

# Sequence Determines the Switch in the Fibril Forming Regions in the Low-Complexity FUS Protein and Its Variants

Abhinaw Kumar, Debayan Chakraborty, Mauro Lorenzo Mugnai, John E. Straub, and D. Thirumalai\*



Cite This: *J. Phys. Chem. Lett.* 2021, 12, 9026–9032



Read Online

ACCESS |



Metrics & More

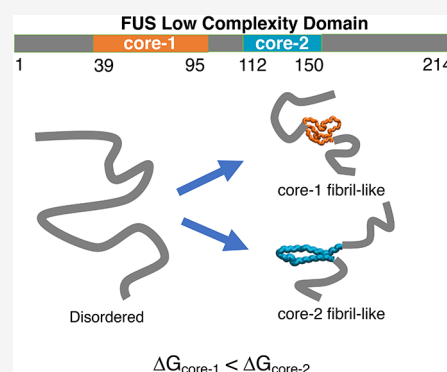


Article Recommendations



Supporting Information

**ABSTRACT:** Residues spanning distinct regions of the low-complexity domain of the RNA-binding protein, Fused in Sarcoma (FUS-LC), form fibril structures with different core morphologies. Solid-state NMR experiments show that the 214-residue FUS-LC forms a fibril with an S-bend (core-1, residues 39–95), while the rest of the protein is disordered. In contrast, the fibrils of the C-terminal variant (FUS-LC-C; residues 111–214) have a U-bend topology (core-2, residues 112–150). Absence of the U-bend in FUS-LC implies that the two fibril cores do not coexist. Computer simulations show that these perplexing findings could be understood in terms of the population of sparsely populated fibril-like excited states in the monomer. The propensity to form core-1 is higher compared to core-2. We predict that core-2 forms only in truncated variants that do not contain the core-1 sequence. At the monomer level, sequence-dependent enthalpic effects determine the relative stabilities of the core-1 and core-2 topologies.



Fused in sarcoma (FUS), an RNA-binding protein, has garnered considerable attention in recent years, not only because it is implicated in neurological disorders and biological functions<sup>1–3</sup> but also because it serves as a prototype for elucidating the principles underlying condensate formation in low-complexity protein sequences.<sup>4–6</sup> The N-terminal region of the 526-residue FUS protein houses the low-complexity (LC) domain, rich in QGSY repeats. The FUS-LC domain (residues 1–214) is intrinsically disordered, and the ground state likely behaves as a random-coil. On the other hand, the C-terminal end of the full-length FUS is ordered, and is involved in protein as well as RNA binding.<sup>7</sup>

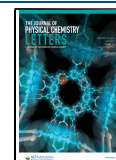
Due to its relevance in liquid–liquid phase separation (LLPS) and fibril formation, the FUS-LC domain has been the subject of several experimental biophysical studies<sup>4,8–21</sup> and simulations.<sup>22–30</sup> The experimental findings that are relevant to our study may be summarized as follows: (i) A combination of fluorescence microscopy and solution-state NMR of a truncated version (residues 1–163) of FUS-LC indicate that it is devoid of persistent structure, both in the monomer as well as in the liquid-droplet state.<sup>8</sup> In contrast to a previous report based on X-ray crystallography and negative stain transmission electron microscopy (TEM)<sup>10</sup> experiments, no signatures of hydrogel formation were immediately apparent in the condensed phase. (ii) Using solid-state NMR, Tycko and co-workers<sup>9</sup> showed that a 50  $\mu$ M solution of FUS-LC does form fibrils, characterized by an S-shaped ordered region involving only residues 39–95. In the S-shaped structure (core-1), the monomers are arranged as parallel  $\beta$ -strands (Figure 1). The rest of the residues are not sufficiently ordered to be resolvable in experiments. They form a fuzzy coat surrounding the fibril

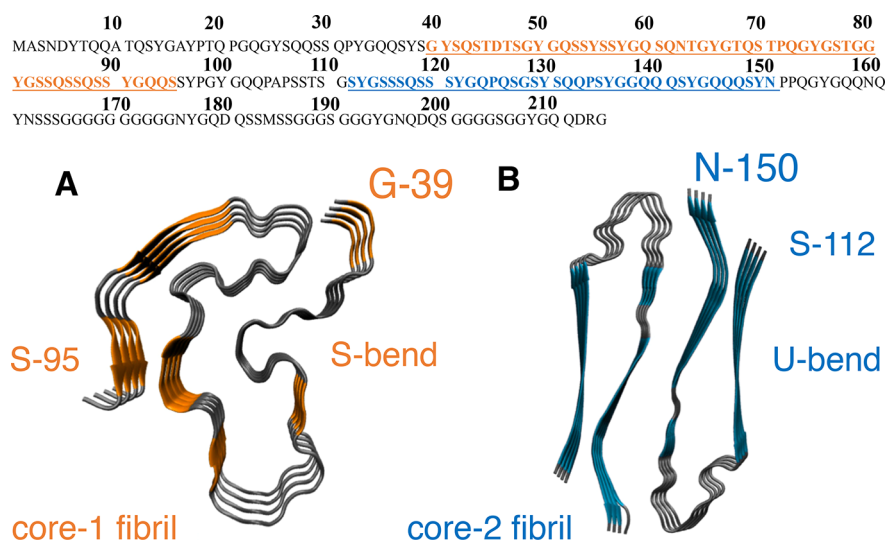
region. Based on the observation that NMR cross-peaks were reproducible in independent measurements, it was concluded that FUS-LC did not show any evidence of fibril polymorphism. (iii) Surprisingly, in a follow-up study, Tycko and co-workers<sup>31</sup> demonstrated that a truncated version of FUS-LC, devoid of the residues forming the core-1 fibril and comprising only the C-terminal region of the protein (FUS-LC-C; containing residues 110–214), also forms a fibril, with an *entirely different morphology* (Figure 1). In this structure, residues 112–150 form a U-shaped fibril (core-2), while the rest of the protein is disordered. To rationalize the findings, it was argued<sup>31</sup> that the formation of core-2 is suppressed in FUS-LC, because in such a structure, the N-terminal and C-terminal disordered segments, constituting the fuzzy coat, would come in close proximity due to the U-bend topology, resulting in a decrease in the effective available volume. In other words, it was conjectured that the lack of core-2 formation for FUS-LC could be attributed to an entropic effect.

Further impetus into this debate, which centers around the probable structures (liquid-droplet, hydrogel, or fibril) of FUS-LC and how transitions between the different forms could occur, comes from a recent work based on magic-angle

Received: July 18, 2021

Published: September 13, 2021





**Figure 1.** Top panel: The FUS-LC sequence, denoted using a one-letter code for the amino acids. Residues 39–95, which form the core-1 fibril, are highlighted in orange, and those within core-2 (112–150) are highlighted in blue. (A) The experimental structure for the core-1 fibril, exhibiting an S-bend morphology (PDB ID: 5W3N).<sup>9</sup> (B) The experimental core-2 fibril structure,<sup>31</sup> exhibiting a U-bend morphology (PDB ID: 6XFM).

spinning NMR spectroscopy and imaging.<sup>11</sup> This study suggests that the liquid-droplet state may represent a metastable state of FUS-LC, and gradual aging could lead to the fibril structure, the likely thermodynamic end-product.

In light of the experimental results summarized above, the following questions arise: (a) Why does core-2 not form in FUS-LC? The nonexistence of core-2 is indeed puzzling, because the solubility of an N-terminal construct derived from the FUS-LC sequence (FUS-LC-N2; residues 2–108), which forms core-1 fibrils, was shown to be similar to FUS-LC-C. Although the rationale<sup>31</sup> based on purely entropic arguments is likely to be correct on the scale of fibrils, there could be subtle enthalpic effects that manifest more readily at the monomer level. Somewhat indirect evidence for such a possibility comes from a previous study, which observed that phosphorylation of six residues within core-1 strongly affected the recruitment of solvated polymers in hydrogel droplets. In contrast, phosphorylation of peripheral residues had no effect.<sup>9</sup> (b) Does a truncated variant of FUS-LC-N (residues 1–163) have the propensity to form the fibril structure? Here, we provide some plausible answers to these questions using computer simulations based on an accurate sequence-specific coarse-grained model for intrinsically disordered proteins (IDPs).<sup>32,33</sup> We probe the conformational landscape of the FUS-LC monomer, and its variants (the truncated and the phosphorylated forms) to ascertain if the free energy excitations within the monomer conformational ensemble (MCE) could provide insights into the conundrum that prevails regarding the feasible structures and polymorphism of FUS-LC.

Our work is based on the premise that the free energy spectrum of the monomer, and the emergent dynamical heterogeneity of the MCE, have robust links to the early events in the assembly cascade, leading to the formation of a liquid-droplet, hydrogel, and ultimately the fibril state.<sup>34,35</sup> The harbingers of self-assembly (referred to as  $N^*$  states) present within the MCE are sparsely populated (usually  $\sim 2\%$  or less) and are challenging to characterize even with advanced NMR techniques.<sup>36</sup> The  $N^*$  states usually have some elements of structural order (akin to fibrils) interspersed with fully disordered segments, which predisposes them to coalesce

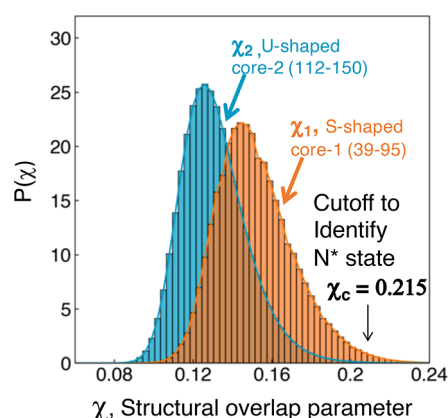
with other assembly-competent conformations to form oligomers of different sizes. In previous studies, we have shown that the  $N^*$  concept accounts for the sequence-specific fibril formation time scales and provides a microscopic basis for fibril polymorphism.<sup>33,37,38</sup>

Within the MCE of FUS-LC, only a very few  $N^*$  conformations having core-2 fibril-like features are found. On the other hand,  $N^*$  states corresponding to core-1 appear with higher probability despite having a more complex morphology than core-2. Our observation suggests that, in addition to entropic effects, sequence-specific enthalpic contributions, apparent at the monomer level, contribute to the free energy difference between the fibril morphologies. The relatively low population of  $N^*$  states within the MCE contributes to the overall destabilization of the core-2 polymorph.

**Fingerprints of fibrillar order within the MCE.** The equilibrium free energy landscape of the FUS-LC monomer is largely featureless like other IDPs and lacks discernible metastable states (Figure S2). To quantify the similarity with respect to the fibril state, we compute the structural overlap parameter,<sup>39</sup>  $\chi$  (see the Supporting Information for definition), between a conformation sampled along the simulation trajectory and a monomer unit within the experimental fibril structure.<sup>9</sup>

The distributions of the overlap parameter,  $P(\chi)$ s, with respect to the core-1 and core-2 fibril structures are shown in Figure 2. In both the cases, the  $\chi$  distributions peak at relatively low values, suggesting that the MCE is conformationally heterogeneous and, on average, bears little or no resemblance to the fibril states. However, a careful structural analysis of the subensembles corresponding to the tails of the  $\chi$  distributions does reveal the existence of a sparse population of fibril-like ( $N^*$ ) conformations.

We define  $N^*$  states as the ensemble of conformations for which  $\chi \geq \chi_c$ , where  $\chi_c$  separates the basin of attraction of fibril-like structures from the disordered states. The relative positions of the  $\chi$  distributions (Figure 2) suggest that the population of core-1 type  $N^*$  conformations is higher than those with core-2 type structures, for any reasonable choice of  $\chi_c$ . This result is unexpected, because from a purely entropic



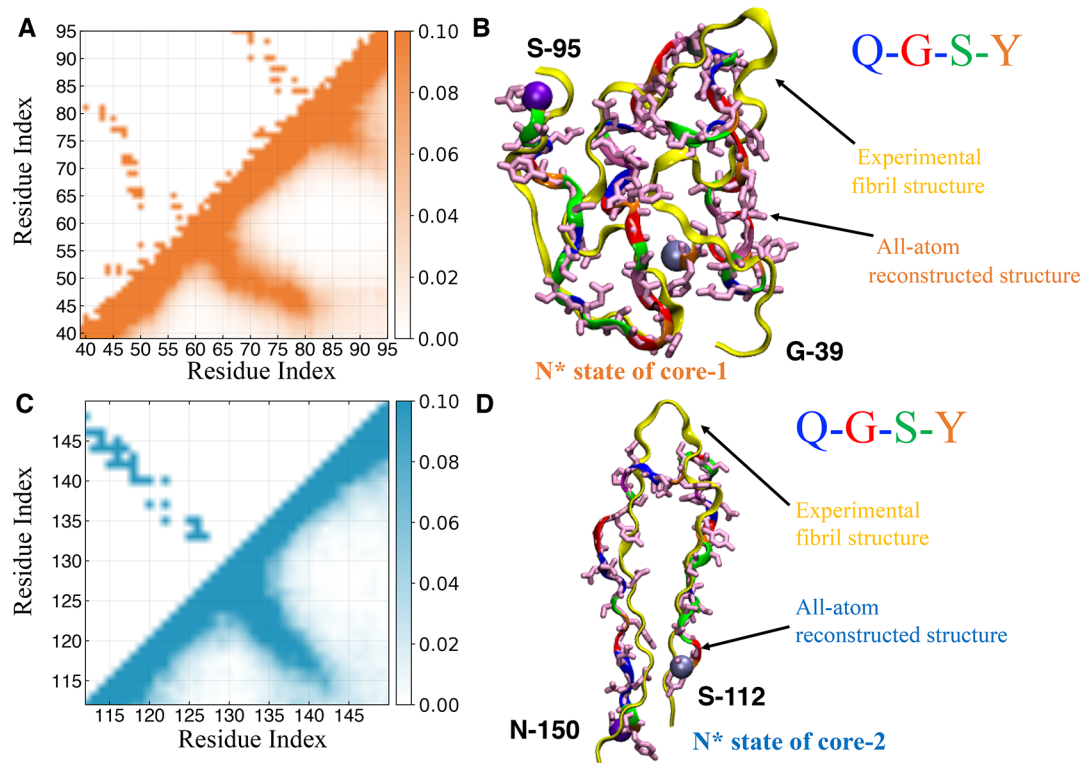
**Figure 2.** Distribution of the structural overlap parameter,  $\chi$ , for the FUS-LC sequence;  $\chi$  (defined in eq S9 in the Supporting Information) was computed with respect to a monomer unit within the experimental fibril structures. For the core-1 fibril, the solid-state NMR structure (PDB ID: 5W3N) for the S-bend morphology<sup>9</sup> is used as the reference. For the core-2 fibril, the cryo-EM structure (PDB ID: 6XFM), the reference structure is the U-bend morphology.

consideration, we would expect that the formation of the complex, longer S-bend shape is less likely than the simpler, shorter U-bend morphology (see the Supporting Information for further discussion of the relative stabilities of core-1 and core-2). But our simulations show the opposite trend, suggesting that sequence-specific effects must play a role.

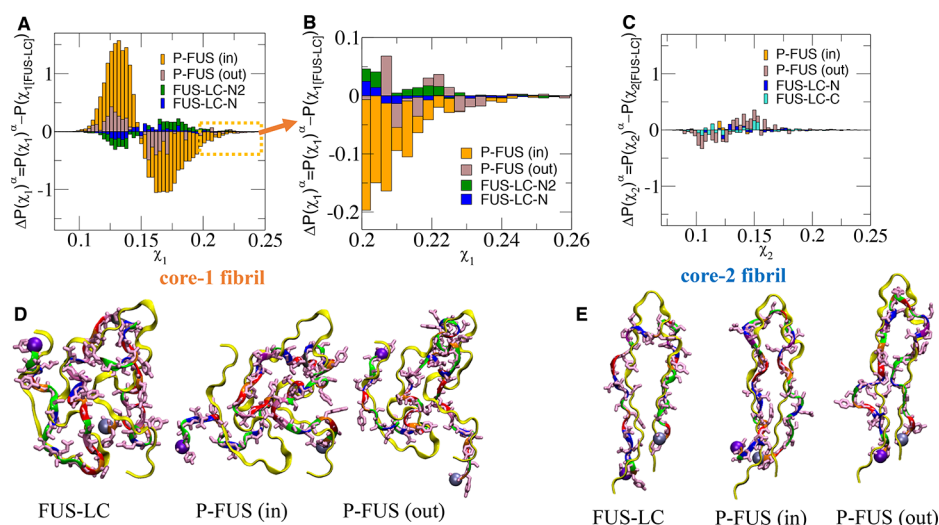
The residue–residue contact maps (Figure 3a,c), reveal that the N\* states (defined for  $\chi \geq \chi_c = 0.215$ ) share a few

structural features in common with the fibril state. In particular, the S-bend and U-bend morphologies of the two cores are captured, showing that the choice of cutoff is reasonable. Nonetheless, the imperfect structural alignment with respect to a monomer unit from the fibril also implies that the N\* states retain some degree of disorder. This is not surprising, because the stability of a monomer in the fibril arises from favorable interactions with its neighbors. Furthermore, N\* states need to be fluid enough to recruit additional monomers and form oligomers of different sizes. We conclude by remarking that the simulations were performed in the absence of potentials biasing the sample toward the N\* state. In this sense, the emergence of fibril-like conformations is a genuine outcome of the model.

**Truncation effects.** In order to quantify the effects of truncation of sequence on the probabilities of core-1 and core-2 formation, we calculated  $\Delta P(\chi_{i[\alpha]}) = P(\chi_{i[\alpha]}) - P(\chi_{i[\text{FUS-LC}]})$  ( $i = 1$  and  $2$  are labels for core-1 and core-2, respectively, and  $\alpha$  is a label for the simulated FUS constructs). It is evident from the difference between the  $\chi$ -distributions of an FUS variant and the WT FUS-LC (blue and green histograms in Figure 4) that there is practically no effect of the two N-terminal truncations (FUS-LC-N2 (residues 1–108); FUS-LC-N (residues 1–163)) on the overall conformational heterogeneity. These findings imply that the population of N\* states (and, by inference, the propensity to form fibrils) is unaffected by sequence truncations. Our observation is consistent with experimental findings,<sup>31</sup> which showed that the FUS-LC-N2 assembles into fibrils similar to those formed by the full-length FUS-LC. Our results also suggest that the



**Figure 3.** (A,C) Contact maps for the N\* ensemble identified from simulations (lower triangle) and a monomer unit from the experimental fibril structure (upper triangle). A structure from the N\* ensemble is shown superimposed on a monomer experimental fibril structure (shown in yellow). For a systematic comparison, all-atom structures were generated from the coarse-grained snapshots (see the Supporting Information for further details). The N\* ensemble corresponding to the core-1 fibril is shown in (B), and that for core-2 is shown in (D). It is clear that the N\* conformations share considerable similarity with the fibril counterparts.



**Figure 4.** Difference in the probability distribution,  $P(\chi)$ , of the structural overlap parameter. (A) Difference plots,  $\Delta P(\chi_{1[\alpha]}) = P(\chi_{1[\alpha]}) - P(\chi_{1[\text{FUS-LC}]})$ , for truncated and phosphomimetic mutants where  $\alpha$  labels the four labeled variants. The four variants are (i) FUS-LC-N (residues 1–163) (blue); (ii) FUS-LC-N2 (residues 1–108) (green); (iii) P-FUS (in) (orange) corresponds to a monomer with phosphorylation inside the core-1 region; (iv) P-FUS (out) (brown) represents the FUS-LC mutant with phosphorylation outside the core-1 region. (B) Zoom-in version of panel (A) showing that the FUS-LC with the mutant with phosphorylation inside the core region (P-FUS (in)) significantly reduces the population of the fibril-like state. (C) Same as (A) except  $\Delta P(\chi_{2[\alpha]}) = P(\chi_{2[\alpha]}) - P(\chi_{2[\text{FUS-LC}]})$  for the four labeled variants corresponding to core-2 are plotted. We used a chi-square test to verify that in panels (A) and (C) the only construct showing a  $P(\chi)$  significantly different from FUS-LC is P-FUS (in) for core-1 ( $p$ -value = 0). This analysis was conducted using the bin-size shown in panels (A) and (C), grouping bins with fewer than five data points, and considering only decorrelated conformations (i.e., separated by about twice the correlation time). (D,E) The left, central, and right panels in panels (D) and (E) compare the overlap of the N\* state obtained using simulation with the experimental fibril structure (core-1 in (C) and core-2 in (D)) for FUS-LC, FUS-LC with a phosphorylation mutant inside the core, and an FUS-LC mutant outside the core. The experimental fibril structures are shown in yellow.

fibril state must be the thermodynamic end-state for the FUS-LC-N as well, a prediction that could be validated using solid-state NMR experiments (Figure S5).

**Effects of phosphorylation.** Using a combination of solution-state NMR and microscopy, it was shown that both phosphorylation and phosphomimetic variations of FUS-LC tend to disrupt aggregation as well as phase separation.<sup>40,41</sup> In particular, 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. To simulate a variant of FUS-LC (denoted as P-FUS (in)), where five serine residues within core-1 (S42, S54, S61, S84, S87) and a threonine (T68) are phosphorylated, we replaced the phosphorylated residues by glutamic acid (see Supporting Information for details). The difference between the distributions,  $\Delta P(\chi_{1[\alpha]})$  (brown and orange histograms in Figure 4), has prominent features. The peak at low values of  $\chi$  and the dip beyond  $\chi_{\text{max}}$  at which  $P(\chi)$  (see Figure 2) is a maximum suggest (orange histograms in Figure 4A,B) that the structural similarity of the MCE relative to the fibril state reduces greatly. Phosphorylation also seems to destabilize some of the key interactions within the N\* state. For the P-FUS (in) variant, not even the conformation structurally closest to the core-1 fibril exhibits the complete S-bend morphology (middle structure in Figure 4D). The features in  $\Delta P(\chi_{1[\alpha]})$  are less prominent (purple histograms in Figure 4A,B) for the construct P-FUS (out), in which T11, T19, and T30 and S112, S117, and S131 outside core-1 were phosphorylated. This suggests that the effect of phosphorylation is less disruptive in this case compared to P-FUS (in). Most of the conformations within the N\* ensemble seem to retain the overall S-bend topology (Figure 4D), which accords

well with experimental findings. 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.

In contrast, the distributions  $\Delta P(\chi_{2[\alpha]})$  for the four variants of FUS-LC displayed in Figure 4C show that neither truncation of the sequence nor phosphorylation affect core-2 formation. Interestingly, representative structures from the N\* ensembles of the phosphorylated variants retain a high degree of structural similarity to the core-2 fibril (Figure 4E), suggesting that sequence perturbations may be better tolerated within the U-bend morphology.

The results in Figure 4C,E show that the formation of core-2 is not precluded by the formation of core-1. The FUS-LC-C construct, which was used to determine the existence of the alternate fibril structure,<sup>31</sup> is practically indistinguishable in terms of  $p_{\text{N}^*}$  (equilibrium population of the N\* state) from the full-length FUS-LC. We infer this result from the small difference observed in the  $\chi$  distributions and probability of forming N\* states in FUS-LC and FUS-LC-C (Figure 4C). Taken together, the simulations using FUS-LC and a number of variants show that, although coexistence of core-1 and core-2 is not ruled out based on thermodynamic considerations, the probability of formation of core-2 structure in FUS-LC is sufficiently decreased, such that it cannot be easily detected. In the absence of the core-1 sequence, when there is no competition, core-2 fibrils can form, and its signature is present even at the monomer level.

We ought to stress that the higher propensity for the formation of the core-1 fibril over core-2 in FUS-LC is a remarkable finding. The entropic penalty associated with the formation of the S-bend motif is certainly larger than for the U-

bend. The 50-residue core-1 structure could be thought of as a juxtaposition of two 35-residue U-bend cores sharing a common edge. Hence, the geometrical requirement to form core-2 is included but is not sufficient to form the core-1 structure. For an athermal homopolymer, lacking any sequence-specificity, the population of core-2 type conformations would be substantially higher than those compatible with the core-1 topology. Therefore, entropic effects alone cannot account for the relative stabilities of the core-1 and core-2 structures at the monomer level. We tested our assertion using the following calculation (see [Supporting Information](#) for further details): the free energy difference between core-1 and core-2-type structures within the MCE is  $\Delta G_{12} = G_1 - G_2 \approx -1.47$  kcal/mol. The difference in energy is  $\Delta U = U_1 - U_2 \approx -2.55$  kcal/mol; the entropy difference between the two states would be  $T\Delta S \approx -1.08$  kcal/mol. In other words, the entropy of core-1 would be less than core-2. Hence, subtle enthalpic effects, which are encoded in the sequence of FUS-LC, determine the stabilities of the aggregation-prone states of the monomer and could be important in understanding the microscopic basis of fibril polymorphism or its absence.

The intermolecular interactions, which are not accounted for in our monomer simulations, could contribute to the conformational selection of the stable fibril morphology. However, the empirical argument introduced by some of us,<sup>33,37</sup> which relates the time scales of fibril formation to the population of aggregation-prone conformations ( $N^*$ ) in the monomer energy landscape (see below), does not depend on the precise details of the intermolecular interactions between protofibril units within a fibril structure. The essence of the  $N^*$  theory is that for aggregation to occur, the signatures of fibril-like states must already be present within the monomer conformational ensemble, regardless of the ultimate fibril morphology.

**Conclusions.** Our simulations answer the questions posed in the introduction: (a) Why does core-2 not form in FUS-LC? Our findings suggest that within the MCE there is a higher population of assembly competent ( $N^*$ ) states corresponding to the core-1 structure, compared to the core-2 structure. The lower free energy gap,  $\Delta G$ , between the ground and the  $N^*$  state, further implies that the initial events in the aggregation cascade are likely to be more conducive to the core-1 formation, rather than core-2. In prior works,<sup>33,37</sup> a correlation between the probability of forming  $N^*$  states and the time required for forming fibrils was demonstrated. Based on this observation, we propose that the higher probability of forming  $N^*$  states exhibiting characteristics of core-1, as opposed to core-2, implies a kinetic preference for the formation core-1 fibrils over core-2 fibrils. (b) Does the FUS-LC-N (residues 1–163) form a fibril? Our simulations show that there is no reason (apart from a kinetic one) why the FUS-LC-N construct should not form a fibril. We propose that this construct should form a fibril having the core-1 morphology, because irrespective of truncation or phosphorylation, the core-1 structure is always thermodynamically favored.

A similar behavior (core-1 formation in one segment that prevents the formation of core-2) is found in the low-complexity TAR DNA-binding protein 43 (TDP43).<sup>42</sup> A common structural characteristic of FUS-LC and TDP43 is the following. The segments that are ordered involve residues that are consecutive to each other. Core-1 in FUS-LC is comprised of residues 39–95, whereas in TDP43, it stretches from 311 to 360. More importantly, all the interactions within core-1 in

these IDPs are only between residues that are consecutive along the sequence. There are no stable contacts between residues in the core-1 region and those that are upstream or downstream along the sequence. This is in contrast to  $\alpha$ -synuclein, for example, in which the fibrils contain contacts between residues that are well-separated (long-range) from each other along the sequence. In other words, the chain folds upon itself in the fibril. The differences may arise because  $\alpha$ -synuclein is not a low-complexity sequence.

We conclude with a few testable predictions. (i) Our findings for FUS-LC and previous theory<sup>37</sup> allow us to qualitatively predict the kinetics of fibril formation. The first passage times for fibril formation starting from the monomer are related to the population of  $N^*$  states as,<sup>37,43</sup>  $\tau_{\text{fib}} \propto \exp(-Cp_{N^*})$ , where the population,  $p_{N^*}$ , of  $N^*$  states is expressed as a percentage, and the value of  $C \approx 1$ . The empirical relationship given does not account for the multiple pathways that may be involved in the aggregation cascade. Nonetheless, it has been exploited in previous studies to predict the experimental trends in fibril formation times nearly quantitatively.<sup>33,38</sup> Based on the relative populations of the  $N^*$  states within the MCE of FUS-LC, we surmise that the core-1 fibril would form faster than the core-2 fibril. A higher value of  $p_{N^*}$  for core-1 also implies a smaller free energy gap,  $\Delta G$ , between the ground and the assembly competent state. Our prediction could be (at least qualitatively) tested by obtaining the rates of fibril formation in various FUS-LC constructs.

(ii) At a first glance, our observations leave us with two complementary explanations for the dominance of core-1 over core-2 in FUS-LC fibrils. The argument based solely on the entropy of the disordered tails, which would hold provided stable oligomers (perhaps even a dimer) first form, does explain the absence of the core-2 structure in the FUS-LC fibril. Our study suggests that the enhanced stability of the core-1 fibril manifests itself already at the monomer level, and it is predominantly a consequence of sequence-specific energetic effects. Thus, on small sizes (monomers and oligomers), stability is determined by enthalpy, whereas on the length scale of the fibril, entropy is likely to be more dominant. In order to test the relative importance of entropy and enthalpy, it would be crucial to investigate whether an FUS-LC-N (residues 1–163) construct forms a fibril and, if so, determine its structure. We predict that for FUS-LC-N, core-1 fibrils will form exclusively and that there should be no difference relative to the full-length FUS-LC construct. The destabilization of core-2 fibrils due to the disordered tails is likely to be weaker in the FUS-LC-N construct, because in this construct, the C-terminal tail is shortened by about 50 residues. From the perspective of core-2, FUS-LC-N should be similar to FUS-LC-C: only one unstructured tail (N-terminal for FUS-LC-N, C-terminal for FUS-LC-C) stems from the core of the fibril.

(iii) It would be interesting to ascertain if in a solution of FUS-LC seeded with FUS-LC-C fibrils, core-2 structures could emerge under appropriate conditions. We anticipate that such a setup would significantly reduce the nucleation barrier for core-2 formation. However, if the disordered tails completely prevent the formation of a core-2 fibril for FUS-LC, we would expect seeding to have no impact on fibril growth. This prediction is also testable.

## ■ ASSOCIATED CONTENT

## ■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.1c02310>.

Details of the model and simulations as well as data analyses (PDF)

## ■ AUTHOR INFORMATION

## Corresponding Author

D. Thirumalai – Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States;

orcid.org/0000-0003-1801-5924;

Email: dave.thirumalai@gmail.com

## Authors

Abhinav Kumar – Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States;

orcid.org/0000-0002-4417-4825

Debayan Chakraborty – Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States;

orcid.org/0000-0003-4339-5818

Mauro Lorenzo Mugnai – Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States;

orcid.org/0000-0002-0267-2279

John E. Straub – Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; orcid.org/0000-0002-2355-3316

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpclett.1c02310>

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge the Texas Advanced Computing Center (TACC) for providing the necessary computing resources. This work was supported by grants from the National Institutes of Health (GM-107703) and National Science Foundation (CHE 19-00033) as well as a grant from the Welch Foundation (F-0019) administered through the Collier-Welch Regents Chair.

## ■ REFERENCES

- (1) Vance, C.; Rogelj, B.; Hortobágyi, T.; De Vos, K. J.; Nishimura, A. L.; Sreedharan, J.; Hu, X.; Smith, B.; Ruddy, D.; Wright, P.; et al. Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science* **2009**, *323*, 1208–1211.
- (2) Kwiatkowski, T. J.; Bosco, D. A.; LeClerc, A. L.; Tamrazian, E.; Vanderburg, C. R.; Russ, C.; Davis, A.; Gilchrist, J.; Kasarskis, E. J.; Munsat, T.; et al. Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science* **2009**, *323*, 1205–1208.
- (3) Matsumoto, T.; Matsukawa, K.; Watanabe, N.; Kishino, Y.; Kunugi, H.; Ihara, R.; Wakabayashi, T.; Hashimoto, T.; Iwatsubo, T. Self-assembly of FUS through its low-complexity domain contributes to neurodegeneration. *Hum. Mol. Genet.* **2018**, *27*, 1353–1365.
- (4) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **2019**, *26*, 637–648.
- (5) Murray, D. T.; Tycko, R. Sidechain hydrogen bonding interactions within amyloid-like fibrils formed by the low-complexity domain of FUS: Evidence from solid state nuclear magnetic resonance spectroscopy. *Biochemistry* **2020**, *59*, 364–378.
- (6) Krainer, G.; Welsh, T. J.; Joseph, J. A.; Espinosa, J. R.; Wittmann, S.; de Csillery, E.; Sridhar, A.; Toprakcioglu, Z.; Gudiskyte, G.; Czekalska, M. A.; et al. Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nat. Commun.* **2021**, *12*, 1085.
- (7) Wang, X.; Schwartz, J. C.; Cech, T. R. Nucleic acid-binding specificity of human FUS protein. *Nucleic Acids Res.* **2015**, *43*, 7535–7543.
- (8) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **2015**, *60*, 231–241.
- (9) Murray, D. T.; Kato, M.; Lin, Y.; Thurber, K. R.; Hung, I.; McKnight, S. L.; Tycko, R. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **2017**, *171*, 615–627.
- (10) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.; Longgood, J.; Pei, J.; et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **2012**, *149*, 753–767.
- (11) Berkeley, R. F.; Kashefi, M.; Debelouchina, G. T. Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophys. J.* **2021**, *120*, 1276.
- (12) Emmanouilidis, L.; Esteban-Hofer, L.; Damberger, F. F.; de Vries, T.; Nguyen, C. K.; Ibáñez, L. F.; Mergenthal, S.; Klotzsch, E.; Yulikov, M.; Jeschke, G.; et al. NMR and EPR reveal a compaction of the RNA-binding protein FUS upon droplet formation. *Nat. Chem. Biol.* **2021**, *17*, 608–614.
- (13) Sahadevan, S.; Hembach, K. M.; Tantardini, E.; Pérez-Berlanga, M.; Hruska-Plochan, M.; Megat, S.; Weber, J.; Schwarz, P.; Dupuis, L.; Robinson, M. D.; et al. Synaptic FUS accumulation triggers early misregulation of synaptic RNAs in a mouse model of ALS. *Nat. Commun.* **2021**, *12*, 3027.
- (14) Jutzi, D.; Campagne, S.; Schmidt, R.; Reber, S.; Mechttersheimer, J.; Gypas, F.; Schweingruber, C.; Colombo, M.; Von Schroetter, C.; Loughlin, F. E.; et al. Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nat. Commun.* **2020**, *11*, 6341.
- (15) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoykov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **2015**, *162*, 1066–1077.
- (16) Bogaert, E.; Boeynaems, S.; Kato, M.; Guo, L.; Caulfield, T. R.; Steyaert, J.; Scheveneels, W.; Wilmans, N.; Haec, W.; Hersmus, N.; et al. Molecular dissection of FUS points at synergistic effect of low-complexity domains in toxicity. *Cell Rep.* **2018**, *24*, 529–537.
- (17) Hofweber, M.; Hutten, S.; Bourgeois, B.; Spreitzer, E.; Niedner-Boblentz, A.; Schifferer, M.; Ruepp, M.-D.; Simons, M.; Niessing, D.; Madl, T.; et al. Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation. *Cell* **2018**, *173*, 706–719.
- (18) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovskiy, A.; Drechsel, D.; et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018**, *174*, 688–699.
- (19) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699.
- (20) Luo, F.; Gui, X.; Zhou, H.; Gu, J.; Li, Y.; Liu, X.; Zhao, M.; Li, D.; Li, X.; Liu, C. Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nat. Struct. Mol. Biol.* **2018**, *25*, 341–346.
- (21) Hughes, M. P.; Sawaya, M. R.; Boyer, D. R.; Goldschmidt, L.; Rodriguez, J. A.; Cascio, D.; Chong, L.; Gonen, T.; Eisenberg, D. S. Atomic structures of low-complexity protein segments reveal kinked  $\beta$  sheets that assemble networks. *Science* **2018**, *359*, 698–701.
- (22) Benayad, Z.; von Bulow, S.; Stelzl, L. S.; Hummer, G. Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput.* **2021**, *17*, 525–537.

- (23) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **2018**, *14*, No. e1005941.
- (24) Chatterjee, S.; Salimi, A.; Lee, J. Y. Insights into amyotrophic lateral sclerosis linked Pro525Arg mutation in the fused in sarcoma protein through in silico analysis and molecular dynamics simulation. *J. Biomol. Struct. Dyn.* **2020**, 1–14.
- (25) Chou, H.-Y.; Aksimentiev, A. Single-protein collapse determines phase equilibria of a biological condensate. *J. Phys. Chem. Lett.* **2020**, *11*, 4923–4929.
- (26) Statt, A.; Casademunt, H.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *J. Chem. Phys.* **2020**, *152*, 075101.
- (27) Das, S.; Eisen, A.; Lin, Y.-H.; Chan, H. S. A lattice model of charge-pattern-dependent polyampholyte phase separation. *J. Phys. Chem. B* **2018**, *122*, 5418–5431.
- (28) Kasahara, K.; Terazawa, H.; Takahashi, T.; Higo, J. Studies on molecular dynamics of intrinsically disordered proteins and their fuzzy complexes: a mini-review. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 712–720.
- (29) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **2020**, *71*, 53–75.
- (30) Zeng, X.; Holehouse, A. S.; Chilkoti, A.; Mittag, T.; Pappu, R. V. Connecting coil-to-globule transitions to full phase diagrams for intrinsically disordered proteins. *Biophys. J.* **2020**, *119*, 402–418.
- (31) 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. *Nat. Commun.* **2020**, *11*, 5735.
- (32) Baul, U.; Chakraborty, D.; Mugnai, M. L.; Straub, J. E.; Thirumalai, D. Sequence effects on size, shape, and structural heterogeneity in Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2019**, *123*, 3462–3474.
- (33) Chakraborty, D.; Straub, J. E.; Thirumalai, D. Differences in the free energies between the excited states of A $\beta$ 40 and A $\beta$ 42 monomers encode their aggregation propensities. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 19926–19937.
- (34) Tarus, B.; Straub, J. E.; Thirumalai, D. Dynamics of Asp23-Lys28 salt-bridge formation in A $\beta$ 10–35 monomers. *J. Am. Chem. Soc.* **2006**, *128*, 16159–16168.
- (35) Straub, J. E.; Thirumalai, D. Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu. Rev. Phys. Chem.* **2011**, *62*, 437–463.
- (36) Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundstrom, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L. E. Structure of an Intermediate State in Protein Folding and Aggregation. *Science* **2012**, *336*, 362–366.
- (37) Li, M. S.; Co, N. T.; Reddy, G.; Hu, C.-K.; Straub, J. E.; Thirumalai, D. Factors governing fibrillogenesis of polypeptide chains revealed by lattice models. *Phys. Rev. Lett.* **2010**, *105*, 218101.
- (38) Zhuravlev, P.; Reddy, G.; Straub, J. E.; Thirumalai, D. Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins. *J. Mol. Biol.* **2014**, *426*, 2653–2666.
- (39) Klimov, D.; Betancourt, M.; Thirumalai, D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of  $\alpha$  helices: effect on stability, cooperativity and kinetics. *Folding Des.* **1998**, *3*, 481–496.
- (40) Monahan, Z.; Ryan, V. H.; Janke, A. M.; Burke, K. A.; Rhoads, S. N.; Zerze, G. H.; O'Meally, R.; Dignon, G. L.; Conicella, A. E.; Zheng, W.; et al. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* **2017**, *36*, 2951–2967.
- (41) Ding, X.; Sun, F.; Chen, J.; Chen, L.; Tobin-Miyaji, Y.; Xue, S.; Qiang, W.; Luo, S.-Z. Amyloid-forming segment induces aggregation of FUS-LC domain from phase separation modulated by site-specific phosphorylation. *J. Mol. Biol.* **2020**, *432*, 467–483.
- (42) 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. *J. Am. Chem. Soc.* **2021**, *143*, 6657–6668.
- (43) Nam, H. B.; Kouza, M.; Zung, H.; Li, M. S. Relationship between population of fibril-prone conformation in the monomeric state and oligomer formation times of peptides: Insights from all-atom simulations. *J. Chem. Phys.* **2010**, *132*, 165104.