Backbone-mediated weakening of pairwise interactions enables percolation in mimics of protein condensates

Xiangze Zeng[†,*] and Rohit V. Pappu[§,*]

[†]Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong and Teaching and Research Division, School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[§]Department of Biomedical Engineering and Center for Biomolecular Condensates, The James McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

Corresponding Authors

*xiangzezeng@hkbu.edu.hk; *pappu@wustl.edu

ABSTRACT

Intrinsically disordered regions (IDRs) of proteins with specific sequence grammars can be drivers of phase separation and percolation that enable the formation of biomolecular condensates. Measurements have shown that dense phases of protein-based condensates are semidilute solutions. In these solutions, the probability of realizing intermolecular interactions is higher than the probability of intramolecular interactions. Accordingly, to zeroth order, dense phases may be viewed as concentrated solutions of peptide-sized motifs. Here, we report results from all-atom molecular dynamics simulations that were used to quantify differences between inter-residue interactions in mimics of dense versus dilute phases. The simulations use the polarizable AMOEBA forcefield for peptides, ions, and water molecules. In simulations that treat coexisting phases as solutions of model compounds, we find that the interactions between aromatic residues are stronger than interactions between cationic and aromatic residues. Cooperativity within dense phases formed by freely diffusing model compounds is manifest as enhanced pairwise interactions and the formation of nanoscale clusters of aromatic sidechains. Simulations that account for the effects of peptide backbones reveal contrasting results. While peptide backbones maintain or enhance pairwise inter-residue interactions in dilute phases, these interactions are weakened within dense phases. Backbone-mediated weakening of pairwise inter-residue interactions within dense phases is accompanied by the gain of higher-order interactions that enables percolation, whereby molecules within a dense phase become part of a system-spanning network. Our findings provide a physico-chemical rationale for phase separation and percolation as joint drivers of condensate formation.

INTRODUCTION

Phase separation of proteins and nucleic acids leads to the formation of biomolecular condensates [1]. In a two-component system comprising a protein and solvent, phase separation yields two coexisting phases namely, a dilute phase that is deficient in protein and a coexisting dense phase that is enriched in protein [2]. Intrinsically disordered regions (IDRs) of proteins that feature specific sequence characteristics [3] can be drivers of protein phase separation [4]. Of relevance to condensates such as stress granules [5] and P-bodies [6] are IDRs known as prion-like low complexity domains (PLCDs) [7]. The sequence determinants of driving forces for phase separation of PLCDs have been the topic of several studies [8].

Recently, Bremer et al. quantified the temperature-dependent driving forces for phase separation of 40 different sequence variants of A1-LCD, which is the PLCD from hnRPNP-A1 [9]. At a given temperature, each A1-LCD variant is defined by a saturation concentration c_{sat}, which is the threshold concentration for phase separation. Lower values of c_{sat} imply stronger driving forces for phase separation and vice versa. The temperature-specific values of c_{sat} were found to be governed by: (i) the number and types of aromatic residues (Phe versus Tyr), also referred to as aromatic stickers, which are uniformly distributed along the linear sequence [8, 10]; (ii) the context-dependent contributions of arginine residues, which function either as auxiliary stickers or spacers; (iii) the types of polar residues, which in PLCDs appear to function as spacers that are interspersed between aromatic and Arg residues; (iv) the destabilizing effects of lysine residues that interfere with π-π interactions; and (v) the contributions of net charge per residue. Away from the critical temperature, changes to sequence features can lead to changes in temperature-specific values of c_{sat} and these changes can span 4-5 orders of magnitude [9]. In contrast, while the material properties of dense phases formed by variants of A1-LCD, specifically the measured and computed viscosities, show a clear correlation with c_{sat} [11, 12], the concentrations of coexisting dense phases vary by no more than a factor of 2-3. These findings suggest that the material properties of dense phases and the values of c_{sat} are driven primarily by the free energy of transfer of PLCDs from dilute to dense phases [13]. The results also imply that interactions must be different between dilute versus dense phases [14].

Dense phases of PLCDs and other systems are equivalent to semidilute solutions of flexible linear polymers (**Figure 1**) [9, 15]. To understand the properties of semidilute solutions of IDRs, we

consider an arbitrarily chosen polymer molecule. In semidilute solutions, the concentrations of polymers are at or above the overlap concentration [14]. Therefore, the pervaded volume of the polymer will be occupied by solvent molecules and segments of other polymers. Each polymer may be viewed as a chain of so-called thermal blobs [16], where each blob comprises g_T residues, and the conformations within a blob are those of an ideal chain [17, 18]. As polymer concentrations increase, the concentrations of thermal blobs increase and this sets up a new length scale ξ, which is the correlation length that is also referred to as the size of a concentration blob or the screening length [19]. For inter-blob separations that are less than ξ, there is an interplay between intra-chain and inter-chain interactions among thermal blobs [20]. For distances larger than ξ, the intra-chain interactions are screened [21], and the only interactions of interest are between blobs from different chain molecules [20, 22]. Away from the critical point, which is the regime of interest, g_T approaches one, and the value of ξ will be on the order of a few residues as polymer concentrations increase above the overlap concentration [14]. Accordingly, as a zeroth order approximation, we model interactions above the overlap regime by approximating dense phases as a fluid of peptide-sized segments in an aqueous solvent [23].

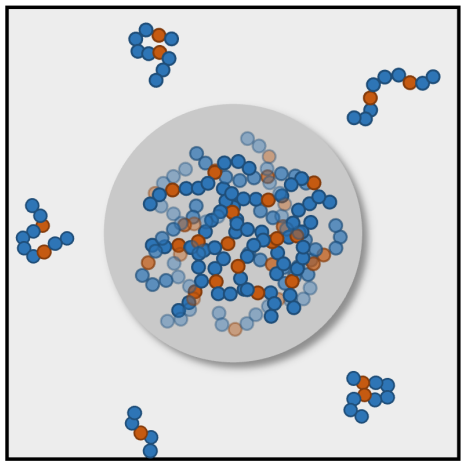

Figure 1. Dense phases of protein-based condensates are semidilute solutions of flexible linear polymers [15]. Here, the volume fraction of polymers is larger than the overlap volume fraction. The overlap volume fraction is defined as the ratio of the occupied volume of the polymer chain $v_{mono}N$ to the pervaded volume V, where N is the number of monomer in one polymer chain and v_{mono} is the occupied volume of one chemical monomer. V can be roughly estimated as R^3, where R is the size of the chain. The blue spheres denote spacers, and the orange spheres denote stickers. Our calculations probe interactions between peptide-sized motifs, focusing on the inter-sticker interactions and their modulation by peptide backbones.

We used all-atom molecular dynamics simulations to investigate the concentration dependence of the strengths of effective pairwise interactions among functional groups that are known to be important as drivers of phase separation of PLCDs [8, 9, 11, 15, 24]. Specifically, our simulations quantify the strengths of interactions among aromatic and cationic residues in mimics of dilute versus dense phases. Volume fractions of sidechain model compounds between 0.002 and 0.003 were used to mimic the dilute phase whereas volume fractions between 0.08 and 0.09 mimic the dense phase.

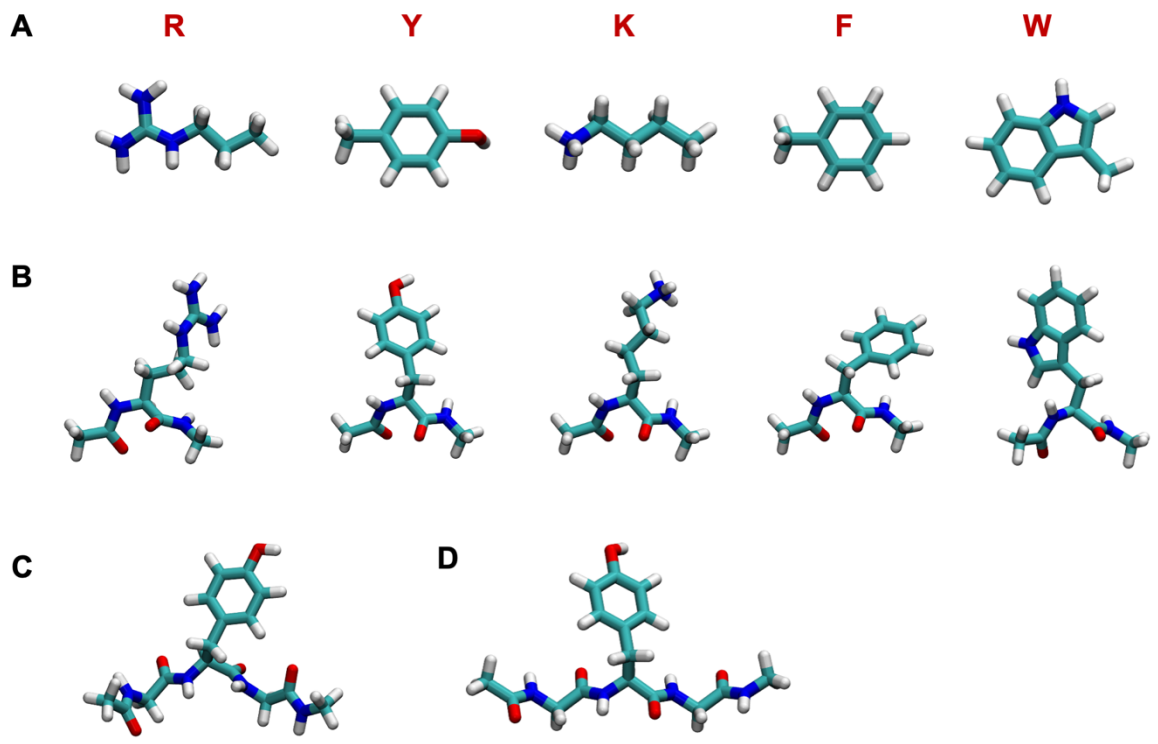

Figure 2. Model compounds and capped amino acids used to compare interactions between pairs of polypeptide units in mimics of dilute versus dense phases. (A) Five model compounds *n*-propylguanidine, 1-butylamine, *p*-Cresol, toluene, and 3-methylindole that mimic the side chains of Arg, Lys, Tyr, Phe, and Trp, respectively. (B) Drawings of capped amino acids, N-acetyl-Xaa-N′-methylamide for Xaa ≡ Arg, Lys, Tyr, Phe and Trp, respectively. Each of these systems feature two peptide units and one pair of backbone ϕ and ψ angles. (C) We also performed simulations for a fully flexible tripeptide (D) N-acetyl-Gly-Tyr-Gly-N′-methylamide GYG and (E) the tripeptide N-acetyl- Gly-Tyr-Gly -N′-methylamide restrained to be in an extended conformation.

The simulations use the polarizable forcefield, AMOEBA [25] and the design is as follows: For different pairs of amino acid sidechains, we performed two sets of simulations. In the first set, we used model compound mimics of specific amino acid side chains (**Figure 2A**) to compute potentials of mean force (PMFs) and quantify the free energies of association / dissociation

between functional groups on sidechains. To assess dilute phase interactions among motifs, we computed PMFs between a pair of model compounds in a cubic box of volume ~125 nm^3, where each model compound represents a specific amino acid sidechain. To quantify interactions among sidechains in dense phases, we computed PMFs from simulations of mixtures of model compounds as a function of concentration. In a second set of simulations, we computed inter-sidechain PMFs using dilute and dense phase simulations of capped amino acids (**Figure 2B**). Here, a capped amino acid Xaa refers to the molecule N-acetyl-Xaa-N′-methylamide, where Xaa is the amino acid residue of interest. Each capped amino acid features two peptide units, and a pair of backbone ϕ and ψ angles that are unrestrained. Tripeptides were also used to test sequence context effects and conformational contributions (**Figure 2C, 2D**). Simulations of capped amino acids and tripeptides help us query the concentration dependent modulatory effects of backbone peptide units and local sequence contexts on the strengths of sidechain interactions. We used a polarizable forcefield because we expect the microenvironments of dense phases to be different from those of dilute phases [26], and a forcefield lacking polarizability does not have the ability to capture changes to interactions that are influenced by changes to the microenvironment.

For model compounds, sans peptide units, the effective pairwise interactions are significantly enhanced in dense phases. This drives the formation of nanoscale clusters of aromatic moieties. For capped amino acids, the presence of the peptide unit enhances the strengths of pairwise interactions between sidechains in the dilute phase. However, in dense phases, pairwise interactions are weakened, and nanoscale clusters formed by aromatic moieties within dense phases of model compounds are dissolved. Weakened pairwise interactions are offset by a gain of higher-order interactions that drive the formation of networks that span the entirety of dense phases. This is a manifestation of percolation [2, 15, 27] that gives rise to the description of condensates being network fluids [28].

RESULTS

PMFs in dilute phase follow well-defined hierarchies: We computed dilute phase PMFs for pairs of model compounds that mimic the following interactions: Phe-Phe, Tyr-Tyr, Trp-Trp, Arg-Phe, Arg-Tyr, Arg-Trp, Lys-Phe, and Lys-Tyr. The model compounds that mimic sidechains of Arg, Lys, Tyr, Phe and Trp are *n*-propylguanidine (R), 1-butylamine (K), *p*-Cresol (Y), toluene (F), and 3-methylindole (W), respectively (**Figure 2A**). These sidechain interaction pairs were

chosen because they are the dominant drivers or destabilizers of phase separation of PLCDs [9, 15, 24].

To quantify inter-sidechain interactions in the dilute phase, we computed PMFs for different pairs of model compounds using umbrella sampling. Volume fractions of sidechain model compounds between 0.002 and 0.003 were used to mimic the dilute phase. Each pair was solvated in a cubic box of water and solution ions, ~5 nm to a side, with periodic boundary conditions. The reaction coordinate chosen for umbrella sampling was the distance between specific atoms on functional groups of each model compound. For R and K, the chosen atoms were the carbon atom in the guanidium group of Arg and the nitrogen atom in the amine of Lys. For Y, F, and W, we used the geometric center of the aromatic ring to define the reaction coordinate.

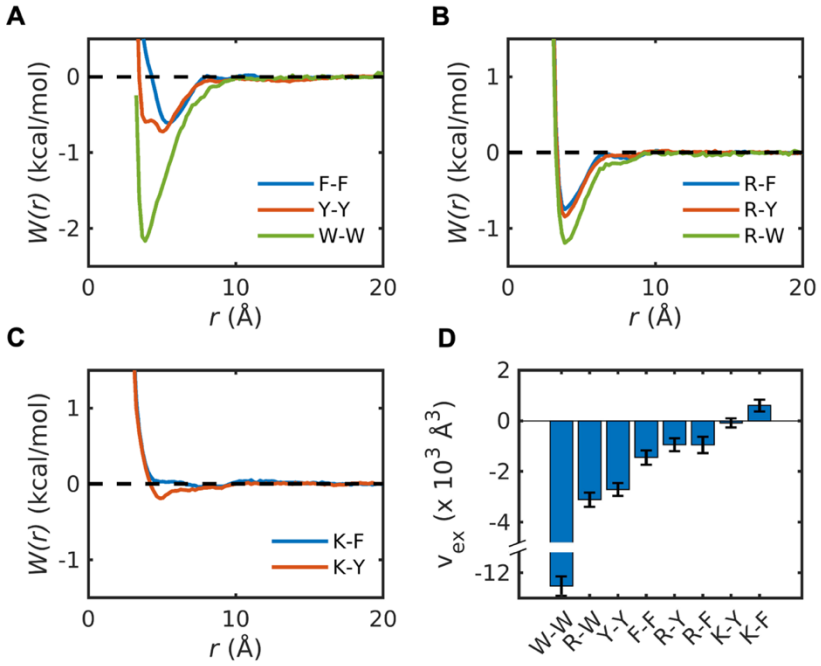

Figure 3. Pairwise PMFs computed for pairs of model compound mimics of sidechains. PMFs, $W(r)$, for interactions between π systems Phe, Tyr, and Trp, (B) Arg-π systems, and (C) Lys-π systems. (D) Each of the pairwise PMFs can be used to compute pair-specific excluded volumes (v_{ex}), as described in the text.

The PMFs for each of the six model compound pairs are shown in **Figures 3A-3C**. The PMF for Y-Y shows two minima at ~3.6 Å and ~5.0 Å, respectively (**Figure 3A**). These minima correspond to two different interaction modes, namely the stacked structure and T-shape structure (**Figure S1**). The locations of these minima are consistent with results reported for benzene dimers using high-level quantum calculations [29]. The PMF for F-F shows a single minimum at ~5.0 Å,

indicating that the T-shape structure dominates the short-range interactions. The PMFs for F-F and W-W have one minimum each, with the minimum for F-F corresponding to the T-shaped structure, and the minimum for W-W corresponding to the stacked structure (**Figure 3A**). The dominant minima in PMFs for R-Y, R-F, and R-W appear at a separation of ~3.6 Å and this corresponds to stacked structures formed by the planar guanidinium group and the rings of the aromatic groups (**Figure S2**). These structures are consistent with reports of interactions between Arg and aromatic residues being of cation-π-π flavor [30]. We also computed PMFs for K-Y and K-F. These interactions are essentially negligible when compared to those of R-Y and R-F (**Figure 3C**).

To enable a single-parameter comparison of the strengths and types of pairwise interactions, we computed excluded volumes (v_{ex}) by integrating the Mayer *f*-function [31]. For this, we used the formalism whereby: $v_{ex} = \int 4\pi r^2 \left[1 - exp\left(-\frac{W(r)}{k_B T}\right)\right] dr$. Here, $W(r)$ is the PMF for pairwise interactions computed from the simulations, and $4\pi r^2 dr$ is the volume element. The value of v_{ex}, which quantifies the volume excluded for interactions with the solvent [31, 32], is an effective two-body interaction coefficient between a pair of moieties in solvent and it is directly proportional to the second virial coefficient [2, 31]. A negative value implies that the interactions are, on average, attractive, whereas positive values imply net repulsions. The magnitude of v_{ex} quantifies the strength of the attraction or repulsion.

The values of v_{ex} computed using the PMFs are shown in **Figure 3D**. The strengths of attractive attractions, quantified by negative values of v_{ex}, follow the hierarchy W-W > R-W > Y-Y > F-F > R-Y > R-F. The v_{ex} values for K-Y and K-F are near zero (K-Y) or positive, albeit small (K-F), implying that these interactions are negligible or slightly repulsive. Note that the direct interactions between water molecules and model compounds are weaker than the solvent-mediated pair interactions between model compounds (**Figure S3**).

Next, we computed dilute phase PMFs between pairs of capped amino acids along reaction coordinates that are identical to those for model compounds (**Figure 4**). Although the locations of minima remain the same as those for model compounds, the well depths increase for all pairs. Additionally, auxiliary minima appear at separations of ~8.5 Å for all pairs. These point to backbone-mediated, longer-range contacts between functional groups of sidechains that are absent for model compounds. Although backbones enhance the overall interaction strengths between sidechains, the overall hierarchy remains the same as can be seen by comparing the excluded

volumes in **Figure 4D** to those in **Figure 3D**. To quantify the influence of backbones on the interactions between side chains, we calculated $\Delta W(r) = W_{\mathrm{AA}}(r) - W_{\mathrm{mc}}(\mathrm{r})$. This refers to the difference between PMFs for capped amino $W_{\mathrm{AA}}(r)$ those for model compounds $W_{\mathrm{mc}}(\mathrm{r})$ (**Figure S4**). These comparisons show that modulation by backbones help enhance attractions between sidechains.

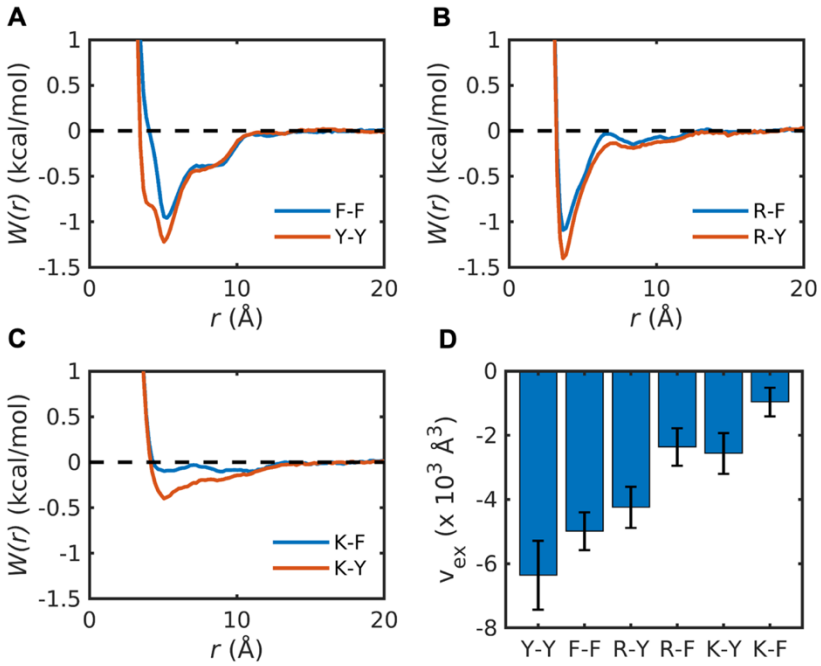

Figure 4. Interactions between sidechains are enhanced by the presence of backbone peptide units in the dilute phase. PMFs, $W(r)$, for (A) π-π, (B) Arg-π, and (C) Lys-π interactions between pairs of capped amino acids. (D) The pairwise PMFs in panels (A) – (C) were used to compute pair-specific excluded volumes (v_{ex}).

Model compounds form nanoscale clusters within dense phases: Next, we quantified the effective pairwise interaction strengths among sidechains in dense phase mimics comprising model compounds. In these simulations, the volume fractions of sidechains are over an order of magnitude larger than in the dilute phase, ranging from 0.08 to 0.09. These values correspond to measured volume fractions of sidechains that place PLCD-like systems in two-phase regimes. Specifically, we performed seven separate simulations for systems comprising different numbers of: R and F, R and Y, R and W, respectively. For each pair, we performed two or three sets of simulations. In 1:1 mixtures, the simulations comprised 40 R and 40 F / Y / W model compound mimics (**Figure 5A**). In 2:1 mixtures the simulations comprised 53 R and 27 F / Y model

compound mimics (**Figure 5B**). Finally, in 1:2 mixtures, the simulations comprised 27 R and 53 F / Y model compound mimics (**Figure 5C**).

When compared to the dilute phase PMFs (**Figure 3A**), the PMFs for the dense phase mimics (**Figures 5A-5C**) show the following trends: The energetics for Y-Y interactions change the least, with the only substantive change being the increased preference for T-shaped structures. The PMFs converge to zero for separations beyond 10 Å, indicating an absence of long-range order in the R-Y mixtures (**Figures 5A-5C**). However, the plateauing to zero of the PMFs for Y-Y interactions is dependent on the R:Y ratio. When the R:Y ratio is 1:2, then the plateauing to zero happens well beyond 10 Å, and this is indicative of long-range ordering. The PMFs for F-F change both qualitatively and quantitatively. When compared to the dilute phase PMFs (**Figure 3A**), the T-shaped structures have higher stability in dense phase mimics. Additionally, we observe long-range, non-random organization for 1:1 and 2:1 ratios of R:Y (**Figures 5A** and **5C**). This is indicative of clustering on the nanometer scale. Increasing the R:F ratio weakens the preference for T-shaped structures and long-range ordering (**Figure 5B**) and decreasing the R:F ratio has the opposite effect (**Figure 5C**). In R:W mixtures, the PMF for W-W interactions shows an equivalent preference for stacked and T-shaped structures (**Figure 5A**), which is different from the PMFs in the dilute phase (**Figure 3A**). This weakens the preference for stacked structures when compared to the dilute phase. In addition to the equivalent preferences of stacked and T-shaped structures, the 1:1 R-W mixtures show the long-range ordering we observe for R-F mixtures. Snapshots derived from simulations of 1:1 mixtures show the formation of nanoscale clusters of aromatic moieties in R-F (**Figure 5D**) and R-W (**Figure 5F**) mixtures, and the weaker clustering in R-Y mixtures (**Figure 5E**).

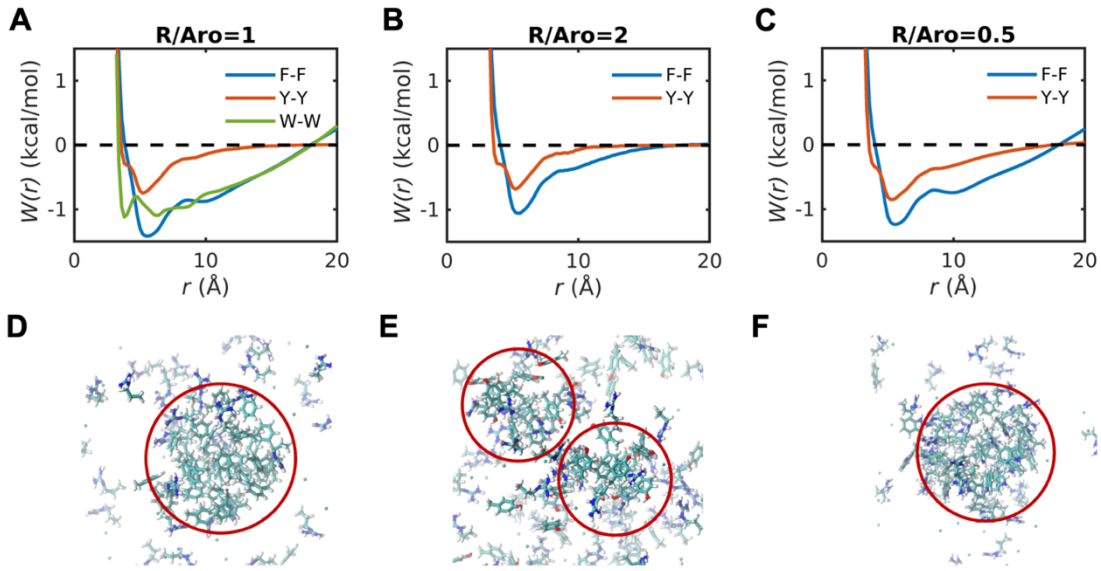

Figure 5. Strengths of interactions between aromatic groups are enhanced in backbone-free dense phases. (A) PMFs, $W(r)$, for F-F, Y-Y, and W-W in 1:1 mixtures of R and F, R and Y, and R and W. Each mixture includes 40 R and 40 aromatic (either F, Y or W) sidechains. (B) $W(r)$ for F-F and Y-Y in 2:1 mixtures of Arg model compound with each aromatic model compound. Each mixture includes 53 Arg and 27 aromatic (either F, Y or W) model compounds. (C) $W(r)$ for F-F and Y-Y in 2:1 mixtures of Arg model compound with each aromatic model compound. Each mixture includes 27 Arg and 53 aromatic (either F, Y or W) model compounds. Snapshots of nanoscale clusters that are formed in 1:1 mixtures of Arg model compound with (D) Phe, (E) Tyr, and (F) Trp.

The weakened clustering of aromatic moieties in R-Y mixtures can be rationalized by comparisons of the R-Y, R-F, and R-W PMFs. The minimum, which corresponds to stacked structures of the guanido group and aromatic moieties is most pronounced in R-Y mixtures and is weaker in R-F and R-W mixtures (**Figure 6A**). The extent to which R-F interactions are weakened depends on the R:F ratio, becoming more pronounced as this ratio decreases (**Figure 6B**). In contrast, the R-Y interactions, although destabilized when compared to the dilute phase, are insensitive to changes in the R:Y ratio in dense phase mimics (**Figure 6C**). These results suggest a correlation between stronger R-Y interactions and weaker clustering of Tyr mimics, whereas the converse is true for R-F and R-W mixtures. In contrast to R-F and R-Y mixtures, the interactions in K-F and K-Y mixtures are such that interactions are either negligible (K-Y, **Figure 6B**) or weakly attractive (K-Y, **Figure 6C**). These interactions in dense phase mimics, which do not depend on concentration or the K:F / K:Y ratios, resemble what we compute in dilute phases (**Figure 3C**).

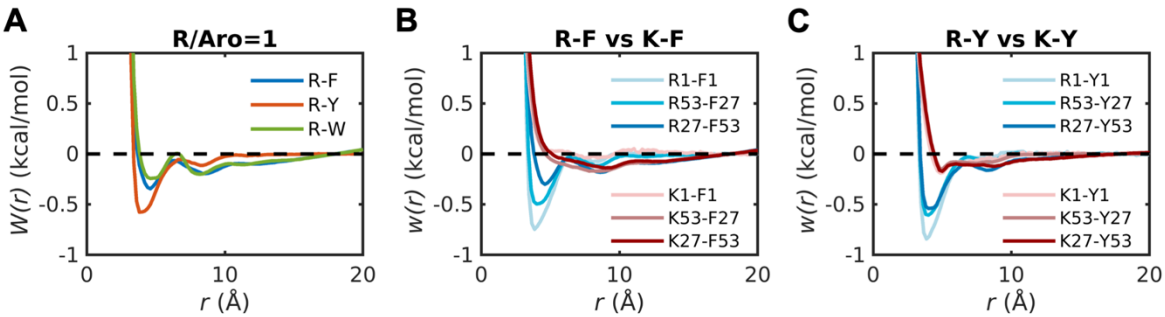

Figure 6. Comparative strengths of cation-π interactions in mimics of dense phases with different ratios of compounds mimicking Arg (R), Lys (K) and aromatic (Aro) sidechains. (A) PMFs, $W(r)$, for R-F, R-Y, and R-W in 1:1 mixtures of Arg model compound with each aromatic model compound. Each mixture includes 40 model compound mimics of Arg and aromatic sidechains (either F, Y or W). (B) Comparison of R-F and K-F in different dense phase mimics. R1-F1 and K1-F1 denote $W(r)$ for R-F and K-F in the dilute phase calculated from umbrella sampling. The other legends denote $W(r)$ in mixtures containing different numbers of model compounds mimicking Arg and aromatic sidechains. In the legend, R53-F27 refers to PMFs extracted from simulations of dense phases with 53 Arg and 27 Phe sidechains. Conversely, R27-F53 refers to PMFs extracted from simulations of dense phases with 27 Arg and 53 Phe sidechains. A similar convention is applied for PMFs computed in dense phases with different mixtures of Lys and Phe sidechains. (C) Comparison of R-Y and K-Y in different dilute and dense phases. Conventions for the legends are the same as those in panel (B).

We also explored the influence of total sticker concentration while maintaining a 1:1 ratio of *n*-propylguanidine to aromatic moieties (**Figure S5**). The effects we observe for Y-Y, F-F, R-Y, and R-F interactions (**Figure 5** and **Figure 6**) become more pronounced with increasing concentration (**Figure S5**). Nanoscale clusters of Tyr moieties, which are absent at lower concentrations, are manifest in mixtures with 60 *n*-propylguanidine and 60 *p*-cresol molecules. This implies that the formation of nanoscale aromatic clusters is likely to be a generic feature of dense solutions with aromatic molecules. However, the specificity of cluster formation, which depends on the solvent- and mixture-mediated strengths of associations among aromatic moieties, will be manifest in the concentration dependence of cluster formation. This point is made by the observation that nanoscale clusters form in 1:1 mixtures with 30 *n*-propylguanidine and 30 toluene molecules.

Pair interactions are weakened in solutions of capped amino acids

Next, we performed simulations for mixtures of capped amino acids to compute pair PMFs for 1:1 mixtures of dense phase mimics of different combinations of capped Arg and aromatic residues (**Figure 7**). In all cases, we observed a weakening of pairwise associations and the extent of weakening increases within increasing concentrations. Furthermore, the pair PMFs for F-F, Y-F

and W-W (**Figures 7A**, **7C**, and **7E**) do not show the long-range ordering that was observed for mixtures of model compounds (**Figure 5** and **Figure S5**).

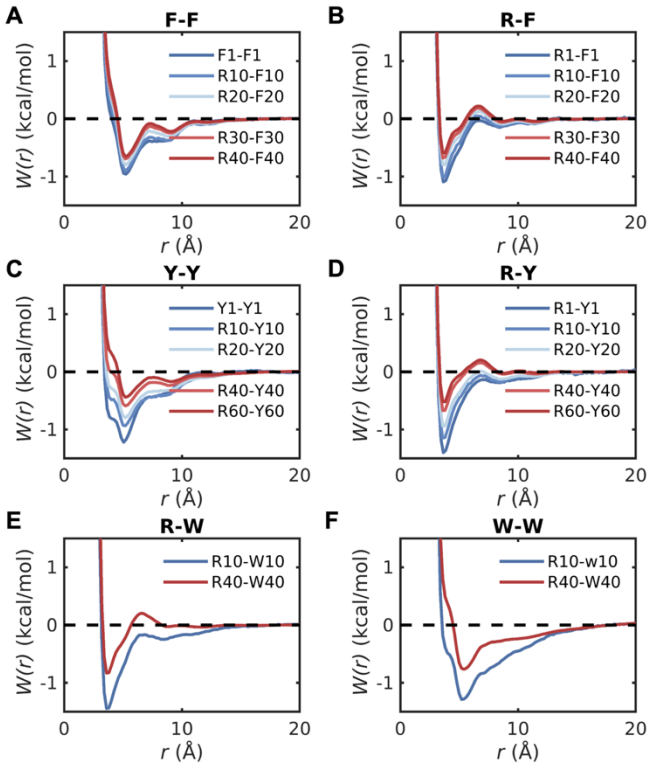

Figure 7. Contributions from backbone peptide units weaken sidechain interactions within dense phases. Comparisons of potentials of mean force $W(r)$ in dilute versus dense phases for pairs of capped amino acids (A) F-F, (B) R-F, (C) Y-Y, (D) R-Y, (E) R-W, and (F) W-W. The number in the legends denotes the amount of model compounds indicated by the preceding letter in the mixture mimicking dense phases. F1-F1, Y1-Y1, R1-YF, and R1-Y1 denote potential of mean force in dilute phase calculated from umbrella sampling. The interactions are weakened as increasing the concentration of amino acid stickers.

The information contained in the different pair PMFs is summarized by computing pair-specific values of v_{ex} for different concentrations of 1:1 mixtures of Arg and aromatic moieties in dense phase mimics based on model compounds (**Figures 8A-8B**) or capped amino acids (**Figures 8C-8D**). For the F-F, Y-Y, and W-W pairs, the magnitudes of v_{ex} increase in dense phase mimics that are based on the use of model compounds, whereas for the same pairs, the magnitudes of v_{ex} decrease with increasing concentrations of capped amino acids. Similar trends prevail when comparing the excluded volumes for pairwise associations of Arg and Tyr, Phe, or Trp (see **Figures 8B** versus **8D**). Overall, the pairwise associations are weakened, long-range attractions are lost, and nanoscale clusters are dissolved in dense phase mimics that use capped amino acids. The implication is that while the peptide unit enhances and maintains the hierarchy of pairwise

associations observed in dilute phase mimics based on model compounds, the interactions are fundamentally altered in dense phases. We asked if the weakening of pairwise associations is offset by a gain in sidechain-backbone or backbone-backbone pairwise associations. Analysis of the pair PMFs that quantify these pairwise associations show the same trends we observed for sidechain interactions (**Figure S6**). Furthermore, we found that the stoichiometry of Arg versus aromatic residues does not alter the PMFs for capped amino acids (**Figure S7**). This points to the robustness of the weakening of pairwise associations caused by the backbone-mediated interactions.

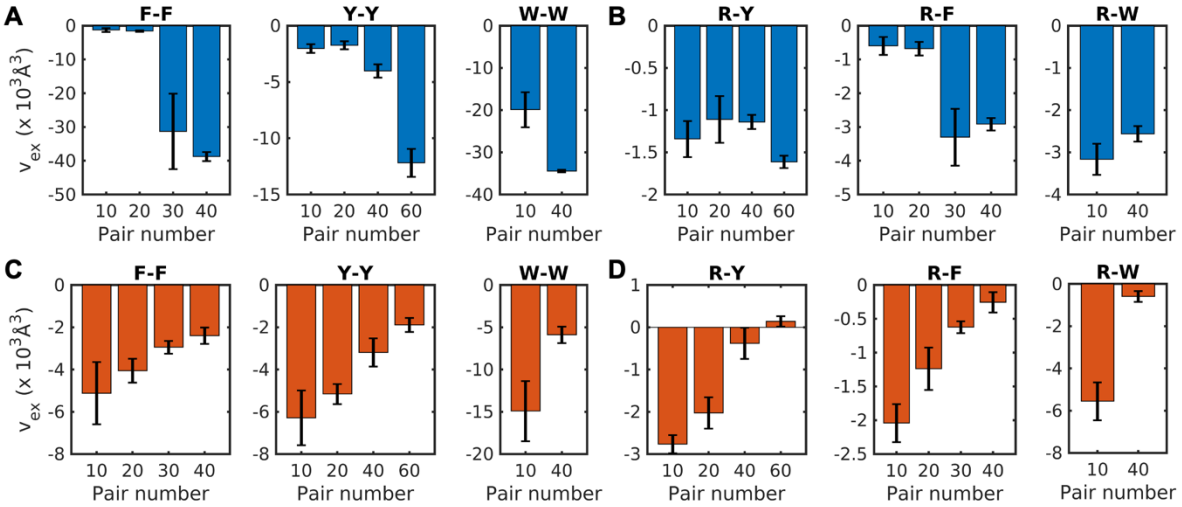

Figure 8. Cooperative effects between backbone and side chain mediated interactions weakens pairwise interactions among amino acids. Excluded volumes v for pairwise interactions of (A)-(C) model compounds and (D)-(F) capped amino acids. (A) F-F and R-F in 1:1 mixture of Arg and Phe model compounds, (B) Y-Y and R-Y in 1:1 mixture of Arg and Tyr model compounds, and (C) W-W and R-W in 1:1 mixture of Arg and Trp model compounds. (D) F-F and R-F in 1:1 mixture of Arg and Phe capped amino acids, (E) Y-Y and R-Y calculated in 1:1 mixture of Arg and Tyr capped amino acids, and (F) W-W and R-W in 1:1 mixture of Arg and Trp model compounds. X-axis represents the number of pairs of Arg and aromatic stickers in the mixture.

Next, we tested the effects of longer peptides on pairwise associations between Arg and aromatic residues, specifically Tyr. In PLCDs, aromatic residues are uniformly distributed along the linear sequence [10]. Motifs such as GYG and GRG are common in PLCDs. Accordingly, we computed pair PMFs, with Arg-Tyr distances as the reaction coordinates, for two different combinations of capped GYG and GRG tripeptides. In one set of simulations, the backbone conformations were unrestrained. In the other set of simulations, the four pairs of backbone ϕ and ψ angles were restrained in extended β-strand-like conformations with values of $\phi = -140^\circ$ and $\psi = +135^\circ$.

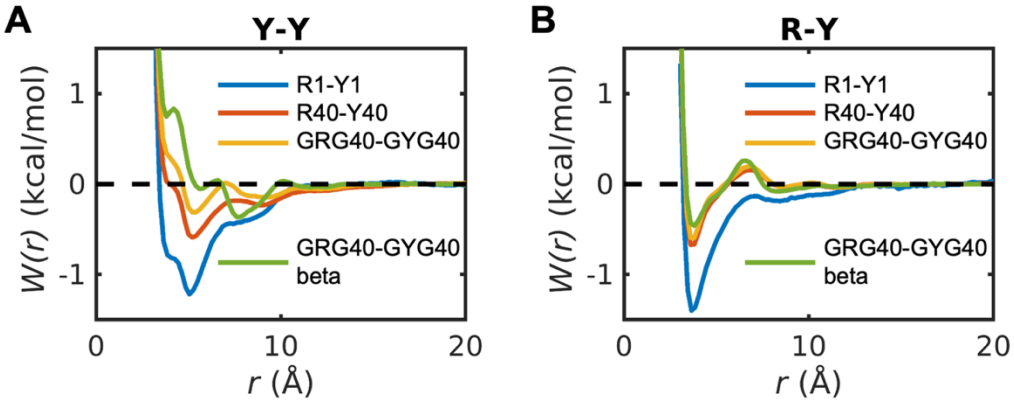

Figure 9. Influence of stoichiometry of cationic versus aromatic amino acids and backbone flexibility on pairwise interactions between capped amino acids. The potential of mean force for (A) Y-Y, and (B) R-Y in mixtures including capped Arg and Tyr. The number in the legends denotes the amount of model compounds indicated by the preceding letter in the mixture mimicking dense phases. Y1-Y1and R1-Y1 denote potential of mean force in dilute phase calculated from umbrella sampling. While the total concentration keeps the same, the stoichiometry of cationic vs. aromatic stickers does not affect the effective pairwise interactions between capped amino acids in the dense phase. The potential of mean force for (A) Y-Y, and (B) R-Y in mixtures including capped tripeptides GRG and GYG. The red curve denotes results when the backbone of tripeptide is constrained in the β region where $\phi = -140°, \psi = 135°$.

The PMFs for Y-Y associations computed in dense phase mimics with 40 GRG and 40 GYG peptides are compared to the PMFs for capped Arg and Tyr in dilute versus dense phases (**Figure 9A**). For the unrestrained pairs of tripeptides, the PMFs share the same profiles as those for capped amino acids, but the interactions are weakened in comparison. Restraining the backbone conformations, elicits a substantial entropic penalty, and this weakens pairwise associations between Tyr residues (**Figure 9A**). In all PMFs, the locations of stacked and T-shaped structures stay invariant, whereas the minimum corresponding to backbone-mediated associations, located at ~ 8.5Å for capped amino acids and unrestrained tripeptide, shifts down to ~7.5Å for the restrained tripeptides. The PMFs that quantify R-Y associations are equivalent for capped amino acids and tripeptides **(Figure 9B**). The totality of the results presented to this point suggest that strengths of pairwise associations and long-range ordering due to strong pair interactions between aromatic residues are weakened by backbone-mediated effects. The loss of long-range ordering points to fluidizing effects of backbone-mediated interactions. However, the pair associations present a conundrum. If they are the only associations to be considered, which is true in theories that rely exclusively on two-body interactions [33], then transferring peptides or proteins into dense phases should be thermodynamically unfavorable [34]. The fact that phase separation is favored above system-specific threshold concentrations suggests that interactions beyond pairwise associations,

[21], must contribute to interactions within dense phases [14, 35]. This point has been underscored recently to explain why compositionally distinct phases are likely to be determined by the contributions of three-body and higher-order interactions [36]. We investigated the contributions of higher-order interactions by quantifying the degree of clustering through connectivity [37], which is the basis for forming network fluids via directional interactions whereby molecules can be viewed as multivalent nodes connected to other molecules via edges on a graph [23, 28]. Indeed, liquid water is an archetypal example of a network fluid, wherein pairwise associations are weakened vis-à-vis the gas phase in favor of forming a system-spanning network [38].

Dense phase mimics modeled using capped amino acids are percolated network fluids

We asked if the weakening of pairwise interactions and loss of long-range ordering in dense phases mimicked using capped amino acids are offset by a gain in higher-order interactions in the form of higher-order clusters and networks that span the entire system. To answer this question, we performed an analysis of the connectivity of molecules within dense phase mimics formed by capped amino acids [23, 28, 37, 39]. The analysis was inspired by the recent work of Dar et al., [28]. Two molecules were considered connected if any of the distances between specific atoms or groups is smaller than a prescribed cutoff. The pairwise distances used to compute connectivity were divided three groups: (1) sidechain-sidechain distances, (2) sidechain-backbone distances, (3) distances between backbone amide nitrogen atoms and backbone carbonyl oxygen atoms. The inter-sidechain distances include those between aromatic rings, between guanidinium groups, and between aromatic ring and guanidinium group. The sidechain-backbone distances include those between aromatic ring and terminal CH_3 groups, between aromatic ring and backbone carbonyl oxygen atoms, and between the guanidinium group and backbone carbonyl oxygen atoms. The geometric centers of carbon atoms in the aromatic ring, the nitrogen atom in the guanidinium group, and the carbon atom in the terminal CH_3 group were used to calculate these distances. The first minimum of the radial distribution function for a specific pair of atoms or groups was chosen as the distance cutoff to determine the connectivity. The values for these distance cutoffs are shown in **Table S1**. If molecule A is connected to molecules B and C, then molecule B and C are also connected via transitive closure [40]. All connected molecules make up a cluster. We used this connectivity analysis to define the cluster size as the number of molecules within a cluster. We also defined the cluster dimension as the maximum distance between alpha carbon atoms within the cluster [41]. For the cluster with only one molecule, we set the cluster dimension to be zero.

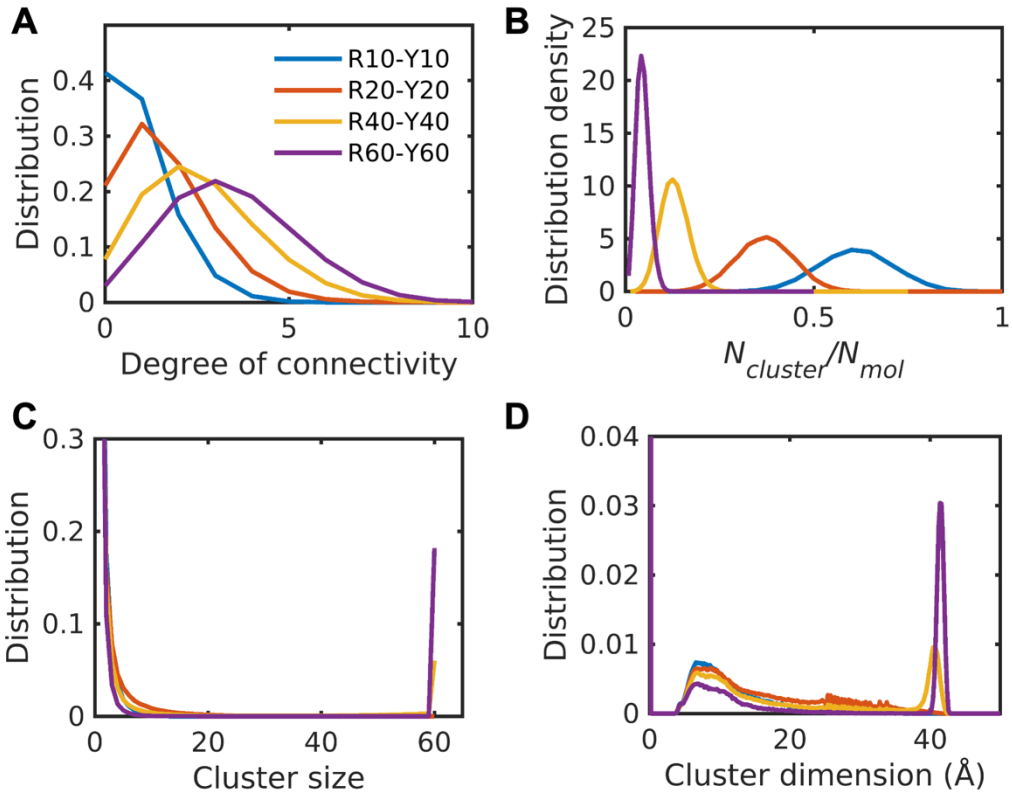

Figure 10: Dense phase mimics modeled using capped amino acids form percolated networks. Distribution of degree of connectivity of each capped amino acid (A), distribution density of number of clusters per molecule (B), and distribution of cluster size (C) and cluster dimension (D) within the dense phase. The degree of connectivity, cluster size and cluster dimension increase with increasing concentration, whereas the number of clusters per molecule decreases.

Results from the connectivity analysis are shown in **Figure 10**. The degree of connectivity, which quantifies the connectedness of a molecule within the dense phase shows a monotonic increase with increasing concentration (**Figure 10A**). Percolation is a continuous connectivity transition [23, 42], whereby above a threshold concentration the connectivity increases to span the entire system. The monotonic increase in the extent of clustering is also evidenced in the distributions of the number of clusters per molecule (**Figure 10B**). The distributions of cluster size show a bimodality at high concentrations, pointing to the incorporation of more than 75% of the molecules into the largest cluster. The largest cluster is system-spanning as evidenced by comparison of the maximum cluster dimension, which is equivalent to the dimension for the central simulation cell. Overall, the weakening in pairwise associations within dense phase mimics formed by capped amino acids is offset by the formation of higher-order connected clusters that

become system spanning at the highest concentrations. Taken together with the absence of long-range ordering, the findings suggest that backbone-mediated interactions drive the formation of a percolated network fluid within dense phases [28].

SUMMARY

In this work, we used all-atom simulations based on the AMOEBA polarizable forcefield to quantify differences between pairwise associations of aromatic and cationic sidechains in mimics of dilute versus dense phases of protein-based condensates. The design of the simulations leverages the fact that protein concentrations in dense phases place these systems in the semidilute regime [17], well above the overlap concentration [14, 15, 18, 24]. This allows us to query how interactions between peptide-sized segments vary between dilute and dense phases. We systematically increased the complexity of the simulations to quantify the strengths and ranges of pairwise interactions between aromatic moieties and cation-π groups. For freely diffusing model compounds, the simulations show evidence of long-range ordering within dense phases that leads to the presence of nanoscale aromatic clusters and dispersed cationic compounds, ions, and water molecules. The hierarchies of interaction strengths are in accord with recent energy decompositions [43]. In our simulations, which are based on the AMOEBA forcefield [44], the pairwise associations are enhanced in dense phases compared to dilute phases. In contrast, for simulations of capped amino acids, we find that backbone-mediated interactions weaken pairwise associations between sidechains, and this causes a loss of long-range ordering, and the absence of nanoscale clusters of aromatic moieties. Instead, the fluidization, which is evidenced in the loss of long-range ordering, drives the formation of system-spanning networks that feature the totality of directional, sidechain-sidechain, sidechain-backbone, and backbone-backbone interactions.

Our simulations paint a new picture regarding the driving forces for phase separation of PLCD-like systems: Strong pairwise associations between sidechains [45] are present in dilute phases, and these, as shown experimentally, are direct determinants of the free energy of transfer of polypeptides into dense phases. These strong pairwise interactions [46] are consistent with the designation of aromatic and Arg residues as stickers. In fact, the intrinsic interaction energies are further enhanced in dense phase mimics based on model compounds, thus further corroborating the designation of these residues as stickers [46, 47]. The favorable inter-residue two-body interactions outcompete chain-solvent interactions and the entropy of mixing, thereby driving phase separation

[48]. However, within dense phases, the interaction patterns are different from the dilute phase. This is true for model compounds and for peptides. For model compounds, the lack of a backbone, which enables a range of rotations and translations of the compounds, enables long-range ordering and the formation of nanoscale clusters that coexist with regions that are devoid of these clusters. Backbones interfere with this clustering to weaken pairwise associations between sidechain moieties. This weakening of pairwise associations enables networking that enables percolation within dense phases. Indeed, recent studies [49] have shown evidence for condensates being percolated network fluids [11, 15, 28, 50]. Furthermore, the network structures directly explain the measured, sequence-specific viscoelastic moduli of different PLCD-based condensates [11].

Our studies lead to two key findings: First, the interactions within dilute and dense phases are quantifiably different and these differences are expected to be generators of interphase properties. Examples of such interphase properties are the recent demonstrations of passive pH gradients across coexisting phases that come about from equalizing the chemical potentials of protons across the phase boundaries [51]. Second, while dense phases of small molecules such as model compounds are likely to drive crystallization transitions, as evidenced by the nanoscale clustering and apparent phase separation within dense phases of model compounds, the peptide backbone intervenes with competing directional interactions such as hydrogen bonds to fluidize dense phases. This suggests that equilibrium liquid-to-solid transitions, which have been reported as being a consequence of the aging of metastable fluid-like condensates [11], involve solidification from the dilute phase [52] or at the interfaces of condensates [53]. Any aging driven by the networking of molecules within condensates must be a non-equilibrium transition [54] that may be thought of as a glass transition [55] or the formation of dynamically arrested networks [56].

In both dilute and dense phases, there is a clear hierarchy of pairwise interactions, and there is a positive correlation between these hierarchies. This explains why models [57] based on transferrable [57, 58] or learned potentials [15, 24] that are pairwise in nature provide accurate descriptions of phase separation driven by homotypic and heterotypic interactions. However, it is worth nothing that our dilute phase calculations, which are concordant with recent quantum mechanical energy decompositions [43], call into question some of the inferences based on simulations suggesting that Arg-π interactions are intrinsically stronger than π-π interactions [59]. We find that the dilute phase hierarchy for strengths of interactions for capped amino acids and model compounds with Arg / Lys as cations and Tyr / Phe as aromatic groups is as follows: Tyr-

Tyr > Phe-Phe > Arg-Tyr > Arg-Phe ≈ Lys-Tyr > Lys-Phe. The Trp-Trp and Arg-Trp interactions are the strongest of those that we have quantified.

Our simulations suggest that dense phases are fluidized and networked by the complex interplay of solvent-mediated sidechain-sidechain, backbone-sidechain, and backbone-backbone interactions. This appears to give rise to dense phases of fluid-like condensates with macromolecular concentrations that vary by less than an order of magnitude across systems even though the concentrations in coexisting dilute phases can vary by 4-7 orders of magnitude. In effect, at least for condensates formed by IDRs that are PLCD-like, the major contributors to the driving forces for phase separation appear to be interactions within the dilute phase. Our results emphasize the importance of backbone-mediated interactions as drivers of weakened pairwise interactions (which we take as a proxy for fluidity) and percolation within dense phases. This highlights the need for using coarse-grained models with explicit representations of polypeptide backbones as recently introduced by Zhang et al., [60]. Alternatively, one can use all-atom descriptions if computational resources are available [61]. As a general and robust method, the approach of Zhang et al., holds significant promise.

Finally, we reiterate that we leveraged the fact that dense phases are semidilute solutions to mimic interactions using peptide-sized segments. However, chain connectivity will impose topological constraints that cannot be captured by freely diffusing peptide-sized segments [62]. Of direct relevance are investigations based on polarizable forcefields that query the impact of chain connectivity on interactions and percolation transitions within dense phases. This is important because percolation will drive emergent and time-dependent properties of condensates.

Methods

Setup of molecular dynamics simulations using AMOEBA polarizable forcefield

All the simulations were performed using TINKER 9 package [63] employing the AMOEBA polarizable forcefield parameters [64]. The size of the central simulation cells, which was cubic, was $\sim 5 \times 5 \times 5\ \mathrm{nm}^3$ for both the dilute and dense phase simulations. The forcefield parameters for model compounds mimicking sidechains of Arg, Lys, Tyr and Phe were taken from our previous work [65]. The forcefield parameters amoebabio18.prm in the TINKER package were used for

capped amino acids. The isothermal-isobaric (NPT) ensemble was used. The temperature and pressure were maintained at 298 K and 1 bar, respectively. Other setup details about MD simulations were identical to our previous work [65]. The model compounds or peptide were solvated in a cubic water box with periodic boundary conditions. Following energy minimization, molecular dynamics simulations were performed using reversible reference system propagator algorithm integrator [66] with an inner time step of 0.25 ps and an outer time step of 2.0 fs in isothermal-isobaric ensemble (NPT) ensemble with the target temperature being between 273 and 400 K depending on the temperature of interest and the target pressure being 1 bar. The temperature and pressure were controlled using a stochastic velocity rescaling thermostat [67] and a Monte Carlo constant pressure algorithm [68], respectively. The particle mesh Ewald (PME) method [69] with PME-GRID being 36 grid units, an order 8 *B*-spline interpolation [70], with a real space cutoff of 7 Å was used to compute long-range corrections to electrostatic interactions. The cutoff for van der Waals interactions was set to be 12 Å. This combination of a shorter cutoff for PME real space and longer cutoff for Buffered-14-7 potential has been verified [71] for AMOEBA free energy simulations [72]. Snapshots were saved every ten ps.

Calculation of PMFs in the dilute phase using umbrella sampling

For the dilute phase simulations, a pair of model compounds or capped amino acids were placed in the simulation box. Umbrella sampling [73] was used to calculate the PMFs of interest. For the dense phase simulations, multiple pairs of model compounds or capped amino acids were placed in the simulation box. The number of pairs were estimated based on the volume fractions of dense phases gleaned from measured binodals of A1-LCD and designed variants thereof [9]. Five independent simulations with different starting structures were performed for simulations that mimic dense phases. Each independent simulation was 100 ns long.

The interactions between pairs of model compounds or capped amino acids in the dilute phase were calculated using umbrella sampling. A harmonic potential with $k = 0.5\ \mathrm{kcal/mol\text{-}Å^2}$

was applied to restrain the distance between two model compounds or capped amino acids. The geometry center of the key functional group in the model compound or the side chain of the capped amino acids is used to calculate the distance. For Arg and Lys , the key functional groups are the guanidinium and ammonium groups, respectively. For Try and Phe, the aromatic ring is the key functional group. The center of the windows for umbrella are 3 Å, and 4 Å to 20 Å with an interval of 2 Å. In each window, multiple independent simulation trajectories with different starting structures are performed. The total simulation length for each window ranges from 240 ns to 260 ns. The weighted histogram analysis method (WHAM) [74] is used to obtain PMFs from umbrella sampling simulations. The distributions of pairwise distances, $P(r)$ were first extracted from the simulations. These were then used to compute the PMFs of interest using: $W(r) = -k_B T \, ln\left(\frac{P(r)}{4\pi r^2}\right)$. Here, k_B is the Boltzmann constant, and $T = 298$ K is the simulation temperature.

Estimation of error bars

For the error estimation of excluded volume in the dilute phase, we used bootstrapping with replacement. First, 160 ns of simulation data in each umbrella sampling window was randomly chosen to calculate the PMF for a specific pair of model compounds or capped amino acids. The PMF was further used to calculate the excluded volume. Next, the above process was performed independently ten times, resulting in ten estimates of the excluded volume. The standard deviation of these ten estimates were the statistical uncertainties for the excluded volume.

For the error estimation of excluded volume in the dense phase, five values of the excluded volume can be calculated using each of the five independent simulation trajectories. Then, the standard deviation of these five estimated values was used to represent the error bars for excluded volume.

ASSOCIATED CONTENT

The Supporting Information includes all supplemental figures and **Table S1**.

AUTHOR INFOMRATION

Corresponding Authors

xiangzezeng@hkbu.edu.hk

pappu@wustl.edu

Author Contributions

Conceptualization: XZ and RVP; Methodology: XZ and RVP; Analysis: XZ and RVP; Funding acquisition: RVP; Project administration: RVP; Supervision: RVP; Writing: and Editing: XZ and RVP.

Notes

RVP is a member of the scientific advisory board and shareholder of Dewpoint Therapeutics Inc. XZ does not have any financial interests to declare.

ACKNOWLEDGMENTS

This work was funded by the US Air Force Office of Scientific Research grant (FA9550-20-1-0241 to RVP), the St. Jude Research Collaborative on the Biology and Biophysics of RNP Granules (to RVP), the US National Science Foundation (MCB-2227268 to RVP), and the Center for Biomolecular Condensates in the James McKelvey School of Engineering at Washington University in St. Louis.

REFERENCES

(1) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews in Molecular and Cell Biology* **2017**, *18* (5), 285-298. DOI: 10.1038/nrm.2017.7.

(2) Pappu, R. V.; Cohen, S. R.; Dar, F.; Farag, M.; Kar, M. Phase Transitions of Associative Biomacromolecules. *Chemical Reviews* **2023**, *123*, 8945-8987. DOI: 10.1021/acs.chemrev.2c00814.

(3) Ruff, K. M.; Choi, Y. H.; Cox, D.; Ormsby, A. R.; Myung, Y.; Ascher, D. B.; Radford, S. E.; Pappu, R. V.; Hatters, D. M. Sequence grammar underlying the unfolding and phase separation of globular proteins. *Molecular Cell* **2022**, *82* (17), 3193-3208.e3198. DOI: 10.1016/j.molcel.2022.06.024.

(4) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; et al. A Molecular Grammar Governing the

Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e616. DOI: 10.1016/j.cell.2018.06.006 (acccessed 2021/07/17).

(5) Protter, D. S. W.; Parker, R. Principles and Properties of Stress Granules. *Trends in Cell Biology* **2016**, *26* (9), 668-679. DOI: 10.1016/j.tcb.2016.05.004.

(6) Currie, S. L.; Rosen, M. K. Using quantitative reconstitution to investigate multicomponent condensates. *RNA* **2022**, *28* (1), 27-35. DOI: 10.1261/rna.079008.121.

(7) Youn, J.-Y.; Dyakov, B. J. A.; Zhang, J.; Knight, J. D. R.; Vernon, R. M.; Forman-Kay, J. D.; Gingras, A.-C. Properties of Stress Granule and P-Body Proteomes. *Molecular Cell* **2019**, *76* (2), 286-294. DOI: 10.1016/j.molcel.2019.09.014.

(8) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367* (6478), 694-699.

(9) Bremer, A.; Farag, M.; Borcherds, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature Chemistry* **2022**, *14* (2), 196-207. DOI: 10.1038/s41557-021-00840-w.

(10) Cohan, M. C.; Shinn, M. K.; Lalmansingh, J. M.; Pappu, R. V. Uncovering Non-random Binary Patterns Within Sequences of Intrinsically Disordered Proteins. *Journal of Molecular Biology* **2022**, *434* (2), 167373. DOI: 10.1016/j.jmb.2021.167373.

(11) Alshareedah, I.; Borcherds, W., M. ; Cohen, S., R. ; Singh, A.; Posey, A. E.; Farag, M.; Bremer, A.; Strout, G. W.; Tomares, D. T.; Pappu, R., V. ; et al. Sequence-specific interactions determine viscoelasticity and aging dynamics of protein condensates. *Nature Physics* **2024**, *https://doi.org/10.1038/s41567-024-02558-1*.

(12) Elbaum-Garfinkle, S.; Kim, Y.; Szczepaniak, K.; Chen, C. C.-H.; Eckmann, C. R.; Myong, S.; Brangwynne, C. P. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences USA* **2015**, *112* (23), 7189-7194. DOI: 10.1073/pnas.1504822112.

(13) Holehouse, A. S.; Ginell, G. M.; Griffith, D.; Böke, E. Clustering of Aromatic Residues in Prion-like Domains Can Tune the Formation, State, and Organization of Biomolecular Condensates. *Biochemistry* **2021**, *60* (47), 3566-3581. DOI: 10.1021/acs.biochem.1c00465.

(14) Wei, M. T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry* **2017**, *9* (11), 1118-1125. DOI: 10.1038/nchem.2803.

(15) Farag, M.; Cohen, S. R.; Borcherds, W. M.; Bremer, A.; Mittag, T.; Pappu, R. V. Condensates formed by prion-like low-complexity domains have small-world network structures and interfaces defined by expanded conformations. *Nature Communications* **2022**, *13*, 7722.

(16) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Archives of Biochemistry and Biophysics* **2008**, *469* (1), 132-141. DOI: https://doi.org/10.1016/j.abb.2007.08.033.

(17) De Gennes, P. G.; Pincus, P.; Velasco, R. M.; Brochard, F. Remarks on polyelectrolyte conformation. *J. Phys. France* **1976**, *37* (12), 1461-1473.

(18) Pincus, P. Excluded Volume Effects and Stretched Polymer Chains. *Macromolecules* **1976**, *9* (3), 386-388. DOI: 10.1021/ma60051a002.

(19) Uematsu, T.; Svanberg, C.; Jacobsson, P. A Unified Picture of Static and Dynamic Length Scales in Polymer Solutions. *Macromolecules* **2005**, *38* (14), 6227-6230. DOI: 10.1021/ma050478t.

(20) De Gennes, P. G. *Scaling concepts in polymer physics*; Cornell University Press, 1979.

(21) Muthukumar, M. Thermodynamics of polymer solutions. *The Journal of Chemical Physics* **1986**, *85* (8), 4722-4728. DOI: 10.1063/1.451748.

(22) Rieger, J. Conformational statistics of polymers: a unifying approach comprising broken rods, blobs, and simple random walks. *Polymer* **1998**, *39* (18), 4417-4420. DOI: 10.1016/S0032-3861(97)10367-6.

(23) Choi, J.-M.; Hyman, A. A.; Pappu, R. V. Generalized models for bond percolation transitions of associative polymers. *Physical Review E* **2020**, *102* (4), 042403. DOI: 10.1103/PhysRevE.102.042403.

(24) Farag, M.; Borcherds, W. M.; Bremer, A.; Mittag, T.; Pappu, R. V. Phase separation of protein mixtures is driven by the interplay of homotypic and heterotypic interactions. *Nature Communications* **2023**, *14*, 5527. DOI: 10.1101/2023.03.15.532828.

(25) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 4046-4063. DOI: 10.1021/ct4003702.

(26) Kilgore, H. R.; Mikhael, P. G.; Overholt, K. J.; Boija, A.; Hannett, N. M.; Van Dongen, C.; Lee, T. I.; Chang, Y.-T.; Barzilay, R.; Young, R. A. Distinct chemical environments in biomolecular condensates. *Nature Chemical Biology* **2024**, *20* (3), 291-301. DOI: 10.1038/s41589-023-01432-0.

(27) Wang, Z.; Qiu, W.; Zhang, Q. Constructing Phase Separation in Polymer Gels: Strategies, Functions and Applications. *Progress in Polymer Science* **2024**, 101847. DOI: 10.1016/j.progpolymsci.2024.101847.

(28) Dar, F.; Cohen, S. R.; Mitrea, D. M.; Phillips, A. H.; Nagy, G.; Leite, W. C.; Stanley, C. B.; Choi, J.-M.; Kriwacki, R. W.; Pappu, R. V. Biomolecular condensates form spatially inhomogeneous network fluids. *Nature Communications* **2024**, *15* (1), 3413. DOI: 10.1038/s41467-024-47602-z.

(29) Sinnokrot, M. O.; Sherrill, C. D. High-Accuracy Quantum Mechanical Studies of π−π Interactions in Benzene Dimers. *The Journal of Physical Chemistry A* **2006**, *110* (37), 10656-10668. DOI: 10.1021/jp0610416.

(30) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *elife* **2018**, *7*, e31486.

(31) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press, 2003.

(32) Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **2017**, *6*, 30294. DOI: 10.7554/eLife.30294.

(33) Huggins, M. L. Solutions of Long Chain Compounds. *The Journal of Chemical Physics* **1941**, *9* (5), 440-440. DOI: 10.1063/1.1750930.

(34) Flory, P. J. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **1942**, *10* (1), 51-61. DOI: 10.1063/1.1723621.

(35) Farag, M.; Holehouse, A. S.; Zeng, X.; Pappu, R. V. FIREBALL: A tool to fit protein phase diagrams based on mean-field theories for polymer solutions. *Biophysical Journal* **2023**, *122* (12), 2396-2403. DOI: 10.1016/j.bpj.2023.05.007.

(36) Luo, C.; Qiang, Y.; Zwicker, D. Beyond pairwise: Higher-order physical interactions affect phase separation in multicomponent liquids. *Physical Review Research* **2024**, *6* (3), 033002. DOI: 10.1103/PhysRevResearch.6.033002.

(37) Choi, J.-M.; Dar, F.; Pappu, R. V. LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology* **2019**, *15* (10), e1007028. DOI: 10.1371/journal.pcbi.1007028.

(38) Stanley, H. E.; Andrade, J. S.; Havlin, S.; Makse, H. A.; Suki, B. Percolation phenomena: a broad-brush introduction with some recent applications to porous media, liquid water, and city growth. *Physica A: Statistical Mechanics and its Applications* **1999**, *266* (1), 5-16. DOI: 10.1016/S0378-4371(99)00029-1.

(39) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annual Review of Biophysics* **2020**, *49*, 107-133.

(40) McColl, W. F.; Noshita, K. On the number of edges in the transitive closure of a graph. *Discrete Applied Mathematics* **1986**, *15* (1), 67-73. DOI: 10.1016/0166-218X(86)90020-X.

(41) Newman, M. E. J. *Networks: An Introduction*; Oxford University Press, 2010.

(42) Stauffer, D.; Aharony, A. *Introduction To Percolation Theory*; Taylor & Francis, 1992.

(43) Carter-Fenk, K.; Liu, M.; Pujal, L.; Loipersberger, M.; Tsanai, M.; Vernon, R. M.; Forman-Kay, J. D.; Head-Gordon, M.; Heidar-Zadeh, F.; Head-Gordon, T. The Energetic Origins of Pi–Pi Contacts in Proteins. *Journal of the American Chemical Society* **2023**, *145* (45), 24836-24851. DOI: 10.1021/jacs.3c09198.

(44) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114* (8), 2549-2564. DOI: 10.1021/jp910674d.

(45) Grimme, S. Do Special Noncovalent π–π Stacking Interactions Really Exist? *Angewandte Chemie International Edition* **2008**, *47* (18), 3430-3434. DOI: 10.1002/anie.200705157.

(46) Tanaka, F. Theory of Molecular Association and Thermoreversible Gelation. In *Molecular Gels: Materials with self-assembled fibrillar networks*, Springer, 2006; pp 17-78.

(47) Semenov, A. N.; Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* **1998**, *31*, 1373-1385.

(48) Rubinstein, M.; Dobrynin, A. V. Solutions of Associative Polymers. *Trends in Polymer Science* **1997**, *5* (6), 181-186.

(49) Alshareedah, I.; Singh, A.; Yang, S.; Ramachandran, V.; Quinn, A.; Potyan, D. A.; Banerjee, P. R. Determinants of Viscoelasticity and Flow Activation Energy in Biomolecular Condensates. *Science Advances* **2024**, *10* (7), eadi6539. DOI: 10.1126/sciadv.adi6539.

(50) Cohen, S. R.; Banerjee, P. R.; Pappu, R. V. Direct computations of viscoelastic moduli of biomolecular condensates. *bioRxiv* **2024**, 2024.2006.2011.598543. DOI: 10.1101/2024.06.11.598543.

(51) King, M. R.; Ruff, K. M.; Lin, A. Z.; Pant, A.; Farag, M.; Lalmansingh, J. M.; Wu, T.; Fossat, M. J.; Ouyang, W.; Lew, M. D.; et al. Macromolecular condensation organizes nucleolar

sub-phases to set up a pH gradient. *Cell* **2024**, *187* (8), 1889-1906.e1824. DOI: 10.1016/j.cell.2024.02.029.

(52) Das, T.; Zaidi, F.; Farag, M.; Ruff, K. M.; Messing, J.; Taylor, J. P.; Pappu, R. V.; Mittag, T. Metastable condensates suppress conversion to amyloid fibrils. *bioRxiv* **2024**, 2024.2002.2028.582569. DOI: 10.1101/2024.02.28.582569.

(53) Linsenmeier, M.; Faltova, L.; Palmiero, U. C.; Seiffert, C.; Küffner, A. M.; Pinotsi, D.; Zhou, J.; Mezzenga, R.; Arosio, P. The interface of condensates of the hnRNPA1 low complexity domain promotes formation of amyloid fibrils. *bioRxiv* **2022**, 2022.2005.2023.493075. DOI: 10.1101/2022.05.23.493075.

(54) Lin, A. Z.; Ruff, K. M.; Dar, F.; Jalihal, A.; King, M. R.; Lalmansingh, J. M.; Posey, A. E.; Erkamp, N. A.; Seim, I.; Gladfelter, A. S.; et al. Dynamical control enables the formation of demixed biomolecular condensates. *Nature Communications* **2023**, *14* (1), 7678. DOI: 10.1038/s41467-023-43489-4.

(55) Jawerth, L.; Fischer-Friedrich, E.; Saha, S.; Wang, J.; Franzmann, T.; Zhang, X.; Sachweh, J.; Ruer, M.; Ijavi, M.; Saha, S. Protein condensates as aging Maxwell fluids. *Science* **2020**, *370* (6522), 1317-1323.

(56) Boeynaems, S.; Holehouse, A. S.; Weinhardt, V.; Kovacs, D.; Van Lindt, J.; Larabell, C.; Van Den Bosch, L.; Das, R.; Tompa, P. S.; Pappu, R. V.; et al. Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proceedings of the National Academy of Sciences USA* **2019**, *116*, 7889-7898. DOI: 10.1073/pnas.1821038116.

(57) Pak, Chi W.; Kosno, M.; Holehouse, Alex S.; Padrick, Shae B.; Mittal, A.; Ali, R.; Yunus, Ali A.; Liu, David R.; Pappu, Rohit V.; Rosen, Michael K. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Molecular Cell* **2016**, *63* (1), 72-85. DOI: 10.1016/j.molcel.2016.05.042 (acccessed 2022/01/26).

(58) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties.

Proceedings of the National Academy of Sciences **2021**, *118* (44), e2111696118. DOI: doi:10.1073/pnas.2111696118.

(59) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **2021**, *1* (11), 732-743. DOI: 10.1038/s43588-021-00155-3.

(60) Zhang, Y.; Li, S.; Gong, X.; Chen, J. Toward Accurate Simulation of Coupling between Protein Secondary Structure and Phase Separation. *Journal of the American Chemical Society* **2024**, *146* (1), 342-357. DOI: 10.1021/jacs.3c09195.

(61) Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. Extreme dynamics in a biomolecular condensate. *Nature* **2023**, *619*, 876-883. DOI: 10.1038/s41586-023-06329-5.

(62) Zeng, X.; Holehouse, A. S.; Chilkoti, A.; Mittag, T.; Pappu, R. V. Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins. *Biophysical Journal* **2020**, *119* (2), 402-418. DOI: 10.1016/j.bpj.2020.06.014.

(63) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018**, *14* (10), 5273-5289. DOI: 10.1021/acs.jctc.8b00529.

(64) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A. Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010**, *114* (8), 2549-2564.

(65) Zeng, X.; Liu, C.; Fossat, M. J.; Ren, P.; Chilkoti, A.; Pappu, R. V. Design of intrinsically disordered proteins that undergo phase transitions with lower critical solution temperatures. *APL Materials* **2021**, *9* (2), 021119.

(66) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *The Journal of Chemical Physics* **1991**, *94* (10), 6811-6815. DOI: 10.1063/1.460259.

(67) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* **2009**, *130* (7), 074101. DOI: 10.1063/1.3073889.

(68) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004**, *384* (4), 288-294. DOI: https://doi.org/10.1016/j.cplett.2003.12.039.

(69) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(70) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577-8593. DOI: 10.1063/1.470117.

(71) Jing, Z.; Liu, C.; Qi, R.; Ren, P. Many-body effect determines the selectivity for Ca²⁺ and Mg²⁺ in proteins. *Proceedings of the National Academy of Sciences* **2018**, *115* (32), E7495. DOI: 10.1073/pnas.1805049115.

(72) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences* **2008**, *105* (17), 6290. DOI: 10.1073/pnas.0711686105.

(73) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23* (2), 187-199. DOI: 10.1016/0021-9991(77)90121-8.

(74) Ferrenberg, A. M.; Swendsen, R. H. Optimized monte carlo data analysis. *Computers in Physics* **1989**, *3* (5), 101-104. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. Temperature weighted histogram analysis method, replica exchange, and transition paths. *The Journal of Physical Chemistry B* **2005**, *109* (14), 6722-6731.