Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior

Benjamin S. Schuster[a,b,1], Gregory L. Dignon[c,d,1], Wai Shing Tang[e], Fleurie M. Kelley[b], Aishwarya Kanchi Ranganath[b], Craig N. Jahnke[f], Alison G. Simpkins[f], Roshan Mammen Regy[c], Daniel A. Hammer[a,f], Matthew C. Good[a,g], and Jeetain Mittal[c,2]

[a]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104; [b]Department of Chemical and Biochemical Engineering, Rutgers University, Piscataway, NJ 08854; [c]Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA 18015; [d]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794; [e]Department of Physics, Brown University, Providence, RI 02912; [f]Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104; and [g]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19104

Edited by B. Montgomery Pettitt, University of Texas Medical Branch, Galveston, TX, and accepted by Editorial Board Member Peter J. Rossky March 30, 2020 (received for review January 9, 2020)

Phase separation of intrinsically disordered proteins (IDPs) commonly underlies the formation of membraneless organelles, which compartmentalize molecules intracellularly in the absence of a lipid membrane. Identifying the protein sequence features responsible for IDP phase separation is critical for understanding physiological roles and pathological consequences of biomolecular condensation, as well as for harnessing phase separation for applications in bioinspired materials design. To expand our knowledge of sequence determinants of IDP phase separation, we characterized variants of the intrinsically disordered RGG domain from LAF-1, a model protein involved in phase separation and a key component of P granules. Based on a predictive coarse-grained IDP model, we identified a region of the RGG domain that has high contact probability and is highly conserved between species; deletion of this region significantly disrupts phase separation in vitro and in vivo. We determined the effects of charge patterning on phase behavior through sequence shuffling. We designed sequences with significantly increased phase separation propensity by shuffling the wild-type sequence, which contains well-mixed charged residues, to increase charge segregation. This result indicates the natural sequence is under negative selection to moderate this mode of interaction. We measured the contributions of tyrosine and arginine residues to phase separation experimentally through mutagenesis studies and computationally through direct interrogation of different modes of interaction using all-atom simulations. Finally, we show that despite these sequence perturbations, the RGG-derived condensates remain liquid-like. Together, these studies advance our fundamental understanding of key biophysical principles and sequence features important to phase separation.

liquid–liquid phase separation | membraneless organelles | molecular simulations

Liquid–liquid phase separation (LLPS) of biomolecules is a highly robust and ubiquitous phenomenon in biology, enabling compartmentalization in the absence of delimiting membranes (1). Biomolecular LLPS commonly occurs within the cell, forming compartments that have been termed biomolecular condensates or membraneless organelles (2) and include stress granules (3–5), P granules (1, 6), nucleoli (7), and numerous others (8–13). Most membraneless organelles contain an overrepresentation of proteins with intrinsically disordered and low-complexity regions (14), which are important drivers of phase-separation behavior (15, 16). Therefore, decoding the sequence determinants of intrinsically disordered protein (IDP) phase separation is important for understanding the biochemistry of biomolecular condensates in physiological and pathophysiological conditions. Characterizing the effects of sequence on phase behavior is also important for the field of protein-based materials (17), wherein proteins can be designed to have desired characteristics and programable assembly (18–20), with applications in biotechnology such as drug delivery, cell engineering, and biomimetics (21–24).

Here we investigate a model IDP sequence from LAF-1, which is a member of the DDX3 family of RNA helicases and is a major component of P granules, membraneless organelles involved in germline specification in *Caenorhabditis elegans* embryos (25). LAF-1 contains an N-terminal domain of 168 residues that is intrinsically disordered, followed by a folded helicase domain, and a short disordered prion-like domain at the C terminus (6). The N-terminal domain contains an abundance of glycine and arginine residues, with several occurrences of the motif RGG, and is hereafter referred to as LAF-1 RGG. Importantly, the RGG domain is necessary and sufficient for phase separation (6), although both experimental and computational

Significance

Membraneless organelles are assemblies of highly concentrated biomolecules that form through liquid–liquid phase separation. These assemblies are often enriched in intrinsically disordered proteins (IDPs), which play an important role in driving phase separation. Understanding the sequence-to-phase behavior relationship of these disordered proteins is important for understanding the biochemistry of membraneless organelles, as well as for designing synthetic organelles and biomaterials. In this work, we explore a model protein, the disordered N-terminal domain of LAF-1, and highlight how three key features of the sequence control the protein's propensity to phase-separate. Combining state-of-the-art simulations with experiments, we find that phase behavior of this model IDP is dictated by the presence of a short conserved domain, charge patterning, and arginine–tyrosine interactions.

Author contributions: B.S.S., G.L.D., D.A.H., M.C.G., and J.M. designed research; B.S.S., G.L.D., W.S.T., F.M.K., A.K.R., C.N.J., A.G.S., and R.M.R. performed research; B.S.S., G.L.D., W.S.T., F.M.K., A.K.R., and R.M.R. analyzed data; B.S.S., G.L.D., and J.M. wrote the paper; and B.S.S., D.A.H., M.C.G., and J.M. supervised research.

The authors declare no competing interest.

This article is a PNAS Direct Submission. B.M.P. is a guest editor invited by the Editorial Board.

[1]B.S.S. and G.L.D. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jeetain@lehigh.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2000223117/-/DCSupplemental.

studies have shown that inclusion of the folded domain increases the protein's ability to phase-separate (26, 27).

LAF-1 RGG is an excellent model system for exploring the sequence determinants of protein phase separation because it is believed to be fully disordered (based on circular dichroism spectroscopy) (6), and it contains a sufficient diversity of amino acids to enable different types of interactions (28, 29). The advantage of a fully disordered sequence is that it allows for relatively distributed interactions between all residues, so the relationship between amino acid composition and phase behavior can be more readily ascertained, as compared to proteins with residues buried in folded domains. LAF-1 was one of the first proteins found in biomolecular condensates in vivo and whose phase behavior was mapped in vitro, yet key questions remain about its properties and function (6, 27). Additionally, we have recently designed constructs based on LAF-1 RGG to generate micrometer-sized protein condensates that can respond to specific stimuli and that can selectively compartmentalize cargo proteins, progressing toward the design of synthetic organelles that may be expressed in cells and that are orthogonal to normal cellular function (21). To advance the design of synthetic organelles in the future, we seek to understand how perturbations to the RGG domain sequence may alter phase behavior in a predictable way (18, 30).

In this work, we use simulations and experiments to characterize the sequence-dependent LLPS of the LAF-1 RGG domain, identifying perturbations that result in significant changes to the phase behavior, and we put forward a mechanistic basis for these changes. First, we have identified a small hydrophobic region that exhibits high contact probability in coarse-grained (CG) molecular dynamics simulations and that contains a well-conserved specific binding site for the eukaryotic translation initiation factor 4E (eIF4E) (31). We demonstrate that removal of this region greatly reduces the phase-separation propensity of the RGG domain in silico, in vitro, and in vivo in a eukaryotic model, suggesting that the hydrophobic interactions within this region are also important to LLPS. Second, we show that shuffling the amino acid residues of the RGG sequence to introduce charge patterning can drastically increase phase-separation propensity and that by simultaneously preserving the conserved hydrophobic region we can further increase it. Third, we investigate alterations to amino acid composition by replacing tyrosine with phenylalanine and arginine with lysine; such alterations affect the phase behavior of Fused in Sarcoma (FUS), a commonly studied protein that is involved in LLPS in vivo and mutations to which are implicated in several diseases, including amyotrophic lateral sclerosis (30, 32). We find that tyrosine-to-phenylalanine and arginine-to-lysine mutations both reduce the phase-separation propensity of the LAF-1 RGG domain. We then identify the interaction mechanisms disrupted by these mutations as being hydrogen bonds, cation–π interactions, and sp^2/π interactions, all three of which are present between arginine and tyrosine and may act cooperatively, whereas at least one of these is diminished upon mutation. Importantly, we rule out a previous model based exclusively on arginine–tyrosine interaction, which cannot predict the critical concentration for LAF-1 RGG phase separation. Finally, we show that the RGG-derived condensates remain liquid-like despite these three classes of sequence perturbations, indicating that phase behavior can be tuned independently from material properties. Our combined results elucidate important sequence determinants of IDP phase separation while demonstrating a computationally guided approach for studying phase behavior of biomolecular condensates.

Results

A Short, Conserved, Hydrophobic Region Is Important for LLPS of the RGG Domain. We focused our efforts on the RGG domain of LAF-1, as it is necessary and sufficient to drive phase separation (6), making it an ideal model system to understand the sequence determinants of LLPS. Phase separation of LAF-1 RGG is hypothesized to be driven by several different modes of interaction, including electrostatic, π–π, and cation–π interactions (27). In addition, hydrogen bonds and hydrophobic contacts may play a role in phase separation for sequences containing residues capable of such interactions (16, 33–37). However, it is difficult to characterize these interactions using experimental techniques due to the dynamic nature of the phase-separated proteins and the high spatiotemporal resolution needed to probe the interactions (35).

To provide insight into the sequence determinants of phase separation, we conducted simulations of a condensed assembly of 100 chains of LAF-1 RGG using a transferrable CG model (*Methods*), which accounts for the combined interaction modes between each amino acid pair (26). The condensed assembly is liquid-like, with chains exhibiting liquid-like diffusion, as we have shown in previous work (26). We then enumerated the average number of intermolecular contacts formed between each residue of the sequence with each residue in all other protein chains, which may represent many different modes of interaction at the atomic scale. The results highlighted a single region spanning residues 21 to 30 (RYVPPHLRGG) having highly enhanced contact probability within the condensed protein assembly (Fig. 1*A*). This region has a considerably different composition from the full RGG sequence, particularly since it contains several hydrophobic residues: This region contains the only two Pro, the only Val, and one of the only two Leu in the entire RGG domain (Fig. 1*B*). Region 21 to 30 is more prone to interaction, not only with itself but also with many regions of the protein (Fig. 1*A*).

Interestingly, subregion 21 to 28 corresponds exactly with the previously identified eIF4E-binding motif (38). We conducted a homology search, which also confirmed this region as an important functional motif due to its high degree of conservation across diverse species (Fig. 1*C* and *SI Appendix*, section 1 and Fig. S1*A*). The level of conservation is likely due to its biological function, rather than its importance to phase separation per se. However, the presence of a domain prone to self-association will still make considerable contributions to phase separation (39). We were curious whether this region alone would undergo LLPS and thus conducted CG simulations on just the eight-residue fragment. We were unable to observe LLPS for the isolated eight-residue fragment, even at very high concentrations and low temperatures (Movie S1); this can be attributed to its small chain length [shorter polypeptide chains have a lower propensity to phase-separate (21, 28)] and its net positive charge.

Next, we considered how deletion of residues 21 to 30 vs. other regions of the RGG domain will affect phase behavior to gain additional insight into the extent to which different regions of the RGG domain contribute to phase separation. Previously, we have shown that the θ-temperature (T_θ), where a single IDP chain behaves as in an ideal solvent, can serve as a good proxy for the critical temperature of phase separation (T_c) (40), above which the IDP will always form a single continuous phase regardless of the protein concentration. Taking advantage of this approximate relationship, we tested the effects of deleting distinct 10-residue segments from the LAF-1 RGG sequence by conducting single-chain simulations across a range of temperatures. We identified T_θ for each deletion and how it deviates from that of the wild-type (WT) RGG sequence (*SI Appendix*, Fig. S1 *B*, *i*). The Δ21 to 30 (Δ21–30) variant shows the greatest change (reduction in this case) of θ-temperature, much more so than any other variant, indicating that it would have the lowest propensity to phase-separate. This strongly suggests that the sticky hydrophobic subregion has an important role in phase separation of the LAF-1 RGG domain.

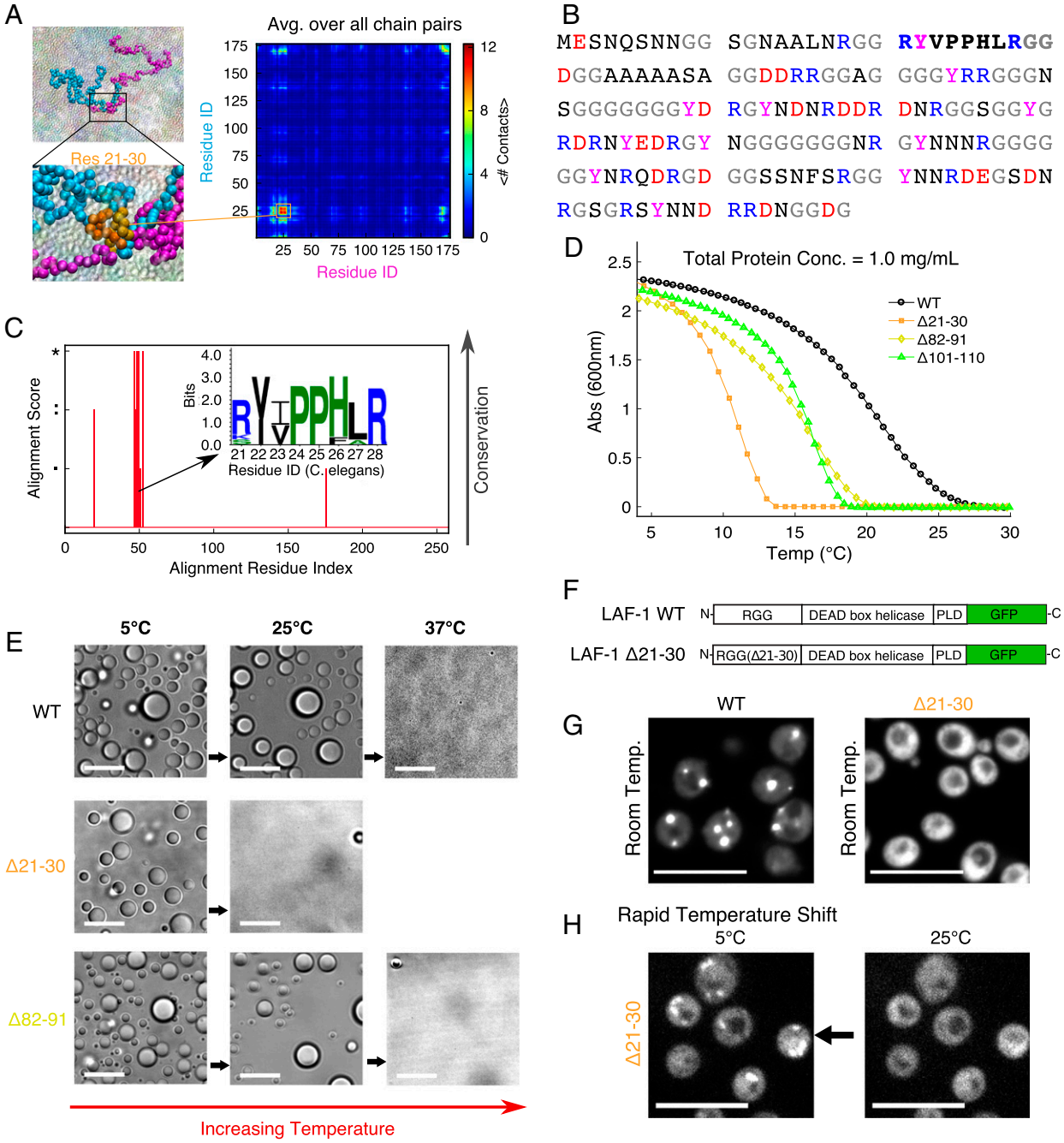

Fig. 1. A short segment of LAF-1 RGG is critical for phase separation. (*A*) CG sequence-specific simulations of LAF-1 RGG highlight a small region where contact probability is enhanced. (*Insets*) The interaction of two protein chains and zoomed view of contacts between residues within the contact-prone region. (*B*) Amino acid sequence of WT LAF-1 RGG with anionic residues colored red, cationic residues in blue, glycine in gray, tyrosine in magenta, and all others in black. Residues 21–30 are bolded. Sequences used for experiments and simulations also contained an additional C-terminal His tag (*SI Appendix*). (*C*) Sequence analysis of LAF-1 and some of its homologs highlight high sequence conservation in the folded helicase domain and poor conservation in the disordered RGG and prion-like domains (*SI Appendix*, Fig. S1*A*). Within the RGG domain, we identify one short region having good conservation, which corresponds to the region highlighted by CG simulations. The amino acids within the sequence are displayed as an inset logo. (*D*) Turbidity measurements show temperature-dependent phase behavior of WT RGG vs. variants with deletion of residues 21–30, 82–91, or 101–110. Proteins phase-separate upon cooling from above to below the phase-transition temperature. Protein concentrations were 1 mg/mL (~60 μM) in 150 mM NaCl buffer, pH 7.5. Data shown are representative of three independent turbidity experiments for each protein (*SI Appendix*, Fig. S2). Similar to previous work (21), we have not averaged the repeats, and therefore we have not added error bars because the temperatures of the measurements from different replicates are not exactly the same. T_{sat} values calculated from turbidity assays and associated statistical tests are presented in Fig. 4. (*E*) RGG Δ21–30 and RGG Δ82–19 condense into spherical liquid droplets, similarly to WT RGG, as shown by bright-field microscopy. Upon heating from 5 °C, RGG Δ21–30 droplets dissolve at a lower temperature compared to WT or RGG Δ82–91. Protein concentration and buffer are the same as for turbidity assay. (Scale bars: 10 μm.) (*F*) Schematic for full-length LAF-1 constructs including C-terminal GFP fluorescent tag. PLD: prion-like domain. (*G*) Full-length LAF-1 phase-separates in yeast at room temperature, with multiple puncta per cell. In contrast, LAF-1 Δ21–30 does not phase-separate at room temperature; delocalized fluorescence in the cytoplasm is observed. (*H*) Upon sufficient cooling, LAF-1 Δ21–30 does exhibit phase separation in yeast: Fluorescent condensates form rapidly upon cooling from 25 °C to 5 °C, consistent with in vitro results in *D*. (Scale bars: 10 μm.)

We then tested these predictions experimentally by recombinantly expressing and purifying RGG and its variants (*SI Appendix*, Fig. S1 *C* and *D*). To study protein phase behavior, we used a temperature-dependent turbidity assay (18, 21, 41, 42), in which protein solutions are cooled from above to below their phase-transition temperature. Proteins transition from well-mixed to demixed upon cooling below the saturation temperature (T_{sat}), defined as the point where we first observe an increase in the measured solution turbidity from that of the well-mixed solution. WT RGG and the deletion variants all exhibited upper-critical solution temperature phase behavior, becoming turbid upon cooling (Fig. 1*D* and *SI Appendix*, Fig. S2*A*), characteristic of IDPs rich in polar and charged amino acids (18, 43). Under these experimental conditions, the T_{sat} of WT RGG is ~26 °C, whereas the variant with the sticky hydrophobic subregion deleted (Δ21–30) has a phase-transition temperature of

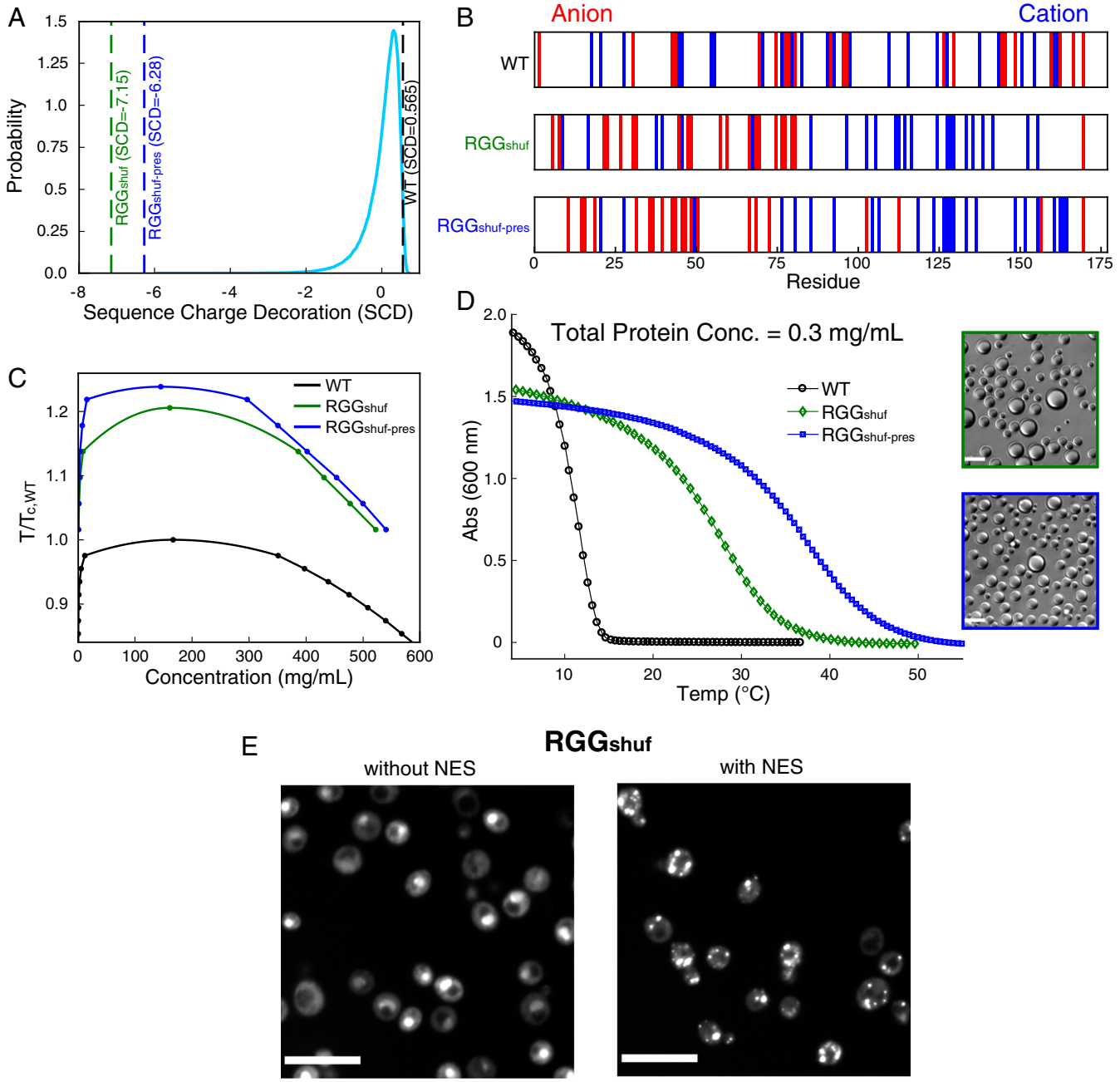

Fig. 2. Charge patterning alters LAF-1 RGG phase transition. (*A*) Probability distribution of SCD values from 1 million random shuffles of LAF-1 RGG. SCD values of WT, RGG_{shuf}, and $RGG_{shuf\text{-}pres}$ are highlighted with dashed lines. (*B*) Location of charged residues in the three sequences. (*C*) Phase diagrams of WT, $RGG_{shuf,}$, and $RGG_{shuf\text{-}pres}$ from CG simulations. Temperatures are normalized to the critical temperature of WT RGG. Errors on the concentration axis are smaller than symbols. (*D*) Turbidity measurements show the temperature-dependent phase behavior of WT RGG vs. RGG_{shuf} and $RGG_{shuf\text{-}pres}$ variants. Data shown are representative of three independent turbidity experiments for each protein (*SI Appendix*, Fig. S2). Protein concentrations were 0.3 mg/mL (~17 μM) in 150 mM NaCl buffer, pH 7.5. Both RGG_{shuf} and $RGG_{shuf\text{-}pres}$ exhibited phase-transition temperatures markedly higher than that of WT RGG, and both appeared as liquid droplet condensates under optical microscopy at room temperature. (*Insets*; scale bars: 10 μm.) T_{sat} values calculated from turbidity assays and associated statistical tests are presented in Fig. 4. (*E*) Expression in yeast of RGG_{shuf}-LAF-1 with C-terminal GFP tag. Charge patterning leads to constitutive import. The addition of NES enables RGG_{shuf}-LAF-1 to be cytosolic, and this variant exhibits protein condensate formation. (Scale bars: 10 μm.)

only ∼14 °C, representing a decrease of 12 °C. We tested two additional deletion variants, the first variant having a glycine-rich region spanning residues 101 to 110 deleted (Δ101–110) and the second variant having residues 82 to 91 deleted (Δ82–91). Residues 82–91 contain the same number of arginine and tyrosine residues as do residues 21–30. Both the Δ82–91 and the Δ101–110 variants display a more modest reduction of T_{sat}, by roughly 6 °C. These results indicate that the eIF4E-binding motif, which was identified from the CG simulation data (Fig. 1*A*), has the effect of promoting phase separation of the LAF-1 RGG domain, in addition to its specific binding function.

We then assessed whether the turbidity was due to the formation of spherical droplets, a hallmark of LLPS. We employed an optical microscope equipped with a temperature controller capable of rapidly setting the sample temperature to above or below room temperature (42, 44, 45). Indeed, we observed that both WT and the deletion variants of RGG assembled into spherical droplets below their respective values of T_{sat}. At low temperature (5 °C), Δ21–30 and the control deletions formed micrometer-scale liquid droplets that were morphologically indistinguishable from those formed by WT RGG (Fig. 1*E*). Notably, Δ21–30 droplets dissolved within 1 min upon increasing the sample temperature from 5 °C to 25 °C, whereas Δ82–91 and WT RGG exhibited slower and incomplete droplet dissolution at 25 °C, requiring a temperature of 37 °C to rapidly and fully dissolve (Fig. 1*E*). In all cases, the process was reversible in that droplets were able to assemble, disassemble, and reassemble upon cycling the temperature (*SI Appendix*, Fig. S3). Thus, both the macroscopic turbidity assays and microscopy confirmed that purified Δ21–30 phase-separates, but with significantly reduced phase-separation propensity as compared to WT RGG and the other deletion variants.

Finally, we assessed the effect of these deletions on the phase behavior of LAF-1 in living cells. For these experiments, we selected *Saccharomyces cerevisiae*, a well-established model for studying protein aggregation (46, 47), and we used full-length LAF-1 tagged with green fluorescent protein (GFP) (Fig. 1*F*). At room temperature, we observed multiple bright cytoplasmic puncta in cells expressing WT LAF-1, whereas we observed only delocalized cytoplasmic fluorescence for LAF-1 Δ21–30 (Fig. 1*G*). We confirmed by Western blot that WT LAF-1 and LAF-1 Δ21–30 expressed at similar levels (*SI Appendix*, Fig. S1*E*). The full-length Δ21–30 variant rapidly formed fluorescent cytoplasmic puncta when cooled to 5 °C, which then rapidly dispersed at 25 °C (Fig. 1*H*). This suggests that residues 21–30 are indeed important for phase separation of full-length LAF-1 in living cells, with their deletion resulting in LAF-1 having a reduced propensity to phase-separate. While deletion of this region would also likely impact the interactions of LAF-1 with eIF4E, the appreciable difference observed in the simulations and in vitro experiments—which do not incorporate the eIF4E protein—indicate that the eIF4E-binding motif itself is contributing to phase separation. It will be interesting to consider in the future how the position of the eIF4E binding region within the disordered LAF-1 RGG domain, in the context of the full-length protein, may affect its phase behavior and function.

In total, our in vitro and in vivo results suggest that LAF-1 phase separation is driven by multivalent interactions in addition to strong interactions with the more hydrophobic eIF4E-binding motif. Although this 10-amino-acid motif is necessary, it is not sufficient to control RGG phase separation, and therefore we sought additional sequence determinants.

Charge Distribution and Sequence Shuffling Can Be Used to Control LLPS. We next sought to understand how the patterning of amino acids can influence the phase separation of LAF-1 RGG, as has been studied previously for other proteins (48, 49), and the joint contributions of charge–charge interactions and the sticky hydrophobic subregion. We constructed one set of sequences having identical amino acid composition to WT RGG, but with the full sequence randomly shuffled, and a second set in which the eIF4E-binding motif (residues 21–28) was preserved. To quantify the extent to which we can expect the sequences to differ, we calculated the sequence charge decoration (SCD) parameter, where a more negative SCD score indicates greater charge segregation for sequences with many positive and negative charges. SCD has been shown to be correlated with disordered proteins' radii of gyration (R_g) (50) and with their critical temperatures (T_c) (48).

To observe the accessible SCD space of polypeptides having the same composition as the LAF-1 RGG domain, we generated 1 million randomly shuffled sequences of LAF-1 RGG and plotted the probability distribution of SCD (Fig. 2*A*). We find that randomly shuffled sequences tend to populate a very small window of SCD values, with 93.6% of the shuffled sequences having SCD scores between −2 and 0.5. For comparison, the theoretical minimum possible value for a sequence of the same length and composition is −28.03, when following the constraints set by experimental procedures (*SI Appendix*, section 1). Notably, the WT RGG sequence does not sit at the center of the distribution, but rather its SCD (0.565) is in the highest 2% of the million randomly generated sequences. This is in contrast to the intrinsically disordered regions of similar helicase proteins such as DDX4, which is more charge-segregated (44), having an SCD value of −1.02. Charge patterning could perhaps regulate phase separation in vivo such that the saturation concentration of LAF-1 is similar to that of the native expression level, and also make it distinguishable from other proteins of similar amino acid composition (51).

We selected the sequence with the lowest SCD value, termed RGG_{shuf}. We did the same for sequences having the eIF4E-binding motif preserved ($RGG_{shuf\text{-}pres}$), to test whether there is an appreciable difference between charge-segregated variants with and without the presence of a sticky hydrophobic subregion. The two sequences are depicted in Fig. 2*B*, which shows that both have an abundance of anionic residues in the first half of the sequence, and an abundance of cationic residues in the second half, in contrast with the WT sequence, which has a relatively even distribution of cationic and anionic residues throughout. We conducted CG molecular simulations for these sequences and determined the phase diagrams as a function of temperature, observing that the protein-rich phase is liquid-like in both cases (Movies S2 and S3). Both shuffled sequences show a drastic increase in the critical temperature compared to WT (Fig. 2*C*), as well as compaction in single-chain simulations (*SI Appendix*, Fig. S4). Interestingly, RGG_{shuf} does not exhibit as large of an upward shift in T_c as does $RGG_{shuf\text{-}pres}$, even though it has a slightly lower SCD value. This indicates that charge patterning is capable of inducing large shifts to the phase diagram, but a combination of charge segregation and preservation of the hydrophobic subregion promotes LLPS even more.

We then tested these predictions experimentally by conducting temperature-dependent turbidity assays on recombinantly expressed and purified WT RGG, RGG_{shuf}, and $RGG_{shuf\text{-}pres}$. These experiments were performed using lower concentrations (0.3 mg/mL) of protein because the two shuffled variants display a much greater propensity to phase-separate (Fig. 2*D* and *SI Appendix*, Fig. S2*B*). Remarkably, whereas WT RGG undergoes LLPS at ∼15 °C under these conditions, RGG_{shuf} demixed at 42 °C, and $RGG_{shuf\text{-}pres}$ demixed at 52 °C. (We observed a lower T_{sat} for WT RGG here compared to Fig. 1*D* due to the need for reduced protein concentration in the case of the shuffled sequences.) This finding nicely agrees with our computational results, which showed that increasing the charge segregation in combination with preserving the eIF4E-binding motif enhances self-association propensity more than simply increasing charge segregation. Importantly, despite such drastic rearrangement of the protein

sequence, both RGG_{shuf} and $RGG_{shuf\text{-}pres}$ formed spherical liquid droplets of normal morphology, as imaged by bright-field microscopy at room temperature (Fig. 2 *D*, *Insets*).

To determine whether altering the charge patterning of the RGG sequence has any unexpected consequences in vivo, we then tested RGG_{shuf} in the context of full-length LAF-1 in live yeast cells. LAF-1 in which the RGG domain was replaced with RGG_{shuf} (RGG_{shuf}-LAF-1) appeared to localize to the nucleus, with a single fluorescent punctum per cell (Fig. 2*E*). This is perhaps unsurprising, as nuclear localization signals characteristically contain stretches of basic amino acids (52, 53). We therefore tagged RGG_{shuf}-LAF1 with a nuclear export signal (NES), which upon expression generated cytoplasmic puncta, thus demonstrating that RGG_{shuf} is capable of self-assembling in living cells. Together, these experimental results support the computational predictions that charge patterning is a critical determinant of LAF-1 RGG phase separation and that this effect can be supplemented by the incorporation of small patches of hydrophobic amino acids. We were unable to conduct the same in vivo experiments on $RGG_{shuf\text{-}pres}$ due to its poor expression in yeast cells.

Arginine and Tyrosine Are Important Determinants for LLPS of LAF-1 RGG. Interactions of tyrosine and arginine can be critically important to protein LLPS (30, 32, 54). The LAF-1 RGG domain contains 24 arginine, 11 tyrosine, and 1 phenylalanine (and no lysine) residues, which are relatively evenly distributed across the 168-residue-long domain (Fig. 3*A*). To test the role of these residues in RGG phase separation, in one construct we mutated all tyrosines to phenylalanine (Y→F), except for a single tyrosine that was mutated to tryptophan to facilitate spectrophotometric detection. In a second construct, we mutated all arginines to lysines (R→K). We then conducted turbidity assays (at 1 mg/mL protein concentration, since the mutations were likely to reduce LLPS propensity) on both constructs.

In contrast to WT RGG, which demixed at ∼26 °C, mutating the tyrosines to phenylalanines lowered the transition temperature to ∼14 °C (Fig. 3*B* and *SI Appendix*, Fig. S2*C*). To confirm that the Y→F mutant still forms normal protein droplets, we imaged it with bright-field microscopy at 5 °C. We observed that the condensates appeared morphologically identical to WT RGG, with many micrometer-scale protein droplets (Fig. 3 *B*, *Insets*). Even more dramatically, upon mutating all arginines to lysines, we observed no phase separation, even below 5 °C (Fig. 3*B* and *SI Appendix*, Fig. S2*C*). The R→K mutant was soluble and did not assemble into protein droplets even under experimental conditions that promote RGG phase separation, including high protein concentration and low salt concentration at low temperature. Thus, the presence of tyrosine and arginine plays a key role in phase separation of the LAF-1 RGG domain, in agreement with studies on FUS (30).

These experimental results suggest that the Y→F and R→K mutations have a significant impact on the overall interactions occurring between LAF-1 RGG molecules. To gain mechanistic insight into these changes, we turned to all-atom simulations with explicit solvent, which can provide highly detailed information on the different types of interactions in which each amino acid may participate (35). Since it is currently impractical to faithfully sample the configurational ensemble of a long IDP like LAF-1 using such high-resolution models, we conducted simulations on a 44-residue region of the LAF-1 RGG domain spanning residues 106 to 149 ($RGG_{106–149}$). This particular contiguous region was selected to provide the highest compositional similarity with the full RGG domain so that the information obtained is most consistent with the expectations for the full-length sequence (*SI Appendix*, Fig. S5). We also simulated two variants in which either all of the tyrosine residues are mutated to phenylalanine (Y→F $RGG_{106–149}$) or all of the arginine residues are mutated to lysine (R→K $RGG_{106–149}$). From single-chain simulations, we find that R_g increases in the following order: WT < Y→F < R→K (Fig. 3*C*). Previous studies provide compelling evidence that chain dimensions or solvent quality can faithfully provide knowledge on protein LLPS (3, 40, 48, 55)—more collapsed chains are expected to be more prone to phase separation. Therefore, the trend in R_g from all-atom simulations is consistent with the experimental LLPS behavior that we observe for these mutants (Fig. 3*B*), which provides further confidence in utilizing these simulations to understand the molecular interactions responsible for the experimental results.

To observe intermolecular interactions and self-association, we conducted simulations of two $RGG_{106–149}$ chains. Consistent with our recent work on the FUS LC domain (35), we used well-tempered metadynamics with the number of intermolecular van der Waals (VDW) contacts as a pertinent collective variable to enhance sampling of intermolecular contacts between the two peptides. The resulting free energy surfaces as a function of the number of intermolecular VDW contacts are shown in *SI Appendix*, Fig. S5*A*. Both WT and Y→F peptides show free energy minima at a finite number of VDW contacts. Interestingly, the R→K variant has a global minimum at zero contacts, suggesting the two chains do not interact, as is consistent with the lack of phase separation in the experiments.

Previous work has suggested the importance to LLPS of cation–π interactions (30, 56) [particularly between arginine and tyrosine (30)], planar interactions between sp^2 hybridized groups (referred to here as sp^2/π interactions) (45), electrostatic interactions (16, 57), and hydrophobic and VDW interactions (35). We calculated the average number of intermolecular contacts between the two chains of the different $RGG_{106–149}$ variants (Fig. 3*D*). In general, WT and Y→F have a much higher number of contacts than R→K, consistent with the free energy profiles, showing that R→K most favors unbound configurations. We also normalized the average number of intermolecular contacts of each type by the average number of intermolecular VDW contacts (*SI Appendix*, Fig. S6*B*) to understand the role of various interaction modes independent of the global contact propensity, which is different between these three variants. Additionally, we provide the unnormalized average number of various contacts formed by each residue (*SI Appendix*, Fig. S7). The number of sp^2/π and cation–π interactions is particularly decreased in R→K, while there is no significant difference between WT and Y→F average contacts. The overall number of contacts, however, may not consider the interaction strengths and thus would not perfectly describe the difference between WT and Y→F.

To further elucidate the differences between the mutants, we considered the effect of the interactions between cationic and aromatic side chains, which were the original target of these designed mutations. By analysis of all simulation snapshots in which arginine or lysine and tyrosine or phenylalanine residues from different chains are in contact (having at least one VDW contact between them), we calculated the probability of occurrence of different interaction types. Three different interaction modes are observed for arginine–tyrosine contacts, while only two are observed for arginine–phenylalanine and lysine–tyrosine contacts (Fig. 3*E*). Importantly, interactions between arginine–tyrosine side chains promoting LLPS could be due to multiple modes of interactions with significant contributions from cation–π, hydrogen bonding, and sp^2/π interactions. The Y→F mutations reduce the extent of these interactions, likely due to loss of hydrogen bonding interactions, as the Y side chain contains a hydroxyl group but F lacks it. The R→K mutations remove planar sp^2/π interactions due to loss of the guanidinium group, which is present in R but not K. Moreover, we observe that these interaction types (particularly cation–π and sp^2/π) overlap, suggesting that they work cooperatively. Multiple studies have observed the important role of R and Y in protein LLPS

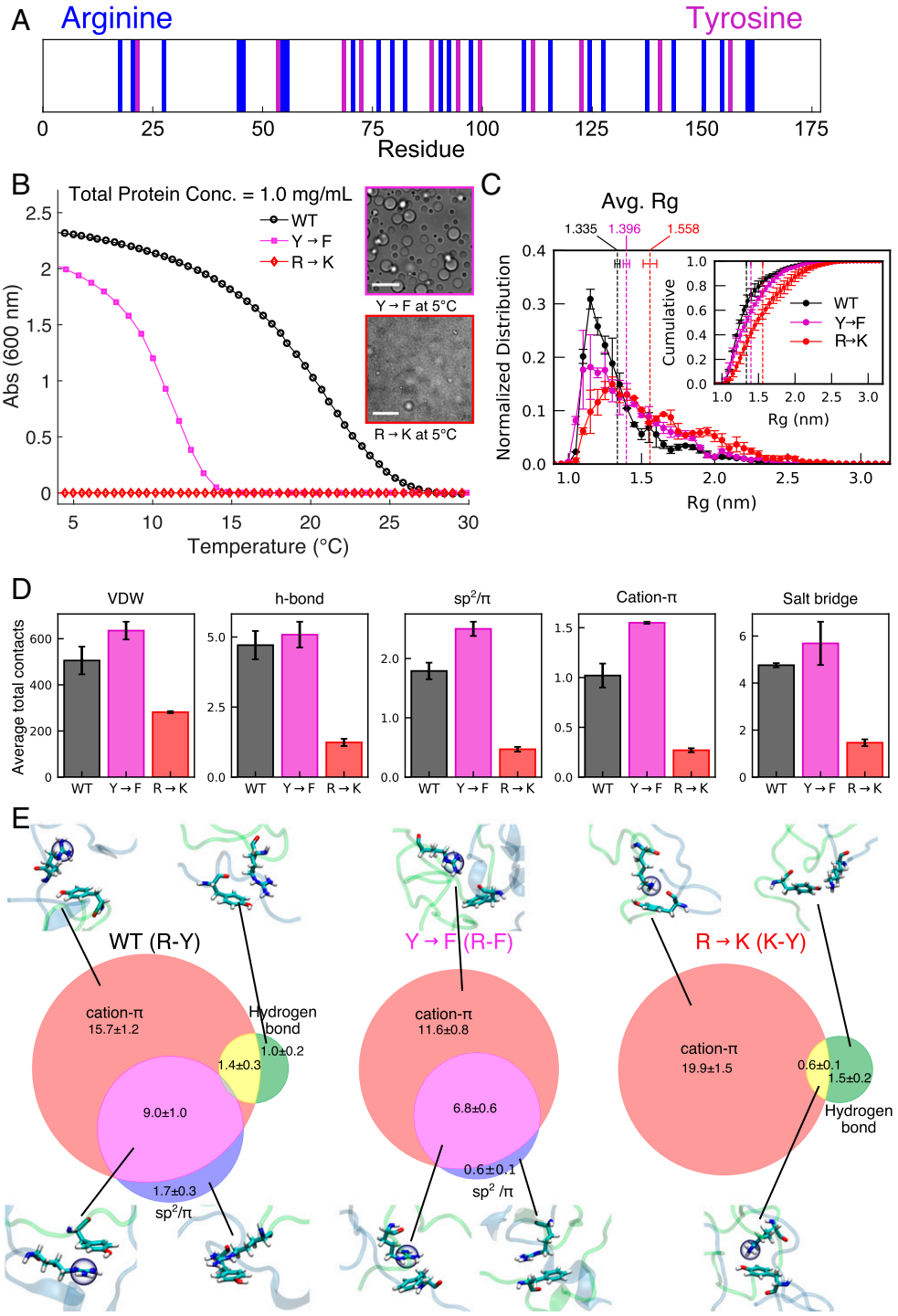

Fig. 3. Contribution of arginine and tyrosine residues to LLPS. (*A*) Arrangement of arginine and tyrosine residues along the RGG sequence. Residues are reasonably well-mixed with the exception that the N-terminal end is relatively void of the two amino acids. (*B*) Turbidity measurements show the temperature-dependent phase behavior of WT RGG vs. Y→F or R→K variants. Data shown are representative of three independent turbidity experiments for each protein (*SI Appendix*, Fig. S2). For turbidity assays, protein concentrations were 1 mg/mL (~60 μM) in 150 mM NaCl buffer, pH 7.5. The Y→F variant assembled into spherical liquid droplets (inset micrograph) at 5 °C. The R→K variant did not phase-separate in the turbidity assay, nor were micrometer-scale protein liquid droplets visible by optical microscopy (bottom inset), even under conditions favorable for phase separation (6.6 mg/mL protein, 50 mM NaCl, pH 7.5, 5 °C). (Scale bars: 10 μm.) (*C*) Normalized distribution of radius of gyration (R_g) of $RGG_{106–149}$ fragments from single-chain simulations for WT, Y→F, and R→K variants. (*Inset*) Cumulative histogram of R_g. (*D*) Average number of intermolecular contacts observed between two chains of $RGG_{106–149}$ in two-chain simulations (*Methods*), where the average is over the simulated ensemble. Backbone and side-chain heavy atoms are included in these calculations. (*E*) Venn diagrams summarizing the interaction types driving the association of R/K and Y/F residues averaged over all instances of intermolecular VDW contact between any pair of these residues. The numbers represent the percentage and only side-chain heavy atoms are included in these calculations. The overlap between different interaction types shows that they may work cooperatively. WT has all three types of interaction, while R→K loses sp^2/π interactions, and Y→F loses hydrogen bonding. Snapshots show an instance of indicated contact type(s) from a two-chain simulation. For simulation data, error bars and uncertainty values are SEM with $n = 2$.

(30, 32, 54, 56), and our results contribute much-needed mechanistic understanding of these observations.

Sequence Perturbations Result in Shifts to Phase Diagram. To more completely map the experimental phase behavior of variants of the LAF-1 RGG domain, we performed temperature-dependent turbidimetry at varying protein concentrations and calculated T_{sat} for each to obtain the low-concentration arm of their phase diagrams. We find that all variants for which we were able to acquire multiple T_{sat} values display an upper critical solution temperature phase diagram, having a region of miscibility at high temperatures and phase separation at low temperatures. By imposing different perturbations to the RGG sequence, we were able to shift the phase diagram upward (Fig. 4*A*) or downward (Fig. 4*B*). A significant increase of LLPS propensity occurs when modifying the sequence such that most cations are localized to one side and anions on the other side, even when the sticky hydrophobic region we identified is lost in the shuffling. We find that designing a shuffled sequence that conserves this region (such conservation has occurred across different organisms) results in the greatest upward shift of the phase diagram (Fig. 4*A*), indicating that both of these types of molecular interactions control phase separation of RGG.

We are also able to shift the phase diagram downward, thus making LLPS less favorable. Mutations of all arginine to lysine result in total loss of LLPS behavior at tested conditions. We suggest the phase diagram has been shifted downward enough that the temperatures or concentrations required to observe LLPS are not practically achievable in vitro (Fig. 4*B*). When deleting residues 21–30, encompassing the eIF4E-binding motif, we find that the phase diagram shifts downward significantly (Fig. 4*B*), much more so than when deleting other regions of 10 residues (*SI Appendix*, Fig. S8). This is also consistent with the predictions of the computational model, which identified the enhanced interactions within that region.

As a control for our experiments on sequence shuffling and residues 21–30, we designed and tested an additional construct, $RGG_{shuf\text{-}control}$. $RGG_{shuf\text{-}control}$ is a shuffled version of the RGG sequence having similar charge distribution as WT (i.e., approximately uniform, with SCD = 0.37), and the eIF4E-binding motif was not preserved in this shuffling process. We observed that $RGG_{shuf\text{-}control}$ has a lower T_{sat} than WT RGG (*SI Appendix*, Fig. S8). These results indicate that T_{sat} decreases upon shuffling the sequence in such a way that maintains WT-like uniform charge distribution while disrupting residues 21–30. This further supports the assertion that the increased T_{sat} of RGG_{shuf} can be attributed to its blocky charge patterning, and again highlights the contribution of residues 21–30 to LAF-1 RGG phase separation.

In previous work, Wang et al. (30) suggested that the saturation concentration (c_{sat}) of a protein may be predicted by counting the number of tyrosine and arginine residues within the sequence as $\boldsymbol{c_{sat} = k(n_{Tyr}n_{Arg})^{-1}}$, where k is a fitting parameter and is equal to 6.5 mM. For the WT RGG sequence, this predicts a saturation concentration of 24.6 μM or 0.439 mg/mL, which also applies to RGG_{shuf} and $RGG_{shuf\text{-}pres}$, as they have an identical composition (Fig. 4*C*). For the R→K and Y→F variants, the denominator becomes zero, so the predicted value is undefined, with the suggestion that c_{sat} is very high. Deletion of residues 21–30 removes two arginine and one tyrosine residue, resulting in a small predicted increase of c_{sat} to 29.6 μM or 0.493 mg/mL. To directly compare with results from this prediction, we calculated saturation concentration at 23 °C using a logarithmic fit to turbidimetry data (*SI Appendix*, Fig. S9 *A* and *B*). Linear fits of the data yield similar c_{sat} values (*SI Appendix*, Fig. S9 *C* and *D*). We find that the equation $\boldsymbol{c_{sat} = k(n_{Tyr}n_{Arg})^{-1}}$ poorly predicts the c_{sat} for RGG_{shuf} or $RGG_{shuf\text{-}pres}$ (Fig. 4*C*). Further, the prediction underestimates the effect of deletion of residues 21–30 from RGG. These results suggest that while the number of arginine and tyrosine residues can sometimes provide a reasonable estimate of c_{sat}, this parameter alone is not predictive, and many other factors, such as charge patterning and hydrophobic interactions, determine LLPS.

Protein Condensates Formed from RGG Variants Retain Liquid-Like Properties. Thus far, we have demonstrated perturbations to the LAF-1 RGG sequence that alter its phase behavior, using molecular simulations to guide experiments and provide a mechanistic understanding of the driving forces of phase separation. We next wondered whether these sequence perturbations would alter the liquid properties of RGG protein condensates. This is important to understand because the material properties of biomolecular condensates are intertwined with their biological function (58). The spherical morphologies of WT RGG and its sequence variants are characteristic of viscous liquids. For all variants, droplets could be seen contacting, fusing, and then rounding into larger spheres (Fig. 5*A*). To determine the liquidity of these droplets, we quantified fusion events, calculating the time τ for the two coalescing droplets to relax to a sphere

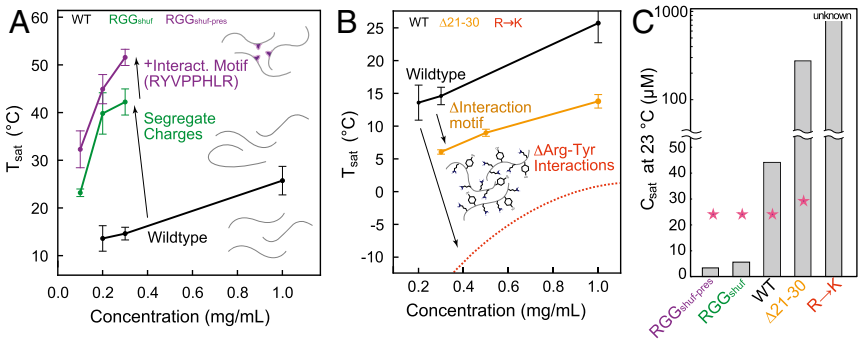

Fig. 4. Phase diagrams illustrate molecular interactions that underlie RGG LLPS. Phase diagrams for different LAF-1 variants. T_{sat} values and associated error bars were calculated from triplicates of the turbidity assays at each concentration. (*A*) Shuffled sequences with a high degree of charge patterning shift phase diagram upward, making phase separation occur at lower concentrations more easily. $RGG_{shuf\text{-}pres}$ features both charge segregation and the self-interaction motif at residues 21–28, allowing for even greater LLPS propensity. (*B*) Deletion of the interaction motif, or mutation of arginine residues to lysine, both result in a drastic decrease of LLPS propensity and downward shift of the phase diagram. Phase diagram for R→K is theoretical and is meant strictly as a visual guide to show that this mutation has a stronger effect on LLPS than the deletion of the interaction motif. T_{sat} of WT, Δ21–30, RGG_{shuf}, and $RGG_{shuf\text{-}pres}$ are all significantly different from one another ($P < 0.005$), based on one-way ANOVA followed by Tukey's post hoc test at 0.3 mg/mL. (*C*) Saturation concentrations from turbidity experiments (gray bars) compared with predictions based on ref. 30 (pink stars).

(*SI Appendix*, Fig. S10*A*). WT RGG and all of the variants examined (RGGΔ21–30, both shuffled versions, and Y→F) exhibited rapid fusion, with droplets of length scale $\ell = 2 \pm 0.25$ μm fusing with $\tau < 100$ ms (*SI Appendix*, Fig. S10*B*). Droplet fusion is driven by surface tension γ and slowed by viscosity η, and the time scale of fusion is also proportional to droplet size ℓ, so $\tau \approx (\eta/\gamma)\ell$ (6, 7, 59). By plotting τ against ℓ for tens of droplet fusion events (Fig. 5*B*), we estimate the ratio η/γ, known as the inverse capillary velocity (Fig. 5*C*). All of the variants tested had η/γ within threefold that of WT RGG, and in all cases $\eta/\gamma < 0.05$ s/μm, indicating faster fusion compared to full-length LAF-1 ($\eta/\gamma = 0.12$ s/μm) (6).

In a complementary approach, we examined dynamics within the droplets through fluorescence recovery after photobleaching (FRAP). For all variants tested, 50% fluorescence recovery was achieved within 30 s of photobleaching a small circular region within a larger droplet (Fig. 5 *D* and *E*). By fitting the FRAP recovery curves to a three-dimensional infinite model, we find diffusion coefficients ranging from $D = 0.01$ μm^2/s to 0.025 μm^2/s, approximately one order of magnitude faster than that for full-length LAF-1 (60) (Fig. 5*E* and *SI Appendix*, Fig. S10*C*). There are modest differences, notably that the construct with deletion of residues 21–30 (lower T_{sat} than WT) exhibited faster FRAP recovery and fusion compared to $RGG_{shuff\text{-}pres}$ (highest T_{sat} of all constructs we tested). However, the main result is that the FRAP and fusion experiments together demonstrate that these variants maintain dynamic, liquid-like condensates, despite the changes to sequence and phase behavior. Consistent with these experimental results, simulation movies of different RGG variants also display liquid-like behavior (Movies S2 and S3). Thus, our results suggest that certain sequence perturbations to LAF-1 RGG may modulate its phase behavior—critical concentration and transition temperature—mostly independently from droplet liquidity.

Discussion

In this work, we elucidate sequence determinants of IDP phase separation, and in so doing we advance a computationally guided approach for rational engineering of protein LLPS. We focus on the RGG domain from LAF-1, a prototypical phase-separating protein of great interest to the LLPS field whose sequence-to-phase behavior relationship has not been mapped in detail previously. By combining simulations and experiments, we identified three important features that govern the propensity of this protein to phase-separate: a short conserved domain, charge patterning, and arginine–tyrosine interactions.

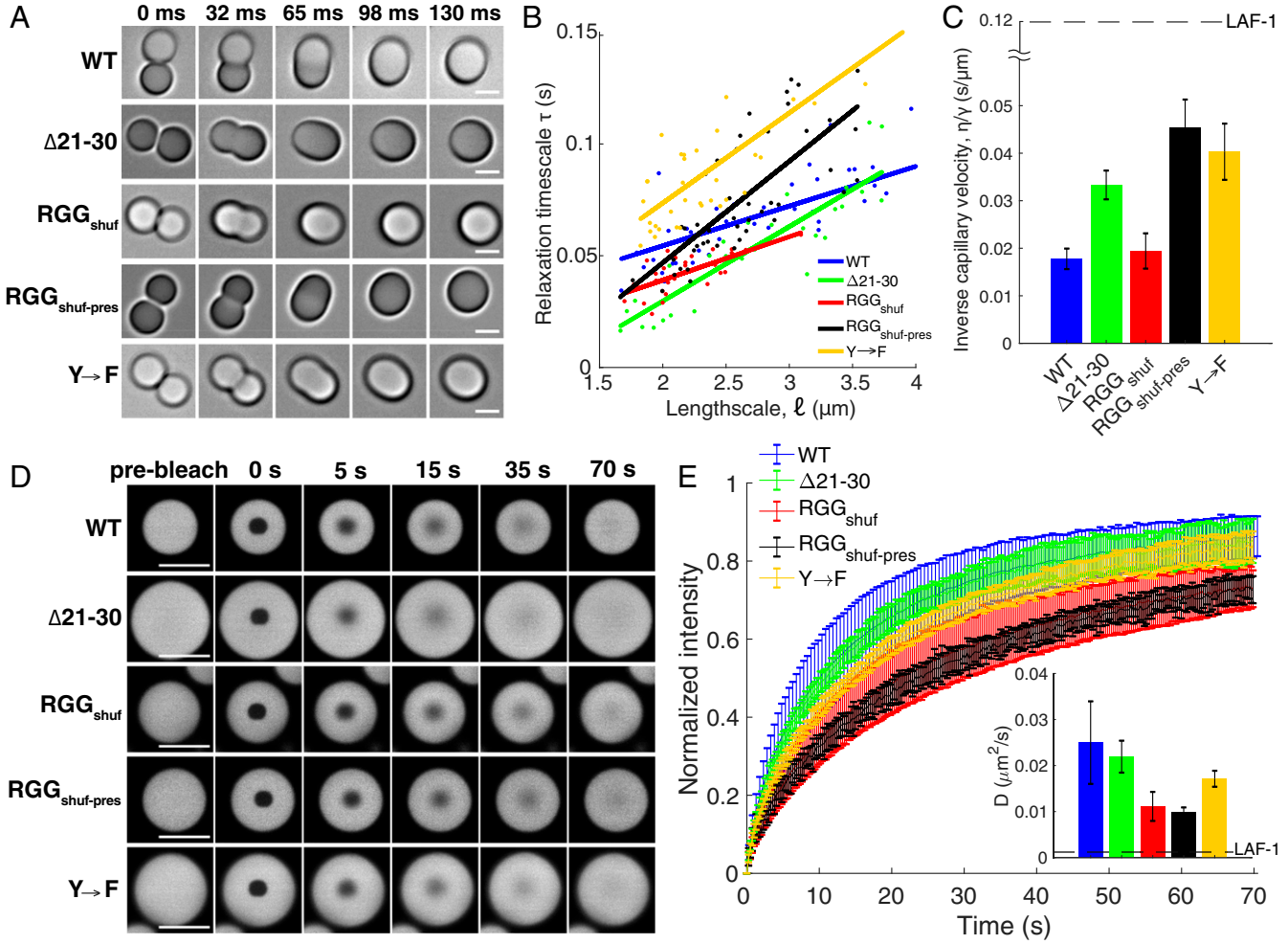

Fig. 5. RGG variants exhibit liquid-like material properties. (*A*) Droplets fuse rapidly to form a single larger sphere. (Scale bars: 2 μm.) (*B*) Relaxation time scale of droplet fusion, plotted against droplet length scale. (*C*) The inverse capillary velocity, η/γ, is the slope of the linear fits to the data in *B*. Dashed horizontal line represents inverse capillary velocity of full-length LAF-1, determined previously (6). (*D*) Representative images from FRAP experiments. (Scale bars: 10 μm.) (*E*) Normalized FRAP recovery curves show >50% recovery within 30 s for all variants. Error bars represent STD ($n > 15$). (*Inset*) Diffusion coefficients, D, calculated by fitting the FRAP recovery curves to an infinite boundary model in three dimensions (60). Dashed horizontal line represents D of full-length LAF-1, determined previously (60).

We first demonstrate that a small conserved domain plays an unexpectedly large role in LAF-1 phase separation, such that the deletion of 10 residues encompassing the identified region decreases the protein's phase separation propensity significantly. Our computational data and in vitro experiments suggest that this region has an intrinsic affinity for itself. This contact-prone region coincides with the previously identified eIF4E-binding motif, although the contribution of this motif to LLPS is likely orthogonal to its specific binding function. Hypothetically, LLPS of LAF-1 might be particularly sensitive to stimuli that may target this region, such as phosphorylation-induced folding that may hide the motif and block its accessibility for self-association (61). More generally, these results suggest that the presence within proteins of functional motifs, such as specific binding motifs (38), may have a nonnegligible effect on LLPS of the full sequence—even if the functional motif is only a small region in a much larger protein.

Second, our results support a revised view of the role of electrostatic interactions in LAF-1 RGG phase separation. Previous views pointed to electrostatic interactions and charge patterning as the driving force for LAF-1 phase separation (6, 13). On the contrary, we found that WT LAF-1 RGG has a well-mixed charge distribution. We therefore asked whether introducing charge patterning could enhance LAF-1 phase separation. We used the SCD metric to identify shuffled versions of LAF-1 RGG having a high degree of charge segregation, and our CG simulations and experiments both show that such charge patterning results in significantly enhanced propensity to phase-separate. Our results extend previous work on this topic (48, 49, 62). DDX4 features blocks of alternating net charge, and scrambling the blocks to remove charge patterning abolishes phase separation (44). Relatedly, complex coacervation of the negatively charged Nephrin intracellular domain (NICD) with positively charged partners is promoted in part by blocks of high charge density in NICD (57). Theoretical work shows as well that block polyampholytes exhibit stronger interactions compared to charge-scattered polyampholytes, as the latter experience repulsion from nearby like charges (48). Thus, it appears that WT RGG may be under negative selection to moderate this mode of blocky electrostatic interaction and maintain a well-mixed charge distribution.

Third, we find that distributed tyrosine and arginine residues are also important to the ability of LAF-1 RGG to phase-separate, and we gain valuable mechanistic insight into this result from all-atom simulations. The importance of these particular residues was attributed in previous work to their propensity to form cation–π interactions (30, 32, 54). Our all-atom simulations confirm the presence of cation–π interactions and, importantly, highlight other important interaction modes as well that change when mutating arginine to lysine or tyrosine to phenylalanine. Our simulations suggest that the loss of planar sp^2/π interactions (45) is likely responsible for reduced LLPS when mutating arginine to lysine. We note that arginine may be particularly prone to promoting LLPS with aromatic-rich sequences due to cooperative cation–π and sp^2/π interactions that cooccur. Another important interaction mode is hydrogen bonding, which has also recently been demonstrated to be important to LLPS (35, 37) and is present in interactions between cationic residues and tyrosine. Our simulations suggest that the reduced LLPS propensity when mutating tyrosine to phenylalanine can be explained by the loss of side-chain hydrogen bonding, as phenylalanine lacks the hydroxyl group. Therefore, we suggest that while the selected mutations likely weaken cation–π interactions (30, 32), one must also consider the loss of several other types of interactions that are responsible for stabilizing the condensed liquid phase (35).

The sequence perturbations investigated here significantly altered c_{sat}—for instance, approximately one order of magnitude decrease in c_{sat} for $RGG_{shuff\text{-}pres}$ compared to WT RGG, and an approximately fivefold increase for Δ21–30. Remarkably, we observed that the RGG variants retained their dynamic liquid material properties, even for a perturbation as drastic as shuffling the sequence. The significant changes in phase behavior would likely have important biological consequences, whereas the modest differences in droplet fluidity are likely of smaller functional significance. Thus, our experiments suggest that in a predictive manner, we can design mutations to an IDP to alter its phase behavior while retaining liquid-like condensate dynamics. It is important to note that rheology and phase behavior may be influenced by interactions with other biomolecules as well. P granules in *C. elegans* are multicomponent systems, containing not just LAF-1 but a multitude of RNAs and other proteins, such as MEG-3 and PGL-3, that collectively determine P-granule properties (63, 64). For instance, the addition of RNA in LAF-1 droplets in vitro decreased droplet viscosity but did not significantly alter phase behavior (6). Recent work showed that P granules are in fact a coassembly of liquid and gel phases, where MEG-3 forms the gel phase (63). The sequence determinants of phase behavior and rheology of these multicomponent systems are a rich topic for future study.

Overall, our combined results elucidate the driving forces of LLPS and highlight how sequence perturbations affect LLPS. This work will inform future studies into the biology of membraneless organelles, aberrant phase transitions in disease, and design of biomaterials and synthetic organelles.

Methods

See *SI Appendix* for details. Genes were cloned into a pET vector in-frame with a C-terminal 6xHis-tag for bacterial expression and into the YIplac211 vector in frame with a C-terminal mEGFP tag for yeast expression. A combination of experimental techniques, including temperature-dependent turbidity assays and microscopy, were used to measure phase behavior, FRAP, and droplet fusion. CG and all-atom simulations were conducted to compute the phase coexistence and to identify the molecular interactions underlying the differences in the behavior of the various proteins.

Data Availability Statement. All of the simulation results discussed in this paper are generated based on computer software that is publicly available. See *SI Appendix* for details. The processed data and associated scripts will be made available by contacting the corresponding author.

ACKNOWLEDGMENTS. We thank Erfei Bi, Kangji Wang, and James Shorter for yeast strains, reagents, and protocols and Cliff Brangwynne and Shana Elbaum-Garfinkle for the full-length LAF-1 gene. We gratefully acknowledge Andrew Tsourkas for use of the temperature-controlled spectrophotometer, Hui Chen for assistance with Western blotting, Ellen Reed for assistance with mass spectrometry, Xinyi Li for assistance with data analysis, and Nick Fawzi for helpful discussions. This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences awards DE-SC0007063 to D.A.H. (experiments) and DE-SC0013979 to J.M. (theory and simulation). J.M. also acknowledges partial support from NIH grant R01-NS116176. We gratefully acknowledge the use of the high-performance computing capabilities of the Extreme Science and Engineering Discovery Environment, which is supported by NSF grant TG-MCB-120014, and the National Energy Research Scientific Computing Center, supported by the Office of Science of the US Department of Energy under contract DE-AC02-05CH11231. B.S.S. received support from an NIH postdoctoral fellowship (F32-GM119430). W.S.T. received support from an NSF grant (1845734). M.C.G. acknowledges support from an NSF Superseed, NIH grant R01-EB028320, and Burroughs Wellcome Fund.

1. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
2. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
3. J. A. Riback *et al.*, Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **168**, 1028–1040.e19 (2017).
4. A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016).
5. 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* **69**, 465–479.e7 (2018).

6. S. Elbaum-Garfinkle *et al*., The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7189–7194 (2015).
7. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4334–4339 (2011).
8. B. R. Sabari *et al*., Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
9. A. G. Larson *et al*., Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
10. I. A. Sawyer, D. Sturgill, M. Dundr, Membraneless nuclear organelles and the search for phases within phases. *Wiley Interdiscip. Rev. RNA* **10**, e1514 (2019).
11. G. Wan *et al*., Spatiotemporal regulation of liquid-like condensates in epigenetic inheritance. *Nature* **557**, 679–683 (2018).
12. M. R. Marzahn *et al*., Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *EMBO J.* **35**, 1254–1275 (2016).
13. 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).
14. A. L. Darling, Y. Liu, C. J. Oldfield, V. N. Uversky, Intrinsically disordered proteome of human membrane-less organelles. *Proteomics* **18**, e1700193 (2018).
15. V. N. Uversky, I. M. Kuznetsova, K. K. Turoverov, B. Zaslavsky, Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.* **589**, 15–22 (2015).
16. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
17. M. Dzuricky, S. Roberts, A. Chilkoti, Convergence of artificial protein polymers and intrinsically disordered proteins. *Biochemistry* **57**, 2405–2414 (2018).
18. F. G. Quiroz, A. Chilkoti, Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* **14**, 1164–1171 (2015).
19. J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. López, Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017).
20. R. A. Kapelner, A. C. Obermeyer, Ionic polypeptide tags for protein phase separation. *Chem. Sci. (Camb.)* **10**, 2700–2707 (2019).
21. B. S. Schuster *et al*., Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 2985 (2018).
22. S. Roberts *et al*., Injectable tissue integrating networks from recombinant polypeptides with tunable order. *Nat. Mater.* **17**, 1154–1163 (2018).
23. J. R. Simon, S. A. Eghtesadi, M. Dzuricky, L. You, A. Chilkoti, Engineered ribonucleoprotein granules inhibit translation in protocells. *Mol. Cell* **75**, 66–75.e5 (2019).
24. H. K. Lau *et al*., Microstructured elastomer-PEG hydrogels via kinetic capture of aqueous liquid-liquid phase separation. *Adv. Sci. (Weinh.)* **5**, 1701010 (2018).
25. A. Hubert, P. Anderson, The C. elegans sex determination gene laf-1 encodes a putative DEAD-box RNA helicase. *Dev. Biol.* **330**, 358–367 (2009).
26. G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, e1005941 (2018).
27. M.-T. Wei *et al*., Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **9**, 1118–1125 (2017).
28. E. W. Martin, T. Mittag, Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478–2487 (2018).
29. E. Gomes, J. Shorter, The molecular language of membraneless organelles. *J. Biol. Chem.* **294**, 7115–7127 (2019).
30. 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).
31. A. Kamenska, C. Simpson, N. Standart, eIF4E-binding proteins: New factors, new locations, new roles. *Biochem. Soc. Trans.* **42**, 1238–1245 (2014).
32. Y. Lin, S. L. Currie, M. K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120 (2017).
33. S. Rauscher, R. Pomès, The liquid structure of elastin. *eLife* **6**, e26526 (2017).
34. E. P. Bentley, B. B. Frey, A. A. Deniz, Physical chemistry of cellular liquid-phase separation. *Chemistry* **25**, 5600–5610 (2019).
35. A. C. Murthy *et al*., Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
36. T. P. Dao *et al*., ALS-linked mutations affect UBQLN2 oligomerization and phase separation in a position- and amino acid-dependent manner. *Structure* **27**, 937–951.e5 (2019).
37. B. Gabryelczyk *et al*., Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. *Nat. Commun.* **10**, 5465 (2019).
38. J.-W. Shih *et al*., Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *Biochem. J.* **441**, 119–129 (2012).
39. G. L. Dignon, W. Zheng, J. Mittal, Simulation methods for liquid-liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.* **23**, 92–98 (2019).
40. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9929–9934 (2018).
41. Y. Yang, H. B. Jones, T. P. Dao, C. A. Castañeda, Single amino acid substitutions in stickers, but not spacers, substantially alter UBQLN2 phase transitions and dense phase material properties. *J. Phys. Chem. B* **123**, 3618–3629 (2019).
42. T. Yoshizawa *et al*., Nuclear import receptor inhibits phase separation of FUS through binding to multiple sites. *Cell* **173**, 693–705.e22 (2018).
43. G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Temperature-controlled liquid-liquid phase separation of disordered proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).
44. T. J. Nott *et al*., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
45. R. M. Vernon *et al*., Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, 1–48 (2018).
46. J. Couthouis *et al*., A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20881–20890 (2011).
47. Z. Sun *et al*., Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol.* **9**, e1000614 (2011).
48. Y.-H. Lin, H. S. Chan, Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys. J.* **112**, 2043–2046 (2017).
49. J. P. Brady *et al*., Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8194–E8203 (2017).
50. L. Sawle, K. Ghosh, A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* **143**, 085101 (2015).
51. Y. H. Lin, J. P. Brady, J. D. Forman-Kay, H. S. Chan, Charge pattern matching as a 'fuzzy' mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* **19**, 115003 (2017).
52. A. Lange *et al*., Classical nuclear localization signals: Definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**, 5101–5105 (2007).
53. S. Kosugi *et al*., Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284**, 478–485 (2009).
54. S. Qamar *et al*., FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* **173**, 720–734.e15 (2018).
55. Z. Monahan *et al*., Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* **36**, 2951–2967 (2017).
56. J. Song, S. C. Ng, P. Tompa, K. A. W. Lee, H. S. Chan, Polycation-π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. *PLoS Comput. Biol.* **9**, e1003239 (2013).
57. C. W. Pak *et al*., Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).
58. M. Feric *et al*., Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
59. H. Zhang *et al*., RNA controls PolyQ protein phase transitions. *Mol. Cell* **60**, 220–230 (2015).
60. N. O. Taylor, M. T. Wei, H. A. Stone, C. P. Brangwynne, Quantifying dynamics in phase-separated condensates using fluorescence recovery after photobleaching. *Biophys. J.* **117**, 1285–1300 (2019).
61. A. Bah *et al*., Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).
62. J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, J.-E. Shea, Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **10**, 1644–1652 (2019).
63. A. Putnam, M. Cassani, J. Smith, G. Seydoux, A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos. *Nat. Struct. Mol. Biol.* **26**, 220–226 (2019).
64. C. S. Lee *et al*., Recruitment of mRNAs to P granules by condensation with intrinsically-disordered proteins. *eLife* **9**, e52896 (2020).