

PROTEIN BIOCHEMISTRY

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

Michael P. Hughes,¹ Michael R. Sawaya,¹ David R. Boyer,¹ Lukasz Goldschmidt,¹ Jose A. Rodriguez,² Duilio Cascio,¹ Lisa Chong,¹ Tamir Gonen,³ David S. Eisenberg^{1*}

Subcellular membraneless assemblies are a reinvigorated area of study in biology, with spirited scientific discussions on the forces between the low-complexity protein domains within these assemblies. To illuminate these forces, we determined the atomic structures of five segments from protein low-complexity domains associated with membraneless assemblies. Their common structural feature is the stacking of segments into kinked β sheets that pair into protofilaments. Unlike steric zippers of amyloid fibrils, the kinked sheets interact weakly through polar atoms and aromatic side chains. By computationally threading the human proteome on our kinked structures, we identified hundreds of low-complexity segments potentially capable of forming such interactions. These segments are found in proteins as diverse as RNA binders, nuclear pore proteins, and keratins, which are known to form networks and localize to membraneless assemblies.

Membraneless organelles, such as P bodies, nuclear paraspeckles, and stress granules (SGs), form and redissolve in mammalian cells in response to stimuli (1, 2). Such phase separation is a property of macromolecules that are capable of multivalent interactions with each other, yielding a liquid phase having ~100 times the concentration of the macromolecule compared to the bulk liquid (3, 4). This type of phase separation is often seen with proteins that bind nucleic acids and contain low-complexity domains (LCDs) (1, 2, 5–8). For example, the SG-associated proteins hnRNPA1, hnRNPA2, and FUS undergo liquid-liquid phase separation (9–12), and they contain LCDs that transition into reversible semisolid phase hydrogels over time or at higher protein concentration (1, 5, 9). LCDs are common in the human proteome; they are largely intrinsically disordered (13), and dramatically underrepresented in the Protein Data Bank (PDB) of known three-dimensional (3D) structures (14).

Electron microscopy reveals that such hydrogels contain protein fibrils, and x-ray diffraction of the hydrogel yields a cross- β pattern (fig. S1, C to E) (5, 15) reminiscent of amyloid. However, whereas the fibrils found in FUS hydrogels are heat and SDS-sensitive (5), amyloid fibrils resist denaturation by SDS and boiling. The spines of amyloid fibrils contain pairs of closely mating β sheets along the fibril axis. Residue side chains tightly

interdigitate with side chains of the opposing β sheet to form a dry interface called a steric zipper, as seen in the structure of NKGAI1 from amyloid beta (A β) (Fig. 1A) (16, 17). The steric zipper explains the extraordinary stability of some pathogenic amyloid. Apparently, the relatively labile multivalent interactions of the hydrogel-forming proteins are different; their atomic-level details are largely unknown, although, importantly, solid-state nuclear magnetic resonance has shown that 57 of the 214 residue LCDs of FUS form an ordered protofilament core, with the remaining residues dynamically disordered (18).

To investigate relatively weak adhesion between LCDs of proteins recruited to SGs, we sought relevant atomic structures. Guided by studies of the LCDs of FUS and RBM14, which show that successive replacement of tyrosine residues by serine lowers their capacity to form hydrogels (1, 5), we scanned the LCD of FUS for tandem sequence motifs of the form [G/S]Y[G/S], finding two such segments: FUS-³⁷SYSGYS⁴² and FUS-⁵⁴SYSSYGQS⁶¹ (fig. S1A). Both segments crystallized as micron-sized needles, and both atomic structures were determined, in addition to structures of three other segments identified by 3D profiling (see below): ²⁴³GYNGFG²⁴⁸ from protein hnRNPA1, ⁷⁷STGGYG⁸² from FUS, and ¹¹⁶GFGNFGTS¹²³ from nup98 (Fig. 1). Confirming the relevance of these structures to adhesion and multivalency of LCDs, a labile hydrogel is formed by a 26-residue synthetic peptide construct linking the three above segments of FUS (Fig. 2). Powder diffraction patterns of all five crystalline segments, this hydrogel, and the FUS-LCD hydrogel suggest that they all share cross- β architecture (figs S2 and S3).

All five segments crystallized as pairs of kinked β sheets (Fig. 1). Each β sheet runs the length of the crystal, formed from the stacking of about 300,000 segments, and all structures show kinks

at either glycines or aromatic residues instead of being extended (fig. S4). The structures share common adhesive features, including hydrogen bonds in-register to an identical segment below it (Fig. 1, B to F, and fig. S5). Aromatic residues predominate, both for intersheet stabilization and intrasheet stabilization. Within sheets, the aromatic side chains stack in an energetically favorable conformation, with the planes of the rings stacked parallel at a separation of 3.4 Å (19–21) (fig. S5). These aromatic “ladders” enhance the stability of each β sheet. The kinks allow close approach of the backbones, providing favorable van der Waals or hydrogen-bond interactions between the sheets (fig. S5). These close interactions are quantified by the structural complementarity (Sc) (Fig. 1), reflecting adhesion between the sheets. However, the kinks prevent side chains from interdigitating across the β -sheet interface so that the kinked interfaces bury smaller surface areas than found in pathogenic amyloid fibrils and presumably have lower binding energies. Because of the distinction of the kinked structures from pathogenic steric zippers, we term them low-complexity aromatic-rich kinked segments (LARKS).

Calculations and experiments support our structural inference that LARKS have smaller binding energies than steric zippers. We estimated energies of separation of the pairs of β sheets in LARKS and steric zippers by applying atomic solvation parameters (22, 23) to our structures: The mean atomic solvation energy for separation of our LARKS interfaces is 567 ± 556 cal/mol/ β strand, whereas it is 1431 ± 685 cal/mol/ β strand for 75 steric zipper structures (fig. S6). These crude estimates suggest that the adhesive energy of one pair of β strands in a LARKS is of the order of thermal energy, so that pairs of β sheets adhere only through multivalent interactions of strands. In contrast, the adhesive energy of one pair of strands in a steric zipper is several times that of thermal energy. Consistent with calculations, the synthetic multi-LARKS construct of Fig. 2 dissolves when gently heated. Thus, paired kinked β sheets of LARKS are less strongly bound than the paired β sheets in amyloid fibrils, yet still produce fibrils with the cross β -diffraction pattern of pathogenic amyloid.

To identify potential LARKS in the human proteome, we used computational 3D profiling, a method that tests the compatibility of query sequences with a template structure (24, 25). Here, we threaded human sequences onto the backbones of SYSGYS, GYNGFG, and STGGYG, placed and optimally repacked side chains, and then evaluated the Rosetta energy (Fig. 3A) (26). We advanced the threading by one residue and repeated the procedure until the end of the query sequence was reached. This 3D profiling predicted that nucleoporin proteins are enriched in LARKS (Fig. 3C). Our confidence in this prediction was bolstered by the success of earlier predictions that GYNGFG and STGGYG could form LARKS, based on threading with only the SYSGYS template. Here, again, we were able to validate our profiling algorithm by determining the structure

¹Department of Biological Chemistry and Department of Chemistry and Biochemistry, University of California Los Angeles (UCLA), Howard Hughes Medical Institute (HHMI), UCLA-Department of Energy (DOE) Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA. ²Department of Chemistry and Biochemistry, UCLA, UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA. ³HHMI and Department of Physiology, UCLA, Los Angeles, CA 90095, USA.

*Corresponding author. Email: david@mbl.ucla.edu

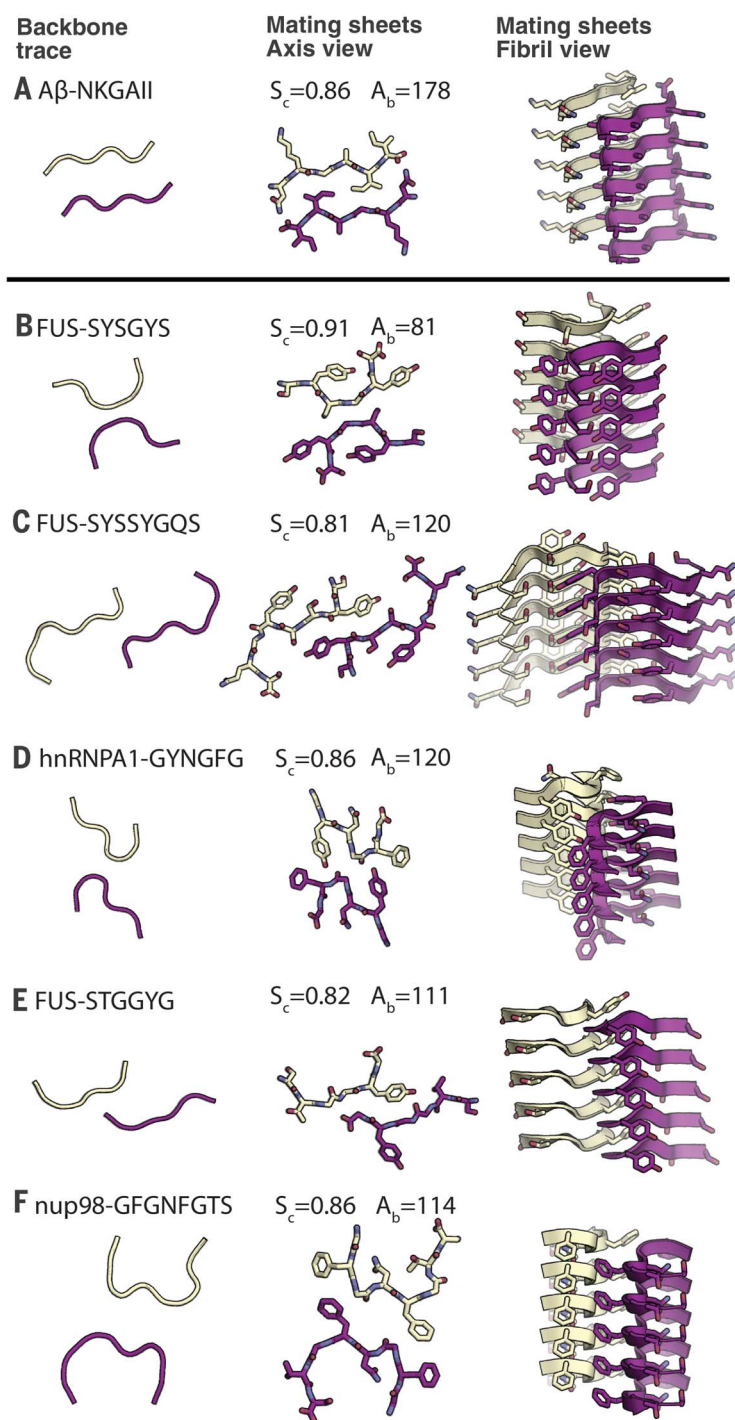


Fig. 1. Structures of LARKS compared to a steric zipper. (A) Steric zipper. **(B)** to **(F)** Structures of LARKS. All structures are composed of two mating β sheets, one purple and the other yellow. The left-hand column shows the trace of the backbones of mating sheets to highlight kinks in the backbones of LARKS and the pleating of the classical β sheets in steric zippers. The second column shows the atomic structures of mating sheets viewed down the fibril axes. The third column shows cartoons of the mating β sheets viewed nearly perpendicular to the fibril axes. Each interface is characterized by the shape complementarity score ($S_c = 1.0$ for perfect complementarity) and buried solvent-accessible surface area (A_b) in \AA^2 between the mated sheets. Carbon atoms are colored purple or yellow, nitrogen is blue, and oxygen is red. Five layers of β sheets are shown of the hundreds of thousands in the crystals. The kinked structures of LARKS are rare among mating β sheets; dozens of other paired β sheets form steric zippers (35).

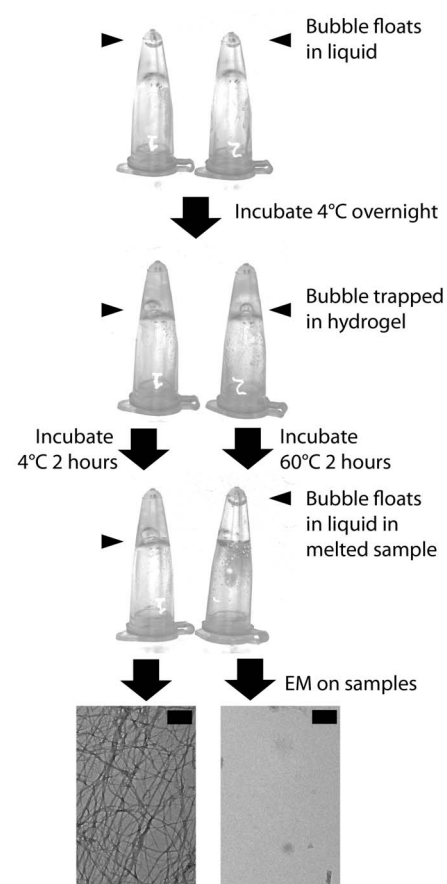


Fig. 2. Synthetic LARKS construct forms a labile hydrogel. A synthetic LARKS construct with the sequence SYSGYSGDTSYS-SYGQSNPSTGGYG (underlined sequences correspond to LARKS) forms a labile hydrogel when dissolved in water at 50 mg/ml and left overnight at 4°C. The hydrogel melts upon heating the sample to 60°C for 2 hours. A bubble (blue arrow) was introduced to the sample to show the difference between the liquid state (bubble rises) and hydrogel state (bubble does not rise). Electron microscopy confirms that fibrils were indeed melted. The hydrogel-forming property of this triple-LARKS sequence suggests that it is the multiple LARKS found in many LCDs that endow their unusual property of forming hydrogels. Scale bars, 200 nm. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

of GFGNFGTS from the porin nup98, confirming LARKS architecture (Fig. 1F) and providing evidence that LARKS are present in a different type of membraneless organelle (27).

Analyzing the nonredundant human proteome of 20,120 sequences from UniProt, we found 5867 proteins with LCDs. Of these, 2500 proteins contain at least one LARKS and 1725 proteins contain two or more LARKS and thus are able to

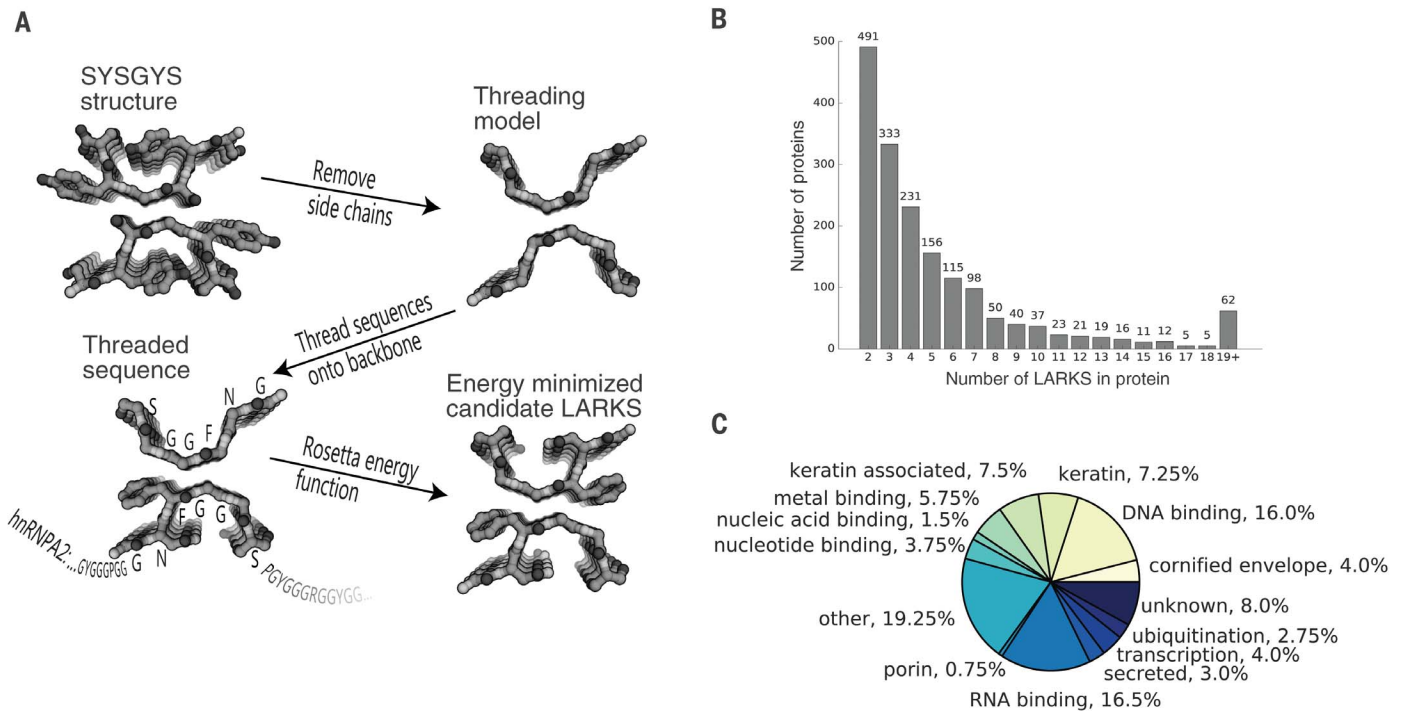


Fig. 3. Three-dimensional profiling to identify LARKS in LCDs of human proteins. (A) Side chains are removed from the backbones of one of our atomic structures of a LARKS. Then the sequence of interest (hnRNPA2 shown) is threaded through the six-residue template by placing the query side chains on the template backbone. Side chains are repacked and a Rosetta energy function is used to estimate whether the

structure is favorable for the threaded sequence. The sequence then advances through the template by one-residue increments, producing successive models. (B) The frequency of the number of LARKS in 1725 human proteins predicted to house at least two LARKS. Proteins having two or more LARKS are predicted to have the capacity to form networks and possibly gels. (C) The annotated functions of the 400 proteins with the most predicted LARKS.

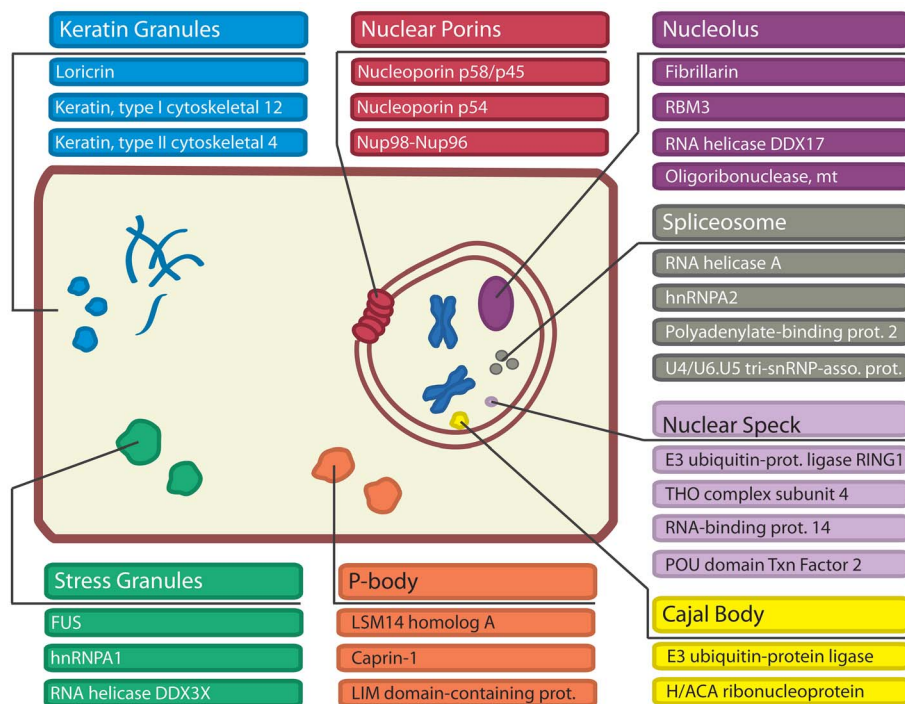


Fig. 4. Functions of proteins among the 400 proteins most enriched in LARKS and dynamic intracellular bodies of which they are known to be a part.

form multivalent interactions and hence protein networks and gels. Hundreds of proteins house three or more LARKS (Fig. 3B). The 400 human LCDs most enriched in LARKS average 14 LARKS, with a median of 10 LARKS.

We assigned cellular function to these 400 proteins based on their UniProt annotations (Fig. 3C): 16% are DNA binding, 17% are RNA binding, and 4% are nucleotide binding, consistent with reports of nucleotide binding proteins in membraneless organelles (2, 8). Keratins (5%), keratin-associated (9%), and cornified envelope proteins (4%) are also enriched in LARKS. The finding of keratins is consistent with experiments (28) showing that keratin granules are trafficked to the cell cortex, where they merge and eventually mature into filaments. Also rich in LARKS are proteins found in ribonucleoprotein particles such as the spliceosome or nucleolus (Fig. 4). Nucleoporins including nup54 and nup98 with FG repeats are enriched in predicted LARKS, and purified FG repeats form a hydrogel (27, 29). The possibility that the FG repeats of nucleoporins may form LARKS in the diffusion barrier of the pore is supported by our structure of GFGNFGTS from nup98. We assigned additional cellular functions to these 400 proteins from their associated gene ontology (GO) terms. We found GO terms enriched in the human proteome

for RNA transport, processing localization, SG assembly, and epithelial cell differentiation due to the numerous keratins enriched in LARKS. Therefore, we propose 3D profiling for LARKS as a tool to identify proteins that may form networks and gels by multivalent interactions and may participate in membraneless organelles (fig. S10).

In conclusion, the prevalence of LCDs within eukaryotic proteomes has long been recognized (30), but the role of these domains has not been fully defined. Previous discoveries include the following: LCDs can “functionally aggregate” (31); proteins with LCDs typically form more protein-protein interactions (32, 33); and proteins can interact homotypically and heterotypically through LCDs (1, 5, 34). Our atomic structures support the hypothesis that LCDs have the capacity to form gel-like networks. LARKS possess three properties that are consistent with their functioning as adhesive elements in protein gels formed from LCDs: (i) high aqueous solubility contributed by their high proportion of hydrophilic residues: serine, glutamine, and asparagine; (ii) flexibility ensured by their high glycine content; and (iii) multiple interaction motifs per chain (Fig. 3B), endowing them with multivalency, enabling them to entangle, forming networks as found in gels (Fig. 2). That each LARKS provides adhesion only comparable to thermal energy suggests that numerous LARKS must cooperate in gel formation and that the interactions must be concentration dependent and may be transient. If steric zippers act as molecular glue, then LARKS in LCDs act as Velcro. These properties are compatible with the hypo-

thesis that LARKS are a protein interaction motif that provides adhesion of LCDs in protein gels and in membraneless assemblies (fig S10).

REFERENCES AND NOTES

1. S. Hennig *et al.*, *J. Cell Biol.* **210**, 529–539 (2015).
2. A. Aguzzi, M. Altmeyer, *Trends Cell Biol.* **26**, 547–558 (2016).
3. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
4. P. Li *et al.*, *Nature* **483**, 336–340 (2012).
5. M. Kato *et al.*, *Cell* **149**, 753–767 (2012).
6. S. Elbaum-Garfinkle, C. P. Brangwynne, *Dev. Cell* **35**, 531–532 (2015).
7. Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, *Mol. Cell* **60**, 208–219 (2015).
8. A. F. Harrison, J. Shorter, *Biochem. J.* **474**, 1417–1438 (2017).
9. T. Murakami *et al.*, *Neuron* **88**, 678–690 (2015).
10. A. Patel *et al.*, *Cell* **162**, 1066–1077 (2015).
11. A. Mollieux *et al.*, *Cell* **163**, 123–133 (2015).
12. S. Xiang *et al.*, *Cell* **163**, 829–839 (2015).
13. B. Kumari, R. Kumar, M. Kumar, *Mol. Biosyst.* **11**, 585–594 (2015).
14. J. C. Wootton, *Comput. Chem.* **18**, 269–285 (1994).
15. J. C. Schwartz, T. R. Cech, R. R. Parker, *Annu. Rev. Biochem.* **84**, 355–379 (2015).
16. R. Nelson *et al.*, *Nature* **435**, 773–778 (2005).
17. M. R. Sawaya *et al.*, *Nature* **447**, 453–457 (2007).
18. D. T. Murray *et al.*, *Cell* **171**, 615–627.e16 (2017).
19. M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, *J. Am. Chem. Soc.* **124**, 10887–10893 (2002).
20. G. B. McGaughey, M. Gagné, A. K. Rappé, *J. Biol. Chem.* **273**, 15458–15463 (1998).
21. E. Arunan, H. S. Gutowsky, *J. Chem. Phys.* **98**, 4294 (1992).
22. D. Eisenberg, A. D. McLachlan, *Nature* **319**, 199–203 (1986).
23. D. E. Eisenberg, M. Wesson, M. Yamashita, *Chem. Scr.* **29A**, 217–221 (1989).
24. J. U. Bowie, R. Lüthy, D. Eisenberg, *Science* **253**, 164–170 (1991).
25. L. Goldschmidt, P. K. Teng, R. Riek, D. Eisenberg, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3487–3492 (2010).
26. A. Leaver-Fay *et al.*, *Methods Enzymol.* **487**, 545–574 (2011).
27. S. Frey, R. P. Richter, D. Görlich, *Science* **314**, 815–817 (2006).
28. R. Windoffer, S. Wöll, P. Strnad, R. E. Leube, *Mol. Biol. Cell* **15**, 2436–2448 (2004).
29. C. Ader *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6281–6285 (2010).
30. K. L. Sim, T. P. Creamer, *Mol. Cell. Proteomics MCP* **1**, 983–995 (2002).
31. J. A. Toretsky, P. E. Wright, *J. Cell Biol.* **206**, 579–588 (2014).
32. A. Coletta *et al.*, *BMC Syst. Biol.* **4**, 43 (2010).
33. V. N. Uversky, C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biophys.* **37**, 215–246 (2008).
34. I. Kwon *et al.*, *Cell* **155**, 1049–1060 (2013).
35. D. S. Eisenberg, M. R. Sawaya, *Annu. Rev. Biochem.* **86**, 69–95 (2017).

ACKNOWLEDGMENTS

D.S.E. is on the scientific advisory board and holds equity in ADRx, Inc. X-ray diffraction data were collected at the Northeastern Collaborative Access Team beamline 24-ID-E, which is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. We thank NSF MCB-1616265, NIH AG-054022, DOE, and HHMI for support. Atomic coordinates and structure factors have been deposited in the PDB with the following accession codes: SYSGYS (6BWZ), SYSSYGQS (6BXV), STGGYG (6BZP), GYNGFG (6BXX), and GFGNFGTS (6BZM).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6376/698/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S3
References (36–47)

13 May 2017; resubmitted 15 November 2017
Accepted 4 January 2018
10.1126/science.aan6398

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 and David S. Eisenberg

Science **359** (6376), 698-701.
DOI: 10.1126/science.aan6398

Interactions of LARKS protein domains

More than 1500 human proteins contain long, disordered stretches of "low complexity"—strings of just a few of the 20 common amino acids. The functions of these low-complexity domains have been unclear. Hughes *et al.* present atomic-resolution structures that suggest that short segments of two such domains can bind weakly to each other by forming a pair of kinked β -sheets. Because aromatic amino acid side chains stabilize these interactions, the interacting motifs are termed LARKS, for low-complexity, aromatic-rich, kinked segments. Numerous proteins associated with membraneless organelles of biological cells contain low-complexity domains housing multiple LARKS.

Science, this issue p. 698

ARTICLE TOOLS

<http://science.sciencemag.org/content/359/6376/698>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/02/07/359.6376.698.DC1>

REFERENCES

This article cites 46 articles, 10 of which you can access for free
<http://science.sciencemag.org/content/359/6376/698#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)