The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores

Masato Kato\textsuperscript{a,b,1} and Steven L. McKnight\textsuperscript{a,1}

\textsuperscript{a}Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390; and \textsuperscript{b}Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, Chiba 263-8555, Japan

Contributed by Steven L. McKnight, September 2, 2021 (sent for review August 5, 2021; reviewed by Arthur L. Horwich and Peter St George-Hyslop)

The low-complexity (LC) domain of the fused in sarcoma (FUS) RNA binding protein self-associates in a manner causing phase separation from an aqueous environment. Incubation of the FUS LC domain under physiologically normal conditions of salt and pH leads to rapid formation of liquid-like droplets that mature into a gel-like state. Both examples of phase separation have enabled reductionist biochemical assays allowing discovery of an N-terminal region of 57 residues that assembles into a labile, cross-β structure. Here we provide evidence of a nonoverlapping, C-terminal region of the FUS LC domain that also forms specific cross-β interactions. We propose that biologic function of the FUS LC domain may operate via the mutually exclusive use of these N- and C-terminal cross-β cores. Neurodegenerative disease–causing mutations in the FUS LC domain are shown to imbalance the two cross-β cores, offering an unanticipated concept of LC domain function and dysfunction.

FUS low-complexity sequence | cross-beta polymer | ALS mutation | neurodegenerative disease

Towards 20% of the proteomes of eukaryotic cells consist of polypeptide sequences that are of low complexity (LC) (1, 2). These atypical polypeptide domains are composed of only a subset of the 20 amino acids normally required for proteins to assume their distinct, three-dimensional shapes. As studied in the monomeric state in biochemical preparations outside of cells, LC sequences remain unfolded. As such, they have been described as intrinsically disordered regions (IDRs) of the proteome.

Unbiased proteomic studies employing thermal perturbation as a probe for structural order within cells have shown that a surprisingly large fraction of protein domains thought to be intrinsically disordered may instead function in states of labile, structural order (3). Numerous reports published over the past several years have also given evidence that LC domains can undergo phase separation upon purification and incubation in test tube assays (4, 5). Purified LC domains can self-associate in a manner causing them to partition out of aqueous solution by forming liquid-like droplets. Following incubation, these droplets solidify into a gel-like state. If the interactions driving phase separation of LC domains properly reflect their biological function, this line of research may help reveal how these unusual protein domains actually work in living cells.

Studies of the fused in sarcoma (FUS) RNA binding protein offered an early example of the unusual behavior of phase separation by an LC domain (6, 7). Incubation of the purified FUS LC domain initially triggers the formation of an opalescent suspension composed of liquid-like droplets. Sustained incubation of the same preparation leads to a more stable, gel-like state. Electron microscopy and X-ray diffraction analysis revealed FUS hydrogels to be composed of uniform, amyloid-like polymers (7). Unlike pathogenic, prion-like amyloids, FUS polymers are labile to disassembly upon dilution. A molecular structure of FUS polymers has been resolved by the use of solid-state NMR (ssNMR) spectroscopy (8). Among the 214 residues constituting the FUS LC domain, polymers were found to form via the organization of in-register, cross-β interactions localized between residues 39 and 95.

Structural studies of FUS polymers revealed two differences from pathogenic amyloids, such as α-synuclein or polymers formed from the Aβ fragment long understood to form hyperstable aggregates in Alzheimer’s disease patients. First, FUS polymers reproducibly form the same, monomorphic structure via an N-terminally localized cross-β core. By contrast, pathogenic fibrils formed from α-synuclein or Aβ can adopt any of a number of different, inordinately stable structures (9, 10). Second, the subunit interface holding α-synuclein or Aβ polymers together are replete with hydrophobic amino acids believed to contribute to extreme polymer stability. The molecular structure of FUS polymers revealed but a single hydrophobic amino acid, proline residue 72, within the 57 residues constituting the subunit interface (8). The paucity of hydrophobic residues at the subunit interface of FUS polymers may, at least in part, explain polymer liability.

Studies of the LC domain of the hnRNPA2 protein have yielded similar findings. The LC domain of hnRNPA2 becomes phase separated into liquid-like droplets that, with time, also mature into a gel-like state (11, 12). Cross-β interactions formed by a region 40 to 50 residues in length also define

**Significance**

Single amino acid changes causative of neurologic disease often map to the cross-β forming regions of low-complexity (LC) domains. All such mutations studied to date lead to enhanced avidity of cross-β interactions. The LC domain of the fused in sarcoma (FUS) RNA binding protein contains three different regions that are capable of forming labile cross-β interactions. Here we describe the perplexing effect of amyotrophic lateral sclerosis (ALS)-causing mutations localized to the LC domain of FUS to substantially weaken its ability to form one of its three cross-β interactions. An understanding of how these mutations abet uncontrolled polymerization of the FUS LC domain may represent an important clue as to how LC domains achieve their proper biological function.

Author contributions: M.K. and S.L.M. designed research; M.K. performed research; M.K. analyzed data; and M.K. and S.L.M. wrote the paper.

Reviewers: A.L.H., Yale University School of Medicine; and P.S.G.-H., University of Cambridge.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

\textsuperscript{1}To whom correspondence may be addressed. Email: Steven.McKnight@UTSouthwestern.edu or masato.kato@utsouthwestern.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2114412118/-/DCSupplemental.

Published October 15, 2021.
hnRNPA2 LC domain polymers (11, 13, 14). Human genetic studies of patients suffering from various forms of neurological disease have identified mutations in the genes encoding hnRNPA1, hnRNPA2, and hnRNPDL (15, 16). These recurrent, disease-causing mutations commonly alter a conserved aspartic acid residue located within the cross-β core known to hold hnRNPA2 LC domain polymers together (11, 13, 17).

A simplistic interpretation of these observations offers that evolution has favored the presence of an aspartic acid residue within the cross-β cores formed from hnRNP LC domains as a means of tuning the balance of polymer stability/lability. Proximal disposition of this aspartic acid residue, as dictated by the in-register organization of the polymer interface, has been hypothesized to impart instability resulting from repulsive charge:charge interactions (13). Mutations removing these destabilizing interactions, especially when replacing the conserved aspartic acid residue with valine, may enhance polymer stability and lead to disease pathophysiology.

The relatively straightforward understanding of disease-causing mutations in the LC domains of three different hnRNP proteins does not translate to human genetic studies of FUS. Two of the most prominent amyotrophic lateral sclerosis (ALS)-causing mutations within the FUS LC domain include a missense mutation, changing glycine residue 156 to glutamic acid, and the deletion of glycine residues 174 and 175 (18–20). Neither of these mutations is anywhere close to the cross-β core of FUS that is located between residues 39 to 95.

The inability of our current understanding of the FUS cross-β core to explain substantive human genetic studies prompted reexamination of the pathway by which soluble monomers are captured by hydrogel samples composed of the intact LC domain of FUS. As will be described, these studies reveal surprising and unanticipated observations. The baroque pathway of subunit recruitment into existing FUS polymers offers an unconventional perspective as to how disease-causing mutations in the FUS LC domain manifest their pathophysiology at a molecular level. These studies further offer a means of understanding how the FUS LC domain is prevented from runaway polymerization via either of its cross-β cores.

Results

Identification of a Region of the FUS LC Domain Critical for Hydrogel Binding. The sequence of the FUS LC domain is quasirepetitive, containing 27 repeats of the tripeptide sequence G/S-Y-G/S (Fig. 1B). Early mutational studies gave evidence that the tyrosine residues of these repeats are comparatively conserved. A C-terminal region of the FUS LC domain required for hydrogel binding was identified through mutagenic studies (8). The yellow region of the schematic diagram of the FUS LC domain corresponds to the location of the NTC determined by ssNMR (8). Native protein (wild type [WT]) and individual point mutants were linked to GFP, expressed in bacteria, purified, and tested in binding assays using hydrogel droplets formed from mCherry linked to the intact LC domain of FUS. Hydrogel binding activity was reduced most significantly by the Y161S and Y177S mutants. (B) Amino acid sequence of the FUS LC domain with tyrosine residues shown in red. Residue numbers for the individual tyrosines are shown at Left. (C) Quantitation of hydrogel binding assays shown in A. Intensities are an average of three measurements.
important for the binding of FUS to hydrogel preparations formed from the LC domain of FUS itself, as well as its recruitment to RNA granules in living cells (7). In these experiments, hydrogels were prepared from a fusion protein linking mCherry to the native LC domain of FUS, and test proteins were GFP fusions to either the native LC domain of FUS or mutated variants thereof. We have repeated these same experiments using 27 variants of the FUS LC domain bearing single tyrosine-to-serine mutations. Little or small effect on hydrogel binding was observed for 24 of the 27 variants. Three variants, wherein tyrosine residues 155, 161, or 177 were individually changed to serine, bound to mCherry:FUS hydrogels considerably less well than the native protein (Fig. 1A and C). It is notable that these mutation-sensitive tyrosine residues localize far distal to the cross-β core housed between residues 39 and 95.

Three sets of deletion mutants were prepared to further investigate regions of the FUS LC domain important for hydrogel binding. One set of deletions systematically truncated the LC domain from its C terminus. As shown in Fig. 2A, removal of 17 residues yielded a protein (ΔC1) that retained strong hydrogel binding activity. The next variant missing 34 residues (ΔC2) retained residual binding, yet all remaining deletion mutants were unable to bind hydrogels formed by the intact LC domain of FUS (Fig. 2A). From these studies we define a boundary around residue 190 that marks the C-terminal location of a region required for monomeric GFP test proteins to bind mCherry hydrogels formed by the intact FUS LC domain.

Analysis of N-terminal truncations of the FUS LC domain yielded the surprising observation that upwards of half of the region specifying the cross-β core, spanning residues 39 to 95, could be deleted without significantly affecting hydrogel binding (Fig. 2B). The ΔN4 variant displayed strong hydrogel binding despite lacking 22 residues of the cross-β core. It was likewise surprising that four deletion mutants missing even greater amounts of the FUS LC domain, two of which eliminated the entire cross-β core (ΔN7 and ΔN8), retained attenuated but readily detectible hydrogel binding activity. In combination, these systematic mutagenesis experiments give evidence that a region of the FUS LC domain located far distal to the cross-β core is
required for an unstructured test protein to bind hydrogel preparations of the FUS LC domain.

In a third set of deletion mutations, we initiated truncation downstream of the N-terminal cross-β core starting at residue 111. Deletions were extended from this point and enlarged at 10 amino acid increments, internally removing as few as 9 residues and extending up to as many as 59 residues. Each variant was expressed as a GFP fusion, purified and tested for binding to mCherry hydrogel samples formed from the intact FUS LC domain. The internal deletion missing 29 amino acids (ΔI3), and all variants missing fewer residues of the FUS LC domain, exhibited hydrogel binding. By contrast, the variant missing 39 internal residues (ΔI4) exhibited reduced hydrogel binding, and those missing 49 or 59 residues (ΔI5 or ΔI6) revealed no binding (Fig. 2C). In combination with studies of C-terminal truncations (Fig. 2A), analysis of these internal deletion mutants define a region of roughly 40 amino acids, located between residues 150 and 190 of the FUS LC domain, required for monomeric test protein to bind mCherry:FUS hydrogels. Interpretation of these experiments assumes that the mCherry:FUS hydrogel droplets are composed of polymers assembled via the same cross-β core whose structure was resolved at the atomic level by solid-state NMR spectroscopy (8). Such studies were performed on the isolated LC domain of FUS not appended to mCherry or GFP. It was formally possible that polymers measured from the mCherry:FUS fusion protein used in this study might be different from polymers formed from the isolated FUS LC domain. To test this possibility we employed in situ chemistry to ligate unlabeled GFP to a uniformly, 13C,15N-labeled form of the FUS LC domain. The latter protein was allowed to polymerize and be evaluated by solid-state NMR spectroscopy. As shown in SI Appendix, Fig. S1, polymers made from the segmentally labeled GFP:FUS chimERIC protein yielded NMR spectra indistinguishable from polymers prepared from the isolated, uniformly labeled LC domain of FUS.

Identification of a Cross-β Core within the C-Terminal Half of the FUS LC Domain. The ΔN8 deletion variant contains residues 111 to 214 of the FUS LC domain. From here forward we designate this as the C-terminal half of the FUS LC domain. This protein fragment of 104 residues was expressed in bacterial cells, purified, and incubated under conditions receptive to phase transition (SI Appendix, Materials and Methods). Upon incubation at neutral pH and physiological concentration of monovalent salt (SI Appendix, Materials and Methods), this C-terminal half of the FUS LC domain became phase separated into a gel-like state. When observed by transmission electron microscopy, the hydrogel was found to be composed of uniform, unbranched polymers (SI Appendix, Fig. S2A). X-ray diffraction analysis of hydrogels formed from the C-terminal half of the FUS LC domain revealed diffraction rings at 4.7 and 10 Å (SI Appendix, Fig. S2B). Finally, when analyzed by semidenaturing agarose gel electrophoresis (SDD-AGE), the observed polymers were labile to disassembly (SI Appendix, Fig. S2C).

Truncations of the C-terminal half of the FUS LC domain were prepared having N termini at residues 141, 145, 150, 155, and 160, together with a common C terminus at residue 214 (Fig. 3A). Following incubation under conditions receptive to phase transition, each variant was evaluated for its capacity to polymerize both by time-dependent acquisition of thioflavin-T fluorescence and transmission electron microscopy (Fig. 3B). All variants were observed to form homogenous polymers, save for the most truncated fragment bearing an N terminus at residue 160. We thus conclude that the region of the FUS LC domain located between residues 155 and 190 specifies a secondary cross-β core distinct from that of the N-terminal half of the FUS LC domain characterized extensively in previous studies (8).

The laboratory of R. Tycko has recently described the structure of cross-β polymers formed from the C-terminal half of the FUS LC domain (21). The structural core of the Tycko polymers, whose atomic fold was resolved by cryoelectron microscopy, is specified by residues 112 to 150 of the FUS LC domain. This region of 39 amino acids is not required for soluble test protein to bind to mCherry hydrogels formed from the intact LC domain of FUS (Fig. 2C). It is likewise clear that the cross-β core described by R. Tycko and colleagues is entirely distinct from the C-terminal cross-β core described herein. Truncated variants of the C-terminal half of the FUS LC domain completely lacking residues 111 to 150 are capable of polymerization (Fig. 3B).

To compare cross-β polymers corresponding to those described by the R. Tycko laboratory (21) with those described in this study, fragments of the FUS LC domain spanning either residues 111 to 160 or 141 to 214 were purified and allowed to polymerize under conditions receptive to phase separation. Both fragments readily formed cross-β polymers that were evaluated by electron microscopy, X-ray diffraction, semidenaturing agarose gel electrophoresis, and solid-state NMR spectroscopy. Both samples yielded homogeneous, unbranched polymers as viewed by electron microscopy (SI Appendix, Fig. S3A). Both samples revealed X-ray diffraction images consistent with cross-β amyloid-like polymers (SI Appendix, Fig. S3B).

In order to investigate the possible relationship between this secondary, C-terminal cross-β core and the capacity of soluble FUS LC domain monomers to bind mCherry hydrogels formed from the intact LC domain of FUS (Figs. 1 and 2), we investigated the effects of single tyrosine-to-serine (Y-to-S) mutations upon polymerization of the fragment of the FUS LC domain spanning residues 141 to 214 (Fig. 3A). As shown in Fig. 3C, four of the single Y-to-S mutants (Y143S, Y149S, Y194S, and Y208S) polymerized in a manner indistinguishable from the wild-type fragment, whereas three of the single Y-to-S mutants (Y155S, Y161S, and Y177S) were severely compromised in polymerization, exactly corresponding to the inability of the same mutants in the intact LC domain to incorporate into hydrogels (see Fig. 1C). The exactly correspondent pattern of polymerization by the C-terminal 141 to 214 segment with hydrogel capture of the intact LC domain thus indicates that the C-terminal region of the LC domain alone can recapture aspects of cross-β behavior in the intact LC domain.

Evidence of Molecular Specificity in Elongation of Both NTC and CTC Cross-β Polymers. In efforts to complete a more thorough analysis of the two regions allowing for self-association of the FUS LC domain, we prepared hydrogels composed of either the NTC (residues 2 to 110) or the CTC (residues 111 to 214) and tested their respective abilities to bind two GFP-tagged test proteins. Both hydrogels were made from mCherry fusion
Fig. 3. Polymerization capacity of a fragment of the FUS LC domain spanning residues 141 to 214. (A) Schematic diagram of a C-terminal region of the FUS LC domain having the capacity to form labile cross-β polymers. Locations of seven tyrosine residues are designated numerically (143, 149, 155, 161, 177, 194, and 208). Truncations incrementally removing 4, 10, 15, or 20 residues from the N terminus are shown below the parental 141 to 214 fragment of the FUS LC domain. (B) Parental and N-terminal truncations were incubated under conditions receptive to polymerization (SI Appendix, Materials and Methods). Assays for time-dependent acquisition of thioflavin-T fluorescence (Left) and electron microscopy (Right) were used to monitor formation of amyloid-like polymers. (Scale bar, 200 μm.) (C) Acquisition of thioflavin-T fluorescence at a function of time was compared for fragments of the FUS LC domain spanning residues 141 to 214 bearing either the native sequence or that of variants carrying a single tyrosine-to-serine mutation. Each graph displays fluorescence increase (y axis) relative to time of incubation (x axis). Four mutants, including Y143S, Y149S, Y194S, and Y208S, revealed evidence of polymerization similar to the wild-type protein. Three mutants, including Y155S, Y161S, and Y177S, revealed substantially impeded capacity for polymerization. arb.: arbitrary unit.
are all understood to function in isolation of remaining parts of their respective polypeptides as oncogenic fusion proteins (24). We cautiously contend, as such, that reductionist studies as exemplified in Fig. 5 may be valid.

**ALS-Causing Variants Destabilize the C-Terminal Cross-β Core of the FUS Low-Complexity Domain.** Variants of the FUS LC domain reported to predispose patients to ALS include a glycine-to-glutamic acid missense mutation of residue 156 (G156E) (18) and the deletion of glycine residues 174 and 175 (ΔG174/G175) (19, 20, 25). Recognizing that these variants map within the C-terminal cross-β core of the FUS LC domain, we initially expressed and purified both variants as GFP fusion proteins in the context of the isolated C-terminal core. Upon incubation under conditions normally leading to phase transition, we were surprised to observe that neither variant was able to form cross-β polymers within a time frame in which the native protein readily polymerized.

In order to investigate these observations more carefully, purified, tag-free monomeric protein was tested for polymerization via assays of thioflavin-T fluorescence. Fig. 6B shows polymerization assays for the truncated C-terminal core (FUS 141 to 214) in its native sequence configuration as compared with variants carrying either the G156E missense mutation or G174/G175 deletion (ΔG174/G175) (Fig. 6A). Although anticipating that the ALS-disposing variants might prompt the formation of abnormally stable or more rapidly forming cross-β polymers, we observed no detectable polymerization for the ΔG174/G175 variant and significantly delayed polymerization for the G156E variant (Fig. 6B). When tested in binding assays to hydrogels formed from the CTC alone, weak binding (G156E) or no binding (ΔG174/G175) was observed for these ALS-disposing mutants (Fig. 6C).

The G156E or ΔG174/G175 ALS-disposing variants were further analyzed in the context of the full-length LC domain of FUS by two assays: 1) Acquisition of thioflavin-T fluorescence and 2) hydrogel binding. For the former assay, tag-free proteins corresponding to the native LC domain of FUS, and both ALS-disposing variants, were incubated under conditions of neutral pH and physiological monovalent salt in the presence of thioflavin T. As shown in Fig. 7A, both ALS-disposing variants acquired thioflavin-T fluorescence considerably more rapidly than the native FUS LC domain. This result, which is consistent with published studies of the G156E ALS-disposing variant of the FUS LC domain (26, 27), was notable in revealing the opposite pattern of polymerization from that observed when the three proteins were studied in the context of the isolated, C-terminal half of the FUS LC domain (Fig. 6B).

For hydrogel binding assays, each of the three proteins was linked to GFP, expressed in bacteria, purified, and incubated with mCherry hydrogels formed from the N-terminal half of the FUS LC domain. To our surprise, both ALS-disposing variants bound more prominently to mCherry hydrogels formed from the NTC alone (Fig. 7B). Why do variants carrying mutations that inactivate the C-terminal cross-β core bind NTG hydrogels more strongly than the intact LC domain of FUS?

This observation, we propose, may offer an unanticipated clue as to how low-complexity domains function in living cells.

**Discussion**

In initiating the experiments described herein, we anticipated a simple pathway for the binding of soluble FUS monomers to precast hydrogels. Given the presence of the structured NTC as the defining feature of mCherry hydrogels formed by the intact LC domain of FUS, our expectation was that the soluble, unstructured test protein would be capable of sampling the contours of the folded NTC located at polymer termini and simply slot itself into the existing protein fold. If so, the amino
The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores.
physiological monovalent salt for 100 h. Thio
Δ domain was attenuated for the G156E variant and absent for the
for the GFP fusion linked to the native C-terminal half of the FUS LC
domain. Hydrogel binding activity evident
proteins were incubated with hydrogels formed from mCherry linked to the
causing mutations as assayed within the isolated C-terminal cross-
Measurements of polymerization and hydrogel binding of ALS-
variants (G156E, purple; ΔG174/175, tan). (B) The region spanning resi-
dues 141 to 214 of the FUS LC domain was expressed in its native form
and 4 and when the isolated CTC is incubated at high concentration (Figs. 3
regions do not adopt the cross-
merizes onto termini of existing polymers. In other words,
C-terminal interactions between test protein and hydrogel may
weakly retain the test protein in a position sufficiently proximal
to polymer termini to enable execution of an unfavorable
molecular event. Even though, according to correlate 1, binding
of an unstructured N-terminal region to the cross-β structure at
polymer termini is disfavored, it eventually transpires.
A more obvious indication of mutual exclusivity derives from
structural studies of hydrogel polymers formed from the intact
LC domain of FUS. Solid-state NMR studies of these polymers
have resolved the monomorphic structure of the N-terminal
cross-β core (8). The in-register conformation of protomers held
together by the NTC cause the unstructured, C-terminal domain
of FUS to protrude laterally from the polymer core in a specified
geometry. These extending, C-terminal regions of the polymer
are flexibly positioned 4.7 Å apart. That the aligned, C-terminal
geometry. These extending, C-terminal regions of the polymer
together by the NTC cause the unstructured, C-terminal domain
mer forms by the NTC cross-β structure cannot assemble on subunits of a poly-
mer formed by the NTC cross-β structure. We instead imagine—as schematized in Fig. 8B—that transient attempts toward
CTC formation weakly adhere the soluble GFP-labeled test protein
to mCherry-labeled hydrogel polymers. Stable hydrogel
binding is limited to the event wherein the test protein copoly-
erizes onto termini of existing polymers. In other words,
C-terminal interactions between test protein and hydrogel may
weakly retain the test protein in a position sufficiently proximal
to polymer termini to enable execution of an unfavorable
molecular event. Even though, according to correlate 1, binding
of an unstructured N-terminal region to the cross-β structure at
polymer termini is disfavored, it eventually transpires.
Finally, the concept of mutual exclusion may explain the per-
plexing data shown in Fig. 7B. mCherry hydrogel droplets com-
posed solely of the NTC of FUS were challenged with GFP-tagged protein corresponding to the full-length LC domain bearing its
native sequence, or that of either ALS variant. We offer that the
enhanced binding of the latter proteins reflects the fact that, unlike
the native LC domain, they cannot form CTC cross-β structures. According to correlate 2, mutual exclusion predicts an impediment
to copolymerization with NTC polymers of the hydrogel if the test
protein can itself form CTC cross-β interactions. Since the native

![Fig. 6. Measurements of polymerization and hydrogel binding of ALS-causing mutations as assayed within the isolated C-terminal cross-β core of the FUS low-complexity domain.](https://doi.org/10.1073/pnas.2114412118)
The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores.
form the CTC cross-β structure (Fig. 5B). The resulting proximity of two unstructured NTC regions is predicted to facilitate a process of isomerization that begets formation of the slightly more stable NTC cross-β structure. Important to the thesis articulated herein, the concept of mutual exclusion demands that the initial CTC cross-β structure be disassembled in order for the NTC structure to form.

Once equilibrium has been reached, we offer two reasons to account for impediments to further growth of NTC polymers. First, if a third FUS LC domain were to use its CTC to invade the free CTC of the dimer held together by NTC cross-β interactions, mutual exclusivity would demand that NTC interactions dissolve (correlate 2). The same idea would guard against the ability of a third FUS LC domain to invade the free and unstructured NTC of a dimer held together by CTC cross-β interactions. Should runaway polymerization by either cross-β core take place, the correlate of mutual exclusion may no longer apply. If a series of NTCs are coassembled into a long polymer, we propose that an incoming monomer is unable to use its CTC to productively coalesce with any of the unstructured CTC domains extending laterally from the polymer. We offer that this option is voided because each NTC internal to the FUS LC domain would prefer interaction with the region of the free CTC of the dimer held together by NTC cross-linking either of the dimers shown in Fig. 8 could be expected to use its CTC to force dissolution of a cross-β assembly bearing the forces of structural order from both sides.

Our second reason for hypothesizing orderly limitation of FUS polymer growth derives from correlate 1. Upon encountering either of the dimers shown in Fig. 8C, an unstructured FUS LC domain would prefer interaction with the region of FUS not already existing in the cross-β structural state. If the dimer were held together by NTC interactions, the incoming protein is predicted to use its CTC in an attempt to form cross-β interactions with the unstructured CTC of the dimer. Reciprocally, if the dimer were held together by CTC interactions, the incoming protein would be expected to use its NTC to attempt self-association with the unstructured NTC of the dimer. These preferences, as specified in correlate 1, are attributed to conformal restriction. If the bias of conformational restriction is strong, and if mutual exclusivity is likewise strong, the combination of these limitations should prevent the FUS LC domain from polymerizing any further than the dimeric state.

We close with emphasis on the multitude of unknowns that cloud our simplistic ideas. The FUS LC domain is subject to many forms of posttranslational modification (PTM) that may influence behavior of the NTC and CTC domains. Differential regulatory effects of PTMs might allow the FUS LC domain to expand either NTC- or CTC-mediated polymerization on demand. The FUS protein likewise shuttles from the nucleus to cytoplasm and back and uses its LC domain for any of a number of heterotypic interactions with other proteins. These variables can be understood to open the opportunity for the FUS LC domain to deviate from the dimeric ground state imposed by correlates 1 and 2. Despite the complexities of this science, we have faith in the value of the reductionist approach exemplified by the experiments described herein.

Materials and Methods
See SI Appendix, Materials and Methods for detailed materials and methods that describe 1) cloning of expression plasmids; 2) protein expression and purification; 3) hydrogel binding assays; 4) polymer extension assays; 5) polymer formation; 6) X-ray diffraction; 7) SDD-AGE; 8) thioflavin-T assays; 9) thermal stability assays; and 10) sample preparation for ssNMR.

Data Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Robert Tycko and Myungwoon Lee for help in recording ssNMR spectra. We thank Deepak Nijhawan and Glen Lissczak for thoughtful discussions regarding the research described herein. We also thank Lillian Sutherland, Lily Sumrow, and Leeloo Wu for plasmid cloning. S.L.M. was supported by National Institute of General Medical Sciences Grant 5R35GM130358 and National Cancer Institute Grant 1U54CA231649, as well as unrestricted funding from an anonymous donor.

References
1. N. Radó-Trilla, M. Albá. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins. BMC Biol. 12, 155 (2016).
2. M. Toli-Riera, N. Radó-Trilla, F. Marty, M. M. Albá. Role of low-complexity sequences in the formation of novel protein coding sequences. Mol. Biol. Evol. 29, 883–886 (2012).
3. P. Leuenberger et al., Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science 355, eaau7825 (2017).
4. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen. Biomolecular condensates: Organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 18, 285–298 (2017).
5. Y. Shin, C. P. Brangwynne. Liquid phase condensation in cell physiology and disease. Science 357, eaaf4382 (2017).
6. T. W. Han et al., Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. Cell 149, 768–779 (2012).
7. M. Kato et al., Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. Cell 148, 753–767 (2012).
8. D. T. Murray et al., Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. Cell 171, 615–627.e16 (2017).
9. R. Tycko, Solid-state NMR studies of amyloid fibril structure. Annu. Rev. Phys. Chem. 62, 279–299 (2011).
10. M. Fändrich, J. Meinhardt, N. Grigorieff. Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. Prion 3, 89–93 (2009).
11. S. Xiang et al., The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. Cell 163, 829–839 (2015).
12. V. H. Ryan, et al., Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. Mol. Cell 65, 456–479.e7 (2018).
13. D. T. Murray et al., Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. Proc. Natl. Acad. Sci. U.S.A. 115, E9782–E9791 (2018).
14. J. Lu et al., CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. Nat. Commun. 11, 4090 (2020).
15. H. J. Kim et al., Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multifunctional proteinopathy and ALS. Nature 495, 467–473 (2013).
16. N. M. Vieira et al., A defect in the RNA-processing protein HNRNPC causes limb-girdle muscular dystrophy 1G (LGMD1G). Hum. Mol. Genet. 23, 4103–4110 (2014).
17. Y. Lin et al., Toxic PR poly-dipeptides encoded by the C1orf72 repeat expansion target LC domain polymers. Cell 167, 789–802.e12 (2016).
18. N. Ticozzi et al., Analysis of FUS gene mutation in familial amyotrophic lateral sclerosis within an Italian cohort. Neurology 73, 1180–1185 (2009).
19. T. J. Kwiatkowski, Jr. et al., Mutations in FUS/LT1 gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323, 1205–1208 (2009).
20. R. Rademakers et al., FUS gene mutations in familial and sporadic amyotrophic lateral sclerosis. Muscle Nerve 42, 170–176 (2010).
21. M. Lee, U. Ghosh, K. R. Thurber, M. Kato, R. Tycko. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. Nat. Commun. 11, 5735 (2020).
22. S. Qamar et al., FUS phase separation is mediated by a molecular chaperone and methylation of arginine cation–s interactions. Cell 173, 720–734.e15 (2018).
23. J. Wang et al., A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell 174, 688–699.e16 (2018).
24. H. Kovar, Dr. Jekyll and Mr. Hyde: The two faces of the FUS/EWS/TAF15 protein family. Sarcoma 2011, 837474 (2011).
25. X. Yan et al., Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. Neurology 75, 807–814 (2010).
26. A. Patel et al., A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell 162, 1066–1077 (2015).
27. T. Murakami et al., ALSFTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. Neuron 88, 678–690 (2015).
28. R. M. Vernon et al., Pi-Pi contacts are an overlooked protein feature relevant to phase separation. eLife 7, e31486 (2018).

Kato and McKnight
The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores

10 of 10 | PNAS

https://doi.org/10.1073/pnas.2114412118

Kato and McKnight
The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores

10 of 10 | PNAS

https://doi.org/10.1073/pnas.2114412118

Kato and McKnight
The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores