experiments by Kato and Mcknight[45] illustrate that in FUS-LC, core-3 forms faster than core-1. The stability of cores was shown to follow exactly the opposite order, based on reverse-phase column chromatography experiments. We predict that core-2, which is less (more) stable than core-1 (core-3) should form after core-3 but ahead of core-1. In other words, the time to form the cores is inversely correlated with stability. This inverse correlation between stability of the

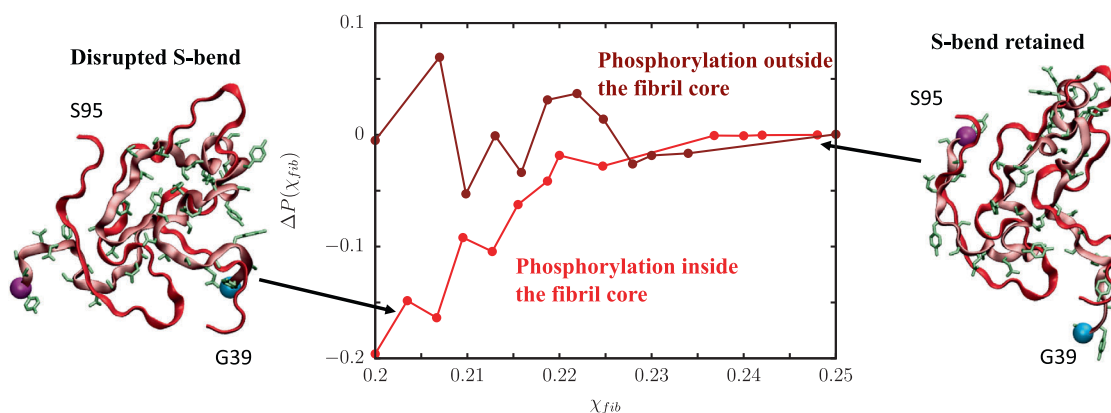## Effect of Phosphorylation is site-specific

FIGURE 4 The effect of phosphorylation on the MCE of FUS-LC is site-specific. When five Serines and a Theronine within the core-1 fibril are phoshorylated, the key interactions within the N* state are destabilized. None of the structures within the MCE retain the complete S-bend morphology. In contrast, phosphorylation of residues outside the fibril core is less disruptive, and most configurations within MCE retain the S-bend. These trends are reflected in $\Delta P(\chi_{fib})$, which denotes the difference in the probability distribution of χ_{fib} (similarity with respect to the experimental core-1 fibril structure) between the wildtype and the phosphorylated variants. For clarity, only the tail of the distribution (which corresponds to the N* states within the MCE) is shown. For the P-FUS (in) variant, there is a dip in $\Delta P(\chi_{fib})$, implying that the structural similarity with respect to the fibril state reduces dramatically.

cores and their rates of formation is consistent with the Ostwald's rule of stages,[42] which we recently showed holds in the formation of distinct polymorphic structures of Aβ_{42} peptides.[46] Ostwald's rule states that if there are two or more polymorphic structures then the least stable structure would form first. In other words, the rate of formation is inversely correlated with the thermodynamic stabilities of the polymorphic structures.[42]

2.7 | Effect of phosphorylation—location matters

Murray et al.,[12] systematically examined the effects of phosphorylation by DNA-dependent protein kinase (DNA-PK) on the stability of the ordered S-bend structure in FUS-LC. Among the 14 phosphorylated residues, six (S42, S44, S61, T68, S84, and S87) are in core-1. By mutating these sites to Ala, mostly two at a time, they[12] discovered that, in all likelihood, core-1 fibril structure is destabilized. Phosphorylation within core-1 sites significantly reduced the propensity of FUS-LC monomers to bind to hydrogel droplets, while phosphorylations outside the fibril core had little impact. If Ser and Thr outside the fibril were mutated to Ala the effect was weaker, which suggests that site-specific phosphorylation plays an important role in the stability of the fibril. A similar conclusion was reached in another experimental study[47] in which the effect of phosphorylation at S61 in a fragment (residues 50–65, referred to as R2) was assessed. Based on a number of measures, the authors concluded that phosphorylation at S61 abolished both intra- and inter-molecular interactions needed for aggregation to occur.

In order to model the effects of phosphorylation, we simulated two FUS-LC variants.[14] In one of them, referred to as P-FUS (in), five Serine residues within core-1 (S42, S54, S61, S84, S87) and T68 were phosphorylated by replacing them with the charged Glu residue,

which is clearly a simplification but is often used in computations.[48,49] The difference between the distributions, $\Delta P(\chi_{fib}) = P\left(\chi_{[\text{P-FUS(in)}]}\right) - P\left(\chi_{[\text{FUS-LC}]}\right)$ shows that the structural similarity of the MCE relative to the fibril state is greatly reduced (Figure 4). Phosphorylation of sites in the fibril core destabilizes some of the key interactions within the N* state. In particular, for the P-FUS (in) variant, the conformation is not structurally close to the core-1 fibril (Figure 4). The features in $\Delta P(\chi_{fib})$ are less prominent for the construct P-FUS (out), in which T11, T19, and T30 and S112, S117, and S131 outside of core-1 were phosphorylated. This suggests that the effect of phosphorylation is less disruptive in P-FUS (out) compared to P-FUS (in). Most of the conformations within the N* ensemble seem to retain the overall S-bend topology, which accords well with experimental findings that if the phosphorylation sites are outside the fibril forming region the effects are small. The conclusion that phosphorylation of residues in the core is destabilizing while the effect is small when residues outside are phosphorylated[12] has also been noted in the simulation of droplets.[15,50]

The reduced aggregation propensity of the phosphorylated variants could be attributed to the enhanced free energy gap between the disordered ground state and the N* state. Our results suggest that phosphorylation effects are site-specific, with disruption occurring if residues within the fibril core are altered either by mutations or phosphorylation. The implication is that specific interactions within and between the FUS-LC polymers stabilize the fibril structure, which was already noted in previous studies.[12,51]

2.8 | FUS aggregates

The lessons one can learn from careful characterization of monomers are limited. At finite concentration of FUS polymers many phases
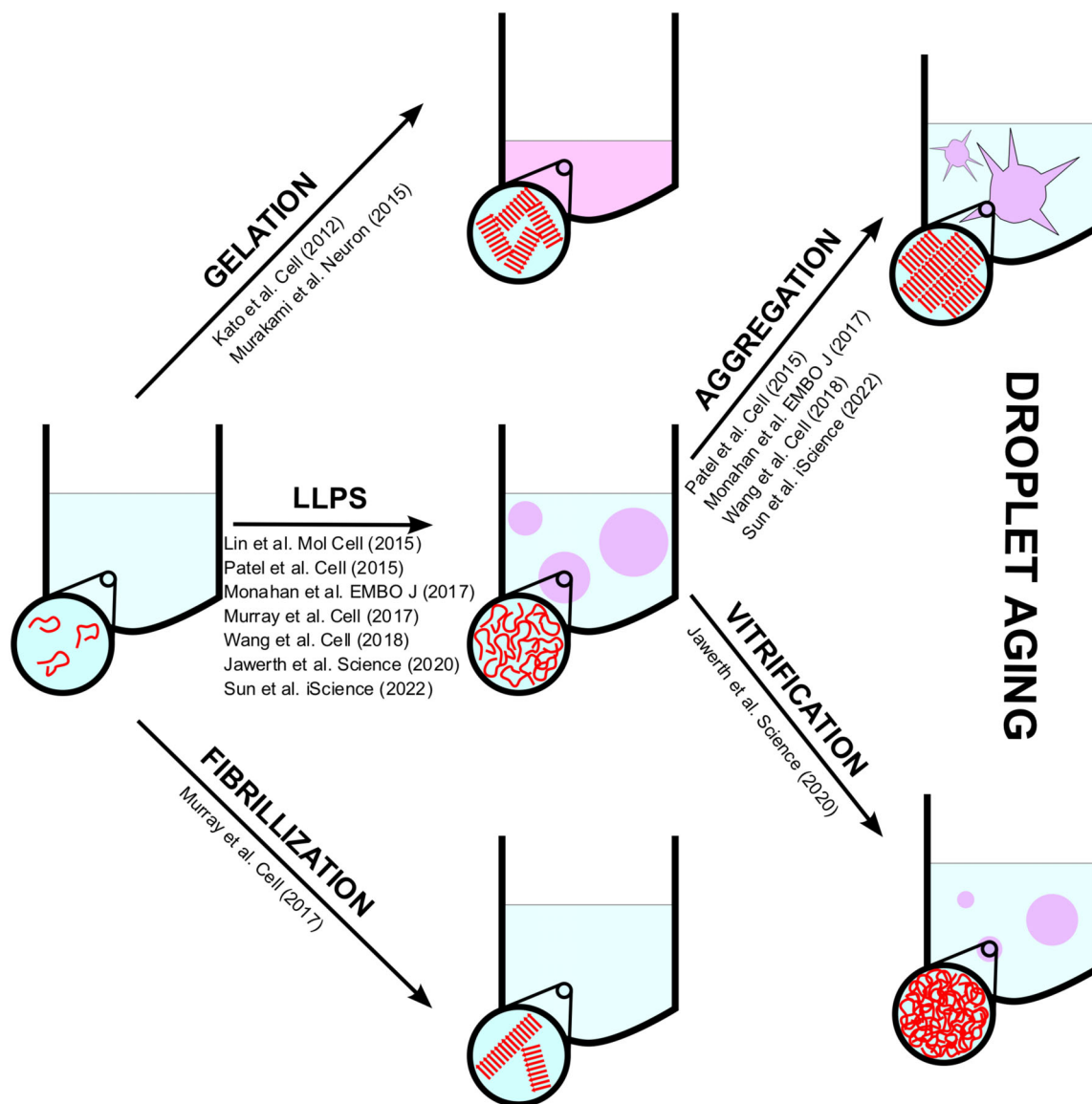
FIGURE 5 A schematic illustrating liquid–solid phase transition in FUS. The liquid droplet state is metastable, and ages with time to form hydrogels or irregularly shaped solid aggregates, which have a characteristic cross-β architecture typical of amyloid fibrils. Recent work also suggests that aging could induce Maxwell glass-like dynamic behavior.[54]

could emerge. For instance, FUS-LC (or FUS) could undergo phase separation into liquid droplets, or form fibrils and gels. The droplet could age to a fibril or a glass, depending on external conditions.[12,52–54] Both fibrils, and glasses are solids (they resist shear). We use the term fibril here to describe the ordered structures found in solid-state NMR experiments[12,13] or more recently by cryo-EM.[55] Computer simulations,[56] using a single-bead per residue coarse grained model,[57] show that as the droplet ages, the morphologies of FUS could change, producing non-equilibrium structures. Some of the scenarios that could arise are schematically shown in Figure 5. At a fixed temperature, various phases of FUS (for that matter any protein or IDP) could emerge as external conditions (for example, salt concentration, crowding agents and pH) are altered. In other words, the phase diagram for FUS-LC or its variants is complicated. A survey of experiments in Table 1 and Figure 5

points to several possibilities. Here, we only discuss phase separation and fibril formation.

Under certain conditions FUS undergoes phase separation at high polymer concentration. In such a phase, a dense condensate (or coarcervate, term first used in 1929 in the same sense as condensate is used today) coexists with a low density dilute phase. Such a generic phase behavior in FUS (and other IDPs) is explained qualitatively by the venerable Flory-Huggins (FH) lattice theory[1–3] that has only the χ_{FH}, a dimensionless parameter, which describes the differential interaction energy (relative to k_BT) between the monomer and solvent with respect to an average of monomer-monomer and solvent-solvent interactions. Although FH is a mean-field theory, it is often remarkably accurate if the degree of polymerization (n) is large ($n \gg 1$) because the mixing entropy associated with the polymer scales as $1/n$. Note that the number of lattice sites, $N = N_1 + nN_2$, where N_1 is

TABLE 1 A list of previous work that have experimentally characterized various phases for the FUS protein. For clarity, the precise experimental conditions, including the inclusion of tags, protein and salt concentration, as well as temperatures, are also summarized.

Experimental references	Sequence	[FUS] (μM)	[Salt] (mM)	Temp. (K)	Observed phase
Kato et al.[52]	Tag (2–214)	2000	200	277	Gel
Lin et al.[58]	SNAP (1–237)	10.75	37.5	298	Liquid droplets
Patel et al.[16]	FUS GFP	10	500	298	Fibrils
	FUS GFP	500	200	277	Gel
Murakami et al.[59]	FUS-LC (2–214)	1000	500	296	Hydrogel
	FUS-LC + RBD	1–5	150	277	Liquid droplets
Murray et al.[12]	FUS-LC (1–214)	30	200	297	Fibril
Monahan et al.[49]	FUS-LC (1–163)	5		298	Liquid droplets
	FUS-LC (1–214)	5	150	298	Liquid droplets
Wang et al.[53]	FUS-LC + RBD	5	150		Liquid droplets
	FUS-LC (1–214)	>10	150		Liquid droplets
Jawerth et al.[54]	FUS-LC + RBD	10	75		Maxwell glass
Berkeley et al.[17]	5%Cy3-FUS (1–163)		150	298	Fibril
Sun et al.[55]	mCeru-FUS-LC (2–214)	100	100	310	Fibril

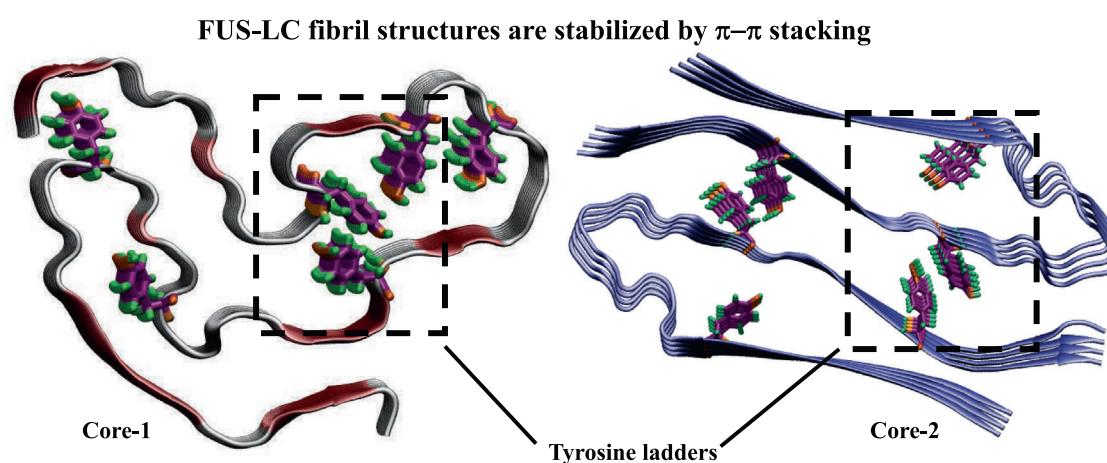### FUS-LC fibril structures are stabilized by π−π stacking

FIGURE 6 Stacking of aromatic Tyrosine side-chains in the core-1 (a) and core-2 fibril (b). The FUS fibril-cores lack the strong hydrophobic interactions found in other fibrils, such as Aβ and α-synuclein. It is likely that the Tyrosines stacked in near-perfect registry (forming Tyrosine ladders) stabilize the fibril cores formed by FUS-LC.

the number of solvent molecules, N_2 is the number of polymer chains, and n is the degree of polymerization, which in FUS-LC is 214. These quantities are easily expressible in terms of number density or molar concentrations. Interestingly, the FH theory describes phase separation in IDPs at least qualitatively (for a lucid review see reference [60]), even though IDPs are not large (in the sense implied in polymer theories). Because IDPs are heteropolymers, which suggests that the sequence, and not merely the overall length, could influence the stability of various structures, as has been shown in experiments involving site-specific phosphorylation.

We focus on two puzzling aspects that pertain to molecular structure of FUS upon aggregation. The survey in Table 1, accompanied by a study of the listed papers, shows that there is a lack of consensus among many different experiments. Some of these are expected because the experimental conditions (concentrations of FUS and salt or pH), and possibly even sequences (FUS-LC and full length FUS that also contains the RNA-binding domain), and the method of preparation are not exactly similar. There are two caveats: (1) The pH used in most of the studies in Table 1 changes over a narrow range (between 7.4 and 7.5). (2) We assume that the ground state should be independent of the preparation method unless FUS aggregates are glass-like that are trapped in one of the many metastable states. For now, we disregard this possibility.

2.9 | Tyrosine ladder

We first describe the salient features of core-1 and core-2, shown in Figure 6. A few observations are worth making: (1) It is self-evident

that the aromatic rings in Tyr are beautifully (seemingly in perfect registry) stacked on top of each other in both core-1 and core-2. There are 8 aromatic ladders in core-1. Because the core-2 protofilament is a dimer, there are 6 Tyr ladders per chain. Naturally, the number of Tyr ladders is exactly equal to the number of Tyr in the two cores. (2) The structures reveal the importance of $\pi - \pi$ interactions in stabilizing the two cores. That $\pi - \pi$ interactions play an important role in materials with aromatic groups is hardly a surprise given the preponderance of such interactions in graphene, organic compounds, and base stacking in nucleic acids.[61–64]

The importance of $\pi - \pi$ interactions has been pointed out in the IDP literature (for an excellent review see reference [65]), although it was established in the context of oligomer formation in Aβ peptides about 20 years ago. Using molecular dynamics simulations, Klimov and Thirumalai[66] showed that the hydrophobic core (KLVFFAE spanning residues 16–22 in Aβ), which adopts an antiparallel fibril structure,[67] is stabilized by $\pi - \pi$ interaction (in this case between two Phe residues). Mutation of Phe to Ser disrupts the stability of Aβ oligomers.[66] A dramatic demonstration of $\pi - \pi$ interactions is found in the self-association of a single amino acid, Phe,[68] which is associated with the autosomal disease, phenyletonuria. It was shown that Phe forms amyloid-like fibrils, almost entirely stabilized by $\pi - \pi$ interactions. In the context of FUS, the role of Tyrosine in mediating multivalent interactions within condensates has been highlighted recently.[53,69] (3) There is only charged residue (D46) in core-1 and none in core-2. It was noted[12] that D46 is either exposed to the solvent or interacts with or S44 and Q52, which are polar. Thus, one can rule out cation-π interactions as the source of stability in core-1 and core-2, although it is suspected to play a role in the self-association of FUS that also includes the RNA-binding domain. (4) Hughes et al.,[70] determined atomic structures of three peptide fragments in core-1 using X-ray crystallography. The peptides crystallized as pairs of kinked β-sheets. The structure of residues 54–61 ([54]SYSSYGQS[61]) containing two Tyr forms exactly two aromatic ladders (Figure 1c in reference [70]). The region, [37]SYSGYS[42] also forms fibril-like structure with Tyr stacks (S37 and Y38 are outside core-1). The kinks in these structures arise possibly because the Tyr residues are separated by only two residues (less than the persistence length of peptides). Consequently, there has to be a kink to accommodate the Tyr ladders. In contrast, in both core-1 and core-2 structures, the Tyr residues are separated by 5 or more residues. (5) Superposition of the X-ray structure for this region with the solid-state NMR structure shows significant deviations. However, the fragments and the core-1 structures both contain the Tyrosine ladder, that must stabilize the ordered structures. In FUS-LC, the favorable inter-peptide interactions in the ordered region (core-1) must compensate for the entropic repulsion between the fuzzy disordered coats that flank the S-bend. The lengths of the two fuzzy coats span residues 1–38 and residues 96–214. More importantly, it is not merely the patterns of sequences or their overall composition, but the location of key residues that determine the propensity of IDPs to self-associate.

3 | DISCUSSION

Although there are several common global features that FUS-LC shares with other monomeric IDPs, it also exhibits certain properties at finite concentrations that are most intriguing. With the current perspective as a background, there are several open questions that we believe are not satisfactorily answered. In what follows we list them in no particular order.

i. From a theoretical perspective, it is puzzling that the formation of non-polymorphic fibrils depends on the length of the FUS-LC. Notwithstanding reports,[24] which assert that aggregates consisting of residues 1–163 are disordered much like the isolated monomers, studies have shown that fibrils indeed form in distinct regions. Why do core-1 and core-2 not coexist? The answer may well be due to entropic repulsion between the fuzzy coats outside the ordered region, but this picture has to be quantified. To make matters interesting, there is a hint that a third fibril structure (core-3, structure as yet undetermined) emerges in a construct with residues beyond 150.[45]

ii. A new serpentine type fold comprising nine β strands, with order in residues between 34 and 124, has been determined using cryo-EM.[55] There is no explanation of the stability of this fourth structure, which contains residues in core-1 and parts of core-2. The serpentine fold is also suspected to be non-polymorphic.

iii. Why are FUS-LC fibrils non-polymorphic whereas Aβ fibrils are not? It is argued that the primary source of stability of Aβ is the inter-peptide hydrophobic interactions. On the other hand, it is suspected that FUS-LC fibrils are stabilized predominantly by hydrophilic interactions. The N-domain of yeast prions, which forms fibrils with cross-β structures that are likely stabilized by hydrophilic interactions, may exhibit polymorphism. What are the general (if there are any) sequence-dependent characteristics that determine the origin of polymorphic structures? We hasten to add that even in FUS-LC, mutation of Tyr (in the Tyr ladder) is likely to disrupt the stability of the fibril.

iv. Aggregation rate for the disease mutation, G156E, in FUS with or without the RNA-binding domain, is enhanced relative to the wild type (WT).[16,17] It is worth noting that the fibril formation rate for the Dutch mutant, E22Q, in Aβ peptides increases compared to the WT. Decrease in the total charge in the Dutch mutant was used to rationalize the enhancement in the rate of fibril formation.[71] A similar reasoning does not explain the enhancement in the aggregation rate for the G156E mutant. In the G156E mutant, which apparently has the same fibril morphology as the WT,[17] there is an increase in charge, which would enhance electrostatic repulsion. Furthermore, association between different molecules could require, minimally, partial desolvation of Glu, another unfavorable process. Thus naively, one would expect that fibril formation rate for the G156E mutant should be less than that for the WT. What are the factors that could quantitatively rationalize this experimental observation?

v. In a most insightful paper[54] it was shown, using passive and active rheology, that IDP/IDR condensates age with time, much like spin glasses,[72] where aging effects have been examined using the trap model. The full length 526-residue FUS, which has no discernible fibril-like order in the interior of the condensate, resembles an amorphous material. Surprisingly, a single relaxation time (τ_c), which increases dramatically with the waiting time (t_w), characterizes the dynamics of the condensates which undergoes no significant structural change over time.[54] By analogy with Maxwell fluids, which describes viscoelastic relaxation using a single time constant, it was argued that protein condensates are like a Maxwell glass. There are two reasons that this result is unexpected. (i) From the perspective of the dynamics in supercooled liquids one expects a spectrum of relaxation times, often expressed in terms of a stretched exponential function.[73,74] (ii) The energy landscape of IDPs and IDR is rugged containing multiple minima, separated by free energy barriers.[75] One would, therefore, expect that transitions between the minima would lead to multiple relaxation times, which would be incompatible with the Maxwell glass picture. In such a scenario, it appears that protein condensates explore only a single rough minimum, but with minima within minima (hierarchical structure).[76] The free energy barriers between the sub-minima increase as the system relaxes to the bottom of the minimum, thus accounting for the increase of τ_c with t_w. A molecular explanation of this finding, which appears to be general[54] in interacting IDPs, would be most interesting.

4 | CONCLUSION

Although we have focussed on the characteristics of FUS-LC and its variants, it is likely that the peculiar features, such as the length-dependent changes in fibril morphology, responses to perturbations (phosphorylation), non-polymorphic nature of the fibrils may be shared with TDP43.[77] Exploring the generic sequence features that translate into the distinct condensate morphologies, which apparently are very sensitive to external conditions, remains a major problem. Finally, what are the common themes in the fibril formation and growth of IDPs, such as Aβ and α-synuclein, which result in polymorphic structures and low-complexity IDP sequences? Answering this question and others raised here should go a long way toward a general theory of condensate formation.

ORCID

D. Thirumalai https://orcid.org/0000-0003-1801-5924
Abhinaw Kumar https://orcid.org/0000-0002-4417-4825
Debayan Chakraborty https://orcid.org/0000-0003-4339-5818
John E. Straub https://orcid.org/0000-0002-2355-3316
Mauro L. Mugnai https://orcid.org/0000-0002-0267-2279

REFERENCES

[1] P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca **1953**.
[2] P. J. Flory, *J. Chem. Phys.* **1941**, *9*, 660.
[3] M. L. Huggins, *J. Phys. Chem.* **1942**, *46*, 151.
[4] J. T. G. Overbreek, M. J. Voorn, *J. Cell. Comp. Physiol.* **1957**, *49*, 7.
[5] H. G. Bungenberg De Jong, H. R. Kruyt, *Proc. R. Acad. Amsterdam* **1929**, *32*, 849.
[6] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, A. A. Hyman, *Science* **2009**, *324*, 1729.
[7] S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285.
[8] S. Alberti, A. A. Hyman, *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 196.
[9] A. A. Hyman, C. A. Weber, F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 39.
[10] D. S. Protter, R. Parker, *Trends Cell Biol.* **2016**, *26*, 668.
[11] A. S. Holehouse, R. V. Pappu, *Biochemistry* **2018**, *57*, 2415.
[12] D. T. Murray, M. Kato, Y. Lin, K. R. Thurber, I. Hung, S. L. McKnight, R. Tycko, *Cell* **2017**, *171*, 615.
[13] M. Lee, U. Ghosh, K. R. Thurber, M. Kato, R. Tycko, *Nat. Commun.* **2020**, *11*, 5735.
[14] A. Kumar, D. Chakraborty, M. L. Mugnai, J. E. Straub, D. Thirumalai, *J. Phys. Chem. Lett.* **2021**, *12*, 9026.
[15] S. Ranganathan, P. Dasmeh, S. Furniss, E. Shakhnovich, *Proc. Natl. Acad. Sci.* **2023**, *120*, e2215828120.
[16] A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, S. Alberti, *Cell* **2015**, *162*, 1066.
[17] R. F. Berkeley, M. Kashefi, G. T. Debelouchina, *Biophys. J.* **2021**, *120*, 1.
[18] A. D. Gitler, J. Shorter, *Prion* **2011**, *5*, 179.
[19] M. Udan, R. H. Baloh, *Prion* **2011**, *5*, 1.
[20] M. D. Michelitsch, J. S. Weissman, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11910.
[21] J. S. Griffith, *Nature* **1967**, *215*, 1043.
[22] M. D. Ter-Avanesyan, A. R. Dagkesamanskaya, V. V. Kushnirov, V. N. Smirnov, *Genetics* **1994**, *137*, 671.
[23] C.-Y. King, R. Diaz-Avalos, *Nature* **2004**, *428*, 319.
[24] K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, *Mol. Cell* **2015**, *60*, 231.
[25] U. Baul, D. Chakraborty, M. L. Mugnai, J. E. Straub, D. Thirumalai, *J. Phys. Chem. B* **2019**, *123*, 3462.
[26] A. G. Kikhney, D. I. Svergun, *FEBS Lett.* **2015**, *589*, 2570.
[27] G. G. Tartaglia, M. Vendruscolo, *Chem. Soc. Rev.* **2008**, *37*, 1395.
[28] R. Linding, J. Schymkowitz, F. Rousseau, F. Diella, L. Serrano, *J. Mol. Biol.* **2004**, *342*, 345.
[29] E. P. O' Brien, B. Brooks, D. Thirumalai, *J. Am. Chem. Soc.* **2012**, *134*, 979.
[30] J. Hong, L. M. Gierasch, *J. Am. Chem. Soc.* **2010**, *132*, 10445.
[31] F. Shewmaker, R. B. Wickner, R. Tycko, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 19754.

[32] R. Tycko, *Neuron* **2015**, *86*, 632.

[33] N. J. Cobb, F. D. Sönnichsen, H. Mchaourab, W. K. Surewicz, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18946.

[34] L. Wang, K. Zhao, H. Y. Yuan, Q. Wang, Z. Guan, J. Tao, X. N. Li, Y. Sun, C. W. Yi, J. Chen, D. Li, D. Z. P. Yin, C. Liu, Y. Liang, *Nat. Struct. Mol. Biol.* **2020**, *27*, 598.

[35] B. Tarus, J. E. Straub, D. Thirumalai, *J. Am. Chem. Soc.* **2006**, *128*, 16159.

[36] M. S. Li, N. T. Co, G. Reddy, C.-K. Hu, J. E. Straub, D. Thirumalai, *Phys. Rev. Lett.* **2010**, *105*, 218101.

[37] D. Chakraborty, J. E. Straub, D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 19926.

[38] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, L. E. Kay, *Science* **2012**, *336*, 362.

[39] S. A. Kotler, V. Tugarinov, T. Schmidt, A. Ceccon, D. S. Libich, R. Ghirlando, C. D. Schwieters, G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 3562.

[40] P. Zhuravlev, G. Reddy, J. E. Straub, D. Thirumalai, *J. Mol. Biol.* **2014**, *426*, 2653.

[41] D. Klimov, M. Betancourt, D. Thirumalai, *Folding Des.* **1998**, *3*, 481.

[42] R. A. Van Stanten, *J. Phys. Chem.* **1984**, *88*, 5768.

[43] N. L. Fawzi, S. H. Parekh, J. Mittal, *Curr. Opin. Struct. Biol.* **2021**, *70*, 78.

[44] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, N. Fawzi, *Nat. Struct. Mol. Biol.* **2019**, *26*, 637.

[45] M. Kato, S. L. McKnight, *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2114412118.

[46] D. Chakraborty, J. E. Straub, D. Thirumalai, *Sci. Adv.* **2023**, *9*, eadd6921.

[47] X. Ding, F. Sun, J. Chen, L. Chen, Y. Tobin-Miyaji, S. Xue, W. Qiang, S.-Z. Luo, *J. Mol. Biol.* **2020**, *432*, 467.

[48] P. M. Maciejewski, F. C. Peterson, P. J. Anderson, C. L. Brooks, *J. Biol. Chem.* **1995**, *270*, 27661.

[49] Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, N. L. Fawzi, *EMBO J.* **2017**, *36*, 2951.

[50] S. Ranganathan, E. Shakhnovich, *Biophys. J.* **2022**, *121*, 2751.

[51] M. Kato, X. Zhou, S. L. McKnight, *RNA* **2022**, *28*, 3.

[52] M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, S. L. Mcknight, *Cell* **2012**, *149*, 753.

[53] J. Wang, J. M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, *Cell* **2018**, *174*, 688.

[54] L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, J. Mahamid, A. A. Hyman, F. Julicher, *Science* **2020**, *370*, 1317.

[55] Y. Sun, S. Zhang, J. Hu, Y. Tao, W. Xia, J. Gu, Y. Li, Q. Cao, D. Li, C. Liu, *iScience* **2022**, *25*, 103701.

[56] A. Garaizar, J. R. Espinosa, J. A. Joseph, G. Krainer, Y. Shen, T. P. Knowles, R. Collepardo-Guevara, *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2119800119.

[57] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, *PLoS Comput. Biol.* **2018**, *14*, e1005941.

[58] Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, *Mol. Cell* **2015**, *60*, 208.

[59] T. Murakami, S. Qamar, J. Q. Lin, G. S. K. Schierle, E. Rees, A. Miyashita, A. R. Costa, R. B. Dodd, F. T. S. Chan, C. H. Michel, D. Kronenberg-Versteeg, Y. Li, S. P. Yang, Y. Wakutani, W. Meadows, R. R. Ferry, L. Dong, G. G. Tartaglia, G. Favrin, W. L. Lin, D. W. Dickson, M. Zhen, D. Ron, G. Schmitt-Ulms, P. E. Fraser, N. A. Shneider, C. Holt, M. Vendruscolo, C. F. Kaminski, P. S. George-Hyslop, *Neuron* **2015**, *88*, 678.

[60] H.-X. Zhou, V. Nguemaha, K. Mazarakos, S. Qin, *Trends Biochem. Sci.* **2018**, *43*, 499.

[61] G. R. Desiraju, A. Gavezzotti, *J. Chem. Soc., Chem. Commun.* **1989**, *10*, 621.

[62] S. K. Burley, G. A. Petsko, *Science* **1985**, *229*, 23.

[63] D. B. Ninkovic, J. P. Blagojevic Filipovic, M. B. Hall, E. N. Brothers, S. D. Zaric, *ACS Cent. Sci.* **2020**, *6*, 420.

[64] J. Sponer, J. E. Sponer, A. Mladek, P. Jurecka, P. Banas, M. Otyepka, *Biopolymers* **2013**, *99*, 978.

[65] E. Gomes, J. Shorter, *J. Biol. Chem.* **2019**, *294*, 7115.

[66] D. K. Klimov, D. Thirumalai, *Structure* **2003**, *11*, 295.

[67] J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, R. Tycko, *Biochemistry* **2000**, *39*, 13748.

[68] L. Adler-Abramovich, L. Vaks, O. Carny, D. Trudler, A. Magno, A. Caflisch, E. Frenkel, E. Gazit, *Nat. Chem. Biol.* **2012**, *8*, 701.

[69] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, *Science* **2020**, *367*, 694.

[70] M. P. Hughes, M. R. Sawaya, D. R. Boyer, L. Goldschmidt, J. A. Rodriguez, D. Cascio, L. Chong, T. Gonen, D. S. Eisenberg, *Science* **2018**, *359*, 698.

[71] F. Massi, D. Klimov, D. Thirumalai, J. E. Straub, *Prot. Sci.* **2002**, *11*, 1639.

[72] J. Bouchaud, *J. Phys. I France* **1992**, *2*, 1705.

[73] L. Berthier, G. Biroli, *Rev. Mod. Phys.* **2011**, *83*, 587.

[74] T. Kirkpatrick, D. Thirumalai, *Rev. Mod. Phys.* **2015**, *87*, 183.

[75] Y. Chebaro, A. J. Ballard, D. Chakraborty, D. J. Wales, *Sci. Rep.* **2015**, *5*, 1.

[76] R. Zwanzig, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2029.

[77] B. D. Fonda, K. M. Jami, N. R. Boulos, D. T. Murray, *J. Am. Chem. Soc.* **2021**, *143*, 6657.