Biophysical Journal
Article

Biophysical Society

Phase Separation of Toxic Dipeptide Repeat Proteins Related to C9orf72 ALS/FTD

Hamidreza Jafarinia,[1] Erik van der Giessen,[1] and Patrick R. Onck[1,*]
[1]Zernike Institute for Advanced Materials, University of Groningen, Groningen, the Netherlands

ABSTRACT The expansion mutation in the C9orf72 gene is the most common known genetic cause for amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). This mutation can produce five dipeptide repeat proteins (DPRs), of which three are known to be toxic: poly-PR, poly-GR, and poly-GA. The toxicity of poly-GA is attributed to its aggregation in the cytoplasm, whereas for poly-PR and poly-GR, several toxicity pathways have been proposed. The toxicity of the DPRs has been shown to depend on their length, but the underlying molecular mechanism of this length dependence is not well understood. To address the possible role of phase separation in DPR toxicity, a one-bead-per-amino-acid (1BPA) coarse-grained molecular dynamics model is used to study the single-molecule and phase-separation properties of the DPRs. We find a strong dependence of the phase-separation behavior on both DPR length and concentration, with longer DPRs having a higher propensity to phase separate and form condensed phases with higher concentrations. The critical lengths required for phase separation (25 for poly-PR and 50 for poly-GA) are comparable to the toxicity threshold limit of 30 repeats found for the expansion mutation in patient cells, suggesting that phase separation could play an important role in DPR toxicity.

SIGNIFICANCE C9orf72 ALS/FTD is caused by a repeat expansion mutation that typically has a length larger than 30 repeats. This expansion produces dipeptide repeat proteins (DPRs) that can induce length-dependent toxicity through several hypothesized pathways. However, no consensus has been reached on the prevailing toxicity pathway, and the underlying molecular mechanisms of the length dependence remain elusive. Here, we use a coarse-grained molecular dynamics model to explore the phase-separation behavior of the DPRs. We find the phase separation of the DPRs ensues only above a critical DPR length that is comparable to the toxicity threshold limit of 30 repeats found in patients, suggesting that phase separation of DPRs might play an important role in C9orf72 ALS/FTD.

INTRODUCTION

Hexanucleotide repeat expansion G_4C_2 in the C9orf72 gene is the most common genetic mutation in familial cases of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (1,2). Healthy individuals typically have less than around 20 repeats of this expansion, whereas in most patient cells, the size of the expansions is estimated to be between several hundred and several thousand repeat units (1–3). There is no consensus on the critical expansion size for the onset of the disease, and different cutoffs between 30 and 80 repeats have been reported for the toxicity threshold (1,2,4,5).

The pathology initiated by the repeat expansion has been proposed to affect a wide range of cellular processes (6). The three main mechanisms of toxicity are loss of function of C9orf72 proteins (1,7) and toxic gain of function from the repeat expansion itself (8,9) or from dipeptide repeat proteins (DPRs) translated from sense and antisense transcripts of the repeat expansion (10–12). It has been shown that DPRs are capable of inducing toxicity without the repeat expansion in different cell types (12–17).

Repeat-associated non-AUG (RAN) translation of the sense and antisense transcripts of the repeat expansion from all reading frames can produce five types of DPRs: poly-PR, poly-GR, poly-GA, poly-GP, and poly-PA (10,11). Poly-PR, poly-GR, and poly-GA can induce length-dependent and dosage-dependent toxicity (12,14,16,18–21), of which especially the R-DPRs, i.e., poly-PR and poly-GR, are highly toxic. Poly-PR is known to be the most toxic DPR (12–15,20,22). Several studies indicate no significant toxicity for poly-GP and poly-PA (12,20,23). In our study, we use the term toxic DPRs to refer to poly-PR, poly-GR, and poly-GA.

Submitted March 25, 2020, and accepted for publication July 6, 2020.
*Correspondence: p.r.onck@rug.nl
Editor: Samrat Mukhopadhyay.
https://doi.org/10.1016/j.bpj.2020.07.005

Poly-GA is the most abundant and the most aggregation-prone DPR (10,22). Poly-GA has only a few interactors in the cell (22). It has been suggested that poly-GA toxicity is due to the formation of cytoplasmic aggregates and direct sequestration of proteins (24–28). Poly-GA aggregates are shown to impair the nuclear import of TDP-43 (29) and enhance DNA damage (27). Poly-PR and poly-GR have many target proteins inside the cell (22,30) and are likely to be involved in several pathology pathways (reviewed in (6)). Significant attention has been drawn to nucleocytoplasmic transport defects, changes in the dynamics of membrane-less organelles through impaired liquid-liquid phase separation (LLPS), and nucleolar dysfunction (13,14,16,17,22,31,32). Recently, the toxicity of poly-PR has been related to changes in LLPS of heterochromatin protein 1α (HP1α) (17) and nucleophasmin (NPM1) (16) through their direct interaction with poly-PR inside the cell nucleus.

Depending on the type of DPR and the toxicity mechanism, the DPR pathology can start in the cytoplasm, nuclear pore complex (NPC), or nucleus. Despite the recent progress, a clear understanding is lacking on how the DPRs cause neurotoxicity in C9orf72 ALS/FTD. Because of the methodological difficulty of synthesizing repetitive sequences, it is highly challenging to study the length-dependent properties of DPRs that might provide insight into the DPRs' length-dependent toxicity. As a result, almost all previous in vitro cell-free studies were limited to DPRs with less than 30 repeats (21,22,33,34) (only one recent study used 60 repeats of poly-PR (16)), which might not mimic the exact role of long DPRs in patients (35). Moreover, the concentration inside the condensed phases, which is important for further maturation of the phase-separated condensates (36), is not easily determined in experiments. These problems can be overcome by using experimentally calibrated coarse-grained molecular dynamics (MD) models that can capture the sequence specificity and are suitable for simulating high-density phases of proteins (37–39). In this study, we use our one-bead-per-amino-acid (1BPA) coarse-grained MD model (37,40,41) to investigate the single-molecule and phase-separation behavior of toxic DPRs in an attempt to identify possible mechanistic roles of DPR phase separation in causing toxicity.

METHODS

Coarse-grained force field

We use our implicit-solvent coarse-grained 1BPA model (37,42), which 1) differentiates between the bonded potentials of glycine, proline, and other residues and 2) is fine-tuned to capture the properties of polyproline, polyglycine, and FG-Nup segments with the highest arginine content. More details are provided in the Supporting Materials and Methods.

Simulations

Langevin dynamics simulations are performed at 300 K and physiological salt concentration of 150 mM using GROMACS. Droplet simulations are performed in a cubic box of size 80 nm. Slab simulations are conducted based on the procedure described in (38). Phase-separation simulations are performed for ~3 μs, which is sufficient to obtain converged density profiles. The density profiles are calculated using discrete cells of thickness 1 nm and time averaged for at least 1 μs at equilibrium. The critical repeat lengths required for phase separation are obtained with an error of less than five repeats. More details are provided in the Supporting Materials and Methods.

RESULTS AND DISCUSSION

Single-molecule properties of toxic DPRs

The dynamics of intrinsically disordered proteins (IDPs) is crucial for their function, including, for example, recognition and binding to target molecules (43–45). The dimension of IDPs can have a large effect on their functional properties. At a fixed temperature and solvent condition, the dimension of an IDP is determined by its amino acid sequence (43,46). The repetitive sequences of toxic DPRs contain only two types of amino acids (see Fig. 1 *a* and Video S2). Therefore, the overall properties of the DPRs strongly depend on the physiochemical properties of

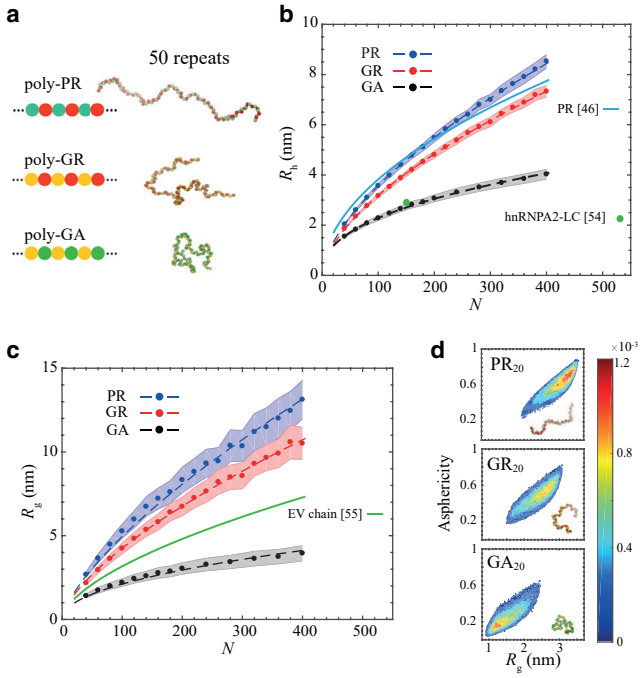

FIGURE 1 Comparison between the single-molecule properties of poly-PR, poly-GR, and poly-GA. (*a*) A schematic representation of the DPRs (*left*) and simulation snapshots (*right*); see also Video S2. (*b*) Hydrodynamic radius R_h of the DPRs plotted against their chain length N (*circles fitted with dashed lines*). The solid light-blue line is a prediction for the R_h of poly-PR based on a closed-form expression (46). Experimental R_h-value for the LC domain of hnRNPA2 is shown by a green circle. (*c*) Radius of gyration R_g of the DPRs plotted against their chain length N (*circles fitted with dashed lines*). R_g of an EV chain is provided for comparison. In (*b*) and (*c*), the shaded regions indicate half of the standard deviation. (*d*) Two-dimensional probability distribution of the asphericity and R_g for PR_{20}, GR_{20}, and GA_{20}. To see this figure in color, go online.

glycine, alanine, proline, and arginine and possibly the patterning of these residues. Alanine is hydrophobic, arginine is positively charged, and proline contributes to the rigidity and glycine to the flexibility of the protein backbone (47–49).

We use our 1BPA MD model (details are provided in Methods) to study the dimensions of the three toxic DPRs. In Fig. 1, *b* and *c*, we show the hydrodynamic radius, R_h, and the radius of gyration, R_g, of poly-PR, poly-GR, and poly-GA. The results are shown for a range of chain lengths N between 40 and 400, with N being two times the number of repeats, n.

For the same chain length, the dimension of poly-PR is larger than poly-GR, which is larger than poly-GA. The electrostatic repulsion between the uniformly distributed arginine residues results in more expanded conformations for the R-DPRs compared with the hydrophobic poly-GA. This observation is consistent with previous studies on the effect of charged residues and their patterning on the dimension of IDPs (50–52). The observed difference between poly-PR and poly-GR, however, cannot be explained by the nonbonded interactions alone. Proline is more hydrophobic than glycine (37,53), and thus poly-PR is expected to have a more compact structure than poly-GR. Therefore, the larger R_h- and R_g-values for poly-PR can only be attributed to the different contribution of proline and glycine to the backbone stiffness. To prove this, we repeated the simulations for poly-GR and poly-PR by interchanging the hydrophobicity values of P and G while keeping the bonded potentials the same. The results in Fig. S2 convincingly show that the difference between poly-GR and poly-PR in Fig. 1, *b* and *c* is almost entirely due to the difference in backbone stiffness. Indeed, proline is much stiffer than glycine because of the cyclic structure of its side chain (47–49), which is incorporated in the 1BPA force field (42). Our results are consistent with the observed correlation between proline content and extended conformation of IDPs (46,48).

In Fig. 1 *b*, we compare our simulation results for R_h of poly-PR with the Marsh and Forman-Kay fit to the experimental R_h-values of 36 IDPs (46). The suggested expression takes into account the proline content and the absolute net charge of the chain. The difference between our simulation results and the prediction by Marsh and Forman-Kay expression is less than 16%. The observed difference for longer chains could be due to the fact that the patterning of amino acids has not been considered as an input variable in the suggested expression (46). To show the importance of sequence patterning, for instance, the R_h of three variants of proline-arginine chains with the same amino acid composition but different patterning of proline and arginine residues are depicted in Fig. S3. These results show that the chain favors a conformation with the highest R_h when the proline and arginine residues are well mixed, as in poly-PR. Poly-GA forms the most compact conformation because of the uniform distribution of hydrophobic alanine residues and the low stiffness of the glycine residues. We also compare the simulated R_h of poly-GA with the experimental R_h of the disordered low-complexity (LC) domain of hnRNPA2 (54); see Fig. 1 *b*. The hnRNPA2 LC domain contains hydrophobic residues (mainly phenylalanine and tyrosine) distributed along the sequence. It has been suggested that the high glycine content (47%) of hnRNPA2 LC domain contributes to its compactness (54). With the same chain length, the hydrodynamic size of poly-GA is very similar to that of the hnRNPA2 LC domain.

Relating the R_g of the DPRs to the chain length N via $R_g \propto N^{\nu}$ (55) leads to scaling exponents of $\nu = 0.70 \pm 0.02$ for poly-PR, 0.67 ± 0.02 for poly-GR, and 0.48 ± 0.02 for poly-GA. Similar scaling exponents of around 0.70 have been obtained for extended variants of prothymosin α (ProTα) (with a mean net charge per residue of -0.46) in water (55). The scaling exponent of poly-GA is close to the value $\nu = 0.5$ expected for a random coil, i.e., a polymer in a θ solvent, and lies in the range 0.46 ± 0.05 obtained for the unfolded state of proteins in experiment (55). The scaling exponent of poly-GA is also comparable to an average scaling exponent of 0.53 obtained for a set of IDPs and unfolded proteins using an improved generation of atomistic force fields (56). The $R_g \propto N^{0.6}$ of an excluded volume (EV) chain is also plotted for comparison in Fig. 1 *c* (55). Poly-PR and poly-GR have an R_g that is larger than an EV chain because of the repulsion of like charges, whereas the flexible and more hydrophobic poly-GA has a lower R_g. In Fig. 1 *d*, we compare the asphericity, which measures the chain shape, versus R_g plots for DPRs with a repeat length $n = 20$. PR_{20} samples conformations with a larger R_g and asphericity than GR_{20} and GA_{20}, showing that poly-PR is more extended and assumes shapes closer to a rod-like conformation.

Phase separation of poly-GA

Poly-GA forms cytoplasmic aggregates (5,18,22,25) that are relatively stable in photobleaching experiments (22). Fourier transform infrared spectroscopy measurements show a random coil structure for GA_{15} right after incubation (21). After a certain incubation time, the GA_{15} molecules form aggregates as indicated by a change in the average particle size in the system (21). After a few hours, GA_{15} starts to form fibrils containing cross-β-sheet structures with disordered molecules of poly-GA still in solution (21,57).

Our results show that poly-GA undergoes a length-dependent phase separation to form a condensed (high-density) phase and a dilute (low-density) phase (Fig. 2). The condensed phases of poly-GA are spherical and exchange molecules with the surroundings (see *snapshots* in Figs. 2 *a* and S4; Video S1). Because hydrogen bonding is not included in the modeling (see the description of our coarse-grained model in the Supporting Materials and Methods), we are not able to predict the final transition into relatively stable aggregates or higher-order β-type

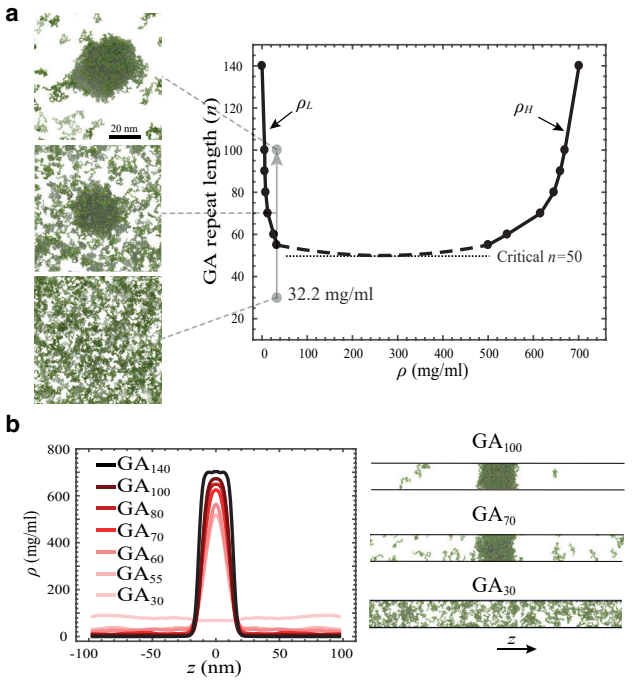

FIGURE 2 Poly-GA length-dependent phase-separation and coexistence phase diagram. (*a*) Coexistence phase diagram of poly-GA obtained from slab simulations. Phase separation occurs above the critical repeat length $n = 50$ inside the coexistence region. ρ_H and ρ_L are the concentrations of the condensed phase and dilute phase, respectively. Simulation snapshots at the left show the results for GA_{100}, GA_{70}, and GA_{30} for a total concentration of 32.2 mg/mL at equilibrium; see also Video S1. (*b*) Poly-GA slab density profiles (*left panel*) and sample slab simulation snapshots (*right panel*). To see this figure in color, go online.

structures of poly-GA as observed in the experiments (21,22,25). However, our results do suggest that long-range hydrophobic interactions drive the formation of high-density phases of poly-GA, which brings the residues close enough for short-range hydrogen bonds to form (58). Further transition of liquid condensates to more solid-like structures have been experimentally observed for FUS and hnRNAPA1 (59–61). To find the critical repeat length required for phase separation, we constructed the coexistence phase diagram of poly-GA (Fig. 2 *a*) by determining the concentrations of the condensed phase (ρ_H) and dilute phase (ρ_L) using the slab method (Fig. 2 *b*; (38,62,63)). In the phase diagram in Fig. 2 *a*, the vertical axis is the repeat length n and the horizontal axis is the concentration ρ. Outside the coexistence curve (i.e., for concentrations lower than ρ_L and higher than ρ_H and lengths shorter than the critical repeat length), there is only a uniform phase, whereas inside it, the poly-GA molecules phase separate (see the *snapshots* at the *left* in Fig. 2 *a*). The black arrows in Fig. 2 *a* show the concentrations of the two phases. The critical repeat length for phase separation of poly-GA is found to be $n = 50$. Below this critical repeat length, no phase separation is observed at any concentration. This critical repeat length is in good agreement with the critical range of $46 < n < 61$ found by Yamakawa et al. for the formation of aggregates of poly-GA in Neuro2a cells (5).

The values of ρ_H and ρ_L depend on the repeat length, with longer poly-GA molecules forming condensed phases with higher concentrations and dilute phases with lower concentrations (see Fig. 2 *a*). The same trend has been observed in previous simulation studies for polymers and IDPs with different lengths (38,62). Also, in experiments using synthetic molecules, the saturation concentration, which is equal to ρ_L at equilibrium, was observed to decrease with chain length (64). For $60 \leq n \leq 100$, we find the concentration range to be 6–33 mg/mL for the dilute phase and 500–670 mg/mL for the condensed phase. Using linear interpolation in Fig. 2 *a*, the molar concentration of the GA_{76} condensed phases is found to be ~34 mM, comparable to that of the 151-residue-long hnRNPA2 LC domain (30–40 mM) obtained in experiment (54). This is expected because the hnRNPA2 LC domain has a similar dimension as GA_{76} (see Fig. 1 *b*). We also observed a lower exchange rate with the surroundings for the condensed phases formed by longer chains (see Table S2), suggesting that longer molecules form more stable condensed phases.

Cluster size distribution analysis of poly-GA at equilibrium shows a strong correlation between the poly-GA total concentration and its ability to phase separate at small concentrations at the left of the phase diagram (Figs. 2 *a* and S5). When phase separation occurs, a further increase of the total concentration has no effect on the ρ_H and ρ_L but only increases the size of the condensed phase (Fig. S6). The cluster size distribution results of Fig. S5 also reveal (see Fig. S7) that at a fixed concentration, the number of free molecules N_{free} is higher for shorter dipeptides and that, as expected, a drop in N_{free} can be taken as an indication of phase separation.

Phase separation of R-DPRs

R-DPRs have been observed to form liquid droplets in the presence of RNA (33,34), phosphate ions (33,57), and several RNA-binding proteins (16,22). Previous studies have pointed at the important role of electrostatic and cation-π interactions in the LLPS of R-DPRs with multivalent proteins and RNA molecules for these cases (16,22,33,34). The phase separation of PR_{30} with different polyanions, known as complex coacervation, has recently been investigated using both in vitro experiments and coarse-grained dynamic Monte Carlo modeling (34).

Our simulations show no phase separation of poly-PR and poly-GR as a consequence of the electrostatic repulsion between arginine residues; see Fig. S8 (*top panels*). The same result has been obtained in in vitro experiments with PR_{20} in the presence of monovalent ions (33). The addition of anionic homopolymers (acidic molecules) of polyaspartate (poly-D) to the simulation box induces the phase separation of R-DPRs into liquid droplets (Fig. S8, *top panels*, and

Video S3). R-DPRs bind to acidic molecules, and upon binding, they become more compact (see Fig. 3 *a*). In Fig. S9 *a*, we compare the potential of mean force associated with binding of the R-DPRs, with repeat lengths $n = 30$, to an acidic molecule of length $N = 60$. Despite the higher hydrophobicity of P compared to G, the binding of poly-GR to the acidic molecule is stronger than poly-PR, and poly-GR makes a greater number of contacts with the acidic molecule than poly-PR; see Fig. S9 and Video S4. This difference is attributed to the difference in the backbone stiffness, as reflected by the persistence lengths in isolation, which we calculate to be 2.06 bond lengths for poly-GR and 5.04 bond lengths for poly-PR. This is confirmed by additional simulations in which we interchanged the hydrophobic values of P and G (see Fig. S9 *a*), showing that the backbone stiffness plays a dominant role in accounting for the different binding strengths of poly-PR and poly-GR.

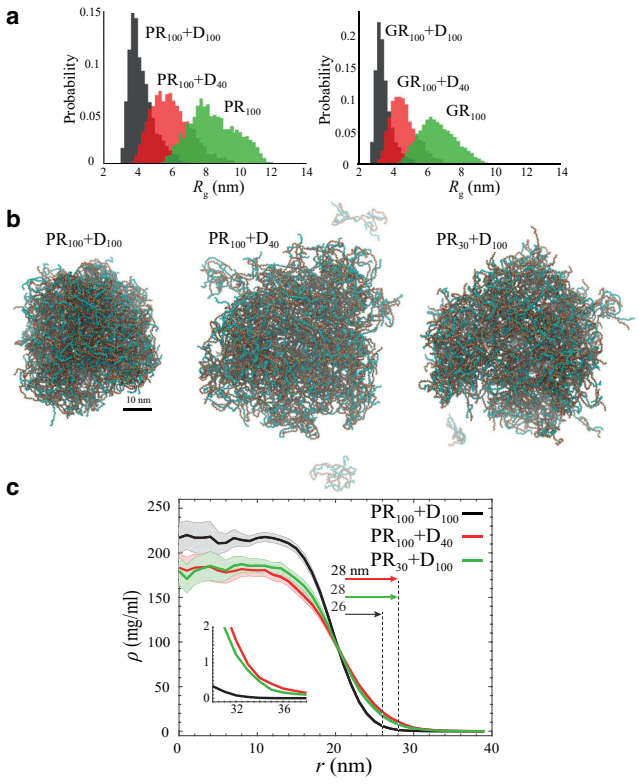

FIGURE 3 Binding of R-DPRs to acidic molecules (poly-D) and length-dependent droplet formation of poly-PR. (*a*) Histograms for R_g of single PR_{100} (*left*) and GR_{100} (*right*) in the presence and absence of a D_{40} and D_{100} molecule. Binding to acidic molecules reduces the R_g of R-DPRs, with longer acidic molecules having a larger effect. (*b*) Simulation snapshots for length-dependent droplet formation of poly-PR with acidic molecules for a total concentration of 14.8 mg/mL and a poly-PR concentration ratio of $r_{PR} = 0.57$. (*c*) Radial density profiles measured from the center of mass of the droplets presented in (*b*). Light shades indicate half of the standard deviation as error bars. The vertical dashed lines indicate the droplet size (for details of the droplet radius calculation, see the Supporting Materials and Methods). The inset shows the zoomed density profiles for $r >$ 30 nm. To see this figure in color, go online.

Long acidic tracts with length ranges of 12–41 amino acids can be found in the disordered regions of two nucleolar proteins: nucleolin ($NCL_{1–300}$) and $NPM1_{120–240}$; see the net charge per residue (NCPR) histograms in Fig. S10 (*top panels*). NCL mislocalization and disruption has been observed in ALS patient cells (65). NSR1, the yeast homolog of NCL, has also been observed to be a strong modifier of PR_{50} toxicity. The toxicity was shown to be suppressed by deletion and enhanced by upregulation of NSR1 (14). A recent study also suggested that poly-PR-induced NPM1 mislocalization generates toxicity (16). Our simulations show binding of the R-DPRs to both $NCL_{1–300}$ and $NPM1_{120–240}$ through interaction with the acidic tracts (Fig. S10, *bottom panels*). These results are consistent with the experimentally observed binding of full-length NPM1 and NCL to the R-DPRs (16,22).

Figs. 3, *b* and *c* and 4 show that the phase separation of poly-PR not only depends on its length but also on the length of the acidic molecules. In these simulations, we use different mass concentration ratios of poly-PR, r_{PR} = (poly-PR mass concentration)/(total mass concentration). With this definition, the poly-PR:poly-D mass ratio can be calculated as $1:(1 - r_{PR})/r_{PR}$. At a fixed total mass concentration and r_{PR}, longer repeats of poly-PR and poly-D phase separate to smaller droplets with a higher concentration surrounded by a dilute phase with a lower concentration (see Fig. 3 *c*). Reduction of the length of acidic molecules (Fig. 3 *b*, *middle panel*) or poly-PR (Fig. 3 *b*, *right panel*) increases the size of the droplet and the concentration of the dilute phase (see Fig. 3 *c*).

To further study the phase separation of poly-PR, we use slab simulations to obtain the coexistence phase diagrams in Fig. 4 *a*. The phase diagrams are constructed for acidic poly-D molecules of lengths $N = 40$ and 100 and three different poly-PR concentration ratios $r_{PR} = 0.57$, 0.62, and 0.67. Similar to the trend observed in Fig. 3, *b* and *c*, we observe that longer repeats form condensed phases with higher concentrations and dilute phases with lower concentrations, in agreement with experimentally observed length-dependent complex coacervation of oppositely charged polyelectrolytes (66). We also observe that the condensed phases of longer PR molecules have a smaller exchange rate with the surroundings; see Table S2. The poly-PR condensed phase concentration ranges from 90 to 211 mg/mL, which is much lower than the concentration range 500–670 mg/mL obtained for poly-GA for $n \leq 100$. As can be seen in the phase diagram, the critical poly-PR repeat length required for phase separation is lower for longer acidic molecules. The results in Fig. 4 *a* show an optimum concentration ratio of $r_{PR} = 0.62$ for phase separation of poly-PR with acidic molecules. At this concentration ratio, with fixed peptide lengths, poly-PR forms condensed phases with the highest concentration; see the inset of Fig. 4 *a*. In Fig. 4 *a*, we constructed the phase diagrams for two other concentration ratios of $r_{PR} = 0.57$ and 0.67 around the optimum value. For

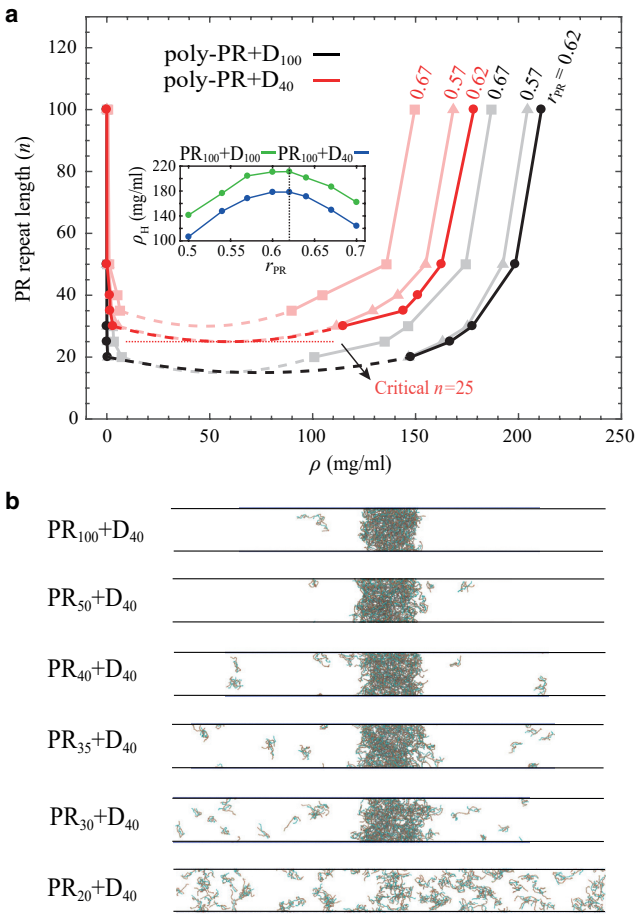

FIGURE 4 Length-dependent phase separation of poly-PR. (*a*) Phase diagrams for poly-PR phase separation with acidic molecules of lengths 40 and 100 for different poly-PR concentration ratios of $r_{PR} = 0.57$, 0.62, and 0.67; see the tilted numbers next to each phase diagram. Light shades are used for the phase diagrams constructed for $r_{PR} = 0.57$ and 0.67. The minimal critical repeat length for phase separation of poly-PR with D_{40} is indicated with a horizontal dotted line. The inset shows the concentration of the high-density phases of $PR_{100} + D_{40/100}$ for different poly-PR concentration ratios. (*b*) Slab simulation snapshots for length-dependent phase separation of poly-PR with D_{40} for a poly-PR concentration ratio of $r_{PR} = 0.62$ are given. No phase separation can be observed below a critical poly-PR repeat length $n = 25$. To see this figure in color, go online.

lower (<0.57) and higher (>0.67) r_{PR}-values, the concentration of the condensed phase is smaller, and the critical repeat length is larger. For $r_{PR} > 0.8$ and $r_{PR} < 0.3$, we observed no phase separation (data not shown). Our results are consistent with electrostatic-driven phase separation of R-DPRs with full-length NPM1, with phase separations that depend on concentration ratios of proteins and length of the R-DPRs (16,22). For D_{40}, which has almost the same length as the longest acidic tract in the nucleolar targets of R-DPRs (see Fig. S10), the critical repeat length of poly-PR required for phase separation is found to be $n = 25$ for the optimum concentration ratio $r_{PR} = 0.62$ (see the *phase diagram* in Fig. 4 *a* and the *slab simulation snapshots* in Fig. 4 *b*). Below this critical repeat length, poly-PR can bind to acidic molecules, but these small clusters are unable to phase separate (see Fig. 4 *b*). The slab density profiles used to obtain the phase diagram are presented in Fig. S11.

In Fig. S12, we compare the coexistence phase diagram of poly-PR and poly-GR with D_{100} obtained for the same concentration ratios of $r_{PR} = r_{GR} = 0.62$. At a fixed repeat length n, poly-GR forms condensed phases with higher concentrations than poly-PR. This observation can be attributed to the sizes of the dipeptides and their free energy of binding to acidic molecules. With the same length, poly-GR is more compact than poly-PR irrespective of the presence of acidic molecules (see Figs. 1 and 3 *a*). This, together with the stronger binding of poly-GR to an acidic molecule (see Fig. S9 *a*), explains the differences in the phase diagrams. These findings provide insight into the experimentally observed differences between poly-PR and poly-GR in binding to their targets. For example, previous experimental measurements have indicated greater association of NCL with poly-GR than poly-PR (22). Consistent with this observation, photobleaching experiments for a single nucleolus labeled by GFP-NCL have shown that poly-GR is more effective than poly-PR in reducing the mobile fraction and fluorescence recovery rate of GFP-NCL (22). Our results suggest that the more stable physical interaction of poly-GR with the nucleolar components can be attributed to the stronger binding of poly-GR to the acidic tracts inside the nucleolus. In another study, poly-GR has been shown to have a higher tendency to accommodate different ligands compared with poly-PR (67). Our results suggest that the more flexible conformation of poly-GR might have contributed to this observation.

CONCLUSIONS

In this study, we used a coarse-grained, 1BPA model to study the properties of toxic DPRs with the aim of providing mechanistic insights into the possible role of LLPS of DPRs in C9orf72 toxicity. We showed that poly-PR favors the most extended conformations among all three toxic DPRs because of the patterning of charged arginine residues and the contribution of proline residues to the backbone rigidity. Our findings revealed a stronger binding of poly-GR to acidic molecules compared with poly-PR that is due to the different backbone stiffnesses of these DPRs. We observed that longer DPRs form condensed phases with higher concentrations and with lower exchange rates to the surroundings. This observation suggests that the more toxic nature of longer chains (12,16) can be due to the more stable structure of the condensates formed by longer DPRs. For poly-GA, we found that increasing the concentration increases the propensity for phase separation at small concentrations and increases the size of the condensed phase at larger concentrations, both consistent with the expected trend for liquid condensates (68). Increasing the length of DPRs increases the number of possible interactions and results in

an increase in the multivalency of the system. Previous experiments and simulation studies have shown that a critical number of valences is required for the formation of biomolecular condensates (49,64,69–71) and that the propensity for phase separation increases with increasing multivalency of the system (38). These observations are in agreement with the phase diagrams presented in Figs. 2 *a* and 4 *a*. We observed an inverse correlation between the ρ_H and R_g of the DPRs. With the same chain lengths and the same concentration ratios of R-DPRs, the ρ_H of poly-PR is lower than poly-GR, which is lower than poly-GA, conceptually consistent with the observed correlation between the compactness of IDPs and their tendency to phase separate (72). A similar inverse correlation also exists between the ρ_H and the scaling exponent ν of DPRs for the same concentration ratios of R-DPRs and the same length of the acidic molecules.

Our results for poly-GA repeat lengths larger than 50 suggest that aggregate nucleation starts with liquid phase separation, as experimentally observed for several RNA-binding proteins (36,59). Care should be taken in comparing our poly-GA droplets with the insoluble aggregates observed in previous studies (21,22,25) because our 1BPA force field does not take into account secondary or higher-order structure formation of poly-GA. However, our coarse-grained MD model does capture the length-dependent phase separation of poly-GA (5). The critical repeat length of $n = 50$ for phase separation of poly-GA is in good agreement with the critical range of $46 < n < 61$ found in Neuro2a cells for the formation of insoluble aggregates of poly-GA (5). Therefore, our findings suggest that the LLPS of poly-GA precedes the formation of insoluble aggregates (21,22,25) and could be the first step in the poly-GA toxicity pathway. The longest acidic tract found in the nucleolar targets of R-DPRs is 41 (see Fig. S10). We found the minimal critical repeat length for the phase separation of the most toxic DPR, i.e., poly-PR, with similar-sized acidic chains D_{40} to be $n = 25$, which is between the expansion size in normal individuals (average around 20 repeats) and the lower bound of the toxicity threshold (30 repeats) found in patient cells (1,2,4). Therefore, our results indicate a substantial change in the poly-PR phase-separation behavior in the range 20–30 repeats, suggesting a likely connection between the phase separation of poly-PR and the toxicity threshold. Note that our results do not rule out a possible role of soluble DPRs in generating toxicity. Further investigation is needed regarding the interaction between DPRs (especially R-DPRs) of different lengths in the soluble phase and their targets such as RNA-binding proteins and transport components (22,67).

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2020.07.005.

AUTHOR CONTRIBUTIONS

H.J., E.v.d.G., and P.R.O. designed research. H.J. performed and analyzed research. H.J., E.v.d.G., and P.R.O. wrote the article.

ACKNOWLEDGMENTS

We would like to acknowledge funding support from the Netherlands Organization for Scientific Research (NWO), Building Blocks of Life programme. We thank Liesbeth Veenhoff for helpful discussions and Anton Jansen for assistance in calculating the hydrodynamic size of the proteins. We are also grateful for use of the computational cluster Peregrine of the University of Groningen.

SUPPORTING CITATIONS

References (73–82) appear in the Supporting Material.

REFERENCES

1. DeJesus-Hernandez, M., I. R. Mackenzie, …, R. Rademakers. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 72:245–256.
2. Renton, A. E., E. Majounie, …, B. J. Traynor; ITALSGEN Consortium. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 72:257–268.
3. Rohrer, J. D., A. M. Isaacs, …, J. D. Warren. 2015. C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol.* 14:291–301.
4. Van Mossevelde, S., J. van der Zee, …, C. Van Broeckhoven. 2017. Relationship between C9orf72 repeat size and clinical phenotype. *Curr. Opin. Genet. Dev.* 44:117–124.
5. Yamakawa, M., D. Ito, …, N. Suzuki. 2015. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum. Mol. Genet.* 24:1630–1645.
6. Balendra, R., and A. M. Isaacs. 2018. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.* 14:544–558.
7. Belzil, V. V., P. O. Bauer, …, L. Petrucelli. 2013. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol.* 126:895–905.
8. Taylor, J. P., R. H. Brown, Jr., and D. W. Cleveland. 2016. Decoding ALS: from genes to mechanism. *Nature.* 539:197–206.
9. Gendron, T. F., K. F. Bieniek, …, L. Petrucelli. 2013. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* 126:829–844.
10. Mori, K., S. M. Weng, …, D. Edbauer. 2013. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science.* 339:1335–1338.
11. Zu, T., Y. Liu, …, L. P. W. Ranum. 2013. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. USA.* 110:E4968–E4977.
12. Mizielinska, S., S. Grönke, …, A. M. Isaacs. 2014. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science.* 345:1192–1194.
13. Kwon, I., S. Xiang, …, S. L. McKnight. 2014. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science.* 345:1139–1145.
14. Jovičić, A., J. Mertens, …, A. D. Gitler. 2015. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* 18:1226–1229.

15. Shi, K. Y., E. Mori, …, S. L. McKnight. 2017. Toxic PR_n poly-dipeptides encoded by the *C9orf72* repeat expansion block nuclear import and export. *Proc. Natl. Acad. Sci. USA.* 114:E1111–E1117.

16. White, M. R., D. M. Mitrea, …, R. W. Kriwacki. 2019. C9orf72 poly(PR) dipeptide repeats disturb biomolecular phase separation and disrupt nucleolar function. *Mol. Cell.* 74:713–728.e6.

17. Zhang, Y. J., L. Guo, …, L. Petrucelli. 2019. Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science.* 363:eaav2606.

18. Zhang, Y. J., K. Jansen-West, …, L. Petrucelli. 2014. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol.* 128:505–524.

19. Kanekura, K., Y. Harada, …, M. Kuroda. 2018. Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. *Sci. Rep.* 8:12740.

20. Wen, X., W. Tan, …, D. Trotti. 2014. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron.* 84:1213–1225.

21. Chang, Y. J., U. S. Jeng, …, Y. R. Chen. 2016. The glycine-alanine dipeptide repeat from C9orf72 hexanucleotide expansions forms toxic amyloids possessing cell-to-cell transmission properties. *J. Biol. Chem.* 291:4903–4911.

22. Lee, K. H., P. Zhang, …, J. P. Taylor. 2016. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell.* 167:774–788.e17.

23. Freibaum, B. D., Y. Lu, …, J. P. Taylor. 2015. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature.* 525:129–133.

24. Schludi, M. H., S. May, …, D. Edbauer; German Consortium for Frontotemporal Lobar Degeneration; Bavarian Brain Banking Alliance. 2015. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol.* 130:537–555.

25. Guo, L., H. J. Kim, …, J. Shorter. 2018. Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell.* 173:677–692.e20.

26. May, S., D. Hornburg, …, D. Edbauer. 2014. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol.* 128:485–503.

27. Nihei, Y., K. Mori, …, C. Haass. 2020. Poly-glycine-alanine exacerbates C9orf72 repeat expansion-mediated DNA damage via sequestration of phosphorylated ATM and loss of nuclear hnRNPA3. *Acta Neuropathol.* 139:99–118, Published online October 23, 2019.

28. Zhang, Y. J., T. F. Gendron, …, L. Petrucelli. 2016. C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat. Neurosci.* 19:668–677.

29. Khosravi, B., H. Hartmann, …, D. Edbauer. 2017. Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. *Hum. Mol. Genet.* 26:790–800.

30. Lin, Y., E. Mori, …, S. L. McKnight. 2016. Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell.* 167:789–802.e12.

31. Zhang, Y. J., T. F. Gendron, …, L. Petrucelli. 2018. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat. Med.* 24:1136–1142.

32. Boeynaems, S., E. Bogaert, …, L. Van Den Bosch. 2016. Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci. Rep.* 6:20877.

33. Boeynaems, S., E. Bogaert, …, L. Van Den Bosch. 2017. Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell.* 65:1044–1055.e5.

34. Boeynaems, S., A. S. Holehouse, …, A. D. Gitler. 2019. Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. USA.* 116:7889–7898.

35. Bennion Callister, J., S. Ryan, …, S. M. Pickering-Brown. 2016. Modelling C9orf72 dipeptide repeat proteins of a physiologically relevant size. *Hum. Mol. Genet.* 25:5069–5082.

36. Lin, Y., D. S. W. Protter, …, R. Parker. 2015. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell.* 60:208–219.

37. Ghavami, A., L. M. Veenhoff, …, P. R. Onck. 2014. Probing the disordered domain of the nuclear pore complex through coarse-grained molecular dynamics simulations. *Biophys. J.* 107:1393–1402.

38. Dignon, G. L., W. Zheng, …, J. Mittal. 2018. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* 14:e1005941.

39. Dignon, G. L., W. Zheng, …, J. Mittal. 2018. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 115:9929–9934.

40. Ananth, A. N., A. Mishra, …, C. Dekker. 2018. Spatial structure of disordered proteins dictates conductance and selectivity in nuclear pore complex mimics. *eLife.* 7:e31510.

41. Ketterer, P., A. N. Ananth, …, C. Dekker. 2018. DNA origami scaffold for studying intrinsically disordered proteins of the nuclear pore complex. *Nat. Commun.* 9:902.

42. Ghavami, A., E. van der Giessen, and P. R. Onck. 2013. Coarse-grained potentials for local interactions in unfolded proteins. *J. Chem. Theory Comput.* 9:432–440.

43. Müller-Späth, S., A. Soranno, …, B. Schuler. 2010. From the Cover: charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:14609–14614.

44. Kriwacki, R. W., L. Hengst, …, P. E. Wright. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA.* 93:11504–11509.

45. Turjanski, A. G., J. S. Gutkind, …, G. Hummer. 2008. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* 4:e1000060.

46. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.

47. Cheng, S., M. Cetinkaya, and F. Gräter. 2010. How sequence determines elasticity of disordered proteins. *Biophys. J.* 99:3863–3869.

48. Martin, E. W., A. S. Holehouse, …, T. Mittag. 2016. Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J. Am. Chem. Soc.* 138:15323–15335.

49. Wang, J., J. M. Choi, …, A. A. Hyman. 2018. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell.* 174:688–699.e16.

50. Mao, A. H., S. L. Crick, …, R. V. Pappu. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:8183–8188.

51. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA.* 110:13392–13397.

52. Beveridge, R., L. G. Migas, …, P. E. Barran. 2019. Ion mobility mass spectrometry uncovers the impact of the patterning of oppositely charged residues on the conformational distributions of intrinsically disordered proteins. *J. Am. Chem. Soc.* 141:4908–4918.

53. Kapcha, L. H., and P. J. Rossky. 2014. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* 426:484–498.

54. Ryan, V. H., G. L. Dignon, …, N. L. Fawzi. 2018. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell.* 69:465–479.e7.

55. Hofmann, H., A. Soranno, …, B. Schuler. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:16155–16160.

56. Zerze, G. H., W. Zheng, …, J. Mittal. 2019. Evolution of all-atom protein force fields to improve local and global properties. *J. Phys. Chem. Lett.* 10:2227–2234.

57. Flores, B. N., M. E. Dulchavsky, …, M. I. Ivanova. 2016. Distinct c9orf72-associated dipeptide repeat structures correlate with neuronal toxicity. *PLoS One.* 11:e0165084.

58. Schmidt, H. B., and D. Görlich. 2016. Transport selectivity of nuclear pores, phase separation, and membraneless organelles. *Trends Biochem. Sci.* 41:46–61.

59. Patel, A., H. O. Lee, …, S. Alberti. 2015. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell.* 162:1066–1077.

60. Murray, D. T., M. Kato, …, R. Tycko. 2017. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell.* 171:615–627.e16.

61. Molliex, A., J. Temirov, …, J. P. Taylor. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* 163:123–133.

62. Silmore, K. S., M. P. Howard, and A. Z. Panagiotopoulos. 2017. Vapour–liquid phase equilibrium and surface tension of fully flexible Lennard–Jones chains. *Mol. Phys.* 115:320–327.

63. Das, S., A. N. Amin, …, H. S. Chan. 2018. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.* 20:28558–28574.

64. Li, P., S. Banjade, …, M. K. Rosen. 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature.* 483:336–340.

65. Haeusler, A. R., C. J. Donnelly, …, J. Wang. 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature.* 507:195–200.

66. Spruijt, E., A. H. Westphal, …, J. van der Gucht. 2010. Binodal compositions of polyelectrolyte complexes. *Macromolecules.* 43:6476–6484.

67. Hayes, L. R., L. Duan, …, J. D. Rothstein. 2020. *C9orf72* arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. *eLife.* 9:e51685.

68. Alberti, S., A. Gladfelter, and T. Mittag. 2019. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell.* 176:419–434.

69. Banani, S. F., A. M. Rice, …, M. K. Rosen. 2016. Compositional control of phase-separated cellular bodies. *Cell.* 166:651–663.

70. Pak, C. W., M. Kosno, …, M. K. Rosen. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell.* 63:72–85.

71. Harmon, T. S., A. S. Holehouse, …, R. V. Pappu. 2017. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife.* 6:e30294.

72. Lin, Y. H., and H. S. Chan. 2017. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys. J.* 112:2043–2046.

73. Yamada, J., J. L. Phillips, …, M. F. Rexach. 2010. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics.* 9:2205–2224.

74. Schuler, B., E. A. Lipman, …, W. A. Eaton. 2005. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 102:2754–2759.

75. Ohnishi, S., H. Kamikubo, …, D. Shortle. 2006. Conformational preference of polyglycine in solution to elongated structure. *J. Am. Chem. Soc.* 128:16338–16344.

76. Perry, S. L., L. Leon, …, M. Tirrell. 2015. Chirality-selected phase behaviour in ionic polypeptide complexes. *Nat. Commun.* 6:6052.

77. Overbeek, J. T., and M. J. Voorn. 1957. Phase separation in polyelectrolyte solutions; theory of complex coacervation. *J. Cell. Physiol. Suppl.* 49 (*Suppl 1*):7–22, discussion, 22–26.

78. Ou, Z., and M. Muthukumar. 2006. Entropy and enthalpy of polyelectrolyte complexation: Langevin dynamics simulations. *J. Chem. Phys.* 124:154902.

79. Lin, Y. H., J. P. Brady, …, H. S. Chan. 2017. Charge pattern matching as a 'fuzzy' mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* 19:115003.

80. Garcia de la Torre, J., S. Navarro, …, J. J. Lopez Cascales. 1994. HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys. J.* 67:530–531.

81. Ghavami, A., E. Van der Giessen, and P. R. Onck. 2018. Sol–gel transition in solutions of FG-Nups of the nuclear pore complex. *Extreme Mech. Lett.* 22:36–41.

82. Ghavami, A., E. van der Giessen, and P. R. Onck. 2016. Energetics of transport through the nuclear pore complex. *PLoS One.* 11:e0148876.

Biophysical Journal, Volume *119*

Supplemental Information

Phase Separation of Toxic Dipeptide Repeat Proteins Related to C9orf72 ALS/FTD

Hamidreza Jafarinia, Erik van der Giessen, and Patrick R. Onck

Table of Contents

Coarse-grained force field

Update of the 1BPA coarse-grained force field. The 1BPA implicit solvent force field has been used before to study intrinsically disordered FG-Nups and nucleocytoplasmic transport (1-3). The bonded interactions, i.e. bending and torsion potentials, in this force field are residue and sequence specific. This force field, interestingly, differentiates between the bending and torsion potentials of Glycine, Proline, and other residues. This feature is highly important since DPRs are rich in Proline and Glycine and it has been shown that these two residues contribute to the rigidity and flexibility of an IDP (4-7). Therefore, using the 1BPA model enables us to distinguish between the properties of poly-PR and poly-GR and to obtain more accurate results for poly-GA. The 1BPA also accounts for the hydrophobic/hydrophilic and electrostatic interactions between different amino acids, polarity of the solvent and screening of free ions. This force field uses the average of several residue-based hydrophobicity scales to describe the effective interactions between the amino acids. The hydrophobic interactions in this force field have been calibrated against the experimentally known R_{h} values of FG-Nup segments (8). For Proline and Glycine the interaction parameters have also been fine-tuned using the end-to-end distance and radius of gyration of poly-Proline (9) and poly-Glycine segments (10) which makes the 1BPA a proper choice for investigating the properties of DPRs.

The majority of FG-Nup segments, however, contain less than 0.6% of Arginine (R) (see the pink shaded band in Fig. S1). For the ones with more than 0.6% of R a correlation can be observed between the R content and the R_{h} error (see black dashed line in Fig. S1a) showing that there is still room for improving the R interaction with other residues in the 1BPA force field. To achieve this we further fine-tune the relative hydrophobic strength ε_i of R. The ε_i value is a residue-specific parameter that ranges between 0 and 1 and is close to 0 for hydrophilic polar residues. Since we also study the interaction of R-DPRs with acidic molecules, we recalibrate the ε_i values of all charged residues, i.e. RDEK. The aim is to obtain an updated 1BPA force field that is more accurate for studying the properties of R-DPRs and their interaction with negatively-charged molecules.

To update the 1BPA force field we slightly increase the ε_i values of charged residues to 0.005 (see Table S1), thus reducing the R_{h} error for all the six FG-Nup segments with R $\mathrm{content} > 0.6\%$ (see Fig. S1a). This choice of parameter for the relative hydrophobic strength of charged residues gives the best results in terms of the total average error and the minimum largest error in our calibration simulations with 16 FG-Nup segments presented here and originally used for the calibration of 1BPA (2). The total average and the largest errors are found to be 8.3% and 21.1% in the 1BPA force field, and 7.5% and 17.1% in the updated 1BPA force field. The correlation mentioned earlier for an R $\mathrm{content} > 0.6\%$ still exists in the updated 1BPA (red dashed line in Fig. S1a) which might be due to the absence of cation-π interactions between R and residues with aromatic rings in our force field. However, this has no effect on our simulation results since the DPRs and acidic molecules studied in this work contain no aromatic residues. A direct comparison between the two force fields is presented in Fig. S1b. At physiological

intracellular pH between 7 and 7.4, i.e more than three pH units away from the pK_a values of Arginine and Aspartic acid, we assume R, D, E, and K to be fully charged (11).

Complex coacervation of R-DPRs. The complex coacervation of polyelectrolytes is driven by a combination of enthalpic and entropic effects (12). Coulombic energy change and counterion release entropy are the main contributors to the free energy of complexation (13). In our single-molecule and phase separation simulations of R-DPRs with stretches of acidic amino acids, we account for the screening effect of ions, but similar to previous theoretical (14, 15) and coarse-grained models (16, 17) used to study complex coacervation, the effect of counterion condensation has not been considered in our modeling. Despite this limitation, our simulations capture the experimentally observed length-dependence of ρ_{L} (concentration of the dilute phase) and ρ_{H} (concentration of the condensed phase) for polyelectrolytes (18). Here we compare the effect of Coulomb energy change and counterion release entropy for the complexation of R-DPRs with acidic molecules by calculating the Coulomb strength parameter as suggested by Ou and Muthukumar (13). In their study the Coulomb strength parameter Γ has been defined as

$$\Gamma = \frac{l_{\mathrm{B}}}{l_0}.$$

Here $l_{\mathrm{B}} = e^2/4\pi\epsilon_0\epsilon_r k_{\mathrm{B}}T$ is the Bjerrum length, where e is the elementary charge, ϵ_0 is the vacuum permitivity, ϵ_r is the relative permitivity, k_{B} is the Boltzmann constant, T is the absolute temperature and l_0 is the charge separation distance along a polymer chain. For $\Gamma < 1$ the entropic term is negligible and the complexation is driven by the change in the Coulomb energy, but for $\Gamma > 1$ the entropic term starts to play a more important role, for $\Gamma < 1.5$ the electrostatic attraction between the oppositely-charged polyelectrolytes still predominantly drives the complexation, while the counterion release entropy only plays a subsidiary role. For $\Gamma > 1.5$, the contribution of the entropic term becomes more significant, until at $\Gamma = 2.5$ the complexation is completely driven by counterion release entropy (13).

The charge separation distance l_0 is 0.76 nm (two times the bond length) for R-DPRs, and 0.38 nm (the bond length) for acidic molecules. Using the average $l_0 = (0.76 + 0.38)/2 = 0.57$ and $l_{\mathrm{B}} \simeq 0.7$ nm for water at 300 K, the Coulomb interaction parameter is found to be $\Gamma = 1.23$ that lies in the range where the Coulomb energy change plays a more significant role than the counterion release entropy, supporting our assumption to neglect the effect of counterion condensation.

Simulations

Single-molecule simulations. Langevin dynamics simulations are performed at 300 K and physiological salt concentration of 150 mM in NVT ensembles with a time-step of 0.02 ps and a Langevin friction coefficient of 0.02 ps^{-1} using GROMACS version 2016. Each simulation is performed for 3 μs and the last 1 μs is used to obtain the ρ_{L} and R_{g}. The Hydro++ program (19) is used to obtain

the R_{h} values from the trajectories. The fits for R_{h} in Fig. 1b are according to bN^{ν}. For R_{g} we use the following equation for the fits in Fig. 1c (20):

$$R_{\mathrm{g}} = \sqrt{\frac{2l_{\mathrm{p}}^{*}b}{(2\nu+1)(2\nu+2)}}N^{\nu},$$

Where $b = 0.38\,\mathrm{nm}$ and $l_{\mathrm{p}}^{*} = 0.40 \pm 0.07\,\mathrm{nm}$. The errors represent the changes in the scaling exponents for $0.33 < l_{\mathrm{p}}^{*} < 0.47$.

The asphericity in Fig. 1d is calculated using the following equation(21):

$$\mathrm{Asphericity} = 1 - 3\left\langle\frac{\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)^2}\right\rangle$$

Where $\lambda_{1,2,3}$ are the eigenvalues of the gyration tensor. Asphericity is zero for a perfect sphere and is 1 for a perfect rod.

Droplet simulations and cluster size distribution analysis. Although the slab method can minimize the finite size effects and produce reliable values for the concentrations of the two phases, this method is not suitable for investigating the properties of the droplets such as the nucleation process, droplet size and interaction between droplets. Due to these limitations we also study the 3D droplet formation of DPRs. To perform droplet simulations the DPRs are placed in a cubic box of size 80 nm and then simulated for $\simeq 3\ \mu s$ which is sufficient to reach the equilibrium state (Figs. 2a left panels, 3b, S4-S8). At equilibrium the number of the residues inside the condensed phase (droplet) and the radial density profile measured from the center of the droplet (e.g. Fig. S6) are well-converged. In our test simulations we observed no significant effect of the initial distribution of the molecules on the properties of the resulting droplets at equilibrium. The radial density profiles (Figs. 3c, S6) and cluster size distribution plots (Figs. S4a, S5) are time-averaged for at least 1 μs at equilibrium. Discrete cells of thickness 1 nm are used to obtain the radial density profiles. From the radial density profiles, one can obtain values of ρ_{H} from the flat region close to the center of the droplet, and ρ_{L} from the region far from the droplet. However, to minimize the finite-size effects (22), we use slab method (see the next subsection 'Slab simulations') to obtain ρ_{H} and ρ_{L}. The droplet radius is obtained from the radial density profile and is the average position of the first point close to the low-density region with $|d\rho/dr| \leq\ 2-5\ \mathrm{mg/ml/nm}$ (see dashed lines in Figs. 3c, S6). This term is the absolute value of the slope of the density profile which is close to zero where the density profile reaches the dilute phase region. Note that using $|d\rho/dr| \simeq 0\ \mathrm{mg/ml/nm}$ gives unrealistically large values for the droplet radius. We performed a sensitivity analysis for the critical value of $|d\rho/dr|$ and observed that selecting slope limits between 2-5 mg/ml/nm results in a maximal change of 4% for all computed values for the droplet radius.

To generate the cluster size distribution plots in Figs. S4 and S5, two chains are considered to be in the same cluster if at least two residues of those chains come closer than 0.7 nm (23). In our cluster size distribution plots (Fig. S4a and S5) the horizontal axis is the logarithm of the number of residues inside a cluster (S) and the vertical axis is the logarithm of the time-averaged number of the clusters (N_c). When phase separation occurs the curves are divided into two regions, a dilute phase containing free molecules and small clusters, and a condensed phase that exchanges molecules with the dilute phase.

Slab simulations. The initial simulations are performed in cubic boxes of size 20 nm for poly-GA and 25 nm for more extended R-DPRs. These box sizes ensure no interaction between DPRs and their periodic images for repeat lengths $n \leq 100$. For longer DPRs the box size is increased accordingly. For the initial equilibration simulations we followed the steps suggested in (24). After equilibration, the box is enlarged to a 10 times larger size than its initial value in the z direction which is sufficient to reduce the finite-size effects and to obtain reliable values for ρ_{H} and ρ_{L}. The system is then simulated for $\simeq$ 3 μs in an NVT ensemble to achieve convergence for the density profiles in the z direction (Figs. 2b, S11, S12). The density profiles are calculated using discrete cells of thickness 1 nm and time-averaged for at least 1 μs at equilibrium. When the system undergoes phase separation, the averaged concentrations in $|z| < 4$ nm region is used to obtain ρ_{H}. To obtain ρ_{L} we use the average concentrations in $|z| > 40$ nm for poly-GA (Fig. 2b) and $|z| > 65$ nm for R-DPRs (Figs. S11, S12). The simulation parameters are similar to the ones we used in the single-molecule and droplet simulations. From the slab simulations we also obtain the time-averaged exchange rates based on the fluctuations of the number of polymer units inside the condensed phases at equilibrium for around 1 μs (Table S1). To obtain the number of the molecules inside the condensed phases, we use cut-offs of 0.7 (23) and 0.9 nm for poly-GA and poly-PR+poly-D. We use a larger cut-off value for the second case since the equilibrium distances between the oppositely-charged residues, determined by both electrostatic and hydrophilic interactions, are almost 0.2 nm larger than the one for non-charged residues in our coarse-grained force field (2). To obtain the exchange rates per unit area, we divide the exchange rate by 2 times the xy-plane area of the slab box.

Phase diagrams. The vertical axis in our phase diagram is the DPR repeat length and the horizontal axis is the concentration (Figs. 2a, 4a, S12). The phase diagram is obtained by connecting values of ρ_{L} and ρ_{H} computed from the slab simulations. To find the critical point, we begin with the smallest repeat length n_1 that has produced converged ρ_{L} and ρ_{H} values. Then we perform slab simulation for repeat length of $n_{-1} = n_1 - 10$ to calculate the time-averaged density profile for 1 μs after 3 μs of simulation time. If the calculated density profile for repeat length n_{-1} is almost flat, with small fluctuations in concentration $\Delta\rho = |\rho_{\max} - \rho_{\min}| < 20$ mg/ml, we report $(n_1 + n_{-1})/2$ as the critical repeat length. With this method we estimate the critical repeat length with an error of less than 5 repeats. Choosing any value larger than 20 mg/ml for $\Delta\rho$ does not change the critical repeat length. In each phase diagram,

for the region close to the critical point, a dashed spline that reaches its minimum at $(n_1 + n_{-1})/2$, $(\rho_{\mathrm{H}}(n_1) + \rho_{\mathrm{L}}(n_1))/2$ is shown as a guide to the eye (Figs. 2a, 4a, S12).

Potential of mean force (PMF) calculation. We use umbrella sampling simulations and the weighted histogram analysis method (WHAM) via the gmx wham utility of GROMACS to calculate the PMF associated with the binding of R-DPRs to the acidic molecules in Fig. S9. The distance r between the center of masses of two molecules is considered as the reaction coordinate. For each window the simulation is conducted for 1 μs. We use $\Delta r = 0.1$ nm for the distance between two windows. For details about the umbrella sampling method the reader is referred to (3, 25).

Protein sequences

Sequences of the disordered parts of NPM1 and NCL used in Fig. S10 are listed below. The negatively charged residues are shown in blue.

$NPM1_{120\text{-}240}$	EEDAESEDEEEEDVKLLSISGKRSAPGGGSKVP QKKVKLAADEDDDDDDEEDDDEDDDDDDFD DEEAEEKAPVKKSIRDTPAKNAQKSNQNGKDS KPSSTPRSKGQESFKKQEKTPKTPKG
$NCL_{1\text{-}300}$	MVKLAKAGKNQGDPKKMAPPPKEVEEDSEDE EMSEDEEDDSSGEEVVIPQKKGKKAAATSAKK VVVSPTKKVAVATPAKKAAVTPGKKAAATPA KKTVTPAKAVTTPGKKGATPGKALVATPGKK GAAIPAKGAKNGKNAKKEDSDEEEDDDSEEDE EDDEDEDEDEDEIEPAAMKAAAAAPASEDEDD EDDEDDEDDDDDEEDDSEEEAMETTPAKGKK AAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE DDEDDDDEDDEEEEEEEEEEPVKEAPGKRKKE MAKQKAAPEAKKQKVEG

Supplementary figures

Figure S1: Coarse-grained force field

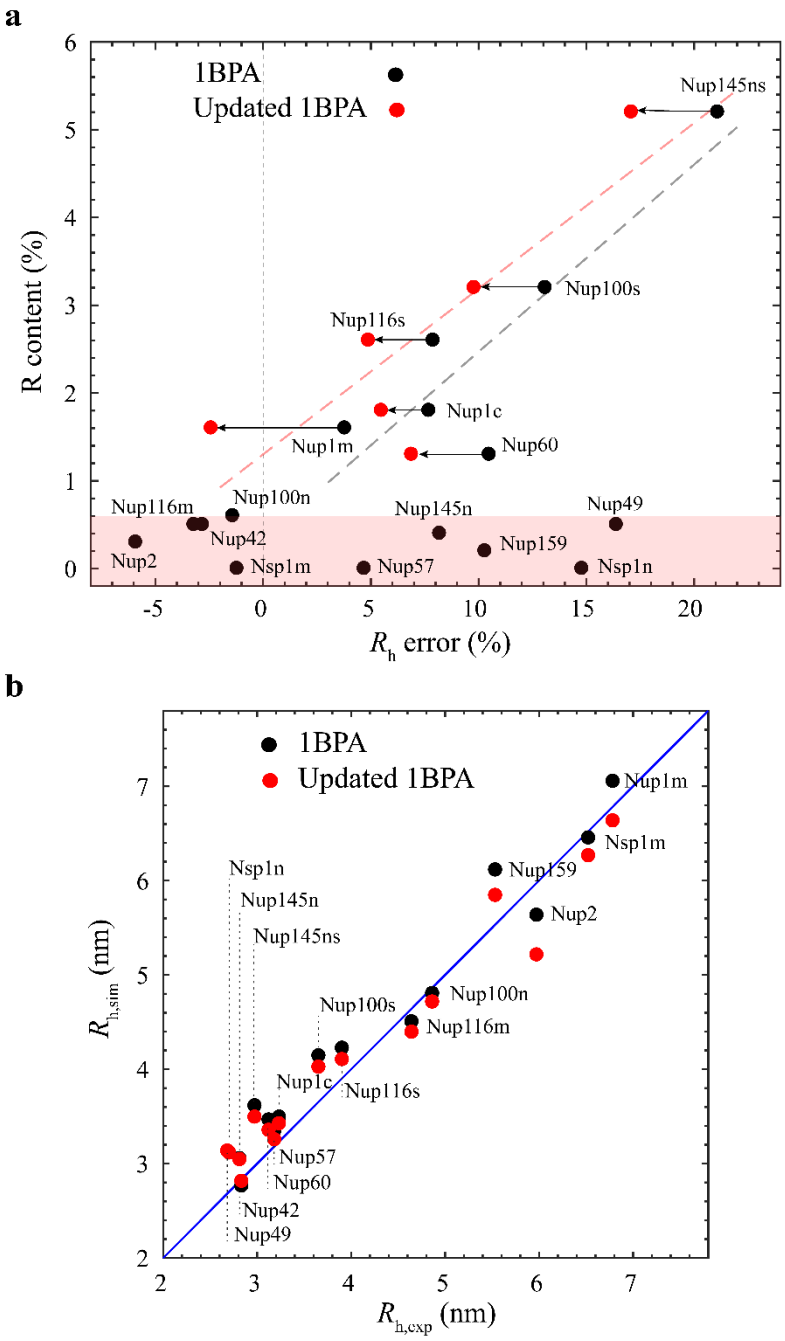

Figure S1: (a) Content of Arginine in FG-Nup segments plotted against their corresponding hydrodynamics radius error: $(R_{\mathrm{h,sim}} - R_{\mathrm{h,exp}})/R_{\mathrm{h,exp}}$ for 1BPA and the updated 1BPA force fields. The pink shaded band contains FG-Nups with R content $< 0.6\%$. The black dashed line shows the correlation between the R content $> 0.6\%$ and the R_{h} error in 1BPA. In the updated 1BPA force field the R_{h} error is reduced for all FG-Nups with R content $> 0.6\%$. The red dashed line shows the correlation between the R content $> 0.6\%$ and the R_{h} error in updated 1BPA. **(b)** A direct comparison

of the two force fields in predicting the hydrodynamic radius of FG-Nups. The total average and the largest errors are found to be 8.3% and 21.1% in the 1BPA force field, and 7.5% and 17.1% in the updated 1BPA force field.

Figure S2: Contribution of the backbone potentials to the size of the R-DPRs

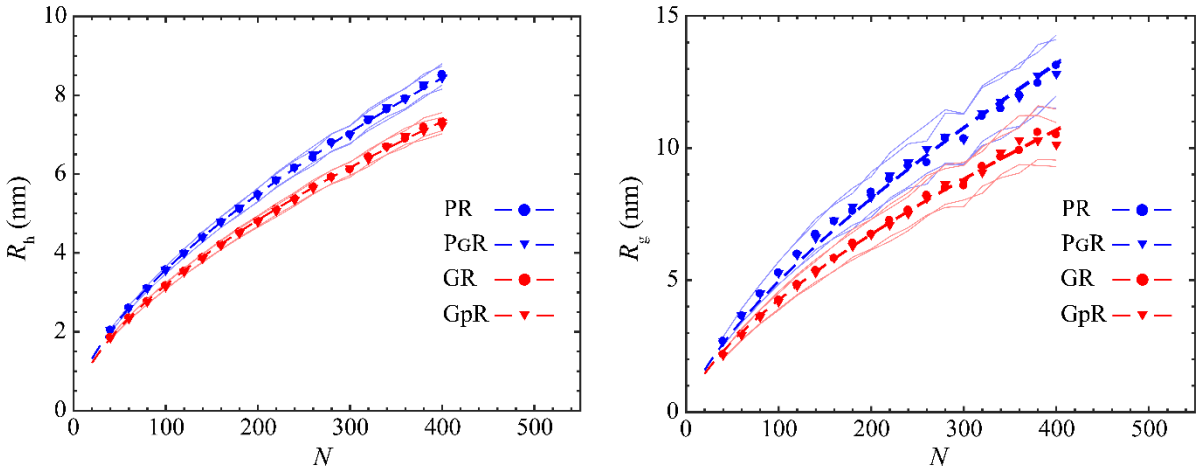

Figure S2: A comparison between the R_h and R_g of poly-PR and -GR (presented in Fig. 1b and 1c) and poly-P_GR and -G_PR. P_G has the backbone rigidity of P and the hydrophobicity of G, G_P has the backbone flexibility of G and the hydrophobicity of P. The error regions between the thin lines indicate half of the standard deviation.

Figure S3: Comparison of the hydrodynamic radius of Proline-Arginine chains

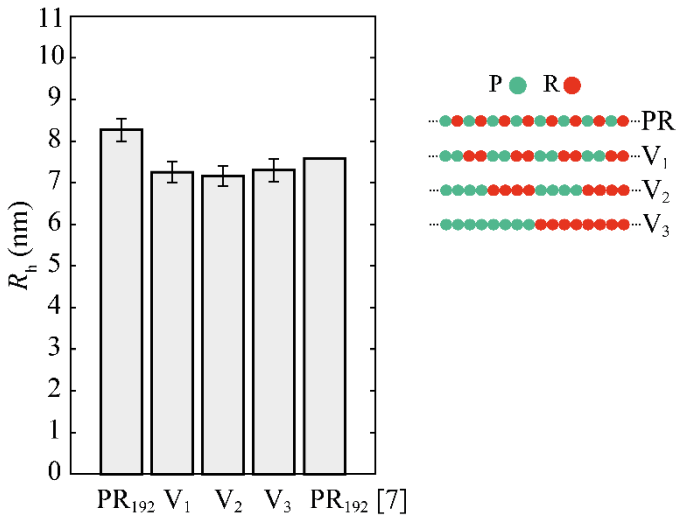

Figure S3: The R_h value of poly-PR and three variants with different patterning of the Proline and Arginine residues. The bar chart at the right has been obtained using the equation reported in (7).

Figure S4: Phase separation of poly-GA

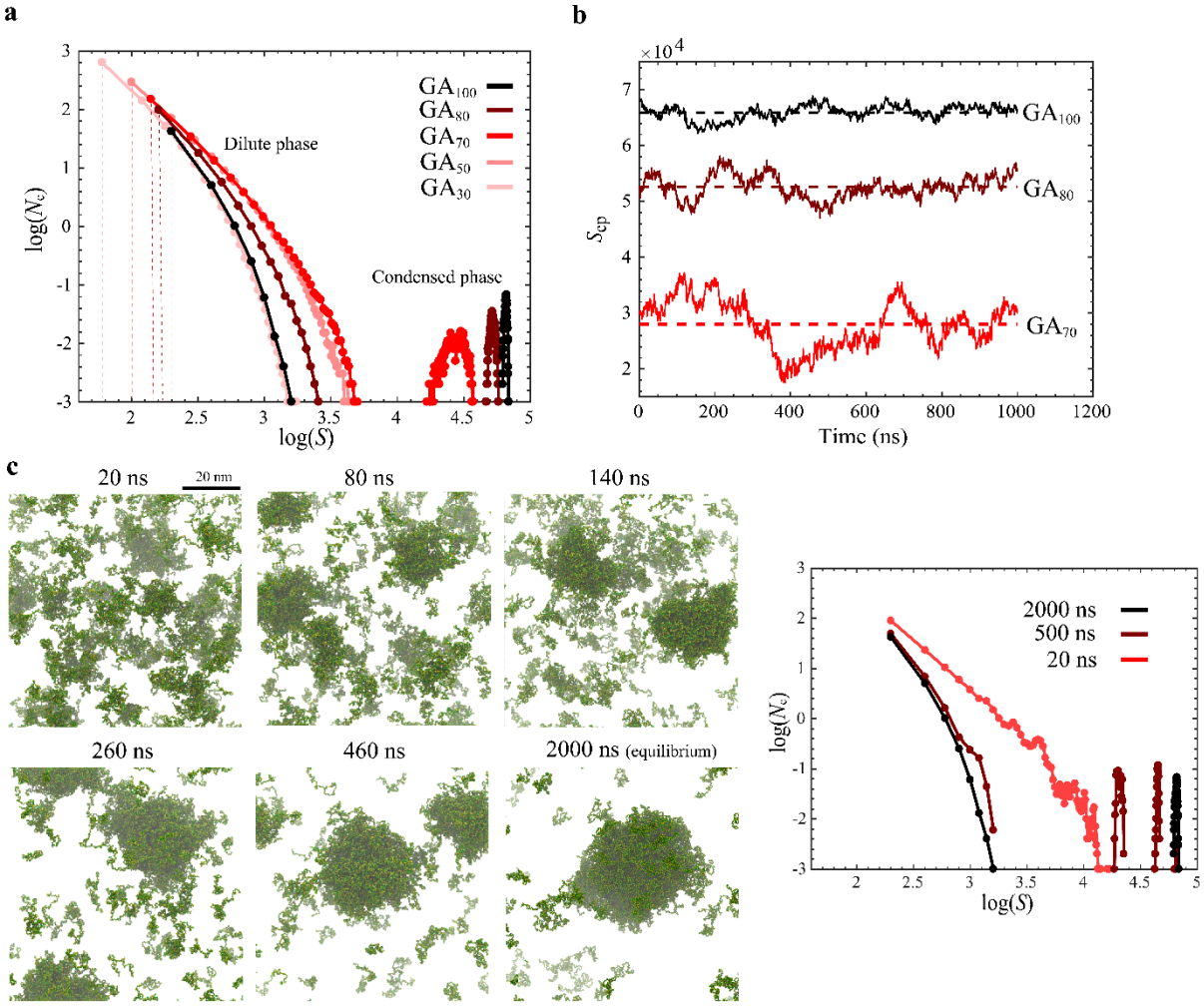

Figure S4: Phase separation of poly-GA for a total concentration of 32.2 mg/ml. **(a)** Cluster size distribution of poly-GA $30 \leq n \leq 100$ at equilibrium. S is the number of residues inside a cluster and N_c is the time-averaged number of the clusters. For comparison, the time-averaged number of free molecules is also included in this plot, indicated with a dashed line for each case. **(b)** The number of residues in the condensed phase (S_{cp}) plotted against time at equilibrium for GA_{100}, GA_{80}, and GA_{70}. Horizontal dashed lines indicate the average values. Longer dipeptides form clusters containing more residues with lower exchange with the surrounding. **(c)** Time evolution for the phase separation of GA_{100} starting from randomly distributed molecules in a cubic box of size 80 nm (left). A similar cluster size distribution analysis as presented in (a) at three different simulation times for GA_{100}. For each data set the average of the last 20 ns is used.

Figure S5: Length- and concentration-dependent phase separation of poly-GA

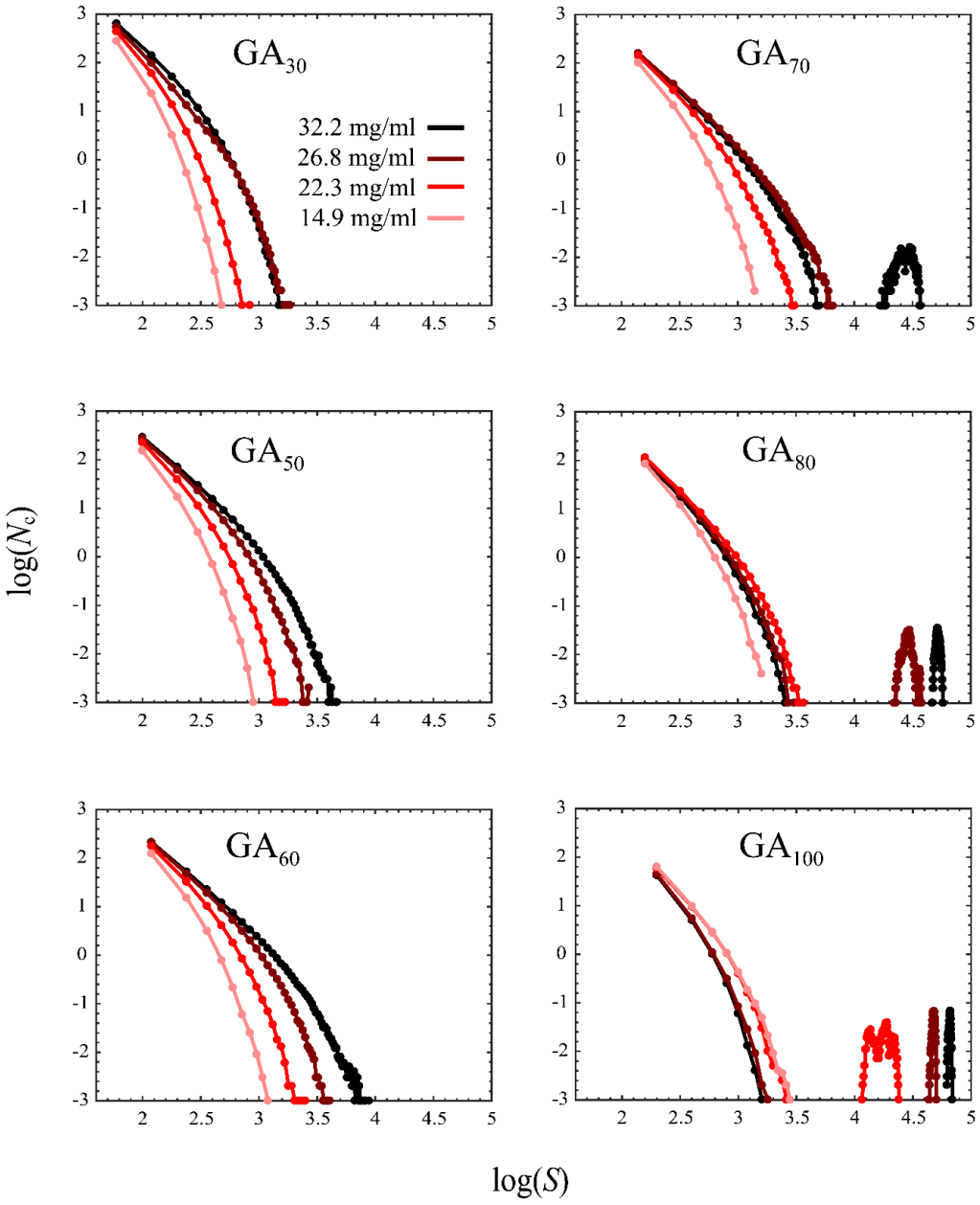

Figure S5: Cluster size distribution analysis of poly-GA at equilibrium for four different total mass concentrations of 32.2, 26.8, 22.3, and 14.9 mg/ml shows a length- and concentration-dependent phase separation. S is the number of residues inside a cluster and N_c is the time-averaged number of the clusters. For a fixed repeat length, increasing the concentration increases the average number of the molecules inside the condensed phase (see the results for GA_{80} and GA_{100}).

Figure S6: Radial density profiles for GA_{100} droplets

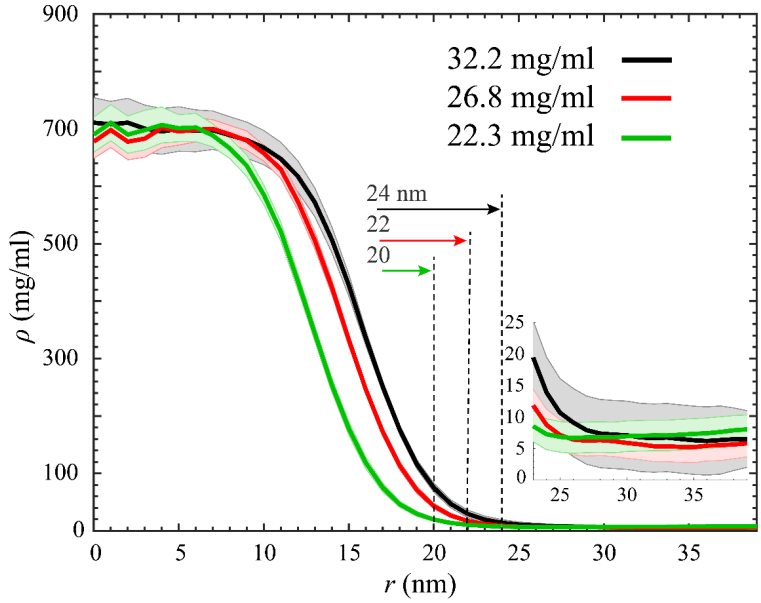

Figure S6: Radial density profiles for GA_{100} droplets for three different total mass concentrations. The shading indicates half of the standard deviations as error bars. The radius of the droplet is shown with dashed lines for each case. The inset figure shows the zoomed density profiles for $r \geq 24$ nm. The size of the droplet increases with increasing the total mass concentration. However, ρ_H (the average concentration for $r < 4$ nm) and ρ_L (the average concentration for $r > 30$ nm) remain unchanged.

Figure S7: Number of free molecules of poly-GA at equilibrium

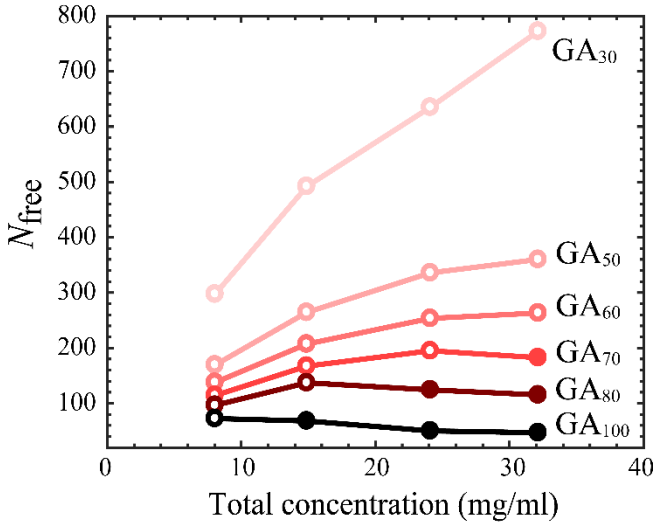

Figure S7: Number of free molecules N_{free} in the dilute phase plotted against the total concentration for different lengths of poly-GA. Filled markers are used when poly-GA molecules undergo phase separation. When phase separation occurs the N_{free} drops. At a fixed concentration, N_{free} is higher for shorter dipeptides. The data in the figure is obtained from the cluster size distribution curves of Fig. S5.

Figure S8: R-DPRs phase separation in the presence of poly-D

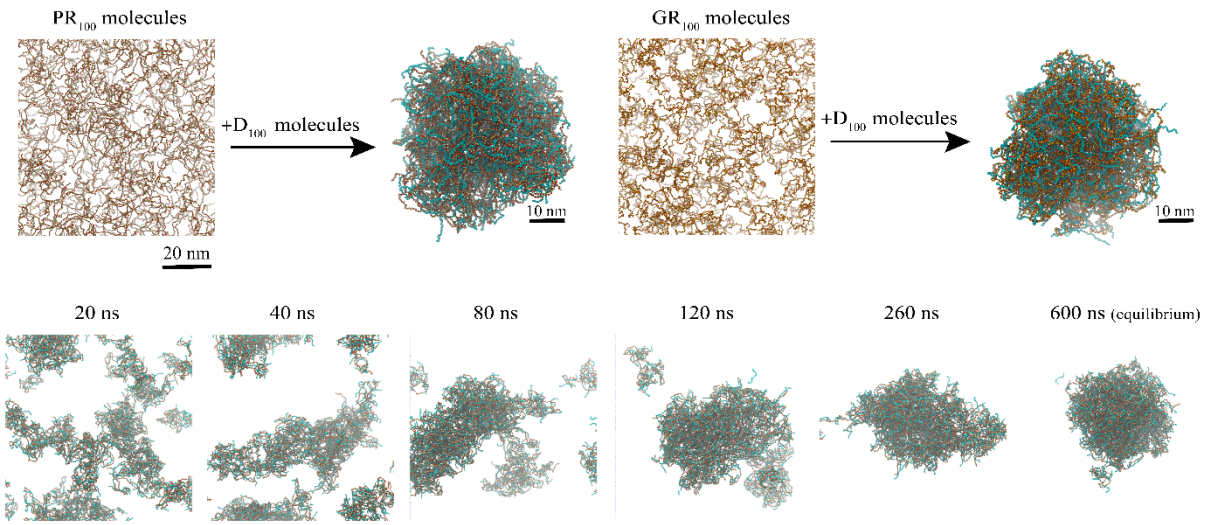

Figure S8: Long chains of poly-PR and poly-GR are not capable of forming clusters. Adding acidic molecules (poly-D) induces the phase separation of R-DPRs (top). Time evolution for the phase separation of $PR_{100} + D_{40}$ starting from randomly distributed molecules (bottom). The concentration ratios and the total concentration are $r_{PR} = r_{GR} = 0.57$ and 14.8 mg/ml for all cases.

Figure S9: PMF calculation for binding of R-DPRs to poly-D

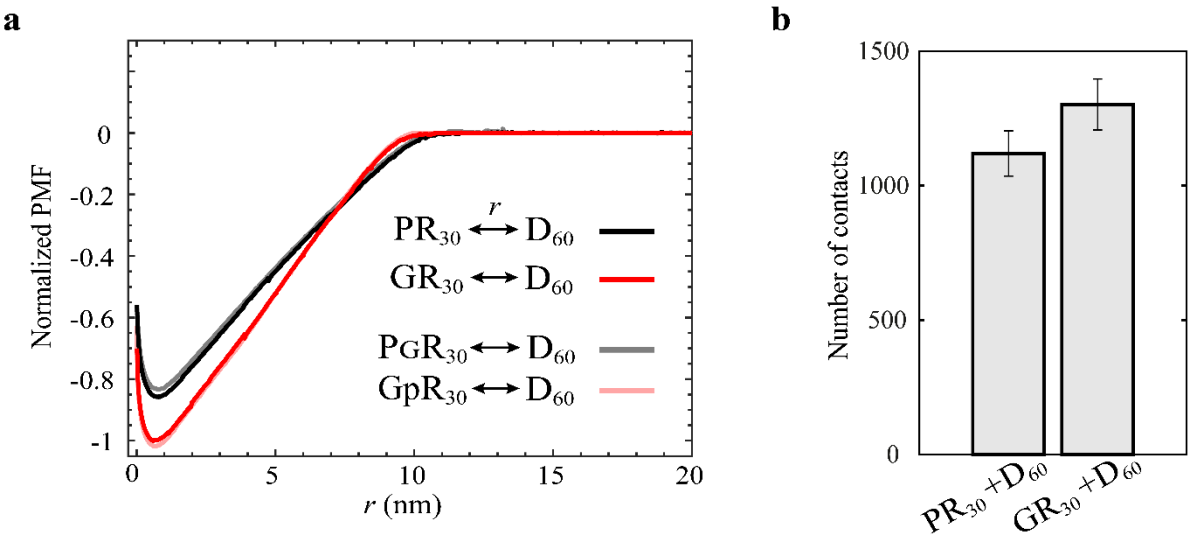

Figure S9: (a) PMF curves for binding of PR_{30} and GR_{30} to D_{60} indicates a larger free energy of binding of poly-GR to acidic molecules. The distance between the center of masses of two molecules is indicated with r. Due to the more compact conformation of poly-GR, its PMF curve vanishes at slightly shorter inter-molecule distances. Results for P_GR_{30} and G_PR_{30} binding to D_{60} show no significant change in the PMF curves. Note that P_G has the backbone rigidity of P and the hydrophobicity of G and G_P has the backbone flexibility of G and the hydrophobicity of P. All curves are normalized with the depth of the GR_{30}–D_{60} binding free energy. **(b)** Time-averaged total number of contacts after binding of PR_{30} and GR_{30} to D_{60} at equilibrium using a cut-off of 2.5 nm. The contact of one residue in the R-DPRs with one residue in the acidic molecule is counted as one contact, see Movie S4 for short trajectories.

Figure S10: NCPR plots and poly-PR binding to the disordered regions of NCL and NPM1

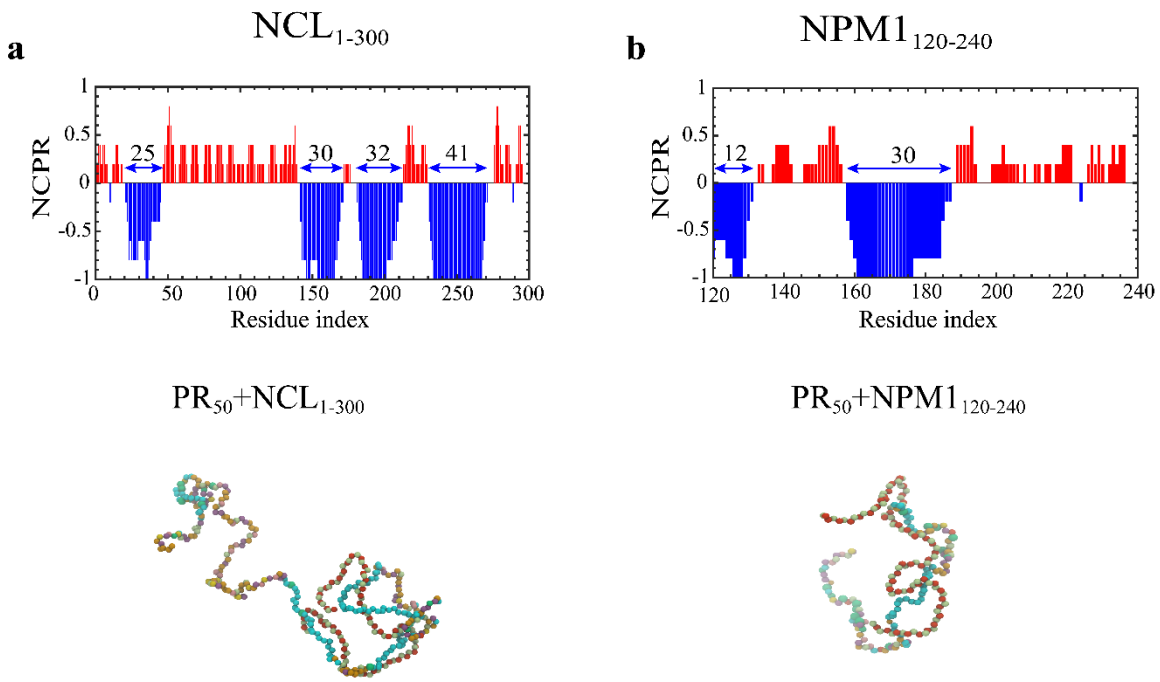

Figure S10: Top: Net charge per residue (NCPR) histograms for disordered parts of two nucleolar proteins: **(a)** nucleolin ($NCL_{1\text{-}300}$) and **(b)** nucleophasmin ($NPM1_{120\text{-}240}$). Blue arrows show acidic tracts with lengths ranging from 12 to 41. To find NCPR we use a sliding window containing 5 residues. Bottom: snapshots of binding of PR50 to $NCL_{1\text{-}300}$ and $NPM1_{120\text{-}240}$. Acidic tracts are indicated in cyan; PR chains comprise of red-green colored beads (as in Fig. 1a). The other aminoacids are given a range of colors according to their aminoacid type.

Figure S11: Slab density profiles for phase separation of poly-PR with poly-D

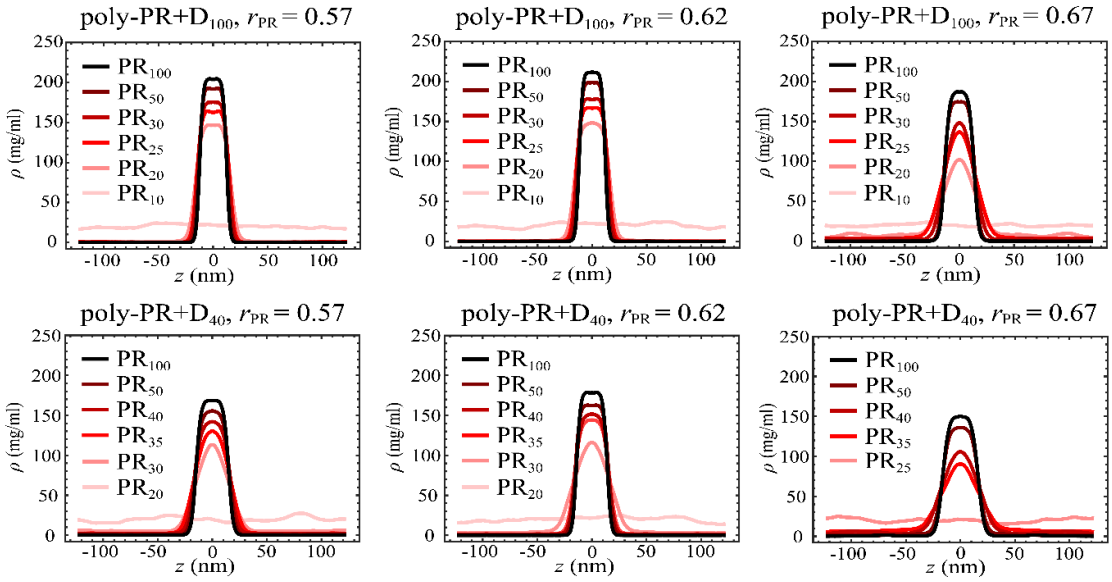

Figure S11: Slab density profiles for phase separation of poly-PR with acidic molecules of lengths 40 and 100 for three different concentration ratios of $r_{\mathrm{PR}} = 0.57, 0.62$ and 0.67. These density profiles are used to obtain the coexistence phase diagrams in Fig. 4a.

Figure S12: Comparison of the phase diagrams of poly-PR and poly-GR

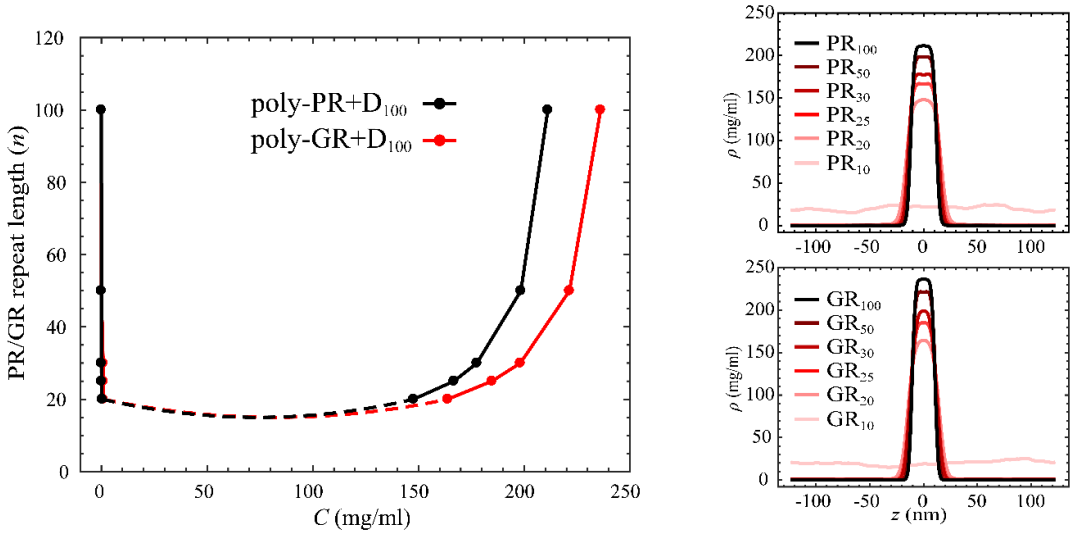

Figure S12: Coexistence phase diagram (left) and the corresponding slab density profiles (right) for phase separation of poly-PR and poly-GR with D_{100} for $r_{PR} = r_{GR} = 0.62$. With the same repeat lengths, poly-GR forms condensed phases with higher concentrations.

Supplementary tables

Table S1: The relative hydrophobic strength values of charged residues

Amino acid	R	D	E	K
ε_i (1BPA)	0	0.0005	0.0005	0.0005
ε_i (updated 1BPA)	0.005	0.005	0.005	0.005

Table S1: The relative hydrophobic strength ε_i values of charged residues in the 1BPA force field (1, 2) and updated 1BPA force field (the current paper).

Table S2: Time-averaged exchange rates

Poly-GA repeat length	Exchange rate (ns^{-1})	Exchange rate per unit area $\times 10^4$ ($ns^{-1}nm^{-2}$)	Poly-PR repeat length	Exchange rate (ns^{-1})	Exchange rate per unit area $\times 10^4$ ($ns^{-1}nm^{-2}$)
55	4.01	50.19	30	4.71	37.72
60	3.33	41.69	35	2.38	19.04
70	1.63	20.32	40	1.56	12.47
80	0.94	11.76	50	0.43	3.41
100	0.37	4.63	100	0.02	0.14

Table S2: The exchange rates between the condensed and dilute phases of poly-GA and poly-PR with different repeat lengths using the slab simulations presented in Fig. 2 and Fig. 4. The details are provided in the section 'Simulations' of the supplementary information. For poly-PR the exchange rates are reported for poly-PR phase separation with D_{40} and $r_{\mathrm{PR}} = 0.62$.

Supplementary movies

Movie S1: Phase separation of GA_{70} (left), GA_{100} (middle) and GA_{140} (right). For better visualization, in each simulation, 10 random molecules of poly-GA are indicated in red.

Movie S2: Single-molecule simulation of PR_{20} (left), GR_{20} (middle), and GA_{20} (right).

Movie S3: Fusion of two liquid droplets formed by PR_{100} and D_{40} with a poly-PR concentration ratio of $r_{\mathrm{PR}} = 0.57$.

Movie S4: Binding of PR_{30} (left) and GR_{30} (right) to D_{60}.

Supporting references

1. Ghavami, A., E. van der Giessen, and P. R. Onck. 2013. Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. J. Chem. Theory Comput. 9(1):432-440.

2. Ghavami, A., L. M. Veenhoff, E. Van Der Giessen, and P. R. Onck. 2014. Probing the disordered domain of the nuclear pore complex through coarse-grained molecular dynamics simulations. Biophys. J. 107(6):1393-1402.

3. Ananth, A. N., A. Mishra, S. Frey, A. Dwarkasing, R. Versloot, E. van der Giessen, D. Görlich, P. Onck, and C. Dekker. 2018. Spatial structure of disordered proteins dictates conductance and selectivity in nuclear pore complex mimics. eLife 7:1-24.

4. Cheng, S., M. Cetinkaya, and F. Gräter. 2010. How sequence determines elasticity of disordered proteins. Biophys. J. 99(12):3863-3869.

5. Martin, E. W., A. S. Holehouse, C. R. Grace, A. Hughes, R. V. Pappu, and T. Mittag. 2016. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. J. Am. Chem. Soc. 138(47):15323-15335.

6. Wang, J., J. M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, and A. A. Hyman. 2018. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell 174(3):688-699.e616.

7. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. Biophys. J. 98(10):2383-2390.

8. Yamada, J., J. L. Phillips, S. Patel, G. Goldfien, A. Calestagne-Morelli, H. Huang, R. Reza, J. Acheson, V. V. Krishnan, S. Newsam, A. Gopinathan, E. Y. Lau, M. E. Colvin, V. N. Uversky, and M. F. Rexacha. 2010. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. Mol. Cell. Proteomics 9(10):2205-2224.

9. Schuler, B., E. A. Lipman, P. J. Steinbach, M. Kumkell, and W. A. Eaton. 2005. Polyproline and the "spectroscopic ruler" revisited with single-molecule flourescence. Proc. Natl. Acad. Sci. U.S.A. 102(8):2754-2759.

10. Ohnishi, S., H. Kamikubo, M. Onitsuka, M. Kataoka, and D. Shortle. 2006. Conformational Preference of Polyglycine in Solution to Elongated Structure. J. Am. Chem. Soc. 128(50):16338-16344.

11. Perry, S. L., L. Leon, K. Q. Hoffmann, M. J. Kade, D. Priftis, K. A. Black, D. Wong, R. A. Klein, C. F. Pierce, K. O. Margossian, J. K. Whitmer, J. Qin, J. J. De Pablo, and M. Tirrell. 2015. Chirality-selected phase behaviour in ionic polypeptide complexes. Nat. Commun. 6(1):1-8.

12. Overbeek, J. T. G., and M. J. Voorn. 1957. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. J. Cell. Comp. Physiol. 49(S1):7-26.

13. Ou, Z., and M. Muthukumar. 2006. Entropy and enthalpy of polyelectrolyte complexation: Langevin dynamics simulations. J. Chem. Phys. 124(15):154902-154902.

14. Lin, Y. H., J. P. Brady, J. D. Forman-Kay, and H. S. Chan. 2017. Charge pattern matching as a 'fuzzy' mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. New Journal of Physics 19(11):115003-115003.

15. Lin, Y. H., and H. S. Chan. 2017. Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. Biophys. J. 112(10):2043-2046.

16. Das, S., A. N. Amin, Y. H. Lin, and H. S. Chan. 2018. Coarse-grained residue-based models of disordered protein condensates: Utility and limitations of simple charge pattern parameters. PCCP 20(45):28558-28574.

17. Boeynaems, S., A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. V. D. Bosch, R. Das, P. S. Tompa, R. V. Pappu, and A. D. Gitler. 2019. Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. Proc. Natl. Acad. Sci. U.S.A. 116(16):7889-7898.

18. Spruijt, E., A. H. Westphal, J. W. Borst, M. A. Cohen Stuart, and J. van der Gucht. 2010. Binodal Compositions of Polyelectrolyte Complexes. Macromolecules 43(15):6476-6484.

19. Garcia de la Torre, J., S. Navarro, M. C. Lopez Martinez, F. G. Diaz, and J. J. Lopez Cascales. 1994. HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys. J. 67(2):530-531.

20. Hofmann, H., A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc. Natl. Acad. Sci. U.S.A. 109(40):16155-16160.

21. Mao, A. H., S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U.S.A. 107(18):8183-8188.

22. Dignon, G. L., W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal. 2018. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. Proceedings of the National Academy of Sciences 115(40):9929.

23. Ghavami, A., E. Van der Giessen, and P. R. Onck. 2018. Sol–gel transition in solutions of FG-Nups of the nuclear pore complex. Extreme Mechanics Letters 22:36-41.

24. Dignon, G. L., W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal. 2018. Sequence determinants of protein phase behavior from a coarse-grained model. PLoS Comp. Biol. 14(1):e1005941-e1005941.

593 25. Ghavami, A., E. van der Giessen, and P. R. Onck. 2016. Energetics of Transport through the
594 Nuclear Pore Complex. PLoS One 11(2):e0148876-e0148876.

595