Decoding Phase Separation of Prion-Like Domains through Data-Driven Scaling Laws

M. Julia Maristany[1], Anne Aguirre Gonzalez[2], Jorge R. Espinosa[3], Jan Huertas[2], Rosana Collepardo-Guevara[1,2,4], Jerelle A. Joseph[5,6*]

[1]Department of Physics, University of Cambridge, Cambridge, UK.
[2]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK.
[3]Universidad Complutense de Madrid, Madrid, Spain.
[4]Department of Genetics, University of Cambridge, Cambridge, UK.
[5]Department of Chemical and Biological Engineering, Princeton University, USA.
[6]Omenn–Darling Bioengineering Institute, Princeton University, USA.

*Corresponding author(s). E-mail(s): jerellejoseph@princeton.edu;

Abstract

Proteins containing prion-like low complexity domains (PLDs) are common drivers of the formation of biomolecular condensates and are prone to misregulation due to amino acid mutations. Here, we exploit the accuracy of our residue-resolution coarse-grained model, Mpipi, to quantify the impact of amino acid mutations on the stability of an unprecedented set of 140 PLD mutants from six proteins (hnRNPA1, TDP43, FUS, EWSR1, RBM14, and TIA1). Our simulations reveal the existence of scaling laws that quantify the range of change in the critical solution temperature of PLDs as a function of the number and type of amino acid sequence mutations. Remarkably, these rules are consistent with the physicochemical properties of the mutations and extend across the entire family tested, suggesting that scaling laws can be used as tools to predict changes in the stability of PLD condensates. Our work offers a quantitative lens into how the emergent behavior of PLD solutions varies in response to physicochemical changes of single PLD molecules.

Keywords: Biomolecular Condensates, Coarse-Grained Simulations, Phase Separation, Prion-like Domains, Big Dataset

1 Introduction

Biomolecular condensates are highly multicomponent systems with remarkable specificity: they exhibit non-random molecular compositions and, concomitantly, contain distinct physicochemical environments defined by the collective behavior of their components [1, 2]. Phase separation has emerged as a leading mechanism to account for the formation of biomolecular condensates. In this thermodynamic process, segregative (density-driven) and associative (connectivity-driven) transitions are coupled to produce protein- and nucleic acid-rich phases suspended in the cellular cytoplasm or nucleoplasm [3]. Within this framework, intrinsically disordered proteins/regions (IDP/IDRs) have emerged as strong contributors to the multivalency and phase separation capacity of many naturally occurring proteins [1, 4–12]. Moreover, IDRs characterized by amino acid sequences of low complexity and enriched in aromatic residues, such as the prion-like low-complexity domains (PLDs) of RNA–binding proteins [8, 13–16], are prone to form condensates that behave like viscoelastic fluids. Furthermore, the PLDs of various naturally occurring proteins, such as Fused in Sarcoma (FUS), Trans-activation response DNA-binding protein 43 (TDP-43), and the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) have been reported to transition from liquid-like condensates into dynamically arrested solids and glasses over time [17, 18]. Such arrested states have been postulated as precursors of pathological protein aggregates and amyloids [19]. Understanding the molecular factors that regulate the ability of PLDs to phase separate is of crucial importance, as these are conjectured to also facilitate their aggregation [20].

Biomolecular condensates are networked fluids sustained by multivalent macromolecules that interconnect with more than two partners simultaneously [1, 2, 21–27]. Thus, the factors that increase the molecular valency, give rise to more densely connected liquid networks, and more stable condensates condensates [1, 21, 22, 25, 28–31]. For PLDs, multivalency has been shown to be strongly promoted by the presence of aromatic residues, which act as "stickers" that enable associative π–π interactions, and also cation–π contacts involving arginine and lysine [8, 9, 32–38]. Indeed, the range of coexistence of PLD condensates has been shown to change most significantly in response to mutations involving aromatic residues [9, 39]. Quantitative experimental phase diagrams for a set of hnRNPA1-IDR variants in combination with mean-field simulations demonstrated that tyrosine is a stronger sticker than phenylalanine, arginine is a context-specific sticker and lysine destabilizes PLD–PLD interactions [35]. More importantly, such study demonstrates that, at least for hnRNPA1, there exist quantitative rules that explain the changes in thermodynamic parameters of condensates in response to precise amino-acid sequence mutations.

In this work, we investigate whether there are scaling laws that can predict the changes in critical parameters of a broader collection of PLDs in response to amino acid sequence mutations. Scaling laws are mathematical relationships that describe how the properties of a system change for a given variable, such as its size or composition [40]. In the context of polymer-like systems such as IDPs, scaling laws have offered invaluable insight into vastly different aspects of their physical behavior: for example, they were instrumental in determining that the relationship between viscosity and radius of gyration in polymer solutions is described by a power law between the two quantities [41]; demonstrating that aging scaling during polymer collapse is a universal property in homopolymers [42]; determining that in single polymer systems, polymer end-to-end distance is directly proportional to their length [43]; showing that foldable sequences with more compact unfolded states are due to protein evolution [44]; and, importantly, illustrating that IDPs indeed behave like a polymer in a good solvent [45]. In the context of protein condensates, we hypothesized that scaling laws, if found, may allow us to quantify how condensates respond to different intrinsic perturbations—such as amino acid mutations or presence/absence of different post-translational modifications, changes in protein length, increased/decreased flexibility—or external stimuli—such as changes to the pH of the medium or the ionic environment. Thus, we focused on on determining if such laws exist for PLDs by exploring six different RNA-binding proteins (i.e., hnRNPA1, TDP43, FUS, EWSR1, RBM14, and TIA1; Fig. 1 A,B), linked to the etiology of neurodegenerative disorders through pathology and genetics [46–56].

Leveraging the accuracy and efficiency of a modern sequence-dependent coarse-grained model [34], we quantify the phase diagrams (temperature-vs-density plane) of an unprecedentedly large set of mutants spanning all six PLDs mentioned above. We performed Direct Coexistence molecular dynamics (MD) simulations (Fig. 1 C), exploring simultaneously the protein-rich and protein-depleted phases of the condensate for each mutant, in the temperature–concentration phase space. Our massive set of simulations encompasses 140 PLD mutants, each sampled at 8 to 12 different temperatures. Accounting for independent replicas to ensure sufficient sampling and convergence, the dataset presented here comprises ∼2000 independent Direct Coexistence simulations (Fig. 1 D), totaling 1.2 milliseconds of accumulated sampling.

Our work puts forward a data-driven scaling-law framework for the study of PLD phase behavior. We reveal that the changes in the critical solution temperatures of PLDs, as a function of the number and type of amino acid sequence mutations (e.g., aromatic, positively charged, etc), can indeed be quantified by scaling laws, which are highly conserved across the family of PLDs we investigate. The addition of aromatic residues yields a positive linear correlation between mutation fraction and critical temperature changes, consistent with the dominant role of aromatic residues in PLD phase separation. We find that protein length dictates scaling effects when aromatic residues are mutated into uncharged, non-aromatic amino acids. We also quantify the effects of arginine mutations on condensate stability, where removing arginines decreases the critical temperature of the condensate—revealing an inverse relationship between the fraction of arginine deletions and critical temperatures. Finally, polar amino acids emerge as subtle modulators of the phase behavior of PLDs. By examining the impact of different amino acid substitutions on the critical solution temperature, our results provide insights into the underlying molecular mechanisms of protein condensation. The scaling laws presented in this study serve two purposes: firstly, to verify the generality of the molecular grammar that governs the phase behavior across various PLD proteins, and secondly, to quantify the extent to which the critical solution temperature of a PLD protein solution will change in response to specific amino acid sequence changes. We anticipate that scaling laws for other families of proteins with very different compositional biases (e.g., transcription factors, nucleoporins, or molecular chaperones) may exist, but could be distinct from those we report here for PLDs. Overall, our work suggests that scaling laws represent useful predictors of the phase behavior of protein families.

2 Results

2.1 Generating and Analyzing an Extensive Dataset of Critical Solution Temperatures for Prion-like Domain Variants

The phase behavior of PLDs and their modulation with amino acid sequence mutations has been extensively investigated with a variety of experimental techniques and with theory and simulation methods for systems including the hnNRPA1 PLD [35, 57], the FUS PLD [8, 52, 54, 58–66], and the TDP-43 PLD [48, 49, 53, 67, 68], among many others. Here, we perform residue-resolution coarse-grained Molecular Dynamics (MD) simulations (see Fig. 1 C, with further details in Methods) with our Mpipi model [34] to quantitatively compare the critical parameters of a total of 140 mutants of the PLDs of TDP-43, FUS, EWSR1, hnRNPA1, RBM14, and TIA1 (Fig. 1 A) under consistent solution conditions. From our simulations, we compute temperature-vs-density phase diagrams and estimate the critical solution temperature for each system. The PLDs of the proteins we investigate have similar composition biases—i.e., are highly rich in neutral, polar, and aromatic residues—and present less than 5% of total charged amino acids in their sequence (Fig. 1 B). Thus, our simulation results allow us to assess if there exists a code—i.e., a simple set of scaling laws—that can predict how the critical parameters of the family of PLDs will scale in response to amino acid sequence mutations.

Our residue-resolution model Mpipi (see Methods 4.2.2) is ideally suited to investigate the phase behavior of PLDs because it achieves the correct balance of the dominant π–π and cation–π interactions, which drive the phase behavior of PLDs. Indeed, the Mpipi model recapitulates the quantitative experimental phase diagrams of hnRNPA1 variants by Bremer and colleagues [35] with remarkable accuracy [34]. It is worth noting that the Mpipi model was not parameterized based on the experimental dataset of Bremer and colleagues, but rather the model parameters were developed independently by combining atomistic umbrella sampling MD simulations of amino acid pairs with the bioinformatics data of Vernon and colleagues for π-driven contacts [37]. Using Flory–Huggins–Staverman theory, critical temperatures for phase separation were extracted from the experimental data of Bremer et al., which were then used to validate the Mpipi model [34].

The amino acid sequences of all the proteins we investigated (Supporting Information) were designed based on five main considerations. (1) The set of mutants should serve to probe the effects of both the strong modulators of PLD phase separation—namely, aromatic and arginine interactions—and the subtle modulators—that is, polar and neutral amino acids, previously identified experimentally [8, 9, 35]. (2) To ensure that the observed changes in the critical solution temperatures can be attributed unequivocally to specific amino-acid-type changes, we mutated only one amino acid type at a time. However, multiple mutations of the same amino acid type were allowed. (3) Mutations were distributed homogeneously across the entire protein. (4) Deleted amino acids (e.g. aromatic deletions) were replaced by weakly interacting amino acids (i.e. serines, threonines, glycines, and alanines). For each PLD protein, the identity of the weakly interacting residues used as replacements were chosen to preserve the percentages of the various types of weakly interacting amino acids present originally in the unmodified PLD. (5) Achieve a data set that spans a wide array of critical solution temperatures, as shown in Fig. 1 D.

2.1.1 Extended Validation of the Mpipi Model for Phase behavior of Prion-like Domains

The availability of quantitative experimental binodals for the hnRNPA1-IDR and various amino acid sequence variants [35] has allowed us to test the accuracy of the Mpipi model to describe the phase diagrams of PLD solutions [34]. Here, we make such a test more stringent by extending the hnRNPA1-IDR validation set from 9 PLD variants to 21, selecting mutations spanning changes in aromatic, charged, polar, and neutral residues. The amino acid sequences of all the hnRNPA1-IDR variants considered are described in the Supporting Information and are denoted using the intuitive nomenclature established by Bremer and colleagues: variants that mutate n residues of type X into m residues of type Z are labeled as $-nX+mZ$, and so forth. For more details on the nomenclature, see Methods.

For each hnRNPA1-IDR variant, we perform Direct Coexistence MD simulations, compute critical temperatures from the simulations, and compare them with the experimental critical temperatures (Fig. 2 A–D). We use the fitting methodology outlined in Methods to estimate the critical solution temperatures from the experimental data.

When comparing with experimental critical temperatures, our model achieves a Pearson correlation coefficient higher than 0.93 for all types of mutations analyzed. Additionally, for all cases the root-mean-square deviation (D) between the experimental and simulated critical temperatures is ≤ 14 Kelvin. Our extended validation set confirms that the Mpipi potential can effectively capture the thermodynamic behavior of

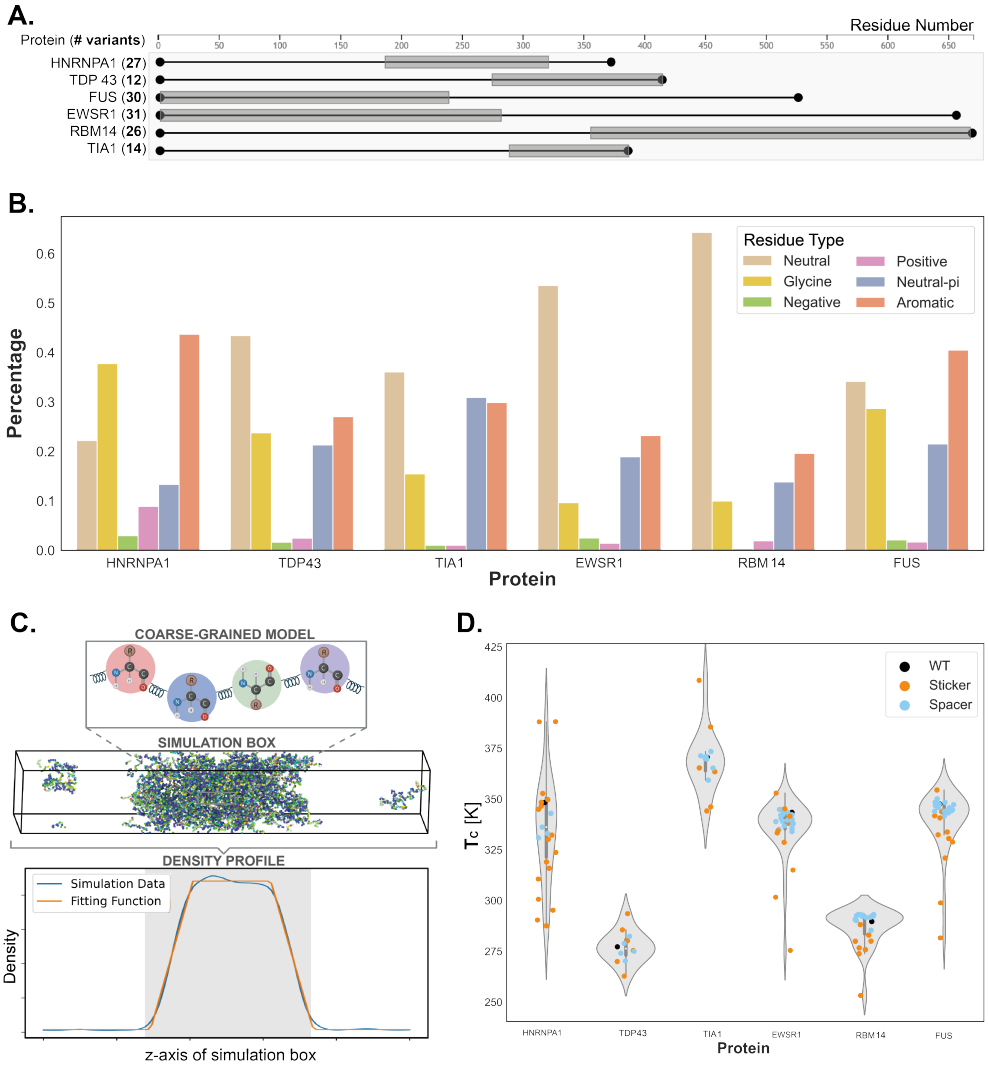

Fig. 1: **Data generation and representation of dataset for probing scaling laws of PLD phase behavior (A)** Representation of the prion-like domains (grey-shaded regions) of each considered protein, alongside the associated number of variants simulated, totaling a set of 140 sequences. **(B)** Composition of the wild-type sequence of the PLDs simulated, in terms of the percentage of neutral amino acids (A, C, I, L, M, P, S, T, and V; no net charge at pH 7 and no π electrons in the side chain), glycine (G), negative (D, E), positive (K, H, R), neutral-π (N and Q; no net charge at pH 7 with π electrons in the side chain) and aromatic (F, W, Y) residues. **(C)** Top panel: Representation of the simulation model: each amino acid of the protein is represented by a single bead, bonded interactions are modeled with springs, and non-bounded interactions are modeled with a combination of the Wang–Frenkel potential and a Coulomb term with Debye-Hückel screening. Middle panel: The Direct Coexistence simulation method, in the slab geometry, is employed, where both the protein-rich and protein-depleted phases are simulated simultaneously. Bottom panel: Finally, to obtain the data point in the concentration–temperature phase space, at a given temperature the density profile is computed and the concentrations of the rich and depleted phases are obtained via a suitable fitting (Methods). **(D)** Representation of the entire computational data set. Orange data points represent variants where charged or aromatic residues were mutated (i.e., stickers), while cyan data points represent all other types of mutations studied (i.e., spacers).

a wide range of hnRNPA1-PLD variants, and suggests that Mpipi is adequate for proteins with similar sequence compositions, as in the set of proteins analyzed in this study.

2.1.2 Large-Scale Data Generation on Critical Parameters of PLD Phase Separation

The goal of generating an unprecedentedly large set of binodal data for protein phase separation presented several unique computational challenges, from the management of considerable computational power to the application of efficient simulation techniques and the implementation of large-scale data analysis.

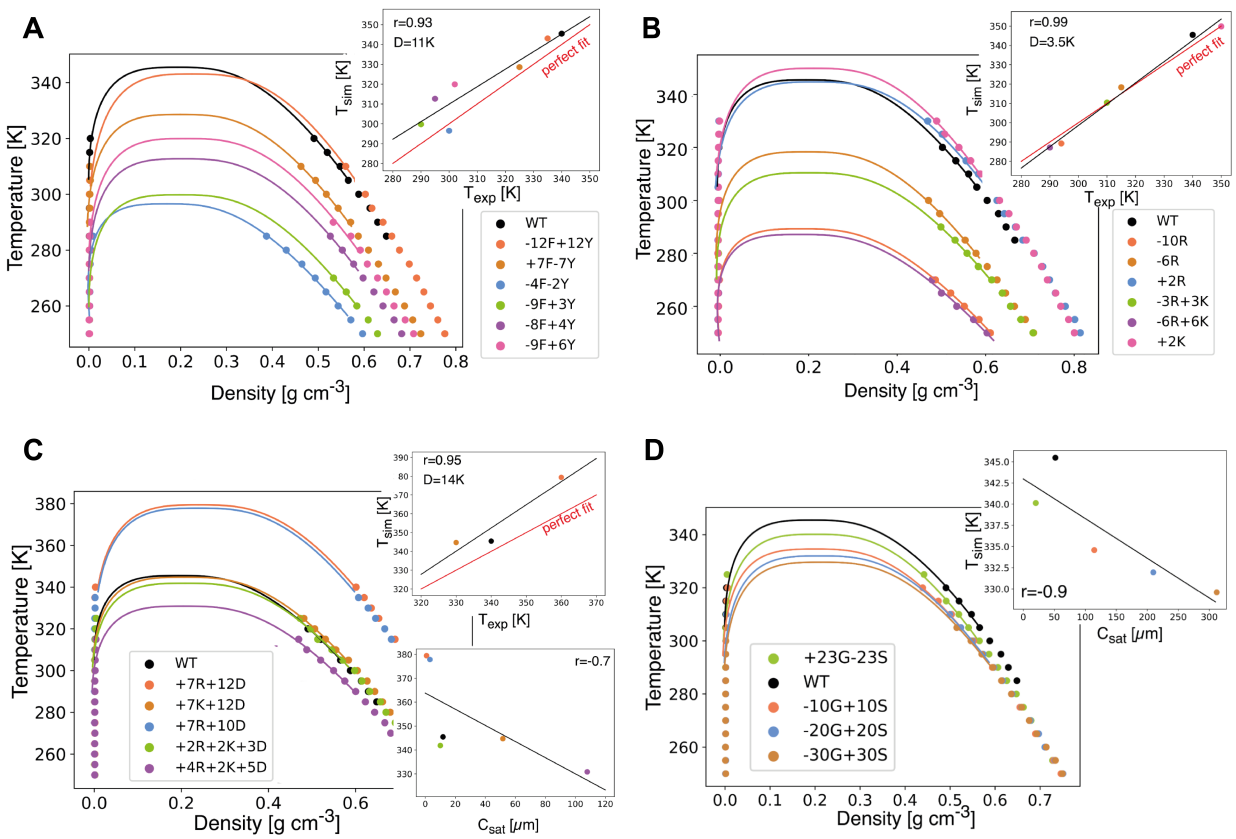

Fig. 2: **Mpipi model quantitatively captures phase behavior for a large set of hnRNPA1 PLD mutants.** Mutations are divided into **A** aromatic, **B** Arg/Lys, **C** charged, and **D** Gly/Ser sequence variants of the PLD of the protein hnRNPA1, and comparison with experimental results. The hnRNPA1 PLD was used to validate the accuracy of the model, by extracting critical temperatures from experimental phase diagrams of a large set of these variants [35]. To assess how well the Mpipi model performed for PLDs, we compared critical temperatures obtained from simulations with the corresponding ones extracted from the experimental data for those variants that were available (depicted in the top right of each binodal, where r is the Pearson coefficient and D is the root mean square deviation between simulated and experimental values). For the variants where the experimental critical temperature could not be determined, we make a comparison between the critical temperature with the saturation concentration, since for these systems we expect them to be inversely correlated. The naming system for the variants is discussed in detail in the text.

Due to the amount of computer resources required for the study, totaling over a quarter million CPU hours for the entire project, our first challenge was optimizing the efficiency of our Direct Coexistence MD simulations. In the Direct Coexistence method, the diluted phase and the condensate are simulated in the same rectangular box separated by an interface. The phase diagrams we compute explore the plane of temperature (vertical axis) versus density (horizontal axis). For each phase diagram, we run a set of about 20 Direct Coexistence simulations–each simulation explores the phase behavior of the system at different constant temperatures. For a given value of the temperature, if two phases are detected in our simulations, we measure a "density profile": the density of the proteins as a function of the long side of the simulation box (Bottom panel in Fig. 1 C). We then plot the two coexistence densities for each temperature we simulate and build the coexistence curve (also known as binodal), Fig. 3. The coexistence curve shows the values of temperature and protein density for which phase separation will occur. The region above the coexistence curve is the "one-phase region" where the temperature is too high for the enthalpic gain coming from the associative protein–protein interactions to overcome the entropic loss upon phase separation. The region below the coexistence curve, the "two-phase coexistence region" represents lower temperatures for which the enthalpic gain due to protein–protein interactions is now sufficiently high to favor phase separation into a condensed (protein-enriched) and a diluted (protein-depleted) liquid phase. The maximum in the coexistence curve is known as the critical point: the upper critical solution temperature (herein, critical solution temperature) and corresponding critical density. For temperatures higher than the critical solution temperature, phase separation is no longer observed.

The protein solutions that we investigate with our coarse-grained simulations must be sufficiently large, in terms of the number of copies of the same protein, to minimize finite size effects—for instance, to adequately

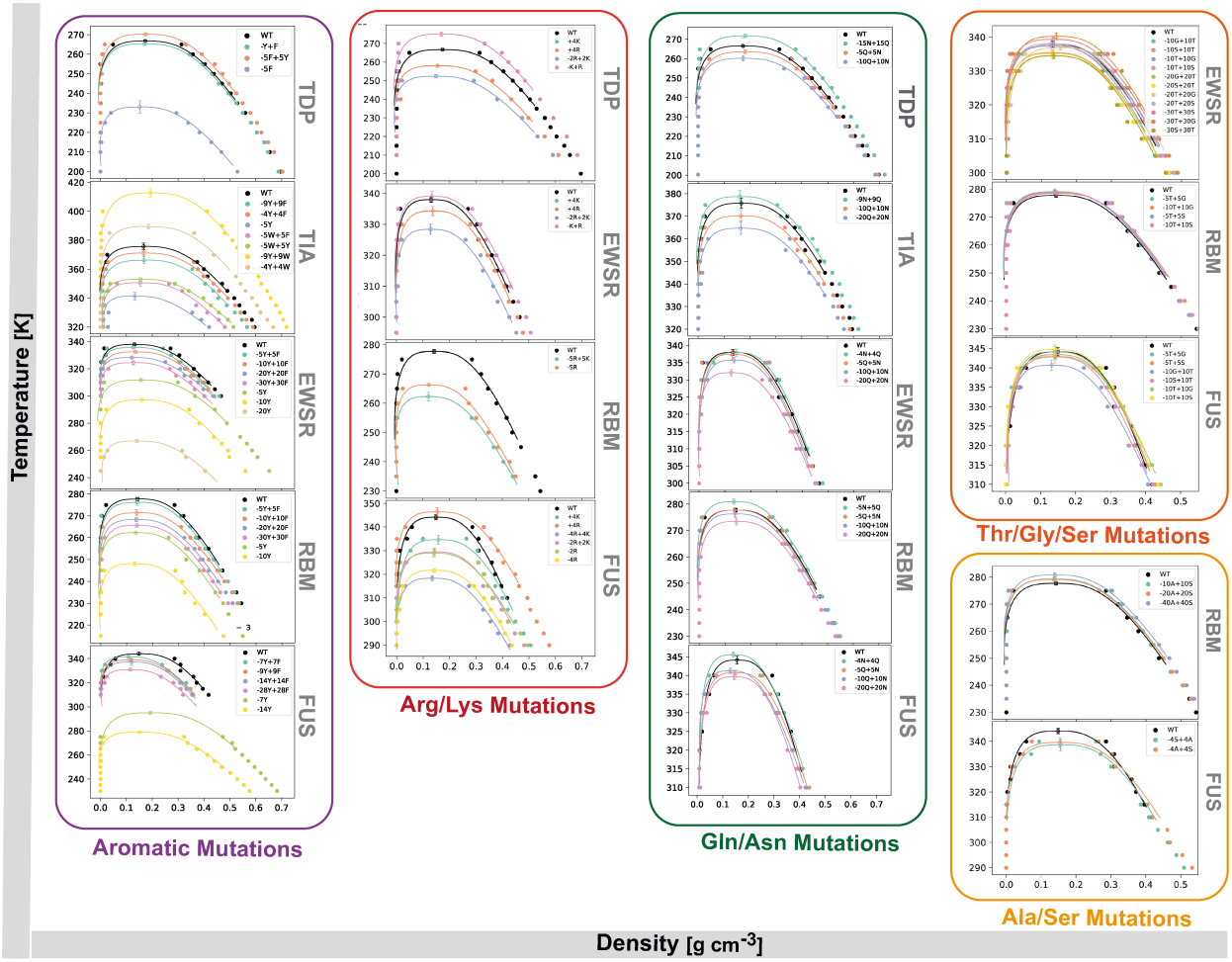

Fig. 3: **Complete set of 140 binodals, in the temperature versus density phase space, generated via largscale molecular dynamics simulations**. Binodals are grouped according to the mutation type (Aromatic, Arg/Lys, Gln/Asn (Polar), Ala/Ser and Thr/Gly/Ser (Neutral)). Each set is further grouped based on the PLD in question.

sample bulk phenomena without dominating interfacial effects. At the same time, our simulations must be small enough to ensure computational convergence in a feasible timescale. Following extensive testing of computational convergence for systems of different sizes and MD trajectories of different time lengths, we determined that a system composed of 64 protein replicas, with trajectories of 200 ns ensured rigorous convergence and reproducible results, while taking into careful consideration the trade-off between accuracy and computational demands. The size of the simulation boxes were also optimized, again to ensure finite size effects did not affect the predictions (See Methods and Materials, and Supporting Information).

To obtain 200 ns of MD trajectories in each computer run, with systems of up to 20,000 atoms, every data point required approximately 12 hours of wall time in 16 CPU nodes. Consequently, each binodal in Fig. 3 required 2,000 CPU hours for computation, and for all 140 protein variants considered, the cumulative computational requirement exceeded a staggering quarter of a million CPU hours. To efficiently manage the workload, since different temperatures of different variants are fully independent runs with no crossover communication between processors, the simulation runs were meticulously scheduled and the resources were dynamically allocated to maximize usage and reduce idle time and resource wastage. More details on technical computational requirements can be found in the Methods section.

Analyzing the data required a complex, multi-step procedure as well. The first step is thorough data cleaning, as well as ensuring that every equilibrium configuration generated via MD is uncorrelated from one another, followed by statistical analysis to interpret the results accurately, and with statistical rigor. All custom scripts developed for this study have been made publicly accessible, to facilitate replication, extension, or application of our methodologies, and are available in our GitHub repository.

Our research marks a significant milestone in the field of computational studies of protein phase behavior, primarily due to our integrated and automated approach focused on streamlining every step of our data generation and processing pipeline—from simulation and storage to analysis and interpretation. Our pipeline was designed to minimize manual intervention, reduce the potential for errors, and maximize the throughput

of data processing. As a result, we generated a dataset of critical parameters of unparalleled scale, of over an order of magnitude higher than previously reported studies [9, 35, 69–72]. This extensive dataset (Fig. 3) not only offers new insights into phase separation phenomena but also sets a new benchmark for data-driven research in this field.

2.2 A Data-Driven Scaling Law Approach for Studying Phase Behavior of Prion-like Domains

Scaling laws can reveal key relationships between system variables, such as composition and critical parameters, providing a useful theory to predict and interpret the regulation of complex biological phenomena such as phase separation. In the context of PLD phase behavior, scaling laws would provide a quantitative theory to analyze how alterations in amino acid composition affect the critical solution parameters—a crucial aspect for understanding protein behavior in biological systems. Here, our objective is to discern which scaling laws can effectively capture the variations of PLD critical parameters as a function of sequence mutations. To accomplish this, we employ a systematic approach of varying the amino acid sequence of the proteins considered, to determine how the phase behavior of PLDs changes with amino acid compositional perturbations.

For each type of mutation we analyze, we define and test three types of scaling laws. These scaling laws illustrate a different relationship between critical solution temperature changes and the number of amino acid mutations. The first and simplest scaling law is a linear relationship between these two quantities,

$$\Delta T_{\mathrm{c}} = S_{X_1 \to X_2} N_{X_1 \to X_2}, \tag{1}$$

where ΔT_{c} is the variation on the critical temperature between the mutated protein and the corresponding unmodified reference sequence, X_i is an amino acid (e.g., Y, F, R, and so forth) and $N_{X_1 \to X_2}$ is the number of X_1 residues that have been mutated to X_2. Thus, $S_{X_1 \to X_2}$ in this case quantifies the variation in the critical solution temperature of the protein solution for a mutation of one X_1 residue to X_2. If the data is well represented by this first scaling law, it would imply that the effects of amino acid mutations on PLD phase behavior depend exclusively on the total number of mutations, but are not affected by other PLD parameters, such as the protein length. Thus, the first scaling law (Eq. 1) is the simplest possible model we test here for how incremental mutations may collectively influence the phase behavior of PLDs.

Next, we define a second type of scaling law, which hypothesizes that the changes to the critical solution temperature upon mutation is proportional to the fraction of residues that are mutated, rather than their raw number. That is,

$$\Delta T_{\mathrm{c}} = S_{X_1 \to X_2} \frac{N_{X_1 \to X_2}}{L} = S_{X_1 \to X_2} f_{X_1 \to X_2}, \tag{2}$$

where, L represents the number of total amino acids in the PLD, or protein length, and $f_{X_1 \to X_2} = \frac{N_{X_1 \to X_2}}{L}$ is the fraction of X_1 residues that have been mutated to X_2. Thus, this second scaling law hypothesizes that the effect of mutations on the critical solution temperature is more pronounced in shorter proteins (where $N_{X_1 \to X_2}$ mutations constitute a larger fraction of the protein) and less so in longer proteins (where $N_{X_1 \to X_2}$ mutations represent a smaller fraction of the total number of amino acids). The expectation is that because longer PLDs have higher valencies than shorter ones, the density of inter-molecular connections they form inside condensates is more robust towards perturbations than that of shorter PLDs [73].

Finally, we define a third scaling law,

$$\Delta T_{\mathrm{c}} = S_{X_1 \to X_2} \frac{N_{X_1 \to X_2}}{\sqrt{L}} \tag{3}$$

where the change in critical solution temperature is hypothesized to be directly proportional to the number of mutations ($N_{X_1 \to X_2}$) but inversely linear to $\sqrt{L}$. Like our second scaling law, this third relationship also presumes that longer PLDs form condensates with critical parameters that are less affected by mutations than shorter PLDs. However, by considering $\sqrt{L}$, instead of L as in Eq. 2, the expectation is that the buffering effects of protein length are more modest than those stated by Eq. 2. Notably, the amino acid composition of PLDs can be understood by the stickers-and-spacers framework [74, 75]. The stickers are amino acids like Tyr or Arg, which confer stability to the liquid network via the formation of associative interactions. Spacers are the residues interspersed among the stickers. Eq. 3 suggests that as a protein lengthens, the number of redundant residues that can establish sufficiently strong inter-protein interactions, i.e. number of stickers, grows too; yet, it grows much more conservatively than L because a fraction of L is made of spacers.

With these scaling laws as ansatz, we set to uncover trends in the dataset we built from our MD simulations of 140 variants of different PLDs. Adopting these laws as foundational hypotheses, we explore the relationship between alterations in composition and changes in critical solution temperatures, offering insights into the molecular drivers of PLD phase separation.

2.2.1 Aromatic Mutations Exhibit a Linear Correlation Between Mutation Fraction and Critical Solution Temperatures

Aromatic amino acids, namely tyrosine (Tyr, Y), phenylalanine (Phe, F), and tryptophan (Trp, W), are widely regarded as essential contributors to phase separation of PLDs, due to their ability to form π–π and cation–π contacts [9, 32, 33, 35, 37, 38, 76–78]. Among aromatics, experiments and simulations agree that Tyr is a stronger contributor to the stability of condensates than Phe [8, 33, 35, 63], and that both the composition and the patterning of aromatic residues in PLDs have an impact on the condensate critical parameters and its material state [18, 70, 79, 80].

Aromatic residues are highly abundant in PLDs, with the fraction of aromatic residues in the PLD variants analyzed in this study ranging from 0.19 to 0.43 (Fig. 1B). Furthermore, because different aromatics are unequal contributors to PLD phase behavior, we analyzed the impact of mutations for Tyr and Phe and Trp separately.

To quantify the impact of mutating Tyr to Phe in shifting the critical solution temperatures of our set of PLDs—i.e., determining which of our proposed scaling laws best describes the data—we designed 17 protein mutants that varied the Tyr to Phe ratio in the family of PLDs and computed their phase diagrams (Fig. 4A). Our simulations reveal that Tyr is stronger than Phe as a driving force for phase separation for the entire PLD family considered, a finding that replicates the results of Bremer *et al.* [35] for hnRNPA1 and confirms that it extends to other PLD proteins. By analyzing the change of critical solution temperature with respect to the unmodified PLD (or wild type, WT) against the number of mutations, our results reveal a linear relationship between the number of Tyr to Phe mutations and the change in the critical temperature of the system; our first scaling law, $S_{Y\to F}$, Eq. 4. Here, ΔT_c represents the temperature difference between the mutated protein and its wild-type sequence, and $N_{Y\to F}$ represents the number of Tyr residues that were mutated to Phe in the sequence.

$$S_{Y\to F} = \frac{\Delta T_c}{N_{Y\to F}} = -0.40 \pm 0.04 \tag{4}$$

Note that, by definition $S_{Y\to F} = -S_{F\to Y}$, since $N_{Y\to F}$ is defined as positive for Tyr to Phe mutations and negative for Phe to Tyr mutations, or, in other words, $N_{Y\to F} = -N_{F\to Y}$. For our data set, we find $S_{Y\to F}$ to be -0.40 ± 0.04, where the error given is the standard error. The way to interpret this result is that a single mutation of a Tyr residue to a Phe residue in a given PLD protein would result in a -0.4 K variation in the critical solution temperature for phase separation. Such a modest variation due to a single-point mutation is smaller than the typical error associated with estimating the absolute value of the critical solution temperature itself using residue-resolution coarse-grained Direct Coexistence simulations (approximately 3 K). Unambiguously measuring the value of the critical solution temperature with sub-Kelvin accuracy poses a considerable challenge also in experiments [81, 82]. However, we can uncover statistically significant trends by analyzing the behavior of the critical solution temperature in a large set of PLDs with diverse numbers of mutations rather than a single case. In particular, our analysis reveals that the impact of Tyr to Phe mutations is modest and the effect of multiple mutations can be described adequately as cumulative.

In addition to these analyses, the role of Tyr/Phe to Trp mutations, albeit Trp being a rare residue in prion domains, was investigated in TIA1 due to the presence of Trp in the wild type of this PLD, Fig. 4B. The simplest scaling law that describes this mutation (Equation 5) is fundamentally equal to the law describing Tyr to Phe mutations. However, a notable distinction is that the $S_{F/Y\to W}$ scaling constant for our data is approximately 4.3 ± 0.1, which is one order of magnitude greater than $S_{Y\to F}$, consistent with the much stronger π–π contacts established by Trp. It is important to note that, in this case, similarly to Try to Phe mutations, multiple Ansatzs align closely with the observed data, which underscores the fact that, in the context of these mutations, the size of the side chains does not significantly alter phase behavior, as the predominance of aromatic perturbations exerts a dominating effect.

$$S_{F/Y\to W} = \frac{\Delta T_c}{N_{F/Y\to W}} = 4.3 \pm 0.1 \tag{5}$$

Across the whole family of PLDs analyzed, our data reveals that changes in aromatic amino acid identity impact the critical solution temperatures modestly (< 1 K per mutations involving F and Y) to strongly (4 K for mutations involving Trp) and independently of the protein length. Such a result is consistent with the mutations in aromatic identity preserving the total number of associative interactions that the PLDs can establish, and hence, the density of molecular connections in the condensed liquid network. That is, aromatic mutations are expected to influence the energetics and lifetimes of the multivalent interactions among PLDs, but not the number of intermolecular contacts they form.

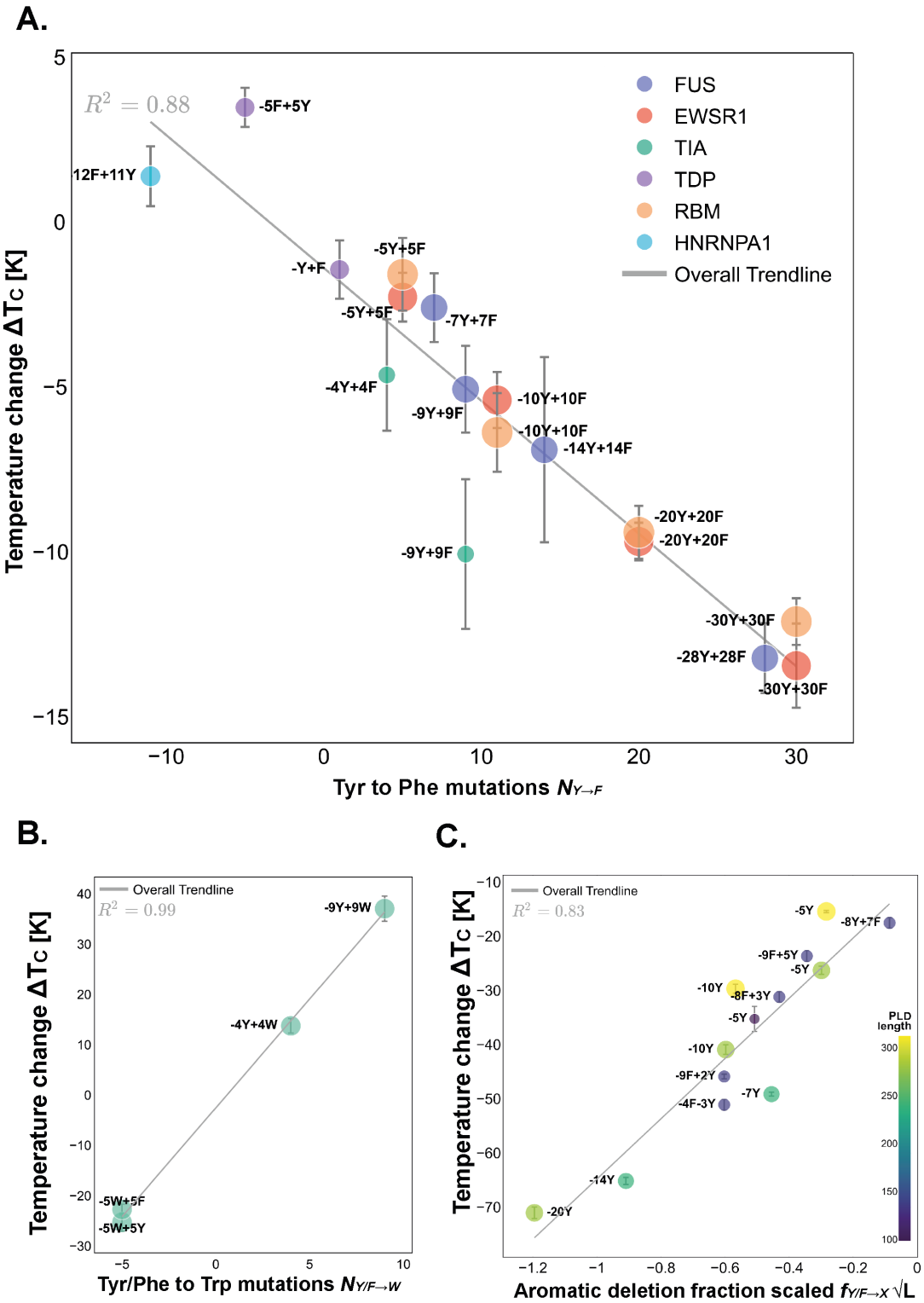

Fig. 4: Mutations in aromatic amino acids have strong effects on the critical solution temperature of PLDs. A. Tyr to Phe mutations. The number of Tyr mutated to Phe (x-axis) versus the change in critical temperature, (y-axis), computed as the critical temperature of the variant minus that of the corresponding PLD wild-type sequence. The trendline defines the simplest, dominating scaling law for this mutation. The R^2 value of 0.88 shows the agreement of our data to the scaling law defined by Equation 4. **B.** Tyr or Phe to Trp mutations. In this case, the overall trendline, defined by equation 5 has an R^2 of 0.99.**C.** Analysis of the variants involving mutations of aromatic residues to uncharged, non-aromatic amino acids, a.k.a. aromatic deletions. In the y-axis, the change in the critical temperature of the variant to that of the wild type sequence, and in the x-axis, a renormalized measure of mutations: the fraction of aromatic residues mutated times the $\sqrt{L}$, to account for the competing physical interactions between PLDs. Error bars indicate the standard error associated with replicas of the simulation.

2.2.2 Strong Critical Temperature Reduction Due To Aromatic Deletions is Buffered By Protein Length

We next examine the role of replacing aromatic residues (Tyr and Phe) with uncharged, non-aromatic residues—such as Gly, Ser, Gln, Ala, and Asn—which are prevalent in PLDs. The mutations were designed to preserve amino acid sequence compositional biases by replacing Tyr and Phe with uncharged, non-aromatic residues in the same proportions as they appear in the wild type of each sequence, i.e., maintaining the relative amounts of each type of residue found in the sequence. All amino acid sequences are available in the Supporting Information.

Our simulations show that aromatic deletions strongly reduce the critical solution temperature of the PLD solutions, Fig. 4C. The best scaling law from our ansätz to describe the data is Eq. 3, which involves the square root of the PLD length. That is, $S_{Y/F\to X}$, where X represents any uncharged, non-aromatic residue.

$$S_{Y/F\to X} = \frac{\Delta T_{\mathrm{c}}\sqrt{L}}{N_{Y/F\to X}} = -56 \pm 5 \tag{6}$$

Here, once again, $N_{Y/F\to X}$ is the number of mutations, and L is the length of the PLD. For our data set, the value of the scaling constant $S_{Y/F\to X}$ is -56 ± 5. To interpret this result, let us consider a model PLD of two hundred amino acids ($L = 200$). If we mutate one Tyr to an uncharged, non-aromatic amino acid, the change in the critical temperature—computed via the formula $S_{Y/F\to X} \times \frac{N_{Y\to X}}{\sqrt{L}}$—is approximately -4 K. Such a value is consistent with the dominance of π–π and cation–π contacts in PLD phase separation reported experimentally [9, 32, 33, 35, 37, 38, 76–78]. As stated above, aromatic residues behave as stickers within PLDs condensates. When these aromatic residues are replaced with uncharged non-aromatic amino acids the density of connections network of molecular connections in the liquid is strongly perturbed, as we have shown earlier [21]. To observe this in more detail, we computed select contact maps (see Fig. 11) – notably, mutations that removed aromatic amino acids strongly altered the primary sites of interactions.

Our results reveal that the length of the PLDs strongly influences how the critical parameters of PLD solutions change in response to aromatic deletions. The effect of mutations is stronger in shorter proteins because each aromatic deletion represents a higher percentage of the total number of stickers in the PLD, than in longer proteins. Longer proteins have higher valencies and redundant stickers that facilitate a network rearrangement.

Our results highlight the delicate balance between amino acid composition and protein length in determining the phase behavior of PLDs, providing a deeper understanding of the molecular basis of protein phase transitions.

2.2.3 Arginine Mutations Drastically Alter Condensate Stability

The distribution and balance of charged amino acids–i.e., aspartic acid (Asp, D), glutamic acid (Glu, E), histidine (His, H), arginine (Arg, R), and lysine (Lys, K)—within an IDR affects its solubility, conformational ensemble, and the range of stability of the ensuing condensates it might form. PLDs are not particularly rich in charged residues: in the wild type PLDs of our data set, the fraction of charged residue ranges from 0.02 to 0.1 (Fig. 1B), and the fraction of positively charged amino acids is narrowly distributed around 0.015, except for hnRNPA1, which has a value of 0.09 (Fig. 1B). However, despite their low prevalence in PLDs, charged residues, in particular Arg, have been shown to be important contributors to the stability of a wide-range of biomolecular condensates [32, 35, 83–85].

While both Arg and Lys are positively charged amino acids at physiological pH, it has been demonstrated that they have unequal roles in the stability of biomolecular condensates [32, 35, 63, 84, 86]. Both residues can form charge–charge and cation–π associative interactions, but the case of Arg is unique: the sp^2-hybridized planar guanidinium group of Arg confers it a special capacity to form strong hybrid π–π/cation–π contacts with the π-electrons of aromatic rings [33]. Thus, to quantify the impact of Arg in the range of stability of PLD condensates, we designed mutants that replaced Arg with Lys and vice-versa (Fig. 5).

The best fitting scaling law associated with mutating Arg to Lys is inversely dependent on the fraction of mutations, rather than the number of residues mutated (Eq. 7). The resulting scaling constant for this particular mutation, i.e., $S_{R\to K}$, has a value of -1300 ± 100, indicating that for a typical PLD of $L = 200$, the mutation of a single Arg into Lys results in a reduction of the critical temperature of the system by 6.5 K. Hence, replacing Arg with Lys is a destabilizing mutation of PLD condensates in agreement with the wide-body of experimental evidence [32, 35, 63, 84, 86].

$$S_{R\to K} = \frac{\Delta T_{\mathrm{c}}}{f_{R\to K}} = -1300 \pm 100 \tag{7}$$

The length dependency of this particular scaling law suggests that the impact of Arg to Lys mutations on the phase behavior of PLDs primarily depends on the relative concentration of these mutations (mutation

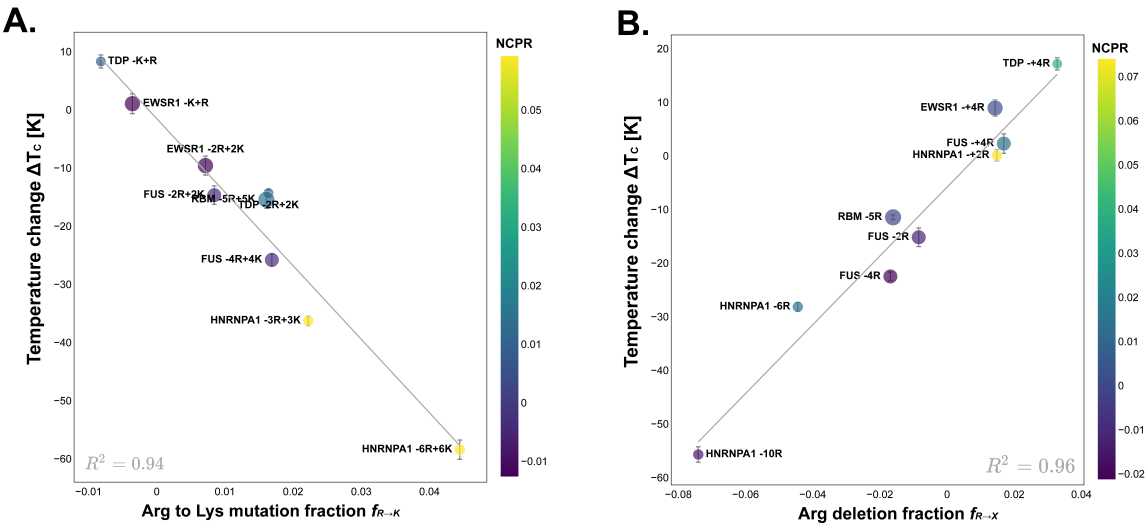

Fig. 5: **Critical effects of Arg mutations in the phase transition temperature of the condensates** **(A)** Analysis of the critical solution temperature of PLDs with Arg to Lys mutations. The number of Arg to Lys mutations divided by PLD length (x-axis) versus the change in the critical solution temperature, that is, $T_c(\text{variant}) - T_c(\text{WT})$ (y-axis). The trendline, representing Equation 7, fits the data with an $R^2 = 0.94$, and defines the stability measure of this perturbation. **(B)** Analysis of the critical solution temperature of PLDs with Arg mutations to uncharged, non-aromatic amino acids that maintain the WT compositional percentages (mainly Gly, Ser, and Ala. For more details see Supporting Information). Arg deletion fraction (x-axis) versus $T_c(\text{variant}) - T_c(\text{WT})$ (y-axis). The trendline, which fits the data with $R^2 = 0.96$, defines the stability measure, represented in Equation 8.

fraction). Additionally, it proposes that longer sequences may exhibit increased resilience against Arg to Lys mutations. This resilience is likely attributed to the greater separation of charge within the longer sequences.

To further analyze the effect of the charge of the PLDs in this set of mutants, in Fig. 5 we include a scalebar showing the net charge per residue (NCPR). Mutants that present higher NCPR present the biggest decrease in their phase-separating propensities, which is consistent with the expectation that condensates with a neutral overall charge are more energetically stable [35, 87, 88].

When we replace Arg with uncharged, non-aromatic residues, the resulting scaling law also depends inversely on the fraction of mutations, with a value of -640 ± 50, amounting to a reduction of 3.2 K in T_c per Arg deletion for a model PLD of length $L = 200$.

$$S_{R\to X} = \frac{\Delta T_c}{f_{R\to X}} = -640 \pm 50 \tag{8}$$

Therefore, our results reveal that it is almost twice more destabilizing to mutate Arg to Lys than to replace Arg with any uncharged, non-aromatic amino acid, namely, Gly, Ser or Ala. Mutating Arg to an uncharged non-aromatic amino acid is less destabilizing than mutating it to Lys likely because the charge-reduced variants decrease the cation–cation repulsion of the PLDs, with respect to the Arg-rich or Lys-rich systems. Furthermore, mutating Arg to Lys decreases the enthalpic gain for condensate formation not only via the replacement of stronger Arg-based associative interactions for weaker Lys-based ones but also by increasing the overall solubility of the Lys-rich PLD [38, 89], given that the hydration free energy of Arg is less favorable than that of Lys [33, 38, 83].

2.2.4 Polar Amino Acids Subtly Modulate PLD Phase behavior

Uncharged, non-aromatic amino acids, including Gly, Ala, proline (P, Pro), Ser, and Thr, are commonly found in PLDs. Because they generally form weaker residue–residue interactions and present smaller excluded volumes than aromatics, they are classified as spacers that contribute to maintaining the solubility of proteins and material states of condensates [70, 74]. These amino acids can also have other functions in proteins, and the role they play as spacers depends on their chemical make-up, but also on the specific context in which they are found [35, 39, 90].

Here, we investigate the role of glutamine (Gln, Q) and asparagine (Asp, N). Both Gln and Asn are polar amino acids, hence their charge distribution is more asymmetric, presenting a significant dipole moment. Consequently, they can establish transient ion–dipole, dipole–dipole and π–dipole interactions with other residues. While the Mpipi model does not account explicitly for polarity, it is worth noting that it was

parametrized based on bioinformatics data, which implicitly incorporates this aspect. Thus, our study allows for the observation of emergent behaviors that stem from this inherent characteristic.

We designed mutations that vary the relative amounts of Gln and Asn, and examined trends in their phase-separating propensity. Increasing the number of Gln per unit length by mutating Asn to Gln increases the ability of the system to phase separate, likely due to Gln having a higher dipole moment than Asn, which is approximated in our model via marginally stronger non-ionic interactions formed by Gln over Asn (Fig. 6 A).

$$S_{N\to Q} = \frac{\Delta T_{\mathrm{c}}}{f_{N\to Q}} = 50 \pm 5 \tag{9}$$

The value of the scaling constant $S_{N\to Q}$ is 50 ± 5, which is significantly lower than that of mutations of residues with a net charge or aromatic deletions (rounding 1300, for example, for arginine to lysine mutations). This result indicates that mutations of Asn to Gln or vice versa have a smaller impact in the critical parameters than aromatic deletions or mutations of charged residues.

Finally, we also explored the role of Ser, Gly, Thr, and Ala on the critical parameters of the condensate. While Ser and Thr are also polar amino acids, their dipole moment is significantly smaller than that of Gln and Asp. Consistently, the scaling constants involving mutations Ser and Thr mutations—i.e., $S_{S\to T}$— are considerably smaller (Fig. 6 B and C) overall. A comparable pattern is noted for $S_{G\to T}$, with G expected to exhibit significantly lower polarity compared to Gln and Asp.

$$S_{S\to T} = \frac{\Delta T_{\mathrm{c}}}{N_{S\to T}} = -0.09 \pm 0.03 \tag{10}$$

$$S_{G\to T} = \frac{\Delta T_{\mathrm{c}}}{N_{G\to T}} = -0.08 \pm 0.02 \tag{11}$$

Regarding Ala, we also tested its relative role with respect to Ser.

$$S_{A\to S} = \frac{\Delta T_{\mathrm{c}}}{N_{A\to S}} = 0.16 \pm 0.05 \tag{12}$$

Amino acids with larger side chains, such as Thr, have a larger excluded volume than those with smaller side chains, such as Gly or Ser. The corresponding value of the scaling constants $S_{S\to T}$ and $S_{G\to T}$ suggests that in the case of Thr mutations, mutations of neutral, non-aromatic amino acids to those with larger excluded volumes inhibit phase separation of the PLD by lowering the corresponding critical temperature, while mutations to those amino acids with smaller excluded volumes have an opposite effect. This is likely due to increased flexibility of the protein backbone, allowing the chain to adopt multiple conformations that promote multivalent contacts. Importantly, these trends cannot be anticipated by inspecting the parameters of the Mpipi model; instead, they emerge from examining the behavior of the collective system.

On the other hand, for mutations involving Ser, Gly, and Ala–all small amino acids with minor side chain size variations–the impact on phase separation seems to stem predominantly from changes in solvation. Solvation plays a key role in phase separation since it can help balance the driving forces of this process [91, 92]. While hydrophobic regions of the protein, with low solvation free energies, tend to form aggregates to reduce interactions with the solvent environment; hydrophilic regions promote interactions with water. Whether these regions inhibit or promote phase separation is likely due to a balance of opposing effects. On the one hand, hydrophilic regions can form energetically favorable contacts with solvent, promoting dispersion and suppressing phase separation. However, when these regions are recruited to condensates the release of water molecules can help to off-set the entropic cost of sequestering hydrophobic regions within a protein droplet, thereby facilitating phase separation.

Specifically, mutations from Gly to Ser (not shown) increase the size of the side chain while also enhancing solvation, while mutations from Ala to Ser do not significantly change the size of the side chain but significantly increase the protein solvation. In the first case, the critical parameters of phase separation remain virtually unchanged, only presenting a very minor decrease. In the case of Ala to Ser mutations, phase separation is slightly enhanced. Both scaling laws $S_{G\to S}$ and $S_{A\to S}$ reflect such an effect, and underlie the context-dependent nature of these mutations.

Overall, mutations of unchanged, non-aromatic amino acids seem to have an impact on phase separation due to two competing effects: a change in side chain size and a change in solvation. However, it is important to note that the magnitude of the scaling laws that govern these mutations is minimal, and thus, likely extremely sensitive to small fluctuations such as changes in the system and solution conditions. Notably, these findings are consistent with experiments that report small changes in phase separation propensity for mutations involving neutral, non-aromatic residues [8, 35].

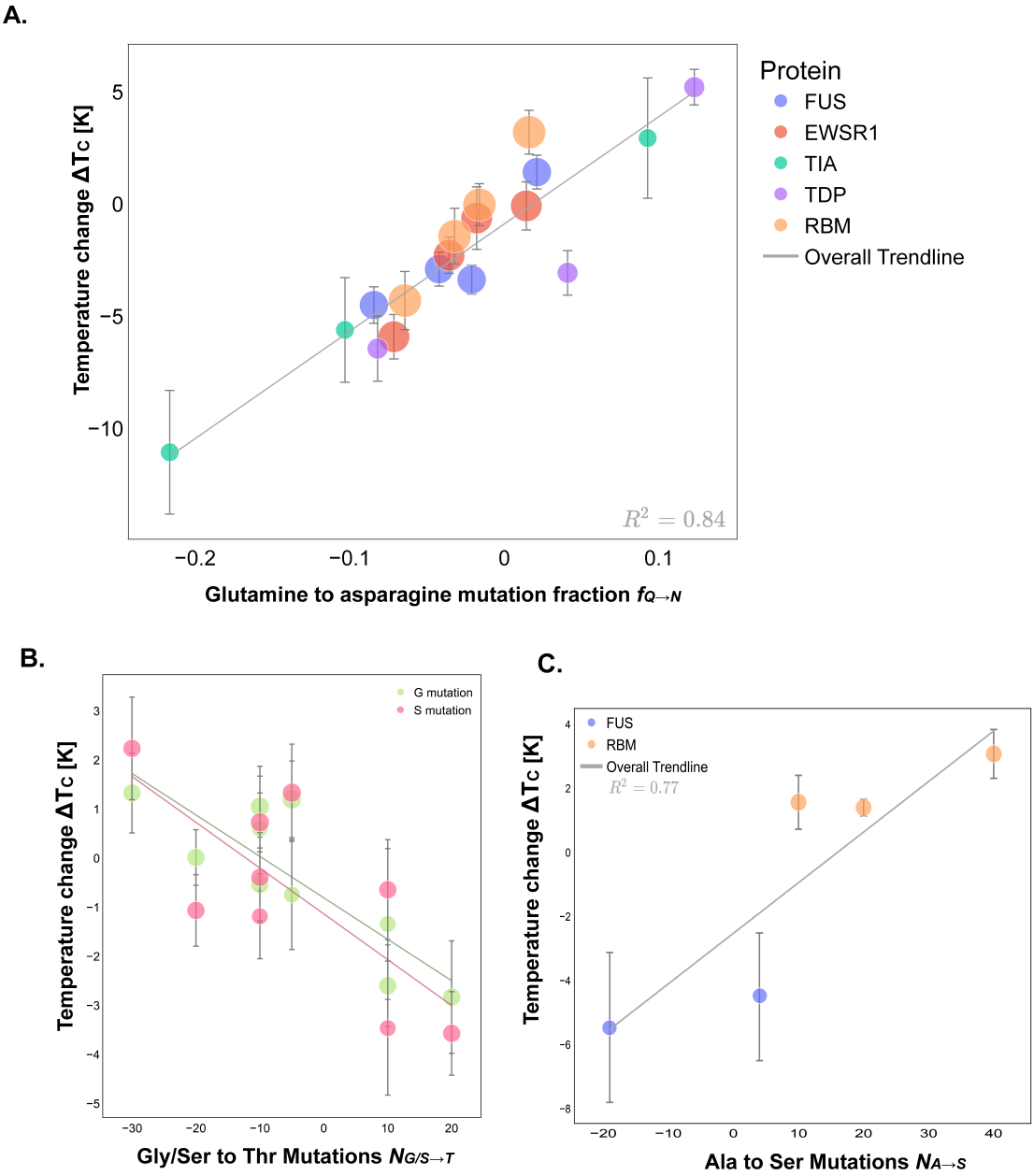

Fig. 6: Polar and neutral amino acid mutations show subtle modulation of the critical solution temperatures of PLDsA) Analysis of the critical solution temperature of PLDs with Asn to Gln mutations. The number of Gln to Asn mutations renormalized by dividing by PLD length (x-axis) versus the temperature change $T_c(\text{variant}) - T_c(\text{WT})$ (y-axis). The trendline, which fits the data with an $R^2 = 0.84$, fits the scaling law Equation 9. **(B)** Analysis of the critical solution temperature of PLDs with Gly/Ser to Thr mutations (x-axis) versus the temperature change $T_c(\text{variant}) - T_c(\text{WT})$ (y-axis). The trendlines, representing Equations 10 and 11 fit the data with an $R^2 = 0.7$. **(C)** The number of Ala to Ser mutations (x-axis) versus the temperature change $T_c(\text{variant}) - T_c(\text{WT})$. The trendline, Equation 12, fits the data with an $R^2 = 0.77$

2.2.5 Beyond composition: patterning considerations

Beyond composition, the relative positioning of amino acids in the sequence, i.e. sequence pattern, also plays a role in protein folding, function and, in this case, phase behavior [9, 93, 94]. Due to the sequence heterogeneity of PLDs, and, more generally, IDPs, it is particularly complex to define adequate patterning measures [88, 95–97].

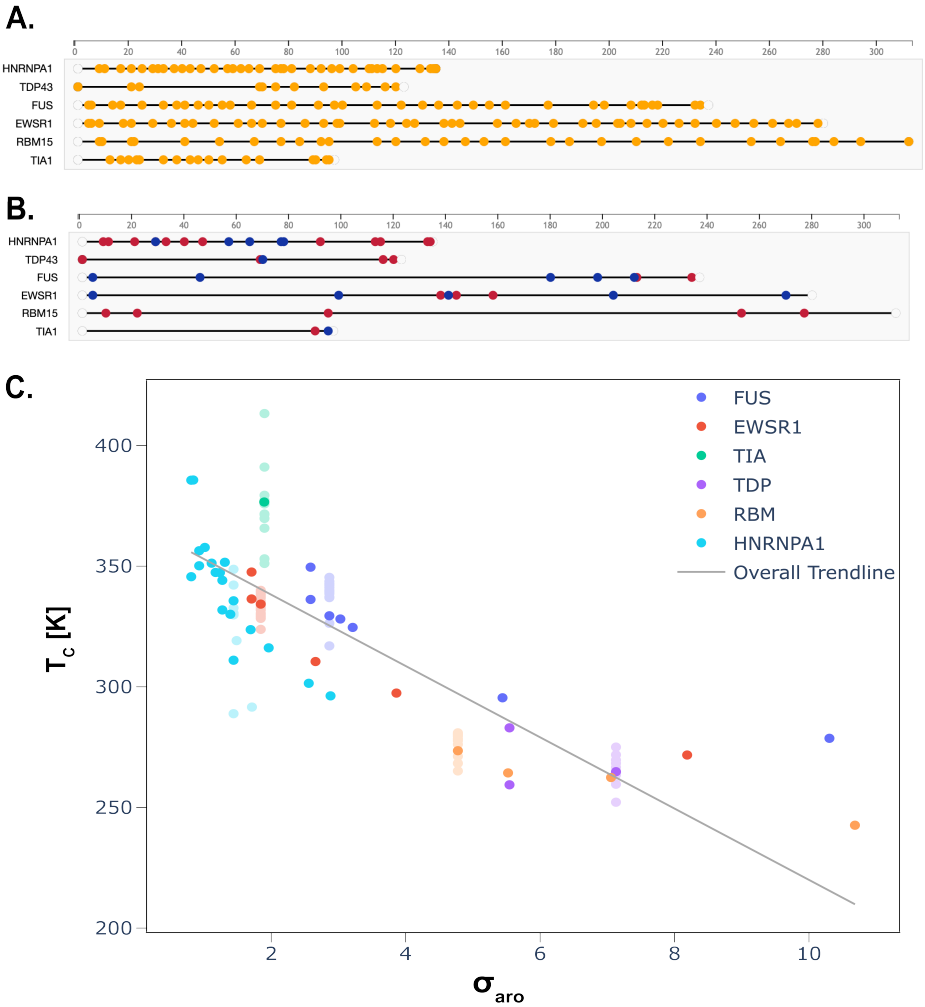

Fig. 7: **Prion-like domains featuring sequences with dispersed aromatic residues show greater propensities for phase separation. A** Representation of the aromatic amino acids present in the wild type sequences of each PLD considered **B** Location of the positively charged (red) and negatively charged (blue) amino acids present in the wild type sequences of each PLD considered **C** Analysis of the critical solution temperature versus the σ order parameter, defined in Equation 13. Those groups of variants with the same 13 value show their average critical temperature highlighted, and softer values for other data points. For lower values of σ, the critical temperature present in the condensates is higher, and vice versa, indicating that those variants with a more homogeneous distribution of aromatic amino acids have condensates more stable at higher temperatures.

In this work, we have taken advantage of our large data set to analyze the impact of amino acid sequence patterning by adopting a patterning order parameter [9, 87, 98], σ_{aro}, that has been designed to describe patterning of aromatic residues in such IDPs.

To compute σ_{aro}, we divide the amino acid sequence into peptide segments of equal length, compute a measure of sequence aromatic asymmetry for each segment, and finally sum the values of such measure over all the segments. Consistent with literature [9, 87, 98], we take l to be 5 and 6. σ_{aro} is then given by:

$$\sigma_{\text{aro}} = \left(\sum_{\text{segments}} \frac{\sigma_i}{\sigma_{\text{max},i}} \right) \frac{N_{\text{stickers}}}{L}, \tag{13}$$

where L is the total number of amino acids in the protein domain, N_{stickers} the number of sticker residues—i.e. charged or aromatic residues—and σ_i has the functional form

$$\sigma_i = \left(\left(2 \frac{N_{\text{aro},i}}{l_i} - 1 \right)^2 - \sigma_{\text{T}} \right)^2. \tag{14}$$

Here, $N_{\text{aro},i}$ is the number of aromatic residues in the i-th segment, l_i the segment length, and σ_{T} is a function of the total number of aromatic residues in the sequence N_{aro}:

$$\sigma_{\mathrm{T}} = \left(2\frac{N_{\mathrm{aro}}}{L} - 1\right)^2 \tag{15}$$

In turn, $\sigma_{\mathrm{max},i}$ is the measure of aromatic amino acid distribution for an artificial sequence with all the aromatic amino acids concentrated in part of the chain.

Notably, our results, presented in Fig. 7, indicate that the sequences with the most dispersed aromatic residues (i.e., low values of σ_{aro}) phase separate more readily. Dispersed residues result in greater effective valency of a PLD, leading to a higher probability of forming a percolated network structure to sustain the condensate. Additionally, this result demonstrates that the patterning order parameter, σ_{aro}, can be used to predict the phase-separating propensity of a PLDs as the function of aromatic residue distribution. This result is consistent with previous studies that demonstrate the importance of aromatic residue patterning on the phase behaviour of PLDs [9, 99].

3 Discussion

Computer simulations have advanced our ability to map emergent properties of protein solutions to the chemical make-up of their individual biomolecules. For instance, modern residue-resolution protein coarse-grained models [34, 76, 100–104] can now be used to quantify the modulation of protein phase behavior as a result of amino acid sequence mutations of their native proteins with near-quantitative agreement with experiments [34, 102, 103]. In this study, we exploit the accuracy and efficiency of our protein model, Mpipi [34], to compute and analyze the temperature-vs-density phase diagrams of an unprecedentedly large set of PLD proteins—140 variants of the proteins hnRNPA1, TDP43, FUS, EWSR1, RBM14, and TIA1. Collectively, our set of phase diagrams reveal conserved scaling laws that quantitatively predict the change in the critical solution temperatures of single-component PLD solutions as a function of the number and type of amino acid mutations of the PLDs.

To facilitate the interpretation of the scaling laws derived in our study, we have organized the data in a table that illustrates the effect of single point mutations on the critical solution temperatures of the family of PLDs, as a function of the PLD length.

L	Y → F	F/Y → W	F/Y → X	R → K	R → X
100	-0.4 ± 0.04	4.3 ± 0.1	-5.6 ± 0.5	-13 ± 1	-6.4 ± 0.5
200	-0.4 ± 0.04	4.3 ± 0.1	-3.9 ± 0.4	-6.5 ± 0.5	-3.2 ± 0.3
300	-0.4 ± 0.04	4.3 ± 0.1	-3.2 ± 0.3	-4.3 ± 0.3	-2.1 ± 0.2

L	N → Q	S → T	G → T	A → S
100	0.5 ± 0.05	-0.09 ± 0.03	-0.08 ± 0.02	0.16 ± 0.05
200	0.25 ± 0.03	-0.09 ± 0.03	-0.08 ± 0.02	0.16 ± 0.05
300	0.17 ± 0.02	-0.09 ± 0.03	-0.08 ± 0.02	0.16 ± 0.05

Table 1: Predicted effect of single point mutations on the critical temperature, in Kelvin, of PLDs, as a function of their length, L

Consistent with experimental evidence and computational predictions, interactions involving aromatic residues—namely, π–π and cation–π—are dominant in stabilizing condensates formed by PLDs [8, 34, 35]. Notably, our simulations reveal the largest decrease in the critical solution temperature of a PLD condensate occurs when an aromatic residue is replaced by an uncharged, non-aromatic amino acid. Our scaling law for such aromatic deletions ($56\,\mathrm{K}/\sqrt{L}$) predicts that the destabilizing effect of replacing an aromatic with a non-aromatic residue is linearly amplified by the number of such mutations and screened by the square root of the length of the protein ($\sqrt{L}$), indicating a disruption in the percolated network of interactions sustaining the condensate. Condensates made of longer PLDs are more robust against aromatic deletions because, by construction, each aromatic residue represents a smaller percentage of the total valency of the PLD in question. In other words, longer PLDs are higher valency molecules than their shorter counterparts and can, thus, form a densely connected percolating condensed-liquid network, even when a few of their aromatic stickers have been deleted.

Our findings also highlight the crucial role of Arg in the critical parameters of PLD condensates. Specifically, our scaling law for Arg deletions ($640\,\mathrm{K}/L$) predicts that the destabilizing effect of an Arg mutation to an uncharged, non-aromatic residue is inversely proportional to the length of the protein (L), as well as linear with respect to the number of mutations. More specifically, it indicates that, for a PLD of the length of 100 amino acids, the change in the critical temperature will be of the order of 6.4 K. While formulaically similar to the scaling law for aromatic deletions, a notable difference is a change in the dependence on the PLD length. In the case of Arg deletions, the effect of mutations is screened by L rather than $\sqrt{L}$, indicating

a stronger screening dependence on the PLD length. This indicates that for long PLDs, the destabilizing effects due to Arg deletions are diminished when compared to aromatic deletions, while in shorter PLDs this trend is reversed with each Arg deletion exerting a more significant destabilizing effect than the aromatic deletions.

These results underscore once again how for PLDs cation–π and π–π interactions are the main drivers of phase separation [9, 32, 33, 35, 37, 38, 76–78]. While it may seem counterintuitive for Arg deletions to have a lesser effect in the critical parameters of longer PLDs than their aromatic counterparts (as Arg residues are known to form strong cation–π interactions [105, 106]) the explanation lies in the type of systems in question, which are particularly rich in aromatic residues. Longer PLDs have, as a result of their higher valency, a larger network of possible pairwise interactions, which can mitigate the destabilizing effect of Arg deletions. In the case of shorter PLDs, each Arg residue represents a larger proportion of the total valency, and thus its deletion has a proportionally larger impact on the critical parameters. In contrast, aromatic residues contribute to the stability of the condensate through both cation–π and π–π interactions, which are predicted to be weaker than their cation–π analogues [33, 34], but more numerous in PLDs. In a longer PLD, the deletion of an aromatic residue results in the loss of more potential interactions, creating a larger destabilizing effect than the deletion of an Arg residue. However, in shorter PLDs, due to their lower total valency, the loss of potential associative interactions that result from the deletion of an aromatic residue is lower. Hence, the destabilizing effect resulting from this mutation is comparatively less than for the Arg deletion.

Our scaling laws have also allowed us to quantify the destabilizing effect of Arg to Lys mutations in the phase separation of PLDs. Concretely, mutations of Arg residues to Lys result in an overall decrease in the critical temperature of phase separation, two-fold in comparison to the impact of mutating Arg into an alternate uncharged non-aromatic residue. Consequently, the scaling law considerably intensifies for mutations from Arg to Lys, nearly doubling the effects of each mutation in the critical temperature of the condensate (See Table 1). This implies that each mutation of Arg to Lys exerts a substantially more destabilizing effect than mutations of Arg to other uncharged, non-aromatic residues, consistent with experimental evidence [32, 35, 84, 107] and highlighting the significant, modulatory role Lys plays in the phase behavior of PLDs.

Uncharged, non-aromatic residues such as Ser, Gly, Asn, Gln, and Thr are found to play a marginal role in PLD condensate stability [35, 39, 70, 74]. Notably, PLD condensates are highly stable in the face of Gly, Ser, Thr and Ala mutations. This behavior is quantified by their respective scaling laws, with values for the scaling constants for mutations from Gly to Thr, Ser to Thr and Ala to Ser at $-0.09\,\mathrm{K}$, $-0.08\,\mathrm{K}$ and $0.16\,\mathrm{K}$ respectively. These numerical values indicate that, for a change of one degree in the critical temperature, approximately 10 amino acids of this set would need to undergo mutation. Notably, these scaling laws do not display a significant length dependency. This observation aligns with the idea that these mutations primarily affect the condensate properties through modulation in excluded volume and solvation, rather than by interrupting specific interaction networks.

In contrast, the condensates display heightened sensitivity towards Asn and Gln mutations, underscored by the Asn to Gln scaling of $50\,\mathrm{K}/L$. This law is also inversely length dependent, indicating that mutations on these amino acids affect interaction networks and that mutations Asn to Gln result in a heightened ability of the system to phase separate. We hypothesize that, in addition to differences in solvation volume, the polarity of each amino acid plays a role by forming charge–dipole and dipole–dipole interactions. Both Gln and Asn are polar amino acids, with Gln having a longer side chain than Asn, which could potentially interact with a larger number of nearby atoms. Furthermore, the additional methylene group (-CH2-) may slightly alter the electron distribution around the amide group, which can lead to the formation of more stable dipole–dipole contacts. It is important to note that our computational model, Mpipi, is a charge-fixed model, and hence does not directly capture the effects of polarization. However, the lack of physical details in the Mpipi model is compensated by the parameters having been derived from bioinformatics data.

Collectively, the definition of a set of scaling laws acts as a unifying framework for the study of the critical parameters of PLD phase separation, by allowing for direct comparison of individual sequence effects. Such laws are cumulative, meaning that the effects of successive mutations can be predicted by the addition of the appropriate laws, as long as the resulting mutation keeps within PLD composition and patterning biases.

Our work explores a new framework to describe, quantitatively, the phase behavior of a large set of protein variants via the identification of scaling laws. Our scaling laws can predict changes in the stability of condensates based on amino acid sequence mutations. This code not only allows for testing the effects of composition bias on PLD phase separation but also serves as a predictive tool, allowing for a better selection of mutations to achieve desired changes in different condensates. This study is focused on PLDs, but we hypothesize similar scaling behavior in other IDPs, with the laws having a dependence on the protein length that reflects the nature of the underlying interaction networks of the system. Our findings have important implications for understanding PLD condensation and for designing modifications to counter their aberrant

aggregation, by setting the basis for characterizing and quantifying the molecular determinants of PLD phase behavior.

4 Methods and Materials

4.1 Design of the dataset

To design sequence variants to test the phase behavior of prion-like domains, we use in-silico targeted mutagenesis to introduce specific changes to the amino acid sequence of the PLD of interest. These changes involve substituting different amino acids, while carefully maintaining patterning and compositional biases, to test the role of given amino acids in the phase-separating propensities of the respective PLDs.

The nomenclature employed is consistent with that of Bremer and colleagues [35]. Namely, if n amino acids of type X are removed from the sequence and replaced with n amino acids of type Y, the variant is labeled $-nX + nY$. If instead, n amino acids of type X are removed from the sequence and replaced with amino acids (often different types) in proportion to the compositional bias of the PLD, the variant is simply labeled $-nX$.

To begin, we select a family of PLDs composed of hNRNPA1, RBM14, FUS, TDP43, TIA1 and EWSR1 and perform mutations on each sequence to generate a set of variants. Accordingly, we assess whether the effects of sequence mutations in PLD phase behaviors hold for the whole family of PLDs. Furthermore, our data set allows us to develop a set of rules that can aid in understanding and predicting PLD phase behavior. The amino acid sequences of the PLDs, and their corresponding variants, investigated herein are provided in the Supporting Information.

In total, we compute phase diagrams of 140 PLD variants via molecular dynamics (MD). This task required a staggering amount of CPU hours, over a quarter million in total. Given these demands, efficient utilization of high-performance computing (HPC) resources and proper parallelization of the computational tasks are paramount to ensuring optimal performance. To this end, we compute MD trajectories using the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) software [108], exploiting its efficient parallelization algorithms. Another significant computational challenge involves the large size of data generated. Each single run produced a significant amount of trajectory, backup, and analysis data, and a large number of files. On average, each run produced over 50 files and a minimum of 5 GB of data. To efficiently handle this, we implement a robust, tiered, data storage and retrieval system, ensuring that the vast amounts of data is organized through a combination of indexing and partitioning, enhancing search efficiency and accessibility. We leveraged the Lustre filesystem for its high performance and scalability, making it the ideal choice for managing and accessing large volumes of active data required by our computational jobs. Custom scripts are employed for automated data processing, enabling us to extract meaningful insights from the complex datasets. All scripts are available in our GitHub repository.

4.2 Critical Parameter Determination Through MD

We used MD simulations to determine the full binodals (i.e., the boundaries that delimit the two-phase region from the one phase region) for the PLDs. Specifically, we perform MD simulations in the canonical ensemble – i.e., fixed volume (V), temperature (T), and number of particles (N). For each binodal, a series of independent NVT simulations are performed at different temperatures and the density of each of the coexisting phases are determined. Thus, this approach yields binodals in the temperature–density phase space.

4.2.1 Direct Coexistence Simulations

Specifically, we use Direct Coexistence simulations, which enables the simultaneous simulation of both high- and low-density phases of a mixture, separated by an interface, within a single simulation box. Additionally, these simulations are performed with periodic boundary conditions in each dimension.

To initiate the NVT Direct Coexistence simulations, we start with one copy of a PLD in a cubic simulation cell and replicate it to the target number of chains (64 in total). We then iteratively scaled the simulation box to compress the chains into a high density direct coexistence simulation (approx. 0.8–1 g/cm^3) [101]. Here, we use a Langevin thermostat [109] (at the target temperature, with a damping time of 40 ps) for temperature control. Keeping the length of the box in the X and Y dimensions fixed, we then elongate the Z dimension of the box. We then simulate each system at the target temperature for over 200 ns (see Convergence test in the SI). From each simulation we compute the density profile in the Z direction and determine the density of the dilute and condense phase. These data are then used to produce the binodals.

4.2.2 Mpipi model

All the simulations are run using the Mpipi coarse-grained force-field [34] with an implicit solvent. In the Mpipi model electrostatic, cation–π and π–π interactions are carefully balanced to capture the phase behavior of PLDs. Notably, Mpipi achieves quantitative agreement with experiments in terms of prediction of critical solution temperatures of such systems. In Mpipi, each amino acid is represented by a single unique bead – i.e., of corresponding mass, molecular diameter (σ), charge (q), and an energy scale that reflects the relative frequency of planar π–π contacts (ϵ) [34].

The potential energy in the Mpipi model is computed as follows:

$$E_{\text{mpipi}} = E_{\text{bond}} + E_{\text{elec}} + E_{\text{pair}}. \tag{16}$$

Beads are connected via harmonic springs

$$E_{\text{bond}} = \sum_i \frac{1}{2} k (r_i - r_i^0)^2. \tag{17}$$

Here k is the bond force constant (19.1 kcal mol^{-1} Å^{-2}) and r_i^0 is the equilibrium bond length, which is set to 3.81 Å.

The electrostatic contribution is modeled via a Coulomb interaction term, with Debye–Hückel screening.

$$E_{\text{elec}} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \exp\left(-\kappa r_{ij}\right), \tag{18}$$

where q_i and q_j are the charges of the bead i and j, respectively. ϵ_0 is the electric constant and $\epsilon_r = 80$ is the relative dielectric constant of water. $\kappa = 0.126\,\text{Å}^{-1}$ is the inverse Debye screening length, which corresponds to monovalent salt concentration of 150 mM.

Finally, short-ranged non-bonded interactions are modeled with the Wang–Frenkel potential

$$E_{\text{pair}} = \sum_{i,j} \phi_{i,j} = \sum_{i,j} \epsilon_{i,j}\alpha_{i,j} \left[\left(\frac{\sigma_{i,j}}{r}\right)^{2\mu_{i,j}} - 1\right] \left[\left(\frac{\sigma_{i,j}}{r}\right)^{2\mu_{i,j}} - 1\right]^{2\nu_{i,j}} \tag{19}$$

where E_{pair} is the potential energy between all pairs of amino acids, $\phi_{i,j}$ is the potential energy between the pairs of amino acids i and j, ϵ_{ij} is an energy parameter that dictates the strength of interaction of that particular pair, $\sigma_{i,j}$ is average the Van der Waals radii of the pair that dictate the position of the energy minima, r is the distance between center of masses and $\nu_{i,j}$ and $\mu_{i,j}$ are fitted exponential parameters that determine the shape of the energy curve. Full details of the Mpipi model are described in 34.

4.2.3 Data fitting

To estimate the critical temperatures of the binodals computed in Section 4.2.1, we use the law of coexistence densities

$$\left(\rho_{\text{high}}(T) - \rho_{\text{low}}(T)\right)^{3.06} = d\left(1 - \frac{T}{T_c}\right) \tag{20}$$

and critical densities are computed by assuming that the law of rectilinear diameters holds, namely

$$\rho_{\text{high}}(T) + \rho_{\text{low}}(T) = 2\rho_c + 2A(T - T_c) \tag{21}$$

where $\rho_{\text{high}}(T)$, $\rho_{\text{low}}(T)$ and ρ_c are the densities of the high-density and low-density phases and the critical density, respectively, T_c is the critical temperature, and d and A are fitting parameters.

4.3 Finite-size scaling

In direct coexistence simulations in the slab geometry finite-size effects can arise from either the cross-sectional area (i.e., size of the interface) or from the long axis (i.e., fraction bulk condensate). To ensure our calculations are robust with respect to system size, we perform finite-size scaling at fixed system number density: (1) varying the size of the cross section and (2) varying the length of the long axis. Using these tests, we confirm that the phase diagrams computed for systems of varying sizes (27, 64, 125 and 216 chains) exhibit no significance difference in predicted densities (see Figure 9 in Supporting Information). Our scaling analysis revealed that 64 chains in a box of dimensions 125x125x1000 is the minimum size that ensures proteins do not interact with their periodic images and that the condensed phase is large enough to represent a true bulk phase.

5 Acknowledgments

We thank Pin Yu Chew for useful comments on the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 803326 to R.C.G.). MJM would like to acknowledge the Winton Programme for Physics of Sustainability for doctoral funding. AAG is funded by the ERC (grant agreement No 803326). JRE acknowledges funding from the Ramon y Cajal fellowship (RYC2021-030937-I) and the Spanish National Agency for Research under the grant PID2022-136919NA-C33. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/X02332X/1 to JHM] under the UK Research and Innovation (UKRI) Postdoctoral Fellowships Guarantee Scheme [project TF-CHROM-LLPS]. JAJ acknowledges research support from departmental start-up funds provided by the Department of Chemical and Biological Engineering and the Omenn–Darling Bioengineering Institute at Princeton University. This project made use of time on HPC granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1).

References

[1] Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **18**, 285–298 (2017).

[2] Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).

[3] Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular cell* **82**, 2201–2214 (2022).

[4] Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).

[5] Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *Journal of Biological Chemistry* **294**, 7115–7127 (2019).

[6] Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **57**, 936–947 (2015).

[7] Smith, J. *et al.* Spatial patterning of p granules by rna-induced phase separation of the intrinsically-disordered protein meg-3. *Elife* **5**, e21337 (2016).

[8] Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688–699 (2018).

[9] Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).

[10] Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerging topics in life sciences* **4**, 307—329 (2020).

[11] Posey, A. E., Holehouse, A. S. & Pappu, R. V. Phase separation of intrinsically disordered proteins **611**, 1–30 (2018).

[12] Elbaum-Garfinkle, S. *et al.* The disordered p granule protein laf-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences* **112**, 7189–7194 (2015).

[13] Franzmann, T. M. *et al.* Phase separation of a yeast prion protein promotes cellular fitness. *Science* **359**, eaao5654 (2018).

[14] Maharana, S. *et al.* Rna buffers the phase separation behavior of prion-like rna binding proteins. *Science* **360**, 918–921 (2018).

[15] Gotor, N. L. *et al.* Rna-binding and prion domains: the yin and yang of phase separation. *Nucleic acids research* **48**, 9491–9504 (2020).

[16] Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).

[17] Jawerth, L. *et al.* Protein condensates as aging maxwell fluids. *Science* **370**, 1317–1323 (2020).

[18] Alshareedah, I. *et al.* Sequence-encoded grammars determine material properties and physical aging of protein condensates. *bioRxiv* 2023–04 (2023).

[19] Linsenmeier, M. *et al.* Dynamic arrest and aging of biomolecular condensates are modulated by low-complexity domains, rna and biochemical activity. *Nature Communications* **13**, 3030 (2022).

[20] Sprunger, M. L. & Jackrel, M. E. Prion-like proteins in phase separation and their link to disease. *Biomolecules* **11**, 1014 (2021).

[21] Espinosa, J. R. *et al.* Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proceedings of the National Academy of Sciences* **117**, 13238–13247 (2020).

[22] Espinosa, J. R., Sanz, E., Valeriani, C. & Vega, C. On fluid-solid direct coexistence simulations: The pseudo-hard sphere model. *Journal of Chemical Physics* **139**, 144502 (2013).

[23] Banani, S. Compositional control of phase-separated cell. bodies. *Cell* **166**, 651–663 (2016).

[24] Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688–699 (2018).

[25] Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics* **49**, 107–133 (2020).

[26] Choi, J.-M., Hyman, A. A. & Pappu, R. V. Generalized models for bond percolation transitions of associative polymers. *Physical Review E* **102**, 042403 (2020).

[27] 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* **6**, e30294 (2017).

[28] Iserman, C. *et al.* Condensation of ded1p promotes a translational switch from housekeeping to stress protein production. *Cell* **181**, 818–831 (2020).

[29] Adame-Arana, O., Weber, C. A., Zaburdaev, V., Prost, J. & Jülicher, F. Liquid phase separation controlled by ph. *Biophysical Journal* **119**, 1590–1605 (2020).

[30] Munder, M. C. *et al.* A ph-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5**, e09347 (2016).

[31] Cinar, H. *et al.* Temperature, hydrostatic pressure, and osmolyte effects on liquid–liquid phase separation in protein condensates: Physical chemistry and biological implications. *Chemistry – A European Journal* **25**, 13049–13069 (2019).

[32] Fisher, R. S. & Elbaum-Garfinkle, S. Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature communications* **11**, 4628 (2020).

[33] Krainer, G. *et al.* Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nature Communications* **12**, 1–14 (2021).

[34] Joseph, J. A. *et al.* Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **1**, 732–743 (2021).

[35] Bremer, A. *et al.* Deciphering how naturally occurring sequence features impact the phase behaviors of disordered prion-like domains. *Nature chemistry* **14**, 196 (2022).

[36] Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **114**, E8194–E8203 (2017).

[37] Vernon, R. Pi–Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7** (2018).

[38] Fossat, M. J., Zeng, X. & Pappu, R. V. Uncovering differences in hydration free energies and structures for model compound mimics of charged side chains of amino acids. *The Journal of Physical Chemistry B* **125**, 4148–4161 (2021).

[39] Martin, E. W. & Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478–2487 (2018).

[40] Ball, R. C. *Scaling in Polymer Physics*, 483–490 (Springer US, Boston, MA, 1991). URL https://doi.org/10.1007/978-1-4757-1402-9_42.

[41] Dunstan, D. E. The viscosity-radius relationship for concentrated polymer solutions. *Scientific Reports* **9** (2019). URL https://doi.org/10.1038/s41598-018-36596-6.

[42] Majumder, S., Christiansen, H. & Janke, W. Understanding nonequilibrium scaling laws governing collapse of a polymer. *The European Physical Journal B* **93**, 1–19 (2020).

[43] Rubinstein, M., Colby, R. H. *et al. Polymer physics* Vol. 23 (Oxford university press New York, 2003).

[44] Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **109**, 16155–16160 (2012).

[45] Mugnai, M. L. *et al.* Sizes, conformational fluctuations, and saxs profiles for intrinsically disordered proteins. *bioRxiv* (2023).

[46] Clarke, J. P., Thibault, P. A., Salapa, H. E. & Levin, M. C. A comprehensive analysis of the role of hnrnp a1 function and dysfunction in the pathogenesis of neurodegenerative disease. *Frontiers in Molecular Biosciences* **8**, 659610 (2021).

[47] Kim, H. J. *et al.* Mutations in prion-like domains in hnrnpa2b1 and hnrnpa1 cause multisystem proteinopathy and als. *Nature* **495**, 467–473 (2013).

[48] Wang, I.-F., Wu, L.-S. & Shen, C. J. Tdp-43: an emerging new player in neurodegenerative diseases. *Trends in molecular medicine* **14**, 479–485 (2008).

[49] Jo, M. *et al.* The role of tdp-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Experimental & molecular medicine* **52**, 1652–1662 (2020).

[50] Ash, P. E. *et al.* Tia1 potentiates tau phase separation and promotes generation of toxic oligomeric tau. *Proceedings of the National Academy of Sciences* **118**, e2014188118 (2021).

[51] Apicco, D. J. *et al.* Reducing the rna binding protein tia1 protects against tau-mediated neurodegeneration in vivo. *Nature neuroscience* **21**, 72–80 (2018).

[52] Deng, H., Gao, K. & Jankovic, J. The role of fus gene variants in neurodegenerative diseases. *Nature Reviews Neurology* **10**, 337–348 (2014).

[53] Lagier-Tourenne, C. & Cleveland, D. W. Rethinking als: The fus about tdp-43. *Cell* **136**, 1001–1004 (2009).

[54] Couthouis, J. *et al.* Evaluating the role of the fus/tls-related gene ewsr1 in amyotrophic lateral sclerosis. *Human molecular genetics* **21**, 2899–2911 (2012).

[55] Lee, J. *et al.* Ewsr1, a multifunctional protein, regulates cellular function and aging via genetic and epigenetic pathways. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1865**, 1938–1945 (2019).

[56] Li, Y. R., King, O. D., Shorter, J. & Gitler, A. D. Stress granules as crucibles of als pathogenesis. *Journal of cell biology* **201**, 361–372 (2013).

[57] Farag, M., Borcherds, W. M., Bremer, A., Mittag, T. & Pappu, R. V. Phase separation in mixtures of prion-like low complexity domains is driven by the interplay of homotypic and heterotypic interactions. *bioRxiv* (2023).

[58] Reber, S. *et al.* The phase separation-dependent FUS interactome reveals nuclear and cytoplasmic function of liquid–liquid phase separation. *Nucleic Acids Research* **49**, 7713–7731 (2021). URL https://doi.org/10.1093/nar/gkab582.

[59] Feng, Z., Chen, X., Wu, X. & Zhang, M. Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *Journal of Biological Chemistry* **294**, 14823–14835 (2019).

[60] Welsh, T. J. *et al.* Surface electrostatics govern the emulsion stability of biomolecular condensates .

[61] Luo, F. *et al.* Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nat. Struct. Mol. Biol.* **25**, 341–346 (2018).

[62] Murray, D. T. *et al.* Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017).

[63] Qamar, S. *et al.* Fus phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* **173**, 720–734 (2018).

[64] Murakami, T. *et al.* Als/ftd mutation-induced phase transition of fus liquid droplets and reversible hydrogels into irreversible hydrogels impairs rnp granule function. *Neuron* **88**, 678–690 (2015).

[65] Patel, A. *et al.* A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).

[66] Murthy, A. C. *et al.* Molecular interactions underlying liquid- liquid phase separation of the fus low-complexity domain. *Nature structural & molecular biology* **26**, 637–648 (2019).

[67] Lie Zbinden, A., Pé Rez-Berlanga, M., De Rossi, P. & Polymenidou, M. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Developmental Cell* **55**, 45–68 (2020).

[68] Guenther, E. L. *et al.* Atomic structures of tdp-43 lcd segments and insights into reversible or pathogenic aggregation. *Nature structural & molecular biology* **25**, 463–471 (2018).

[69] Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *Journal of Cell Biology* **210**, 529–539 (2015). URL https://doi.org/10.1083/jcb.201504117.

[70] 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* **60**, 3566–3581 (2021).

[71] Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell* **60**, 208–219 (2015).

[72] Ryan, V. H. *et al.* Mechanistic view of hnrnpa2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Molecular cell* **69**, 465–479 (2018).

[73] Sanchez-Burgos, I., Espinosa, J. R., Joseph, J. A. & Collepardo-Guevara, R. Valency and binding affinity variations can regulate the multilayered organization of protein condensates with many components. *Biomolecules* **11**, 278 (2021).

[74] Choi, J.-M. & Pappu, R. V. The stickers and spacers framework for describing phase behavior of multivalent intrinsically disordered proteins. *Biophysical Journal* **118**, 492a (2020).

[75] Ginell, G. M. & Holehouse, A. S. in *An introduction to the stickers-and-spacers framework as applied to biomolecular condensates* 95–116 (Springer, 2022).

[76] Das, S., Lin, Y. H., Vernon, R. M., Forman-Kay, J. D. & Chan, H. S. Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 28795–28805 (2020).

[77] Dyson, H. J., Wright, P. E. & Scheraga, H. A. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences* **103**, 13057–13061 (2006).

[78] Andrew, C. D. *et al.* Stabilizing interactions between aromatic and basic side chains in α-helical peptides and proteins. tyrosine effects on helix circular dichroism. *Journal of the American Chemical Society* **124**, 12706–12714 (2002).

[79] Li, W. & Jiang, H. Nuclear protein condensates and their properties in regulation of gene expression. *Journal of Molecular Biology* **434**, 167151 (2022). URL https://www.sciencedirect.com/science/article/pii/S0022283621003806. Protein Interactions in Liquid-Liquid Phase Separation.

[80] Tejedor, A. R., Collepardo-Guevara, R., Ramirez, J. & Espinosa, J. R. Time-dependent material properties of aging biomolecular condensates from different viscoelasticity measurements in molecular dynamics simulations. *The Journal of Physical Chemistry B* (2023).

[81] Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

[82] Ray, S. & Buell, A. K. Emerging experimental methods to study the thermodynamics of biomolecular condensate formation. *The Journal of Chemical Physics* **160** (2024).

[83] Hong, Y. *et al.* Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins. *Nature Communications* **13**, 7326 (2022).

[84] Greig, J. A. *et al.* Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *Molecular cell* **77**, 1237–1250 (2020).

[85] Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).

[86] Mitchell, D. J., Steinman, L., Kim, D. T., Fathman, C. G. & Rothbard, J. B. Polyarginine enters cells more efficiently than other polycationic homopolymers. *Journal of Peptide Research* **56**, 318–325 (2000).

[87] Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **110**, 13392–13397 (2013).

[88] 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* **434**, 167373 (2022). URL https://www.sciencedirect.com/science/article/pii/S0022283621006100.

[89] Zeng, X., Ruff, K. M. & Pappu, R. V. Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **119** (2022). URL https://doi.org/10.1073/pnas.2200559119.

[90] De Sancho, D. Phase separation in amino acid mixtures is governed by composition. *Biophysical Journal* **121**, 4119–4127 (2022).

[91] Ahlers, J. *et al.* The key role of solvent in condensation: Mapping water in liquid-liquid phase-separated fus. *Biophysical Journal* **120**, 1266–1275 (2021). URL https://www.sciencedirect.com/science/article/pii/S0006349521000606.

[92] Nomoto, A., Nishinami, S. & Shiraki, K. Solubility parameters of amino acids on liquid–liquid phase separation and aggregation of proteins. *Frontiers in Cell and Developmental Biology* **9**, 691052 (2021).

[93] Zheng, W., Dignon, G. L., Brown, M. L., Kim, Y. C. & Mittal, J. Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins. *Journal of Physical Chemistry Letters* **11**, 3408–3415 (2020).

[94] Weiner, B. G., Pyo, A. G., Meir, Y. & Wingreen, N. S. Motif-pattern dependence of biomolecular phase separation driven by specific interactions. *PLoS computational biology* **17**, e1009748 (2021).

[95] Huihui, J. & Ghosh, K. Intrachain interaction topology can identify functionally similar intrinsically disordered proteins. *Biophysical Journal* **120**, 1860–1868 (2021).

[96] Ghosh, K., Huihui, J., Phillips, M. & Haider, A. Rules of physical mathematics govern intrinsically disordered proteins. *Annual Review of Biophysics* **51**, 355–376 (2022). URL https://doi.org/10.1146/annurev-biophys-120221-095357.

[97] Rana, U., Brangwynne, C. P. & Panagiotopoulos, A. Z. Phase separation vs aggregation behavior for model disordered proteins. *The Journal of Chemical Physics* **155**, 125101 (2021). URL https://doi.org/10.1063/5.0060046.

[98] Martin, E. W. *et al.* Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *Journal of the American Chemical Society* **138**, 15323–15335 (2016).

[99] Holehouse, A. S., Ginell, G. M., Griffith, D. & Boke, E. Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates: published as part of the biochemistry virtual special issue "protein condensates". *Biochemistry* **60**, 3566–3581 (2021).

[100] Regy, R. M., Thompson, J., Kim, Y. C. & Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **30**, 1371–1379 (2021). URL https://doi.org/10.1002/pro.4094.

[101] Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Computational Biology* **14** (2018).

[102] 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* **118** (2021). URL https://doi.org/10.1073/pnas.2111696118.

[103] Regy, R. M., Zheng, W. & Mittal, J. Using a sequence-specific coarse-grained model for studying protein liquid–liquid phase separation 1–17 (2021). URL https://doi.org/10.1016/bs.mie.2020.07.009.

[104] Latham, A. P. & Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *Journal of Chemical Theory and Computation* **16**, 773–781 (2019). URL https://doi.org/10.1021/acs.jctc.9b00932.

[105] Park, S., Lee, Y., Jho, Y. & Hwang, D. S. Molecular hydration tunes the cation–π interaction strength in aqueous solution. *Advanced Materials Interfaces* **10**, 2201732 (2022). URL https://doi.org/10.1002/admi.202201732.

[106] Gallivan, J. P. & Dougherty, D. A. A computational study of cation-π interactions vs salt bridges in aqueous media: implications for protein engineering. *Journal of the American Chemical Society* **122**, 870–874 (2000). URL https://doi.org/10.1021/ja991755c.

[107] Ukmar-Godec, T. *et al.* Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nature Communications* **10** (2019). URL https://doi.org/10.1038/s41467-019-10792-y.

[108] Thompson, A. P. *et al.* LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.* **271**, 108171 (2022).

[109] Grønbech-Jensen, N. & Farago, O. A simple and effective verlet-type algorithm for simulating langevin dynamics. *Molecular Physics* **111**, 983–991 (2013).

Supporting Information

A Amino acid Sequences of PLDs

RBM14

```
WT(Q96PK6)   1 PMTAQAASYRAQPSVSLGAPYRGQLASPSSQSAAASSLGPYGGAQPSASALSSYGGQAAAASSLNSYGAQGSSLASYG
-20Y         1 ........G...........S...................Q............G.......................S.
-10Y         1 ........G...........S...................Q.......................................
-5Y          1 ......................S.......................................................
-30Y+30F     1 ........F...........F...................F............F............F.........F.
-20Y+20F     1 ........F...........F...................................F............F...........
-10Y+10F     1 ........F...........................................F...........................
-5Y+5F       1 ........F...........................................................................
-10T+10S     1 ..S.....................................................................................
-10T+10G     1 ..G.....................................................................................
-5T+5S       1 ..S.....................................................................................
-5T+5G       1 ..G.....................................................................................
-20Q+20N     1 ....N...................N.................................N.......................
-10Q+20N     1 ....N.........................N.......................................N........
-5Q+5N       1 ....N...................................................................................
-4N+4Q       1 ...................................................................Q.............

WT(Q96PK6)  79 NQPSSYGAQAASSYGVRAAASSYNTQGAASSLGSYGAQAASYGAQSAASSLAYGAQAASYNAQPSASYNAQSAPYAAQ
-20Y        79 .....G.......Q........Q...........G......S.................G.......Q......S...
-10Y        79 .............Q....................S.........................G...................
-5Y         79 .............G.......................................................................
-30Y+30F    79 .....F.......F........F...........F......F..........F......F.......F......F...
-20Y+20F    79 .....F.......F....................F......F.................F.......F..........
-10Y+10F    79 .....F.............................F..........................F...................
-5Y+5F      79 .............F.......................................................................F...
-10T+10S    79 ........................S..........................................................
-10T+10G    79 ........................G..........................................................
-5T+5S      79 ........................................................................................
-5T+5G      79 ........................................................................................
-20Q+20N    79 .N.......................N..................N.................N...............N
-10Q+20N    79 ........................N.............................N......................N
-5Q+5N      79 .N.......................................................................N.......
-4N+4Q      79 ........................Q.......................................................Q.........

WT(Q96PK6) 157 QAASYSSQPAAYVAQPATAAAYASQPAAYAAQATTPMAGSYGAQPVVQTQLNSYGAQASMGLSGSYGAQSAAAATGSY
-20Y       157 ....Q.................S......S...........Q............S..........................
-10Y       157 ............................S...........Q.......................................
-5Y        157 ......................S.................................Q..........................
-30Y+30F   157 ....F......F.........F......F...........F............F...........F............F
-20Y+20F   157 ....F......F.................F...........F..........................F............F
-10Y+10F   157 ....F.......................F.......................................F............
-5Y+5F     157 ........................................................................F............
-10T+10S   157 ..................S...............SS.............S...........................S...
-10T+10G   157 ..................G...............GG.............G...........................G...
-5T+5S     157 ..................S................S...........................................S...
-5T+5G     157 ..................G................G...........................................G...
-20Q+20N   157 N.............N.........N..................N...N........N.....................
-10Q+20N   157 N.......................................................N........N.....................
-5Q+5N     157 ...............................................N....................................
-4N+4Q     157 ........................................................................................

WT(Q96PK6) 235 GAAAAYGAQPSATLAAPYRTQSSASLAASYAAQQHPQAAASYRGQPGNAYDGAGQPSAAYLSMSQGAVANANSTPPPY
-20Y       235 .....S...........Q...........Q...........S.......S.........Q..................S
-10Y       235 ..................S.............................S.............................
-5Y        235 ........................................S.......................................
-30Y+30F   235 .....F...........F...........F...........F.......F.........F..................F
-20Y+20F   235 ..................F...........F...................F.........F...................
-10Y+10F   235 ..................F.........................F.......F.............................
-5Y+5F     235 ...................................................................F...................
-10T+10S   235 ............S......S.......................................................S....
-10T+10G   235 ............G......G.......................................................G....
-5T+5S     235 ....................S.......................................................
-5T+5G     235 ....................G.......................................................
-20Q+20N   235 ........N.........................NN..........N.........N.........N.............
-10Q+20N   235 ....................N.......................................................
-5Q+5N     235 ........................................N...........................................
-4N+4Q     235 ...................................................Q.........................Q......
```

TDP43

```
WT(Q13148)    1 R G G G A G L G N N Q G S N M G G G M N F G A F S I N P A M M A A A Q A A L Q S S W G M M G M L A S Q Q N Q S G P S G N N
-Y+F          1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5F+5Y        1 . . . . . . . . . . . . . . . . . . . . Y . . Y . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5F           1 . . . . . . . . . . . . . . . . . . . . G . . S . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-20Q+20N      1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-10Q+10N      1 . . . . . . . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . N . . . N . . . . . . . . . . . N N . N . . . . . . .
-5Q+5N        1 . . . . . . . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . N . . . . . . . . . . . . N . . . . . . . .
-15N+15Q      1 . . . . . . . . Q . . . . Q . . . . . Q . . . . . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . Q . . . . . . Q Q
-10N+10Q      1 . . . . . . . . Q . . . . Q . . . . . Q . . . . . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Q Q
-5N+5Q        1 . . . . . . . . Q . . . . . . . . . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Q .

WT(Q13148)   62 Q N Q G N M Q R E P N Q A F G S G N N S Y S G S N S G A A I G W G S A S N A G S G S G F N G G F G S S M D S K S S G W G M
-Y+F         62 . . . . . . . . . . . . . . . . . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5F+5Y       62 . . . . . . . . . . . . . Y . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5F          62 . . . . . . . . . . . . . G . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-20Q+20N     62 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-10Q+10N     62 N . N . . . N . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5Q+5N       62 N . . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-15N+15Q     62 . Q . . Q . . . . . Q . . . . . . Q Q . . . . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-10N+10Q     62 . Q . . . . . . . . Q . . . . . . Q Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5N+5Q       62 . . . . Q . . . . . . . . . . . . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

TIA1

```
WT(P31483)    1 M I N P V Q Q Q N Q I G Y P Q P Y G Q W G Q W Y G N A Q Q I G Q Y M P N G W Q V P A Y G M Y G Q A W N Q Q G F N Q T Q
-12Y          1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-6Y           1 . . . . . . . . . . . . Q . . . G . . . . . . S . . . . . . . . . . . . . . . . . . Q . . . . . . . . . . . . . . . .
-9Y+9F        1 . . . . . . . . . . . . F . . . F . . . . . . F . . . . . . . . F . . . . . . . . . F . . F . . . . . . . . . . . . .
-4Y+4F        1 . . . . . . . . . . . . F . . . . . . . . . F . . . . . . . . . . . . . . . . . . F . . . . . . . . . . . . . . . . .
-9Y+9W        1 . . . . . . . . . . . . W . . . W . . . . . . W . . . . . . . . W . . . . . . . . . W . . W . . . . . . . . . . . . .
-4Y+4W        1 . . . . . . . . . . . . W . . . . . . . . . . W . . . . . . . . . . . . . . . . . . W . . . . . . . . . . . . . . . . .
-5W+5F        1 . . . . . . . . . . . . . . . . . . . . F . . F . . . . . . . . . . . . . . F . . . . . . . . . . . . F . . . . . . . . .
-5W+5Y        1 . . . . . . . . . . . . . . . . . . . . Y . . Y . . . . . . . . . . . . . . Y . . . . . . . . . . . . Y . . . . . . . . .
-20Q+20N      1 . . . . . N N N . N . . . . N . . . N . . N . . . . . N N . . N . . . . . . N . . . . . . . . N . . . N N . . . N . N
-10Q+10N      1 . . . . . N N N . . . . . . . . . . . . . . . . . . N N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . N . .
-9N+9Q        1 . . Q . . . . . Q . . . . . . . . . . . . . . . . Q . . . . . . . . . Q . . . . . . . . . . . . . . . Q . . . . Q . . .

WT(P31483)   60 S S A P W M G P N Y G V Q P P Q G Q N G S M L P N Q P S G Y R V A G Y E T N
-12Y         60 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-6Y          60 . . . . . . . . . . G . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-9Y+9F       60 . . . . . . . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-4Y+4F       60 . . . . . . . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-9Y+9W       60 . . . . . . . . . . W . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-4Y+4W       60 . . . . . . . . . . W . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5W+5F       60 . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-5W+5Y       60 . . . . Y . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-20Q+20N     60 . . . . . . . . . . . . . N . . N . N . . . . . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-10Q+10N     60 . . . . . . . . . . . . . N . . . . N . . . . . . . . N . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-9N+9Q       60 . . . . . . . . Q . . . . . . . . . Q . . . . . Q . . . . . . . . . . . . . . -
```

914 **FUS**

```
(P35637)      1 MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQS
-28Y+28F      1 .....F.......F..F.......F.......F....F..F........F....F..F...
-14Y+14F      1 .....F..........F................F.......F.............F......
-7Y+7F        1 .....F.........................F.......................F......
-14Y          1 ............Q..........G............S...........Q.......G...
-7Y           1 ........................Q.........................G...........
-10T+10S      1 ......S...S.......S.........................S.S..............
-10T+10G      1 ......G...G.......G.........................G.G..............
-5T+5S        1 ......S...........S.........................S...............
-5T+5G        1 ......G...........G.........................G...............
-10G+10T      1 ..............T.......................................T..........
-10S+10T      1 ..T.......................................T..................T....
-20Q+20N      1 .......N...N..........N....N......N.......N................N.
-10Q+10N      1 .......N..............N...........N........................N.
-5Q+5N        1 .......N.............................N.........................
-4N+4Q        1 ..........................................................................

(P35637)     62 QNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSSSY
-28Y+28F     62 ....F........F.....F.........F.....F..F............F........F
-14Y+14F     62 ....F..............F................F................F.........
-7Y+7F       62 ...................F...................................F.........
-14Y         62 ............S................Q........G......................S
-7Y          62 ............S.........................Q.......................
-10T+10S     62 ..S...S..S......S..............................S..............
-10T+10G     62 ..G...G..G......G..............................G..............
-5T+5S       62 ......S.........S...........................................
-5T+5G       62 ......G.........G...........................................
-10G+10T     62 ..................T.................................T...........
-10S+10T     62 .........................T.....................T.........T..T..
-20Q+20N     62 .......N................N.......N........N................N....
-10Q+10N     62 .........................N................N.....................
-5Q+5N       62 ..............................N.............................
-4N+4Q       62 .Q.........................................................................

(P35637)    123 GQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNSSSGGGGGGGGGGNYGQDQSS
-28Y+28F    123 .......F.....F......F.....F.....F.....F................F......
-14Y+14F    123 .......F............F...........F.......................F......
-7Y+7F      123 ....................F.....................................F......
-14Y        123 ............Q............S...........G.......................
-7Y         123 ............S.........................G.......................
-10T+10S    123 ..........................................................................
-10T+10G    123 ..........................................................................
-5T+5S      123 ..........................................................................
-5T+5G      123 ..........................................................................
-10G+10T    123 ......................T........................T.....T........
-10S+10T    123 ............T.............................T....................
-20Q+20N    123 ...N......N......N....N.N.........N..N....................N..
-10Q+10N    123 ...N.............N......N............N.......................
-5Q+5N      123 ................N.......................................N..
-4N+4Q      123 ........................................Q..Q.............Q.......

(P35637)    184 MSSGGGSGGGYGNQDQSGGGGSGGYGQQDRGGRGRGGSGGGGGGGGGGYNRSSG
-28Y+28F    184 ..........F.............F.......................F.....
-14Y+14F    184 ........................F.............................
-7Y+7F      184 ..........................................................
-14Y        184 ..........G.......................................G.....
-7Y         184 ..........G.......................................
-10T+10S    184 ..........................................................
-10T+10G    184 ..........................................................
-5T+5S      184 ..........................................................
-5T+5G      184 ..........................................................
-10G+10T    184 ........T...........T..................T................
-10S+10T    184 ..T.......................................................
-20Q+20N    184 ..........................................................
-10Q+10N    184 ..........................................................
-5Q+5N      184 ..........................................................
-4N+4Q      184 ............Q.........................................
```

EWSR1

```
WT(Q01844)     1 MASTDYSTYSQAAAQQGYSAYTAQPTQGYAQTTQAYGQQSYGTYGQPTDVSYTQAQTTATYGQTAYATSYGQP
-20Y           1 .....T..Q........S..G....................T...........Q..............S.......
-10Y           1 .....T..Q........S..G.......................................................T.......
-5Y            1 .....T.......................................................................Q...
-30Y+30F       1 .....F..F........F.......F.....F.....F.........F........F....F...F....F...
-20Y+20F       1 .....F...........F..........F...........F.........F.............F.......
-10Y+10F       1 .....F...............F...................F.....................F............
-5Y+5F         1 ........F.......................................................F............
-30T+30S       1 ...S................S...S.....S.........S....S....S...S.S...S.........
-30T+30G       1 ...G................G...G.....G.........G....G....G...G.G...G.........
-20T+20S       1 ...S................S.........S.........S.........S...S.....S.........
-10T+10S       1 .......S.......................S.......................S...........S.....
-20T+20G       1 ...G................G.........G.........G.........G...G.....G.........
-10T+10G       1 ...G.......................G.....................G...........G.........
-30S+30T       1 ..T...T..T.....................T...........T..................T....
-20G+20T       1 ................T..........T............T..T.......................T..
-10G+10T       1 ................T.......................T...............................T..
-20S+20T       1 ..T......T.......................T...............................T....
-10S+10T       1 ..T..............T...............................................T....
-20Q+20N       1 ..........N....N..............N......N.......N.......N..................N.
-10Q+10N       1 ..........N...................N......................N...................
-5Q+5N         1 ..........N......................................N...................
-4N+4Q         1 ....................................................................

WT(Q01844)    74 PTGYTTPTAPQAYSQPVQGYGTGAYDTTTATVTTTQASYAAQSAYGTQPAYPAYGQQPAATAPTRPQDGNKPT
-20Y          74 ...T........G...........Q.......................T..S...................
-10Y          74 ...Q................S...................G.......................................
-5Y           74 ....................S...................G.......................................
-30Y+30F      74 ............F......F....F.............F.....F........F...................
-20Y+20F      74 ...F................F....F.................F........F...................
-10Y+10F      74 ............F.......................F...............F...................
-5Y+5F        74 .......................F..............................................
-30T+30S      74 .S..SS.S...................SSS...SSS.............................S........S
-30T+30G      74 ....GG.G...................GG..G.GG............G.............G...........G
-20T+20S      74 .S...S................S.....S..S..S............S................S.........
-10T+10S      74 .......S.......................S.....S...................................S
-20T+20G      74 .G...G................G.....G..G..G............G................G.........
-10T+10G      74 .....G......................G.....G..............................G.........
-30S+30T      74 ......................................T....T...................................
-20G+20T      74 ..T..................T.T.............................T.............T....
-10G+10T      74 ....................T.......................T.......................T....
-20S+20T      74 ......................................T.......................................
-10S+10T      74 ....................................................................
-20Q+20N      74 ...............N....................N...................N...........N......
-10Q+10N      74 ...............N.....................................N..................
-5Q+5N        74 ....................................................N..................
-4N+4Q        74 ....................................................................Q...

WT(Q01844)   147 ETSQPQSSTGGYNQPSLGYGQSNYSYPQVPGSYPMQPVTAPPSYPPTSYSSTQPTSYDQSSYSQQNTYGQPSS
-20Y         147 ......................Q.G......T.......................S..........Q......
-10Y         147 ...........T...............................S.......................................
-5Y          147 ................T.......................................................
-30Y+30F     147 ..........F......F....F........F..........F....F.......F....F.....F.....
-20Y+20F     147 ................F......F...............F............F..........F.....
-10Y+10F     147 ......................F...................F...................F............
-5Y+5F       147 ................F.....................................F.................
-30T+30S     147 .S.................................S.......S....S..S...........S......
-30T+30G     147 .G......G..........................G.......G....G..G...........G......
-20T+20S     147 .S.................................S............S..............S......
-10T+10S     147 ..................................................S.......................
-20T+20G     147 .G.................................G............G..............G......
-10T+10G     147 ...................................G.......................G.....
-30S+30T     147 ..T....T.......T....T.........T..........T....T..T....T...TT.T........TT
-20G+20T     147 ..........T......T.T.........T.....................................
-10G+10T     147 .........T.................T.....................................
-20S+20T     147 ..T....T.............T.........T.......................T..T........T..T.....T
-10S+10T     147 ..T................T.......................T.....................T.
-20Q+20N     147 .....N.....................N.....................N..........N.........
-10Q+10N     147 ............N.....................................N.................
-5Q+5N       147 ....................................................N.................
-4N+4Q       147 ............Q.........Q.......................................Q.......

WT(Q01844)   220 YGQQSSYGQQSSYGQQPPTSYPPQTGSYSQAPSQYSQQSSSYGQQSSFRQDHPSSMGVYGQ
-20Y         220 T.....................G.....................S.....................
-10Y         220 ....................................................................
-5Y          220 ....................................................................
-30Y+30F     220 F.....F......F.......F......F.......................................
-20Y+20F     220 ......F.............F.............F.......................F..
-10Y+10F     220 F...................................................................
-5Y+5F       220 ....................................................................
-30T+30S     220 ................S.....S.............................................
-30T+30G     220 ................G.....G.............................................
-20T+20S     220 ....................S...............................................
-10T+10S     220 ................S...................................................
-20T+20G     220 ....................G...............................................
-10T+10G     220 ....................................................................
-30S+30T     220 .....T....TT.............T.....T..T....T.....T.......................
-20G+20T     220 .T.....T................T...............T............T..T.
-10G+10T     220 .......T...................................T.....................
-20S+20T     220 .....T.....T.............T.....T.....T.T.....................
-10S+10T     220 ...........T...............T...........T.....................
-20Q+20N     220 ..N.....N......N............N......N.....................
-10Q+10N     220 ...N...................N...................N.....................
-5Q+5N       220 ....................N...................................
-4N+4Q       220 ....................................................................
```

B Convergence tests

Molecular Dynamic simulations are inherently sensitive to both the length of simulated time and system scaling factor such as the number of particles present in the system and the size of the simulation box. In this study, we systematically tested all these aspects to ensure proper convergence of the results.

The system was placed in the middle of the simulation box, and the length of X and Y was the same as the length of the condensate, while the box was elongated in Z, to about ten times the size of X, Y(see 1). Importantly, the cross-section is large enough to avoid protein self-interactions through the periodic boundary conditions (e.g., cross-section sizes are at least twice the protein radius of gyration in condensate bulk conditions).

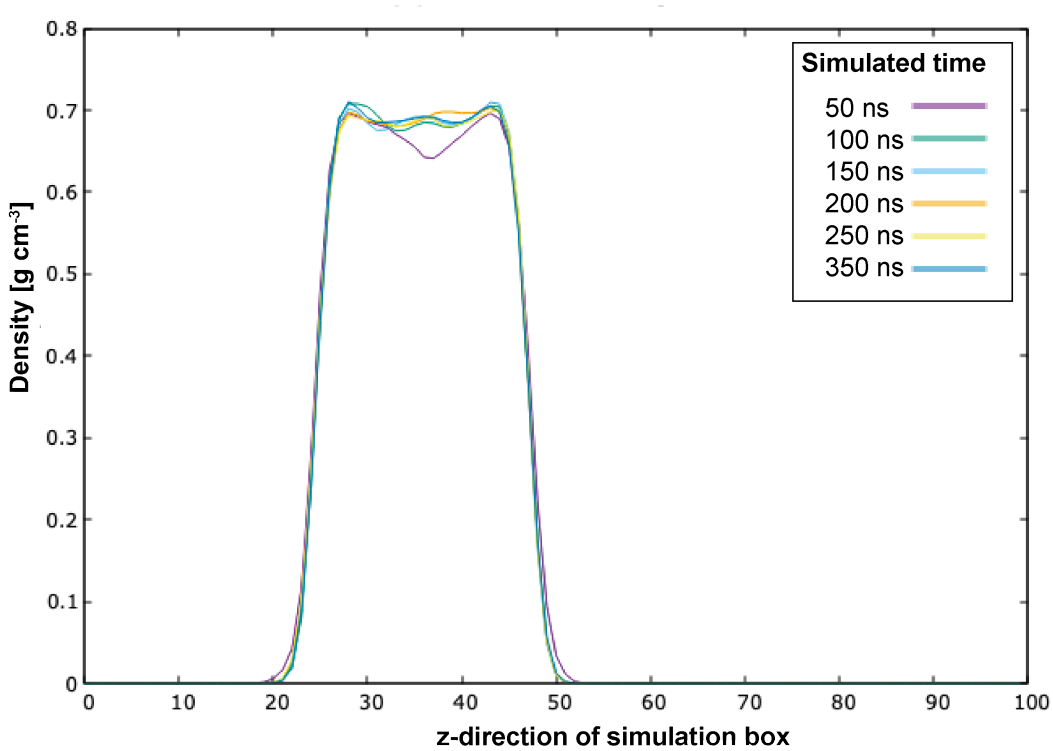

Fig. 8: Convergence tests for the WT sequence of the PLD of FUS. The density profiles of different and independent simulations are plotted across the perpendicular axis to the condensate interfaces. Each simulation was performed for different timescales (as specified in the legend) to check for proper convergence.

We tested for time convergence by plotting the obtained density of the protein wild types after increasing simulation times, from 50 ns to 350 ns. On simulated timesteps above 100 ns, the density of the condensate exhibited no significant differences. We thus concluded that, to optimize speed-up, 200 ns provided sufficient sampling for each point of the binodal.

Finally, we decided to keep the number of particles at a constant value of 64, which we found to be a good balance between minimizing finite-size effects and optimizing simulation speed (see Figure 9). We tested this by varying the corresponding number of protein chains of the FUS WT as a model protein from 27 to 216 and found that in these systems there was not a significant difference in the simulated density. Additionally, with 64 protein chains we could ensure a bulk large enough to sustain the condition that the cross section of the box should be higher than twice the radii of gyration of the protein chain.

B.1 Normalised and comparative contact maps

Further study of the differences between variants at the chemical level was carried out by computing the inter-molecular contacts per residue inside a phase-separated droplet. For each variant, the contact matrix was computed at a temperature around ten percent below the T_c. The cutoff to define a pairwise distance as a contact was dynamically calculated for each pair, as 1.2 times the average Van der Waals radius of both pair residues. Given residues i and j belong to different molecules, the contact maps were computed as shown below:

$$c_{i,j} = \sum_{i,j} \chi_{ij} \tag{22}$$

$$\chi_{ij} = \begin{cases} 1 & \text{if } |r_i - r_j| \leq r_{ij}^{\text{cutoff}} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

To get a more accurate picture of the main interactions driving phase separation for each variant, we normalised the contact maps to account for the occurrence of each residue in the sequence. This allowed us to account for the disparities in residue distributions in the sequences, and extract the relevant residue–residue

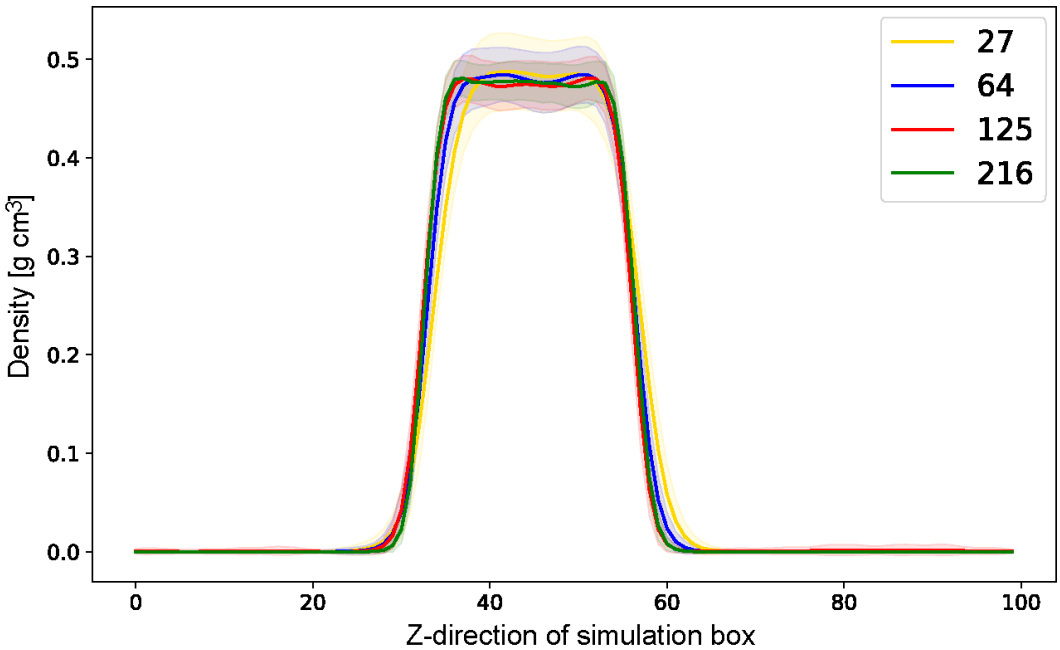

Fig. 9: Convergence tests for the WT sequence of the PLD of FUS. The density profiles of different and independent simulations are plotted across the perpendicular axis to the condensate interfaces. Each simulation was performed for different system sizes (aka, number of chains, as specified in the legend) to check for proper convergence.

interactions. For each residue–residue pair, the normalisation was performed as follows:

$$C = \begin{vmatrix} \alpha_{M,M} & \alpha_{M,G} & \dots & \alpha_{M,I} \\ \alpha_{G,M} & \alpha_{G,G} & \dots & \alpha_{G,I} \\ \dots & \dots & \dots & \dots \\ \alpha_{I,M} & \alpha_{I,G} & \dots & \alpha_{I,I} \end{vmatrix} \tag{24}$$

$$\alpha_{i,j} = \frac{c_{i,j}}{n_i \cdot n_j} \tag{25}$$

where n_i is the number of i residues in the IDP sequence. The resulting contact matrix is min–max normalised:

$$\beta_{i,j} = \frac{\alpha_{i,j} - \min C}{\max C - \min C} \tag{26}$$

$$C_{\text{norm}} = \begin{vmatrix} \beta_{M,M} & \beta_{M,G} & \dots & \beta_{M,I} \\ \beta_{G,M} & \beta_{G,G} & \dots & \beta_{G,I} \\ \dots & \dots & \dots & \dots \\ \beta_{I,M} & \beta_{I,G} & \dots & \beta_{I,I} \end{vmatrix} \tag{27}$$

The comparative contact maps (Fig. 10 and 11), on the other hand, are computed as the difference in contacts of a protein variant respective to its wild type, as:

$$C_{\text{comparative}} = C_{\text{variant}} - C_{\text{WT}} \tag{28}$$

which is finally mean-normalised.

$$\gamma_{i,j} = \frac{\alpha_{i,j} - \overline{C}_{\text{comparative}}}{\max C_{\text{comparative}} - \min C_{\text{comparative}}} \tag{29}$$

$$C_{\text{comparative,normed}} = \begin{vmatrix} \gamma_{M,M} & \gamma_{M,G} & \dots & \gamma_{M,I} \\ \gamma_{G,M} & \gamma_{G,G} & \dots & \gamma_{G,I} \\ \dots & \dots & \dots & \dots \\ \gamma_{I,M} & \gamma_{I,G} & \dots & \gamma_{I,I} \end{vmatrix} \tag{30}$$

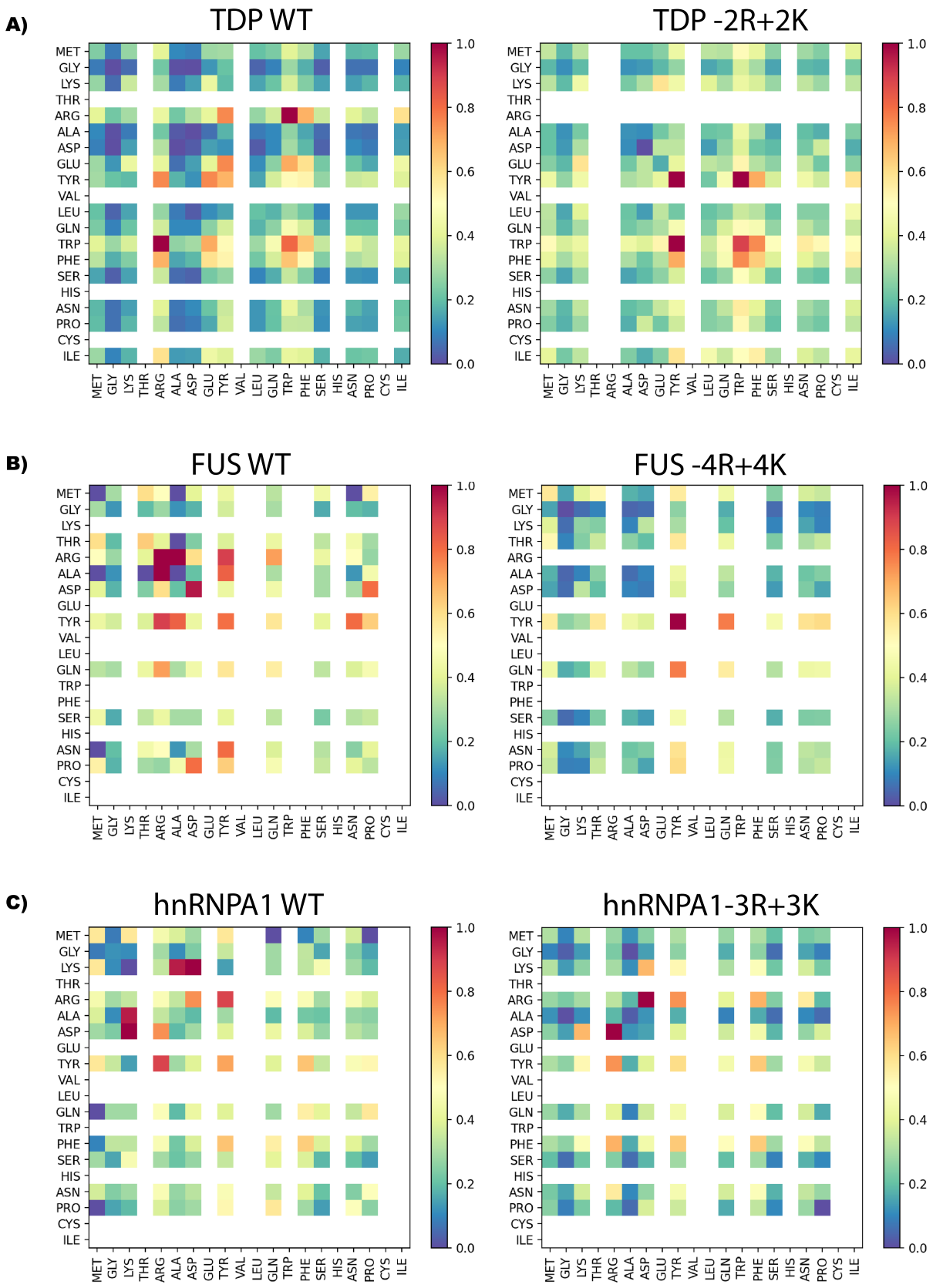

Fig. 10: Normalised contact maps of different *R to K* variants of PLDs: A) TDP B) FUS, C) hnRNPA1

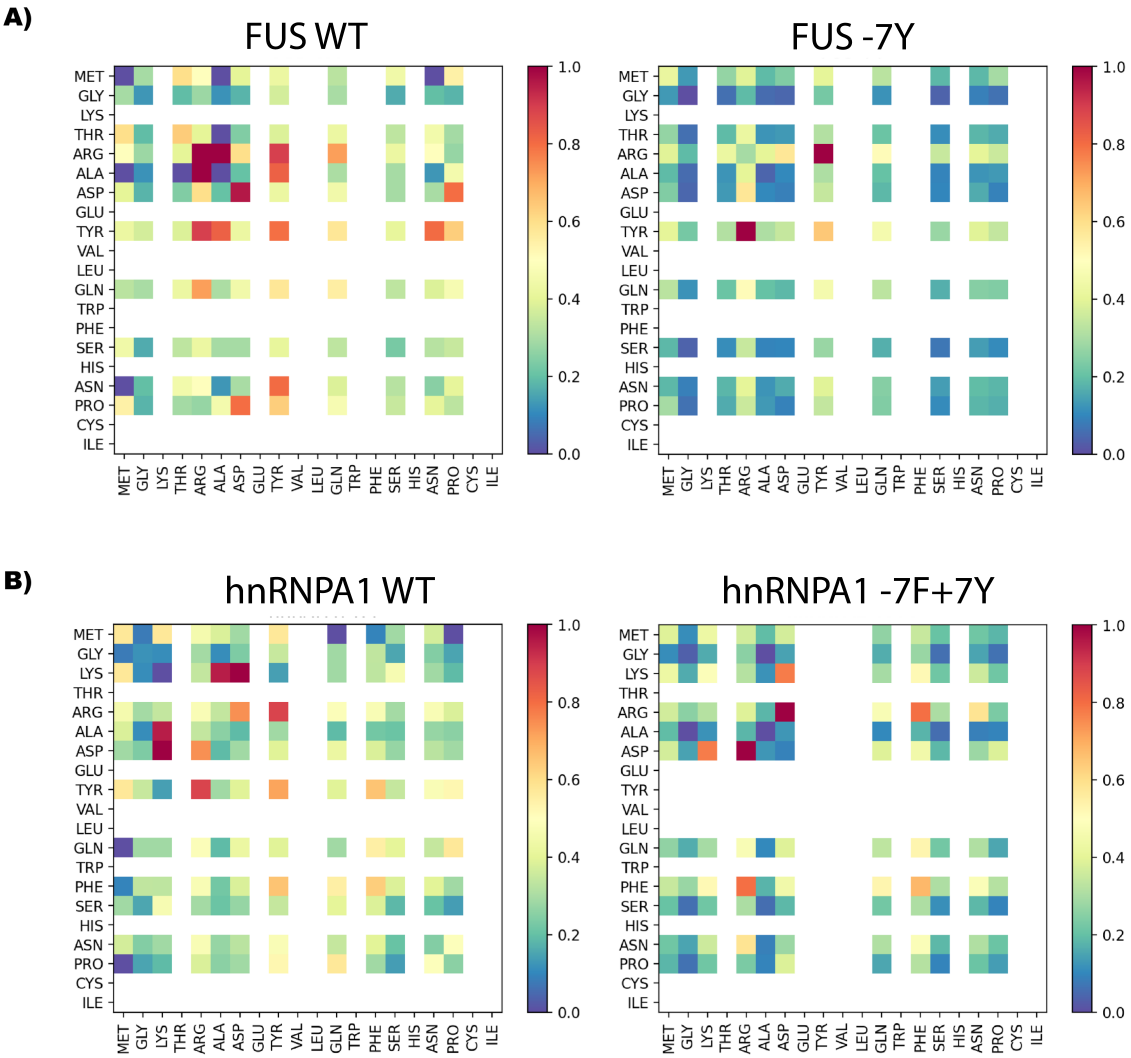

Fig. 11: Normalised contact maps of different aromatic variants of PLDs: A) FUS B) hnRNPA1.