This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2019, 5, 821−830

http://pubs.acs.org/journal/acscii

Temperature-Controlled Liquid−Liquid Phase Separation of Disordered Proteins

Gregory L. Dignon,[†,∥] Wenwei Zheng,[‡,∥] Young C. Kim,[§] and Jeetain Mittal*[,†]

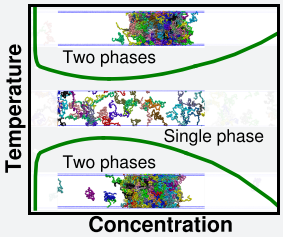
[†]Department of Chemical and Biomolecular Engineering, Lehigh University, 111 Research Drive, Bethlehem, Pennsylvania 18015, United States

[‡]College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona 85212, United States

[§]Center for Materials Physics and Technology, Naval Research Laboratory, Washington, D.C. 20375, United States

Supporting Information

ABSTRACT: The liquid−liquid phase separation (LLPS) of intrinsically disordered proteins (IDPs) is a commonly observed phenomenon within the cell, and such condensates are also highly attractive for applications in biomaterials and drug delivery. A better understanding of the sequence-dependent thermoresponsive behavior is of immense interest as it will aid in the design of protein sequences with desirable properties and in the understanding of cellular response to heat stress. In this work, we use a transferable coarse-grained model to directly probe the sequence-dependent thermoresponsive phase behavior of IDPs. To achieve this goal, we develop a unique knowledge-based amino acid potential that accounts for the temperature-dependent effects on solvent-mediated interactions for different types of amino acids. Remarkably, we are able to distinguish between more than 35 IDPs with upper or lower critical solution temperatures at experimental conditions, thus providing direct evidence that incorporating the temperature-dependent solvent-mediated interactions to IDP assemblies can capture the difference in the shape of the resulting phase diagrams. Given the success of the model in predicting experimental behavior, we use it as a high-throughput screening framework to scan through millions of disordered sequences to characterize the composition dependence of protein phase separation.



INTRODUCTION

It is now well recognized that cellular compartments may form in the absence of lipid membranes through liquid−liquid phase separation (LLPS), driven by proteins, nucleic acids, and other biomolecules.[1,2] These "membraneless organelles" or "biomolecular condensates" have since been shown to be highly diverse and ubiquitous within biological systems and constitute organelles such as the nucleolus,[3,4] ribonucleoprotein (RNP) granules,[5,6] stress granules,[7,8] and many others.[1,9−11] Protein LLPS is commonly associated with proteins containing regions that are intrinsically disordered[12,13] and is mediated by a myriad of interaction types such as electrostatic attraction, cation−π, π−π, hydrogen bonding, and hydrophobic interactions.[14−18] External stimuli such as changes in salt concentration, pH, other biomolecules such as RNA or ATP, and temperature are all factors that may be used to modulate protein LLPS.[5,8,17,19−21]

Intrinsically disordered protein-based polymers have been used for decades in the design of functional polymeric materials for applications in biomaterials and drug delivery.[22−29] The advantages of protein-based polymers include direct control over the sequence and length by using recombinant expression[24] and the ability to directly encode functional domains such as enzyme-cleavage sites,[30] light-activated domains,[31,32] cross-linking motifs,[33] and substrate-specific binding motifs.[30] The high degree of control over the protein sequence also allows one to finely tune the protein LLPS in response to solution conditions and external stimuli.[34]

As temperature is a factor that is easily controlled in vitro, there is a large interest in thermoresponsive LLPS.[35,36] Thermoresponsive protein-based polymers can be designed such that they are miscible at high temperatures and demix at low temperatures, showing an upper critical solution temperature (UCST), or such that they demix at high temperatures and are miscible at lower temperatures, displaying a lower critical solution temperature (LCST).[37] Tropoelastin and resilin are two proteins which are commonly used as templates to design elastin-like and resilin-like peptides (ELPs and RLPs) exhibiting LCST and UCST phase behaviors, respectively.[37] Some variants of RLPs have also been shown to exhibit a dual-response phase separation and will condense upon both heating and cooling, with a region of miscibility in between.[38,39] The amino acid composition and sequence have been implicated as being responsible for the differences in phase behavior.[37]

Designing intrinsically disordered proteins (IDPs) with controllable LLPS is a nontrivial task due to the near-infinite selection of possible IDP sequences. Computational modeling can be an effective approach to inform experimental design and to gain insights about the sequence-determinants of temperature-dependent LLPS and the underlying physics.[40,41] Temperature-dependent amino acid properties have previously been used in understanding properties of both folded and

Received: February 1, 2019
Published: May 1, 2019

Downloaded via 207.241.233.164 on October 30, 2020 at 13:13:32 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

unfolded/disordered proteins including cold denaturation[42] and temperature-induced collapse.[43] All-atom explicit-solvent simulations can in principle provide an atomically detailed view of the interactions driving phase separation[15,44] but are computationally demanding and prohibitive to the direct simulation of protein LLPS. To overcome this obstacle, coarse-grained (CG) models, in which amino acids are simplified into CG beads, and solvent is accounted for implicitly, can be used to handle sufficiently large systems and to compute phase behavior of a large number of protein sequences.[45,46] In these cases, the solvent-mediated interactions are indirectly captured via interactions between the CG beads composing the protein molecules.[46−49] Most existing CG models were built at a specific temperature (e.g., room temperature)[47,49,50] without taking into account the temperature dependence of such solvent-mediated interactions.[51−55] These models are not able to capture properties like disordered protein collapse with increasing temperature[43,56] and the emergence of LCST behavior.[35,37] Therefore, there is an urgent need for a temperature-dependent CG model, given the compelling prospect of using IDP LLPS in designing thermoresponsive materials.

In this paper, we take advantage of our recently developed CG model in which amino acid hydrophobicity was used in modeling the pairwise interactions between different amino acids[49] and introduce an amino-acid-type-specific temperature dependence into the hydrophobicity scale. We then tune the model parameters using knowledge from both single molecule Förster Resonance Energy Transfer (smFRET) experiment and all-atom simulations on the dimensions of disordered proteins across a wide range of temperatures. The optimized model successfully predicts the experimentally known phase behavior of a large library of ELPs and RLPs qualitatively by distinguishing between UCST and LCST. This strongly suggests that the difference in the thermoresponsive behavior of a protein sequence is encoded in its amino-acid-specific solvent-mediated interactions and how these change with temperature. Using this newfound knowledge, we apply the new model to propose sequence-determinants of the protein LLPS in terms of their UCST or LCST characteristics, which should allow for the design of protein-based polymers with controllable thermoresponsive phase behavior.

■ RESULTS AND DISCUSSION

Using Amino Acid Contact Potential and IDP Size as a Function of Temperature To Tune Solvent-Mediated Interactions. Given our recent success using amino acid resolution CG models to study IDP LLPS,[45,49] we apply a similar philosophy to build a model based on amino acid hydrophobicity with temperature-dependent interactions. We previously used a top-down approach to parametrize the CG model to reproduce experimental measurables using the radius of gyration (R_g) of a large number of IDPs[49] at a single temperature (300 K). Building a model that accurately captures temperature-dependent IDP dimensions, and therefore LLPS, necessitates a set of data for IDPs spanning a range of temperatures. The temperature-induced expansion and collapse of a diverse set of IDPs and denatured proteins have been observed previously in smFRET experiments[43] and all-atom simulations.[56] The proteins examined in these studies include the cold-shock protein from *Thermotoga maritima* (CspTm), the N-terminal domain of HIV integrase, the DNA binding domain of λ-repressor, and the N- and C-terminal segments of human Prothymosin-α (ProTαN and ProTαC). For protein sequences, please refer to the Supporting Methods 1.1 section in the Supporting Information. These two data sets are complementary as the experimental study is limited to a smaller temperature range, whereas the all-atom results were obtained using an older force field which results in protein dimensions smaller than expected.[57] Here, we merge the two data sets to a single reference data set to take into account the desirable features of each, i.e., the wider temperature range from simulation and the quantitative accuracy of experiment. More details can be found in the Supporting Methods 1.2 section and Figure S1.

The amino acid composition of a protein is largely responsible for the differences in the observed phase behavior.[37] To account for the sequence-dependent behavior of proteins, we aim to develop a physics-based CG model which can capture amino acid residue-specific changes induced by temperature. Van Dijk et al. used a library of solved protein structures at different temperatures to build a knowledge-based contact potential as a function of temperature between protein residues.[58] They used a reduced classification by lumping the temperature dependence of all 20 amino acids into five different types, generally having similar responses to temperature within each group (Table 1). We note that the temperature dependence from this knowledge-based potential is also consistent with the changes in solvation free energy of amino acid side chain analogues as a function of temperature.[51−53,59] The resulting potential was successfully used to obtain the estimates of protein thermal stability[60] and to explain protein cold denaturation by invoking the changes in solvent-mediated interactions with temperature.[42] For further use, we fit the temperature-dependent contact potential to a parabolic function for each amino acid type (Supporting Methods 1.3 in the Supporting Information). At low temperatures, the interactions between hydrophobic groups will strengthen with increasing temperature, whereas these interactions will weaken with a further increase in temperature after a point of maximum strength. This behavior arises from the dominance of the enthalpic component of the free energy at low temperatures and its entropic part at higher temperatures.[51,59,61] The parabolic functional forms fit the bioinformatic contact potential[58] within a small temperature range and allow us to extrapolate to a wider range of temperatures relevant to the experimental studies of thermoresponsive phase separation. We anticipate this is a reasonable assumption to capture the qualitative changes in single chain compaction and phase behavior.

Table 1. List of Amino Acids and Type Classifications

hydrophobic (H)	Ala, Ile, Leu, Met, Val
aromatic (A)	His, Phe, Trp, Tyr
other (O)	Cys, Gly, Pro
polar (P)	Asn, Gln, Ser, Thr
charged (C)	Arg, Asp, Glu, Lys

The nonbonded interactions in our original CG model are based on the hydrophobicity (λ) of each amino acid (see the Methods section).[49] Therefore, the contact potential described above can be used to introduce temperature dependence on λ in the model by an appropriate scale as

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

Table 2. Optimized Parameters for the 3-Parameter and 15-Parameter Model

parameter	hydrophobic	aromatic	other	polar	charged	3-parameter model
α ($k_B T^{-1}$)	0.995	2.0	2.0	2.0	2.0	0.7836
T_{ref} (K)	250.0	308.8	250.0	253.6	250.0	296.7
T_{shift} (K)	97.07	48.72	100.0	100.0	49.24	61.97

$$\lambda_i(T) = \lambda_{i,\mathrm{HPS}} + \frac{1}{\epsilon}[E_X(T) - E_X(T_{\mathrm{ref}})] \quad (1)$$

where i is the amino acid, X is the amino acid type, E_X is the corresponding parabolic function from eq S2, ϵ is 0.337 $k_B T$ (0.2 kcal/mol), as in the original HPS model[49] (see the Methods section) to convert to the correct units for use with the LJ-like functional form, T_{ref} is the reference temperature (300 K) for which the model will be equal to the original HPS model, and $\lambda_{i,\mathrm{HPS}}$ is the hydrophobicity value for residue i in the original HPS model (Table S1). To test the new model, we simulate the five proteins for which the radius of gyration is available as a function of temperature from experiment and all-atom simulations.[43,56] In contrast to the original HPS model, we are able to observe a nonmonotonic trend in R_g as seen in experiment (Figure S2), although it is in poor qualitative agreement. Specifically, CspTm, integrase, and λ-respressor do initially collapse and then expand; however, the turning point of R_g is at about 300 K instead of about 350 K seen in the reference data set, which suggests that allowing for the shifting of the contact potential extrema is necessary for further refinements.

To better capture the reference data, we modified our approach to empirically define a temperature-dependent model which quantitatively agrees with the reference data by making T_{ref} a free parameter and introducing two additional free parameters as the prefactor (α) and a shift along the temperature axis (T_{shift}) into the function:

$$\lambda_i(T) = \lambda_{i,\mathrm{HPS}} + \alpha[E_X(T - T_{\mathrm{shift}}) - E_X(T_{\mathrm{ref}} - T_{\mathrm{shift}})] \quad (2)$$

To find the optimal parameters for eq 2 with minimized deviation from the reference data, we need a way to estimate the R_g from our CG model in an efficient manner. Toward this goal, we use a homopolymer-based predictor which can be used to quickly calculate the R_g for a specific protein sequence on the basis of its chain length and average hydrophobicity (see the Methods section). Using a standard global optimization method,[62] we minimize the difference between the predicted R_g and that from the reference data set. We optimize two different models, one where the free parameters are allowed to vary for all five amino acid types (Table 1), yielding a 15-dimensional optimization, and the other where the free parameters are made universal for the different amino acid types, yielding a 3-dimensional optimization problem. Optimized parameters for both models are listed in Table 2. We find that the 3-parameter model is sufficient to achieve good agreement with the reference set (Figure 1). By allowing parameters for each amino acid type to vary independently in the 15-parameter model, we are able to match the empirical predictions to the reference data very well, while imposing the following constraints on the parameters: $0 < \alpha < 2$; $250 < T_{ref} < 350$; $-100 < T_{shift} < 100$, to search through the physically meaningful parameter space. While the empirical R_g predictions become more similar to the reference data, we find that results from simulations are actually less accurate than the 3-parameter model (Figure S3). We believe this is due to the homopolymer-based predictor not accounting for the greater heterogeneity of interaction strengths within this model.

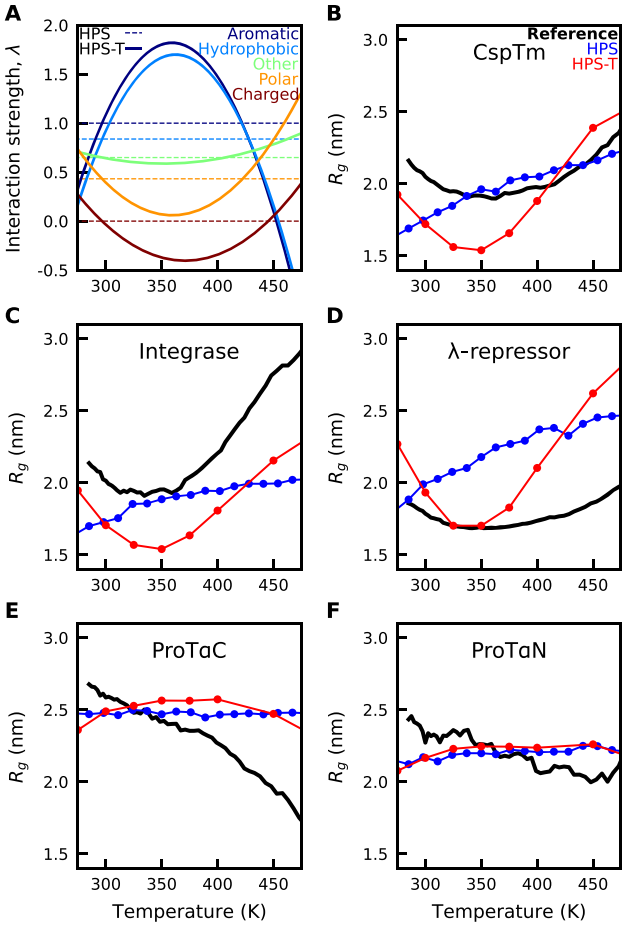


Figure 1. Temperature-dependent interaction potential and protein dimensions. (A) Original λ values from HPS potential are shown as dashed lines, and the new temperature-dependent model is shown as solid lines. Example HPS values are shown for phenylalanine (aromatic), methionine (hydrophobic), glycine (other), asparagine (polar), and arginine (charged). (B−F) Experimental/all-atom radius of gyration data for 5 proteins used to fit the temperature-dependent (HPS-T) model.

We also consider the use of a physics-based model from Dill, Alonso, and Hutchinson[63] as suggested recently by Lin, Forman-Kay, and Chan.[64] Upon implementing this temperature dependence into the HPS model (see the Supporting Methods 1.4 section in the Supporting Information), we find reasonable agreement with the training set from predictions and simulations for the first three sequences, but the collapse of ProTαC and ProTαN is not observed (Figure S4), because this model does not capture the temperature dependence of hydrophilic amino acid interactions. Thus, we conclude that the use of the bioinformatics-based temperature dependence from van Dijk is advantageous in its ability to capture the temperature dependence of all types of amino acids. Future

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

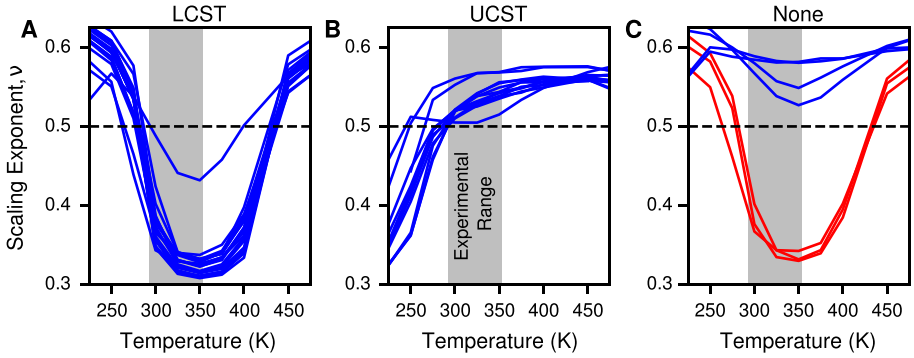


Figure 2. Experimental verification of the model with QC sequences. (A) Simulated ν as a function of T for QC sequences with experimental LCST behavior. (B) QC sequences with experimental UCST behavior. (C) QC sequences without phase behavior in experiment. The red lines show QC3, QC6, and QC7 with crossing points to ν = 0.5 in simulations. The gray bar shows the range of temperature scanned by the experiment.[37]

work, however, may focus on a more fundamental approach to estimate temperature dependence and even pressure dependence of pairwise interactions from all-atom explicit-solvent simulations.[65]

To demonstrate that the use of temperatures outside the realm of the experimentally realizable range in the reference data is not negatively affecting the model, we also conducted the optimization using only temperature points below 400 K. We find that using this truncated temperature range results in a nearly identical set of parameters as the full reference data set (Figure S5). The use of a scaling parameter (eq S1) to create the reference set may also modify the results due to the magnitude of R_g variation in response to a change in temperature. To assess this effect, we created a separate reference data set for the five test proteins, by setting the scaling parameter equal to 1 (see the Supporting Methods 1.2 section in the Supporting Information). Optimizing to this reference set results in a similar model to the initial 3-parameter model, with a somewhat weaker temperature dependence (Figure S6). We additionally attempt to fit directly to the experimental data to avoid having to combine with simulation data but find that the even more limited range of temperatures does not account for the full shape of the temperature dependence of R_g (Figure S7). Thus, using only the limited experimental data available is likely not suitable for describing the thermoresponsive behavior of phase separating IDPs.

The resulting temperature-dependent interaction strengths from the 3-parameter CG model are shown in Figure 1A. The simplified functional forms for the temperature dependence can be found in the Methods section and in the Supporting Methods 1.5 section in the Supporting Information. Hereafter, we refer to this temperature-dependent hydrophobicity-based model (3-parameter model in Table 2) as the HPS-T model. Given our previous work, intramolecular interactions driving single chain collapse and intermolecular interactions driving LLPS are fundamentally related;[45] thus, we expect this model should be sufficient to capture the thermoresponsive phase behavior of IDPs.

Temperature-Dependent Solvent-Mediated Interactions Can Help Distinguish between UCST and LCST Proteins. Garcia-Quiroz and Chilkoti synthesized a large number of low-complexity IDPs mimicking the short, repetitive amino acid motif characteristic of tropoelastins and resilins, with a highly diverse range of amino acid compositions.[37] Through this, they provided an excellent characterization of how amino acid composition can influence the thermoresponsive protein phase behavior.[37] They found that RLPs are generally composed of charged and polar amino acids and show UCST behavior, while ELPs tend to contain more hydrophobic amino acids and exhibit LCST behavior. Due to the large number of sequences spanning a wide range of amino acid compositions and the direct observance of thermoresponsive phase behavior, this data set is ideal for testing the applicability of the HPS-T model. We classified the 39 sequences, termed QC sequences, into three groups: LCST, UCST, and no phase separation (Table S2).

Since it is impractical to conduct coexistence simulations for all 39 sequences in the QC data set, we take advantage of a recently suggested correlation between the critical temperature T_c (which separates the two-phase region from the single-phase region in the phase diagram) and the Θ temperature (T_θ) (the temperature at which the Flory scaling exponent, ν, is equal to 0.5[45]). For conditions where ν < 0.5, the effective intrachain or interchain interactions are attractive causing chain collapse or phase separation, whereas ν values larger than 0.5 imply that repulsive interactions are dominant, causing chain expansion and inability to phase separate. One can also approximately calculate T_c from the Boyle temperature (T_B), the temperature at which the osmotic second virial coefficient (B_{22}) goes to zero. We found these relationships to be non-model-specific as they were identified using two different potential energy functions[45] and therefore should, in principle, be able to predict both UCST and LCST behavior from T_θ or T_B in the new HPS-T model.

In Figure 2, we present the Flory scaling exponent (ν) for each of the QC sequences for a wide temperature range to determine what conditions will allow for phase separation and to predict whether these will display UCST or LCST behavior. For the first set of QC sequences (LCST), we first observe chain collapse (decreasing ν) with increasing temperature and a subsequent expansion (increasing ν) from our simulation, with the initial collapse occurring at the range of temperatures tested in experiment (Figure 2A). For the second set (UCST), an initial chain expansion is followed by collapse highlighting UCST behavior within the experimental temperature range (Figure 2B). Most of these sequences show a dual-response phase behavior with two crossing points, which was not observed in experiment. In general, the two T_θ values are quite far apart which would be difficult to observe in experiment

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

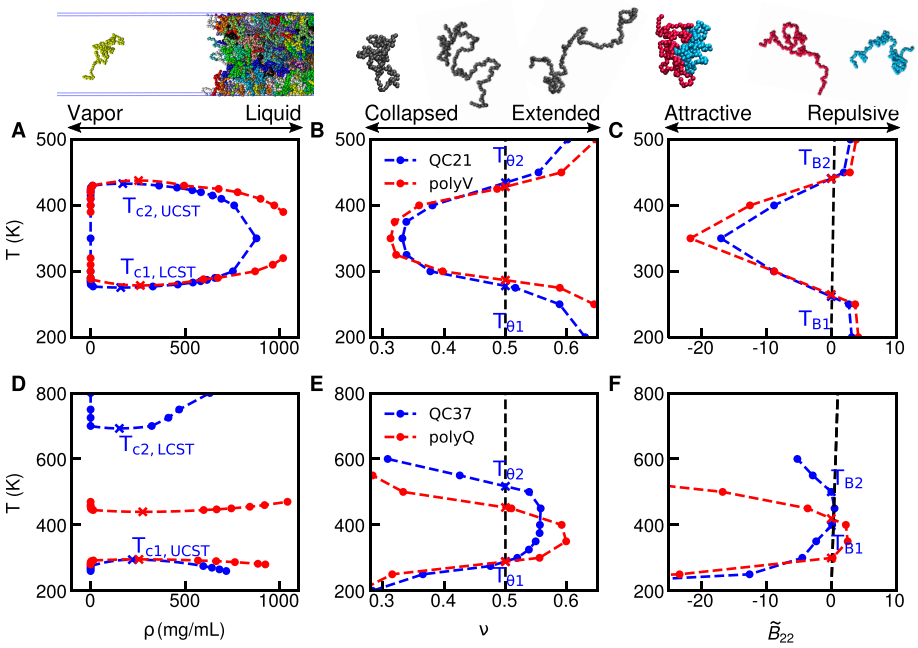


Figure 3. Dual-phase behavior of IDP sequences. (A) Temperature-dependent ν, (B) B_{22}, and (C) phase coexistence of a hydrophobic homopolymer (V_{50}) and an elastin-like LCST sequence from Garcia-Quiroz et al.[37] (D) Temperature-dependent ν, (E) B_{22}, and (F) phase coexistence of a hydrophilic homopolymer (Q_{50}) and a resilin-like UCST sequence from Garcia-Quiroz et al.[37]

without making other perturbations to the system. Thus, the experimental studies would only observe the single phase transition we see near the experimental range.

For the third group of QC sequences which were shown to not phase separate in experiment, we find that the ν values for four of the seven proteins never decrease below 0.5, suggesting that these particular proteins are not expected to phase separate within the broad temperature range tested (Figure 2C). The three discrepant sequences are all in the set of proteins mimicking the content of elastin, for which simulation results predict LCST behavior at experimental conditions as predicted for the majority of the other proteins in that set (Figure 2A). The simulation results are at odds with the experimentally documented behavior for these three proteins, which raises questions about the general validity of the HPS-T model despite its strong predictive capabilities for 36 out of 39 proteins. A careful look at these protein sequences highlights similarities with other sequences in the QC data set, some of which are nearly identical to QC3, QC6, and QC7 in composition. As these analogous sequences (QC2 ∼ QC3, QC4/5 ≡ QC6, and QC9 ∼ QC7 in Table S2) show LCST behavior, the predictions of the model are not entirely unreasonable. Another possibility is that the experimental temperature range is not sufficiently broad to induce phase separation at relatively low concentrations, and our results may provide impetus to revisit these experiments in the future.

We also note that use of the Dill−Alonso−Hutchinson model captures the sequences which undergo LCST but not those with UCST, supporting our assertion that temperature-dependent interactions between polar amino acids must also be accounted for (Figure S8).

Reentrant Phase Behavior as a Function of Temperature. The qualitative agreement of the HPS-T model and experimental results indicates it is a promising approach to directly study the LLPS of proteins undergoing UCST and LCST. It is therefore instructive to ask if the UCST versus LCST phase behavior predicted by changes in ν as a function of temperature due to solvent-mediated interactions is present in the thermodynamic phase diagram. The appearance of different phases as a function of temperature in a multiprotein simulation will also allow one to probe the differences in the molecular structure and dynamics directly from the simulated trajectories. We select one QC sequence from each of the first two groups (QC21 and QC37) and conduct slab coexistence simulations to obtain the thermodynamic phase diagram as well as two-chain simulations to determine B_{22} at a range of temperatures and to estimate T_B following the same protocol as in previous work.[45,49]

QC21 exhibits a dual-response phase behavior described by an LCST at 275 K and a UCST at 432 K, with a region in between where LLPS is observed, having the shape of a closed-loop phase diagram (Figure 3A).[36,66,67] The closed-loop phase diagram is analogous to the predicted cold denaturation of folded proteins which unfold at both extreme high and low temperatures.[42] This observed phase behavior is also qualitatively consistent with the collapse and expansion of a single protein chain with increasing temperature (Figure 3B) as well as with the preference for intermolecular attraction ($B_{22} < 0$) and repulsion ($B_{22} > 0$) between two proteins as a function of temperature (Figure 3C). Moreover, there is a good correspondence between the different transition temperatures that can be identified from Figure 3A−C. This suggests that the previously proposed correlations[45] as well as the homopolymer predictor model used to fine-tune the CG model parameters can be accurate enough for future use.

QC37, on the other hand, phase separates at low temperatures and is miscible at temperatures above 294 K. With a further increase in temperature to a very high value (692 K), which is not physically meaningful, the system shows a reentrant behavior by phase separating again into two phases (Figure 3D). Such dual-responsive phase behavior has been observed experimentally for various RLPs such as Rec1[38] and

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

An16 resilin[39] within temperature ranges accessible to experiment. The qualitative behavior observed from the other two transition temperatures based on protein collapse (Figure 3E) and intermolecular interactions between a pair of protein molecules (Figure 3F) is also consistent with this phase diagram. However, only the lower transition temperatures (T_{c1}, $T_{\theta 1}$, and T_{B1}) are in quantitative agreement with each other, while the LCST (T_{c2}), is significantly higher than $T_{\theta 2}$ or T_{B2} (Figure 3D−F).

A closer examination of the QC37 multiprotein system between temperatures $T_{\theta 2}$ and T_{c2} suggests that these proteins prefer to form intramolecular contacts (leading to collapsed globular conformations) as opposed to the intermolecular contacts required to stabilize a condensed protein-rich phase (Figure S9). A possible explanation for this behavior is the relative importance of enthalpy and entropy in the free energy of the system. We hypothesize that the entropic cost of incorporating protein chains into a condensed phase, which is not appropriately accounted for in a single chain simulation to estimate T_θ, becomes more important at higher temperatures. The system free energy is thus minimized through maximizing intramolecular contacts by forming collapsed globules and maximizing the system entropy by keeping the proteins dispersed in a larger volume. If this is indeed the case, one would expect the proteins to adopt conformations such that hydrophobic residues are deeply buried inside, and the protein surface is more hydrophilic and therefore less likely to form favorable contacts with other proteins. Indeed, we find that a single QC37 chain will isolate the more hydrophobic amino acids toward the center of the globule, while the more repulsive/hydrophilic residues occupy the surrounding region at high temperatures (Figure S10). Considering the average and standard deviation of λ values for the amino acids in the QC sequences, we see that the variation between different amino acids is much higher for QC37 at $T_{\theta 2}$ than it is for $T_{\theta 1}$ or either T_θ of QC21 (Figure S11), thus facilitating the collapse of more attractive amino acids to the center with repulsive residues at the exterior.

A simple test to determine whether the variation of attraction and repulsion within an IDP sequence is causing the unfavorability of the LCST phase transition is to simulate a simple homopolymeric protein expected to display a similarly shaped phase diagram. Therefore, we conduct simulations of a polyglutamine (Q_{50}) protein sequence to compute the phase diagram as well as the single-chain and two-chain properties as a function of temperature as shown in Figure 3D−F. Our observed ν value for Q_{50} at room temperature is consistent with the expectation from the work of Singh and Lapidus on polyglutamine peptides,[68] though we do not expect our model to be in perfect agreement with all available experimental data.[69] In this case, we find that all the transition temperatures are in quantitative agreement with each other, and the LCST is also much lower than the heteropolymer QC37 sequence. This suggests that the heterogeneity of the sequence, having large variance in attraction and repulsion, contributes to the breakdown of the general correlations between T_c, T_θ, and T_B.[45,70,71] However, we note that the phase behavior of a polyvaline (V_{50}) sequence expected to have a closed-loop phase diagram is quite similar to the heterogeneous sequence QC21 within this model (Figure 3A−C).

The protein assemblies formed by QC37 at extremely high temperatures resemble solid aggregates when visualized (Figure S9). Interestingly, we find that the diffusion of protein chains within the condensed phase formed above the LCST is significantly slower than within the condensed phases formed below the UCST (Figure S12). This behavior is reminiscent of experimental findings on RLPs which undergo similar reentrant phase transitions upon cooling and heating, having slower recovery from the high-temperature LCST.[38,39] It stands to reason that having few strong interaction sites within a sequence would lead to slower dynamics than having many weaker interaction sites. Thus, we postulate that the variation of attraction and repulsion within a sequence can be used to manipulate the dynamics within the condensed phase, which may be tuned by sequence, and temperature-dependent hydrophobicity.

Role of Amino Acid Composition in the Thermoresponsive Behavior of Disordered Proteins. Given the success of our new HPS-T model in distinguishing UCST from LCST sequences with the help of a simple predictor, we have a unique opportunity to identify the molecular determinants of the temperature-dependent phase behavior of IDPs. We scan a large number of sequences (≈1 million) with the chain length the same as CspTm (66 amino acids) on the basis of the relative abundance of each amino acid in the intrinsically disordered proteome[72] (Figure S13) and compute ν for these proteins as a function of temperature (see the Supporting Methods 1.6 section in the Supporting Information). We can use this information to infer the shape of the phase diagram regarding their transition temperatures, number of such transitions, and their type (UCST or LCST). On the basis of this analysis, we can classify IDP sequences into four groups: none (ν > 0.5 always) without phase behaviors like QC sequences in Figure 2C; single UCST with monotonically decreasing ν when increasing temperature; closed-loop with UCST higher than LCST (Figure 3A); and hourglass with UCST lower than LCST (Figure 3D).

To understand the role of specific amino acids in the marked preference for a given type of phase behavior, we compute the probability of their occurrence with respect to the probability of those amino acids for a typical IDP sequence from a bioinformatics study (Figure S13). As shown in Figure 4 and Table S3, the amino acid probabilities in the types "closed-loop" and "none" are most similar to a typical IDP sequence, whereas an enhanced polar and charged amino acids content would be needed to observe single UCST or hourglass type behavior. These results present a path forward for the design of thermoresponsive materials with tunable properties by changing their amino acid content. However, we caution that the use of empirical predictions may not be directly applicable to all IDPs due to sequence-specific effects such as patterning of charges or hydrophobic regions. Rather, we hope this analysis serves as a demonstration of the possibilities, with more work to follow using direct MD simulations.

■ CONCLUSION

In this study, we provide a direct interrogation of the thermoresponsive phase behavior of IDPs through use of a novel coarse-grained model which explicitly represents the amino acid sequence and accounts for the temperature-dependent solvent-mediated interactions of each type of amino acid. We validate the model using experimental and all-atom data on the R_g of several disordered proteins, as well as the thermoresponsive phase behavior of a large library of designed RLP and ELP sequences. The qualitative capture of the sequence-encoded phase behavior shows promise for the

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

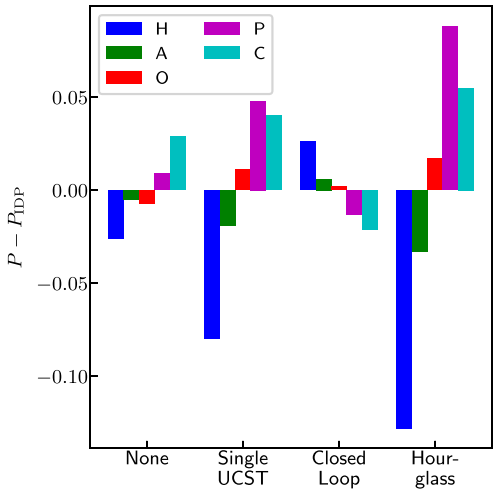


Figure 4. Difference between probabilities of H/A/O/P/C type amino acids (see Table 1) in sequences with a specific phase-diagram shape (labeled on the x-axis) and the probabilities of those in a typical IDP sequence from a bioinformatics study[72] (see Figure S13). The definition of the phase-diagram shape is shown in Table S3. Errors are shown in Table S3 and are not noticeable in the figure.

model to extend to the furthest reaches of the IDP sequence space when coupled with an empirical homopolymer-based predictor. From this, we learn that a typical IDP sequence will undergo phase separation with a closed-loop phase diagram, having LCST at the more physiological conditions. Sequences with an hourglass-shaped phase diagram generally contain more polar or charged residues than a typical IDP sequence.

METHODS

HPS-T Model. The HPS-T model is mostly identical to our original framework, which represents proteins as flexible chains of amino acids with harmonic bonds, screened electrostatics, and a nonbonded pairwise interaction potential to account for different amino acid types.[49] The full energy function of the system is

$$\Phi(\mathbf{r}) = \sum_{\text{bonds}} \Phi_{\text{bond}}(r_{ij}) + \sum_{i<j} [\Phi_{\text{elec}}(r_{ij}) + \Phi_{\text{nb}}(r_{ij})] \tag{3}$$

where Φ_{bond} is a standard harmonic spring:

$$\Phi_{\text{bond}}(r) = k_{\text{spring}}(r - r_0)^2 \tag{4}$$

with k_{spring} = 10 kcal/(mol Å) and r_0 = 3.8 Å. The electrostatic term is represented using Debye−Hückel electrostatic screening:[73]

$$\Phi_{\text{elec}}(r) = \frac{q_i q_j}{4\pi D r} e^{-r/\kappa} \tag{5}$$

where q_i and q_j are the net charges of formally charged amino acids (D, E = −1; K, R = 1; H = 0.5), D is the dielectric constant, which is set to 80 for water, and κ is the screening length, which we set to 10 Å to represent a salt concentration of 100 mM. For the nonbonded pairwise interactions, we utilize a Lennard-Jones-like functional form with a tunable well depth as used by Ashbaugh and Hatch:[50,74]

$$\Phi_{\text{nb}}(r) = \begin{cases} \Phi_{\text{LJ}}(r) + (1 - \lambda(T))\epsilon & \text{if } r \le 2^{1/6}\sigma \\ \lambda(T)\Phi_{\text{LJ}}(r) & \text{otherwise} \end{cases} \tag{6}$$

where Φ_{LJ} is the standard Lennard-Jones potential, and $\lambda(T)$ is the temperature-dependent interaction strength. The finalized HPS-T model uses the optimized set of equations for the temperature dependence of each type of amino acid:

$$\lambda_{i,\text{H}}(T) = \lambda_{i,\text{HPS}} - 25.475 + 0.14537T - 0.00020059T^2 \tag{7a}$$

$$\lambda_{i,\text{A}}(T) = \lambda_{i,\text{HPS}} - 26.189 + 0.15034T - 0.00020920T^2 \tag{7b}$$

$$\lambda_{i,\text{O}}(T) = \lambda_{i,\text{HPS}} + 2.4580 - 0.014330T + 0.000020374T^2 \tag{7c}$$

$$\lambda_{i,\text{P}}(T) = \lambda_{i,\text{HPS}} + 11.795 + 0.067679T + 0.000094114T^2 \tag{7d}$$

$$\lambda_{i,\text{C}}(T) = \lambda_{i,\text{HPS}} + 9.6614 - 0.054260T + 0.000073126T^2 \tag{7e}$$

where i is the amino acid; H, A, O, P, and C correspond to the type of amino acid, which i is included in (see Table 1); and λ_{HPS} is the original value used for λ in the temperature-independent model,[49] adapted from a standard hydrophobicity scale.[75] We originally optimized the free parameter ϵ to 0.2 kcal/mol based on the agreement between the model and experimentally determined radius of gyration (R_g) of a set of IDPs. For further details, please refer to our previous work.[49]

We account for protein−solvent interactions through the protein−protein interaction term as more hydrophobic amino acids will have a stronger attraction, and hydrophilic will be more repulsive. The use of Debye−Hückel screened electrostatics, in addition to a standard nonbonded potential based on the amino acid contact probability, is justified by the expectation that charge−charge interactions are not fully captured in data based on folded protein structures and that attractive and repulsive electrostatic interactions would be averaged for the charged amino acids. Similar approaches have been used extensively in the protein CG modeling literature and have provided accurate information on protein binding thermodynamics and structure.[47]

Molecular Dynamics Simulations. We conduct single chain simulations using the LAMMPS software package[76] and two-chain simulations using an in-house Monte Carlo code[47] with umbrella sampling to enhance sampling of binding and unbinding events.[77] Results from umbrella sampling were reweighted using a weighted histogram analysis method (WHAM).[78] To efficiently sample phase coexistence, we conduct coexistence simulations in slab geometry[49,79,80] using the HOOMD-blue v2.1.5 software package.[81] Single chain simulations were conducted for 1 μs at each temperature, and two-chain simulations were conducted for 5 × 10^7 Monte Carlo steps; coexistence simulations are carried out for up to 5 μs.

Empirical R_g and ν Predictions. To empirically predict R_g of an IDP sequence based on its average sequence properties, we conducted simulations on a large number of homopolymers using the HPS model, varying two important sequence descriptors, the chain length and the average hydropathy. We use 8 chain lengths from 25 to 450 residues and 16 average interaction strengths ($\langle\lambda\rangle$) from −3 to 3 and simulated each at 12 temperatures ranging from 150 to 600 K. For each of the 1536 systems, we calculated both R_g and the Flory scaling exponent (ν),[82] to approximate the dependence of chain dimensions on each of these three factors. Using this data set,

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

we are able to use a 3D linear grid interpolation approximate R_g and ν for any sequence of a given $\langle\lambda\rangle$ and chain length at any temperature within the range of the data set. The dimensions of the homopolymers are visualized as a function of T, λ, and chain length in Figures S14 and S15. To validate the accuracy of predictions from this method, we tested 2000 randomly generated sequences with a chain length of 80 and measured R_g and ν from the simulation to compare with estimates from the predictor (Figure S16). We find the largest source of error to be sequences with a high net charge, which is expected since our predictor only takes into account average hydropathy of the sequence. However, the predictor gives less than 10% error on R_g for most sequences with a low net charge.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00102.

Supporting text, figures, and tables including protein sequences, details on reference data set, details on additional models, details about empirical R_g predictor, and statistics from randomly generated sequences (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jeetain@lehigh.edu.

ORCID

Jeetain Mittal: 0000-0002-9725-6402

Author Contributions

‖G.L.D. and W.Z. contributed equally to this work

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research is supported by US Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0013979. This research used resources of the NERSC Center (a DOE Office of Science User Facility supported under Contract DE-AC02-05CH11231) and XSEDE (supported by the NSF Project TG-MCB120014). Y.C.K. was supported by the Office of Naval Research via the U.S. Naval Research Laboratory base program. W.Z. acknowledges the startup support from Arizona State University. J.M. and W.Z. both acknowledge the support of NVIDIA Corporation for the donation of a Titan Xp GPU.

■ REFERENCES

(1) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729−1732.

(2) Ditlev, J. A.; Case, L. B.; Rosen, M. K. Who's in and who's out—compositional control of biomolecular condensates. *J. Mol. Biol.* **2018**, *430*, 4666−4684.

(3) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **2016**, *165*, 1686−1697.

(4) Falahati, H.; Wieschaus, E. Independent active and thermodynamic processes govern the nucleolus assembly in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 1335−1340.

(5) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **2015**, *60*, 231−241.

(6) Ryan, V. H.; Dignon, G. L.; Zerze, G. H.; Chabata, C. V.; Silva, R.; Conicella, A. E.; Amaya, J.; Burke, K. A.; Mittal, J.; Fawzi, N. L. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* **2018**, *69*, 465−479.

(7) Riback, J. A.; Katanski, C. D.; Kear-Scott, J. L.; Pilipenko, E. V.; Rojek, A. E.; Sosnick, T. R.; Drummond, D. A. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **2017**, *168*, 1028−1040.

(8) Kroschwald, S.; Munder, M. C.; Maharana, S.; Franzmann, T. M.; Richter, D.; Ruer, M.; Hyman, A. A.; Alberti, S. Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery. *Cell Rep.* **2018**, *23*, 3327−3339.

(9) Marzahn, M. R.; et al. Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *EMBO J.* **2016**, *35*, 1254−1275.

(10) Sabari, B. R.; et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **2018**, *361*, eaar3958.

(11) Milovanovic, D.; Wu, Y.; Bian, X.; De Camilli, P. A liquid phase of synapsin and lipid vesicles. *Science* **2018**, *361*, 604−607.

(12) Kato, M.; et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **2012**, *149*, 753−767.

(13) Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K.; Zaslavsky, B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.* **2015**, *589*, 15−22.

(14) Pak, C. W.; Kosno, M.; Holehouse, A. S.; Padrick, S. B.; Mittal, A.; Ali, R.; Yunus, A. A.; Liu, D. R.; Pappu, R. V.; Rosen, M. K. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **2016**, *63*, 72−85.

(15) Rauscher, S.; Pomès, R. The liquid structure of elastin. *eLife* **2017**, *6*, e26526.

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

(17) Shakya, A.; King, J. T. DNA local-flexibility-dependent assembly of phase-separated liquid droplets. *Biophys. J.* **2018**, *115*, 1840−1847.

(18) Lu, H.; Yu, D.; Hansen, A. S.; Ganguly, S.; Liu, R.; Heckert, A.; Darzacq, X.; Zhou, Q. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **2018**, *558*, 318−323.

(19) Elbaum-Garfinkle, S.; Kim, Y.; Szczepaniak, K.; Chen, C. C.-H.; Eckmann, C. R.; Myong, S.; Brangwynne, C. P. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7189−7194.

(20) Conicella, A. E.; Zerze, G. H.; Mittal, J.; Fawzi, N. L. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **2016**, *24*, 1537−1549.

(21) Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A. ATP as a biological hydrotrope. *Science* **2017**, *356*, 753−756.

(22) Partridge, S.; Davis, H.; Adair, G. The chemistry of connective tissues. 2. Soluble proteins derived from partial hydrolysis of elastin. *Biochem. J.* **1955**, *61*, 11.

(23) Urry, D. W. Entropic elastic processes in protein mechanisms. I. Elastic structure due to an inverse temperature transition and

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

elasticity due to internal chain dynamics. *J. Protein Chem.* **1988**, *7*, 1−34.
(24) Meyer, D. E.; Chilkoti, A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. *Biomacromolecules* **2002**, *3*, 357−367.
(25) Bellingham, C. M.; Lillie, M. A.; Gosline, J. M.; Wright, G. M.; Starcher, B. C.; Bailey, A. J.; Woodhouse, K. A.; Keeley, F. W. Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. *Biopolymers* **2003**, *70*, 445−455.
(26) Lyons, R. E.; Nairn, K. M.; Huson, M. G.; Kim, M.; Dumsday, G.; Elvin, C. M. Comparisons of recombinant resilin-like proteins: repetitive domains are sufficient to confer resilin-like properties. *Biomacromolecules* **2009**, *10*, 3009−3014.
(27) Muiznieks, L. D.; Weiss, A. S.; Keeley, F. W. Structural disorder and dynamics of elastin. *Biochem. Cell Biol.* **2010**, *88*, 239−250.
(28) Garcia Garcia, C.; Kiick, K. L. Methods for producing microstructured hydrogels for targeted applications in biology. *Acta Biomater.* **2019**, *84*, 34−48.
(29) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; de la Cruz, M. O.; Xu, T. Random heteropolymers preserve protein function in foreign environments. *Science* **2018**, *359*, 1239−1243.
(30) Schuster, B. S.; Reed, E. H.; Parthasarathy, R.; Jahnke, C. N.; Caldwell, R. M.; Bermudez, J. G.; Ramage, H.; Good, M. C.; Hammer, D. A. Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **2018**, *9*, 2985.
(31) Shin, Y.; Berry, J.; Pannucci, N.; Haataja, M. P.; Toettcher, J. E.; Brangwynne, C. P. Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell* **2017**, *168*, 159−171.
(32) Zhang, H.; Aonbangkhen, C.; Tarasovetc, E. V.; Ballister, E. R.; Chenoweth, D. M.; Lampson, M. A. Optogenetic control of kinetochore function. *Nat. Chem. Biol.* **2017**, *13*, 1096.
(33) Lau, H. K.; Li, L.; Jurusik, A. K.; Sabanayagam, C. R.; Kiick, K. L. Aqueous liquid−liquid phase separation of resilin-like polypeptide/polyethylene glycol solutions for the formation of microstructured hydrogels. *ACS Biomater. Sci. Eng.* **2017**, *3*, 757−766.
(34) Roberts, S.; Harmon, T. S.; Schaal, J.; Miao, V.; Li, K. J.; Hunt, A.; Wen, Y.; Oas, T. G.; Collier, J. H.; Pappu, R. V.; Chilkoti, A. Injectable tissue integrating networks from recombinant polypeptides with tunable order. *Nat. Mater.* **2018**, *17*, 1154−1163.
(35) Li, L.; Luo, T.; Kiick, K. L. Temperature-triggered phase separation of a hydrophilic resilin-like polypeptide. *Macromol. Rapid Commun.* **2015**, *36*, 90−95.
(36) Ruff, K. M.; Roberts, S.; Chilkoti, A.; Pappu, R. V. Advances in understanding stimulus responsive phase behavior of intrinsically disordered protein polymers. *J. Mol. Biol.* **2018**, *430*, 4619−4635.
(37) Quiroz, F. G.; Chilkoti, A. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* **2015**, *14*, 1164.
(38) Dutta, N. K.; Truong, M. Y.; Mayavan, S.; Roy Choudhury, N.; Elvin, C. M.; Kim, M.; Knott, R.; Nairn, K. M.; Hill, A. J. A genetically engineered protein responsive to multiple stimuli. *Angew. Chem.* **2011**, *123*, 4520−4523.
(39) Balu, R.; Dutta, N. K.; Choudhury, N. R.; Elvin, C. M.; Lyons, R. E.; Knott, R.; Hill, A. J. An16-resilin: An advanced multi-stimuli-responsive resilin-mimetic protein polymer. *Acta Biomater.* **2014**, *10*, 4768−4777.
(40) Das, P.; Matysiak, S.; Mittal, J. Looking at the disordered proteins through the computational microscope. *ACS Cent. Sci.* **2018**, *4*, 534−542.
(41) Dignon, G. L.; Zheng, W.; Mittal, J. Simulation methods for liquid-liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.* **2019**, *23*, 92.
(42) van Dijk, E.; Varilly, P.; Knowles, T. P.; Frenkel, D.; Abeln, S. Consistent treatment of hydrophobicity in protein lattice models accounts for cold denaturation. *Phys. Rev. Lett.* **2016**, *116*, 078101.
(43) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5213−5218.
(44) Zhao, B.; Li, N. K.; Yingling, Y. G.; Hall, C. K. LCST behavior is manifested in a single molecule: elastin-like polypeptide (VPGVG)n. *Biomacromolecules* **2016**, *17*, 111−118.
(45) Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 9929−9934.
(46) Das, S.; Amin, A. N.; Lin, Y.-H.; Chan, H. S. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.* **2018**, *20*, 28558−28574.
(47) Kim, Y. C.; Hummer, G. Coarse-grained models for simulation of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* **2008**, *375*, 1416−1433.
(48) Nguyen, T. D.; Qiao, B.; de la Cruz, M. O. Efficient encapsulation of proteins with random copolymers. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 6578−6583.
(49) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **2018**, *14*, e1005941.
(50) Ashbaugh, H. S.; Hatch, H. W. Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. *J. Am. Chem. Soc.* **2008**, *130*, 9536−9542.
(51) Privalov, P. L.; Makhatadze, G. I. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **1993**, *232*, 660−679.
(52) Makhatadze, G. I.; Privalov, P. L. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **1993**, *232*, 639−659.
(53) Garde, S.; Hummer, G.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of entropy convergence in hydrophobic hydration and protein folding. *Phys. Rev. Lett.* **1996**, *77*, 4966.
(54) Cheung, J. K.; Shah, P.; Truskett, T. M. Heteropolymer collapse theory for protein folding in the pressure-temperature plane. *Biophys. J.* **2006**, *91*, 2427−2435.
(55) Patel, B. A.; Debenedetti, P. G.; Stillinger, F. H.; Rossky, P. J. A water-explicit lattice model of heat-, cold-, and pressure-induced protein unfolding. *Biophys. J.* **2007**, *93*, 4116−4127.
(56) Zerze, G. H.; Best, R. B.; Mittal, J. Sequence-and temperature-dependent properties of unfolded and disordered proteins from atomistic simulations. *J. Phys. Chem. B* **2015**, *119*, 14622−14630.
(57) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(58) van Dijk, E.; Hoogeveen, A.; Abeln, S. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. *PLoS Comput. Biol.* **2015**, *11*, e1004277.
(59) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640−647.
(60) Pucci, F.; Bourgeas, R.; Rooman, M. Predicting protein thermal stability changes upon point mutations using statistical potentials: introducing HoTMuSiC. *Sci. Rep.* **2016**, *6*, 23257.
(61) Huang, D. M.; Chandler, D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8324−8327.
(62) Wales, D. J.; Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(63) Dill, K. A.; Alonso, D. O.; Hutchinson, K. Thermal stabilities of globular proteins. *Biochemistry* **1989**, *28*, 5439−5449.

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830

(64) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Theories for sequence-dependent phase behaviors of biomolecular condensates. *Biochemistry* **2018**, *57*, 2499−2508.
(65) Dias, C. L.; Chan, H. S. Pressure-dependent properties of elementary hydrophobic interactions: ramifications for activation properties of protein folding. *J. Phys. Chem. B* **2014**, *118*, 7488−7509.
(66) Jung, J. G.; Bae, Y. C. Liquid-liquid equilibria of polymer solutions: Flory-huggins with specific interaction. *J. Polym. Sci., Part B: Polym. Phys.* **2010**, *48*, 162−167.
(67) Clark, E.; Lipson, J. LCST and UCST behavior in polymer solutions and blends. *Polymer* **2012**, *53*, 536−545.
(68) Singh, V. R.; Lapidus, L. J. The intrinsic stiffness of polyglutamine peptides. *J. Phys. Chem. B* **2008**, *112*, 13172−13176.
(69) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16764−16769.
(70) Panagiotopoulos, A. Z.; Wong, V.; Floriano, M. A. Phase equilibria of lattice polymers from histogram reweighting Monte Carlo simulations. *Macromolecules* **1998**, *31*, 912−918.
(71) Wang, R.; Wang, Z.-G. Theory of polymer chains in poor solvent: Single-chain structure, solution thermodynamics, and Θ point. *Macromolecules* **2014**, *47*, 4094−4102.
(72) Coeytaux, K.; Poupon, A. Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics* **2005**, *21*, 1891−1900.
(73) Debye, P.; Hückel, E. De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Physikalische Zeitschrift* **1923**, *24*, 185−206.
(74) Miller, C. M.; Kim, Y. C.; Mittal, J. Protein composition determines the effect of crowding on the properties of disordered proteins. *Biophys. J.* **2016**, *111*, 28−37.
(75) Kapcha, L. H.; Rossky, P. J. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* **2014**, *426*, 484−498.
(76) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(77) Torrie, G. M.; Valleau, J. P. Non-physical sampling distributions in monte-carlo free-energy estimation. *J. Comput. Phys.* **1977**, *23*, 187−199.
(78) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(79) Silmore, K. S.; Howard, M. P.; Panagiotopoulos, A. Z. Vapour-liquid phase equilibrium and surface tension of fully flexible Lennard-Jones chains. *Mol. Phys.* **2017**, *115*, 320−327.
(80) Jung, H.; Yethiraj, A. A simulation method for the phase diagram of complex fluid mixtures. *J. Chem. Phys.* **2018**, *148*, 244903.
(81) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **2008**, *227*, 5342−5359.
(82) Flory, P. J. The configuration of real polymer chains. *J. Chem. Phys.* **1949**, *17*, 303−310.

DOI: 10.1021/acscentsci.9b00102
ACS Cent. Sci. 2019, 5, 821−830