PHASE SEPARATION

Valence and patterning of aromatic residues determine the phase behavior of prion-like domains

Erik W. Martin[1]*, Alex S. Holehouse[2,3]*, Ivan Peran[1]*, Mina Farag[2,3], J. Jeremias Incicco[4], Anne Bremer[1], Christy R. Grace[1], Andrea Soranno[3,4], Rohit V. Pappu[2,3]†, Tanja Mittag[1]†

Prion-like domains (PLDs) can drive liquid-liquid phase separation (LLPS) in cells. Using an integrative biophysical approach that includes nuclear magnetic resonance spectroscopy, small-angle x-ray scattering, and multiscale simulations, we have uncovered sequence features that determine the overall phase behavior of PLDs. We show that the numbers (valence) of aromatic residues in PLDs determine the extent of temperature-dependent compaction of individual molecules in dilute solutions. The valence of aromatic residues also determines full binodals that quantify concentrations of PLDs within coexisting dilute and dense phases as a function of temperature. We also show that uniform patterning of aromatic residues is a sequence feature that promotes LLPS while inhibiting aggregation. Our findings lead to the development of a numerical stickers-and-spacers model that enables predictions of full binodals of PLDs from their sequences.

Membraneless biomolecular condensates coordinate a variety of cellular processes such as stress responses (1–3), RNA splicing (4), mitosis (5), chromatin organization (6, 7), and the clustering of receptors at membranes (8). Several condensates form through reversible phase transitions that are driven by key protein and RNA molecules (9, 10). Multivalence (i.e., the number of interaction domains or motifs) is a defining hallmark of proteins that drive phase transitions (9). Many of these proteins encompass intrinsically disordered prion-like domains (PLDs) that are often necessary and sufficient for driving intracellular phase transitions (3, 11). PLDs have distinctive amino acid compositions: They are enriched in polar amino acids and are often punctuated by aromatic residues (12). There have been various explanations for how aromatic residues and other polar moieties contribute to phase transitions of PLDs (13, 14). Models based on high-resolution structural studies of hydrogels formed by PLDs suggest that the formation of β-sheeted structural motifs is obligatory for driving phase transitions in PLDs (15–17), whereas other experiments do not detect ordered structures in dense phases (18, 19).

PLDs can be described using a stickers-and-spacers framework adapted from the field of associative polymers (20, 21). These systems are characterized by noncovalent, intra- and intermolecular cross-links between stickers, whereas spacers either enable or suppress the formation of these cross-links (20, 22, 23). Above a threshold concentration known as the percolation threshold, the formation of a criticial number of intermolecular cross-links leads to the emergence of system-spanning networks (22, 24). The percolation threshold can be predicted from knowledge of the number of complementary stickers (21). If percolation is a cooperative process, then it is driven by a density transition known as phase separation (23). In this scenario, the percolation threshold becomes a suitable proxy for the saturation concentration, defined as the threshold concentration for phase separation (21). The cooperativity of percolation and phase separation gives rise to dense-phase condensates that are akin to viscoelastic network fluids (25) in which individual molecules are incorporated into condensate-spanning networks within dense phases that coexist with nonpercolated dilute solutions (23). In the interest of brevity, we shall refer to the combination of percolation and phase separation as liquid-liquid phase separation (LLPS).

Stickers can be patches on folded domains or sequence motifs within disordered regions that can be as small as individual residues (23). Spacers are residues that are interspersed between stickers (25). Previous studies identified arginine and tyrosine as stickers in FUS and other FET family proteins. An analytical model shows that the saturation concentrations of these proteins are inversely proportional to the product of the numbers of arginine and tyrosine residues in a given sequence (21). Although this is useful, a complete characterization of sequence-encoded driving forces for phase transitions requires knowledge of coexistence curves (binodals) of PLDs. Binodals quantify the dilute and dense phase concentrations as a function of temperature or other control variables. This allows us to predict how condensates spontaneously form and dissolve in response to changes to protein concentrations and solution conditions. In this work, we combined a multipronged experimental and computational approach to develop a predictive stickers-and-spacers model for constructing sequence-specific binodals. In doing so, we developed a protocol to identify stickers in an unbiased manner. Furthermore, we quantified how the sticker valence (number), patterning (relative positions along the sequence), and interaction strengths contribute to sequence-specific binodals with a numerical stickers-and-spacers model. Our work is focused on the archetypal PLD or low-complexity domain (LCD) from heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) (A1-LCD) (Fig. 1A and fig. S1), which shares sequence features with PLDs from an assortment of RNA-binding proteins (21).

Analysis of conformational ensembles in dilute solutions allowed us to identify putative stickers within the A1-LCD. First, we used nuclear magnetic resonance (NMR) spectroscopy to interrogate the conformational ensembles of monomeric forms of A1-LCD. To minimize artifacts due to aggregation, we used an A1-LCD construct from which a hexapeptide that acts as a steric zipper (residues 259 to 264) was removed (26). The ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectrum of this A1-LCD construct displays the characteristic narrow proton chemical shift dispersion of disordered proteins (Fig. 1B). Resonance assignment (fig. S2A) and the experimental C^α and C^β chemical shifts demonstrate that the A1-LCD does not form a persistent secondary structure (fig. S2B).

Next, we used size exclusion chromatography–coupled small-angle x-ray scattering (SEC-SAXS) measurements to quantify the dimensions of monomeric A1-LCD in dilute solutions. We also used all-atom simulations based on the ABSINTH implicit solvation model and force-field paradigm (27) to generate ensembles of conformations of the A1-LCD. For flexible polymers in the long-chain limit, the radii of gyration (R_g) may be quantified as: $R_g(T) = R_0(T)N^{\nu(T)}$, where N is the number of repeating units and the prefactor, $R_0(T)$, is determined by the temperature (T)–dependent excluded volume per residue, the average size of each residue, and the thickness of the chain (28). The temperature-dependent solvent quality is characterized by the exponent ν, which takes on limiting values of 0.33 and 0.59 well below and well above the theta temperature T_θ, where $\nu = 0.5$ because polymer-solvent and intrapolymer interactions counterbalance one another. Polymer-solvent interactions are favored above T_θ, whereas intrapolymer interactions are favored below T_θ (12). For systems that undergo continuous globule-to-coil transitions, ν takes on values between 0.33 and 0.5 corresponding to the crossover regime between the poor and theta solvent limits (29).

[1]Department of Structural Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA. [2]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA. [3]Center for Science & Engineering of Living Systems (CSELS), Washington University in St. Louis, St. Louis, MO 63130, USA. [4]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: pappu@wustl.edu (R.V.P.); tanja.mittag@stjude.org (T.M.)
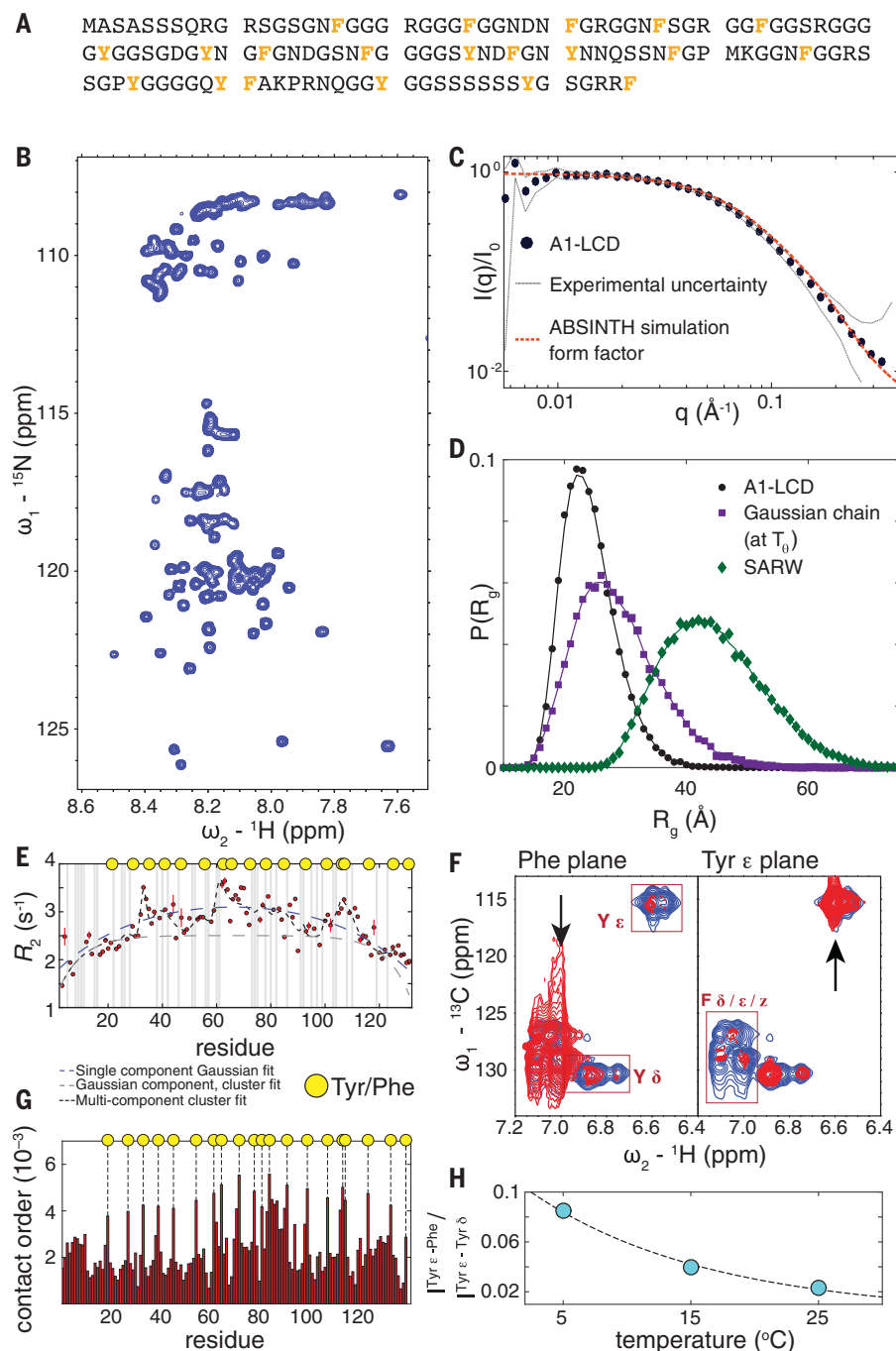
A

```
MASASSSQRG RSGSGNFGGG RGGGFGGNDN FGRGGNFSGR GGFGGSRGGG
GYGGSGDGYN GFGNDGSNFG GGGSYNDFGN YNNQSSNFGP MKGGNFGGRS
SGPYGGGGQY FAKPRNQGGY GGSSSSSSYG SGRRF
```

Fig. 1. Aromatic residues are the stickers in the PLD derived from hnRNPA1. (**A**) The sequence of the PLD or LCD from hnRNPA1 (A1-LCD); aromatic residues are indicated in orange. (**B**) ^{1}H-^{15}N HSQC spectrum recorded at 800 MHz and 25°C in pH 6 MES buffer. For assignments see fig. S2A. ω, chemical shift; ppm, parts per million. (**C**) SEC-SAXS data for A1-LCD. Calculated scattering profiles from simulated ensembles are overlaid in red. $I(q)/I_0$, scattering intensity normalized by zero-angle scattering; q, the momentum transfer vector, which is related to scattering angle. (**D**) R_g distribution from all-atom simulations of A1-LCD (black) versus Gaussian chain (violet) and self-avoiding random walk (SARW, green) reference states. P(R_g), probability density distribution of R_g. (**E**) ^{15}N amide transverse (R_2) relaxation rates recorded at 800 MHz and 25°C. Overlaid are fits to the data assuming a pure Gaussian-like profile (blue dashed line) or multiple regions of enhanced relaxation centered at aromatic residues (black dashed line) and the underlying Gaussian-like profile from this fit (gray dashed line) with a persistence length of 7.8 amino acid residues. The yellow circles indicate the positions of the aromatic residues. Gray bars indicate positions for which data were not analyzed owing to unresolvable overlap in 2D spectra. Monte Carlo sampling of the location of group centers shows a clear positive correlation between the quality of fit and positions of aromatic amino acids within the sequence (fig. S5A). (**F**) ^{13}C-^{1}H planes from the aromatic-edited 3D NOESY (red) recorded at 800 MHz and 25°C. Planes correspond to the Phe ^{1}Hδ/ε/ζ (left) and Tyr ^{1}Hε (right) frequencies and their corresponding diagonal signals are indicated by arrows. In both planes, red boxes show signals at Tyr ^{1}Hδ/^{1}Hε (left) and Phe (right), which exist in this plane only because of NOE transfer. The ^{1}H,^{13}C-HSQC aromatic region is shown superimposed in blue. [For NOEs in a A1-LCD variant with uniformly spaced aromatic residues (Aro^{Perfect}), see fig. S5, E and F.] (**G**) Contact order from simulations. The dashed lines and yellow circles indicate the positions of all aromatic amino acids. (**H**) Normalized intensity of Tyr-Phe NOEs as a function of temperature. The Tyr ^{1}Hε−Phe NOE is displayed normalized to the Tyr ^{1}Hδ−^{1}Hε NOE of fixed distance at 5°, 15°, and 25°C. The dashed line is a power-law fit. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; F, Phe; G, Gly; K, Lys; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; and Y, Tyr.

SEC-SAXS experiments provided a direct measurement of the global dimensions of the A1-LCD (Fig. 1C). Guinier transformation of the SAXS data yields an ensemble-averaged R_g value of 26.1 ± 1.1 Å (fig. S3). We extracted an apparent scaling exponent of v^{app} = 0.45 from the experimental scattering data by fitting to an empirically derived molecular form factor (MFF) (30) (fig. S3). Results from all-atom simulations agree with data from SEC-SAXS experiments (Fig. 1C and fig. S4). The distribution of R_g values obtained for the A1-LCD from all-atom simulations (movie S1) is

biased toward compact conformations when compared with those of self-avoiding random walks and polymers at their theta temperatures (Fig. 1D). These analyses lead us to conclude that the A1-LCD adopts ensembles that lie in the crossover regime between poor and theta solvents.

We used NMR spectroscopy to uncover the putative stickers that determine the features of the conformational ensembles of the A1-LCD. NMR transverse relaxation rates (R_2) are sensitive to internal motions slower than the rotational correlation time and can be used to

identify regions of restricted motion. Fitting the experimentally measured R_2 rates as a function of sequence position within the A1-LCD to a simple Gaussian chain model requires the use of unrealistic values for the persistence length (Fig. 1E). This suggests that groups of residues have enhanced relaxation. Allowing groups of enhanced R_2 rates along the sequence gave good agreement with the experimental data using values for the persistence length that are in line with those reported previously for denatured proteins (31). Fixing the group centers at aromatic residues results

in a qualitatively good fit to the R_2 profile. The R_2 profiles are largely concentration-independent (fig. S5B), suggesting that the observed R_2 rates are dominated by intramolecular interactions. Such interactions were directly observed as nuclear Overhauser effects (NOEs) between phenylalanine and tyrosine side chains in a three-dimensional (3D) aromatic-edited NOESY (NOE spectroscopy) spectrum (Fig. 1F and fig. S5, C to F). These types of long-range NOEs are not typically observed in disordered regions in the absence of well-defined secondary and tertiary structures (32, 33). Given the absence of a persistent secondary or tertiary structure, we interpreted the NOEs to be evidence of transient clustering among aromatic residues, identifying them as the putative stickers in this sequence.

Next, we analyzed the simulated conformational ensembles to quantify the patterns of intramolecular interactions. In accordance with the NMR data (fig. S2B), the ensembles show weak preferences for persistent secondary structure (fig. S6A) and are characterized by interactions among networks of aromatic residues that are distributed along the chain (fig. S6B). Analysis of the contact order shows that many spikes are found at the positions coincident with aromatic residues (Fig. 1G). In addition, we calculated normalized distance maps, which quantify the average distance between pairs of residues normalized by the distance expected in a Gaussian chain. The distance map displays a checkerboard pattern, with pronounced spatial clustering of residues at the C terminus (fig. S6C). The observed pattern of distances indicates that the intramolecular contacts primarily involve aromatic residues, and these are indeed the stickers along the A1-LCD. Charged and polar residues interspersed between stickers do not display strong interaction patterns but instead act as spacers that mediate the contacts among stickers.

The A1-LCD undergoes phase separation with an upper critical solution temperature (3). Given the well-known coupling between the driving forces for phase separation and the determinants of single-chain dimensions (34), we inferred that the contraction of individual A1-LCD molecules in dilute solutions should be temperature dependent. Indeed, NOEs between aromatic residues measured as a function of temperature increased in magnitude when the temperature was lowered (Fig. 1H and fig. S5F). The increase in intensity is clearly visualized when the intensity of the NOE between Tyr $^1H\epsilon$ and Phe protons (which have distinct resonance frequencies) is normalized by the $^1H\epsilon$-$^1H\delta$ NOE within Tyr residues. This is suggestive of the intersticker interactions becoming stronger as temperature decreases, and this in turn promotes compaction as temperature is lowered. The temperature dependence of the interaction strength and protein size is also manifest in the R_2 relax-

ation profiles (fig. S7A) and in the translational diffusion coefficients determined from pulsed field gradient NMR diffusion experiments (fig. S7B).

Our results identify aromatic residues that are distributed along the sequence of A1-LCD as the stickers. To test the accuracy of our inferences regarding the identities of the stickers, we designed variants to quantify how R_g changes
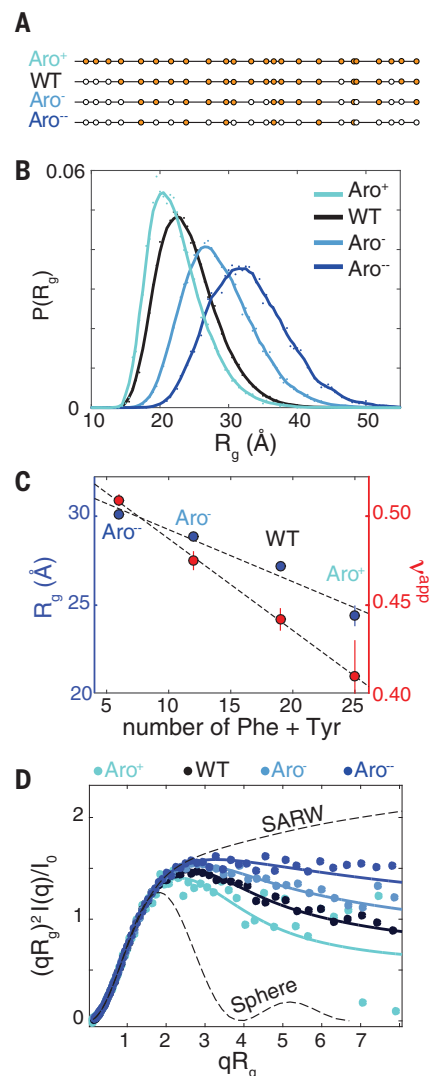

Fig. 2. Sticker valence directly determines the single-chain behavior of the A1-LCD. (**A**) Schematic showing the position of aromatic residues indicated as circles, where orange and white reflect the presence and absence of aromatic residues, respectively. (**B**) The R_g distributions from all-atom simulations of A1-LCD variants. (**C**) Values of R_g (blue) and v^{app} (red) derived from the MFF fits in (D). Dashed lines are the lines of best fit through the four points. (**D**) Raw SEC-SAXS data in normalized Kratky representation (logarithmically smoothed into 60 bins). Solid lines are fits to an empirical MFF (30). The MFFs for a SARW and a solid sphere are overlaid as dashed lines.

with changes to the number (valence) of aromatic residues (Fig. 2A). All-atom simulations indicate a systematic chain expansion that accompanies a decrease in the valence of aromatic residues (variants Aro⁻ and Aro⁻⁻) and further compaction when the valence of aromatic residues increases beyond the wild-type (WT) A1-LCD (variant Aro⁺) (Fig. 2B). SEC-SAXS measurements of the three variants confirm the simulation results (fig. S8); the dependence of the mean R_g and v^{app} values on the valence of aromatic residues shows that they provide the cohesive interactions that drive chain contraction (Fig. 2C). Similarly, comparing normalized Kratky plots confirms the dependence of chain contraction and expansion on the valence of aromatic residues (stickers) (Fig. 2D).

Guided by our observations at the single-chain level, we developed a model for the phase behavior of A1-LCD using a lattice-based coarse-grained description that uses a single bead per residue. In this model, the sticker beads correspond to the aromatic residues, whereas the spacer beads correspond to the nonaromatic residues (Fig. 3A). We generated a single model by parameterizing the strengths of the sticker-sticker, sticker-spacer, and spacer-spacer interactions to reproduce the average R_g values from SEC-SAXS measurements and the R_g distributions obtained from all-atom simulations for the WT and three variant A1-LCD sequences. Simulations using a single, parameterized stickers-and-spacers lattice model reproduced the results obtained from all-atom simulations and SAXS experiments (Fig. 3B).

We used the parameterized stickers-and-spacers lattice model to perform Monte Carlo simulations of hundreds of polymers to quantify phase behavior as a function of simulation temperature that modulates the sticker-sticker, sticker-spacer, and spacer-spacer interaction strengths that are referenced to k_BT, where k_B is the Boltzmann constant and T is temperature. Interaction strengths are inversely proportional to simulation temperature. Simulations of different variants reveal that phase separation occurs in a sequence-, concentration-, and temperature-dependent manner (fig. S9 and movies S2 to S4; note that no phase separation was observed for Aro⁻⁻ at this simulation temperature). Computed binodals are shown in terms of simulation temperatures (in units of k_BT) and volume fractions for the WT A1-LCD, Aro⁻, Aro⁻⁻, and Aro⁺ variants in fig. S10. As the valence of stickers decreases, the location of the low-concentration arm of the binodals shifts to the right, lowering the critical temperature and reducing the overall width of the two-phase regime. By contrast, if the valence of aromatic residues is increased, the opposite changes occur. Accordingly, calculated binodals directly link the valence of aromatic stickers to the phase behavior of A1-LCD and its designed variants.
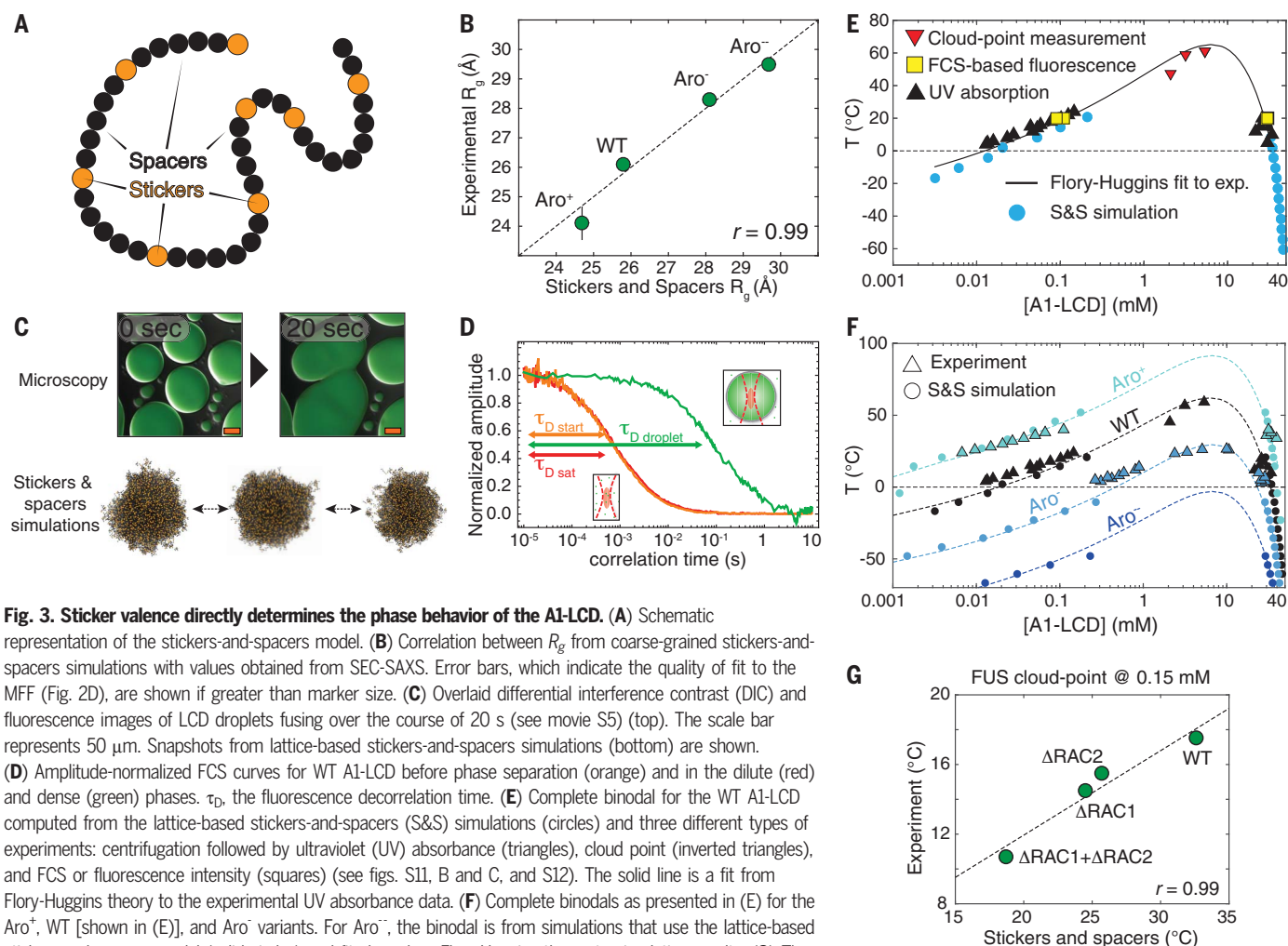
Fig. 3. Sticker valence directly determines the phase behavior of the A1-LCD. (**A**) Schematic representation of the stickers-and-spacers model. (**B**) Correlation between R_g from coarse-grained stickers-and-spacers simulations with values obtained from SEC-SAXS. Error bars, which indicate the quality of fit to the MFF (Fig. 2D), are shown if greater than marker size. (**C**) Overlaid differential interference contrast (DIC) and fluorescence images of LCD droplets fusing over the course of 20 s (see movie S5) (top). The scale bar represents 50 μm. Snapshots from lattice-based stickers-and-spacers simulations (bottom) are shown. (**D**) Amplitude-normalized FCS curves for WT A1-LCD before phase separation (orange) and in the dilute (red) and dense (green) phases. τ_D, the fluorescence decorrelation time. (**E**) Complete binodal for the WT A1-LCD computed from the lattice-based stickers-and-spacers (S&S) simulations (circles) and three different types of experiments: centrifugation followed by ultraviolet (UV) absorbance (triangles), cloud point (inverted triangles), and FCS or fluorescence intensity (squares) (see figs. S11, B and C, and S12). The solid line is a fit from Flory-Huggins theory to the experimental UV absorbance data. (**F**) Complete binodals as presented in (E) for the Aro⁺, WT [shown in (E)], and Aro⁻ variants. For Aro⁻, the binodal is from simulations that use the lattice-based stickers-and-spacers model (solid circles) and fits based on Flory-Huggins theory to simulation results. (**G**) The correlation between the experimentally reported saturation concentrations and those calculated by stickers-and-spacers simulations for WT and three FUS variants with deleted RACs. (*37*). *r*, Pearson correlation coefficient.

To test the predictions from the lattice-based stickers-and-spacers model, we performed in vitro experiments to quantify the temperature-dependent phase behavior of the A1-LCD and designed variants. Monitoring the temperature-dependent, reversible phase separation of the A1-LCD (fig. S11A) provided the basis for accurate mapping of full binodals. Fluorescence microscopy of a small proportion of labeled A1-LCD in the presence of unlabeled A1-LCD showed droplets that diffuse and fuse to form larger droplets (Fig. 3C and movie S5), providing evidence for LLPS. We used fluorescence correlation spectroscopy (FCS) to probe the mobility of protein molecules inside and outside the droplets (Fig. 3D). The increase in the correlation time of the protein molecules reflects the viscosity increase due to the concentration (fig. S11B). Amplitudes of the correlation curve, as well as the fluorescence intensities, allowed us to determine the concentrations and the molecular brightness of the diffusing species in the coexisting dilute and dense phases (Fig.

3E and fig. S11, C to F). The concentration of A1-LCD in its dense phase is approximately three orders of magnitude larger than its concentration in the dilute phase (Fig. 3E and table S1). Analysis of the brightness of the diffusing species indicates that A1-LCD molecules within the droplet are freely diffusing monomers.

Next, we obtained experimentally derived binodals, achieved for only a small number of disordered LCDs (*18*, *35*), by measuring the concentrations (*c*) within coexisting dilute and dense phases as a function of temperature (Fig. 3, E and F). For WT, Aro⁻, and Aro⁺ variants, coexistence points in the (*T*,*c*) space were mapped to quantify the locations of the dilute and dense phase arms of binodals (Fig. 3F and fig. S12A). To locate the critical point, we fit the measured binodals using a modified Flory-Huggins model for phase separation that includes two- and three-body interaction coefficients (*36*) to estimate the critical temperatures (*T*_c) for each system. The quality of the fits shown in Fig. 3E for the WT A1-LCD and in Fig. 3F for

Aro⁻ and Aro⁺ suggest that the PLDs can be approximated as effective homopolymers. Using the predicted values for T_c as a guide, we measured coexistence points close to the predicted critical temperatures using a cloud-point assay (see fig. S12, B and C) and found the predicted values (for WT A1-LCD and Aro⁻) to be within a few degrees of the measured values (Fig. 3, E and F).

The measured binodals of WT A1-LCD were also fit to data from simulations that use the lattice-based stickers-and-spacers model (Fig. 3, E and F). Fits to the experimental data for the WT A1-LCD were used to rescale the simulation temperature to units of degrees Celsius (Fig. 3E) and to convert concentrations from volume fractions into molar units. This allowed us to compare calculated binodals for the WT, Aro⁻, and Aro⁺ sequences to the binodals extracted from experiments (Fig. 3F). These comparisons highlight the phenomenological accuracy of the stickers-and-spacers model. We also calculated the binodal for Aro⁻, and these
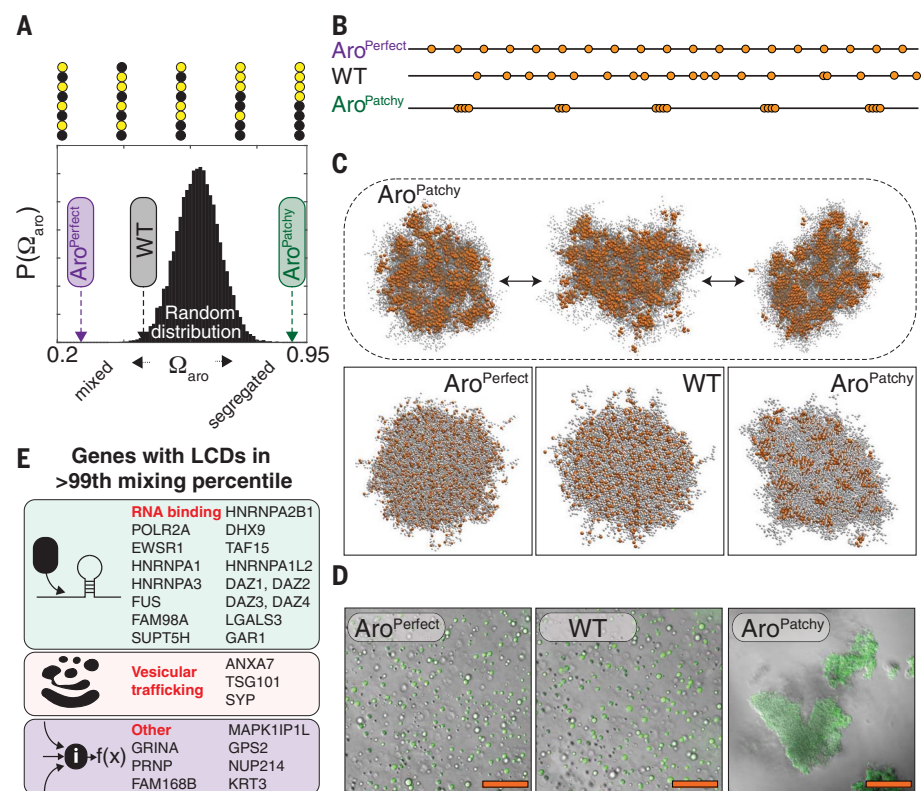
Fig. 4. Linear patterning of stickers versus spacers determines the ability of LCDs to undergo LLPS versus aggregation. (**A**) The aromatic residues in the WT A1-LCD are more uniformly distributed than 99.99% of sequence variants with the same composition as quantified by a mixing parameter Ω_{aro}. The positions of the Aro[Perfect], WT, and Aro[Patchy] LCDs are indicated by arrows on the distribution. (**B**) A schematic showing the positions of aromatic amino acids as orange circles in the Aro[Perfect], WT, and Aro[Patchy] LCDs. (**C**) Snapshots from stickers-and-spacers simulations of Aro[Perfect], WT, and Aro[Patchy] LCDs. The Aro[Patchy] LCD forms amorphous structures (top), whereas Aro[Perfect] and WT both form spherical droplets (bottom). Stickers are orange, and spacers are either gray (bottom) or transparent (top). (**D**) Overlaid DIC and fluorescence images of Aro[Perfect], WT, and Aro[Patchy] LCDs at identical concentrations and solution conditions. The scale bar represents 50 μm. (**E**) Functional annotation of proteins with PLDs that have similarly well-mixed distributions of aromatic residues.

effective homopolymers. Accordingly, we asked if there was a general sequence pattern that characterizes PLDs and LCDs with aromatic stickers. Using a patterning parameter Ω_{aro} ($0 \leq \Omega_{aro} \leq 1$) (see methods), we performed a statistical analysis to determine how likely it would be for the evenly spaced aromatic residues observed in the A1-LCD to occur by random chance. This analysis was motivated by previous studies that connected sequence patterns to changes in conformational features of disordered regions and driving forces for phase separation (14, 40–42). We found that aromatic residues are more uniformly spaced than 99.99% of randomly generated sequences (Fig. 4A). This suggests a clear bias toward a uniform, nonrandom patterning of aromatic stickers within the A1-LCD sequence.

To test the impact of the apparent preference for uniform distribution of stickers along the linear sequence, we used the parameterized version of the stickers-and-spacers model and performed simulations for two variants, Aro[Perfect] and Aro[Patchy] (Fig. 4B). These sequences are of identical composition when compared with the A1-LCD; they are distinguished by the patterning of stickers and hence values of Ω_{aro}. The simulations show that increased linear clustering of stickers in Aro[Patchy] leads to the formation of micellar substructures within the droplet (Fig. 4C and movie S6). Theories predict that these micelles can aggregate to form amorphous precipitates as opposed to liquid-like droplets (43) because the increased linear clustering of stickers increases the apparent intersticker interaction strengths. In contrast to Aro[Patchy], Aro[Perfect] forms spherical droplets that are indistinguishable from WT (Fig. 4C).

Our results suggest that increasing Ω_{aro} by clustering stickers together along the linear sequence will affect the interplay between LLPS and aggregation, with the latter becoming prominent as Ω_{aro} increases. We tested this prediction by performing fluorescence microscopy measurements using a small proportion of labeled molecules in the presence of unlabeled versions for the two designed variants. Whereas Aro[Perfect] formed spherical droplets (Fig. 4D), Aro[Patchy] formed large amorphous aggregates (Fig. 4D). Importantly, we observed aggregation of Aro[Patchy] even under conditions in which the WT A1-LCD remains in the one-phase regime (fig. S14). In accordance with predictions, the experiments indicate that the uniform distribution of aromatic residues along LCDs favors solubility and LLPS over aggregation.

The preceding analysis suggests that there may be selection pressure against the linear clustering of aromatic stickers along the sequences of PLDs or a selection for uniformly distributed aromatic stickers along PLD sequences. To test this conjecture, we quantified the patterning of aromatic residues within PLDs

calculations predict T_c for Aro[--] to be below the freezing point of water. This explains why Aro[--] does not undergo LLPS over all temperature and concentration ranges that were titrated ($T > 5°C$; $c < 0.8$ mM).

We next asked if the stickers-and-spacers model was generalizable to PLDs from other proteins. Short motifs known as low-complexity aromatic rich kinked segments (LARKS) and/or reversible amyloid cores (RACs) have been proposed to drive phase separation of the FUS-LCD (16, 37). The removal of RACs leads to measurable changes in the driving forces for phase separation of the FUS-LCD (37). We used our lattice-based stickers-and-spacers model, parameterized using simulation results and experimental data for the A1-LCD, and simulated the phase behavior for four sequence variants of the FUS-LCD (WT, ΔRAC1, ΔRAC2, and ΔRAC1+ΔRAC2) (fig. S13). The nomenclature ΔRAC1 and ΔRAC2 reflects the deletion of RACs 1 and 2 that were identified in the FUS-

LCD by Luo et al. (37). For each of the four constructs, previous measurements quantified the cloud point temperatures at a concentration of ~150 μM. The phase behavior of FUS[1-163] has been studied extensively in published work (19, 38, 39). In our simulations, all constructs formed well-defined, spherical, liquid-like assemblies, and we back-calculated the cloud point at ~150 μM. We observed a 1:1 correlation between experimentally measured cloud points and those estimated using simulation results rescaled to be in absolute temperature units and molar concentrations (Fig. 3G). The simulations do not use any specific information regarding the FUS-LCD other than the relative positions of the aromatic stickers nor do they invoke β sheet–dependent interactions. Accordingly, these results point to the transferability of our model to other PLDs with similar compositional biases.

The quality of the fits of measured binodals to a simple Flory-Huggins model indicate that the sequences studied here can be reduced to

from different proteins that are known drivers of LLPS. We identified a strong bias toward uniform distribution of aromatic stickers along linear sequences for the PLDs of RNA-binding proteins such as FUS, TAF15, EWSR1, hnRNPA2B1, and hnRNPA3 (Fig. 4E) in which the patterning of aromatic residues is highly conserved despite very low levels of sequence conservation (fig. S15, A to C). A proteome-wide analysis of the patterning of aromatic residues within disordered PLDs showed a similar bias toward nonrandom, uniform patterning of aromatic stickers in PLDs from an assortment of dissimilar proteins that are known drivers of LLPS (3, 21) (Fig. 4E and fig. S15, D and E). We also identified disordered regions from proteins involved in vesicular trafficking and signal transduction that have a similar nonrandom bias toward uniform patterning of aromatic residues (Fig. 4E). Intriguingly, the PLD of Xvelo, a protein that drives the formation of solid-like Balbiani bodies (44), is a prominent outlier in terms of its large value of Ω_{aro}.

We propose that the uniform spacing of aromatic stickers along the linear sequences of LCDs ensures that strong interactions among aromatic residues are weakened by the favorable solvation of the spacers. The preferential solvation of spacers likely dilutes the effects of aromatic stickers. Similar considerations are likely to apply to other nonaromatic stickers, such as hydrophobic motifs or those that contribute to cation-π interactions and/or complementary electrostatic interactions (41). Our findings point to interactions encoded in PLDs that are strong enough to drive LLPS and yet weak enough to suppress aggregation—a balance that is likely disrupted through mutations that (i) increase the valence of aromatic or other stickers; (ii) disrupt the nonrandom, well-mixed patterning of aromatic stickers; and/or (iii) add other cohesive interactions through mutations to spacers that weaken their preference for being well solvated.

We have demonstrated the applicability of the stickers-and-spacers framework for quantitative descriptions of sequence-binodal relationships of archetypal PLDs. We have converged on a transferable protocol (fig. S16) for identifying stickers; quantifying the relative strengths of sticker-sticker, sticker-spacer, and spacer-spacer interactions as a function of temperature or other equivalent control parameters; and using this information to generate sequence-to-binodal relationships. These methods will help in mapping cellular concentrations for PLDs to positions relative to their measured or calculated binodals, thus allowing the prediction of how condensates spontaneously form and dissolve in response to changes in protein concentration and cellular conditions (1).

REFERENCES AND NOTES

1. T. M. Franzmann et al., Science **359**, eaao5654 (2018).
2. J. A. Riback et al., Cell **168**, 1028–1040.e19 (2017).
3. A. Molliex et al., Cell **163**, 123–133 (2015).
4. A. Wang et al., EMBO J. **37**, e97452133 (2018).
5. J. B. Woodruff, J. Mol. Biol. **430**, 4762–4772 (2018).
6. A. G. Larson et al., Nature **547**, 236–240 (2017).
7. A. R. Strom et al., Nature **547**, 241–245 (2017).
8. S. Banjade, M. K. Rosen, eLife **3**, e04123 (2014).
9. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Nat. Rev. Mol. Cell Biol. **18**, 285–298 (2017).
10. Y. Shin, C. P. Brangwynne, Science **357**, eaaf4382 (2017).
11. A. Patel et al., Cell **162**, 1066–1077 (2015).
12. C. P. Brangwynne, P. S. Tompa, R. V. Pappu, Nat. Phys. **11**, 899–904 (2015).
13. R. M. Vernon, J. D. Forman-Kay, Curr. Opin. Struct. Biol. **58**, 88–96 (2019).
14. T. J. Nott et al., Mol. Cell **57**, 936–947 (2015).
15. X. Gui et al., Nat. Commun. **10**, 2006 (2019).
16. M. P. Hughes et al., Science **359**, 698–701 (2018).
17. M. Kato et al., Cell **149**, 753–767 (2012).
18. J. P. Brady et al., Proc. Natl. Acad. Sci. U.S.A. **114**, E8194–E8203 (2017).
19. K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, Mol. Cell **60**, 231–241 (2015).
20. M. Rubinstein, A. V. Dobrynin, Trends Polym. Sci. **5**, 181–186 (1997).
21. J. Wang et al., Cell **174**, 688–699.e16 (2018).
22. A. N. Semenov, M. Rubinstein, Macromolecules **31**, 1373–1385 (1998).
23. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, eLife **6**, e30294 (2017).
24. P. J. Flory, J. Am. Chem. Soc. **63**, 3083–3090 (1941).
25. A. E. Posey, A. S. Holehouse, R. V. Pappu, in Methods in Enzymology, E. Rhoades, Ed. (Academic Press, 2018), vol. 611, pp. 1–30.
26. H. J. Kim et al., Nature **495**, 467–473 (2013).
27. A. Vitalis, R. V. Pappu, J. Comput. Chem. **30**, 673–699 (2009).
28. A. S. Holehouse, R. V. Pappu, Annu. Rev. Biophys. **47**, 19–39 (2018).
29. G. Raos, G. Allegra, J. Chem. Phys. **107**, 6479–6490 (1997).
30. J. A. Riback et al., Science **358**, 238–241 (2017).
31. J. Klein-Seetharaman et al., Science **295**, 1719–1722 (2002).
32. P. A. Evans, K. D. Topping, D. N. Woolfson, C. M. Dobson, Proteins **9**, 248–266 (1991).
33. D. Neri, M. Billeter, G. Wider, K. Wüthrich, Science **257**, 1559–1563 (1992).
34. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Proc. Natl. Acad. Sci. U.S.A. **115**, 9929–9934 (2018).
35. M. T. Wei et al., Nat. Chem. **9**, 1118–1125 (2017).
36. M. Rubinstein, R. H. Colby, Polymer Physics (Oxford Univ. Press, 2003).
37. F. Luo et al., Nat. Struct. Mol. Biol. **25**, 341–346 (2018).
38. Z. Monahan et al., EMBO J. **36**, 2951–2967 (2017).
39. A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, Structure **24**, 1537–1549 (2016).
40. R. K. Das, R. V. Pappu, Proc. Natl. Acad. Sci. U.S.A. **110**, 13392–13397 (2013).
41. C. W. Pak et al., Mol. Cell **63**, 72–85 (2016).
42. E. W. Martin et al., J. Am. Chem. Soc. **138**, 15323–15335 (2016).
43. F. Tanaka, M. Ishida, Macromolecules **30**, 1836–1844 (1997).
44. E. Boke et al., Cell **166**, 637–650 (2016).
45. A. Holehouse, R. V. Pappu, PIMMS (0.24 pre-beta), Zenodo (2019); doi:10.5281/zenodo.3588456