Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins

Xiangze Zeng, Kiersten M. Ruff and Rohit V. Pappu[1]

Department of Biomedical Engineering and Center for Science & Engineering of Living Systems (CSELS), Washington University in St. Louis, St. Louis, MO 63130

[1] To Whom Correspondence May be Addressed: pappu@wustl.edu

Author Contributions: XZ and RVP designed the research; XZ performed the simulations; XZ, KMR, and RVP contributed to the analysis; XZ and RVP wrote the manuscript; All authors edited the manuscript.

Competing Interest Statement: The authors do not have any competing interests to declare

Classification: Biophysics

Keywords: Polyampholytes; Polyzwitterions; Intrinsically Disordered Proteins; Self-Avoiding Walk; Globule; Flory random coil.

Abstract

The most commonly occurring intrinsically disordered proteins (IDPs) are polyampholytes, which are defined by the duality of low net charge per residue and high fractions of charged residues. Recent experiments have uncovered nuances regarding sequence-ensemble relationships of model polyampholytic IDPs. These include differences in conformational preferences for sequences with lysine vs. arginine, and the suggestion that well-mixed sequences form a range of conformations, including globules, conformations with ensemble averages that are reminiscent of ideal chains, or self-avoiding walks. Here, we explain these observations by analyzing results from atomistic simulations. We find that polyampholytic IDPs generally sample two distinct stable states, namely globules and self-avoiding walks. Globules are favored by electrostatic attractions between oppositely charged residues, whereas self-avoiding walks are favored by favorable free energies of hydration of charged residues. We find sequence-specific temperatures of bistability at which globules and self-avoiding walks can coexist. At these temperatures, ensemble averages over coexisting states give rise to statistics that resemble ideal chains without there being an actual counterbalancing of intra-chain and chain-solvent interactions. At equivalent temperatures, arginine-rich sequences tilt the preference toward globular conformations whereas lysine-rich sequences tilt the preference toward self-avoiding walks. We also identify differences between aspartate and glutamate containing sequences, whereby the shorter aspartate sidechain engenders preferences for metastable, necklace-like conformations. Finally, although segregation of oppositely charged residues within the linear sequence maintains the overall two-state behavior, compact states are highly favored by such systems.

Significance Statement

Intrinsically disordered regions (IDRs) of proteins, when tethered to folded domains, function either as flexible tails or as linkers between domains. Most IDRs are polyampholytes that comprise

a mixture of oppositely charged residues. Recent measurements of tethered polyampholytes showed that tendency of arginine- and lysine-rich sequences to behave very differently from one another. Using computer simulations, we show that these differences are determined by differences in free energies of hydration, steric volumes, and other considerations. Further, the interplay between electrostatic attractions and favorable free energies of hydration creates distinct stable states for polyampholytic IDRs. These findings have implications for switch-like transitions and the regulation of effective concentrations of interaction motifs by IDRs.

Significant fractions of eukaryotic proteomes are made up of intrinsically disordered regions (IDRs) (1). Conformational heterogeneity (2) is a defining hallmark of IDRs (3-5). Studies over the past decade have helped quantify relationships (6) that connect sequence-encoded information within IDRs to properties of conformational ensembles such as overall sizes and shapes, the amplitudes of spontaneous conformational fluctuations, and the dynamics of interconverting between distinct conformational states (7-19). These sequence-ensemble relationships have direct functional consequences that have been uncovered via studies based on biophysical, biochemical, and engineering approaches (5, 20-38). Our work, which is focused on physical principles underlying sequence-ensemble relationships of IDRs, is of direct relevance to understanding how IDRs function.

Charged residues are key determinants of sequence-ensemble relationships of IDRs (19, 39-41). They contribute through highly favorable free energies of hydration (42) and long-range electrostatic interactions. Net charge per residue (19, 39, 40) and the patterning of oppositely charged residues (43-45) have been used as order parameters for describing sequence-ensemble relationships and interactions of charge-rich IDRs (46). Both features can be modulated through post-translational modifications (47-51), charge renormalization by solution ions (52), and charge regulation through context- and conformation-dependent uptake and release of protons (53).

Polyampholytes feature roughly equivalent numbers of oppositely charged residues, and they make up more than 70% of known IDRs (7, 17). For a given amino acid composition, which sets the fraction of charged residues and the net charge per residue, it has been shown that the linear mixing vs. segregation of oppositely charged residues can have a profound impact on sequence-ensemble relationships of polyampholytic IDRs (31, 32, 43). Specifically, for a given set of solution conditions, sequences featuring uniform linear distributions of oppositely charged residues are predicted to favor more expanded conformations compared to sequences with identical amino acid compositions where the oppositely charged residues are segregated into distinct blocks along the linear sequence. These predictions made using simulation and theory (43, 44, 54), have been confirmed using different experiments (31-33, 41, 55).

The ensemble-averaged radii of gyration (R_g) of flexible polymers follow scaling relationships of the form $R_g \sim N^{\nu}$. Here, N denotes the number of residues and the scaling exponent ν is a measure of the length-scale over which conformational fluctuations are correlated. For homopolymers or systems that are effective homopolymers, ν has four limiting values *viz*., 0.33, 0.5, 0.59, or 1, corresponding to globules, Flory random coils, self-avoiding walks, and rod-like conformations, respectively (56). Atomistic simulations performed at fixed temperatures suggest that $\nu \approx 0.59$ (43) for strong, well-mixed polyampholytes (17). The explanation for this behavior is as follows: Electrostatic attractions and repulsions are realized on similar length scales for well-mixed sequences. These interactions screen one another, and the highly favorable free energies of hydration become the main determinants of overall sizes and shapes of well-mixed strong polyampholytes (17). In contrast, compact conformations are formed by strong polyampholytes where oppositely charged residues are segregated into distinct blocks. Here, the electrostatic attractions between oppositely charged blocks can outcompete opposing effects of favorable solvation. These inferences were gleaned using sequences comprising 1:1 ratios of Lys and Glu (43). In the original simulations, the reference free energies of hydration of all charged residues were treated as being quantitatively equivalent and highly favorable. This leads to the hypothesis

that Lys and Arg are interoperable with one another as determinants of sequence-ensemble relationships of IDRs (17). A similar inference emerges regarding the interoperability of Asp and Glu with respect to one another. The recent work of Sørensen and Kjaergaard has challenged these inferences (57). Using a system where model IDRs were deployed as flexible linkers between interaction domains, Sørensen and Kjaergaard used their measurements to estimate the relationships between amino acid sequence and the scaling exponent ν (57). Inferences from their experiments suggest that the $\nu \approx 0.33$ for $(GRESRE)_n$ and $\nu \approx 0.5$ for $(GKESKE)_n$. Here, n is the number of repeats of the hexapeptides GRESRE or GKESKE. The results point to significant differences between Arg- and Lys-containing sequences. Further, while globularity of $(GRESRE)_n$ has precedent in mean-field theories for polyampholytes, the mechanism by which Flory random coil-like behavior of $(GKESKE)_n$ is achieved is unclear. Here, we develop a physical explanation for the findings of Sørensen and Kjaergaard (57). Our work is based on atomistic simulations and the ABSINTH implicit solvation model and forcefield paradigm (58-61).

Results

Conformational ensembles of polyampholytic intrinsically disordered proteins (IDPs) show two-state behavior: Simulations were performed using an adaptation of recently recalibrated free energies of hydration (42, 62). These new values highlight the more hydrophobic nature of Arg when compared to Lys, and more favorable hydration of Asp / Glu when compared to Arg or Lys (see details in the *SI Appendix*). Using the recalibrated free energies of hydration (42, 62), we computed free energy profiles with $x = (R_g / N^{0.5})$ as the reaction coordinate (see details in *SI Appendix*). Note that $x \equiv x_{FRC} \approx 2.5$ Å per residue sets a useful reference length scale (63). Here, x_{FRC} is the value we obtain from our numerical approximation of the Flory random coil (FRC), which is an ideal chain model where all non-nearest neighbor interactions are ignored (2, 63). The free energy profile $W(x)$ quantifies the free energy change associated with converting a Flory random coil to more expanded ($x > x_{FRC}$) or compact ($x < x_{FRC}$) conformations.

The free energy profiles, calculated at different simulation temperatures, are shown in Fig. 1A for $(GKESKE)_7$. Seven repeats are short enough to be computationally tractable and long enough to observe the full spectrum of transitions without being confounded by finite size effects (43). The probability distribution function $P(x)$, resolved along x as the reaction coordinate, is shown in Fig. 1B. Both $W(x)$ and $P(x)$ show that there are two stable states, one for $x > x_{FRC}$ and another for $x < x_{FRC}$ (*SI Appendix*, Fig. S3). In support of this two-state behavior, we note that the density of $P(x)$ at $x = x_{FRC}$ is essentially zero. The width of the well for $x < x_{FRC}$ ranges from 1.4 to 2.2 Å per residue, and the minimum is located at $x \approx 1.6$ Å per residue. Based on the scaling of internal distances, which shows plateauing behavior for low temperatures (*SI Appendix*, Fig. S2), we designate $x \approx 1.6$ Å per residue, the location of the one of the minima on the free energy profile, as $x_{globule}$ (Fig. 1A). The width of the well for $x > x_{FRC}$ ranges from 2.2 to 4.5 Å per residue. The minimum in $W(x)$ and the corresponding peak in $P(x)$ are located at $x \approx 3.4$ Å per residue. Scaling analysis shows that this free energy minimum is defined by a value of 0.6 for ν, implying that the minimum for $x > x_{FRC}$ corresponds to a self-avoiding walk (*SI Appendix*, Fig. S2). Accordingly, we designate the minimum at $x \approx 3.4$ Å per residue as x_{SAW} (Fig. 1A).

As temperature increases, the positions of the two minima in $W(x)$ remain roughly fixed, while the relative depths, and the barrier separating the minima, change with temperature. At low

temperatures, $(GKESKE)_7$ favors globules. At high temperatures, the preferred state is the self-avoiding walk. At an intermediate temperature, globules and self-avoiding walks are of equivalent stability. The scaling exponent ν approaches 0.5 as the temperature of bistability is approached. To extract the temperature of bistability, we fit a two-state model to $P(x)$ using a lever-rule: $P(x;T) = a_T P_{\text{globule}}(x) + (1 - a_T) P_{\text{SAW}}(x)$. Here, $0 \leq a_T \leq 1$; $P_{\text{globule}}(x)$ and $P_{\text{SAW}}(x)$ represent unimodal distributions peaked around x_{globule} and x_{SAW}, respectively (Fig. 1C). These distributions were extracted from simulations at 298 K and 350 K, respectively. Using these reference distributions for each state, the estimated temperature of bistability, defined as the temperature where $a_T = 0.5$, is ≈ 315 K for $(GKESKE)_7$ (inset in Fig. 1C).

We propose that the apparent FRC behavior reported by Sørensen and Kjaergaard (57) for the $(GKESKE)_n$ system might derive from there being a bistability at the specific solution conditions that were investigated. Two-state behavior arises from competing interactions, namely, the highly favorable free energies of hydration of charged residues, which favor self-avoiding walks, and electrostatic attractions, which favor compaction. In contrast to charge-rich polyampholytic IDRs, the $P(x)$ distributions for neutral polymers lacking charged groups should be continuous, tracking with the continuous change of the two-body interaction coefficient (6, 64). To demonstrate this, we computed temperature-dependent distributions for an artificial sequence designated as $(GkeSke)_7$. Here, k and e are deprotonated and protonated versions of Lys and Glu, respectively. The free energies of hydration of deprotonated Lys and protonated Glu are ~20 times lower in magnitude than the charged versions (*SI Appendix*, Table S2). Therefore, both competing interactions that are present in $(GKESKE)_7$ are lost in $(GkeSke)_7$, setting the two-body interactions due to short-range interactions as the main determinants of globule-to-coil transitions. As expected, the distributions for $(GkeSke)_7$ show a continuous evolution from globules to self-avoiding walks, with there being at least one transition temperature where the distribution is peaked at x_{FRC} (Fig. 1D and *SI Appendix*, Fig. S4). Comparing the results in Fig. 1B to Fig. 1D suggests clear differences between the neutral, uncharged polymers vs. polyampholytes.

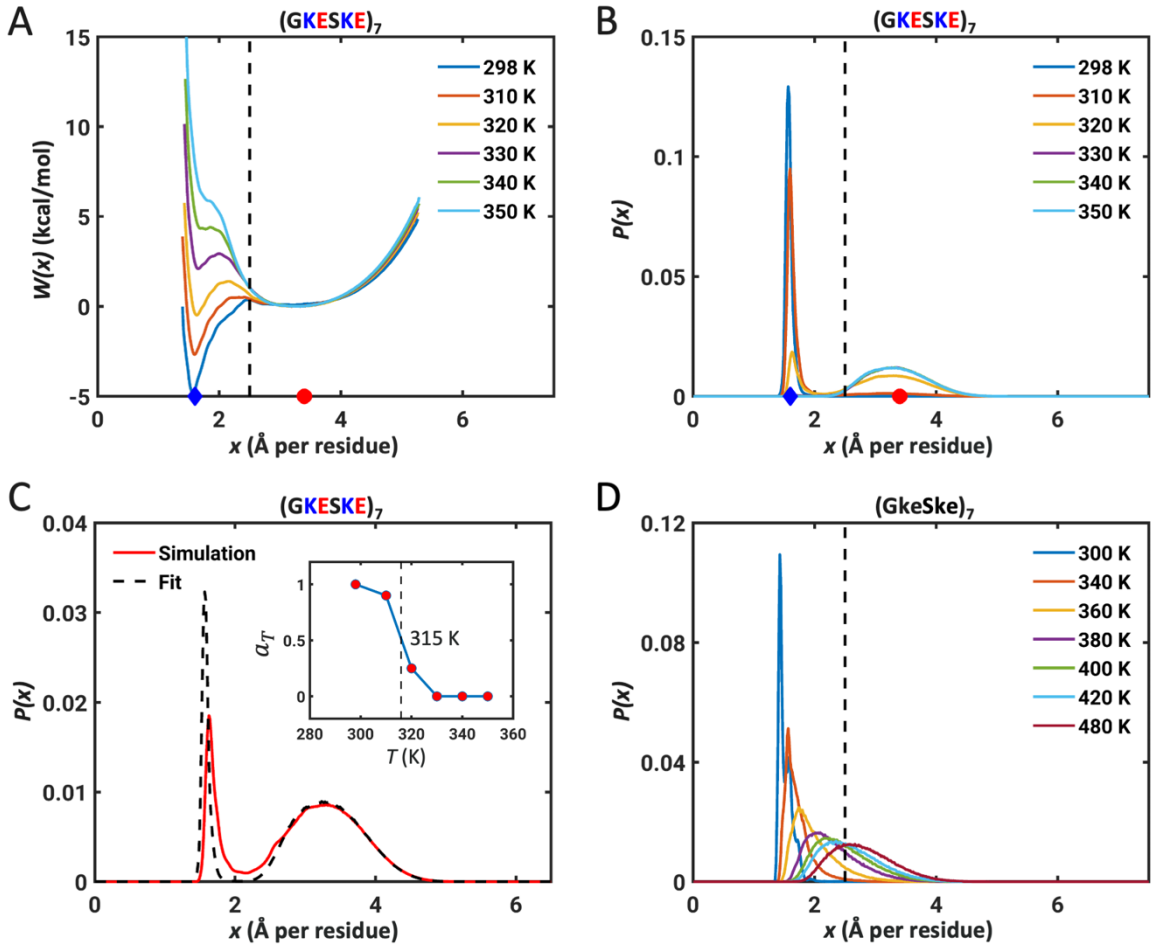

Fig. 1. Neutral polyampholytic IDPs show two-state behavior. (A) The free energy profile $W(x)$ for $(GKESKE)_7$ calculated for different simulation temperatures. (B) Probability distribution functions $P(x)$ obtained from the free energy profiles show the presence of two distinct peaks corresponding to the distinct minima in panel (A). The dashed line in (A) and (B) indicates the reference length scale of x_{FRC} = 2.5 Å per residue. And in each of the panels, the blue diamonds, and red circles mark positions of $x_{globule}$ and x_{SAW}, respectively. (C) The bimodal distribution $P(x)$ can be fit to a two-state model (see main text). The fit is shown here for an intermediate simulation temperature of 320 K. The inset shows a_T as a function of temperature and the dashed vertical line corresponds to $T = 315$ K where $a_T \approx 0.5$. (D) The distributions $P(x)$ for the neutral polymer lacking charged residues change continuously as temperature increases. Note the existence of temperatures where distributions can be peaked around x_{FRC}, which is not the case for the parent sequence $(GKESKE)_7$ that has charged residues.

There is theoretical precedent for the observed two-state behavior of polyampholytes. For stiff homopolymers, the transition between globules and coils is sharp and the sharpness of this transition is governed by the interplay between the two- and three-body interactions (65-67). Muthukumar has shown the existence of an effective temperature of bistability for polyelectrolytes in solution, where the bistability is governed by the interactions with solution ions (68). Recently,

Ghosh and coworkers have extended the approach of Muthukumar to describe the collapse transitions of finite-sized heteropolymers. They report the existence of an effective temperature of bistability that results from the choice they make for the free energy functional for polyampholytic systems (54, 69). Using a mean-field description for random polyampholytes, an approach that ignores chain connectivity and sequence details, Higgs and Joanny showed that in the long-chain limit of $N \rightarrow \infty$, individual random polyampholytes collapse to form globules irrespective of the sign or magnitude of the two-body interaction coefficient (70). The globular state is favored by electrostatic attractions. In contrast to the long-chain limit, finite-sized chains in the theory of Higgs and Joanny are predicted to form expanded conformations that are akin to self-avoiding walks (70). The preference for collapsed states (71, 72), whereby electrostatic attractions drive chain collapse, was also predicted by the Flory theory for individual chains of random polyampholytes that was developed by Dobrynin and Rubinstein (73). Our observations of two-state behavior closely mimic those of Moldakarimov et al., who noted that neutral polyampholytes undergo jump-like coil-to-globule transitions due to the formation of intra-chain ion pairs (74). The two-state behavior and the nature of the collapse transition can be modulated by an excess of charge of one type (75-77).

The main finding is that a scaling exponent of 0.5 can arise because the conformations are mixtures of globules, defined by a negative excluded volume, and SAWs, defined by a positive excluded volume. In contrast, observations of scaling exponents of 0.5 are typically taken to mean that the polymer in question is at its theta temperature with an excluded volume of zero. The challenge is to discern between these two scenarios, both of which yield an apparent scaling exponent of 0.5. This requires the deployment of measurements that directly access or allow one to infer the distributions of polymer sizes. This seems feasible using single molecule measurements that access distributions of radii of gyration or internal distances (39, 78, 79). Advances in single molecule detection might make such experiments feasible (80). A single molecule variant of the measurements performed by Sørensen and Kjaergaard could be deployed to uncover the distributions of effective concentrations. The measurements will have to be of coil-to-globule transitions as a function of temperature or denaturant. Unlike the continuous transitions that have been reported for most systems, the expectation is of fixed two-state behavior (if temperature is the perturbant) or variable two-state behavior (if denaturant is the perturbant) (81). The signatures of bistable vs. continuous transitions can be gleaned by quantifying the distributions of R_g and effective concentration values at the system-specific apparent theta temperatures. This is shown in **Fig. 2A** and **Fig. 2B**.

Can differences between the two types of behaviors be detected using an ensemble measurement such as small angle x-ray scattering (SAXS)? To answer this question, we computed Kratky profiles (82) for $(GKESKE)_7$ vs. $(GkeSke)_7$. For each system, three specific temperatures were chosen, and these correspond to temperatures where the systems of interest are characterized by apparent scaling exponents (ν_{app}) of ≈ 0.33, ≈ 0.5, and ≈ 0.59. For the $(GKESKE)_7$ system we observe Kratky profiles with two distinct peaks located at similar positions, and the three Kratky profiles crossover at a common wavenumber. The presence of a single crossover length scale, circled in **Fig. 2C**, as a function of temperature or other perturbants is to be contrasted with the Kratky profiles for the $(GkeSke)_7$ system (**Fig. 2D**). For the system without charged residues, the continuous nature of the globule-to-SAW transitions give rise to non-coincident cross-over length

scales for the globule and coil vs. globule and SAW, plus an additional crossover between the profiles for coils and SAWs.

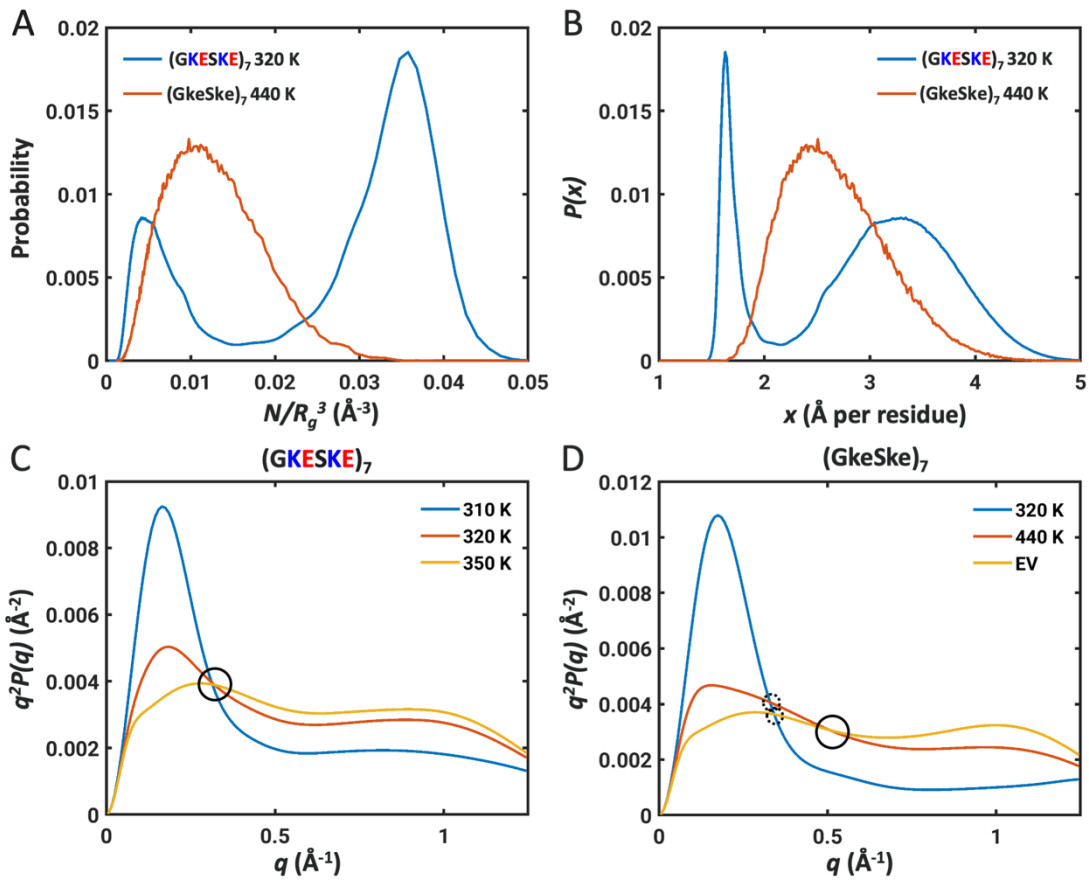

Fig. 2: R_g distributions and Kratky profiles at the apparent theta temperature for two types of systems where the scaling exponents would be 0.5. (A) Differences in R_g distributions, plotted as distributions of effective concentrations at the apparent theta temperatures for $(GKESKE)_7$ (blue) vs. $(GkeSke)_7$ (red). (B) Differences in distributions of x, the normalized R_g distributions between the two systems at their respective apparent theta temperatures. The bimodality of a bistable system is replaced by a Gaussian behavior for system that lacks charged residues. (C) Kratky profiles for globules (blue), coils (red), and yellow (SAWs) for $(GKESKE)_7$ computed at the relevant temperatures. The single crossover length scale is circled for ease of identification. (D) Kratky profiles for globules (blue), coils (red), and yellow (SAWs) for $(GKESKE)_7$ computed at the relevant temperatures. Non-coincidence of crossover length scales is shown as dashed circles, and the presence of an additional crossover length scale is identified using a solid circle.

Temperature of bistability and the nature of the collapse transition are affected by sidechain chemistries: The two-state behavior and the presence of a temperature of bistability arises from the competing effects of globules arising from electrostatic attractions and favorable free energies of hydration that prefer self-avoiding walks. Therefore, we propose that the sidechain

specific reference free energies of hydration, which arise from differences in chemical structures of sidechains, combined with chemistry-specific interactions of sidechains should affect the temperature of bistability and the nature of the collapse transition. We tested this hypothesis using ABSINTH-based simulations of $(GRESRE)_7$, $(GKDSKD)_7$, and $(GRDSRD)_7$. Being able to capture distinct preferences of Arg- vs. Lys-containing sequences requires a formal accounting of the differences in free energies of hydration that were derived recently using a combination of experimentally derived quantities and free energy calculations (42, 62). These results indicate that despite being a strong base, arginine is more hydrophobic than lysine, the physical basis for which has been discussed recently by Fossat et al., (42).

We obtained free energy profiles and probability density functions, $W(x)$ and $P(x)$, for $(GRESRE)_7$, $(GKDSKD)_7$, and $(GRDSRD)_7$ as shown in *SI Appendix*, Fig. S6. For these simulations, we used recalibrated free energies of hydration that are summarized in *SI Appendix*, Table S2. The qualitative behaviors of all three systems resemble those of $(GKESKE)_7$. These results suggest that the two-state behavior is a generic attribute of well-mixed polyampholytes. However, there are quantitative differences among the systems. These are quantified in **Fig. 3**, where we plot the ensemble average of x against reduced simulation temperatures for $(GKESKE)_7$, $(GRESRE)_7$, $(GKDSKD)_7$, and $(GRDSRD)_7$.

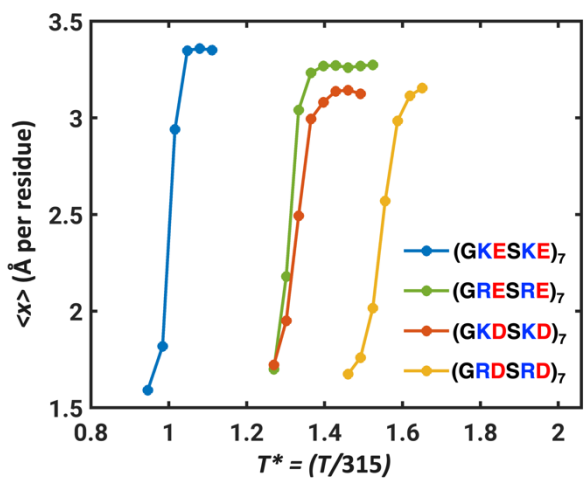

Fig. 3. Ensemble-averaged values of x, the normalized R_g, plotted against simulation temperature for $(GKESKE)_7$, $(GRESRE)_7$, $(GKDSKD)_7$, and $(GRDSRD)_7$.

The reduced temperature is $T^* = (T / 315\text{ K})$ where T is the simulation temperature, and 315 K corresponds to the inferred simulation temperature of bistability for $(GKESKE)_7$. The temperatures in our simulations are not to be taken literally given the mean-field nature of the solvation model. Instead, interpretation of simulation temperatures in terms of realistic solution conditions will require some prior calibration against experimental data for the system of interest – as has been shown in recent work, albeit in different contexts (46, 83). Comparisons in terms of the reduced temperature T^* indicate that replacing Lys with Arg shifts the temperature of bistability upward by a factor of ~1.3 for $(GRESRE)_7$. Likewise, the temperature of bistability is higher by a factor of ~1.2 for $(GRDSRD)_7$ compared to $(GKDSKD)_7$. It follows that Arg as the

basic residue tilts the balance toward globules when compared to Lys, which tilts the balance toward self-avoiding walks (*SI Appendix*, Fig. S8).

In addition to differences between Lys and Arg, we find that replacing Glu with Asp also has an impact on the transition between globules and self-avoiding walks. Specifically, the apparent temperature of bistability is higher by a factor of ~1.3 for $(GKDSKD)_7$ when compared to $(GKESKE)_7$. Likewise, the apparent temperature of bistability is higher by a factor of ~ 1.25 for $(GRDSRD)_7$ compared to $(GRESRE)_7$. These differences cannot be explained in terms of differences in free energies of hydration (see *SI Appendix*, Table S2). Instead, we considered two plausible explanations for the differences we observe: First, the longer sidechain of Glu and the higher sidechain entropy could lower the temperature of bistability. Second, the longer sidechain of Glu and the higher steric volume will likely weaken electrostatic attractions within the interior of globules.

To test the two hypotheses, we compared profiles for temperature-dependent collapse transitions of two artificial sequences $(GkeSke)_7$ and $(GkdSkd)_7$. Here, k is the deprotonated Lys, whereas e and d are protonated versions of Glu and Asp, respectively. If differences in sidechain entropy are the main contributors, then it should follow that $(GkdSkd)_7$ will have a higher transition temperature than $(GkeSke)_7$. However, as shown in *SI Appendix*, Fig. S9, the transition temperature for $(GkdSkd)_7$ is lower than that of $(GkeSke)_7$, thereby ruling out the first hypothesis. We tested the second hypothesis by comparing the transition temperature of $(GKESKE)_7$ to that of $(GZESZE)_7$ where Z is 2,4-diamino butyric acid or Dab. This maintains the charge and chemical structure of the amine while replacing the longer sidechain in Lys with the shorter sidechain of Dab. As shown in *SI Appendix*, Fig. S10, the transition temperature for $(GZESZE)_7$ is ~1.3 times higher than $(GKESKE)_7$. This result implies that charged residues with equivalent functional groups and shorter sidechains stabilize compact conformations when compared to sequences with charged residues that have longer sidechains. Accordingly, we ascribe the differences between $(GRDSRD)_7$ and $(GRESRE)_7$ to the fact that the longer sidechain of Glu and the higher steric volume weaken electrostatic attractions within the interior of globules.

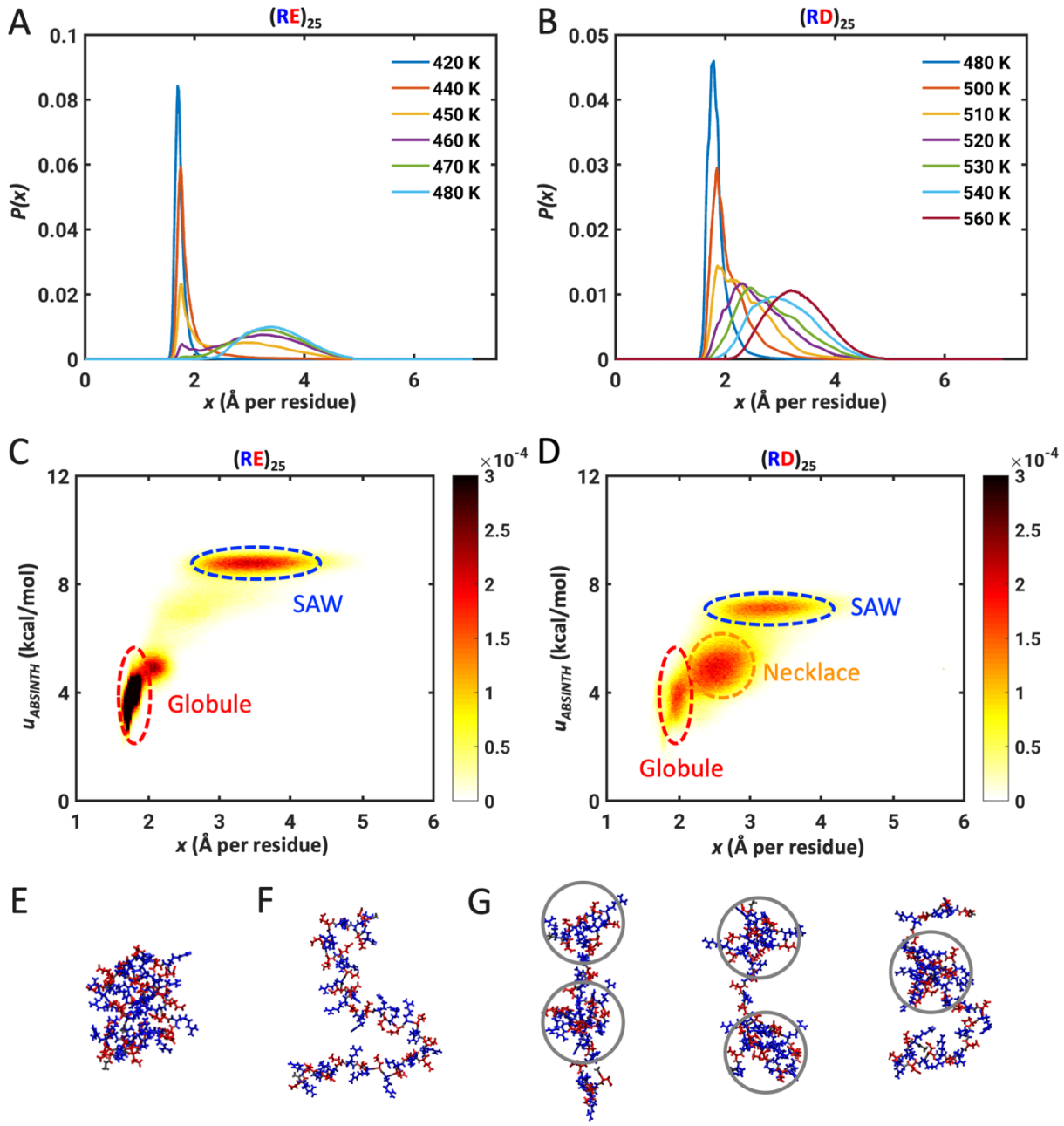

Fig. 4. **Asp residues modulate the two-state behaviors of polyampholytic IDPs.** Calculated distributions $P(x)$ for (A) $(RE)_{25}$ and (B) $(RD)_{25}$. Two-dimensional histograms $p(x, u_{\mathrm{ABSINTH}})$ for (C) $(RE)_{25}$ and (D) $(RD)_{25}$. Here, u_{ABSINTH} is relative potential energy per residue, referenced to the system-specific lowest energy conformation. The histograms were computed as an average over unbiased replica exchange simulations at three temperatures near the transition temperature, namely 450 K, 460 K and 470 K for $(RE)_{25}$, 520 K, 530 K and 540 K for $(RD)_{25}$. The bin widths for x and u_{ABSINTH} are 0.004 Å and 0.02 kcal/mol per residue, respectively. (E) Snapshot of a representative globular conformation for $(RE)_{25}$. (F) Snapshot of a representative self-avoiding-walk conformation for $(RE)_{25}$. (G) Three snapshots, showing representative necklace-like

conformations, with gray circles delineating the "pearls" formed by $(RD)_{25}$. Structures were drawn using VMD (84). Lys is shown in blue and Glu / Asp residues are in red.

To further interrogate the differences between different well-mixed polyampholytes, we computed free energy profiles and distribution functions for sequences that are essentially polyzwitterions (85). These include $(RE)_{25}$, $(KE)_{25}$, $(RD)_{25}$, and $(KD)_{25}$ (Figs. 4A-B and *SI Appendix*, Figs. S11, S12). Our investigations of these systems were motivated by recent work showing the relevance of RD and RE repeats for nuclear speckle assembly (86). This work also highlighted profound differences in the phase behavior of Arg vs. Lys containing polyzwitterions. The Glu-containing polyzwitterions show two-state behavior, whereas the Asp-containing sequences show the presence of an intermediate state that is favorably populated at the transition temperature. We examined two-dimensional histograms $p(x, u_{\mathrm{ABSINTH}})$ computed in terms of x and the conformation-specific values of the ABSINTH potential energies. These histograms confirm the two-state behavior of $(RE)_{25}$ (Fig. 4C). However, for $(RD)_{25}$, we observe an intermediate state between globules and self-avoiding walks (Fig. 4D). These intermediate states correspond to "necklace-like" conformations (Fig. 4G) that have been predicted to be either stable or metastable states by different theories for polyampholytes (75). We compute differences in the numbers of hydrogen bonds formed by Asp containing polyzwitterions. These differences are most pronounced for conformations with values of R_g that are on a par with or larger than Flory random coils (*SI Appendix*, Fig. S13). The Asp containing sequences feature more hydrogen bonds, in expanded conformations, when compared to Glu containing sequences. Accordingly, it appears that necklace-like conformations formed (RD) and even (KD) repeats are stabilized by hydrogen bonds enabled by the shorter Asp sidechain.

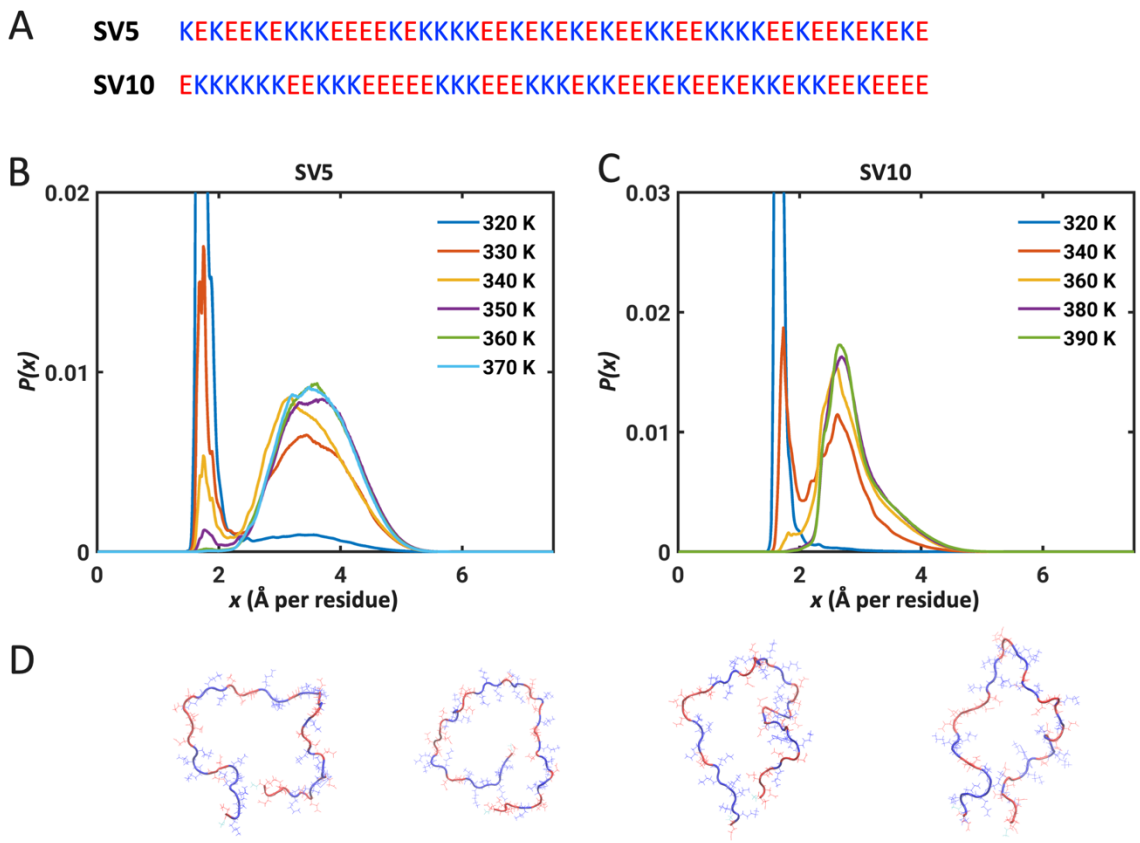

Fig. 5. Two-state behavior is preserved for sequences with blocks of oppositely charged residues. (A) Amino acid sequences of sv5 and sv10 – nomenclature as in Das and Pappu (43). Panels (B) and (C) show the distributions $P(x)$ computed for two sequences designated as sv5 and sv10. The corresponding free energy profiles of (B) and (C) are shown in *SI Appendix* Fig. S14 (B) $P(x)$ for sv5 shows that the two-state behavior is preserved for sequences with mild degrees of segregation of oppositely charged residues. (C) $P(x)$ for sv10 shows that there are temperatures where intermediate states, corresponding to well-defined metastable structures, become prominent as the linear segregation of oppositely charged residues increases. (D) Four Ω-loop-like structures (87) observed as stable intermediates for sv10. The structures were extracted from simulations performed at 340 K and were drawn using VMD (84). Lys is in blue and Glu residues are in red.

Influence of sequence patterning on the two-state behavior of polyampholytic IDRs: Next, we investigated the impact of the linear segregation of oppositely charged residues on the two-state behavior of polyampholytes. For this, we performed simulations for two Lys and Glu containing sequences designated as sv5 and sv10 (**Fig. 5A**) (43). The results in **Figs. 5 B-C** show that these two sequences also exhibit two-state behavior. However, the peak in the distribution $P(x)$ at high temperatures is much smaller (~2.7 Å per residue) for sv10 when compared to that of sv5 (3.6 Å). For self-avoiding walk-like conformations, the value of x should have a peak around 3.4 Å per residue. This is not the case for sv10. Instead, the distribution $P(x)$ at high temperature for sv10 has a wide right shoulder that can be fit to a mixture of two Gaussian distributions, one

peak of 2.7 Å per residue, and the other peaked at 3.2 Å per residue. This implies that the blockier sequence populates a metastable state, which for sv10 corresponds to Ω-loop-like structures (87) that are shown in **Fig. 5D**. These metastable structures are stabilized by electrostatic attractions between the N-terminal Lys patch and C-terminal Glu patch. The close contacts between N-terminal and C-terminal are clearly shown in the distance and scaling maps (*SI Appendix,* Fig. S16). Overall, increased charge segregation increases the stability of different compact structures that are globule-, loop-, and hairpin-like (*SI Appendix*, Fig. S17).

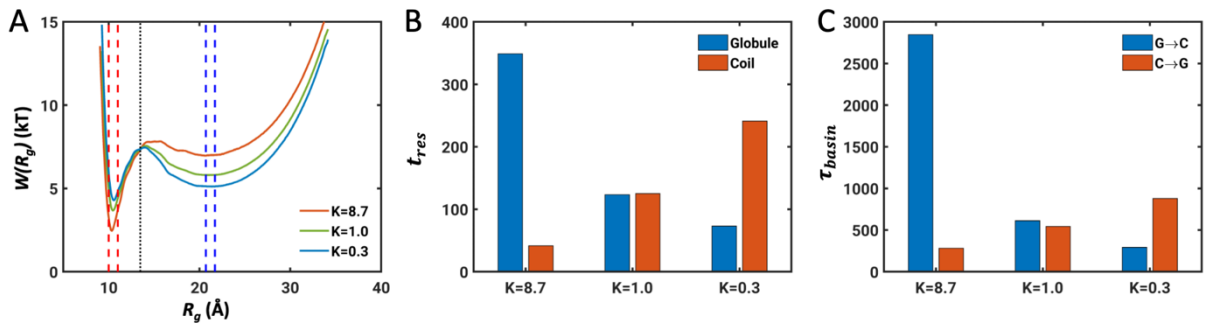

Fig. 6. Simulated dynamics for transitions between globules and self-avoiding walks. (A) Free energy profiles $W(R_g)$ used for Langevin dynamics simulations. The dashed line indicates the boundary between the globule and coil state. Regions of equal width within the globule and SAW basins are indicated by the red and blue dashed windows, respectively. (B) Mean residence times (t_{res}) within the globule and coil (self-avoiding walk) states. (C) Mean basin-to-basin transit times (τ_{basin}) for forward and reverse transitions between globule (red region in panel A) to the coil (SAW) basins. Details of how we calculated τ_{basin} are described in the *SI Appendix*.

Dynamics of interconversion between globules and self-avoiding walks: To investigate the dynamics of interconversion between globules and self-avoiding walks, we performed Langevin dynamics simulations using the free energy profiles extracted for $(GKESKE)_7$ (see details in *SI Appendix*). These simulations were performed using three different profiles (Fig. 6A), one corresponding to the case where the globule is more stable than the self-avoiding walk (K=8.7), the second corresponding to the case of bistability (K=1.0), and the third corresponding to the case where the self-avoiding walk is more stable than the globule (K=0.3). From the simulation trajectories, we calculated the mean residence times to the left and right of the dividing line, depicted as a dotted line in Fig. 6A. The mean residence times track with the equilibrium constants, being longer in the globule region (left of the barrier) for K = 8.7 and longer in the coil region (right of the barrier) for K = 0.3 (Fig. 6B). In addition to the mean residence times (t_{res}), we computed the mean time for transitioning from the globule-to-coil basins and vice versa. To account for the asymmetries of the globule and coil basins, we computed the basin-to-basin transit times (τ_{basin}) as the time it takes for transitioning from the window striped in red to that in blue and vice versa. Given that a barrier must be negotiated for these transitions to occur, we find that τ_{basin} is greater than t_{res} by an order of magnitude (Fig. 6C). For our analysis of τ_{basin} and t_{res} we used R_g as the reaction coordinate. This is relevant because the transitions between globules and self-avoiding walks are associated with the change in internal density and large fluctuations in density around the barrier (delineated in Fig. 6A) as noted by Kuznetsov and Grosberg (88). It is worth

emphasizing that estimation of τ_{basin} rests on the choice of two regions of equal width within the globular and coil regions. However, the potential of mean force is an asymmetrical double well, with the coil basin being broad and shallow when compared to the globule basin – even at the temperature of bistability. In such systems, the mean first passage time for crossing the barrier will be influenced by the diffusive search within the broad and shallow basin, thereby yielding an effective kinetic readout that might confound simple interpretations of two-state behavior (89).

Discussion

In this work, we have uncovered the two-state behavior of strong polyampholytic IDPs. At a given temperature, the relative stabilities of the two states are modulated by the identity of the basic residue, Lys vs. Arg (Fig. 2). Our findings help explain the observations of Sørensen and Kjaergaard (57). We show that the apparent Flory random coil-like scaling behavior reported for the $(GKESKE)_n$ system arises from an averaging over the contributions of coexisting globules and self-avoiding walks. Further, we find that the two-state behavior is altered to accommodate metastable necklace-like, conformations when the acidic residue is Asp as opposed to Glu. Although the two-state behavior is preserved for polyampholytes with blocky architectures where oppositely charged residues are segregated from one another, compact, globule-, loop-, and hairpin-like conformations are highly stable due to intra-chain electrostatic attractions (*SI Appendix*, Fig. S17). Detection of the proposed bistable behavior will require experiments that use temperature, salt, denaturant, or pH as perturbants. Ideally, the experiments will be able to uncover population distributions, not just ensemble averages, although in systems with high fractions of charged residues even ensemble averaged measurements might be valuable (Fig. 2).

Our findings establish that competing effects of favorable hydration of charged groups, differences in steric volumes (*SI Appendix*, Fig. S8), and intra-chain electrostatic interactions among oppositely charged residues give rise to two distinct states namely, self-avoiding walks and globules. The relative stabilities are influenced by: (i) temperature, which is a proxy for solvent quality; (ii) the identity of the basic residues, with Arg favoring globules more so than Lys; (iii) the linear segregation vs. mixing of oppositely charged residues, where increased segregation gives rise to stronger preferences for compact states; (iv) the fraction of charged residues and the net charge per residue, with lower fractions of charged residues enabling conventional, continuous coil-to-globule transitions, and increased charge of one kind driving the preference for self-avoiding walk-like conformations (19, 39, 40). The differences between Arg and Lys as well as the differences between Asp and Glu (albeit to a lesser extent) have received recent attention in the context of work on the molecular grammar underlying the driving forces for phase separation of IDPs (46, 86, 90-92). As noted by Tesei et al., (93) different hydrophobicity scales show a lack of consensus regarding the apparent hydrophobicity of Arg vs. Lys. Arriving at a consensus will require the assessment of differences in intrinsic differences in free energies of hydration and in the context dependent effects of Arg vs. Lys and Glu vs. Asp (42, 62). Additionally, the sensitivity of a sequence to the effects of charge regulation are also likely to be highly dependent on the identities of ionizable residues as well as sequence contexts (53, 94).

Metastable, necklace-like conformations are accessible to sequences featuring differences in the lengths of sidechains of acidic vs. basic residues. Examples of such systems include RD repeats that are found in IDRs within nuclear speckle proteins (86). RD repeats form spherulitic

fibrils through intermolecular associations, and the formation of fibrillar solids is governed by the number of RD repeats and overall charge neutrality (86). We propose that necklace-like conformations are likely to be the drivers of intermolecular associations that give rise to fibrillar structures. This hypothesis is based on the observation that the necklace conformation might be an excited state corresponding to the so-called N* state proposed by Thirumalai and coworkers (95, 96) as being essential for driving fibril formation. Greig et al., (86) also showed that Asp and Glu play different roles in the nuclear speckle condensation. Whereas RD repeats form fibrils, the RE repeats form non-fibrillar condensates. Further, the driving forces for assembly are stronger for the RD repeats when compared to RE repeats. Our studies establish clear differences in the energy landscapes of these systems. It remains to be ascertained if these differences help explain the finding of Greig et al. (86).

For infinitely long homopolymers or effective homopolymers the scaling exponent ν has four limiting values *viz*., 0.33, 0.5, 0.59, or 1. These values correspond to globules, Flory random coils, self-avoiding walks, and rod-like conformations, respectively (56). Deviations from these exponents are expected for finite sized homopolymers, and the magnitude of finite size corrections that are required can be formally estimated (97, 98). Unlike homopolymers, unfolded proteins and IDPs are complex, finite sized, heteropolymers. Such systems feature a spectrum of intra-chain and chain-solvent interactions. Accordingly, a value for ν_{app} that is not 0.33 or 1 could arise from a mixture of stable states – a point that his made in this work for strong polyampholytes. Our findings highlight the importance of going beyond estimates for ν_{app} as a device for comparative assessments of sequence-ensemble relationships of unfolded proteins and IDPs.

The two-state behavior of polyampholytic IDPs allows for the formal possibility of switch-like transitions between globules and self-avoiding walks. These transitions can be spontaneous if the solution conditions are such that the system of interest is in the vicinity of the temperature of bistability. Away from this temperature, switch-like transitions can be driven by energy inputs. Overall, the dynamics of interconversion between stable states will be influenced by sequence-intrinsic properties of the energy landscapes such as the barrier for interconversion and the widths of conformational basins around the stable states (Fig. 6).

Scanning for solution conditions (78, 99, 100) that place polyampholytic IDPs in the vicinity of the temperature of bistability would appear to be a way to enable switch-like interconversions between dramatically different conformational states of equivalent stability. Precedent for switch-like transitions and dynamics that span a spectrum of timescales has been reported for different IDRs (10, 13). Such transitions, especially in the context of tethered systems, are likely to make important contributions to the spatial and temporal control over biochemical reactions where IDRs modulate the effective concentrations of ligands around binding sites or of substrates around active sites of enzymes (31, 32, 34, 57, 101-105). While our results are presented for IDRs studied as autonomous units, we expect that our findings will be transferable to tethered IDRs, albeit with context-dependent modifications (103, 106).

The simulations reported here were performed in the absence of excess salt or solution ions (if the systems were electroneutral, see *SI Appendix*). To probe the effects of excess salt, we investigated the effects of adding 50 mM NaCl for the $(GRESRE)_7$ system. The results are summarized in *SI Appendix* Fig. S7. The temperature of bistability shifts down in the presence of

salt, and this can be rationalized as being mainly due to the screening effects of monovalent salts. Solutions ions are likely to play an important role in affecting conformational transitions, conformational equilibria, binding equilibria, and phase equilibria of polyampholytic and polyelectrolytic IDRs (24, 39, 52, 78, 90, 107, 108). Besides the screening effect, predicted by Higgs and Joanny (70) and illustrated in *SI Appendix* Fig. S7, there is the expectation that ions that preferentially accumulate around blocks of the same type of charge are released when the ion-chain interactions are replaced by intra-chain interactions (68, 108). The effects of solution ions are also likely to be directly relevant to the phase behaviors of polyampholytic IDRs (109, 110).

Finally, we asked if sequences that are likely to show bistability are present in the disordered proteome. For this, we calculated the distribution of fraction of charged residues (FCR) for IDRs in DisProt 2022-03 database (*SI Appendix*, Fig. S18). We also calculated FCR values of strong polyampholytes (17), and the fractions of specific charged residues. This analysis uncovered 202 out of 3236 IDRs with lengths ≥ 30 and FCR values equal to or larger than 0.5. Further, 26 out of 299 strong polyampholytes with lengths ≥ 30 have FCR values equal or larger than 0.5. Asp disfavors the bi-stable distribution, but it has lower frequency of being observed when compared to Glu. Higher fractions of Lys when compared to Arg also indicate that strong polyampholytes are likely to show a preference for bistability since Arg rich polyampholytes tend to prefer globules. *SI Appendix* Table S6 lists sequences for the top 15 strong polyampholytes with length ≥ 30 ranked by FCR in the DisProt 2022-03 database. Annotations of functions suggest that some of the IDRs are thought to function as flexible linkers that are regulated by multisite phosphorylation.

Materials and Methods

Monte Carlo simulations. All simulations were performed using version 2.0 of the CAMPARI molecular modeling software (http://campari.sourceforge.net/). The forcefield parameters were derived based on OPLS-AA/L forcefield as implemented in abs3.2_opls.prm. Free energies of hydration for the charged residues were based on recently recalibrated values. Na^+ or Cl^- ions were added to the simulation droplet to neutralize the system if the net charge of the peptide is not zero. All the details are described in *SI Appendix.*

Free energy profiles. Umbrella sampling was used to obtain free energy profiles using x as the reaction coordinate. Thermal replica exchange Monte Carlo simulations were used to enhance the sampling of globular structures in the windows with low R_g. Standard Metropolis Monte Carlo simulations were performed at the rest of windows. The weighted histogram analysis method was used to derive the free energy profile. The details of the setup for each system, the umbrella sampling, replica exchange, and reweighting are described in *SI Appendix.*

Acknowledgments

This work was supported by grants from the US National Institutes of Health (R01NS121114 and 5R01NS056114) and the US Air Force Office of Scientific Research (FA9550-20-1-0241). We thank Samuel Cohen, Martin Fossat, Alex Holehouse, Greg Jedd, Ammon Posey, Min Kyung Shinn, and Andrea Soranno for useful discussions regarding charge-rich IDRs. XZ would like to thank Louis Smith for helpful discussions and Jared Lalmansingh, Stephen Tahan and Mark Bober for support in the using McKelvey Engineering Compute Cluster and RIS cluster at Washington University in St. Louis.

References

1. R. van der Lee *et al.*, Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114**, 6589-6631 (2014).

2. N. Lyle, R. K. Das, R. V. Pappu, A quantitative measure for protein conformational heterogeneity. *The Journal of chemical physics* **139**, 09B607 (2013).

3. M. Guharoy, P. Bhowmick, P. Tompa, Design Principles Involving Protein Disorder Facilitate Specific Substrate Selection and Degradation by the Ubiquitin-Proteasome System *The Journal of biological Chemistry* **291**, 6723-6731 (2016).

4. M. Fuxreiter, P. Tompa, "Fuzzy Complexes: A More Stochastic View of Protein Function" in Fuzziness: Structural Disorder in Protein Complexes*,* M. Fuxreiter, P. Tompa, Eds. (Springer US, New York, NY, 2012), 10.1007/978-1-4614-0659-4_1, pp. 1-14.

5. B. Schuler *et al.*, Binding without folding – the biomolecular function of disordered polyelectrolyte complexes. *Current opinion in structural biology* **60**, 66-76 (2020).

6. Albert H. Mao, N. Lyle, Rohit V. Pappu, Describing sequence–ensemble relationships for intrinsically disordered proteins. *Biochemical Journal* **449**, 307-318 (2012).

7. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current opinion in structural biology* **32**, 102-112 (2015).

8. T. Mittag, L. E. Kay, J. D. Forman-Kay, Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition* **23**, 105-116 (2010).

9. M. R. Jensen, R. W. H. Ruigrok, M. Blackledge, Describing intrinsically disordered proteins at atomic resolution by NMR. *Current opinion in structural biology* **23**, 426-435 (2013).

10. G. Parigi *et al.*, Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **136**, 16201-16209 (2014).

11. M. Schwalbe *et al.*, Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and alpha-Synuclein in Solution from NMR and Small Angle Scattering. *Structure* **22**, 238-249 (2014).

12. T. Lazar *et al.*, PED in 2021: a major update of the protein ensemble database for intrinsically disordered proteins. *Nucleic Acids Research* **49**, D404-D411 (2020).

13. M. Tsytlonok *et al.*, Dynamic anticipation by Cdk2/Cyclin A-bound p27 mediates signal integration in cell cycle regulation. *Nature communications* **10**, 1676 (2019).

14. R. B. Best, Computational and theoretical advances in studies of intrinsically disordered proteins. *Current opinion in structural biology* **42**, 147-154 (2017).

15. G. H. Zerze, R. B. Best, J. Mittal, Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. *The Journal of Physical Chemistry B* **119**, 14622-14630 (2015).

16. N. Rezaei-Ghaleh *et al.*, Local and Global Dynamics in Intrinsically Disordered Synuclein. *Angewandte Chemie International Edition* **57**, 15262-15266 (2018).

17. A. S. Holehouse, R. K. Das, J. N. Ahad, M. O. Richardson, R. V. Pappu, CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical journal* **112**, 16-21 (2017).

18. M. C. Cohan, K. M. Ruff, R. V. Pappu, Information theoretic measures for quantifying sequence–ensemble relationships of intrinsically disordered proteins. *Protein Engineering, Design and Selection* **32**, 191-202 (2019).

19. A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, R. V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **107**, 8183-8188 (2010).

20. L. Mollica *et al.*, Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences* **3** (2016).

21. R. Schneider *et al.*, Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *Journal of the American Chemical Society* **137**, 1220-1229 (2015).

22. P. Tompa, Multisteric Regulation by Structural Disorder in Modular Signaling Proteins: An Extension of the Concept of Allostery. *Chemical Reviews* **114**, 6715-6732 (2014).

23. D. De Sancho, R. B. Best, Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1α with CBP. *Molecular BioSystems* **8**, 256-267 (2012).

24. A. Sottini *et al.*, Polyelectrolyte interactions enable rapid association and dissociation in high-affinity disordered protein complexes. *Nature communications* **11**, 5736 (2020).

25. A. Borgia *et al.*, Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61-66 (2018).

26. C. Cragnell, L. Staby, S. Lenton, B. B. Kragelund, M. Skepö, Dynamical Oligomerisation of Histidine Rich Intrinsically Disordered Proteins Is Regulated through Zinc-Histidine Interactions. *Biomolecules* **9**, 168 (2019).

27. W. Borcherds *et al.*, Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nature Chemical Biology* **10**, 1000-1002 (2014).

28. A. C. M. Ferreon, J. C. Ferreon, P. E. Wright, A. A. Deniz, Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390-394 (2013).

29. L. Staby *et al.*, Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochemical Journal* **474**, 2509-2532 (2017).

30. A. L. Turner *et al.*, Highly disordered histone H1−DNA model complexes and their condensates. *Proceedings of the National Academy of Sciences* **115**, 11964 (2018).

31. R. K. Das, Y. Huang, A. H. Phillips, R. W. Kriwacki, R. V. Pappu, Cryptic Sequence Features Within the Disordered Protein p27Kip1 Regulate Cell Cycle Signaling. *Proceedings of the National Academy of Sciences USA* **113**, 5616-5621 (2016).

32. K. P. Sherry, R. K. Das, R. V. Pappu, D. Barrick, Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proceedings of the National Academy of Sciences* **114**, E9243-E9252 (2017).

33. M. V. Staller *et al.*, A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Systems* **6**, 444-455.e446 (2018).

34. C. L. Cuevas-Velazquez *et al.*, Intrinsically disordered protein biosensor tracks the physical-chemical effects of osmotic stress on cells. *Nature communications* **12**, 5438 (2021).

35. R. Vancraenenbroeck, Y. S. Harel, W. Zheng, H. Hofmann, Polymer effects modulate binding affinities in disordered proteins. *Proceedings of the National Academy of Sciences* **116**, 19506 (2019).

36. L. Y. Beh, L. J. Colwell, N. J. Francis, A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proceedings of the National Academy of Sciences* **109**, E1063 (2012).

37. M. C. Cohan, A. M. P. Eddelbuettel, P. A. Levin, R. V. Pappu, Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ. *J Mol Biol* **432**, 3205-3221 (2020).

38. S. Banjade *et al.*, Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proceedings of the National Academy of Sciences* **112**, E6426 (2015).

39. S. Müller-Späth *et al.*, Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **107**, 14609 (2010).

40. J. A. Marsh, J. D. Forman-Kay, Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophysical Journal* **98**, 2383-2390 (2010).

41. F. Wiggers *et al.*, Diffusion of a disordered protein on its folded ligand. *Proceedings of the National Academy of Sciences* **118**, e2106690118 (2021).

42. M. J. Fossat, X. Zeng, R. V. Pappu, 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).

43. R. K. Das, R. V. Pappu, 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).

44. L. Sawle, K. Ghosh, A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *The Journal of Chemical Physics* **143**, 085101 (2015).

45. D. Srivastava, M. Muthukumar, Sequence Dependence of Conformations of Polyampholytes. *Macromolecules* **29**, 2324-2326 (1996).

46. A. Bremer *et al.*, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature Chemistry* **14**, 196-207 (2021).

47. A. Bah, J. D. Forman-Kay, Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *The Journal of biological Chemistry* **291**, 6696-6705 (2016).

48. K. K. Kristensen *et al.*, A disordered acidic domain in GPIHBP1 harboring a sulfated tyrosine regulates lipoprotein lipase. *Proceedings of the National Academy of Sciences* **115**, E6020 (2018).

49. P. R. Banerjee, D. M. Mitrea, R. W. Kriwacki, A. A. Deniz, Asymmetric Modulation of Protein Order–Disorder Transitions by Phosphorylation and Partner Binding. *Angewandte Chemie International Edition* **55**, 1675-1679 (2016).

50. E. W. Martin *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).

51. F. Jin, F. Gräter, How multisite phosphorylation impacts the conformations of intrinsically disordered proteins. *PLOS Computational Biology* **17**, e1008939 (2021).

52. B. I. M. Wicky, S. L. Shammas, J. Clarke, Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proceedings of the National Academy of Sciences* **114**, 9882 (2017).

53. M. J. Fossat, A. E. Posey, R. V. Pappu, Quantifying charge state heterogeneity for proteins with multiple ionizable residues. *Biophysical Journal* **120**, 5438-5453 (2021).

54. T. Firman, K. Ghosh, Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *The Journal of chemical physics* **148**, 123305 (2018).

55. R. Beveridge *et al.*, Ion mobility mass spectrometry uncovers the impact of the patterning of oppositely charged residues on the conformational distributions of intrinsically disordered proteins. *Journal of the American Chemical Society* **141**, 4908-4918 (2019).

56. A. S. Holehouse, R. V. Pappu, Collapse Transitions of Proteins and the Interplay Among Backbone, Sidechain, and Solvent Interactions. *Annual Review of Biophysics* **47**, 19-39 (2018).

57. C. S. Sørensen, M. Kjaergaard, Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proceedings of the National Academy of Sciences* **116**, 23124 (2019).

58. A. Vitalis, R. V. Pappu, A simple molecular mechanics integrator in mixed rigid body and dihedral angle space. *The Journal of Chemical Physics* **141**, 034105 (2014).

59. A. Vitalis, R. V. Pappu, ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry* **30**, 673-699 (2009).

60. A. Radhakrishnan, A. Vitalis, A. H. Mao, A. T. Steffen, R. V. Pappu, Improved Atomistic Monte Carlo Simulations Demonstrate That Poly-l-Proline Adopts Heterogeneous Ensembles of Conformations of Semi-Rigid Segments Interrupted by Kinks. *The Journal of Physical Chemistry B* **116**, 6862-6871 (2012).

61. J.-M. Choi, R. V. Pappu, Improvements to the ABSINTH Force Field for Proteins Based on Experimentally Derived Amino Acid Specific Backbone Conformational Statistics. *Journal of Chemical Theory and Computation* **15**, 1367-1382 (2019).

62. X. Zeng *et al.*, Design of intrinsically disordered proteins that undergo phase transitions with lower critical solution temperatures. *APL Materials* **9**, 021119 (2021).

63. K. M. Ruff, "Predicting Conformational Properties of Intrinsically Disordered Proteins from Sequence" in Intrinsically Disordered Proteins: Methods and Protocols, B. B. Kragelund, K. Skriver, Eds. (Springer US, New York, NY, 2020), 10.1007/978-1-0716-0524-0_18, pp. 347-389.

64. M. O. Steinhauser, A molecular dynamics study on universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties. *The Journal of Chemical Physics* **122**, 094901 (2005).

65. P. G. De Gennes, Collapse of a polymer chain in poor solvents. *J. Physique Lett.* **36**, 55-57 (1975).

66. P. G. de Gennes, Collapse of a flexible polymer chain II. *J. Physique Lett.* **39**, 299-301 (1978).

67. M. Muthukumar, Collapse transition of a stiff chain. *The Journal of Chemical Physics* **81**, 6272-6276 (1984).

68. M. Muthukumar, Theory of counter-ion condensation on flexible polyelectrolytes: Adsorption mechanism. *The Journal of Chemical Physics* **120**, 9343-9350 (2004).

69. J. Huihui, K. Ghosh, An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *The Journal of Chemical Physics* **152**, 161102 (2020).

70. P. G. Higgs, J. F. Joanny, Theory of polyampholyte solutions. *The Journal of Chemical Physics* **94**, 1543-1554 (1991).

71. Y. Kantor, M. Kardar, Polymers with Random Self-Interactions. *Europhysics Letters (EPL)* **14**, 421-426 (1991).

72. Y. Kantor, H. Li, M. Kardar, Conformations of polyampholytes. *Physical Review Letters* **69**, 61-64 (1992).

73. A. V. Dobrynin, M. Rubinstein, Flory Theory of a Polyampholyte Chain. *J. Phys. II France* **5**, 677-695 (1995).

74. S. B. Moldakarimov, E. Y. Kramarenko, A. R. Khokhlov, S. E. Kudaibergenov, Formation of Salt Bonds in Polyampholyte Chains. *Macromolecular Theory and Simulations* **10**, 780-788 (2001).

75. Y. Kantor, M. Kardar, Instabilities of charged polyampholytes. *Physical Review E* **51**, 1299-1312 (1995).

76. Y. Kantor, M. Kardar, Excess Charge in Polyampholytes. *Europhysics Letters (EPL)* **27**, 643-648 (1994).

77. A. M. Gutin, E. I. Shakhnovich, Effect of a net charge on the conformation of polyampholytes. *Physical Review E* **50**, R3322-R3325 (1994).

78. R. Wuttke *et al.*, Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proceedings of the National Academy of Sciences* **111**, 5213 (2014).

79. H. Hofmann *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).

80. M. F. Nüesch *et al.*, Single-molecule Detection of Ultrafast Biomolecular Dynamics with Nanophotonics. *Journal of the American Chemical Society* **144**, 52-56 (2022).

81. K. A. Dill, D. Shortle, DENATURED STATES OF PROTEINS. *Annual Review of Biochemistry* **60**, 795-825 (1991).

82. A. Banks, S. Qin, K. L. Weiss, C. B. Stanley, H.-X. Zhou, Intrinsically Disordered Protein Exhibits Both Compaction and Expansion under Macromolecular Crowding. *Biophysical Journal* **114**, 1067-1079 (2018).

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

84. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38 (1996).

85. R. Kumar, G. H. Fredrickson, Theory of polyzwitterion conformations. *The Journal of Chemical Physics* **131**, 104901 (2009).

86. J. A. Greig *et al.*, Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Molecular cell* **77**, 1237-1250.e1234 (2020).

87. J. F. Leszczynski, G. D. Rose, Loops in Globular Proteins: A Novel Category of Secondary Structure. *Science* **234**, 849-855 (1986).

88. A. Y. Grosberg, D. V. Kuznetsov, Quantitative theory of the globule-to-coil transition. 2. Density-density correlation in a globule and the hydrodynamic radius of a macromolecule. *Macromolecules* **25**, 1980-1990 (1992).

89. K. Ghosh, S. B. Ozkan, K. A. Dill, The Ultimate Speed Limit to Protein Folding Is Conformational Searching. *Journal of the American Chemical Society* **129**, 11920-11927 (2007).

90. J. P. Brady *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 (2017).

91. Timothy J. Nott *et al.*, Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular cell* **57**, 936-947 (2015).

92. J. Wang *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e616 (2018).

93. G. Tesei, T. K. Schulze, R. Crehuet, K. Lindorff-Larsen, 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**, e2111696118 (2021).

94. M. J. Fossat, R. V. Pappu, q-Canonical Monte Carlo Sampling for Modeling the Linkage between Charge Regulation and Conformational Equilibria of Peptides. *The Journal of Physical Chemistry B* **123**, 6952-6967 (2019).

95. M. S. Li, D. K. Klimov, J. E. Straub, D. Thirumalai, Probing the mechanisms of fibril formation using lattice models. *The Journal of Chemical Physics* **129**, 175101 (2008).

96. M. S. Li *et al.*, Factors Governing Fibrillogenesis of Polypeptide Chains Revealed by Lattice Models. *Physical Review Letters* **105**, 218101 (2010).

97. M. E. Fisher, M. N. Barber, Scaling Theory for Finite-Size Effects in the Critical Region. *Physical Review Letters* **28**, 1516-1519 (1972).

98. B. Dünweg, D. Reith, M. Steinhauser, K. Kremer, Corrections to scaling in the hydrodynamic properties of dilute polymer solutions. *The Journal of Chemical Physics* **117**, 914-924 (2002).

99. A. S. Holehouse, S. Sukenik, Controlling Structural Bias in Intrinsically Disordered Proteins Using Solution Space Scanning. *Journal of Chemical Theory and Computation* **16**, 1794-1805 (2020).

100. D. Moses *et al.*, Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *The Journal of Physical Chemistry Letters* **11**, 10131-10136 (2020).

101. L. Staby *et al.*, Flanking Disorder of the Folded αα-Hub Domain from Radical Induced Cell Death1 Affects Transcription Factor Binding by Ensemble Redistribution. *Journal of Molecular Biology* **433**, 167320 (2021).

102. K. Bugge *et al.*, Interactions by Disorder – A Matter of Context. *Frontiers in Molecular Biosciences* **7** (2020).

103. A. Mittal, A. S. Holehouse, M. C. Cohan, R. V. Pappu, Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *Journal of Molecular Biology* **430**, 2403-2421 (2018).

104. T. Chen, J. Song, H. S. Chan, Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Current opinion in structural biology* **30**, 32-42 (2015).

105. B. Szabo *et al.*, Intrinsically Disordered Linkers Impart Processivity on Enzymes by Spatial Confinement of Binding Domains. *International Journal of Molecular Sciences* **20**, 2119 (2019).

106. I. Taneja, A. S. Holehouse, Folded domain charge properties influence the conformational behavior of disordered tails. *Current Research in Structural Biology* **3**, 216-228 (2021).

107. A. G. Kozlov, M. K. Shinn, E. A. Weiland, T. M. Lohman, Glutamate promotes SSB protein-protein Interactions via intrinsically disordered regions. *J Mol Biol* **429**, 2790-2801 (2017).

108. P. O. Heidarsson *et al.*, Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. *Nature Chemistry* **10.1038/s41557-021-00839-3** (2022).

109. L.-W. Chang *et al.*, Sequence and entropy-based control of complex coacervates. *Nature communications* **8**, 1273 (2017).

110. Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Physical Review Letters* **117**, 178101 (2016).