Identifying Sequence Perturbations to an Intrinsically Disordered Protein that Determine Its Phase Separation Behavior

Benjamin S. Schuster[1,a,b], Gregory L. Dignon[1,c,d], 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], 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
[g] Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, 19104
[1] These authors contributed equally to this work
[2] Corresponding author

Abstract (Currently 250/250 words)

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 bio-inspired 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. By altering the wild-type sequence, which contains well-mixed charged residues, to increase charge segregation, we designed sequences with significantly increased phase separation propensity. 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 in spite of these sequence perturbations, the RGG-derived condensates remain liquid-like. Together, these studies advance a predictive framework and identify key biophysical principles of sequence features important to phase separation.

Significance Statement (Currently 120/120 words)

Membraneless organelles are assemblies of highly concentrated biomolecules that form through a liquid-liquid phase separation process. These assemblies are often enriched in intrinsically disordered proteins, 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 predictive 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.

Introduction

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 *C. 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 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, 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, 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. 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 mutating tyrosine to phenylalanine and arginine to lysine, mutations that affect LLPS propensity of FUS[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 impossible 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 independent from material properties. Our combined results elucidate fundamentally new and important sequence determinants of IDP phase separation while demonstrating a computationally-guided approach for studying phase behavior of biomolecular condensates. These results promise a framework toward the rational design of LLPS-enabled IDPs.

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 (see 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 the 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 21-30 (RYVPPHLRGG) having highly enhanced contact probability within the condensed protein assembly (Fig. 1A). 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. Region 21-30 is more prone to interaction, not only with itself but also with many regions of the protein (Fig. 1A).

Interestingly, subregion 21-28 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. 1B, S1A; Supporting Methods). 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 8-residue fragment. Due to the small chain length and net charge, we were unable to observe LLPS for the fragment alone, even at very high concentrations and low temperatures (Movie S1).

We, therefore, conducted simulations to compare how deletion of residues 21-30 vs. other regions of the RGG domain 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 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 WT RGG sequence (Fig. 1C). The Δ21-30 variant shows the greatest reduction of θ-temperature, indicating that it would have the lowest propensity to phase separate. In contrast, most other deletions had little effect or actually raised T_θ. This strongly suggests that the sticky hydrophobic subregion has an important role in phase separation of the LAF-1 RGG domain. We note that many of the deletion sequences have a higher T_θ than the full-length RGG, counter to the expectation that longer chain length generally favors LLPS. We believe this effect in the simulation model can be attributed to a subtle balance between the changes in hydrophobicity, net charge, and SCD rather than a single sequence descriptor (Figure S1B). Given the simplicity of our simulation model and the errors associated with predicting phase separation based solely on T_θ, it is possible our computational framework can distinguish sequences such as Δ21-30 which have more significant changes to LLPS behavior, but cannot capture smaller changes as with the other sequences.

We then tested these predictions experimentally by recombinantly expressing and purifying RGG and its variants (Fig. S1C, S1D). To study protein phase behavior, we used a temperature-dependent turbidity assay, 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. 1D, S2A), characteristic of IDPs rich in polar and charged amino acids[18,41]. Under these experimental conditions, the T_{sat} of WT RGG is approximately 26 °C, whereas the variant with the sticky hydrophobic subregion deleted (Δ21-30) has a phase transition temperature of only approximately 14 °C, representing a decrease of 12 °C. We tested two additional deletion variants, the first having residues 101-110 deleted (Δ101-110) which displayed the highest T_θ value according to simulations (Fig. 1C), and a control sequence having residues 82-91 deleted (Δ82-91), which contains the same number of arginine and tyrosine residues as Δ21-30. Both of these display a more modest reduction of T_{sat}, by roughly 6 °C. These results indicate that the eIF4E-binding motif 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. 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. 1E). Notably, Δ21-30 droplets dissolved within 1 minute 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. 1E). In all cases, the process was reversible in that droplets were able to assemble, disassemble, and reassemble upon cycling the temperature (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 *S. cerevisiae*, a well-established model for studying protein aggregation[42,43], and we used full-length GFP-tagged LAF-1 (Fig. 1F). 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. 1G). We confirmed by western blot that WT LAF-1 and LAF-1 Δ21-30 expressed at similar levels (Fig. S1E). The full-length Δ21-30 variant rapidly formed fluorescent cytoplasmic puncta when cooled to 5 °C, which then rapidly dispersed at 25 °C (Fig. 1H). 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[44,45], 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)[46], and with their critical temperatures (T_c)[44].

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. 2A). 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 minimum possible value for a sequence of the same length and composition is -28.03, when following the constraints set by experimental procedures (Supporting Methods). 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 IDRs of similar helicase proteins such as DDX4, which is more charge-segregated[47], 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[48].

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. 2B, 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. Both shuffled sequences show a drastic increase in the critical temperature compared to WT (Fig. 2C), as well as compaction in single-chain simulations (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. 2D, S2B). Remarkably, whereas WT RGG undergoes LLPS at approximately 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. 1D 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 brightfield microscopy at room temperature (Fig. 2D 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. 2E). This is perhaps unsurprising, as nuclear localization signals characteristically contain stretches of basic amino acids[49,50]. 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 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[30,32,51]. 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. 3A). 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 approximately 26 °C, mutating the tyrosines to phenylalanines lowered transition temperature to approximately 14 °C (Fig. 3B, S2C). To confirm that the Y→F mutant still forms normal protein droplets, we imaged it with brightfield microscopy at 5 °C. We observed that the condensates appeared morphologically identical to WT RGG, with many micrometer-scale protein droplets (Fig. 3B, insets). Even more dramatically, upon mutating all arginines to lysines, we observed no phase separation, even below 5 °C (Fig. 3B, S2C). 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 conduct simulations on a 44-residue region of the LAF-1 RGG domain spanning residues 106-149 ($RGG_{106-149}$). 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 (Fig. S5). We also simulate two variants in which either all the tyrosine residues are mutated to phenylalanine (Y→F $RGG_{106-149}$) or all the arginine residues are mutated to lysine (R→K $RGG_{106-149}$). From single-chain simulations, we find we find that R_g increases in the following order: WT < Y→F < R→K (Fig. 3C). Previous studies provide compelling evidence that chain dimensions or solvent quality can faithfully provide knowledge on protein LLPS[3,40,44,52] – 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. 3B), 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 use 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 Fig. S5A. 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 of cation-π interactions[30,53], particularly between arginine and tyrosine[30]; planar interactions between sp^2 hybridized groups (referred to here as sp^2/π interactions)[54]; electrostatic interactions[16,55]; and hydrophobic and VDW interactions[35] to LLPS. We calculate the average number of intermolecular contacts between the two chains of the different $RGG_{106-149}$ variants (Fig. 3D). 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 normalize

the average number of intermolecular contacts of each type by the average number of intermolecular VDW contacts (Fig. S6B) 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 (Fig. S7). The number of sp2/pi and cation-pi 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 sidechains 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. 3E). Importantly, interactions between arginine-tyrosine sidechains 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 the loss of hydrogen bonding interactions, as phenylalanine sidechain lacks a hydroxyl group, unlike tyrosine. On the other hand, R→K mutations remove sp^2/π interactions due to the removal of the guanidine group present in arginine. These results provide a much-needed mechanistic understanding of the importance of arginine and tyrosine residues to protein LLPS.

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 a UCST 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. 4A) or downward (Fig. 4B). 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. 4A), 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. When deleting residues 21-30, encompassing the eIF4E-binding motif, we find that the phase diagram shifts downward significantly (Fig. 4B), much more so than when deleting other regions of 10 residues (Fig. S8). This also validates the predictions of the computational model, which identified the enhanced interactions within that region. 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. 4B).

In previous work, Wang et al. suggest 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 $\boldsymbol{c_{sat} = k\left(n_{Tyr}n_{Arg}\right)^{-1}}$ where k is a fitting parameter and is equal to 6.5 mM[30]. 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. 4C). 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 2 arginine, and 1 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 (Fig. S9A,B). Linear fits of the data yield similar c_{sat} values (Fig. S9C,D). We find that the equation $\boldsymbol{c_{sat} = k\left(n_{Tyr}n_{Arg}\right)^{-1}}$ poorly predicts the c_{sat} for RGG_{shuf} or $RGG_{shuf\text{-}pres}$ (Fig 4C). 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[56]. 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. 5A). 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. S10A). WT RGG and all the variants examined (RGGΔ21-30, both shuffled versions, and Y→F) exhibited rapid fusion, with droplets of lengthscale ℓ = 2 $\pm$ 0.25 μm fusing with τ < 100 ms (Fig. 5B). Droplet fusion is driven by surface tension γ and slowed by viscosity η, and the timescale of fusion is also proportional to droplet size ℓ, so $\tau \approx \frac{\eta}{\gamma}\ell$ [6,7,57]. By plotting τ against ℓ for tens of droplet fusion events (Fig. 5C), we estimate the ratio η/γ, known as the inverse capillary velocity (Fig. S10B). All the variants tested had η/γ within 3-fold that of WT RGG, and in all cases η/γ < 0.05 s/μm, indicating faster fusion compared to LAF-1 (η/γ = 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. 5D,E). By fitting the FRAP recovery curves to a 3D 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[58]. There are modest differences, notably that the construct with deletion of residues 21-30 (ie lower Tsat that WT) exhibited faster fusion and FRAP recovery compared to that of $RGG_{shuff\text{-}pres}$ (ie highest Tsat of all constructs we tested). Together, the FRAP and fusion experiments demonstrate that these variants maintain dynamic, liquid-like condensates, despite the changes to sequence and phase behavior. Thus, phase behavior – critical concentration and transition temperature – can be modulated mostly independently from the alteration of 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 predictive 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.

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 show 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[59]. More generally, these results suggest that the presence within proteins of functional motifs, such as specific binding motifs[38], may have a non-negligible 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[44,45,60]. Ddx4 features blocks of alternating net charge, and scrambling the blocks to remove charge patterning abolishes phase separation[47]. 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[55]. 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[44]. 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[30,32,61]. 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 interaction[54] 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 co-occur. 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 sidechain 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 approximately 5-fold 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. Future work will continue to explore sequence-to-rheology relationships[30].

Overall, our combined results elucidate the driving forces of LLPS and highlight how the sequence perturbations affect LLPS, promising a framework toward the rational design of LLPS-enabled IDPs. This work will inform studies into the biology of membraneless organelles, aberrant phase transitions in disease, and design of biomaterials and synthetic organelles.

Methods

Cloning

The WT, full-length LAF-1 gene was a gift of Shana Elbaum-Garfinkle and Clifford Brangwynne. WT RGG was amplified by PCR from LAF-1. All modified versions of the RGG domain were ordered as synthetic double-stranded DNA fragments (gBlocks; IDT). Plasmids were constructed using either In-Fusion cloning (Takara Bio) or NEBuilder HiFi DNA Assembly (New England BioLabs). For bacterial expression, genes were cloned into a pET vector in-frame with a C-terminal 6xHis-tag. For yeast expression, genes were cloned into the YIplac211 vector in frame with a C-terminal mEGFP (monomeric enhanced GFP) tag. YIplac211 is a yeast integrating plasmid with a URA3 marker[62]. Gene sequences were verified by Sanger sequencing (GENEWIZ).

Protein expression and purification

For bacterial expression, plasmids were transformed into BL21(DE3) competent *E. coli* (New England BioLabs). Colonies picked from fresh plates were grown for 8 h at 37 °C in 1 mL LB + 1% glucose while shaking at 250 rpm. This starter culture (0.5 mL) was then used to inoculate 0.5 L cultures. Cultures were grown overnight in 2L baffled flasks in Terrific Broth auto-induction medium (Formedium; supplemented with 4 g/L glycerol) at 37 °C while shaking at 250 rpm. The pET vectors used contained a kanamycin resistance gene; kanamycin was used at concentrations of 50 µg/mL in starter cultures and 100 µg/mL in the auto-induction medium[63]. After overnight expression, bacterial cells were pelleted by centrifugation. Pellets were resuspended in lysis buffer (1 M NaCl, 20 mM Tris, 20 mM imidazole, Roche EDTA-free protease inhibitor, pH 7.5) and lysed by sonication. Lysate was clarified by centrifugation at 15,000 g for 30-60 minutes. Lysis was conducted on ice, but other steps were conducted at room

temperature to prevent phase separation. Proteins were purified using an AKTA FPLC with 1 mL nickel-charged HisTrap columns (GE Healthcare Life Sciences) for affinity chromatography of the His-tagged proteins. The column was washed with 500 mM NaCl, 20 mM Tris, 20 mM imidazole, pH 7.5. Proteins were eluted with a linear gradient up to 500 mM NaCl, 20 mM Tris, 500 mM imidazole, pH 7.5. Proteins were dialyzed overnight using 7 kDa MWCO membranes (Slide-A-Lyzer G2, Thermo Fisher) into 500 mM NaCl, 20 mM Tris, pH 7.5 or 150 mM NaCl, 20 mM Tris, pH 7.5. Proteins were dialyzed at temperatures (25 °C - 42 °C) high enough to inhibit phase separation because phase-separated protein bound irreversibly to the dialysis membrane. Proteins were snap frozen in liquid N_2 in single-use aliquots and stored at -80 °C. For turbidity and microscopy experiments, protein samples were prepared as follows: Protein aliquots were thawed above the phase transition temperature. Proteins were then mixed with buffer (20 mM Tris, pH 7.5, 0 – 150 mM NaCl) to obtain solutions containing the desired protein and NaCl concentrations. Protein concentrations were measured based on their absorbance at 280 nm using a Nanodrop spectrophotometer (ThermoFisher). Proteins were mixed in a 1:1 ratio with 8 M urea to prevent phase separation during concentration measurements.

MALDI-TOF mass spectrometry

Molecular weights of purified proteins were measured by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry on an Ultraflextreme mass spectrometer (Bruker). Protein samples were applied as spots to an MPT 384 polished steel target plate. Spots consisted of 1 µL protein solution (approximately 10 µM protein in 50 mM NaCl) plus 1 µL matrix solution (10 mg/mL sinapinic acid dissolved in a 50:50 acetonitrile:water mixture with 0.1% trifluoroacetic acid added).

Turbidity assays

Temperature-dependent turbidity assays were conducted using a UV-Vis spectrophotometer (Cary 100 Bio; Agilent) equipped with a multicell Peltier temperature controller. Protein samples were assayed in quartz cuvettes with 1 cm path length (Thorlabs). Samples were first equilibrated above the phase transition

temperature (25-60 °C depending on the sample) and blanked. Then, the samples were cooled at a rate of 1 °C per minute until reaching 2 °C. Absorbance was measured at λ = 600 nm every 0.5 °C throughout the temperature ramp. Upon cooling below the phase transition temperature, the samples changed from clear to turbid.

SDS-PAGE and western blot

For chromatographically purified proteins, SDS-PAGE was run using NuPAGE 4-12% Bis-Tris gels (Invitrogen) and stained using a Coomassie stain (SimplyBlue SafeStain; Invitrogen). For western blotting, yeast cells were lysed as follows[64]: Cell cultures were pretreated with 2 M lithium acetate for 5 minutes on ice, then with 0.4 M NaOH for 5 minutes on ice. The cell cultures were then resuspended in SDS sample buffer, heated at 95 °C for 5 minutes, and centrifuged to remove cell debris. The supernatant was stored at -80 °C until use. The supernatant was run on a Novex 10% Tris-Glycine gel, WedgeWell format (Invitrogen), then transferred to a nitrocellulose membrane (0.2 µm pore size). The membrane was then incubated with two primary antibodies: rabbit polyclonal antibody to GFP (Invitrogen, catalog #A11122) for detection of the GFP-tagged LAF-1 constructs, and mouse monoclonal antibody to PGK1 (Invitrogen, catalog #459250) as a loading control. Secondary antibodies used for detection were IRDye 680RD goat anti-rabbit IgG (LI-COR, catalog #926-68071) and IRDye 800CW goat anti-mouse IgG (LI-COR, catalog #926-32210). Blots were visualized on a LI-COR Odyssey CLx infrared imaging system.

Yeast transformation and yeast cultures

YIplac211 plasmids were prepared for yeast chromosomal integration by restriction digest with EcoRV, which cuts in the URA3 marker. Linearized plasmids were transformed into *S. cerevisiae* YEF473A strain[65] using the Frozen-EZ Yeast Transformation II Kit (Zymo Research). Transformed yeast cells were cultured at 30°C in uracil dropout synthetic defined medium (-Ura dropout supplement was purchased from Takara Bio). To induce expression of genes under the control of the GAL1 promoter, yeast cultures were first grown overnight in dropout medium + 2% glucose, then grown for 8-10 hours in dropout medium + 2% raffinose, and finally grown overnight in dropout medium + 2% galactose with a target OD_{600} = 0.3 – 0.5 for imaging.

Microscopy: phase behavior, FRAP, and fusion

Imaging of temperature-dependent phase behavior in vitro and in yeast was performed on an Olympus IX81 inverted microscope equipped with a Yokogawa CSU-X1 spinning disk confocal unit and an iXon3 EMCCD camera (Andor). The microscope stage was outfitted with a Cherry Temp microfluidic temperature controller (Cherry Biotech), which enabled imaging samples over the temperature range 5 to 42 °C, with rapid switching (approximately 10 s) between temperature extremes. Imaging was conducted with a 100x/1.4 NA plan-apochromatic oil-immersion objective.

FRAP experiments were performed on a Zeiss Axio Observer 7 inverted microscope equipped with an LSM900 laser scanning confocal module and a 63x/1.4 NA plan-apochromatic oil-immersion objective. LAF-1 RGG and its variants were mixed with 5% of RGG-GFP-RGG, which partitions into the RGG droplets and serves as a FRAP probe[21]. GFP was imaged with a 488 nm laser and bleached with a 405 nm laser. Circular bleach regions of approximate radius R = 1.5 μm were drawn in the center of protein droplets whose radii were at least 2.5R. Recovery curves were fit to an infinite boundary model in three dimensions to calculate the recovery timescale τ[58]. The diffusion coefficient was calculated as $D = R^2/\tau$. The same Zeiss microscope was used for droplet fusion experiments, but using brightfield transillumination and imaging onto an Axiocam 702 sCMOS camera at a frame rate of approximately 62 Hz. Droplet fusion was analyzed by first fitting the image of the fusing droplets to an ellipse and calculating the aspect ratio of the ellipse. The aspect ratio was then plotted against time and the decreasing portion of the curve was fit to an exponential decay to calculate the relaxation time[7,27,57]. The droplet length scale was defined as the radius of the droplet after completion of fusion, when the merged droplet was circular (aspect ratio 1). FRAP and droplet fusion experiments were conducted at room temperature of 16-18 °C using protein concentrations above c_{sat} at that temperature. Image analysis and data processing were performed in MATLAB.

All other imaging was performed on a Leica DMi8 inverted microscope equipped with a spinning disk confocal unit (Spectral Applied Research) and an sCMOS camera (Orca Flash 4.0; Hamamatsu) using a 63x/1.4 NA or 100x/1.4 NA plan-apochromatic oil-immersion objective.

For imaging purified RGG proteins, the protein samples were placed in chambers on glass coverslips (#1.5 glass thickness) that had been passivated for >1 hr by incubation with 5% Pluronic F127 (for FRAP and droplet fusion experiments) or bovine serum albumin. Coated coverslips were thoroughly

rinsed with buffer prior to the addition of RGG protein solutions. For imaging yeast, the glass surface was pretreated by incubation with 0.4 mg/mL concanavalin A (ConA; Sigma) for 5-10 minutes. After removing the ConA solution, yeast was pipetted into the imaging chamber and allowed to settle for several minutes before imaging.

Coarse-grained simulations

Coarse-grained simulations were conducted using an amino-acid-resolution model with 20 residue types to capture sequence specificity, having interactions based on relative hydropathies of each amino acid. Each system was simulated at a range of temperatures using constant volume and temperature using a Langevin thermostat, following similar protocols to our previous work[26]. Simulations of phase coexistence were conducted using HOOMD-Blue v2.1.5 software package[66].

All-atom simulations

Atomic-resolution simulations were conducted for systems containing either one or two copies of a 44-residue fragment of the LAF-1 RGG domain ($RGG_{106-149}$). Simulations were of 44-residue fragments as we have found this size to be computationally tractable for single- and two-chain simulations in the previous studies[5,35]. We selected residues 106-149 by calculating the overall sequence composition of all possible 44-residue fragments and comparing them with the total composition of the 168-residue RGG domain (Fig. S5A). The region having the overall composition most similar to that of the full RGG domain was residues 106-149. Notably, this fragment contains 6 arginine and 3 tyrosine residues constituting 13.6% and 6.8% of the 44-residue sequence, comparable to the 14.3% and 6.5% composition in the full RGG (Fig. S5B,C).

Simulations were conducted with either a single $RGG_{106-149}$ chain solvated in explicit water and ~100 mM NaCl or two chains at the same conditions. We used a modified version of the state-of-the-art Amber99SBws force field[67] with improved residue-specific dihedral corrections (unpublished), tip4p/2005 water[68] and improved salt parameters from Luo and Roux[69]. To efficiently sample the configurational ensemble and contacts between amino acid residues, we employed enhanced sampling using parallel tempering in the well-tempered ensemble (PT-WTE) which couples replica exchange molecular dynamics (REMD)[70] and well-tempered metadynamics[71] applied to the total system energy to enhance fluctuations

and reduces the number of replicas required for good replica exchanges[72]. For two-chain simulations, we also applied a well-tempered metadynamics bias on the intermolecular VDW contacts between heavy nonpolar atoms (i.e. $|q| < 0.25$) as we have done previously to improve sampling of binding and unbinding events[35]. Simulations were conducted using GROMACS 2016 software package[73] with PLUMED 2.4 plugin[74].

We calculated the free energy surface of the two-chain systems from the metadynamics bias using the built-in function (sum_hills) in PLUMED, and an alternative time-independent method from Tiwary and Parrinello[75], then subtract the difference between the two results to generate error bars for Fig. S6A. Contact propensities in all-atom two-chain PT-WTE simulations were reweighted based on free energy surface.

VDW contacts were considered as any two heavy atoms being within 6 Å of each other. Hydrogen bonds were considered as a donor atom and an acceptor atom being within 3 Å and the donor-hydrogen-acceptor angle being larger than 120°. Sp^2/π interactions were calculated as presented by Vernon et al.[54] and considered as any two sp^2-hybridized groups having at least two pairs of atoms being within 4.9 Å and the angle between the normal axes of the two sp2-planes being less than 60°. Cation-π interactions were considered as a cationic atom being within 7 Å of the center of an aromatic ring and less than 60° from the normal axis of the π face. Salt bridges are considered as a cationic atom and an anionic atom being within 6 Å of each other.

Competing Interests

The authors declare no competing interests.

Author Contributions

BSS, GLD, MCG, and JM conceived and designed research. BSS, FMK, AKR, CNJ, and AGS conducted experiments. GLD, WST, and RMR performed and analyzed simulations. BSS, DAH, MCG, and JM supervised research. BSS, GLD, and JM wrote manuscript.

Acknowledgements

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 U.S. Department of

624 Energy, Office of Science, Basic Energy Sciences awards DE-SC0007063 to D.H. (experiments) and DE-
625 SC0013979 to J.M. (theory and simulation). We gratefully acknowledge the use of the high-performance
626 computing capabilities of the Extreme Science and Engineering Discovery Environment (XSEDE), which
627 is supported by NSF grant TG-MCB-120014, and the National Energy Research Scientific Computing
628 Center, supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-
629 05CH11231. B.S. received support from an NIH postdoctoral fellowship (F32GM119430). W.S.T. received
630 support from a National Science Foundation Grant (1845734). M.G. acknowledges support from a
631 National Science Foundation Superseed, NIH R01-EB028320, and Burroughs Wellcome Fund.

632

633

634

635

636

Figures

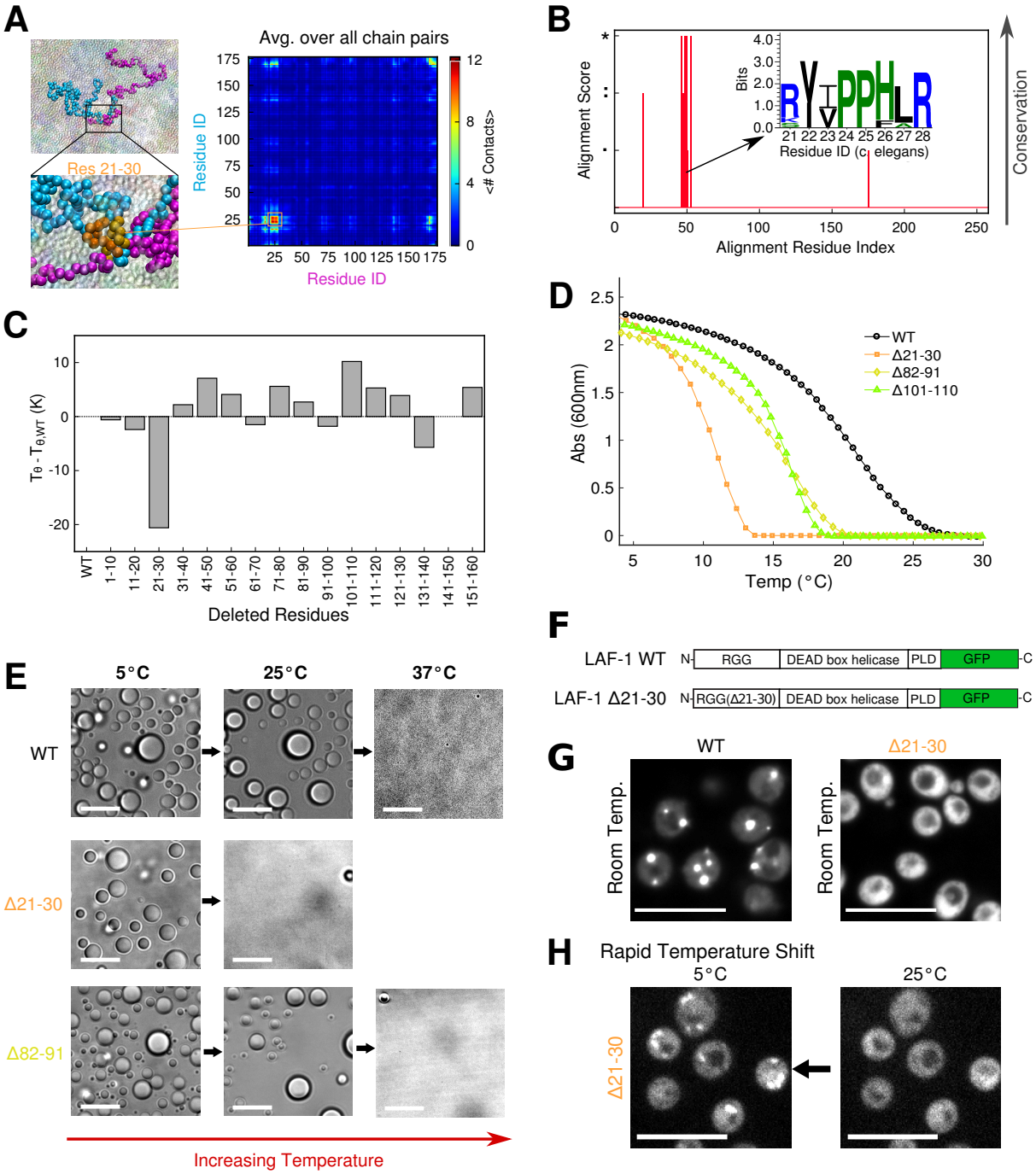


Figure 1: A short segment of LAF-1 RGG is critical for phase separation: A) Coarse-grained sequence-specific simulations of LAF-1 RGG highlight a small region where contact probability is enhanced. Insets show the interaction of two protein chains and zoomed view of contacts between residues within the contact-prone region. B) Sequence analysis of LAF-1 and some of its homologs

644 highlight high sequence conservation in the folded helicase domain, and poor conservation in the
645 disordered RGG and prion-like domains (Fig. S1A). Within the RGG domain, we identify one short
646 region having good conservation, which corresponds to the region highlighted by CG simulations. The
647 amino acids within the sequence are displayed as an inset logo. C) Results of deleting 10 amino
648 segments, scanning across the sequence of RGG; T_θ from CG simulations. Errors are very small and
649 would not show up well on the bar plot. D) Turbidity measurements show temperature-dependent
650 phase behavior of WT RGG vs. variants with deletion of residues 21-30, 82-91, or 101-110. Proteins
651 phase separate upon cooling from above to below the phase transition temperature. Protein
652 concentrations were 1 mg/mL (approximately 60 μM) in 150 mM NaCl buffer, pH 7.5. Data shown is
653 representative of three independent turbidity experiments for each protein (Fig. S2). Similar to previous
654 work[18], we have not averaged the repeats, and therefore, we have not added error bars because the
655 temperatures of the measurements from different replicates are not exactly the same. E) RGG Δ21-30
656 and RGG Δ82-19 condense into spherical liquid droplets, similarly to WT RGG, as shown by brightfield
657 microscopy. Upon heating from 5 °C, RGG Δ21-30 droplets dissolve at a lower temperature compared
658 to WT or RGG Δ82-91. Protein concentration and buffer are the same as for turbidity assay. Scale
659 bars: 10 μm. F) Schematic for full-length LAF-1 constructs including C-terminal GFP fluorescent tag.
660 (PLD: prion-like domain.) G) Full-length LAF-1 phase separates in yeast at room temperature, with
661 multiple puncta per cell. In contrast, LAF-1 Δ21-30 does not phase separate at room temperature;
662 delocalized fluorescence in the cytoplasm is observed. H) Upon sufficient cooling, LAF-1 Δ21-30 does
663 exhibit phase separation in yeast: fluorescent condensates form rapidly upon cooling from 25 °C to 5
664 °C, consistent with in vitro results in (D). Scale bars: 10 μm.
665
666

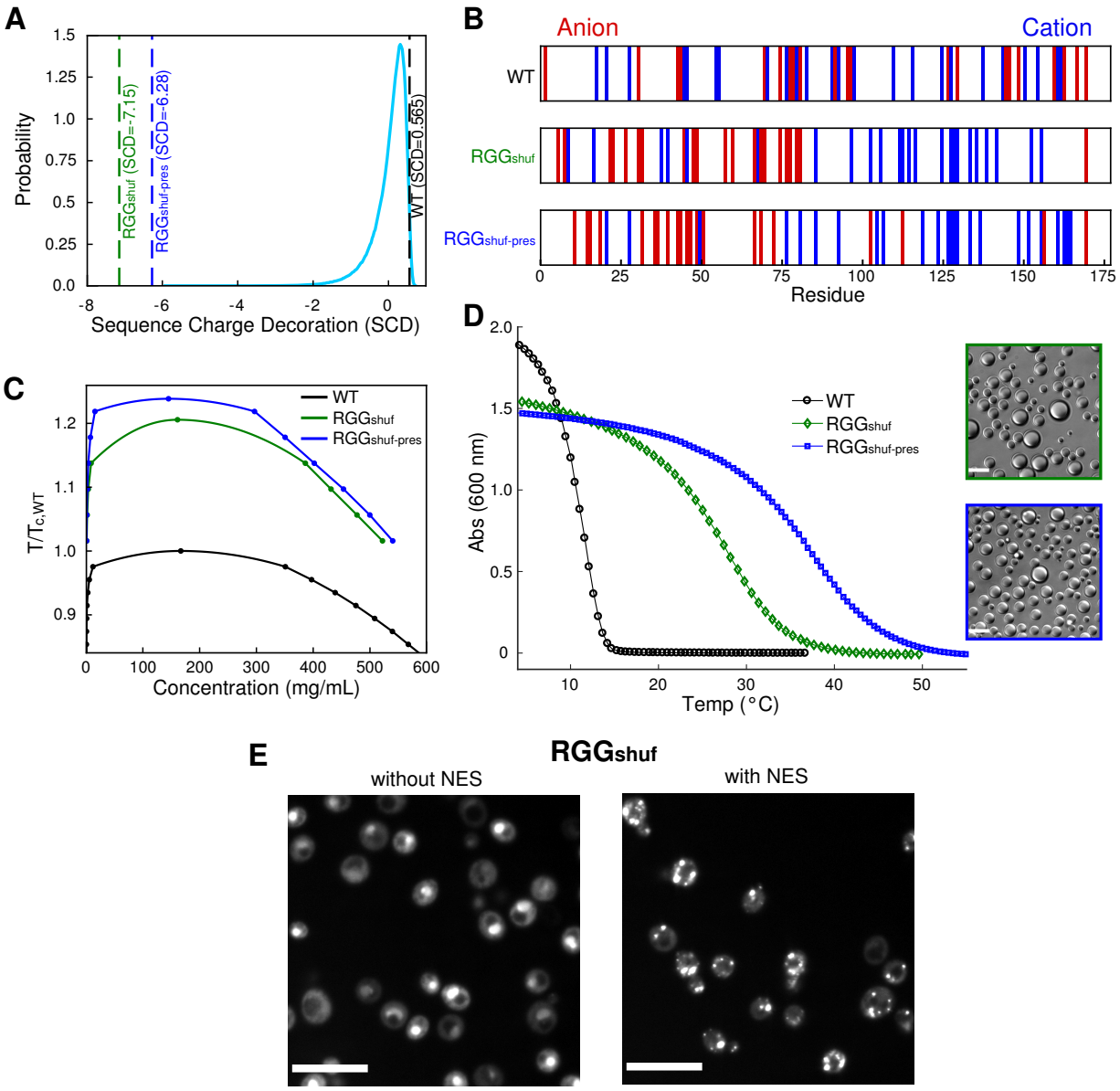


Figure 2: Charge patterning alters LAF-1 RGG phase transition: A) Probability distribution of sequence charge decoration (SCD) values from 1 million random shuffles of LAF-1 RGG. SCD values of WT, RGG_{shuf}, and $RGG_{shuf\text{-}pres}$ are highlighted with dashed lines. B) Location of charged residues 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

675 independent turbidity experiments for each protein (Fig. S2). Protein concentrations were 0.3 mg/mL
676 (approximately 17 μM) in 150 mM NaCl buffer, pH 7.5. Both RGG_{shuf} and $RGG_{shuf\text{-}pres}$ exhibited phase
677 transition temperatures markedly higher than that of WT RGG, and both appeared as liquid droplet
678 condensates under optical microscopy at room temperature (insets; scale bars are 10 μm). E) LAF-
679 1_{shuf}-GFP expression in yeast. Charge patterning leads to constitutive import. The addition of NES
680 enables LAF-1_{shuf}-GFP to be cytosolic, and this variant exhibits protein condensate formation. Scale
681 bars: 10 μm.

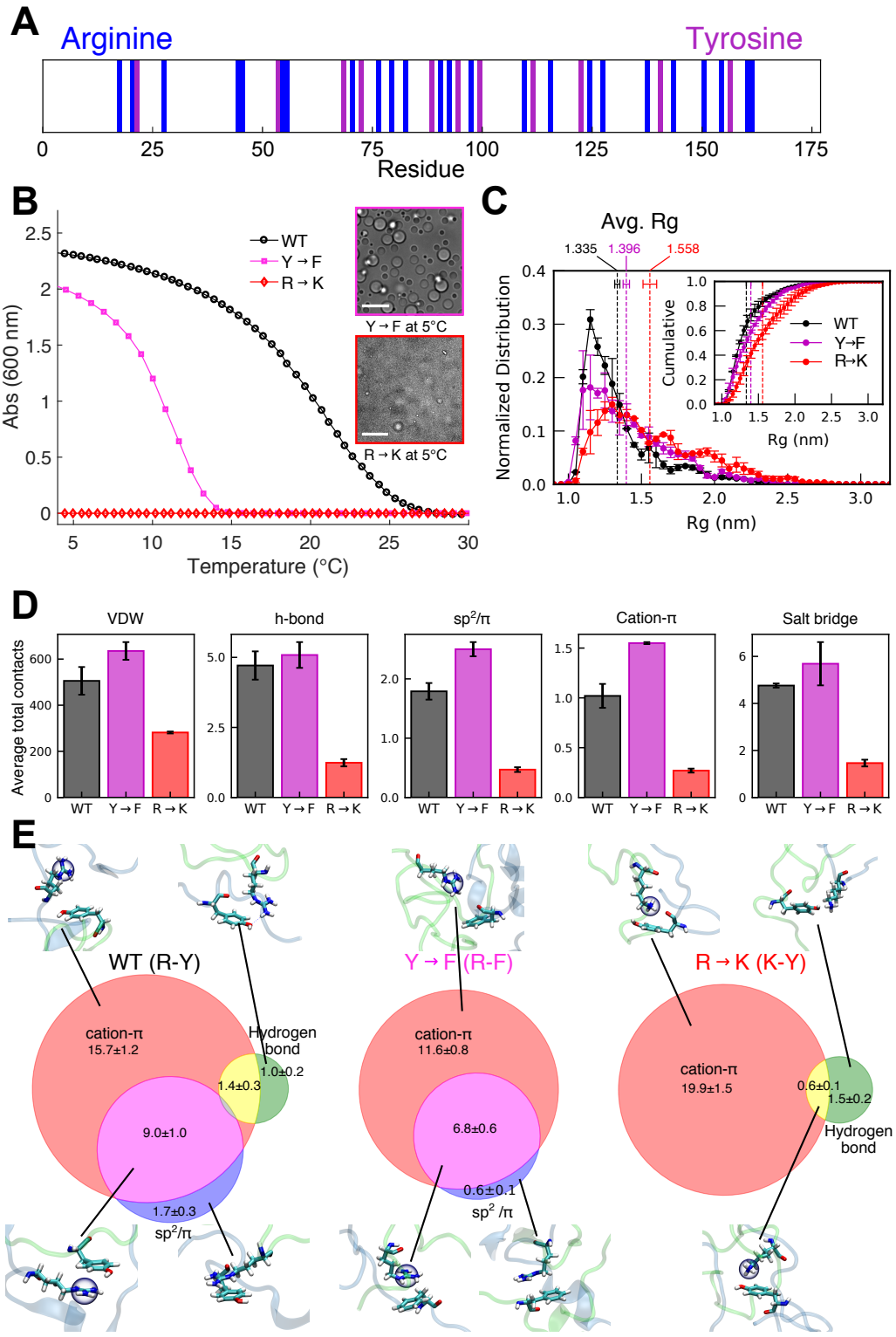


Figure 3: Contribution of arginine and tyrosine residues to LLPS: A) Arrangement of arginine and tyrosine residues along the RGG sequence. Residues are reasonably well-mixed with the exception

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 (Fig. S2). For turbidity assays, protein concentrations were 1 mg/mL (approximately 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 $RGG_{106\text{-}149}$ fragments from single-chain simulations for WT, Y→F, and R→K variants. Inset shows cumulative histogram of R_g. D) Average number of intermolecular contacts observed between two chains of $RGG_{106\text{-}149}$ in two-chain simulations (see Methods), where the average is over the simulated ensemble. Backbone and sidechain 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 sidechain 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.

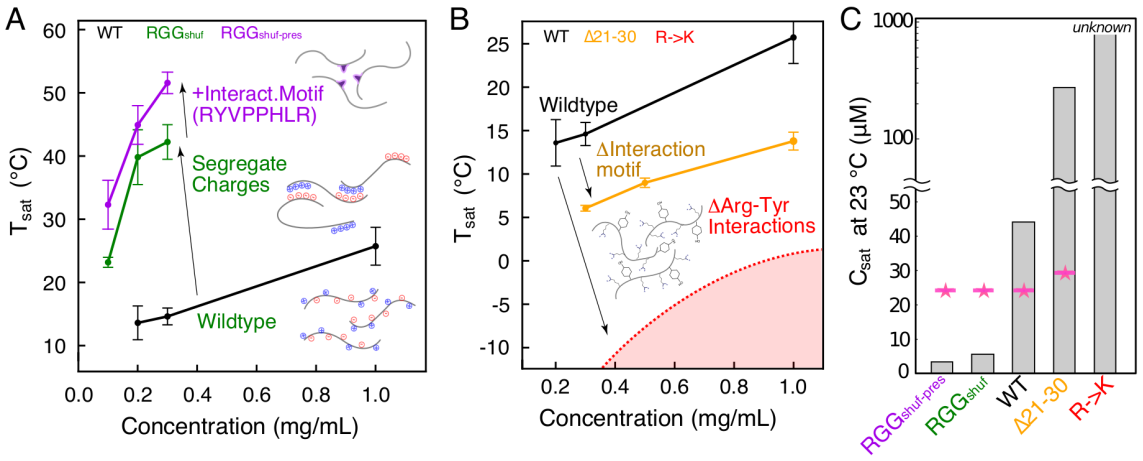


Figure 4: Phase diagrams illustrate molecular interactions that underlie RGG LLPS: Phase diagrams for different LAF-1 variants. T_{sat} values and associated error bars were calculated from triplicates of the turbidity assays at each concentration. A) Shuffled sequences with a high degree of charge patterning 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 than 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 compared with predictions presented in ref [30].

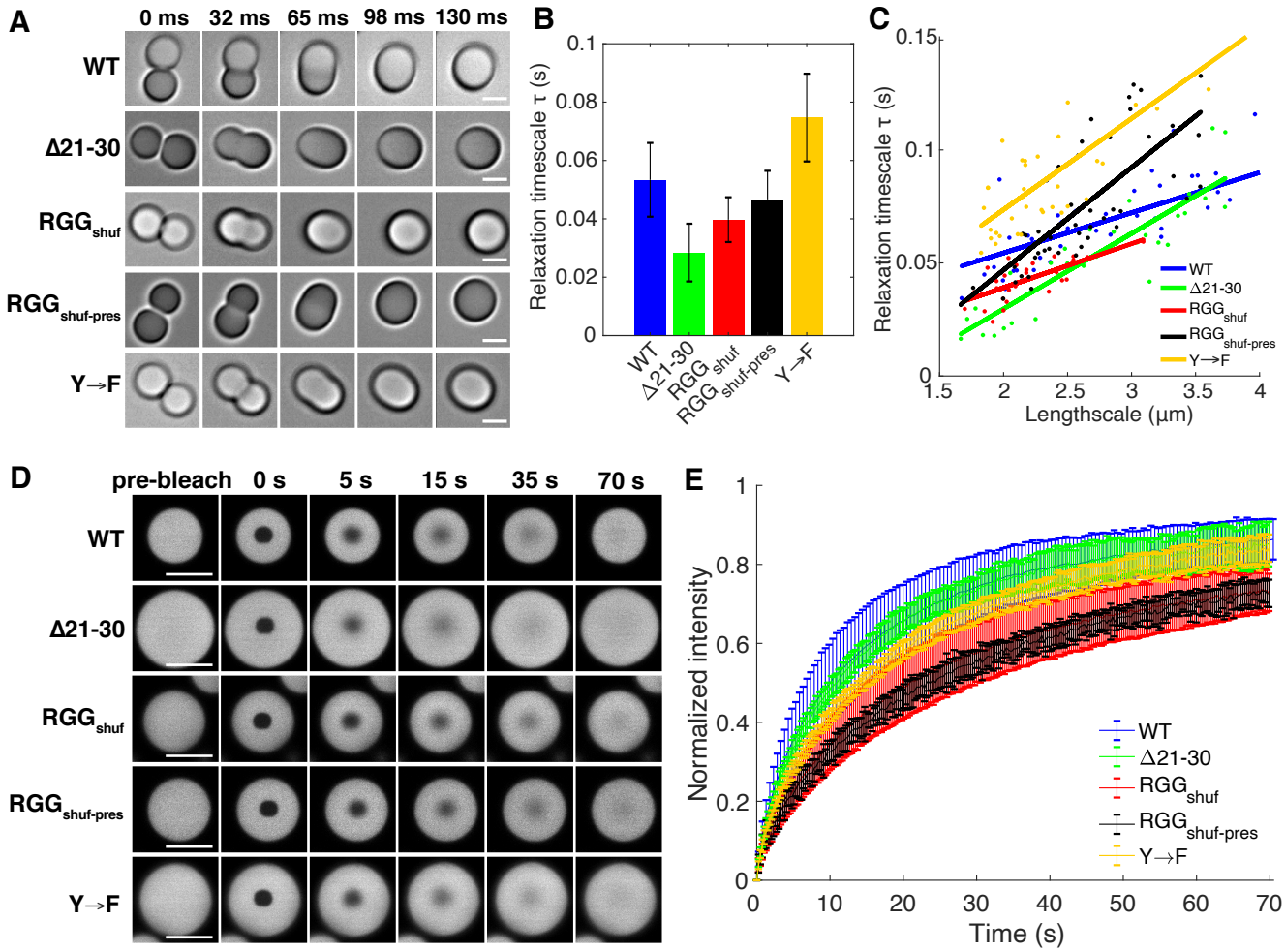


Figure 5: RGG variants exhibit liquid-like material properties. (A) Droplets fuse rapidly to form a single larger sphere. Scale bars: 2 μm. (B) The timescale of droplet fusion for droplets of lengthscale range 1.75-2.25 μm. Error bars represent STD (n ≥ 9). (C) Plots of relaxation timescale against length scale, which allows determination of the inverse capillary velocity. (D) Representative images from FRAP experiments. Scale bars: 5 μm. (E) Normalized FRAP recovery curves show > 50% recovery within 30 s for all variants. Error bars represent STD (n > 15).

References

1. C. Brangwynne, C. Hoege, J. Gharakhani, F. Jülicher, A. A. Hyman Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science (80-.).* **324**, 1729–1732 (2009).
2. Y. Shin, C. P. Brangwynne Liquid phase condensation in cell physiology and disease. *Science (80-.).* **357**, eaaf4382 (2017).
3. J. A. Riback, et al. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028–1040 (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* **39**, 465–479 (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.* **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.* **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**, 1–20 (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**, 1700193 (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.* doi:10.1039/C8SC04253E (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**, 1–10 (2019).

24. H. K. Lau, et al. Microstructured Elastomer-PEG Hydrogels via Kinetic Capture of Aqueous Liquid–Liquid Phase Separation. *Adv. Sci.* **5**, 1–13 (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 (2018).

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 (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. Pomes 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. *Chem. - A Eur. J.* **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.* doi:10.1038/s41467-019-13469-8.

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**, 119–29 (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.* **115**, 9929–9934 (2018).

41. G. L. Dignon, W. Zheng, Y. Kim, J. Mittal Temperature-Controlled Liquid−Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).

42. J. Couthouis, et al. A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl. Acad. Sci.* **108**, 20881–20890 (2011).

43. 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).

44. 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).

45. J. P. Brady, et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ-cell specific protein upon phase separation. *Proc. Natl. Acad. Sci.* **114**, E8194–E8203 (2017).

46. L. Sawle, K. Ghosh A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* **143**, 085101 (2015).

47. T. J. Nott, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).

48. 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).

49. A. Lange, et al. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**, 5101–5 (2007).

50. S. Kosugi, et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284**, 478–85 (2009).

51. S. Qamar, et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **173**, 720–734 (2018).

52. Z. Monahan, et al. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* **36**, 2951–2967 (2017).

53. 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** doi:10.1371/journal.pcbi.1003239 (2013).

54. R. M. Vernon, et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**, 1–48 (2018).

55. C. W. Pak, et al. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **63**, 72–85 (2016).

56. M. Feric, et al. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).

57. H. Zhang, et al. RNA Controls PolyQ Protein Phase Transitions. *Mol. Cell* **60**, 220–230 (2015).

58. 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).

59. A. Bah, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).

60. J. 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).

61. S. Qamar, et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell*, 720–734 (2018).

62. R. D. Gietz, A. Sugino New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene* **74**, 527–534 (1988).

63. F. W. Studier Stable Expression Clones and Auto-Induction for Protein Production in E. coli. *Methods in Molecular Biology (Clifton, N.J.)*, pp 17–32. (2014).

64. T. Zhang, et al. An improved method for whole protein extraction from yeast Saccharomyces cerevisiae. *Yeast* **28**, 795–798 (2011).

65. E. Bi, J. R. Pringle *ZDS1 and ZDS2, Genes Whose Products May Regulate Cdc42p in Saccharomyces cerevisiae* Available at: http://mcb.asm.org/ [Accessed July 24, 2019] (1996).

66. J. A. Anderson, C. D. Lorenz, A. Travesset General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **227**, 5342–5359 (2008).

67. R. B. Best, W. Zheng, J. Mittal Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **10**, 5113–5124 (2014).

68. J. L. Abascal, C. Vega A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505 (2005).

69. Y. Luo, B. Roux Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett.* **1**, 183–189 (2010).

70. Y. Sugita, Y. Okamoto Replica exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).

71. A. Barducci, G. Bussi, M. Parrinello Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 1–4 (2008).

72. M. Bonomi, M. Parrinello Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **104**, 1–4 (2010).

73. B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).

74. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).

75. P. Tiwary, M. Parrinello A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **119**, 736–742 (2015).

76. 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).

1. Supporting Text

1.1 Sequences used in in vitro work (including His tag and XhoI restriction site).

LAF-1 RGG WT [Highlighted residues: 21-30 (red); 82-91 (blue); 101-110 (green)]:

MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD
RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG
YNNRDEGSDN RGSGRSYNND RRDNGGDGLE HHHHHH

LAF-1 RGG Δ21-30:

MESNQSNNGG SGNAALNRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD RGYNDNRDDR
DNRGGSGGYG RDRNYEDRGY NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG YNNRDEGSDN
RGSGRSYNND RRDNGGDGLE HHHHHH

LAF-1 RGG Δ82-91:

MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD
RGYNDNRDDR DDRNYEDRGY NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG
YNNRDEGSDN RGSGRSYNND RRDNGGDGLE HHHHHH

LAF-1 RGG Δ101-110:

MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD
RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG YNNRDEGSDN
RGSGRSYNND RRDNGGDGLE HHHHHH

LAF-1 RGG_{shuf}:

MNNSGDNDRG SGNYGLRNSF GDDGYGDNGN DEGNSGYRNR GLGGDRADEY GNSGGNGDNE AAPNASDRDD
AHYYDSDDYD DGGGGRGSGG AGGGGARGPG SNRAGRYGGG GRRGRGRGNG YNGNRSQRRR GGGRGRGNRG
YRVGNGNGQS GGRNSRGGGG GNGGANYGLE HHHHHH

LAF-1 $RGG_{shuf\text{-}pres}$:

MGGYGYGSSG DGGGDDYGDA **RYVPPHLR**GY GDGAGDDGGD NNDDSDDADR DYNGGLSGGA GGNSGGDGEN
GGDGNGRNNA RSGNNRGGNG NYRYFGANYG AGEGRGRNGQ GGEGSGNNRG GGGRYGRRRR QGSRGGRGSG
GNYGGNSNRS GRAGGRDNNA RNRRRNGSLE HHHHHH

LAF-1 RGG R to K:

MESNQSNNGG SGNAALNKGG KYVPPHLKGG DGGAAAAASA GGDDKKGGAG GGGYKKGGGN SGGGGGGGYD
KGYNDNKDDK DNKGGSGGYG KDKNYEDKGY NGGGGGGGNK GYNNNKGGGG GGYNKQDKGD GGSSNFSKGG
YNNKDEGSDN KGSGKSYNND KKDNGGDGLE HHHHHH

LAF-1 RGG Y to F:

MESNQSNNGG SGNAALNRGG RFVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGFRRGGGN SGGGGGGGFD
RGFNDNRDDR DNRGGSGGFG RDRNWEDRGF NGGGGGGGNR GFNNNRGGGG GGFNRQDRGD GGSSNFSRGG
FNNRDEGSDN RGSGRSFNND RRDNGGDGLE HHHHHH

1.2 LAF-1 homologs used in sequence alignment (accession numbers)

LAF-1, C. elegans (NP_001254859.1):

MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD
RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG
YNNRDEGSDN RGSGRSYNND RRDNGGDGQN TRWNNLDAPP SRGTSKWENR GARDERIEQE LFSGQLSGIN
FDKYEEIPVE ATGDDVPQPI SLFSDLSLHE WIEENIKTAG YDRPTPVQKY SIPALQGGRD LMSCAQTGSG
KTAAFLVPLV NAILQDGPDA VHRSVTSSGG RKKQYPSALV LSPTRELSLQ IFNESRKFAY RTPITSALLY
GGRENYKDQI HKLRLGCHIL IATPGRLIDV MDQGLIGMEG CRYLVLDEAD RMLDMGFEPQ IRQIVECNRM
PSKEERITAM FSATFPKEIQ LLAQDFLKEN YVFLAVGRVG STSENIMQKI VWVEEDEKRS YLMDLLDATG
DSSLTLVFVE TKRGASDLAY YLNRQNYEVV TIHGDLKQFE REKHLDLFRT GTAPILVATA VAARGLDIPN

VKHVINYDLP SDVDEYVHRI GRTGRVGNVG LATSFFNDKN RNIARELMDL IVEANQELPD WLEGMSGDMR
SGGGYRGRGG RGNGQRFGGR DHRYQGGSGN GGGGNGGGGG FGGGGQRSGG GGGFQSGGGG GRQQQQQQRA
QPQQDWWS

DDX3X, H. sapiens (NP_001180345.1):
MSHVAVENAL GLDQQFAGLD LNSSDNQSGG STASKGRYIP PHLRNREATK GFYDKDSSGW SSSKDKDAYS
SFGSRSDSRG KSSFFSDRGS GSRGRFDDRG RSDYDGIGSR GDRSGFGKFE RGGNSRWCDK SDEDDWSKPL
PPSERLEQEL FSGGNTGINF EKYDDIPVEA TGNNCPPHIE SFSDVEMGEI IMGNIELTRY TRPTPVQKHA
IPIIKEKRDL MACAQTGSGK TAAFLLPILS QIYSDGPGEA LRAMKENGRY GRRKQYPISL VLAPTRELAV
QIYEEARKFS YRSRVRPCVV YGGADIGQQI RDLERGCHLL VATPGRLVDM MERGKIGLDF CKYLVLDEAD
RMLDMGFEPQ IRRIVEQDTM PPKGVRHTMM FSATFPKEIQ MLARDFLDEY IFLAVGRVGS TSENITQKVV
WVEESDKRSF LLDLLNATGK DSLTLVFVET KKGADSLEDF LYHEGYACTS IHGDRSQRDR EEALHQFRSG
KSPILVATAV AARGLDISNV KHVINFDLPS DIEEYVHRIG RTGRVGNLGL ATSFFNERNI NITKDLLDLL
VEAKQEVPSW LENMAYEHHY KGSSRGRSKS RFSGGFGARD YRQSSGASSS SFSSSRASSS RSGGGGHGSS
RGFGGGGYGG FYNSDGYGGN YNSQGVDWWG N

DEAD box helicase 3b isoform 5X, D. rerio (XP_005168849.1)
MSHVAVENVH GLDQQLAALD LSSADVQGVT GRRYIPPHLR NKEAAKNDAP GGWDNGRSNG FVNGYHDGRD
NRMNGGSSFA GRGPIRSDRG GRGGFRGKST ASYNPIQPMQ SAGFGYDNKE AGGWNVPKDN AYNSFGGRSD
RGKSSFFNDR GSSSRGRYER GGFGGGGNSR WVEECRDEDW SKPLPPNERL EHELFSGSNT GINFEKYDDI
PVEATGHNGP QPIDRFHDLE MGEIIMGNIN LSRYTRPTPV QKHAIPIIKS KRDLMACAQT GSGKTAAFLL
PVLSQIYTDG PGEALQAAKN SAQENGKYGR RKQYPISLVL APTRELALQI YDEARKFSYR SHVRPCVVYG
GADIGQQIRD LERGCHLLVA TPGRLVDMME RGKIGLDYCN YLVLDEADRM LDMGFEPQIR RIVEQDTMPP
KGLRQTMMFS ATFPKEIQIL ARDFLEDYIF LAVGRVGSTS ENITQKVVWV EENDKRSFLL DLLNATGKDS
LTLVFVETKK GADALEDFLY REGYACTSIH GDRSQRDREE ALHQFRSGRC PILVATAVAA RGLDISNVKH
VINFDLPSDI EEYVHRIGRT GRVGNLGLAT SFFNDKNGNI TKDLLDILVE AKQEVPSWLE SLAYEHQHKS
SSRGRSKRFS GGFGARDYRQ NSSSGGGGFG GRGGRSTGGH GGNRGFGGGG FGNFYSSDGY GGNYSQVDWWG
N

DEAD-box helicase 3 X-linked L homeolog, X. laevis (NP_001080283.1)
MSHVAVENVL NLDQQFAGLD LNSADAESGV AGTKGRYIPP HLRNKEASRN DSNWDSGRGG NGYINGMQDD
RDGRMNGYDR GGYGSRGTGR SDRGFYDREN SGWNSGRDKD AYSSFGSRGE RGKGSLFNDK GSGSRRPDES
RPDGFDGVGN RGNNSSFGRF DRGNSRWSDE RNDEDDWSKP LAPNDRVEQE LFSGSNTGIN FEKYDDIPVD
ATGSNCPPHI ECFQDVDMGE IIMGNIQLTR YTRPTPVQKH AIPIIIGKRD LMACAQTGSG KTAAFLLPIL
SQIYADGPGD AMKHLKDNGR YGRRKQFPLS LVLAPTRELA VQIYEEARKF AYRSRVRPCV VYGGADIGQQ
IRDLERGCHL LVATPGRLVD MMERGKIGLD FCKYLVLDEA DRMLDMGFEP QIRRIVEQDT MPPKGVRQTM
MFSATFPKEI QILARDFLDE YIFLAVGRVG STSENITQKV VWVEEMDKRS FLLDLLNATG KDSLTLVFVE
TKKGADALED FLYHEGYACT SIHGDRSQRD REEALHQFRS GKCPILVATA VAARGLDISN VKHVINFDLP
SDIEEYVHRI GRTGRVGNLG LATSFFNEKN INITKDLLDL LVEAKQEVPS WLENMAYEQH HKSSSRGRSK
SRFSGGFGAK DYRQSSSAGS SFGSSRGGRS SGHGGSRAFG GGYGGFYNSD GYGGNYGGSS QVDWWGN

Belle isoform B, D. melanogaster (NP_001262379.1)
MSNAINQNGT GLEQQVAGLD LNGGSADYSG PITSKTSTNS VTGGVYVPPH LRGGGGNNNA ADAESQGQGQ
GQGQGFDSRS GNPRQETRDP QQSRGGGGEY RRGGGGGGRG FNRQSGDYGY GSGGGGRRGG GGRFEDNYNG
GEFDSRRGGD WNRSGGGGGG GRGFGRGPSY RGGGGGSGSN LNEQTAEDGQ AQQQQQPRND RWQEPERPAG
FDGSEGGQSA GGNRSYNNRG ERGGGGYNSR WKEGGGSNVD YTKLGARDER LEVELFGVGN TGINFDKYED
IPVEATGQNV PPNITSFDDV QLTEIIRNNV ALARYDKPTP VQKHAIPIII NGRDLMACAQ TGSGKTAAFL
VPILNQMYEL GHVPPPQSTR QYSRRKQYPL GLVLAPTREL ATQIFEEAKK FAYRSRMRPA VLYGGNNTSE
QMRELDRGCH LIVATPGRLE DMITRGKVGL ENIRFLVLDE ADRMLDMGFE PQIRRIVEQL NMPPTGQRQT
LMFSATFPKQ IQELASDFLS NYIFLAVGRV GSTSENITQT ILWVYEPDKR SYLLDLLSSI RDGPEYTKDS
LTLIFVETKK GADSLEEFLY QCNHPVTSIH GDRTQKEREE ALRCFRSGDC PILVATAVAA RGLDIPHVKH
VINFDLPSDV EEYVHRIGRT GRMGNLGVAT SFFNEKNRNI CSDLLELLIE TKQEIPSFME DMSSDRGHGG
AKRAGRGGGG RYGGGFGSRD YRQSSGGGGG GRSGPPPRSG GSGSGGGGGS YRSNGNSYGK FGGNSGGGGY
YGGGAGGGSY GGSYGGGSAS HSSNAPDWWA Q

DDX3X-like RNA helicase, E. pallida (XP_020899200.1)
MSHVAPGNQQ SLDQRFAGLD LNSGVGNNPD AGHNQRQQRY VPPHLRRNPQ ELFHNDPRNP VNFPSGGAPQ

QFQGGGRDGA FRGMNYGGKY NNFGGGGGYG GGGGGYGGRG GYGGAGYRRG GGGGNWRERG GNNYWGNNSG
YDDRDSYAKT ARPEDWSKLL PKNDRIEREL FGGHNTGINF EKYDDIPVEA TGQDCPQNIE SFTDVDLGEI
LTHNIQLANY SKPTPVQKYA IPIVKHKRDL MACAQTGSGK TAAFLIPILS RIYQEGPPPA PDAKHTSRRR
QYPVCLVLAP TRELAVQIFD EARKFAYCSL VRPCVVYGGA DIGSQLRELD RGCHLLVATP GRLVDMMDRG
RIGLDVIKFL VLDEADRMLD MGFEPQIRRI VDQDTMPKAG DRQTLMFSAT FPKEIQILAR DFLDNYIFLA
VGRVGSTSEN ITQKIVWVDE YDKRSFLLDL LNASGPDALT LVFVETKKGA DSLELFLYKD GYQCTSIHGD
RSQSEREEAL RSFRSGKTPI LVATAVAARG LDINNVRHVI NFDLPSDIEE YVHRIGRTGR VGHTGLATSF
FNEKNKNVAK DLLSLVTETG QEVPSWLESI AYESNQNSKR GPRRYGGFGG SRDYRQQRGN SAQMNQMHGY
GGYGGGGGGY MHYGGYSGGG GGGGSGGRYH GGGGGGGGQD WWN

Hypothetical protein, M. brevicolis (XP_001747837.1)

MSNGANPNGS DLSQHMADLD LTKTKPSGGS RYVPPHLRNR QPSGPAPPSG GRTAAPPVSA PPPSSNGGGR
DFGSSRPPRG SRDGSRDMGG SRPPRDGGRG GSWDVQPRFQ QEDWTRPLKR NERMEEELFG SNHRTGGINF
EKYDDIPVEA SGNNVPAHIS EFATAGLCEL MTGNLELARY TVPTPVQKYS IPIVQAKRDL MACAQTGSGK
TAAFLVPILN RVYETGPVPP PPNARRSQQF PVALILAPTR ELAIQIYGEA QKFSYRSRVR ICCVYGGASP
RDQIQDLRRG CQLLVATPGR LVDFMERGVI GLDSIRFLVL DEADRMLDMG FEPQIRRIVE EDNMPQVGIR
QTLMFSATFP KDIQMLAQDF LDDYVHLSVG RVGSTSENIQ QIVHWIDEAD KRPSLLDLIS AASSEDLFLI
FVETKKAADA LEYYLTMQGR PATSIHGDRT QYEREEALAD FRAGRRPILV ATAVAARGLD IPNVKHVINF
DLPSDIDEYV HRIGRTGRAG HKGTAVSFFN DKNRNVARDL LN

Dbp1p, S. cerevisiae (AJW08300.1)

MADLPQKVSN LSINNKENGG DGGKSSYVPP HLRSRGKPSF ERSTPKQEDK VTGGDFFRRA GRQTGNNGGF
FGFSKERNGG TSANYNRGGS SNYKSSGNRW VNGKHIPGPK NAKLEAELFG VHDDPDYHSS GIKFDNYDDI
PVDASGKDVP EPILDFSSPP LDELLMENIK LASFTKPTPV QKYSIPIVTK GRDLMACAQT GSGKTGGFLF
PLFTELFRSG PSPVPEKAQS FYSRKGYPSA LVLAPTRELA TQIFEEARKF TYRSWVRPCV VYGGAPIGNQ
MREVDRGCDL LVATPGRLND LLERGKVSLA NIKYLVLDEA DRMLDMGFEP QIRHIVEECD MPSVENRQTL
MFSATFPVDI QHLARDFLDN YIFLSVGRVG STSENITQRI LYVDDMDKKS ALLDLLSAEH KGLTLIFVET
KRMADQLTDF LIMQNFKATA IHGDRTQAER ERALSAFKAN VADILVATAV AARGLDIPNV THVINYDLPS
DIDDYVHRIG RTGRAGNTGV ATSFFNSNNQ NIVKGLMEIL NEANQEVPTF LSDLSRQNSR GGRTRGGGGF
FNSRNNGSRD YRKHGGSGSF GSTRPRNTGT SNWGSIGGGF RNDNEKNGYG SSNASWW

Sum3, S. pombe (NP_588033.1)

MSDNVQQQVD SVGSVTEKLQ KTNISRPRKY IPPFARDKPS AGAAPAVGDD ESVSSRGSSR SQTPSEFSSN
YGGRREYNRG GHYGGGEGRQ NNYRGGREGG YSNGGGYRNN RGFGQWRDGQ HVIGARNTLL ERQLFGAVAD
GTKVSTGINF EKYDDIPVEV SGGDIEPVNE FTSPPLNSHL LQNIKLSGYT QPTPVQKNSI PIVTSGRDLM
ACAQTGSGKT AGFLFPILSL AFDKGPAAVP VDQDAGMGYR PRKAYPTTLI LAPTRELVCQ IHEESRKFCY
RSWVRPCAVY GGADIRAQIR QIDQGCDLLS ATPGRLVDLI DRGRISLANI KFLVLDEADR MLDMGFEPQI
RHIVEGADMT SVEERQTLMF SATFPRDIQL LARDFLKDYV FLSVGRVGST SENITQKVVH VEDSEKRSYL
LDILHTLPPE GLTLIFVETK RMADTLTDYL LNSNFPATSI HGDRTQRERE RALELFRSGR TSIMVATAVA
SRGLDIPNVT HVINYDLPTD IDDYVHRIGR TGRAGNTGQA VAFFNRNNKG IAKELIELLQ EANQECPSFL
IAMARESSFG GNGRGGRYSG RGGRGGNAYG ARDFRRPTNS SSGYSSGPSY SGYGGFESRT PHHGNTYNSG
SAQSWW

1.3 Homolog sequence alignment

```
NP_001254859.1      ------------------MESN--QS----NNGGSGNAALNRGGRYVPPHLRGGDGGAAA        36
NP_001262379.1      MSNAINQNGTGLEQQVAGLDLNGGSADYSGPITSKTSTNSVTGGVYVPPHLRGGGGNNNA        60
XP_001747837.1      MSNGANPNGSDLSQHMADLDLTKTKP--------------SGGSRYVPPHLRNRQPSGPA        46
XP_020899200.1      MSHVAPGNQQSLDQRFAGLDLNSGVGNNP------DAGHNQRQQRYVPPHLRRNPQELFH        54
XP_005168849.1      MSHVAVENVHGLDQQLAALDLSSADV--Q------G----VTGRRYIPPHLRNKEAAKN-        47
NP_001180345.1      MSHVAVENALGLDQQFAGLDLNSSDN--Q------SGGSTASKGRYIPPHLRNREA----        48
NP_001080283.1      MSHVAVENVLNLDQQFAGLDLNSADA--E------SG-VAGTKGRYIPPHLRNKEASRN-        50
AJW08300.1          MAD--------LPQKVSNLS-I---------N-NKENGGDGGKSSYVPPHLRSRGKPSFE        41
NP_588033.1         MSDN-------VQQQVDSVGSV---------TEKLQKTNISRPRKYIPPFARDKPSAGAA        44
                                          :                         *:**. *

NP_001254859.1      AASAGGDDRRGGAGGGGYRRG------------------------GGNS-----------        61
NP_001262379.1      ADAESQGQGQ--GQGQGFDSRSGNPRQETRDPQQSRGGGGEYRRGGGGGGRGFNRQSGDY        118
```

```
1070  XP_001747837.1      ----PPS---GGRTAAPPVS----------------------------------------A        60
1071  XP_020899200.1      NDPRNPVNFPSGGAPQQFQGG-----------------GRDGAFRGMNYGG---------        88
1072  XP_005168849.1      ---DAPGGWDNGR-SNGFVNG-----------------YHDGRDNRMNGGSSF---AGRG        83
1073  NP_001180345.1      ------------------------------------------------------------        48
1074  NP_001080283.1      ---DS--NWDSGRGGNGYING-----------------MQDDRDGRMNGYDR--------        80
1075  AJW08300.1          RSTPKQED-----------------KV--------------------------------        51
1076  NP_588033.1         PAVGDDESVSSR----------GSSRSQ--------------------------------        62
1077
1078  NP_001254859.1      ----GGGGGGG----YDRGYNDNRDD------------------RDNRGGSGGYGRDRNY        95
1079  NP_001262379.1      GYGSGGGGRRGGGGRFEDNY--------NGGEFDSRRGGDWNRSGGGGGGGRGFGRGPSY       170
1080  XP_001747837.1      PPPSSNGG-----------------------------------------GRDFGSSRPP        78
1081  XP_020899200.1      ------------------------------------------------KYNNFGGGGGY        99
1082  XP_005168849.1      PIRSDRGGRGGFRGKSTASYNPIQPMQSAGFGYDNKEAGGWNVPKD--NAYNSFGGRSD-       140
1083  NP_001180345.1      ----------------------------TKGFYDKDSSGWSSSKDK-DAYSSFGSRSDS        78
1084  NP_001080283.1      ------G---GYGSRG---------TGRSDRGFYDRENSGWNSGRDK-DAYSSFGSRGE-       120
1085  AJW08300.1          ----------------------------TGGDF-----------------FR-RAGRQTG        65
1086  NP_588033.1         ----------------------------TPSEF-----------------SSNYGGRREY        77
1087                                                                          .
1088
1089  NP_001254859.1      EDRGYNGGGGGGGNRGYNNNRGGGGGG-----YNRQDRGDGGSSNFSRGGYNNRDEGSDN       150
1090  NP_001262379.1      -----RGGG-----GGSGSNLNEQTAEDGQAQQQQQPRNDRWQEPERPAGFDGSEGG-QS       219
1091  XP_001747837.1      -----RGSR-----D-----------------------------------------G-----        84
1092  XP_020899200.1      -----GGGG-----GGYG----G-------------------------RGGYGGAG-----       116
1093  XP_005168849.1      -----RGKS-----SFFN----DRGSS-----------SRGRYER---------------       160
1094  NP_001180345.1      -----RGKS-----SFFS----DRGSG-----------SRGRFDDRGRSDYDGIG-----       108
1095  NP_001080283.1      -----RGKG-----SLFN----DKGSG-----------SRRP-DESRPDGFDGVG-----       149
1096  AJW08300.1          NNGGFF---------GFSKERNG-------------------------------GT------        81
1097  NP_588033.1         NRGGHYGGG-----EGRQNNYRG-------------------------------GR------        97
1098
1099
1100  NP_001254859.1      RGSGRSYNNDRRDNGGDGQNTRWNNLDA-------PPSRGTSKWENRGARDERIEQELFS       203
1101  NP_001262379.1      AGGNRSYNN-RGERGGGGYNSRWKE-----------GGGSNVDYTKLGARDERLEVELFG       267
1102  XP_001747837.1      ---SRDMGG-SRPPRDGGRGGSWDV----------QPRFQQEDWTRPLKRNERMEEELFG       130
1103  XP_020899200.1      ---YRRGGG-GGNWRERGGNNYWGNNSGYDDRDSYAKTARPEDWSKLLPKNDRIERELFG       172
1104  XP_005168849.1      ------------GGFGGGGNSRWV-----------EEC-RDEDWSKPLPPNERLEHELFS       196
1105  NP_001180345.1      ---SRGDRS-GFGKFERGGNSRWC------------DKSDEDDWSKPLPPSERLEQELFS       152
1106  NP_001080283.1      ---NRGNNS-SFGRFDRGN-SRWS-----------DERNDEDDWSKPLAPNDRVEQELFS       193
1107  AJW08300.1          ---SANYNR-GGSSNYKSSGNRWVN-----------------GKHIPGPKNAKLEAELFG       120
1108  NP_588033.1         ---EGGYSN-GGGYRNNRGFGQWRD-----------------GQHVIGARNTLLERQLFG       136
1109                                      *                          .  :* :**.
1110
1111  NP_001254859.1      GQ------LSGINFDKYEEIPVEATGDDVPQPISLFSDLSLHEWIEENIKTAGYDRPTPV       257
1112  NP_001262379.1      VG------NTGINFDKYEDIPVEATGQNVPPNITSFDDVQLTEIIRNNVALARYDKPTPV       321
1113  XP_001747837.1      SNH----RTGGINFEKYDDIPVEASGNNVPAHISEFATAGLCELMTGNLELARYTVPTPV       186
1114  XP_020899200.1      GH------NTGINFEKYDDIPVEATGQDCPQNIESFTDVDLGEILTHNIQLANYSKPTPV       226
1115  XP_005168849.1      GS------NTGINFEKYDDIPVEATGHNGPQPIDRFHDLEMGEIIMGNINLSRYTRPTPV       250
1116  NP_001180345.1      GG------NTGINFEKYDDIPVEATGNNCPPHIESFSDVEMGEIIMGNIELTRYTRPTPV       206
1117  NP_001080283.1      GS------NTGINFEKYDDIPVDATGSNCPPHIECFQDVDMGEIIMGNIQLTRYTRPTPV       247
1118  AJW08300.1          VHDDPDYHSSGIKFDNYDDIPVDASGKDVPEPILDFSSPPLDELLMENIKLASFTKPTPV       180
1119  NP_588033.1         AVADGTKVSTGINFEKYDDIPVEVSGGDI-EPVNEFTSPPLNSHLLQNIKLSGYTQPTPV       195
1120                              **:*::*::***:.:* :    :  *    : . :  *:  : :  ****
1121
1122  NP_001254859.1      QKYSIPALQGGRDLMSCAQTGSGKTAAFLVPLVNAILQDGPDAVHRSVT------SSGGR       311
1123  NP_001262379.1      QKHAIPIIINGRDLMACAQTGSGKTAAFLVPILNQMYELGHVPPPQST-------RQYSR       374
1124  XP_001747837.1      QKYSIPIVQAKRDLMACAQTGSGKTAAFLVPILNRVYETGPVPPPPNA----------RR       236
1125  XP_020899200.1      QKYAIPIVKHKRDLMACAQTGSGKTAAFLIPILSRIYQEGPPPAPDA--------KHTSR       278
1126  XP_005168849.1      QKHAIPIIKSKRDLMACAQTGSGKTAAFLLPVLSQIYTDGPGEALQAAKNSAQENGKYGR       310
1127  NP_001180345.1      QKHAIPIIKEKRDLMACAQTGSGKTAAFLLPILSQIYSDGPGEALRAM----KENGRYGR       262
1128  NP_001080283.1      QKHAIPIIIGKRDLMACAQTGSGKTAAFLLPILSQIYADGPGDAMKHL----KDNGRYGR       303
1129  AJW08300.1          QKYSIPIVTKGRDLMACAQTGSGKTGGFLFPLFTELFRSGPSPVPEKA-------QSFYS       233
1130  NP_588033.1         QKNSIPIVTSGRDLMACAQTGSGKTAGFLFPILSLAFDKGPAAVPVDQ----DAGMGYRP       251
1131                      ** :** :   ****:*********..**.*:..     *
1132
1133  NP_001254859.1      KKQYPSALVLSPTRELSLQIFNESRKFAYRTPITSALLYGGRENYKDQIHKLRLGCHILI       371
1134  NP_001262379.1      RKQYPLGLVLAPTRELATQIFEEAKKFAYRSRMRPAVLYGGNN-TSEQMRELDRGCHLIV       433
1135  XP_001747837.1      SQQFPVALILAPTRELAIQIYGEAQKFSYRSRVRICCVYGGAS-PRDQIQDLRRGCQLLV       295
1136  XP_020899200.1      RRQYPVCLVLAPTRELAVQIFDEARKFAYCSLVRPCVVYGGAD-IGSQLRELDRGCHLLV       337
1137  XP_005168849.1      RKQYPISLVLAPTRELALQIYDEARKFSYRSHVRPCVVYGGAD-IGQQIRDLERGCHLLV       369
1138  NP_001180345.1      RKQYPISLVLAPTRELAVQIYEEARKFSYRSRVRPCVVYGGAD-IGQQIRDLERGCHLLV       321
1139  NP_001080283.1      RKQFPLSLVLAPTRELAVQIYEEARKFAYRSRVRPCVVYGGAD-IGQQIRDLERGCHLLV       362
```

```
AJW08300.1          RKGYPSALVLAPTRELATQIFEEARKFTYRSWVRPCVVYGGAP-IGNQMREVDRGCDLLV        292
NP_588033.1         RKAYPTTLILAPTRELVCQIHEESRKFCYRSWVRPCAVYGGAD-IRAQIRQIDQGCDLLS        310
                     : :*  *:*:*****  **. *::** * : :  . :***      *::.:  **.::

NP_001254859.1      ATPGRLIDVMDQGLIGMEGCRYLVLDEADRMLDMGFEPQIRQIVECNRMPSKEERITAMF        431
NP_001262379.1      ATPGRLEDMITRGKVGLENIRFLVLDEADRMLDMGFEPQIRRIVEQLNMPPTGQRQTLMF        493
XP_001747837.1      ATPGRLVDFMERGVIGLDSIRFLVLDEADRMLDMGFEPQIRRIVEEDNMPQVGIRQTLMF        355
XP_020899200.1      ATPGRLVDMMDRGRIGLDVIKFLVLDEADRMLDMGFEPQIRRIVDQDTMPKAGDRQTLMF        397
XP_005168849.1      ATPGRLVDMMERGKIGLDYCNYLVLDEADRMLDMGFEPQIRRIVEQDTMPPKGLRQTMMF        429
NP_001180345.1      ATPGRLVDMMERGKIGLDFCKYLVLDEADRMLDMGFEPQIRRIVEQDTMPPKGVRHTMMF        381
NP_001080283.1      ATPGRLVDMMERGKIGLDFCKYLVLDEADRMLDMGFEPQIRRIVEQDTMPPKGVRQTMMF        422
AJW08300.1          ATPGRLNDLLERGKVSLANIKYLVLDEADRMLDMGFEPQIRHIVEECDMPSVENRQTLMF        352
NP_588033.1         ATPGRLVDLIDRGRISLANIKFLVLDEADRMLDMGFEPQIRHIVEGADMTSVEERQTLMF        370
                    ****** *.: :* :.:   .:*******************:**:   *     * * **

NP_001254859.1      SATFPKEIQLLAQDFLKENYVFLAVGRVGSTSENIMQKIVWVEEDEKRSYLMDLLDAT--        489
NP_001262379.1      SATFPKQIQELASDFLS-NYIFLAVGRVGSTSENITQTILWVYEPDKRSYLLDLLSSIRD        552
XP_001747837.1      SATFPKDIQMLAQDFLD-DYVHLSVGRVGSTSENIQQIVHWIDEADKRPSLLDLISAA--        412
XP_020899200.1      SATFPKEIQILARDFLD-NYIFLAVGRVGSTSENITQKIVWVDEYDKRSFLLDLLNAS--        454
XP_005168849.1      SATFPKEIQILARDFLE-DYIFLAVGRVGSTSENITQKVVWVEENDKRSFLLDLLNAT--        486
NP_001180345.1      SATFPKEIQMLARDFLD-EYIFLAVGRVGSTSENITQKVVWVEESDKRSFLLDLLNAT--        438
NP_001080283.1      SATFPKEIQILARDFLD-EYIFLAVGRVGSTSENITQKVVWVEEMDKRSFLLDLLNAT--        479
AJW08300.1          SATFPVDIQHLARDFLD-NYIFLSVGRVGSTSENITQRILYVDDMDKKSALLDLLSA---        408
NP_588033.1         SATFPRDIQLLARDFLK-DYVFLSVGRVGSTSENITQKVVHVEDSEKRSYLLDILHTL--        427
                    ***** :** ** ***. :*:.*:************ * :  : : :*:  *:*:: :

NP_001254859.1      ----GDSSLTLVFVETKRGASDLAYYLNRQNYEVVTIHGDLKQFEREKHLDLFRTGTAPI        545
NP_001262379.1      GPEYTKDSLTLIFVETKKGADSLEEFLYQCNHPVTSIHGDRTQKEREEALRCFRSGDCPI        612
XP_001747837.1      ----SSEDLFLIFVETKKAADALEYYLTMQGRPATSIHGDRTQYEREEALADFRAGRRPI        468
XP_020899200.1      ----GPDALTLVFVETKKGADSLELFLYKDGYQCTSIHGDRSQSEREEALRSFRSGKTPI        510
XP_005168849.1      ----GKDSLTLVFVETKKGADALEDFLYREGYACTSIHGDRSQRDREEALHQFRSGRCPI        542
NP_001180345.1      ----GKDSLTLVFVETKKGADSLEDFLYHEGYACTSIHGDRSQRDREEALHQFRSGKSPI        494
NP_001080283.1      ----GKDSLTLVFVETKKGADALEDFLYHEGYACTSIHGDRSQRDREEALHQFRSGKCPI        535
AJW08300.1          ----EHKGLTLIFVETKRMADQLTDFLIMQNFKATAIHGDRTQAERERALSAFKANVADI        464
NP_588033.1         ----PPEGLTLIFVETKRMADTLTDYLLNSNFPATSIHGDRTQRERERALELFRSGRTSI        483
                          . * *:*****: *. *  :*   .   .:**** .* :**. *  *::.   *

NP_001254859.1      LVATAVAARGLDIPNVKHVINYDLPSDVDEYVHRIGRTGRVGNVGLATSFFNDKNRNIAR        605
NP_001262379.1      LVATAVAARGLDIPHVKHVINFDLPSDVEEYVHRIGRTGRMGNLGVATSFFNEKNRNICS        672
XP_001747837.1      LVATAVAARGLDIPNVKHVINFDLPSDIDEYVHRIGRTGRAGHKGTAVSFFNDKNRNVAR        528
XP_020899200.1      LVATAVAARGLDINNVRHVINFDLPSDIEEYVHRIGRTGRVGHTGLATSFFNEKNKNVAK        570
XP_005168849.1      LVATAVAARGLDISNVKHVINFDLPSDIEEYVHRIGRTGRVGNLGLATSFFNDKNGNITK        602
NP_001180345.1      LVATAVAARGLDISNVKHVINFDLPSDIEEYVHRIGRTGRVGNLGLATSFFNERNINITK        554
NP_001080283.1      LVATAVAARGLDISNVKHVINFDLPSDIEEYVHRIGRTGRVGNLGLATSFFNEKNINITK        595
AJW08300.1          LVATAVAARGLDIPNVTHVINYDLPSDIDDYVHRIGRTGRAGNTGVATSFFNSNNQNIVK        524
NP_588033.1         MVATAVASRGLDIPNVTHVINYDLPTDIDDYVHRIGRTGRAGNTGQAVAFFNRNNKGIAK        543
                    :******:***** :* ****:***:*:::********** *: * *.:*** .* .:

NP_001254859.1      ELMDLIVEANQELPDWLEGMSGDMRSGGGYRGRGGRG----NGQRFGGRDHRYQGGSGNG        661
NP_001262379.1      DLLELLIETKQEIPSFMEDMSSDRGHGGAKRAGRGGG--GRYGGGFGSRDYRQSSGGGGG        730
XP_001747837.1      DLLN----------------------------------------------------------        532
XP_020899200.1      DLLSLVTETGQEVPSWLESIAYESNQNSKRG-----P--RRYGGFGGSRDYRQQRGNSAQ        623
XP_005168849.1      DLLDILVEAKQEVPSWLESLAYEHQHKSSSRG--RSK---RFSGGFGARDYRQNSSSGGG        657
NP_001180345.1      DLLDLLVEAKQEVPSWLENMAYEHHYKGSSRG--RSK--SRFSGGFGARDYRQSSGASSS        610
NP_001080283.1      DLLDLLVEAKQEVPSWLENMAYEQHHKSSSRG--RSK--SRFSGGFGAKDYRQSSSAGSS        651
AJW08300.1          GLMEILNEANQEVPTFLSDLSRQNSRGGRTRGGGGF----FNSRNNGSRDYRKHGGSGSF        580
NP_588033.1         ELIELLQEANQECPSFLIAMARESSFGGNGRGGRYSGRGGRGGNAYGARDFRRPTNSSSG        603
                     *:.

NP_001254859.1      GGG----NGGGGGFGGGG-----------------QRSGGGG---GFQSGG--GGGRQQQ        695
NP_001262379.1      GRSGPPPRSGGSGSGGGGGSYRS---NGNSYGKFGGNSGGGGYYGGGAGGGSYGGSYGGG        787
XP_001747837.1      ----------------------------------------------------------        532
XP_020899200.1      MNQM----HGYGGYGGGGGGYMHY----GGYSG----GGGGGGSGGRYH-----GGGGG-        665
XP_005168849.1      ------------GFGGRGG----RSTGGHGGNR----GFGGGGFGNFYSSDGYGGNY---        694
NP_001180345.1      ------------SFSSSRASSSRSGGGGHGSSR----GFGGGGYGGFYNSDGYGGNYNS-        653
NP_001080283.1      -------------FGSSRG----GRSSGHGGSR----AFG-GGYGGFYNSDGYGGNYGG-        688
AJW08300.1          GSTRP-RNTGTSNWGSIGGGFRND-------------------------NEKNGYG--        610
NP_588033.1         YSSGP-SYSG---Y----GGFESRT-------------------------PHHGNTYN--        628


NP_001254859.1      QQQRAQPQQDWWS-      708
```

```
NP_001262379.1      SASHSSNAPDWWAQ      801
XP_001747837.1      --------------      532
XP_020899200.1      -----GGGQDWWN-      673
XP_005168849.1      ------SQVDWWGN      702
NP_001180345.1      ------QGVDWWGN      661
NP_001080283.1      -----SSQVDWWGN      697
AJW08300.1          -----SSNASWW--      617
NP_588033.1         ----SGSAQSWW--      636
```

1.4 Calculation of minimum possible SCD for sequence with same composition as LAF-1 RGG

To obtain a sequence with the minimum possible SCD value, the charged amino acids must be clustered at the very ends of the sequence with positive charges at one end, negative charges at the other, and uncharged amino acids in between. Since we consider histidine in our model to have a +0.5 charge, the +1 charged amino acids should be at the very end with histidine residues following.

We also must consider that for in vitro studies, the initial methionine residue and the LEHHHHHH tag must be conserved. Thus a sequence with minimum possible SCD is:

MDDDDDDDDDDDDDDDDDEEE.....HRRRRRRRRRRRRRRRRRRRRRRRRLEHHHHHH

with all of the uncharged residues in between the negatively-charged N-terminal, and the positively-charged C-terminal, and having an SCD of -28.032. Note that since D and E have the same charge, any permutation of residues 2-21 would not change the SCD value.

The probability of randomly sampling a sequence with the minimum SCD value can be calculated by considering the number of residues being shuffled as 176 – 9 = 167. Then one must consider the four regions that must be correct:

1. All D and E residues within 2-21
2. All R residues within 145-168
3. H residue at 144
4. All uncharged residues within 22-143

To account for these values and the degeneracies, we can calculate the probability of randomly sampling such a sequence as

$$p_{minSCD} = \frac{n_{DE}! \cdot n_{RK}! \cdot n_H! \cdot (n - n_{DE} - n_{RK} - n_H)!}{n!} = \frac{20! \cdot 24! \cdot 1! \cdot 122!}{167!} = 0.9914x10^{-56}$$

2. SI Figures

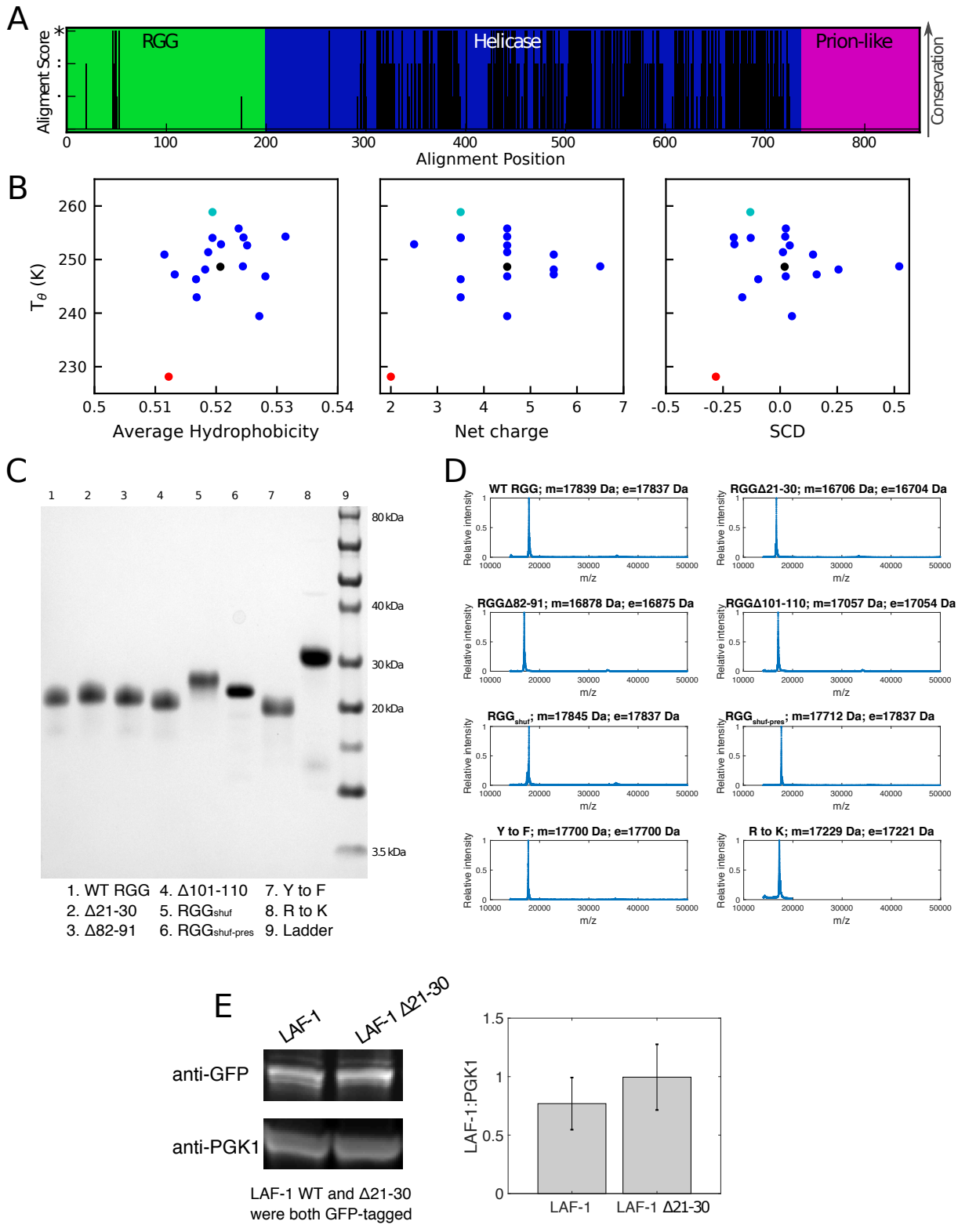


Figure S1: Characterization of deletion variants of LAF-1 RGG: A) Sequence conservation of full-

1254 length LAF-1, showing a high degree of conservation in folded helicase domain, and poor conservation in
1255 disordered RGG and prion-like domains. B) T_θ calculated from single-chain simulations of the deletion
1256 series (Fig. 1C) compared to sequence descriptors. In general, higher T_θ is expected to be associated
1257 with higher average hydrophobicity, smaller absolute net charge, and more negative SCD. The symbol
1258 colors correspond to WT (black), Δ21-30 (red), Δ101-110 (cyan), with all other variants represented as
1259 blue. C) SDS-PAGE gel of purified RGG and its variants. D) MALDI-TOF mass spectra of RGG domain
1260 and its variants, where m denotes measured and e denotes expected molecular mass. The only
1261 discrepancy > 10 Da is $RGG_{shuf\text{-}pres}$ which is likely due to the loss of initiating methionine. E) Western blot
1262 shows a similar expression level of LAF-1 WT and LAF-1 Δ21-30 in yeast.

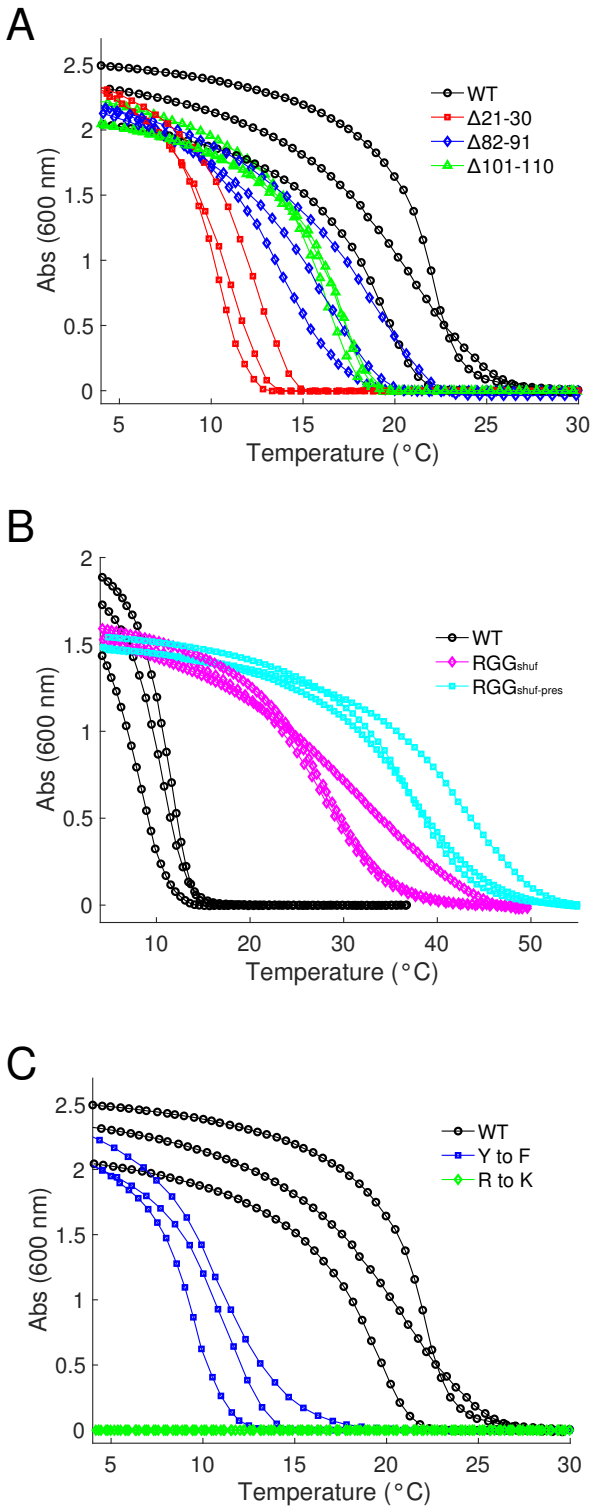


Figure S2: Replicates of turbidity experiments (corresponding to Fig. 1D, 2D, 3B): Turbidity curves of WT and A) deletion variants, B) shuffled sequences, and C) bulk mutations. Protein concentrations were 1 mg/mL for (A) and (C) and 0.3 mg/mL in (B). WT data is the same for (A) and (C). In all cases, proteins

were in 150 mM NaCl, 20 mM Tris, pH 7.5.

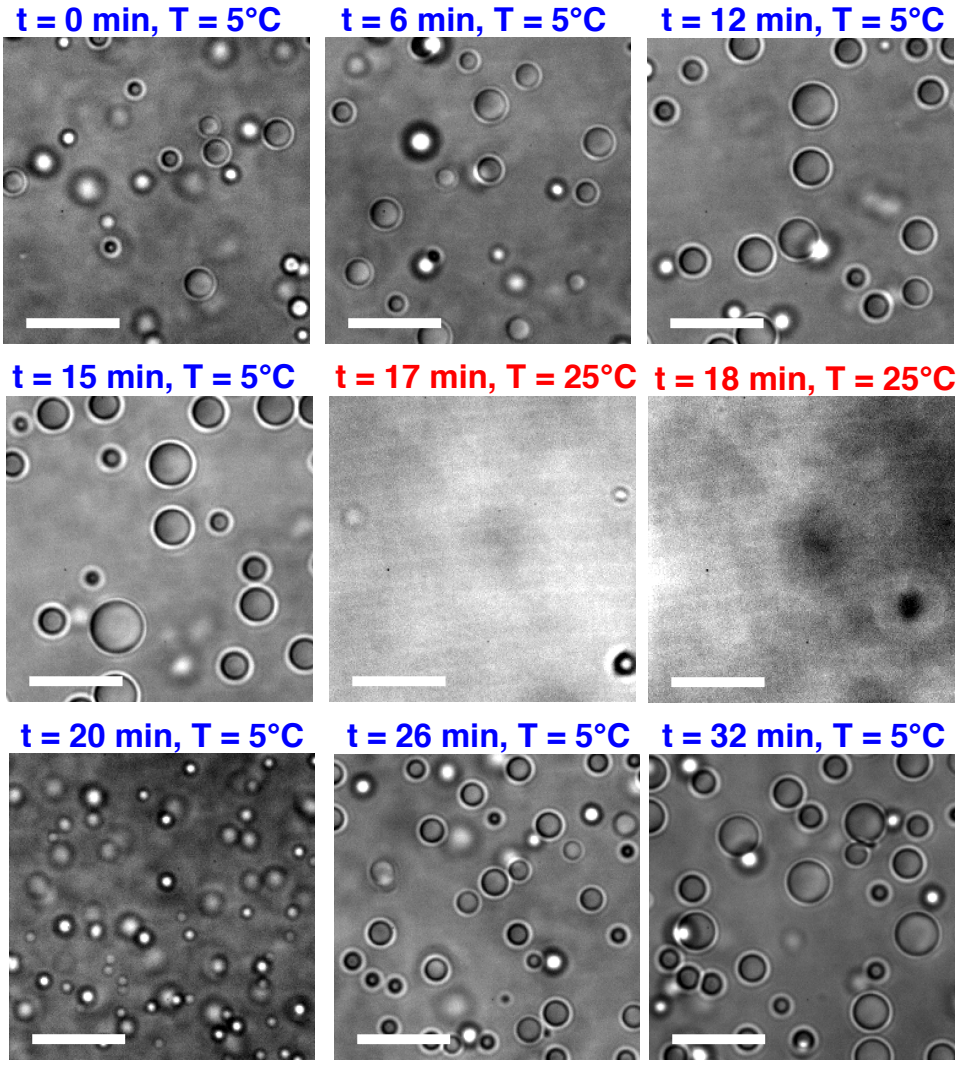


Figure S3: Reversible LLPS of Δ21-30 variant: The Δ21-30 variant of RGG undergoes reversible, temperature-dependent LLPS. Snapshots follow the formation of droplets over time starting at low temperature (5 °C), then rapidly increasing temperature from 5 °C to 25 °C to disperse the droplets, and then rapidly decreasing the temperature back to 5 °C to induce phase separation again. Scale bars: 10 μm.

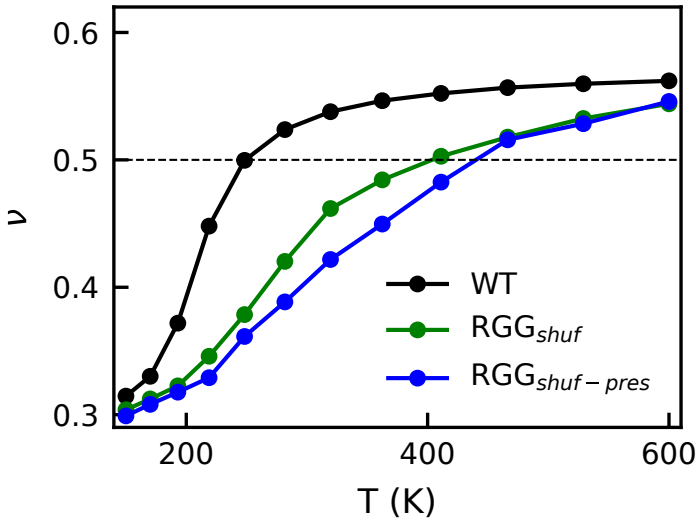


Figure S4: Single-chain compactness of RGG and shuffled variants: We calculate the Flory scaling exponent (v) of the three variants of RGG as in previous work[40,76] and see that the WT is significantly more extended than the shuffled variants at a wide range of temperatures. We also see that $RGG_{shuf-pres}$ is marginally more compact than RGG_{shuf}, consistent with our experimental results showing that $RGG_{shuf-pres}$ has the greatest LLPS propensity.

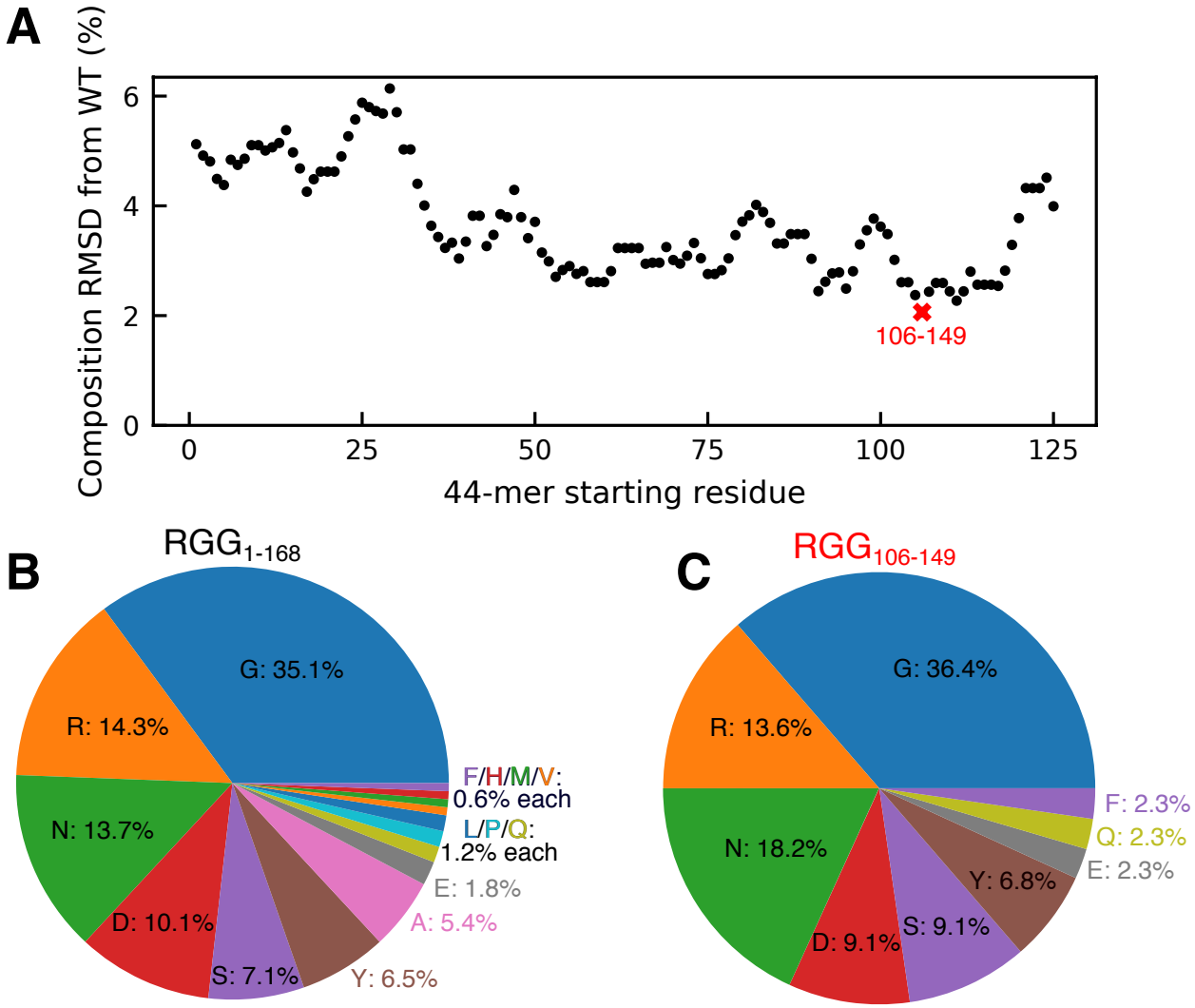


Fig S5: Sequence composition of WT RGG and 44-residue fragments: A) Composition-based RMSD is calculated for all continuous 44-residue fragments of LAF-1 RGG, showing the overall compositional similarity with the full 168-residue sequence. A total of 168 – 44 + 1 = 125 sequences of 44 residues were tested. B) Pie chart of amino acid composition of WT RGG is highly similar to C) pie chart of the lowest-RMSD 44-mer, $RGG_{106\text{-}149}$.

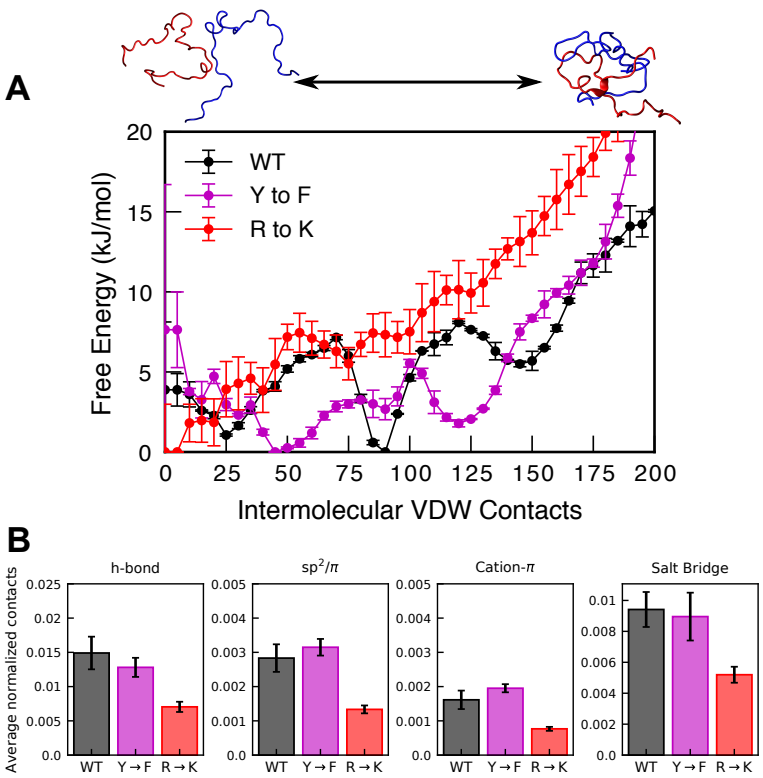


Figure S6: All-atom simulations of RGG$_{106\text{-}149}$ show R→K has lower self-association: A) Free energy profile of contact formation between two identical RGG$_{106\text{-}149}$ chains from simulations using well-tempered metadynamics. Both WT and Y→F show global minima at a finite number of contacts, while R→K has a global minimum at 0 contacts, indicating unfavorable self-interaction. B) Average total number of intermolecular contacts from two-chain simulations normalized by the average total number of VDW contacts for that system. Errorbars for all plots are SEM with n = 2.

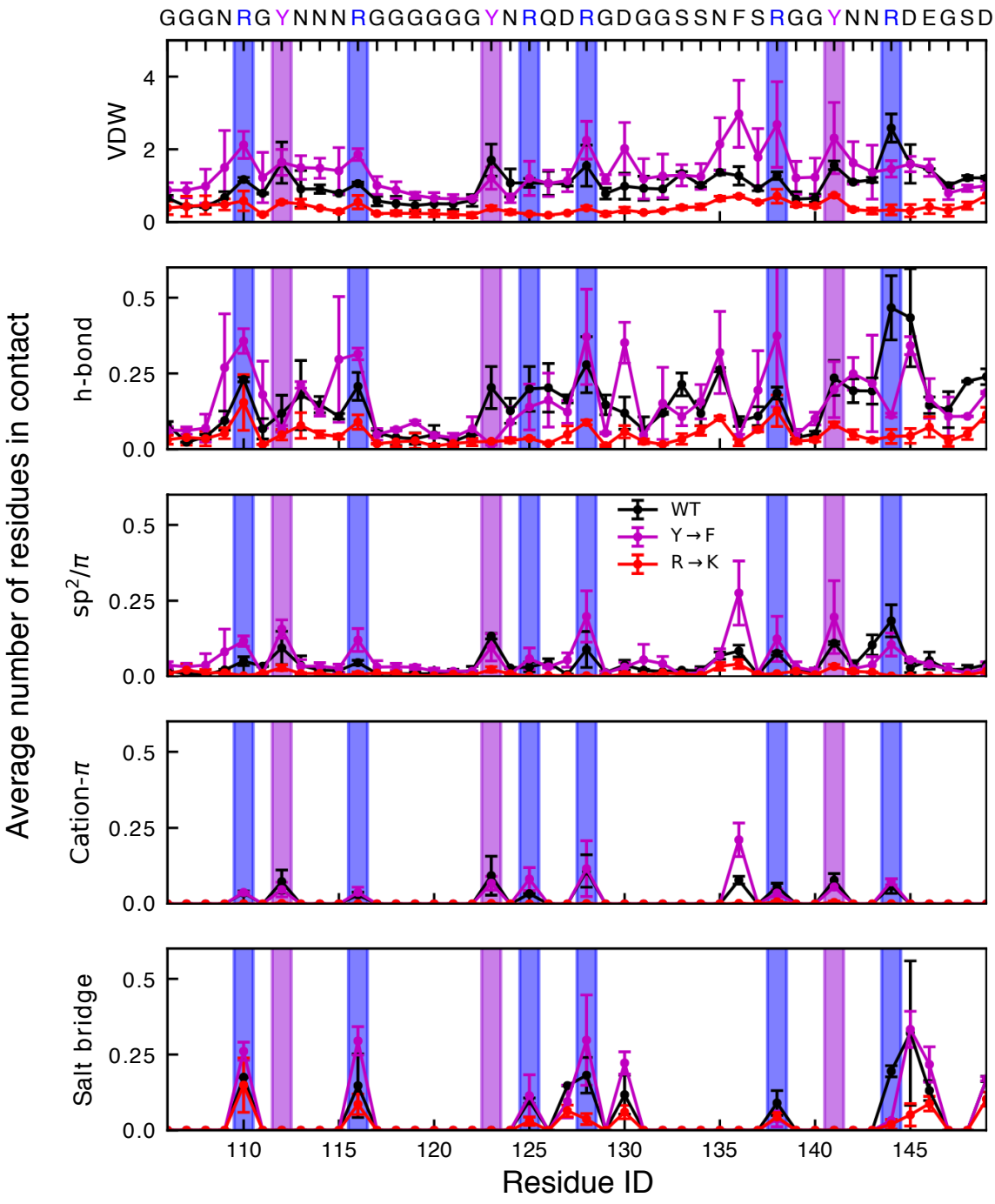


Figure S7: Per-residue contacts from all-atom simulations: Average number of intermolecular residue-residue pairs for each residue of the $RGG_{106\text{-}149}$ sequence. Two residues are considered to be in contact if there is at least one atom from each residue in contact (VDW) or at least one hydrogen bond, sp^2/π interaction, cation-π interaction, or salt bridge between the two residues. In the case of VDW, multiple residues may be in contact with a single residue as only one atom needs to be in contact, and

residues may have VDW interactions with many other amino acids on the other protein chain. Generally, we see that VDW interactions and hydrogen bonds are well-distributed throughout the sequence for all variants of RGG106-149. Cation-pi, sp^2/π, and salt-bridge interactions are less well-distributed due to their dependence on certain amino acid side chains. To highlight the contribution of aromatic and cationic residues, we have highlighted the arginine and tyrosine residues in these plots. Error bars are SEM with n = 2.

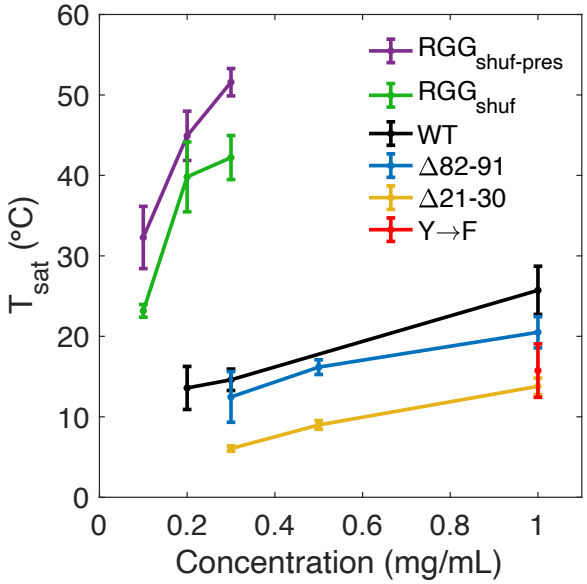


Figure S8: Phase diagrams of RGG mutants: Saturation temperature as a function of total protein concentration for the two shuffled variants, two deletion variants, Y→F, and WT RGG. T_{sat} values are determined from turbidity curves where absorbance first exceeds 0.02. Error bars are STD with n = 3. T_{sat} of WT is significantly different than that of Δ21-30, RGG_{shuf}, and $RGG_{shuf\text{-}pres}$ ($p \leq 0.005$), but not significantly different than that of Δ82-91 ($p = 0.73$), based on one-way ANOVA followed by Tukey's post-hoc test at 0.3 mg/mL. T_{sat} of WT is significantly different than that of Δ21-30 and Y→F ($p \leq 0.005$), but not Δ82-91 ($p = 0.12$), based on one-way ANOVA followed by Tukey's post-hoc test at 1 mg/mL.

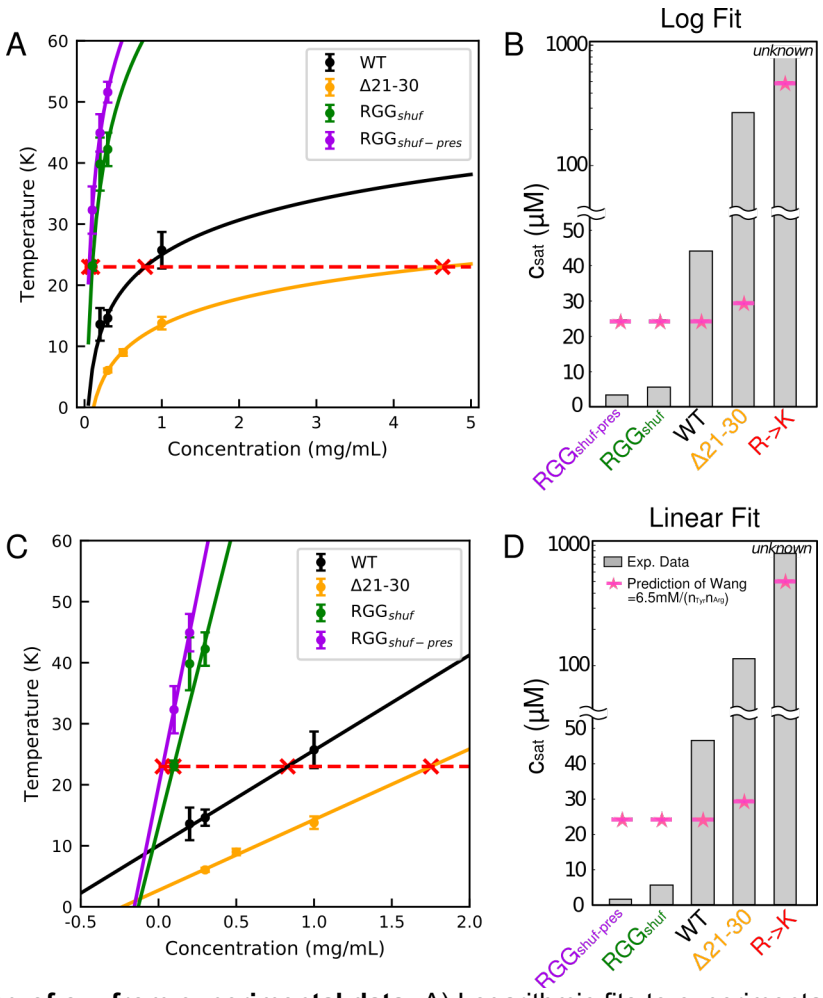
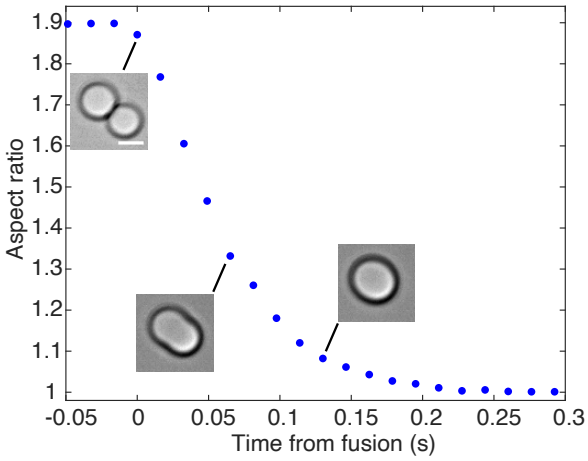


Figure S9: Fitting of c_{sat} from experimental data: A) Logarithmic fits to experimental data to calculate c_{sat} (red X's) at 23°C, (red dashed line). B) bar plot of saturation concentrations for the different variants of RGG and comparison to empirical predictions using relationship from Wang et al. C) Linear fits to experimental data to calculate c_{sat} as before. For $RGG_{shuf\text{-}pres}$, one data point was removed from the fitting such that the extrapolated c_{sat} value would be positive. D) bar plot of saturation concentrations for the different variants of RGG using the linear fit and compared to empirical predictions using relationship from Wang et al. Error bars are STD with n = 3.

A



B

	Fusion: η/γ (s/μm)	FRAP: τ (s)	FRAP: D (μm²/s)
WT	0.018 ± 0.002	112 ± 45	0.025 ± 0.009
Δ21-30	0.033 ± 0.003	113 ± 20	0.022 ± 0.004
RGG_{shuf}	0.019 ± 0.004	238 ± 80	0.011 ± 0.003
$RGG_{shuf\text{-}pres}$	0.045 ± 0.006	244 ± 28	0.010 ± 0.001
Y→F	0.040 ± 0.006	142 ± 13	0.017 ± 0.002

Figure S10: Measurements of droplet material properties: A) Example trace showing aspect ratio of fusing droplets relaxing exponentially to a sphere, from which the relaxation timescale is calculated. The data shown corresponds to the Y→F droplet fusion event in Fig. 5A, several images of which are reproduced here as insets beside their corresponding data points (scale bar: 2 μm). B) Table summarizing measurements of inverse capillary velocity η/γ from droplet fusion experiments, as well as recovery timescale τ and diffusivity D from FRAP.