



Supplementary Materials for

Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks

Michael P. Hughes, Michael R. Sawaya, David R. Boyer, Lukasz Goldschmidt,
Jose A. Rodriguez, Duilio Cascio, Lisa Chong, Tamir Gonen, David S. Eisenberg*

*Corresponding author. E-mail: david@mbi.ucla.edu

Published 9 February 2018, *Science* **359**, 698 (2018)
DOI: 10.1126/science.aan6398

This PDF file includes:

Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References

Other Supplementary Material for this manuscript includes the following:

(available at www.sciencemag.org/cgi/content/full/359/6376/698/DC1)

Table S3 as Excel file

Materials and Methods:

FUS purification and hydrogel formation

The plasmid with pHis-parallel-mCherry-1FUS214 was a generous gift from the McKnight Lab (UT Southwestern, Dallas, TX) and FUS was purified and formed into a gel according to the work presented in Kato et al. (5). A hydrogel was formed by placing purified mCherry-FUS concentrated to 100mg/ml in 4° for 2 months.

Negative stain EM

A pipette was used with remove 2µl of gel and then mix with 798µl H₂O and pipetted vigorously to break up gel. This sample was pipetted on a copper mesh EM grid and stained with 2% uranyl acetate. The grid was scanned for fibrils and images were collected on a T12 120kV electron microscope (FEI) with a Gatan 2kX2k CCD.

Protein segment crystallization and structure determination

FUS | 37-SYSGYS-42 | peptide was ordered from GenScript (Piscataway, NJ). A peptide stock solution was prepared by adding nano-pure H₂O to achieve a concentration of 10mg/ml. Solubility was enhanced by autoclaving, and then the peptide solution was immediately used for crystallization. Crystals grew at room temperature by hanging drop vapor diffusion. The reservoir contained 0.1 M Bis-Tris pH 6.0, 0.2 M magnesium formate dihydrate. Crystals were cryoprotected by transferring them into 50% glycerol and then mounted on a loop. Diffraction data were collected at the Advanced Photon Source (APS) on beamline 24-ID-E at 100 K. Phases were determined by molecular replacement using the program Phaser (36). The successful search model was a truncated form of the peptide MIHFGND (PDB ID code 3NVH) with sidechains and methionine removed. Refinement was performed with Refmac (37).

FUS | 54-SYSSYGQS-61 | peptide was ordered from GenScript (Piscataway, NJ). The peptide was solubilized by adding nano-pure H₂O to achieve a concentration of 75mg/ml, and then autoclaved. The peptide solution was immediately used for crystallization. Crystals grew at room temperature by hanging drop vapor diffusion. The reservoir contained 0.5 M ammonium tartrate dibasic pH 7.0. Crystals were cryoprotected by transferring them into 50% glycerol and then mounted on a loop. Diffraction data were collected at the APS beamline 24-ID-E at 100 K. Phases were determined by direct methods using the program ShelxD (38). Refinement was performed with Refmac (37).

hnRNPA1 | 243-GYNGFG-248 | peptide was ordered from Innoprep (San Diego, CA). The peptide was solubilized by adding nano-pure H₂O to achieve a concentration of 10mg/ml, and then autoclaved. The peptide solution was immediately used for crystallization. The crystals grew at room temperature by hanging drop vapor diffusion. The reservoir contained 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate pH 6.5, and 30% (v/v) MPD. Crystals were mounted dry on the ends of pulled glass needles. Data were collected at the APS beamline 24-ID-E at 100 K. Phases were determined by direct methods using the program ShelxD (38). Refinement was performed with Refmac (37).

All X-ray diffraction data sets were collected on an ADSC Q315 CCD detector. All diffraction data were processed with XDS (39).

FUS | 77-STGGYG-82 | peptide was ordered from Innopep (San Diego, CA). The peptide was dissolved in nano-pure H₂O at 150mg/ml. Crystals grew at room temperature by hanging drop vapor diffusion. The reservoir contained 0.1 M sodium acetate pH 4.6, 0.15 M ammonium sulfate, and 25% (v/v) PEG 2000 MME. Several drops were then collected using a pipette and put in a 200µl PCR tube and placed in a water bath sonicator. The solution was sonicated 20min with ice to prevent melting the peptide crystals by temperature increase. A 2-3µl drop of resulting suspension was deposited onto a Quantifoil holey-carbon EM grid. Grids were then blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (FEI). On a per-crystal basis, blotting times and forces were optimized to keep a desired concentration of crystals on the grid and to avoid damaging the crystals. Frozen grids were then either immediately transferred to liquid nitrogen for storage or placed into a Gatan 626 cryo-holder for imaging. Images and diffraction patterns were collected from crystals using an FEG-equipped FEI Tecnai F20 TEM operating at 200 keV. All data were recorded using a bottom mount TVIPS F416 CMOS camera with a sensor size of 4000 by 4000 pixels, each 15.6 by 15.6 µm in size. Diffraction patterns were recorded by operating the detector in a continuous capture mode termed 'rolling shutter' with 2 x 2 pixel binning. This produced images 2000 by 2000 pixels in size. Exposure times for these images ranged from 1-2 seconds per image. During each exposure, crystals were unidirectionally rotated within the electron beam at a fixed number of degrees per second, corresponding to a fixed angular wedge per frame.

Nup98 | 116-GFGNFGTS-123 | peptide was ordered from GenScript (Piscataway, NJ). The peptide was solubilized by adding nano-pure H₂O with 20% DMSO to achieve a concentration of 12mg/ml. The peptide solution was immediately used for crystallization. Crystals grew at room temperature by hanging drop vapor diffusion. The reservoir contained 10% ethanol and 0.1M CHES pH 9.5. Several drops were then collected using a pipette and put in a 200µl PCR tube and placed in a water bath sonicator. The solution was sonicated 20min with ice to prevent melting the peptide crystals by temperature increase. A 2-3µl drop of resulting suspension was deposited onto a Quantifoil holey-carbon EM grid. Grids were then blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (FEI). On a per-crystal basis, blotting times and forces were optimized to keep a desired concentration of crystals on the grid and to avoid damaging the crystals. Frozen grids were then either immediately transferred to liquid nitrogen for storage or placed into a Gatan 626 cryo-holder for imaging. Images and diffraction patterns were collected from crystals using an FEG-equipped FEI Tecnai F20 TEM operating at 200 keV. All data were recorded using a bottom mount TVIPS F416 CMOS camera with a sensor size of 4000 by 4000 pixels, each 15.6 by 15.6 µm in size. Diffraction patterns were recorded by operating the detector in a continuous capture mode termed 'rolling shutter' with 2 x 2 pixel binning. This produced images 2000 by 2000 pixels in size. Exposure times for these images ranged from 1-2 seconds per image. During each exposure, crystals were unidirectionally rotated within the electron beam at a fixed number of degrees per second, corresponding to a fixed angular wedge per frame.

Crystals that appeared undistorted and that were 100-300nm thick produced the best diffraction. Since our crystals are needle-like and lie preferentially oriented on EM grids, data sets were collected from multiple crystals in different orientations and these sets were merged in

order to achieve sufficient data completeness and redundancy. Data from individual crystals spanned a wedge of reciprocal space ranging from 60-80°, within 100-200 frames. We used a selected area aperture with an illuminating spot size of approximately one micron. The geometry detailed above equates to an electron dose of less than 0.1 e-/Å² per second being deposited onto our crystals. Measured diffraction images were converted from tiff or tvips format into SMV crystallographic format, using in-house software available for download at <https://cryoem.janelia.org/downloads/> (40).

Electron diffraction images were indexed and integrated with XDS (39) and data sets originating from different crystals were scaled together with XSCALE (39). Phases for FUS-STGGYG and nup98-GFGNFGTS were determined by direct methods using the program ShelxD (38). Crystallographic refinement were performed using phenix (41) and buster (42) and a library of electron scattering factors. All models were built with Coot (43).

Crystal powder diffraction

All images were collected from crystal clusters or hydrogel formed according to above protocols. All samples were shot wet and without cryo-stream. Images were collected with a RIGAKU R-AXIS HTC imaging plate detector using Cu K(alpha) radiation from a FRE+ rotating anode generator with VARIMAX HR confocal optics (Rigaku, Tokyo, Japan). The sample-to-detector distance was 200mm. The sample rotated 5° during the exposures.

Atomic solvation energy calculations

For each structure, crystal symmetry was used to generate a fibril model with two β-sheets each composed of 10 strands. The exposed surface area for each atom was measured with a 1.4Å diameter probe using the program areaimol from the CCP4 package (44). One β-sheet was translated away and the exposed surface area was measured again. These values were subtracted to find the difference in exposed surface area for each atom (ΔA_i). The exposed surface area was then multiplied by an atomic solvation parameter (22) to get the estimated free energy for solvating that atom (σ_i).

Atomic solvation parameters (σ_i) for each atom:

C = 18

N or O = -9

O⁻ = -37

N⁺ = -38

All atom atomic solvation energies were then summed for a strand:

$$\Delta G^\theta = \sum_{atoms\ i} \Delta A_i \sigma_i$$

N-terminal and C-terminal charges on the peptide backbone were considered neutral, as they would be in the full length protein. The atomic solvation energy then reflects the calculated energy to melt the fibril interface. Energies are taken from a single, most central strand in one β-sheet.

Identifying LC domains

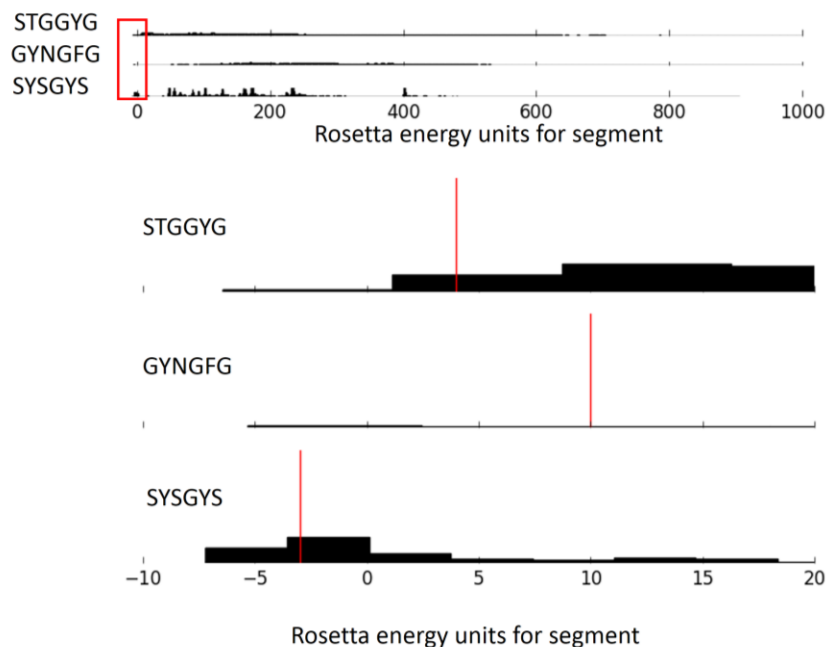
To identify LC domains on proteins we used the SEG algorithm (14) with normal parameters. We looked for stretches of at least 25 adjacent residues to consider the region to be an LC domain, but allowed for an interruption by up to five residues not considered LC to get rid of artifacts in the algorithm. This allowed us to identify 178 proteins with LC domains out of a UniProt proteome of 3990 proteins in *Escheria coli*; 1202 proteins with LC domains out of a UniProt proteome 6720 proteins in *Saccharomyces cerevisiae*; and 5893 proteins with LC domains out of a UniProt proteome of 20197 proteins in *Homo sapiens*. Minimal proteomes were downloaded from the UniProt consortium (45) : *Homo sapiens* was downloaded March 28, 2016; *Escheria coli* was downloaded April 2, 2016, and *Saccharomyces cerevisiae* was downloaded April 4, 2016.

Template based threading

Using a process analogous to our 3D profile method for predicting steric-zipper-forming segments (25), we used RosettaDesign for a high-throughput computational search of other sequences compatible with the STGGYS, SYSGYS, and GYNGFG backbones. Each structural backbone template consists of the assembly of two (STGGYS, GYNGFG) or three (SYSGYS) beta-sheets, with each sheet made up of ten hexapeptides with the same residue sequence. Method for identifying interface for STGGYG structure is shown in Fig. S9. Interfaces for LARKS and zippers are selected for the following reasons: There is usually only 1 significantly interacting cross-beta interface. We define the steric zipper or LARKS as the interface with the strongest interaction, as judged by the area buried and shape complementarity. The one exception in this paper is with SYSGYS where the two interfaces seem roughly equivalent but only the selected one can be part of a longer protein; in the other, the connection to the rest of the chain is sterically blocked. We threaded the sequences of 20,120 proteins from a non-redundant human proteome (45) that collectively contains 7,900,599 unique 6-residue segments. Segments were evaluated with the standard Rosetta energy function (Talaris2014) except with changes to the Dunbrack rotamer and Lennard-Jones potential terms for the following reasons:

1. The Dunbrack energy term in the Rosetta scoring function represents the probability that a chosen residue rotamer is native-like given the backbone phi, psi angles. This knowledge-based term is derived from structures of proteins in the PDB. Observed angles are placed into wells with steep energy barriers between them. Therefore, this term is rather sensitive to small deviations from the optimal angles observed in globular proteins. Steric zippers peptides, however, are small and have the freedom to adapt less common confirmations. To capture this in our modeling protocol, we made the acceptable angle wells wider by increasing the cutoffs from optimal values to 4.0 standard deviations.
2. LJ switch distance: the repulsive van der Waals energy in Rosetta is computed using a combination of a steep energetic term ($1/d^{12}$) and a softer linear term at short atom distances. The LJ switch parameter defines the distance at which the transition between these calculations occurs. Rosetta uses this approach to improve performance in minimization and overall structure prediction and design calculations. We have increased the switch distance from 0.6 to 0.8 Angstroms because the interfaces of steric zippers in crystal structures tend to be packed tighter than those of globular proteins.

Models with energies below: -3.5 for NKGAIL, 4.0 for STGGYG, 10.0 for GYNGFG, and -3.0 for SYSGYS (modified Rosetta Energy Units) were considered compatible with these structure backbones. Energy cutoffs for the steric zipper NKGAIL were determined as in Goldschmidt et al. (25) with about 20% of sequences scoring favorably. Energy cutoffs for LARKS were determined by looking at a histogram of scores and choosing the most energetically favorable segments. For the STGGYG and SYSGYS backbones the most favorable segments of the first peak in the histogram of segment scores was chosen. Since fewer segments scored well on the GYNGFG backbone, the cutoff was put on the right side of the first peak. We believe these to be the most conservative cutoffs we can use, only opting to consider the most favorable of scoring segments as LARKS. Histograms for energies of segments are shown below, with the region in the red box blown up for fine detail. Vertical lines indicate the corresponding score cutoff for that backbone.



We determined the fraction of compatible segments in each protein. If any segment scored well on any of the three LARKS backbones, and it was in a LC domain, it was considered favorable for forming a LARKS. We only considered LARKS in LC domains because in globular domains, it would be unlikely that they would be available for interaction.

Identifying top 400 proteins enriched in LARKS

LARKS were found using the above methods. If at least six adjacent residues were predicted to be in LARKS, it was scored as a single LARKS. Proteins were ordered by the largest number of LARKS, and the top 400 were qualitatively categorized (Fig. 3C) based upon their UniProt GO annotations. For more rigorous methodology we used GO to find enriched GO terms in these top 400 proteins.

Gene ontology on top 400 proteins enriched in LARKS

The gene ontology tree and annotations for the human proteome were downloaded from the Gene Ontology Consortium (46). The GO terms for the top 400 proteins were found 4412 unique GO terms represented by these top 400 proteins. We used standard bootstrapping methods to identify statistically enriched terms. For bootstrapping, to test if a GO term was enriched, 400 random proteins were chosen from the human proteome and we counted how many times that GO term was found in those 400 proteins. This procedure was repeated 15,000 times to create a distribution of the number of times that GO term was found in a random selection from the proteome. If the actual number of times that GO term was found in the top 400 proteins rich in LARKS was greater than the 95th percentile of times that GO term was found in our random distribution, then it was considered enriched. P-values are from the score on the randomly selected distribution. This yielded 496 enriched GO terms. Several were qualitatively selected for display in supplemental table S2.

Venn diagram overlap of proteins rich in structures predicted by different threading backbones.

For fig. S7 we threaded the human proteome on four different backbones: SYSGYS, STGGYG, GYNGFG, and NKGAI. The first three being examples of kinked backbones to search for LARKS and the latter being a steric zipper to search for traditional amyloid forming segments. For each peptide backbone, all proteins in the human proteome were ranked by number of hits for that backbone, and then the top 400 were taken. Overlap of proteins predicted by each backbone was found, and presented on a 4-way Venn diagram. Low-complexity domains were not considered in determining if a segment was a hit, as in prediction for LARKS in the rest of the paper.

Venn diagram of sequence bias and predicted structures

For fig. S8, we ranked the human proteome according to these four methods:

- 1) Highest fraction on the protein in predicted LC domains (as described in Materials and Methods)
- 2) Highest content of Glycine or Serine (these residues are highly over represented in human LC domains)
- 3) Most number of predicted LARKS (as described in Materials and Methods)
- 4) Most number of predicted steric zippers (disregards LC domains)

Then the top 400 proteins for each category were taken and overlap between the 4 methods was represented on a 4-way Venn diagram.

Hydrogel from synthetic LARKS construct

The synthetic LARKS construct with sequence: SYSGYSGDTSYSSYGQSNPSTGGYG was chosen by linking the three LARKS structures from FUS with three residue linkers that were representative of the amino acid bias found in the FUS-LC domain and did not introduce any new predicted LARKS into the peptide.

To form the hydrogel in Fig. 2, the peptide (from GenScript (Piscataway, NJ)) was dissolved in nano-pure water to 50mg/ml (concentrations 100 mg/ml were also used for experiments on Fig. S2) in a 200 μ l PCR tube. A hydrogel was formed after the solution was incubated at 4°C overnight. Sample was melted by incubating the sample for two hours in a PCR thermocycler set to heat the sample to 60°C and with the lid heated to 96°C. To identify of the sample was a liquid or a gel, a 5 μ l air bubble was introduced with a micropipette. The 200 μ l tube was then inverted to see if the air bubble would rise to the top. Pictures were taken with the sample tubes inverted, resting on a black surface. The pictures were converted to grayscale, then the contrast was increased to wash at the background of the image. The image color was then inverted to give a white background. Negative stain in Fig. 2 was performed by pipetting 2 μ l of the liquid samples or scraping a chunk of hydrogel onto a glow discharged copper mesh EM grid and stained with 2% uranyl acetate for 2 minutes. Images were collected on a 200kV electron microscope (FEI) with a Gatan 2kX2k CCD.

Additional Author notes:

Author contributions: DSE and MPH conceived of the project; MPH and LC carried out all crystallographic experiments; MPH, MRS, JAR, DRB, and DC collected diffraction data and determined structures; MPH, MRS, and LG carried out computations; and MPH, MRS, and DSE wrote the manuscript with contributions from all authors.

Supplementary Figures:

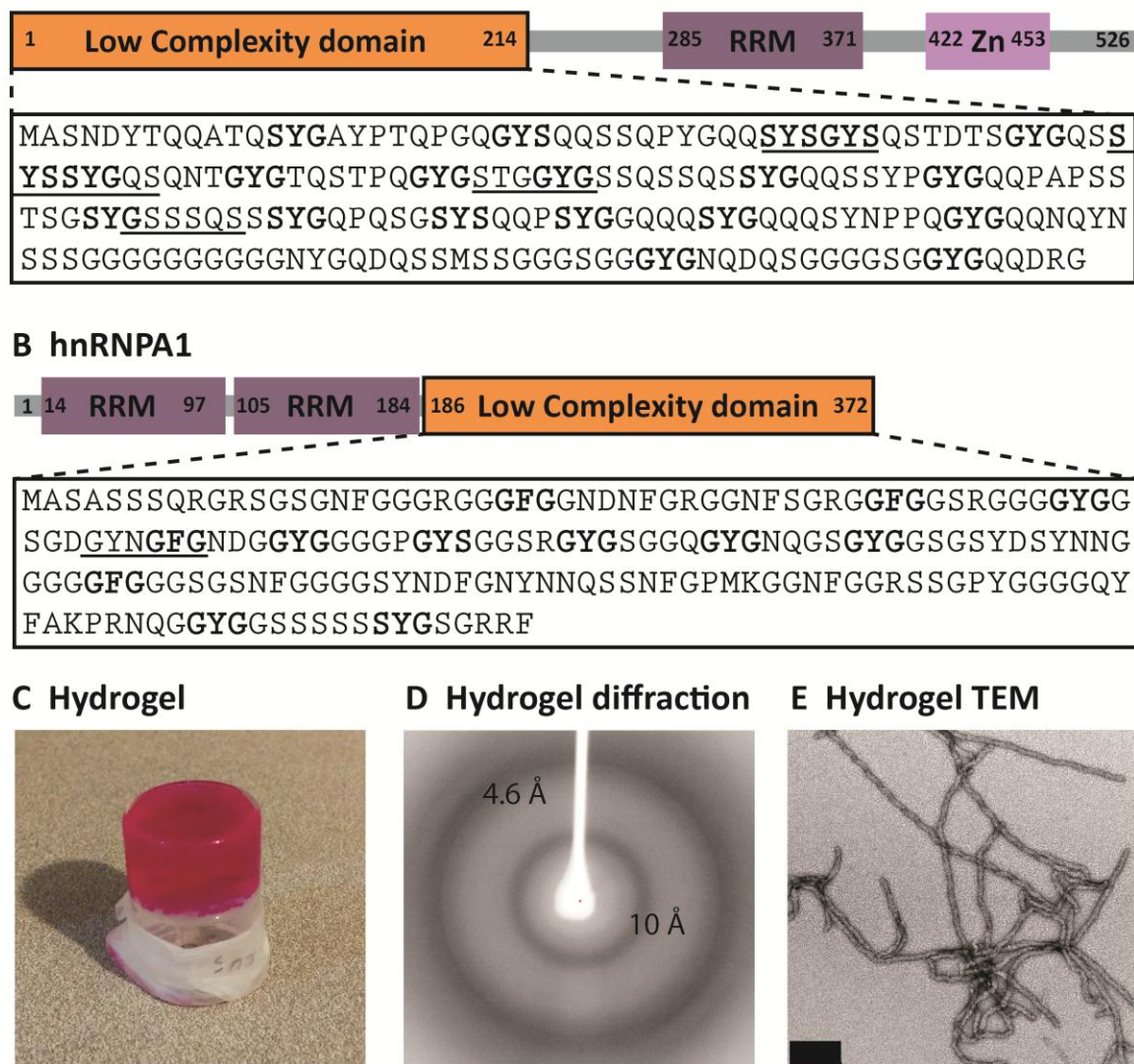


Fig. S1. Low complexity (LC) domains of FUS and hnRNPA1. (A and B) Amino acid sequences of the LC domains of FUS and hnRNPA1. Residues [G/S][Y/F][G/S] motifs are in bold and sequences of structures determined in this work are underlined. (C) Hydrogel formed from the LC domain of FUS fused with mCherry. The hydrogel sits atop a 5 ml beaker which had previously been inverted. When stirred, the hydrogel reliquified. (D) X-ray fiber diffraction pattern of the mCherry-FUS hydrogel, displaying the cross- β diffraction pattern characteristic of amyloid structure. (E) Negatively stained transmission electron micrograph of 400x diluted mCherry-FUS hydrogel shows protofilamentous structures. Scale bar is 0.2 μ m.

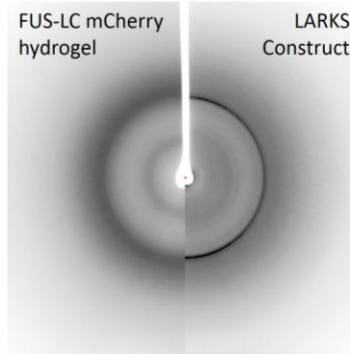
A Synthetic multi-LARKS construct

SYSGYSGDTSYSSYGQSNGPSTGGYG

B Hydrogel



C Hydrogel diffraction



D Hydrogel TEM

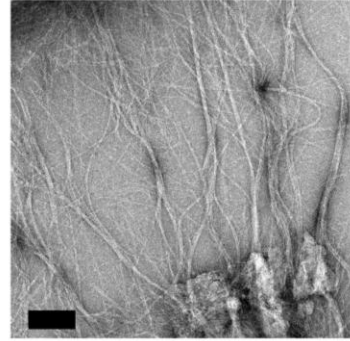


Fig. S2. Synthetic multi-LARKS construct. (A) Sequence of synthetic construct used in experiments. Three LARKS structures from FUS (underlined) are fused in tandem by linker sequences that reflect the amino acid composition bias of the FUS LC domain. (B) Hydrogel formed by synthetic construct at 100 mg/ml, similar to the concentration to form the FUS-LC mCherry hydrogel, spread on a glass slide under a light microscope. (C) X-ray diffraction of the synthetic construct is similar to the FUS-LC mCherry hydrogel. (D) Negative stain EM of the synthetic construct reveals a dense mesh of protofilaments interacting with each other.

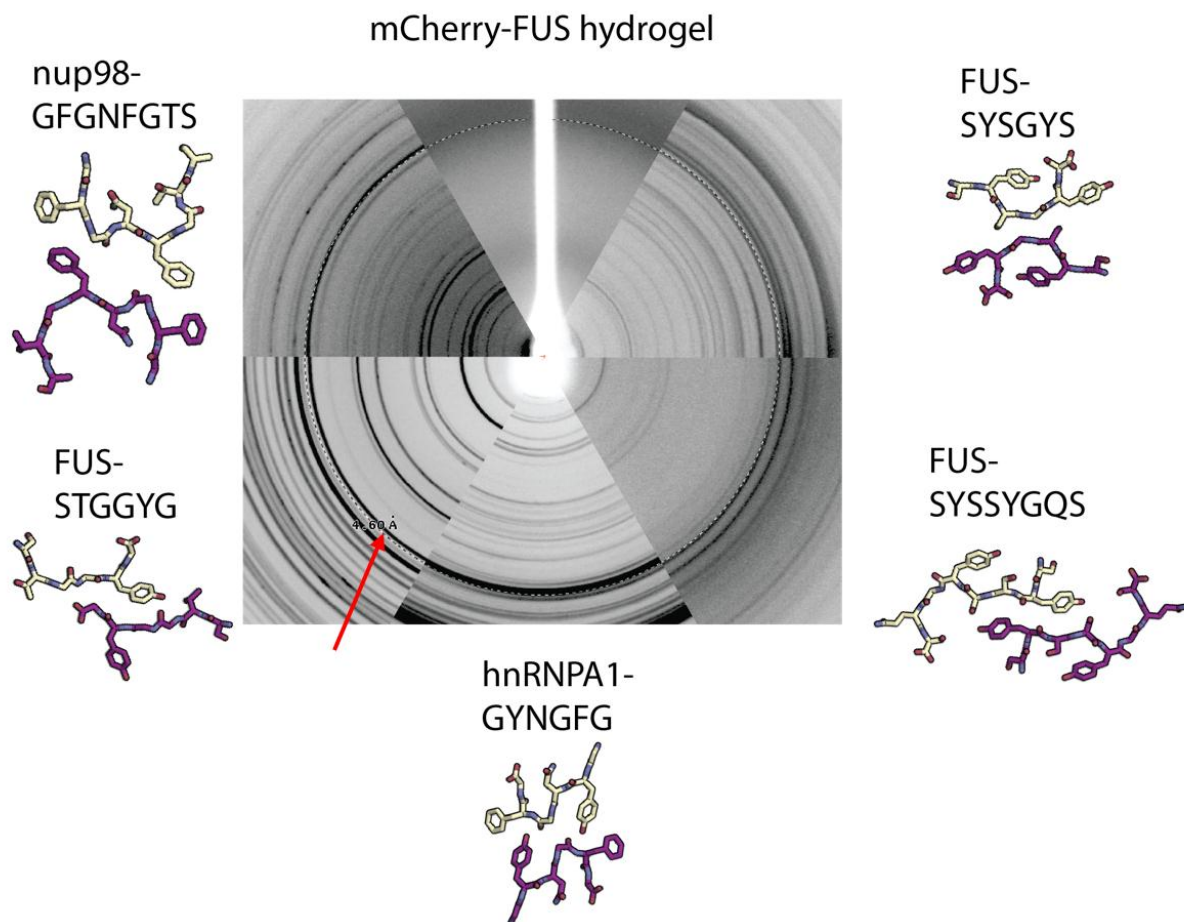


Fig. S3. Fiber X-ray diffraction of LARKS compared to that of FUS hydrogel. Each wedge shows the fiber diffraction of the corresponding samples recorded under identical conditions (distance to detector, X-ray wavelength, and detector) to allow direct comparison of Bragg spacings. The patterns from the five LARKS are the crystalline powder diffraction patterns. The dotted circle pointed to by the red arrow shows spacing of 4.60 Å. All LARKS display a ring at, or very near, 4.60 Å indicating their consistency with the diffraction pattern from the FUS LC domain. The greater width of the 4.60 Å hydrogel ring reflects the relative disorder of the gel compared to the crystals.

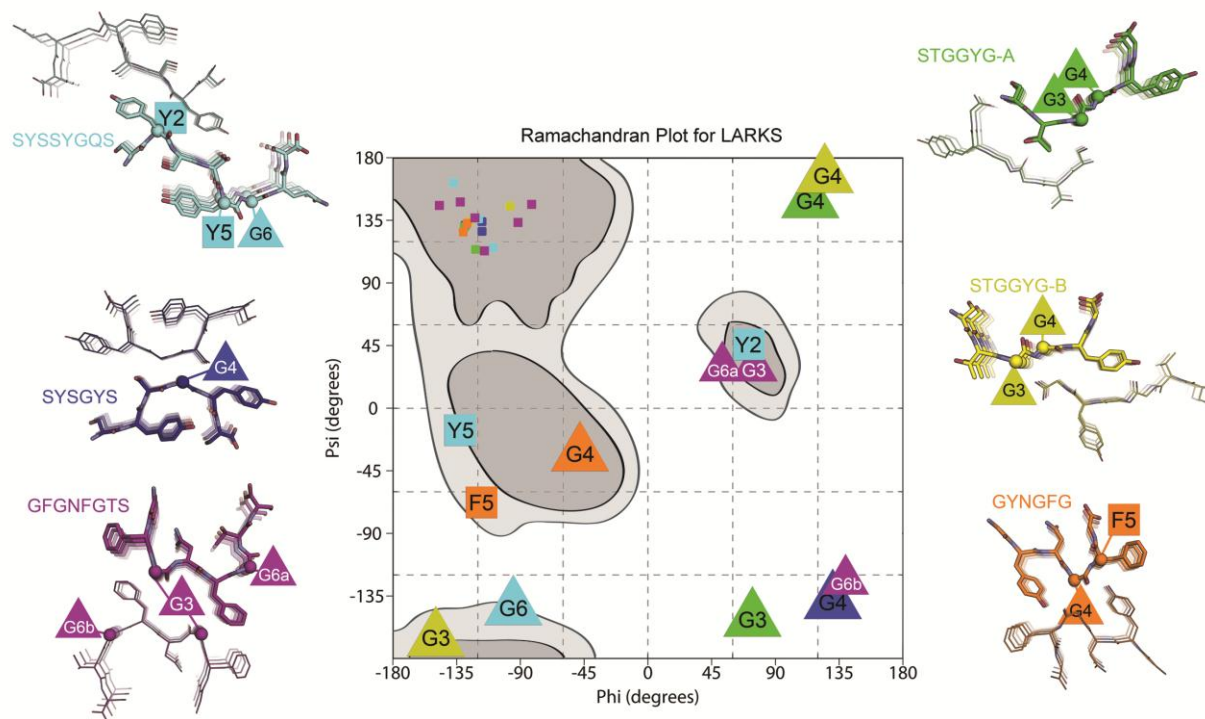


Fig. S4. Ramachandran plot of LARKS. Dihedral (Phi-Psi) angles for all residues in our atomic structures of LARKS shown in a Ramachandran diagram. The Phi-Psi angles are color coded by structure. Spheres and corresponding labels indicate the alpha-carbon atoms of residues with large deviations from standard β -sheet Phi-Psi values. Triangles indicate glycine residues that occupy non- β -sheet Phi-Psi angles, and squares represent aromatic residues that occupy non- β -sheet Phi-Psi angles. All Phi-Psi angles fall in allowed regions of the Ramachandran diagram.

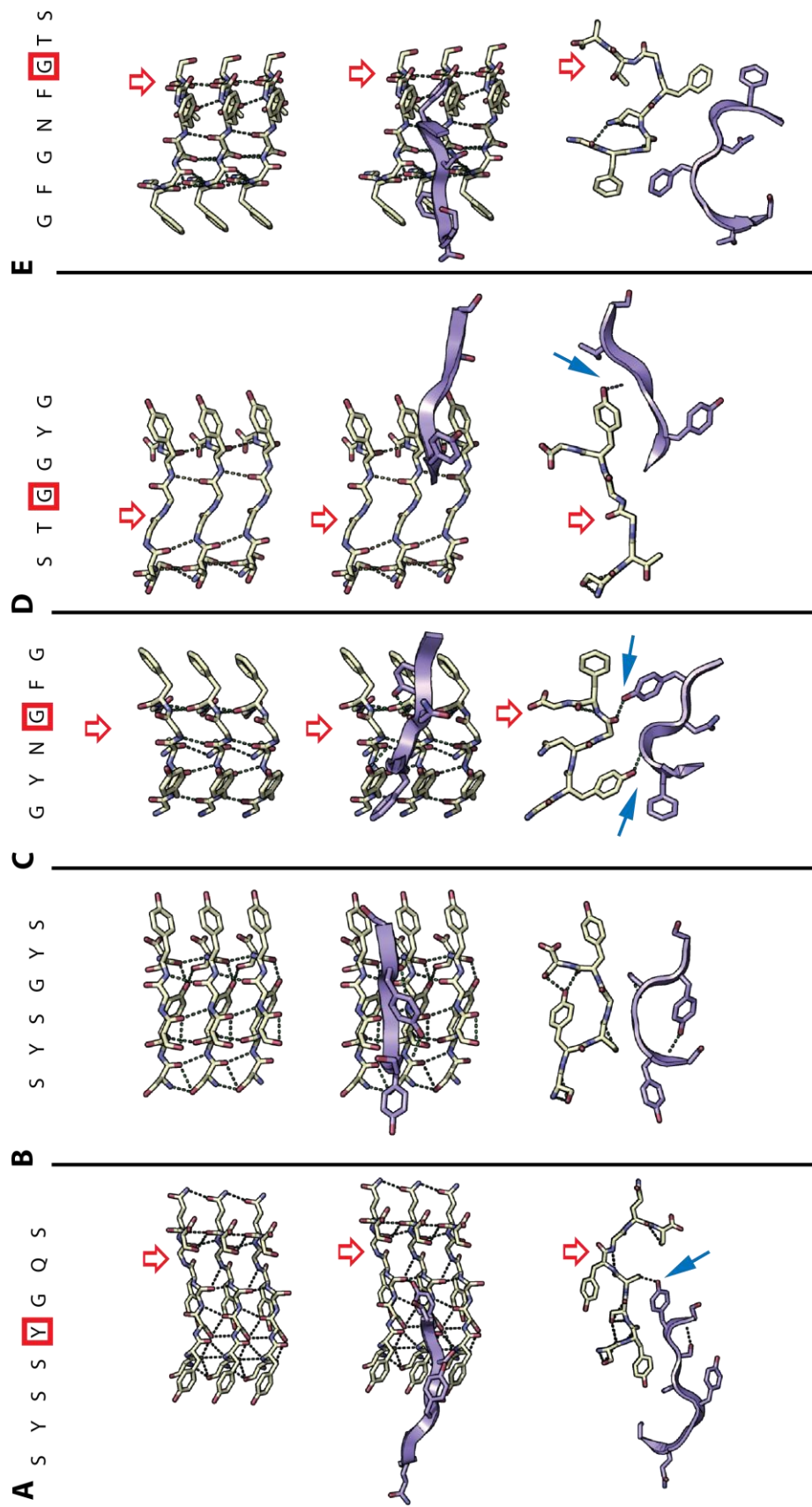
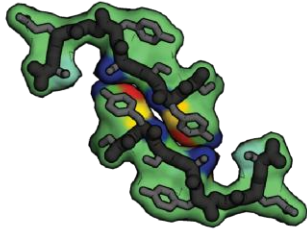


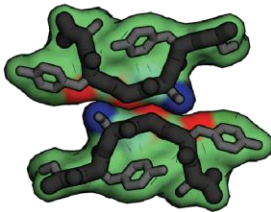
Fig. S5. Intra-sheet and inter-sheet patterns of hydrogen bonding in five LARKS. The top row displays three stacked beta-strands in our five LARKS viewed perpendicular to their fibril axes with the corresponding segment sequence above. The second row displays one strand of the mating sheet, colored blue, in the foreground. The third row views the fibril parallel to the fibril axis. Intra-sheet hydrogen bonds are shown by dotted lines; inter-sheet hydrogen bonding from unsatisfied backbone donors/acceptors are highlighted with blue arrows. Red boxes on the sequence indicate residues with carbonyl/amide backbone hydrogen bond donors/acceptors that are not satisfied by hydrogen bonding along the fibril axis, and instead form hydrogen bonds with mating sheets. Red arrows point to unsatisfied hydrogen bond donors and acceptors. **(A)** Structure of SYSSYGQS **(B)** structure of SYSGYS **(C)** structure of GYNGFG **(D)** structure of STGGYG **(E)** structure of GFGNFGTS.

LARKS

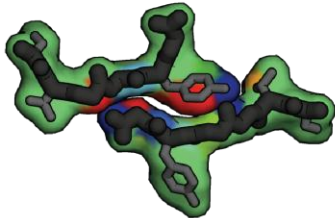
FUS-SYSSYGQS



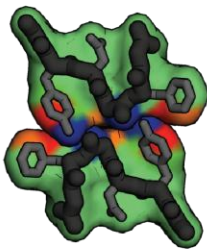
FUS-SYSGYS



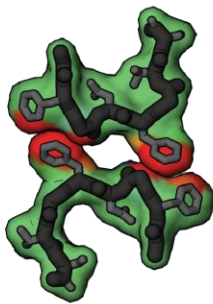
FUS-STGGYG



hnRNPA1-GYNGFG



nup98-GFGNFGTS



Zippers

A β -NKGAI

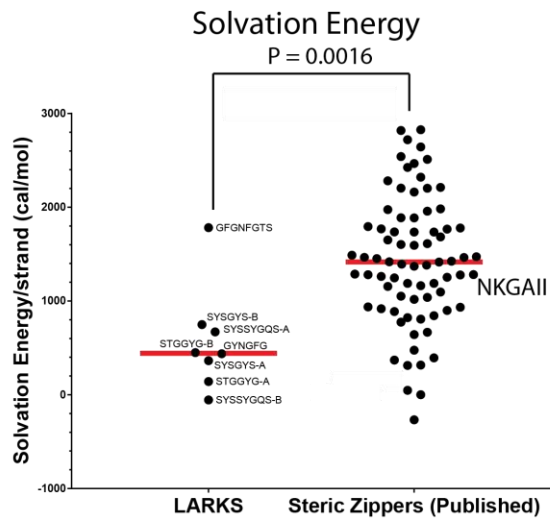
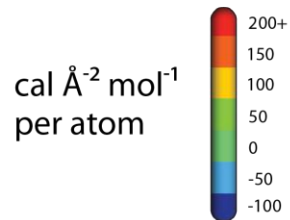
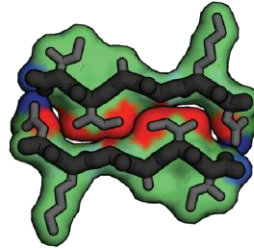


Fig. S6. Solvation energies for LARKS and steric zippers. We calculated solvation energies for the process of separating (exposing to solvent) two mated sheets. Values are in calories per mol per strand for the buried surface area. Plot compares the solvation energy for five LARKS published in this work and 76 published steric zipper structures. Steric zippers have significantly more stable interfaces as judged by these energy values ($P = 0.0016$ by student's t-test). Horizontal red bars indicate medians: 443 for LARKS and 1417 for steric zippers.

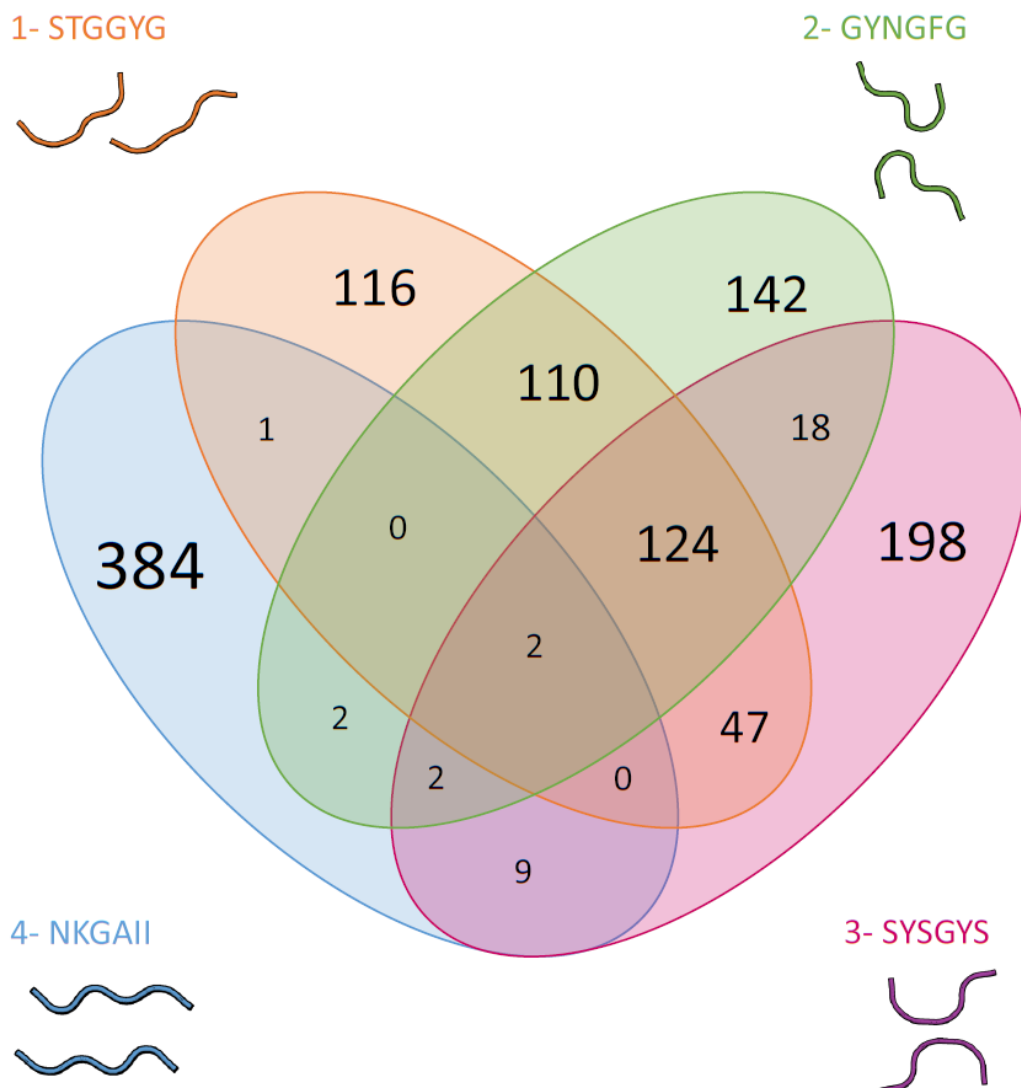


Fig. S7. 3D profiling on three different LARKS backbones largely identify the same proteins as rich in LARKS. The top 400 proteins rich in LARKS as judged by threading the human proteome on backbones: 1) STGGYG, 2) GYNGFG, 3) SYSGYS, or 4) NKGAI – a steric zipper for reference. The 3 LARKS backbones identify the same 126 (=124+2) proteins within the top 400 proteins enriched in LARKS, indicating that threading results from the 3 backbones largely agree. 384 proteins are identified only by the steric zipper backbone, indicating that the steric zippers are distinct structural motifs from LARKS.

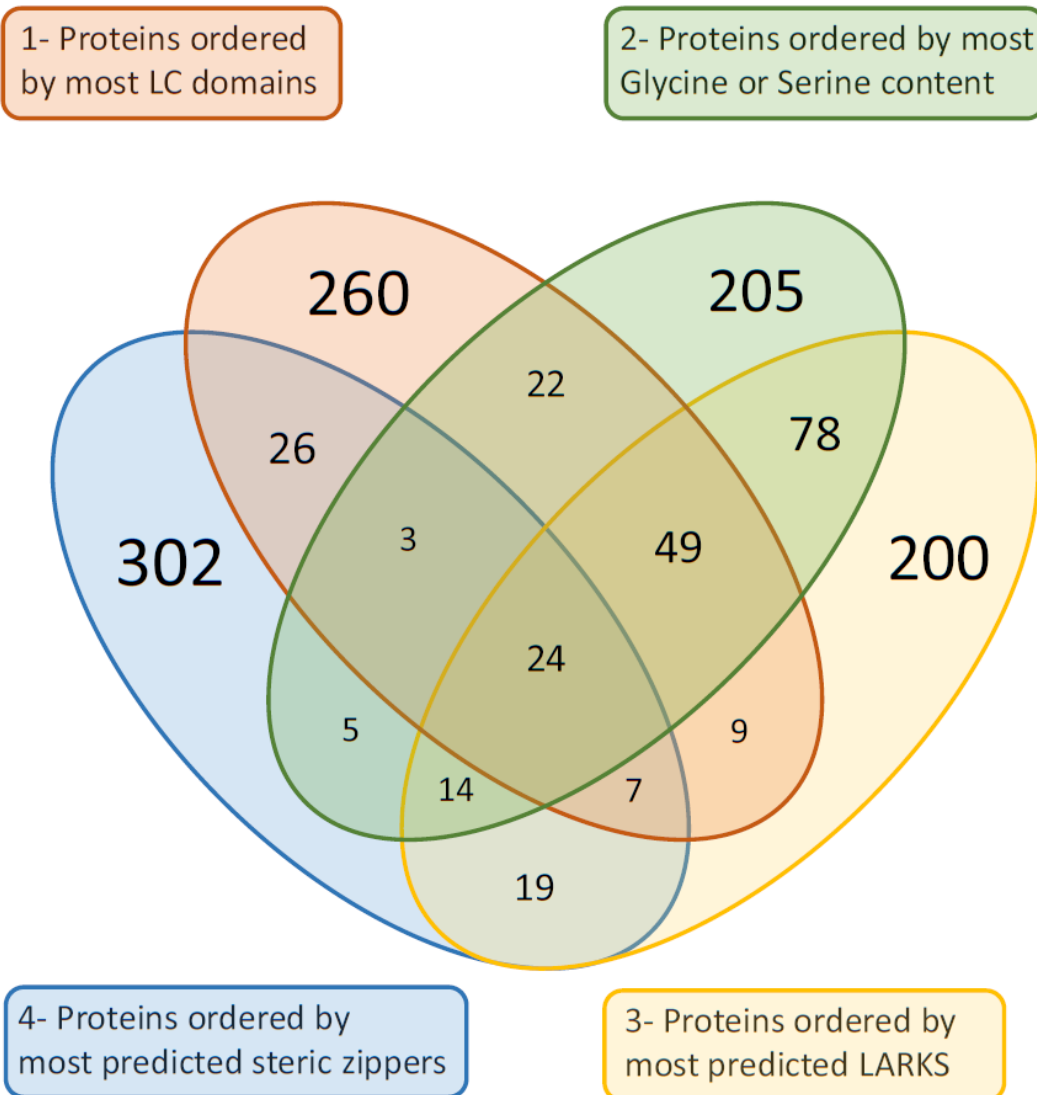


Fig. S8. 3D profiling identifies a unique set of LARKS-containing proteins, which are not simply LC-containing proteins or serine and glycine containing proteins. We compared the identities of the top 400 proteins enriched in predicted LARKS (yellow oval) to the top 400 proteins by fraction of residues in steric zippers (blue oval), in LC domains (orange oval), or by serine/glycine sequence bias - some of the most over-represented amino acids in LC domains in humans (green oval). The lack of large overlap between the circles indicates that that 3D profiling identifies a relatively unique set of 400 proteins, which are not selected by LC domains, by sequences rich in serine/glycine, or rich in steric zippers. In other words, 3D profiling selects proteins by 3D structure and not by sequence.

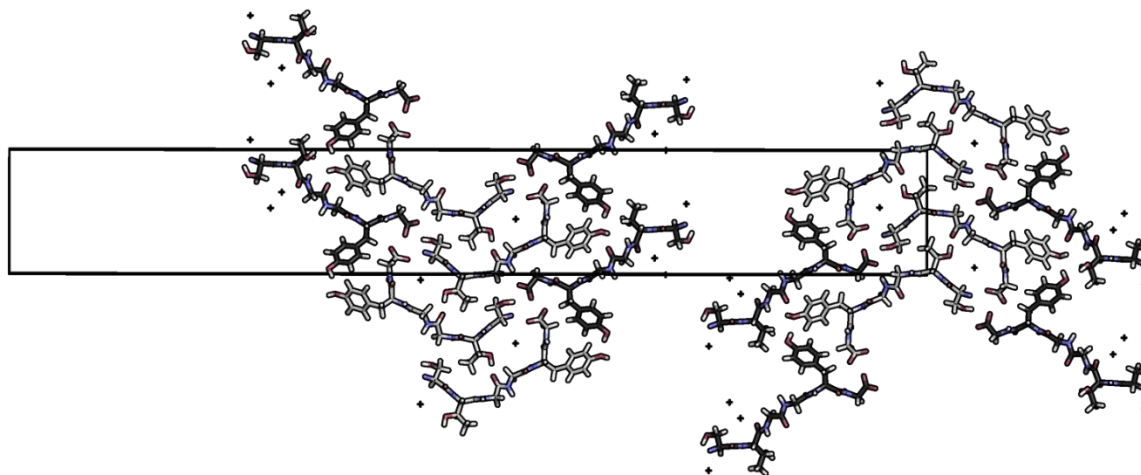


Fig. S9. Crystal structure of STGGYG determined by microED reveals an asymmetric unit with two segments of STGGYG in different conformations. The unit cell is outlined by the black box. Nitrogen atoms are blue; oxygens are red; waters are represented by small crosses; carbon atoms of STGGYGa are light gray; and carbons of STGGYGb are colored dark gray. Two paired interfaces between STGGYG segments exist in the crystal structure: an interface between STGGYGa and STGGYGa, and an interface between STGGYGa and STGGYGb. We used the interface between STGGYGa and STGGYGb as a LARKS in our 3D profiling because it has the more extensive and hydrophobic interface.

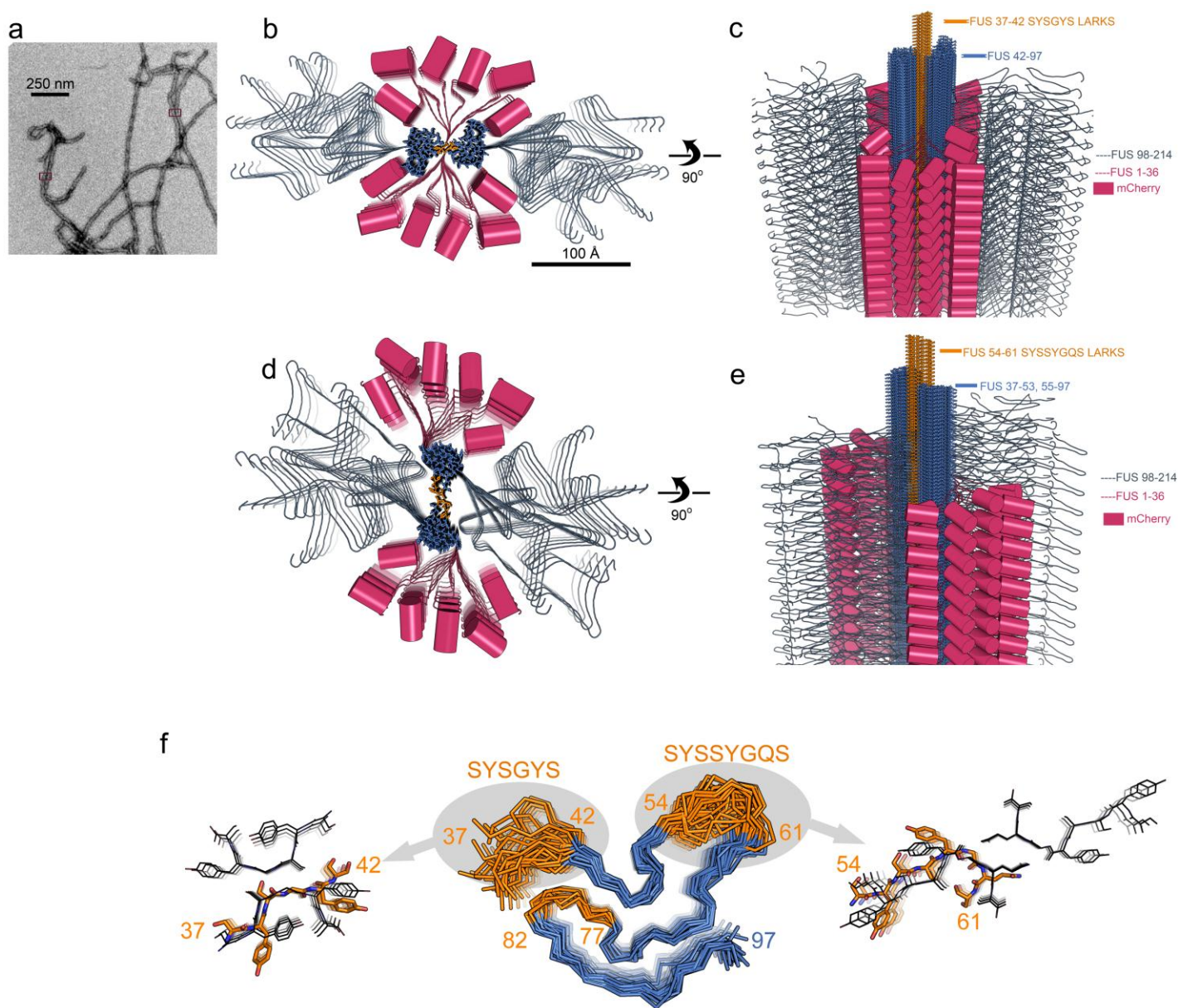


Fig. S10. Adhesive connections among a network of mCherry-FUS β -sheets, modeled as LARKS. (a) Electron micrograph of a filamentous network of mCherry-FUS molecules in a hydrogel, as in Fig S1. The boxes highlight two network connections that we have modeled as distinct LARKS segments. (b and c) An atomic model for one of the adhesive connections between a pair of filamentous kinked β -sheets, viewed (b) down the hydrogen-bonding axis and (c) perpendicular to it. The LARKS motif at the center of the connection is adapted from the SYSGYS crystal structure (orange sticks), superimposed on model 7 from FUS 37-97 determined by Murray et al (blue sticks). The remaining C-terminal 117 residues of FUS LC domain are modeled as unstructured strands (blue ribbon) and mCherry protomers (cherry barrels) are connected to the N-term of FUS through disordered linkers (cherry ribbon). The pairing of sheets is weak and extends only about 100-200 nm, before splitting apart. (d and e) A second network connection, modeled as an adaption of the SYSSYGQS structure (orange sticks), superimposed on model 11 from FUS 37-97 determined by Murray et al (blue sticks). Because LCDs may contain multiple LARKS, protein networks assemble a gel (47). (f) FUS 37-97 NMR structure represented as an ensemble of 20 C-alpha traces (PDB ID code 5W3N). Orange colored segments correspond to those crystallized in this study. The LARKS crystal structures (black lines) are overlaid on individual

members of the ensemble to which they are the most structurally similar (orange sticks). The RMS deviations for all-atom superimpositions are 3.8 and 2.9 Å for SYSGYS and SYSSYGQS, respectively. These overlays were used to model the pairing of filamentous kinked sheets shown in panels b-e. The sheet-sheet pairings in LARKS do not require lengthy sequence motifs, rather just short segments able to kink in complementary shapes. This lack of requirement for sequence complexity suggests that LARKS may not be limited to homotypic interfaces, but also may include heterotypic interfaces composed of segments of FUS distant in linear sequence, or between FUS and other proteins that also contain low complexity domains. Such heterotypic LARKS might provide the adhesive interactions that enable the copolymerization of distinct low complexity domains (such as hnRNPA2, RBM3, FUS, and CIRBP), as observed in Kato et al., 2012, Figs 7D and S5.

Peptide	SYSGYS	SYSSYGQS	STGGYG	GYNGFG	GFGNFGTS
Protein	FUS	FUS	FUS	hnRNPA1	Nup98
Crystal size (in μm)	150 x 6 x 6	200 x 8 x 8	1 x 0.2 x 0.2	200 x 7 x 7	150 x 5 x 5
Data Collection					
Diffraction type	X-ray	X-ray	Electron	X-ray	Electron
Beamline	APS 24-ID-E	APS 24-ID-E	JFRC TF20	APS 24-ID-E	JFRC TF20
Space group	P 1 21 1	C 1 2 1	P 21 21 21	P 21 21 21	P1
Resolution (\AA)	1.1	1.1	1.1	1.1	0.9
Unit cell dimensions: a, b, c (\AA)	17.57, 4.82, 18.28	49.58, 4.79, 21.81	13.79, 4.93, 101.9	16.61, 4.77, 40.97	4.79, 18.24, 26.44
Unit cell angles: $\alpha, \beta, \gamma(^{\circ})$	90.00, 91.03, 90.00	90.00, 95.46, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	93.14, 92.73, 97.15
molecules per asymmetric unit	1	1	2	1	2
measured reflections	5518	18360	23,271	32383	31674
unique reflections	1176	2166	3220	1584	5800
completeness (%)	82.0 (28.4)	92.8 (83.3)	95.5(74.9)	81.0 (36.5)	88.0(47.8)
Redundancy	4.7 (2.2)	8.5 (4.4)	7.2(3.9)	25.9 (6.8)	5.5(2.8)
Rmerge	0.179 (0.468)	0.135 (0.547)	0.250(0.553)	0.130 (0.250)	0.289(0.588)
I/s	7.8 (2.4)	15.0 (4.0)	4.15(1.5)	21.1 (8.7)	3.51(1.12)
CC1/2	0.989 (0.497)	0.995 (0.800)	0.985(0.729)	0.998 (0.962)	0.976(0.546)
Refinement					
Phasing method	MR	DM	DM	DM	DM
Rwork	0.148	13.7	0.231	0.061	0.226
Rfree	0.176	14.8	0.251	0.083	0.264
RMSD bond length (\AA)	0.03	0.004	0.01	0.018	0.014
RMSD angle ($^{\circ}$)	2.27	0.755	0.85	2.22	1.16
Number of peptide atoms	94	124	134	44	112
Number of solvent atoms	2	24	10	4	2
Average B factor of peptide (\AA^2)	4.2	8.3	4.5	1.4	8.0
Average B factor of solvent (\AA^2)	6	20.1	16.6	11.7	14.4
PDB ID code	6BWZ	6BWV	6BZP	6BXX	6BZM

*MR = Molecular Replacement ; DM = Direct Methods

Table S1. Crystallographic statistics. Statistics about data collection and structure determination for new structures presented in this work.
MR = Molecular Replacement; DM = Direct Methods.

GO term ID	GO term description	log10 p-value
GO:0051028	mRNA transport	-3.875
GO:0006397	mRNA processing	-4.176
GO:0030855	epithelial cell differentiation	-4.176
GO:0071765	nuclear inner membrane organization	-1.689
GO:0034063	stress granule assembly	-1.860
GO:0006403	RNA localization	-3.398
GO:0007004	telomere maintenance via telomerase	-2.257
GO:0016074	snoRNA metabolic process	-2.745

Table S2. Selected enriched GO terms in the top 400 proteins rich in LARKS. The top 400 proteins rich in LARKS were analyzed for enriched GO terms. 496 GO terms were found to have P values below log10 -1.301 (equivalent to a p-value of 0.05). The first column is GO ID, the second column gives the name of the GO term. The P-value is the confidence of enrichment determined by standard bootstrapping methods.

Table S3. Top 400 proteins rich in LARKS. Column one has the UniProt accession codes, followed by protein name, followed by number of LARKS.

References

1. S. Hennig, G. Kong, T. Mannen, A. Sadowska, S. Kobelke, A. Blythe, G. J. Knott, K. S. Iyer, D. Ho, E. A. Newcombe, K. Hosoki, N. Goshima, T. Kawaguchi, D. Hatters, L. Trinkle-Mulcahy, T. Hirose, C. S. Bond, A. H. Fox, Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015). doi:10.1083/jcb.201504117 [Medline](#)
2. A. Aguzzi, M. Altmeyer, Phase separation: Linking cellular compartmentalization to disease. *Trends Cell Biol.* **26**, 547–558 (2016). doi:10.1016/j.tcb.2016.03.004 [Medline](#)
3. 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). doi:10.1038/nrm.2017.7 [Medline](#)
4. P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. T. Nixon, M. K. Rosen, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012). doi:10.1038/nature10879 [Medline](#)
5. 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-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012). doi:10.1016/j.cell.2012.04.017 [Medline](#)
6. S. Elbaum-Garfinkle, C. P. Brangwynne, Liquids, fibers, and gels: The many phases of neurodegeneration. *Dev. Cell* **35**, 531–532 (2015). doi:10.1016/j.devcel.2015.11.014 [Medline](#)
7. Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015). doi:10.1016/j.molcel.2015.08.018 [Medline](#)
8. A. F. Harrison, J. Shorter, RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474**, 1417–1438 (2017). doi:10.1042/BCJ20160499 [Medline](#)
9. 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. Schneider, C. Holt, M. Vendruscolo, C. F. Kaminski, P. St George-Hyslop, ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678–690 (2015). doi:10.1016/j.neuron.2015.10.030 [Medline](#)
10. A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoyanov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakowski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, S. Alberti, A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015). doi:10.1016/j.cell.2015.07.047 [Medline](#)

11. A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, J. P. Taylor, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015). doi:10.1016/j.cell.2015.09.015 [Medline](#)
12. S. Xiang, M. Kato, L. C. Wu, Y. Lin, M. Ding, Y. Zhang, Y. Yu, S. L. McKnight, The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**, 829–839 (2015). doi:10.1016/j.cell.2015.10.040 [Medline](#)
13. B. Kumari, R. Kumar, M. Kumar, Low complexity and disordered regions of proteins have different structural and amino acid preferences. *Mol. Biosyst.* **11**, 585–594 (2015). doi:10.1039/C4MB00425F [Medline](#)
14. J. C. Wootton, Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285 (1994). doi:10.1016/0097-8485(94)85023-2 [Medline](#)
15. J. C. Schwartz, T. R. Cech, R. R. Parker, Biochemical properties and biological functions of FET proteins. *Annu. Rev. Biochem.* **84**, 355–379 (2015). doi:10.1146/annurev-biochem-060614-034325 [Medline](#)
16. R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riek, R. Grothe, D. Eisenberg, Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005). doi:10.1038/nature03680 [Medline](#)
17. M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riek, D. Eisenberg, Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007). doi:10.1038/nature05695 [Medline](#)
18. D. T. Murray, M. Kato, Y. Lin, K. R. Thurber, I. Hung, S. L. McKnight, R. Tycko, Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017). doi:10.1016/j.cell.2017.08.048 [Medline](#)
19. M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, Estimates of the ab initio limit for pi-pi interactions: The benzene dimer. *J. Am. Chem. Soc.* **124**, 10887–10893 (2002). doi:10.1021/ja025896h [Medline](#)
20. G. B. McGaughey, M. Gagné, A. K. Rappé, π -Stacking interactions: Alive and well in proteins. *J. Biol. Chem.* **273**, 15458–15463 (1998). doi:10.1074/jbc.273.25.15458 [Medline](#)
21. E. Arunan, H. S. Gutowsky, The rotational spectrum, structure and dynamics of a benzene dimer. *J. Chem. Phys.* **98**, 4294 (1992).
22. D. Eisenberg, A. D. McLachlan, Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986). doi:10.1038/319199a0 [Medline](#)
23. D. E. Eisenberg, M. Wesson, M. Yamashita, Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.* **29A**, 217–221 (1989).

24. J. U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170 (1991). [doi:10.1126/science.1853201](https://doi.org/10.1126/science.1853201) [Medline](#)
25. L. Goldschmidt, P. K. Teng, R. Riek, D. Eisenberg, Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3487–3492 (2010). [doi:10.1073/pnas.0915166107](https://doi.org/10.1073/pnas.0915166107) [Medline](#)
26. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011). [doi:10.1016/B978-0-12-381270-4.00019-6](https://doi.org/10.1016/B978-0-12-381270-4.00019-6) [Medline](#)
27. S. Frey, R. P. Richter, D. Görlich, FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006). [doi:10.1126/science.1132516](https://doi.org/10.1126/science.1132516) [Medline](#)
28. R. Windoffer, S. Wöll, P. Strnad, R. E. Leube, Identification of novel principles of keratin filament network turnover in living cells. *Mol. Biol. Cell* **15**, 2436–2448 (2004). [doi:10.1091/mbc.E03-09-0707](https://doi.org/10.1091/mbc.E03-09-0707) [Medline](#)
29. C. Ader, S. Frey, W. Maas, H. B. Schmidt, D. Görlich, M. Baldus, Amyloid-like interactions within nucleoporin FG hydrogels. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6281–6285 (2010). [doi:10.1073/pnas.0910163107](https://doi.org/10.1073/pnas.0910163107) [Medline](#)
30. K. L. Sim, T. P. Creamer, Abundance and distributions of eukaryote protein simple sequences. *Mol. Cell. Proteomics MCP* **1**, 983–995 (2002). [doi:10.1074/mcp.M200032-MCP200](https://doi.org/10.1074/mcp.M200032-MCP200) [Medline](#)
31. J. A. Toretsky, P. E. Wright, Assemblages: Functional units formed by cellular phase separation. *J. Cell Biol.* **206**, 579–588 (2014). [doi:10.1083/jcb.201404124](https://doi.org/10.1083/jcb.201404124) [Medline](#)
32. A. Coletta, J. W. Pinney, D. Y. Solís, J. Marsh, S. R. Pettifer, T. K. Attwood, Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, 43 (2010). [doi:10.1186/1752-0509-4-43](https://doi.org/10.1186/1752-0509-4-43) [Medline](#)
33. V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008). [doi:10.1146/annurev.biophys.37.032807.125924](https://doi.org/10.1146/annurev.biophys.37.032807.125924) [Medline](#)
34. I. Kwon, M. Kato, S. Xiang, L. Wu, P. Theodoropoulos, H. Mirzaei, T. Han, S. Xie, J. L. Corden, S. L. McKnight, Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013). [doi:10.1016/j.cell.2013.10.033](https://doi.org/10.1016/j.cell.2013.10.033) [Medline](#)
35. D. S. Eisenberg, M. R. Sawaya, Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* **86**, 69–95 (2017). [doi:10.1146/annurev-biochem-061516-045104](https://doi.org/10.1146/annurev-biochem-061516-045104) [Medline](#)

36. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi:10.1107/S0021889807021206 [Medline](#)
37. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997). doi:10.1107/S0907444996012255 [Medline](#)
38. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008). doi:10.1107/S0108767307043930 [Medline](#)
39. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). doi:10.1107/S0907444909047337 [Medline](#)
40. J. Hattne, F. E. Reyes, B. L. Nannenga, D. Shi, M. J. de la Cruz, A. G. W. Leslie, T. Gonen, MicroED data collection and processing. *Acta Crystallogr. Sect. A Found. Adv.* **71**, 353–360 (2015). doi:10.1107/S2053273315010669 [Medline](#)
41. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). doi:10.1107/S0907444912001308 [Medline](#)
42. E. Blanc, P. Roversi, C. Vonnrhein, C. Flensburg, S. M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221 (2004). doi:10.1107/S0907444904016427 [Medline](#)
43. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi:10.1107/S0907444910007493 [Medline](#)
44. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011). doi:10.1107/S0907444910045749 [Medline](#)
45. The UniProt Consortium, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017). doi:10.1093/nar/gkw1099 [Medline](#)
46. Gene Ontology Consortium, Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015). doi:10.1093/nar/gku1179 [Medline](#)
47. R. Halfmann, A glass menagerie of low complexity sequences. *Curr. Opin. Struct. Biol.* **38**, 18–25 (2016). doi:10.1016/j.sbi.2016.05.002 [Medline](#)