Hydrophobic residues advance the onset of simple coacervation in intrinsically disordered proteins at low densities: Insights from field theoretical simulations studies

Satwik Ramanjanappa¹, Sahithya S. Iyer², Anand Srivastava^{2,*}

¹Department of Physics, Indian Institute of Science-Bangalore, KA 560012

²Molecular Biophysics Unit, Indian Institute of Science-Bangalore, KA 560012

* anand@iisc.ac.in

Abstract

Intrinsically disordered proteins (IDPs) have engendered a definitive change in the way we think about the classical “sequence-structure-function” dogma. Their conformational pliability and rich molecular recognition features endow them with the ability to bind to diverse partners and predispose them to an elaborate functional armory. And of late, with studies on IDP-based liquid-liquid phase separation (LLPS) leading to formation of functional subcellular coacervates - best described as “membrane-less organelles (MLOs)”, IDPs are also bringing about paradigmatic changes in the way we think about biomolecular assemblies and subcellular organization. Though it is well recognized that the phase behavior of a given IDP is tightly coupled to its amino-acid sequences, there are only a few theories to model polyampholyte coacervation for IDPs. Recently, Joan-Emma Shea and co-workers used field theoretical simulations (FTS) to elucidate the complete phase diagram for LLPS of IDPs by considering different permutations of 50-residues chain representing 25 Lysine and 25 Glutamic acid [1]. Our work is an extension of that FTS framework where we develop and solve an augmented Hamiltonian that also accounts for hydrophobic interactions in the chain. We show that incorporation of hydrophobic interactions result in an advanced onset of coacervation at low densities. The patterning of hydrophobic, positive and negative residues plays important role in determining relative differences in the onset of phase separation. Though still very coarse-grained, once additional chemical specificities are incorporated, these high throughput analytical theory methods can be used as a starting point for designing sequences that drive LLPS.

1 Introduction

Intrinsically Disordered Proteins (IDPs) are functionally important despite lacking a native 3-dimensional fold [2–4]. Since they populate a range of conformations, they are best described by an ensemble rather than a single fold [5–8]. IDPs conformational landscape primarily deviates from a classical formula of sequence-based structure and function relationship generally applicable to folded proteins [2]. The

conformational flexibility and multivalency of IDPs allows them to bind with diverse partners [9, 10] and make them multifunctional in nature [11, 12]. In living cells, these IDPs/IDRs-containing proteins undergo spatio-temporal separation leading to the formation of membrane-less organelles (MLOs) through multivalent interactions with other IDPs/IDRs, folded proteins, and nucleic acids [13–16]. In general, formation of these functional and dynamic biomolecular condensates require multiple proteins and nucleic acids molecules. However, there are studies where one IDP species is enough to form liquid droplets.

One of the fundamental questions in the field is how the phase-separation behaviour is encoded in the IDP/IDR sequences [17–21]. The relationship between amino-acid sequence distribution and ability of IDPs to form LLPS droplets have been explored extensively in the recent past using experimental, computational and theoretical approaches [22–25]. While the underlying design principles are still a matter of intense exploration, the one that stands out is the abundant presence of charged amino-acids in the sequence. Strategically placed positive and negative residues in these high charge content sequences generally account for high conformational entropy of the IDPs. And the shorter range interactions from the hydrophobic and aromatic residues that are interspersed across the sequence play important role in the formation of large stable and dynamic complexes. The delicate interplay between the above two potentially competing contributions [19, 26] namely the chain conformational entropy and the multiple interchain enthalpic interactions give rise to multivalent interactions between the IDPs that sometimes lead to LLPS and formation of liquid droplets.

IDPs typically interact with binding partners through short sequence motifs; short linear motifs (SLiMs), molecular recognition features (MoRFs), and low-complexity domain (LCD) sequences. While the role of charge content and patterning is undisputed, hydrophathy patterning also plays an important role in dictating the multivalent nature and phase separation propensities of IDP complexes. In this context, LLPS in Tau protein is an interesting subject of study, where the LLPS is thought to be an intermediate state – capable of redissolving back into soluble state or transitioning into irreversible aggregates or amyloid fibrils [27–31]. A recent study shows that the fibril-prone and the reversible LLPS depends

on the nature of interactions driving Tau condensation [32]. The study shows that Tau amyloid aggregation is favoured if the LLPS is hydrophobically driven while the electrostatically driven LLPS is largely reversible and aggregation-independent. The sequence patterning, evolved over millions of years, needs to be understood in terms of continuum of intermolecular interactions that ranges from electrostatic effects to shorter-ranged polar and dipole interactions, PI-orbital based interactions (PI-PI and cation-PI) and the hydrophobic effects, which are largely overlooked so far. Hydrophobic effects are emergent in nature and they are particularly important while the LLPS is underway and their role is not just limited to short-ranged stabilization of the already formed condensates.

Though it is well recognized that the phase behaviour of a given IDP is tightly coupled to its amino acids sequences, there are only a few theories to model polyampholyte coacervation for IDPs. The available analytical theories are largely derived from the rich field of polymer physics and are predominantly based on homopolymers. Given the importance of chemical specificity of the sequence and also the importance of solvation, computer simulations with detailed atomic models with explicit solvent would be ideal to test and/or predict LLPS propensities of a given solution. However, the length and time scales are formidable and computationally intractable even for a simple model systems [33]. Coarse-grained (CG) models, with reduced representations of amino acids and solvents, provide a computationally tractable alternative. And some recent CG simulations studies have been providing deeper insights into the relationship between polypeptide sequence and its phase behaviour [34–37]. With rising computational powers, faster sampling algorithms and better AAMD and CGMD forcefield parameters [36, 38–42], molecular simulations have the promise to unravel the design principles of sequence patterning that leads to LLPS in biomolecular solutions. Complementary to molecular simulation methods, purely analytical theories such as random phase approximation (RPA) [19, 43] and field theoretical simulations (FTS) [1, 44–49] go beyond simple mean-field approaches such as Flory-Huggins and Overbeek-Voorn theories [50]. These theories can be used with as high throughput first-pass methods to make broad-level predictions of the phase behaviour of IDPs. This is because these models take into account the chain connectivity and local fluctuations in polymer density, which makes them “sequence sensitive” and are theoretically well-grounded and can be formally derived by considering Gaussian fluctuations in the

fields.

In general, classical self consistent field theory (SCFT) [44–49], a method widely used in polymer Physics field, becomes unreliable when there is significant density fluctuations and systems are close to phase transition. This often happens when one introduces electrostatic interactions and higher chemical specificity (detailed molecular interactions) on chains with limited degrees of polymerization. These inaccuracies can be avoided by using a fully fluctuating FTS models that promotes the Hamiltonian to complex values [44, 45, 51]. Though this leads to the famous "sign problem", which can be effectively treated by adopting the complex Langevin (CL) sampling method widely used in nuclear physics [52, 53]. In a wide-ranging and rich body of work, Glenn Fredrickson's group has used FTS with great effect to study a variety of problems in polymer physics [44–49]. Recently, together with Joan-Emma Shea's group, they have applied the method to elucidate the complete phase diagram for LLPS of IDPs where the authors have considered different permutations of 50-residues chain (25 Lysine and 25 Glutamic acid) and their study is focused on the role of charge-charge interaction in driving self-coacervation. Our work is an extension of the framework laid out in this recent publication where we develop an augmented Hamiltonian to account for apolar residues in the chain. Borrowing from the seminal work in the field of SASA and hydrophobic solvation calculation [54–58], we include an additional term in the Hamiltonian that is a function of solvent accessible surface area (SASA) and surface tension. These inclusion invokes approximations for making the functional integrals and partition function tractable for FTS. The new Hamiltonian and extended partition function makes the systems under consideration more realistic in their sequence decoration. We consider a variety of sequence with different patterns of positively charged, negatively charged and neutral apolar residues. To uniquely annotate the sequence decorations, we use a modification of scoring prescription proposed by Das and Pappu [59] to account for neutral residues in the score. Our work clearly shows how consideration of apolar residues (and hydrophobic solvation energy) affects the phase diagram and in particular the onset of the LLPS.

This article is organised as follows. In section 2 we describe the methods of formulation of hydrophobic interactions with which we score patterns on chains that have both charged and apolar sites along with

the field theoretic Hamiltonian and associated equations. After that we present and describe the key insights that we arrive at using our methods and how this can be used and refined to better models in other fields of biophysics and molecular organisations. In the appendix section more details of solving methods and how Gibbs ensemble was performed are described. All codes related to the methods development and example files are made available on Github (<https://github.com/codesrivastavalab/fts-idplls>).

2 Methods

2.1 Sequence patterning with neutral beads and systems under consideration

We assign a unique score (m) to each of the sequence. The value of m helps in distinguishing the different patterns of positive, negative and neutral residues on the sequences. The scoring method used is an extension to "patterning parameter" κ introduced by Das and Pappu [59]. We modify the charge asymmetry parameter (σ) to include neutral residues. σ is now defined as $\sigma = \frac{(f_+ - f_-)^2}{(f_+ + f_- + f_0)}$, where f_+ , f_- and f_0 are the fraction of positive, negative and neutral residues within the blob respectively. We find that a blob size of three instead of the commonly used size of five ensures that the uniqueness of the score in the sequences. The total number of blobs (N_{blob}) is calculated by the number of sets of residues that is obtained by moving the window of size three with overlapping blobs. Asymmetry parameter is calculated for each sets of three residues. δ is the standard deviation of σ given as $\delta = \frac{\sum_{i=1}^{N_{blob}} [(\sigma_i - \sigma)^2]}{N_{blob}}$. The sequence score κ is calculated by normalising δ , $\kappa = \frac{\delta}{\delta_{max}}$. m ranges from 0 to 1 with $\kappa = 1$ for a system with a block polymer kind of arrangement of positive, neutral and negative residues.

For this study, we mainly consider two types of chains. In the first set of chains, we use only charge residues and the systems are same as the ones used by Joan-Emma Shea and co-workers [1]. We chose these systems both to validate our work as well as to show how the inclusion of explicit hydrophobic interactions in the FTS Hamiltonian affects the phase diagram even in case of pure polyampholytes. In the second set of chain patterning, we also consider neutral beads in the sequence. Here again, we have two broad class of sequence decoration. In the first set, we keep the neutral beads in the middle of the chain with flanking positive and negative beads and change the size of the central neutral band. In the

Figure 1: (Top panel) Value of m computed for various types of chains with no neutral sites as done in previous studies. (Bottom panel) Chain number for the chains with neutral sites as well, considered in following sections.

second set, we have an interspersed neutral beads across the chain. These systems are designed to shed more light on the nature of phase diagram as a function of more realistic sequences as found in IDPs. Please see Fig. 1 for different patterning that we have used in this work.

2.2 Theory and numerical implementation

2.2.1 Hamiltonian for the fluctuating field theoretical model

For sake of completeness and to highlight the augmentation made to the existing model, we re-present an overview of polymer model with notations that are consistent with the original paper [1]. The system consists of n continuous Gaussian chains of volume V with N monomers per chain. This results in a monomer number density ρ of nN/V . Each monomer carries a monomer specific charge in terms of electronic charge e such that monomer j of chain i has a charge pattern $\sigma_{i,j}$. The polymer chains are held together by a harmonic interaction potential given as:

$$\beta U_{\text{chain}} = \frac{3}{2b^2} \sum_{i=1}^g \sum_{j=1}^N (|\mathbf{r}_{i,j} - \mathbf{r}_{i,j-1}|)^2 \quad (1)$$

where $r_{i,j}$ refers to the position of monomer j on chain i , b is the statistical segment length, and $\beta = 1/k_B T$. The monomers interact with a purely repulsive excluded volume interaction given as:

$$\beta U_{vol} = v \delta(\mathbf{r}) \quad (2)$$

where v is the excluded volume parameter. Electrostatic interactions between charged monomers in a polymer chain as given as:

$$\beta U_{el} = \frac{l_B \sigma_i \sigma_j}{r} \quad (3)$$

where $l_B = e^2 / (4\pi\epsilon_0\epsilon_r k_B T)$ is the Bjerrum length, ϵ_0 is the vacuum permittivity, ϵ_r is the relative dielectric strength of the solvent, and e is the fundamental charge of electron.

The total microscopic monomer number density and charge density are given as:

$$\hat{\rho}(\mathbf{r}) = \sum_i^g \sum_j^N \delta(\mathbf{r} - \mathbf{r}_{i,j}) \quad (4)$$

and

$$\hat{\rho}_e(\mathbf{r}) = \sum_i^g \sum_j^N \sigma_{ij} \delta(\mathbf{r} - \mathbf{r}_{i,j}) \quad (5)$$

respectively. In order to maintain continuity, the number and charge density are smeared by performing a Gaussian convolution on Eqn 4 and 5. The smeared densities $\bar{\rho}(\mathbf{r})$ and $\bar{\rho}_e(\mathbf{r})$ are given as:

$$\bar{\rho}(\mathbf{r}) = \int d\mathbf{r}' \Gamma(|\mathbf{r} - \mathbf{r}'|) \hat{\rho}(\mathbf{r}') \quad (6)$$

and

$$\bar{\rho}_e(\mathbf{r}) = \int d\mathbf{r}' \Gamma(|\mathbf{r} - \mathbf{r}'|) \hat{\rho}_e(\mathbf{r}') \quad (7)$$

with

$$\Gamma(r) = \left(\frac{1}{2\pi a^2} \right)^{3/2} \exp(-r^2/2a^2) \quad (8)$$

where "a" is normalised Gaussian smear width.

In this work, we add to the Hamiltonian an additional term to account for hydrophobic interaction. The formulation of the hydrophobic interaction terms come from approximations that make the formulations pliable for the FTS and we explain the same in greater detail in next section. The final form of the hydrophobic interaction potential that we arrive at is given as:

$$\beta U_s = \sigma_H A \bar{\rho}(r)(1 - \bar{\rho}_0(r)) \quad (9)$$

where σ_H , scaled by b^2 , is the solvation surface tension term and A is the surface area of the monomer sphere. The value of σ_H determines the energy per unit surface area of the solute molecule exposed to a particular solvent. It is taken to be $0.6 \text{ J } \text{\AA}^{-2}$ (unscaled) for this work. In this work, we ignore the effect of hydrophobic interactions at very small distances. The Hamiltonian function has no particular form and the noise term also competes with this term leading to randomly staggered values.

Taken together, the total interaction energy Hamiltonian of the system written as the sum of bonded, excluded volume, Coulombic and hydrophobic interactions is given as:

$$\begin{aligned} \beta \bar{U}[\mathbf{r}] = & \frac{3}{2b^2} \sum_{i=1}^g \sum_{j=1}^N (|\mathbf{r}_{i,j} - \mathbf{r}_{i,j-1}|)^2 + \frac{v}{2} \int d\mathbf{r} \bar{\rho}^2(\mathbf{r}) \\ & + \frac{l_B}{2} \iint d\mathbf{r} d\mathbf{r}' \frac{\bar{\rho}_e(\mathbf{r}) \bar{\rho}_e(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ & + \sigma_H A \int d\mathbf{r} \bar{\rho}(\mathbf{r})(1 - \bar{\rho}_0(r)) \end{aligned} \quad (10)$$

2.2.2 Formulation details of Hydrophobic interactions

We use the the classical solvent accessible surface area (SASA) approach to account for hydrophobic interactions into the field theoretic hamiltonian framework. The choice of discrete Gaussian density makes it convenient to work with the integral form of the field theoretical approach. This formulation is mainly build using solid angle calculation and trigonometric manipulations and we demonstrate that in following text. For this work, we assume that all the molecules are spherical in shape. To illustrate the formulations, we will consider two different cases of solute-solute interactions. In the first case, two

solute/monomer molecules are in close proximity. This is possible for two adjacent sites on a chain, two well-separated sites on the same chain making contact or intermolecular contacts. The maximum area of the solute molecule is blocked (due to presence of other molecule) from solvent when the molecule is in contact with the other spherical solute molecule. The magnitude of blocked area is computed by calculating solid angle area of angular separation the molecule creates on the considered solute molecule plus the angular separation the solvent molecule creates as shown in Fig. 2a. This would simply be given as $\bar{A}_0(r = 2R_s) = 4\pi(\sin^2 \frac{\theta}{2})R_s^2$ where R_s is radius of solute. Here, the area is expressed in terms of centre to centre distance between the solute molecules. For cases where there is a gap between proximal solutes that could allow a solvent space, we also require the information of blocked area for distances ranging from $2R_s$ to $2(R_s + r_s)$, with r_s being the radius of solvent molecule as shown in Fig. 2b. The blocked area calculated as a function of distance between solute molecules is in such a situation is given as $\bar{A}_0(\mathbf{r}) = A \left(\frac{K}{4} \left(\frac{4R_s|\mathbf{r}| - 3R_s^2}{|\mathbf{r}|^2 + R_s^2} \right) \right)$, where K is sine squared of the angle subtended by solute and solvent molecule on the surface of considered solute as shown in Fig. 2b. This is expressed in terms of R_s and r_s where A is the surface area of a solute molecule equal to $4\pi R_s^2$ and $\bar{A}_0(r)$ is the blocked surface area due to single molecule of solute as a function of centre to centre distance between the molecules as shown in the Fig. 2. More details, in terms of the underlying algebra associated with the formulation, is provided in the following paragraph.

Starting from the simplest case for a solute placed over the surface of other as depicted in main text, the area blocked in this scenario is given by $\bar{A}_0(r = 2R_s) = 4\pi(\sin^2 \frac{\theta}{2})R_s^2$. On calculating this sine of the angle in terms of R_s we get $\sin^2 \frac{\theta}{2} = \frac{R_s^2}{(2R_s)^2 + R_s^2}$. When we consider the second situation, where the proximal solutes can accommodate solvent molecules, we can take the difference in the blocked area from the first case to get the new blocked area as depicted in Fig. 2 and this is given as $\bar{A}_0(\theta) = A((\sin^2 \frac{\theta}{2}) - [\sin^2 \frac{(\theta - \theta')}{2}])$. Here, we approximate $\sin \frac{\theta}{2}$ as $\frac{\sin \theta}{2}$ and also assume the limit of 90 degrees as the angle that is subtended by both the points at centre of the mobile sphere. On expanding the $\sin^2 \frac{(\theta - \theta')}{2}$ and putting back everything in terms of r and R_s , for the sake of field theoretical formulation, we arrive at $\bar{A}_0(r) = A[\frac{(K=0.2)}{4} - \frac{(K)}{4} \frac{((r - R_s)^2)}{r^2 + R_s^2}]$, which can be rewritten as $\bar{A}_0(r) = A \frac{(K)}{4} [\frac{(r^2 + R_s^2 - (r - R_s)^2)}{r^2 + R_s^2}]$. With this, we end up as the final form of $\bar{A}_0(r) = A \frac{K}{4} [\frac{4rR_s - 3R_s^2}{r^2 + R_s^2}]$.

Figure 2: (Left panel) Case where maximum area is blocked by the solute molecule to interact with solvent. (Middle panel) Situation where magnitude of separation between the molecules is r . (Right panel) Function in Eqn 9 vs r by keeping $\bar{\rho} = 1$ for simplicity of demonstration for $R_s = 3.5\text{\AA}$ and $r_s = 2.5\text{\AA}$.

The simple system of two solute molecules can then be used and generalised to normal system with inclusion of density of solute at that point to find the total contribution to blocked area due to those molecules and then add them over complete region. We find that the local fraction of blocked area can be computed using the density of solutes at that point as:

$$\bar{\rho}_0(\mathbf{r}) = \int_{\mathbf{r}+2R_s}^{\mathbf{r}+2(R_s+r_s)} d\mathbf{r}' \bar{\rho}(\mathbf{r}') \left(\frac{K}{4} \left(\frac{4R_s|\mathbf{r}' - \mathbf{r}| - 3R_s^2}{|\mathbf{r}' - \mathbf{r}|^2 + R_s^2} \right) \right) \quad (11)$$

where $\bar{\rho}$ is the Gaussian smeared number density (Eqn 8). This form can be integrated over the whole region to get the total fraction of the blocked area blocked expressed as $\rho_0(\mathbf{r}) = \int d\mathbf{r} \bar{\rho}(\mathbf{r})(1 - \bar{\rho}_0(\mathbf{r}))$ and the total surface exposed area for the whole system can be expressed as $A_0(\mathbf{r}) = A \int d\mathbf{r} \bar{\rho}(\mathbf{r})(1 - \bar{\rho}_0(\mathbf{r}))$ and incorporated in the full augmented Hamiltonian as in Eq. 10.

We subject the dependence of the blocked area to a particular range as the expression diverges for distance lower than the contact distance and becomes intractable in the integrable form. This approximation is acceptable since this distance has no influence on calculations of the blocked area. As shown

above in equations above, formulation for computing SASA suggests complicated dependence on distance between molecules. This complication can be avoided for easier and convenient computational analysis by just considering the functional part only concerned with present work (i.e. the variation of this function for a range from $2R_s$ to $2(R_s + r_s)$). The trend suggests that Fig. 2(c) (for $R_s = 3.5\text{\AA}$) can be approximated to a linear function with definite slope and intercept. Values of slope and the intercept change depending on the value of (R_s) and (r_s) , expression for $\bar{A}_0$ would become a simple dependence on r as $\bar{A}_0(\mathbf{r}) = A \left[\left(\frac{K}{4} \right) + (|\mathbf{r}| - 2R_s)m \right]$. In the above demonstration value of R_s was taken to be $3.5\text{\AA}$ for which m was equal to 0.00168. For this work R_s was taken equal to $2.5\text{\AA}$ for which m was equal to 0.00123. The values of slope and intercepts to linearise can be found out by taking sufficiently good number of data points from original function and fitting it with a linear function. This kind of linear trend has reduced the computational and theoretical expense to minimal. The simplified linearized expression for $\bar{A}_0(\mathbf{r})$ is used in the entire work as an account to blocked surface area.

2.2.3 The partition function and modified diffusion equation (MDE)

For our augmented Hamiltonian (Eq. 10), we write the partition function by keeping the ideal gas term as a constant and we introduce auxiliary field variables using functional integrals for excluded volume and electrostatic interactions terms as done by Shea and co-workers [1]. The auxiliary field variables for chemical potential field (ω) and electrostatic potential field (φ) are incorporated through Hubbard-Stratonovich transformation [60, 61]. We incorporate the hydrophobic interactions terms ($\bar{\rho}_0$) in the single chain formulation [62]. The partition function is expressed as:

$$\mathcal{Z} = \mathcal{Z}_0 \int D\omega \int D\varphi e^{-H[\omega, \varphi, \bar{\rho}_0]} \quad (12)$$

where $\mathcal{Z}$ is the partition function for an ideal gas with discrete Gaussian chains and other dependent parameters. The Hamiltonian, in terms of auxiliary field variables, is now rewritten as:

$$H[\omega, \varphi, \bar{\rho}_0] = \frac{1}{2} \int d\mathbf{r} \left(\frac{[\omega(\mathbf{r})]^2}{v} + \frac{|\nabla \varphi(\mathbf{r})|^2}{4\pi l_B} \right) - n \ln Q[\omega, \varphi, \bar{\rho}_0] \quad (13)$$

where $Q[\omega, \varphi, \bar{\rho}_0] = \frac{1}{V} \int d\mathbf{r} q_N(\mathbf{r}; [\omega, \varphi, \bar{\rho}_0])$ is a single chain partition function, which can be computed by the Gaussian chain propagator q as a solution to modified differential equation (MDE):

$$\frac{\partial}{\partial s} q_l(\mathbf{r}, s) = \left[\frac{b^2}{6} \nabla^2 - (\sigma_H A \Gamma \star (1 - \bar{\rho}_0) + \psi_l)(\mathbf{r}, s) \right] q_l(\mathbf{r}, s) \quad (14)$$

with initial condition $q_l(\mathbf{r}, s=0) = 1$ and $\psi_l(\mathbf{r}, s)$ entering Eqn 14 is given by $\psi_l(\mathbf{r}, s) = i\Gamma \star (\omega + \sigma_l \varphi)$, where $i = \sqrt{-1}$, $\star$ denotes a spatial convolution and σ_l is the charge of l^{th} monomer. $q_N(\mathbf{r}; [\omega, \varphi, \bar{\rho}_0])$ can also be constructed from Chapman-Kolmogorov equation [63]. The above differential equation was solved using operator splitting approach through discrete Fourier transform to get the initial configuration dependent field values. We also used the finite-difference approach and arrive at almost similar results for all cases. Field theoretical simulations are implemented by stochastically sampling the fields in complex Langevin dynamics where the variables ω and φ are propagated to complex space and given as:

$$\begin{aligned} \frac{\partial \omega}{\partial t} &= -\lambda_1 \frac{\delta H[\omega, \varphi, \bar{\rho}_0]}{\delta \omega(\mathbf{r}, t)} + N_1(\mathbf{r}, t) \\ \frac{\partial \varphi(\mathbf{r}, t)}{\partial t} &= -\lambda_2 \frac{\delta H[\omega, \varphi, \bar{\rho}_0]}{\delta \varphi(\mathbf{r}, t)} + N_2(\mathbf{r}, t) \end{aligned} \quad (15)$$

with

$$\begin{aligned} \frac{\delta H[\omega, \varphi, \bar{\rho}_0]}{\delta \omega(\mathbf{r}, t)} &= \frac{\omega(\mathbf{r})}{v} - \frac{g}{Q} \frac{\delta Q[\omega, \varphi, \bar{\rho}_0]}{\delta \omega(\mathbf{r}, t)} \\ \frac{\delta H[\omega, \varphi, \bar{\rho}_0]}{\delta \varphi(\mathbf{r}, t)} &= \frac{\nabla^2 \varphi(\mathbf{r})}{4\pi l_B} - \frac{g}{Q} \frac{\delta Q[\omega, \varphi, \bar{\rho}_0]}{\delta \varphi(\mathbf{r}, t)} \end{aligned} \quad (16)$$

$N(\mathbf{r}, t)$ terms are real valued Gaussian White noise with zero mean and variance depending on dimensions of the simulation box, time stepping and volume segment. We update the field for every $t = 0.02$ time stepping in CL simulations. The modified diffusion equation (MDE) was solved with fixed contour stepping of $\Delta s = 0.01$. We provide more details on numerical implementation for MDE and CL formulations in the next section. With partition function firmly in place, we proceed to calculate the various necessary quantities. Here chemical potential was computed with $\tilde{\mu} = \ln \frac{\rho}{N} - \ln Q[\varphi, \omega, \bar{\rho}_0]$ being chemical potential operator and ρ was the monomer number density. Numerical details of MDE and CL dynamics are discussed in Appendix. Also, following the CG dynamics calculations, we use the regular Gibb's ensemble method to arrive at the phase coexistence curve. For sake of completeness, Gibb's ensemble workflow is also presented in the Appendix.

Figure 3: Chemical potential plots for various different charge decorated chains.

3 Results and Discussion

In both molecular simulations and analytical theory work, role of charge patterning in IDP chains has been extensively studied to elucidate the conformational preferences and phase separation behaviour of the IDP. There has been relatively little progress in accounting for the role of other types of interactions such as hydrogen bonding interactions, weak non-covalent interactions ($\pi - \pi$, Cation- π) and hydrophobic interactions. In this work, we explore the effect of hydrophobic interactions and as well as hydrophobic residues patterning on the phase co-existence behavior.

3.1 Effect of hydrophobic interactions on phase coexistence

The effective charge distribution in sequences on the binodal curve was studied extensively by Shea and co-workers [1]. Sequences with different degree of “patchy” distribution of charges, quantified by the κ values showed an increase in critical Bjerrum length with decrease in κ of the sequences. A direct comparison to the experimental phase diagrams can be made by considering the inverse relation of Bjer-

rum length and temperature. The sequences show increasing upper critical solution temperatures with increase in the κ values. This behaviour is driven by the predominance of intra-sequence electrostatic interactions in sequences with larger “patches” of contiguously charged residues. In Figure 4, we show the effect on the binodal when hydrophobic interactions is considered for a polyelectrolyte where we show the phase diagram for two sequences at extremes of the charge decoration “patchiness”. The points on the coexistence curves were determined using Gibb’s ensemble [46] method as described in the methods section. While the behaviour of increasing upper critical solution temperature of sequences with higher κ score is preserved with the inclusion of hydrophobic interactions, the onset of phase separation occurs at lower densities. This can be attributed to higher contribution of the hydrophobic interactions than the electrostatic interactions to the Hamiltonian at lower densities. Since field theoretical simulations do not provide information about geometries of the sequences, it is difficult to comment on how addition of hydrophobic interactions effect single chain collapse at lower densities.

Figure 4: Phase diagram for sequence with lower and higher κ values. Lower κ have higher critical Bjerrum length. Sequences corresponding to different κ values is shown in Figure 1.

3.2 Effect of inclusion of hydrophobic interactions in binodal of sequences with apolar residues

The role of hydrophobic interactions in self- coacervation leading to LLPS has been studied extensively in elastin-like peptides and tau protein [64–66]. The effect of inclusion of hydrophobic interactions is more evident in sequences that have apolar sites. These sites do not have an electrostatic interaction term, hence making the hydrophobic interactions the only other non-bonded interaction term in the Hamiltonian governing sequence organization. In the following section, we explore the effect of different hydrophobic patterning as well as different hydrophobic stretch-size on the phase co-existence behaviour.

3.2.1 Different decorations

Similar to the effect of “patches” of distribution of charge beads on the critical Bjerrum length, the decorations of apolar residues also determine the critical Bjerrum length. Figure 6 shows the phase diagram of sequences with different patterns of charged and neutral beads arrangement. The arrangement of sequences with κ value of 0.001, 0.1 and 1 are shown in Figure 1. All these sequences have conserved number of positive, neutral and negative beads. The critical Bjerrum length is higher for sequences with lower κ values. Bjerrum length can be an indicator of temperature of the system to some extent though we note that a direct connection between Bjerrum length and temperature cannot be made in these systems due to the inclusion of hydrophobic interaction term in the Hamiltonian.

Apart from the effect on critical Bjerrum length, inclusion of hydrophobic residues also result the onset of LLPS at lower monomer densities. This is similar to the effect of inclusion of hydrophobic interactions in sequences with no neutral residues. In Figure 6, the phase diagram obtained for sequences without hydrophobic interaction term is shown in open circles. The onset of LLPS at lower monomer densities for sequences after accounting for hydrophobic interactions is evident from the figure. Furthermore, the effect of hydrophobic interactions on the onset of LLPS at lower densities has more effect on polymer chains with less charge decorations i.e. block type polymers than in distributed systems.

Figure 5: Phase diagram for sequence with lower and higher κ values. Lower κ have higher critical Bjerrum length. Effect of HI is seen more evident in systems enriched with contiguous neutral patches.

3.2.2 Different stretches of hydrophobic centers

We study another aspect of inclusion of neutral residues with hydrophobic interactions by varying the length of apolar neutral block in the triblock system while keeping the total chain length same. As expected, with increase in length of the neutral hydrophobic stretch, the onset of LLPS occurs at lower densities. The possible entropic origin of these self-coacervation at lower densities is better examined using particle based simulation studies. Though an extensive study of low-complexity regions like the RGG/RG motifs in driving LLPS has been documented [67,68], to the best of our knowledge, the nature of length of hydrophobic amino acid motifs that drive LLPS is yet to be thoroughly studied.

Generally, a diblock polymer of opposite charges always tends to be in a hair pin like structure. On the contrary, polymers with distributed charges tend to be in a coiled state and are more collapsed. One would expect similar behavior in IDPs. Chains with distributed charges are in considerably collapsed or coiled state than the ones with stretches of like-residues in blocks. For a noticeable hydrophobic stretch,

SASA contribution becomes more effective for phase separation to happen. This can also be seen from the changes in chemical potential with concentration as seen in Figure 6, which clearly portrays unique behavior of triblock chain, (black dots) with that trend fading with more distributed systems. As a result the changes observed in later system is less.

Figure 6: Phase diagram for triblock polymers with different lengths of stretches of neutral residues for sequences with same chain length

3.2.3 Different values of solvation coefficient

Apart from Bjerrum length, solute-solvent interactions governed by exposed surface area, also determine the coexistence points on the binodal. As per the definition, with larger value of σ_H , the chains would avoid exposure to the solvent resulting in faster phase separation. This was verified by increasing the value of σ_H in above calculations. Figure 7 shows that an increase in σ_H results in the shift of phase boundaries to lower densities.

Figure 7: Phase diagram for triblock polymers with different σ_H .

4 Conclusion and Outlook

In this work, we have attempted to describe phase coexistence with incorporating hydrophobic interactions through SASA approach, which gives us better understanding of LLPS for studies in biology. This entropically driven effect with more sophisticated studies can give biologists a better understanding of formation of cell organelles and biological globules formed by IDP's. We would need better formulations for hydrophobic interactions for more intrinsic details, as seen earlier we have ignored the contributions from lower distances which FTS for this formulation fail to give accurate values, better models need to address this issue.

5 Appendix

Numerical details of MDE and CL dynamics:

We first calculate the initial values for the auxiliary fields through Fourier transform by moving into the reciprocal space. We solve MDE using operator splitting method (as shown below) as well as well known

finite difference algorithm which happen to agree within appreciable error. In case of operator splitting algorithm, the formal solution to the diffusion equation is expanded to second order in monomer step Δs such that:

$$\begin{aligned} q(\mathbf{r}, s + \Delta s) &= \\ &\exp \left[\Delta s \left(\nabla^2 - (\sigma_H A \Gamma * (1 - \bar{\rho}_0) + \psi(\mathbf{r}, s)) \right) \right] q(\mathbf{r}, s) \\ &= \exp \left[-\frac{\Delta s}{2} (\sigma_H A \Gamma * (1 - \bar{\rho}_0) + \psi(\mathbf{r}, s)) \right] * \exp [\Delta s \nabla^2] * \\ &\exp \left[-\frac{\Delta s}{2} (\sigma_H A \Gamma * (1 - \bar{\rho}_0) + \psi(\mathbf{r}, s)) \right] q(\mathbf{r}, s) + O(\Delta s^3) \end{aligned} \quad (17)$$

The solution of the diffusion equations indeed represents the most computationally demanding component of FTS. Runge-Kutta algorithm was used to propagate CL equations with time step of $\Delta t = 0.1$. The chemical potential operator as described in the main text was used to compute chemical potential and pressure operator as by Shea and co-workers [1] with small modifications was

$$\begin{aligned} \tilde{\Pi}[\omega, \varphi] &= \frac{C}{N} + \frac{C}{NV} \int d\mathbf{r} \left(\frac{2}{3} \Phi_1(\mathbf{r}) + \sum_m \Phi_m(\mathbf{r}) \sigma_H A \right. \\ &= \left(\Gamma_2 - \frac{\Gamma}{2} \right) * (1 - \bar{\rho}_0) + i \left(\Gamma_2 - \frac{\Gamma}{2} \right) * \omega + i \sigma_m \left(\Gamma_2 - \frac{\Gamma}{6} \right) * \varphi \end{aligned} \quad (18)$$

where C is the scaled monomer density (ρb^3) and $*$ denotes a spatial convolution. $\Gamma_2(\mathbf{r}) = (1 - r^2/3a^2) \Gamma(\mathbf{r})$ with $\Gamma(\mathbf{r})$ a normalized Gaussian. In the above equation $\Phi_1(\mathbf{r})$ is given as $\Phi_1(\mathbf{r}) = Q^{-1} \sum_{j=1}^N q^\dagger(\mathbf{r}, j) \nabla^2 q(\mathbf{r}, j)$ where $q(\mathbf{r}, j)$ is the forward Gaussian chain propagator, and $q^\dagger(\mathbf{r}, j)$ is a similar reverse chain propagator, and $\Phi_m(\mathbf{r})$ represents the concentration of all the non interacting fields $\Phi_m(\mathbf{r}) = Q^{-1} \sum_{j \in m} q^\dagger(\mathbf{r}, j) q(\mathbf{r}, j)$, which is a monomer-type m density field computed by counting monomer segments of monomer-type m . Here in this work, we used only three types of monomers $\sigma = -1, 0, 1$.

Gibbs ensemble and phase separation equilibrium:

After the CL runs, we constructed the binodal curves by performing FTS in Gibbs ensemble. This is conventionally done by partitioning the polymer in solution of total volume V into two phases of V_I and V_{II} with corresponding number of chains g_I and g_{II} , respectively such that $V_I + V_{II} = V$ and

$g_I + g_{II} = g$ [1]. We arrive at the equilibrium condition when the chemical potential and the osmotic pressure of both phases are equal. In Gibbs ensemble, the total simulation box of volume V is partitioned into two regions corresponding to a dense and dilute phase. The concentration and phase volumes change continuously till we reach equilibrium between the two phases. We scaled all the lengths by $R_0 = \frac{b}{\sqrt{6}}$ with b as segment length and reduced chain density $C = \frac{gNR_0^3}{V} = \rho R_0^3$, which we used to plot phase diagrams. From the equation of balances, we get $C_T = v_I C_I + v_{II} C_{II}$ with C_T, C_I and C_{II} as total chain density, chain densities of region I and II, respectively. v_I and v_{II} are the volume fractions defined as $v_I = V_I/V$ and $v_{II} = 1 - v_I$. Differentiating with time and with concentration rate proportional to difference in chemical potential of either medium and similarly correlating between the rate of change of volume and pressure change, we get:

$$\begin{aligned}\frac{dC_I}{dt} &= -(\mu_I - \mu_{II}) \\ \frac{dv_I}{dt} &= -(\Pi_{II} - \Pi_I)\end{aligned}\tag{19}$$

which can be solved numerically by simple Euler stepping algorithms. We used Runge-Kutta methods to solve and update the scheme as:

$$\begin{aligned}C_I^{i+1} &= C_I^i - \Delta t_c (\mu_I - \mu_{II}) \\ C_{II}^{i+1} &= \frac{C_T - v_I C_I^{i+1}}{v_{II}} \\ v_I^{i+1} &= v_I^i - \Delta t_v (\Pi_{II} - \Pi_I) \\ v_{II}^{i+1} &= 1 - v_I^{i+1}\end{aligned}\tag{20}$$

In our work, we maintain charge neutrality and we run our simulations for 10^6 CL steps, updating concentrations and volume according to above equations for every 1000 steps with $\Delta t_c = 0.05$ and $\Delta t_v = 2$. We recorded the concentrations C_I and C_{II} at the completion of Gibbs ensemble runs as this dataset form the points of boundary for the binodal curve for particular temperature in the phase diagram.

Acknowledgments

A.S. thanks the Indian Institute of Science and the Ministry of Human Resource Development of India for the start up grant and the Department of Science and Technology of India for the early career grant. This research was also supported by the Department of Biotechnology, Government of India in the form of IISc-DBT partnership programme. Support from FIST program sponsored by the Department of Science and Technology and UGC, India – Centre for Advanced Studies and Ministry of Human Resource Development, India is gratefully acknowledged by the authors.

References

- [1] J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, and J.-E. Shea, “Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins,” *The Journal of Physical Chemistry Letters*, vol. 10, no. 8, pp. 1644–1652, 2019.
- [2] P. E. Wright and H. Dyson, “Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm,” *Journal of Molecular Biology*, vol. 293, no. 2, pp. 321–331, 1999.
- [3] V. N. Uversky, J. R. Gillespie, and A. L. Fink, “Why are “natively unfolded” proteins unstructured under physiologic conditions?,” *Proteins: Structure, Function, and Bioinformatics*, vol. 41, no. 3, pp. 415–427, 2000.
- [4] A. Dunker, J. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic, “Intrinsically disordered protein,” *Journal of Molecular Graphics and Modelling*, vol. 19, no. 1, pp. 26 – 59, 2001.
- [5] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samorì, “Conformational equilibria in monomeric alpha-synuclein at the single-molecule level,” *PLOS Biology*, vol. 6, pp. 1–10, 01 2008.

- [6] M. Schwalbe, V. Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M. Jensen, J. Biernat, S. Becker, E. Mandelkow, M. Zweckstetter, and M. Blackledge, “Predictive atomic resolution descriptions of intrinsically disordered htau40 and alpha-synuclein in solution from nmr and small angle scattering,” *Structure*, vol. 22, no. 2, pp. 238 – 249, 2014.
- [7] J. Henriques, L. Arleth, K. Lindorff-Larsen, and M. Skepö, “On the calculation of saxs profiles of folded and intrinsically disordered proteins from computer simulations,” *Journal of Molecular Biology*, vol. 430, no. 16, pp. 2521 – 2539, 2018.
- [8] G.-N. W. Gomes, M. Krzeminski, A. Namini, E. W. Martin, T. Mittag, T. Head-Gordon, J. D. Forman-Kay, and C. C. Gradinaru, “Conformational ensembles of an intrinsically disordered protein consistent with nmr, saxs, and single-molecule fret,” *Journal of the American Chemical Society*, vol. 142, no. 37, pp. 15697–15710, 2020.
- [9] W.-L. Hsu, C. Oldfield, B. Xue, J. Meng, F. Huang, P. Romero, V. Uversky, and A. Dunker, “Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding,” *Protein science : a publication of the Protein Society*, vol. 22, 03 2013.
- [10] H. J. Dyson and P. E. Wright, “Intrinsically unstructured proteins and their functions,” *Nature Reviews Molecular Cell Biology*, vol. 6, p. 197–208, 2005.
- [11] P. E. Wright and H. J. Dyson, “Intrinsically disordered proteins in cellular signaling and regulation,” *Nature Reviews Molecular Cell Biology*, vol. 16, no. 1, pp. 18–29, 2015.
- [12] M. M. Babu, “The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease,” *Biochemical Society Transactions*, vol. 44, pp. 1185–1200, 10 2016.
- [13] G. L. Dignon, R. B. Best, and J. Mittal, “Biomolecular phase separation: From molecular driving forces to macroscopic properties,” *Annual Review of Physical Chemistry*, vol. 71, no. 1, pp. 53–75, 2020.
- [14] J.-M. Choi, A. S. Holehouse, and R. V. Pappu, “Physical principles underlying the complex biology of intracellular phase transitions,” *Annual Review of Biophysics*, vol. 49, no. 1, pp. 107–133, 2020.

- [15] A. A. Hyman, C. A. Weber, and F. Jülicher, “Liquid-liquid phase separation in biology,” *Annual Review of Cell and Developmental Biology*, vol. 30, no. 1, pp. 39–58, 2014.
- [16] S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillén-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y.-T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman, and S. Alberti, “Rna buffers the phase separation behavior of prion-like rna binding proteins,” *Science*, vol. 360, no. 6391, pp. 918–921, 2018.
- [17] B. S. Schuster, G. L. Dignon, W. S. Tang, F. M. Kelley, A. K. Ranganath, C. N. Jahnke, A. G. Simpkins, R. M. Regy, D. A. Hammer, M. C. Good, and J. Mittal, “Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 21, pp. 11421–11431, 2020.
- [18] J. Forman-Kay and T. Mittag, “From sequence and forces to structure, function, and evolution of intrinsically disordered proteins,” *Structure*, vol. 21, no. 9, pp. 1492–1499, 2013.
- [19] Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, “Theories for sequence-dependent phase behaviors of biomolecular condensates,” *Biochemistry*, vol. 57, no. 17, pp. 2499–2508, 2018.
- [20] I. Pritišanac, R. M. Vernon, A. M. Moses, and J. D. Forman Kay, “Entropy and information within intrinsically disordered protein regions,” *Entropy*, vol. 21, no. 7, 2019.
- [21] R. M. Vernon and J. D. Forman-Kay, “First-generation predictors of biological protein phase separation,” *Current Opinion in Structural Biology*, vol. 58, pp. 88 – 96, 2019.
- [22] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, and A. A. Hyman, “A molecular grammar governing the driving forces for phase separation of prion-like rna binding proteins,” *Cell*, vol. 174, no. 3, pp. 688 – 699, 2018.
- [23] T. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. Bazett-Jones, T. Pawson, J. Forman-Kay, and A. Baldwin, “Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles,” *Molecular Cell*, vol. 57, no. 5, pp. 936 – 947, 2015.

- [24] E. W. Martin and T. Mittag, “Relationship of sequence and phase separation in protein low-complexity regions,” *Biochemistry*, vol. 57, no. 17, pp. 2478–2487, 2018.
- [25] C. E. Sing and S. L. Perry, “Recent progress in the science of complex coacervation,” *Soft Matter*, vol. 16, pp. 2885–2914, 2020.
- [26] H. Y. J. Fung, M. Birol, and E. Rhoades, “Idps in macromolecular complexes: the roles of multi-valent interactions in diverse assemblies,” *Current Opinion in Structural Biology*, vol. 49, pp. 36 – 43, 2018. Theory and simulation • Macromolecular assemblies.
- [27] S. Ambadipudi, J. Biernat, D. Riedel, E. Mandelkow, and M. Zweckstetter, “Liquid–liquid phase separation of the microtubule-binding repeats of the alzheimer-related protein tau,” *Nature Communications*, vol. 8, 12 2017.
- [28] X. Zhang, Y. Lin, N. A. Eschmann, H. Zhou, J. N. Rauch, I. Hernandez, E. Guzman, K. S. Kosik, and S. Han, “Rna stores tau reversibly in complex coacervates,” *PLOS Biology*, vol. 15, pp. 1–28, 07 2017.
- [29] Y. Lin, J. Mccarty, J. Rauch, K. Delaney, K. Kosik, G. Fredrickson, J. Shea, and S. Han, “Narrow equilibrium window for complex coacervation of tau and rna under cellular conditions,” *eLife*, vol. 8, 04 2019.
- [30] Y. Lin, Y. Fichou, Z. Zeng, N. Y. Hu, and S. Han, “Electrostatically driven complex coacervation and amyloid aggregation of tau are independent processes with overlapping conditions,” *ACS Chemical Neuroscience*, vol. 11, no. 4, pp. 615–627, 2020.
- [31] A. Patel, H. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Hein, S. Stoyanov, J. Mahamid, S. Saha, T. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. Royer, M. Weigert, E. Myers, S. Grill, D. Drechsel, A. Hyman, and S. Alberti, “A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation,” *Cell*, vol. 162, no. 5, pp. 1066 – 1077, 2015.
- [32] Y. Lin, Y. Fichou, and A. P. Longhini, “Liquid-liquid phase separation of tau driven by hydrophobic interaction facilitates fibrillization of tau,” *bioRxiv*, 2020.

- [33] M. Paloni, R. Bailly, L. Ciandrini, and A. Barducci, “Unraveling molecular interactions in liquid–liquid phase separation of disordered proteins by atomistic simulations,” *The Journal of Physical Chemistry B*, vol. 124, no. 41, pp. 9009–9016, 2020.
- [34] M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne, “Coexisting liquid phases underlie nucleolar subcompartments,” *Cell*, vol. 165, no. 7, pp. 1686 – 1697, 2016.
- [35] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, “Relation between single-molecule properties and phase behavior of intrinsically disordered proteins,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 40, pp. 9929–9934, 2018.
- [36] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, “Sequence determinants of protein phase behavior from a coarse-grained model,” *PLOS Computational Biology*, vol. 14, pp. 1–23, 01 2018.
- [37] S. Das, A. Amin, Y.-H. Lin, and H. Chan, “Coarse-grained residue-based models of disordered protein condensates: Utility and limitations of simple charge pattern parameters,” *PCCP*, 08 2018.
- [38] R. B. Best, W. Zheng, and J. Mittal, “Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association,” *Journal of Chemical Theory and Computation*, vol. 10, no. 11, pp. 5113–5124, 2014.
- [39] J. Huang, S. Rauscher, G. Nawrocki, R. Ting, M. Feig, B. de Groot, H. Grubmüller, and A. MacKerell, “Charmm36m: An improved force field for folded and intrinsically disordered proteins,” *Biophysical Journal*, vol. 112, pp. 175a–176a, 02 2017.
- [40] P. Robustelli, S. Piana, and D. E. Shaw, “Developing a molecular dynamics force field for both folded and disordered protein states,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. E4758–E4766, 2018.
- [41] J. Huang and A. D. MacKerell, “Force field development and simulations of intrinsically disordered proteins,” *Current Opinion in Structural Biology*, vol. 48, pp. 40 – 48, 2018. Folding and binding in silico, in vitro and in cellula • Proteins: An Evolutionary Perspective.

- [42] H. Wu, P. G. Wolynes, and G. A. Papoian, “Awsem-idp: A coarse-grained force field for intrinsically disordered proteins,” *The Journal of Physical Chemistry B*, vol. 122, no. 49, pp. 11115–11125, 2018.
- [43] Y.-H. Lin, J. Song, J. D. Forman-Kay, and H. S. Chan, “Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins,” *Journal of Molecular Liquids*, vol. 228, pp. 176 – 193, 2017. From simple liquids to macromolecular solutions: recent experimental and theoretical developments. In Honor of the 70th birthday of Vojko Vlachy.
- [44] A. Alexander-Katz, A. G. Moreira, S. W. Sides, and G. H. Fredrickson, “Field-theoretic simulations of polymer solutions: Finite-size and discretization effects,” *The Journal of Chemical Physics*, vol. 122, no. 1, p. 014904, 2005.
- [45] Y. Popov, J. Lee, and G. Fredrickson, “Field-theoretic simulations of polyelectrolyte complexation,” *Journal of Polymer Science Part B*, vol. 45, pp. 3223–3230, 2007.
- [46] R. A. Riggleman and G. H. Fredrickson, “Field-theoretic simulations in the gibbs ensemble,” *The Journal of Chemical Physics*, vol. 132, no. 2, p. 024104, 2010.
- [47] R. A. Riggleman, R. Kumar, and G. H. Fredrickson, “Investigation of the interfacial tension of complex coacervates using field-theoretic simulations,” *The Journal of Chemical Physics*, vol. 136, no. 2, p. 024903, 2012.
- [48] M. C. Villet and G. H. Fredrickson, “Efficient field-theoretic simulation of polymer solutions,” *The Journal of Chemical Physics*, vol. 141, no. 22, p. 224115, 2014.
- [49] A. Arora, J. Qin, D. C. Morse, K. T. Delaney, and G. H. Fredrickson, “Broadly accessible self-consistent field theory for block polymer materials discovery,” *Macromolecules*, vol. 49, no. 13, pp. 4675–4690, 2016.
- [50] R. A. L. Jones, *Soft condensed matter*. Oxford master series in condensed matter physics, Oxford [England] ;: Oxford University Press, 2002.

- [51] M. W. Matsen, “Field theoretic approach for block polymer melts: Scft and fts,” *The Journal of Chemical Physics*, vol. 152, no. 11, p. 110901, 2020.
- [52] T. Hirayama, “Classical Statistical simulation of Quantum Field Theory,” *arXiv*, 2019.
- [53] L. L. Salcedo, “Representation of complex probabilities,” *Journal of Mathematical Physics*, vol. 38, no. 3, pp. 1710–1722, 1997.
- [54] S. J. Wodak and J. Janin, “Analytical approximation to the accessible surface area of proteins,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 4, pp. 1736–1740, 1980.
- [55] T. J. Richmond, “Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect,” *Journal of Molecular Biology*, vol. 178, no. 1, pp. 63 – 89, 1984.
- [56] D. Eisenberg and A. McLachlan, “Solvation energy in protein folding and binding,” *Nature*, vol. 319, no. 6050, p. 199—203, 1986.
- [57] T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga, “Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 10, pp. 3086–3090, 1987.
- [58] W. Hasel, T. F. Hendrickson, and W. Still, “A rapid approximation to the solvent accessible surface areas of atoms,” *Tetrahedron Computer Methodology*, vol. 1, no. 2, pp. 103 – 116, 1988.
- [59] R. K. Das and 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*, vol. 110, no. 33, pp. 13392–13397, 2013.
- [60] J. Hubbard, “Calculation of partition functions,” *Phys. Rev. Lett.*, vol. 3, pp. 77–78, Jul 1959.
- [61] R. L. Stratonovich, “On a method of calculating quantum distribution functions,” *Soviet Physics Doklady*, vol. 2, p. 416, 1957.

- [62] D. Weingarten, “Complex probabilities on R^N as real probabilities on C^N and an application to path integrals,” *Phys. Rev. Lett.*, vol. 89, p. 240201, Nov 2002.
- [63] A. Kolmogoroff, “Über die analytischen methoden in der wahrscheinlichkeitsrechnung,” *Mathematische Annalen*, vol. 104, pp. 415–458, 1931.
- [64] H. Cai, B. Gabryelczyk, M. S. Manimekalai, G. Grüber, S. Salentinig, and A. Miserez, “Self-coacervation of modular squid beak proteins—a comparative study,” *Soft Matter*, vol. 13, no. 42, pp. 7740–7752, 2017.
- [65] Y. Lin, Y. Fichou, A. P. Longhini, L. C. Llanes, Y. Yin, G. C. Bazan, K. S. Kosik, and S. Han, “Liquid-liquid phase separation of tau driven by hydrophobic interaction facilitates fibrillization of tau,” *bioRxiv*, 2020.
- [66] B. Gabryelczyk, H. Cai, X. Shi, Y. Sun, P. J. Swinkels, S. Salentinig, K. Pervushin, and A. Miserez, “Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides,” *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019.
- [67] P. A. Chong, R. M. Vernon, and J. D. Forman-Kay, “Rgg/rg motif regions in rna binding and phase separation,” *Journal of Molecular Biology*, vol. 430, no. 23, pp. 4650 – 4665, 2018. Phase Separation in Biology and Disease.
- [68] S. Xue, R. Gong, F. He, Y. Li, Y. Wang, T. Tan, and S.-Z. Luo, “Low-complexity domain of u1-70k modulates phase separation and aggregation through distinctive basic-acidic motifs,” *Science advances*, vol. 5, no. 11, p. eaax5349, 2019.