New Journal of Physics
The open access journal at the forefront of physics

PAPER
Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins

Tyler S Harmon1,2,3, Alex S Holehouse© and Rohit V Pappu©

1 Max-Planck-Institute for the Physics of Complex Systems, D-01187 Dresden, Germany
2 Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130, United States of America
3 Author to whom any correspondence should be addressed.
E-mail: pappu@wustl.edu

Keywords: biomolecular condensates, multiphase equilibria, spatially organized droplets, lattice model, intrinsically disordered linkers

Abstract
Intracellular biomolecular condensates are membraneless organelles that encompass large numbers of multivalent protein and nucleic acid molecules. The bodies assemble via a combination of liquid–liquid phase separation and gelation. A majority of condensates included multiple components and show multilayered organization as opposed to being well-mixed unitary liquids. Here, we put forward a simple thermodynamic framework to describe the emergence of spatially organized droplets in multicomponent systems comprising of linear multivalent polymers also known as associative polymers. These polymers, which mimic proteins and/or RNA have the architecture of domains or motifs known as stickers that are interspersed by flexible spacers known as linkers. Using a minimalist numerical model for a four-component system, we have identified features of linear multivalent molecules that are necessary and sufficient for generating spatially organized droplets. We show that differences in sequence-specific effective solvation volumes of disordered linkers between interaction domains enable the formation of spatially organized droplets. Molecules with linkers that are preferentially solvated are driven to the interface with the bulk solvent, whereas molecules that have linkers with negligible effective solvation volumes form cores in the core–shell architectures that emerge in the minimalist four-component systems. Our modeling has relevance for understanding the physical determinants of spatially organized membraneless organelles.

Introduction
The molecular components of cellular matter comprise of biomacromolecules, salts, and osmolytes [1–3]. Subcellular organelles house specific subsets of cellular matter. Many organelles are membrane-bound and the membranes serve as physical barriers to separate the organelar components from the cellular milieu. However, there are many more occurrences of membraneless organelles when compared to membrane-bound organelles [4–8]. Although knowledge of membraneless organelles dates back to the early characterizations of cellular ultra-structures [9], the material properties and physical driving forces for the assembly of membraneless organelles largely remained unclear until the past decade [10, 11]. Starting with the physical characterization of germline P-granules [12], which are protein-RNA bodies, there has been a surge of interest in characterizing a variety of membraneless organelles in different organisms [10, 11].

Recent advances have helped in uncovering many of the key determinants of material properties and assembly mechanisms of well-established membraneless organelles [10, 11]. Most membraneless organelles are referred to as biomolecular condensates [10, 11] because they appear to be non-stoichiometric assemblies of proteins and nucleic acids that are akin to networked fluids that form via liquid–liquid phase separation (LLPS) and gelation [13–16]. Here, LLPS refers to a density transition, which leads to the formation of spherical droplets.

© 2018 The Author(s). Published by IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft
rich in macromolecules that are in equilibrium with dilute phases that are rich in solvent [17]. Gelation is a
networking transition defined by the formation of droplet-spanning networks [18]. These networks are
characterized by physical crosslinks amongst the constituent macromolecules. The extent of crosslinking and
average timescales of crosslinks contributes to the material properties of condensates.

Protein and nucleic acid molecules that make up networked fluid-like condensates are classifiable into two
main categories, namely scaffolds and clients [11]. Scaffolds are macromolecules that drive the formation of
condensates whereas clients selectively partition into condensates based on their chemical potentials and the
chemical potentials of scaffolds. Bona fide scaffolds are multivalent associative polymers [19]. These molecules
comprise of multiple structural domains or short linear motifs (SLiMs) that participate in stereospecific homo-
or heterotypic interactions giving rise to physical crosslinks that define the extent of networking within a
condensate. Co-opting nomenclature from the literature on associative polymers, we designate domains and/or
SLiMs as stickers [20]. In archetypal scaffolds such as linear multivalent proteins, the spacers are often intrinsically disordered regions where sequence–intrinsic conformational heterogeneity is a defining feature of these regions [21]. We recently developed a model for the
phase behavior of linear multivalent proteins and proposed that details of the phase diagrams are jointly
determined by the number (valence) of stickers, the affinities between stickers, and the sequence-encoded
effective solvation volumes of linkers/spacers [16].

An improved understanding of the driving forces for phase transitions of scaffold proteins and RNA
molecules has emerged from in vitro reconstitutions using individual or pairs of scaffold molecules and/or by
monitoring selectively labeled scaffold molecules in vivo [10, 11]. These approaches essentially treat condensates
as pseudo two-component systems [13]. However, in reality, condensates are multicomponent entities. Indeed,
many membraneless organelles encompass more than 10^2 types of scaffold and client molecules [22–25]. A true
understanding of the driving forces for condensate formation and the determinants of material properties of
condensates requires a deeper understanding of multiphasic equilibria of multicomponent systems.

A simple assessment of the complexity may be stated as follows: if there are n_s supersaturated scaffold
molecules in the cellular milieu, then the Gibbs phase rule tells us that at constant temperature and pressure, and
in the absence of chemical reactions, there can be no more than n_s coexisting phases. The Gibbs phase rule places
a precise numerical constraint on the number of thermodynamic degrees of freedom within multicomponent,
multiparticip systems. The degrees of freedom refer to the number of chemical potentials of scaffold molecules
that can be independently specified by gene expression, degradation, modification, or other ways of regulating
intracellular concentrations of scaffold molecules in a closed system.

A key question is as follows: in the absence of active processes, will all n_s scaffold molecules condense into a
single, homogeneously mixed unitary droplet? Alternatively, will there be n_p phases, each enriched in individual
scaffold molecules or will reality be specific to distinct organelles with the situation being somewhere between
the unitary condensed phase and n_p distinct phases? These questions were the recent focus of a mean field model
developed by Jacobs and Frenkel [26]. In their model, each component may be defined by a distinct preferential
interaction energy that is sampled from a Gaussian distribution. Here, preferential interaction defines the mean
interaction strength between pairs of components after considering the interplay amongst the components, the
bulk solvent, and all combinations of components. These preferential interactions can be attractive or repulsive.
Two limiting scenarios emerge from the mean field model of Jacobs and Frenkel: in one limiting scenario, all
components undergo condensation into a single well-mixed multicomponent droplet that is in equilibrium
with a dispersed phase. In the opposite scenario, a small number of components undergo demixing to form
individual assemblies that coexist with one another. The key parameters that determine the differences between
the two limiting scenarios are the parameters of the Gaussian distributions from which preferential interaction
energies are randomly selected. Jacobs and Frenkel find that for n_s supersaturated scaffold molecules, there is a
narrow range of preferential interactions delineating the sharp boundary between the formation of a single,
homogeneous, well-mixed, multicomponent albeit unitary condensate versus the formation of ∼n_p demixed
phases. Similar insights, albeit with a detailed accounting for sequence-specific effects have emerged from the
mean field random phase approximation developed by Lin et al [27]. The key message emerging from mean field
models is that sequence-encoded interactions ascribed to scaffold molecules provide the requisite tunability and
specificity for describing the extent of condensation/demixing in multicomponent systems.

Membraneless organelles provide a mechanism for compartmentalization of specific cellular components.
Interestingly, membraneless organelles are themselves often characterized by internal spatial organization
[28–31]. Recent studies have shown that protein and RNA molecules can show distinct spatial distributions
within key membraneless organelles. In the nucleolus, which is the largest nuclear body, there are distinct
nucleolar sub-compartment, which have been described as co-existing liquid phases [28]. Similarly, nuclear
speckles, paraspeckles, and stress granules have either simple core–shell architectures or more complex multi-
layered structures with a distinct radial distribution of different components [29, 30, 32]. The key question is if
there is a simple thermodynamic framework for describing the observation of spatial organization in
membraneless organelles as has been recently proposed [28, 29] or if the maintenance of energy gradients is essential for realizing such structures [30]?

As a simple starting point, we consider the case of a three-component system comprising of the bulk solvent (encompassing water, salts, and osmolytes), and a pair of scaffold molecules $s_1$ and $s_2$. The concentrations or more precisely the activities of $s_1$ and $s_2$, designated as $a_1$ and $a_2$, are the thermodynamic degrees of freedom at fixed temperature and pressure. There are eight distinct scenarios for phase equilibria in this single three-component system and these are depicted in figure 1. Mean field theories based on interactions drawn from a statistical distribution or averaged over specific sequences account for the scenarios depicted in figures 1(a)–(e). However, the biologically relevant scenarios (figures 1(f), (g)) of spatially organized core–shell architectures cannot be described by mean field theories [33]. In this work, we show that explicit consideration of sequence-encoded solvation preferences of disordered linkers in linear multivalent proteins provides a necessary and sufficient minimalist numerical framework for modeling the emergence of spatially organized coexisting phases.

We consider a simple four-component generalization of the synthetic three-component system of linear multivalent proteins that has been established by the Rosen lab as a minimalist system for observing phase separation and gelation in vitro [15]. In the original system, linear multivalent proteins, which are concatemers of SH3 domains designated as poly-SH3 proteins undergo LLPS and gelation through cooperative interactions with concatemers of proline-rich modules (PRMs). These multivalent ligands are designated as poly-PRMs. This system must be modeled as a three-component system given the presence of solvent. This point was made clear in recent work, in which we developed a coarse-grained numerical description of the poly-SH3/poly-PRM system [16]. We showed that while valence is important for determining the gel point, it is the physical properties of linkers between domains in poly-SH3 and poly-PRM proteins, and the affinities between SH3 domains and PRMs that determine if gelation occurs with or without phase separation. Linker length has an important role, but for biologically relevant, finite-sized systems the crucial parameter is the sequence-specific effective solvation volume ($v_{se}$) of each linker. This parameter is defined as: $v_{se} = \int (1 - \exp[-\beta W(r)]) d^3r$. Here, $W(r)$ is the potential of mean force for effective solvent-mediated interaction between pairs of linker residues.

The effective solvation volume, which is identical to the excluded volume ($v_{es}$) used in the polymer physics literature [17], can be negative, zero, or positive. The effective solvation volume is negative ($v_{es} < 0$) for compact linkers that prefer self-interactions when compared to interactions with the solvent. Conversely, $v_{es} > 0$ for stiff/expanded linkers that interact preferentially with the surrounding solvent; if linker–solvent and linker–linker interactions are counterbalanced, then $v_{es} \approx 0$. The amino acid compositions and sequence features of linkers determine the sign and magnitude of $v_{es}$. Our recent analysis showed that $v_{es} \geq 0$ for a majority of naturally occurring linkers in linear multivalent proteins [17]. Importantly, for a given linker length, especially if lengths are in the 20–50 residue range, gelation is enabled by LLPS in the Flory random coil (FRC) limit where $v_{es} \approx 0$. We use the term FRC as opposed to Gaussian chain because we explicitly used Flory’s rotational

![Figure 1](image-url)

Figure 1. Realizable phase equilibria for a three-component system comprising of solvent (white), scaffold molecule $s_1$ (blue), and scaffold molecule $s_2$ (yellow). (a) Depicts the case of a uniformly mixed one-phase system; (b) depicts the condensation limit of Jacobs and Frenkel whereby the dense droplet phase is enriched in the scaffold molecules whereas the dispersed phase is enriched in solvent; (c) depicts the demixing of $s_1$ from the bulk phase comprising of solvent and $s_2$; (d) depicts the opposite scenario of demixing of $s_2$ from the bulk phase comprising of solvent and $s_1$; (e) depicts the demixing limit of Jacobs and Frenkel whereby distinct scaffold-enriched droplets coexist with a solvent-enriched bulk phase; (f) depicts the formation of spatially organized droplet with preferential solvation of $s_3$ leading to an interface between the $s_3$-enriched shell and the solvent, which is to be contrasted by wetting of the $s_2$-enriched core by the shell because in the hierarchy of interactions involving $s_2$, the homotypic $s_2-s_2$ interactions are most favorable, when compared to the heterotypic $s_2-s_3$ interactions, which are of intermediate strength, followed by the unfavorable $s_2$-solvent interactions; (g) depicts the complement of the scenario depicted in (f); (h) depicts the scenario that is a specialization of the scenario in (e) whereby the distinct $s_1$ and $s_2$-rich droplets can wet one another due to relatively favorable heterotypic $s_1-s_2$ interactions when compared to the unfavorable interactions of the scaffold molecules with solvent.
isomeric approximation to generate distributions that were compatible with the theta-solvent limit [34, 35]. As \( v_{\text{c}} \) becomes positive, due to preferential solvation of linkers, the driving forces for phase separation are weakened and the gel point is shifted to higher protein concentrations. In the limit of linkers that are self-avoiding random coils (SARC) where \( v_{\text{c}} = v_{\text{c,min}} \), gelation occurs without phase separation. The SARCs are distinct from self-avoiding walks [36], which are mathematical models to describe paths on a lattice that do not intersect with one another. We do not enumerate all possible self-avoiding configurations of a chain on a lattice. Instead, SARC refers to atomistic ensembles that we obtain using a combination of Metropolis Monte Carlo simulations and potentials based on the purely repulsive arm of the Lennard–Jones potential [16, 37, 38]. Therefore, although SARC belongs to the same universality class as self-avoiding walks (SAWs) [39], we choose the designation of SARC to avoid conflation with the mathematical attributes of SAWs.

The valence of domains/interaction modules, the affinities of these stickers, the lengths of linkers, and their effective solvation volumes open the door to multiple routes for designing a rudimentary four-component system based on the poly-SH3 and poly-PRM system. For example, we can keep the linkers fixed and change the valence of SH3 domains for some fraction of poly-SH3 proteins. Alternatively, we can keep the valence of SH3 domains fixed and change the physical properties of linkers or keep the valence and linker properties fixed and modulate the affinities of some SH3 domains for PRMs. Here, we investigate the phase behavior of each of these different four-component systems by generalizing the model developed to study the impact of disordered linkers on the phase behavior of the binary poly-SH3/poly-PRM system [17].

We show that a minimal four-component system (consisting of two distinct types of poly-SH3 proteins, a poly-PRM protein, and an effective solvent) supports the spontaneous formation of spatially organized droplets. Specifically, droplet organization requires (i) poly-SH3 proteins with SARC linkers, which we designate as [poly-SH3]_i, (ii) poly-SH3 proteins with FRC linkers, which we designate as [poly-SH3]_f, and (iii) poly-PRM proteins with FRC linkers. The core–shell architecture comprises of [poly-SH3]_i proteins in the core, [poly-SH3]_f in the shell, and poly-PRM proteins partitioning between the core and shell. The core–shell structure emerges as a consequence of the preferential solvation of SARC linkers in [poly-SH3]_i proteins and the zero effective solvation volumes of FRC linkers in [poly-SH3]_f proteins. The FRC linkers enable cooperative density transitions required for forming the core, whereas a clean delineation of the interface with the bulk solvent leads to the formation of a shell that is enriched in [poly-SH3]_f proteins. Other properties such as differences in valence, interaction strengths, and linker length do not influence spatial organization.

Our findings provide a conceptual framework for identifying sequence-encoded features of disordered linkers that might drive the formation of spatially organized droplets in multi-component systems comprising of multivalent associative polymers. The effective solvation volumes of linkers in linear multivalent proteins can be dynamically modulated by post-translational modifications, through interactions with binding partners, or through active enzyme-catalyzed energy-dependent processes. Thus spatial organization of droplets is, in principle, biologically tunable. In the sections that follow, we present the details of the design of our model, a detailed account of the results, and a discussion placing our results in the broader biological context.

**Methods**

We developed and implemented a lattice based coarse-grained simulation engine to explore the impact of intrinsically disordered linkers on the spatial organization of droplets formed through LLPS. This model has been employed to describe spatial organization in a number of biological systems [16, 28, 29], and to examine how disordered linkers influence the coupling between phase separation and gelation [16]. The model uses a periodic cubic lattice, where multivalent proteins consist of interacting domains connected by flexible linkers. Each domain occupies a single site on the lattice. There are two limiting forms of linkers. In the case of SARC linkers, each linker bead occupies a single site on the lattice, the linker beads must be directly adjacent to one another, and the linker may be an arbitrary number of beads long. The linker beads have a finite volume corresponding to that of a single lattice site. They are otherwise inert in that their interactions with one another, and the linker may be an arbitrary number of beads long. The linker beads have a finite volume corresponding to that of a single lattice site. They are otherwise inert in that their interactions with one another or with domains are neither attractive nor repulsive pairwise interactions. Each linker bead corresponds to roughly 7-residues [16]. In the case of FRC linkers, the linker is described by a three-dimensional infinite square well potential that tethers the two domains it connects together, where the linker length is the maximum distance the domains can be separated along any axis. Unlike SARC linkers, FRC linkers do not occupy any volume on the lattice. For an FRC linker connecting two domains, this ensures that the domain beads are always within a set distance of one another but does not exert any other influence. Modeling FRC linkers using distance restraints alone effectively means that such linkers are phantom tethers, which is in accord with the Flory designation of a random walk [35]. As noted in the introduction, SARC linkers belong to the universality class of self-avoiding walks, whereby the radii of gyration and mean end-to-end distances scale as \( N^{0.59} \) with linker length \( N \) [36, 38]. The FRC linker belongs to the universality class of three-dimensional random walks, whereby
the radii of gyration and mean end-to-end distances scale as $N^{0.5}$. This formalism is identical to that of our previous work [16].

Domains engage in inter- and intra-molecular interactions with adjacent domains. The interactions are defined based on a binding energy, measured in units of thermal energy ($kT$). Here, the terms inter- and intra-refer to interactions between domains on adjacent proteins (inter), or on the same protein (intra). Each domain can only engage in a single interaction at a time, but can be spatially adjacent to multiple domains without engaging in an interaction. Domains and linkers do not engage in any additional form of interaction, beyond the excluded volume effects of the SARC linkers. The solvent is treated quasi-implicitly. Lattice sites that are unoccupied by the polymers are sites that are occupied by the solvent. In this work, we are assessing how the phase behavior of linear multivalent polymers is impacted as we fix the solvent-mediated affinities between interaction domains and assess how the effective solvation volumes of linkers, which are governed by the sequences of linkers, alter the overall phase behavior. In this model, the sequence of the linker dictates whether the solvent is poor (negative $v_{es}$), theta ($v_{es} \approx 0$), or good (positive $v_{es}$) for the linker. We are deploying a coarse-grained model to assess the effects of linkers and there are various ways to implement such a model. One approach would be the titration of solvent quality by assigning a set number of sites to all linker residues/beads and modulating the effective interaction strengths from being negative, zero, or positive, as a means to titrate the solvent quality and hence the effective solvation volume. Such an approach is in fact equivalent to the model we have developed. Given this equivalence, we selected the computationally inexpensive route that avoids the need for extensive parameterization of the pairwise attractions between linker beads to find the theta point, which would correspond to zero $v_{es}$.

The pairwise interactions we assign are effective interactions, in that they represent the favorable interactions between SH3 and PRM sites upon consideration of the multi-way interplay amongst protein sites (SH3 and PRM), linker sites, and the solvent sites. Setting an effective interaction parameter to zero does not mean that the sites to do not interact. It means, that they contribute to the overall configurational and solvent entropies. The model, which is intended to mimic so-called associative polymers, asserts that the specific interactions between SH3 and PRM sites are favored by the chosen amount over the interactions between linker sites, between linker sites and the domains, or even among the domains themselves. Adding SH3:SH3 or PRM:PRM interactions would lead to increased ruggedness on the energy landscape. Similarly, setting the other energies to a non-zero value will simply renormalize the baseline, especially if we follow the experimental data and set the SH3:PRM interaction to be more favorable than SH3:SH3, PRM:PRM, SH3: linker, PRM: linker, or linker:linker interactions, which is the basis for our model that mimics associative polymers and their phase behavior.

The system is propagated using Monte Carlo sampling where moves are accepted/rejected via Metropolis acceptance criteria with a correction for move proposal probabilities for satisfying detailed balance, as described previously [16]. In addition to a series of moves that perform local chain re-arrangement and chain translation/rotation, specific moves also randomly change the interaction state of two adjacent beads without updating their coordinate positions on the lattice. Such moves represent binding/unbinding events. The details of the Monte Carlo move sets were discussed in detail in recent published work [16].

We examined three distinct types of multivalent proteins: a poly-PRM scaffold component, and two types of poly-SH3 proteins. The only attractive interactions in the system come from SH3-PRM interaction, while all other interactions are solely due to excluded volume effects. Given the number of possible model parameter combinations, we chose to focus on ‘inert’ scaffolds, where the reaction coordinates of interest are the protein valence, the effective solvation volumes of linkers, and the strengths of SH3-PRM interactions. For simulations used to examine the role of linker type (SARC versus FRC) on spatial organization, we used linkers of length 6 unless noted otherwise. To interpolate between SARC and FRC linkers, we systematically replace individual SARC linker beads with an implicit square-well potential, creating a hybrid linker. These square-well potentials are distributed as evenly as possible over the linker while ensuring the linker length remains fixed. Poly-PRM proteins always encompass FRC linkers, while poly-SH3 proteins utilize both FRC and SARC linkers, as designated. Unless explicitly noted, each protein has a valence of five domains.

All simulations were run for $10^{10}$ Monte Carlo steps. Simulation results are taken from the average of the second last half of each simulation, over which global features are no longer changing. We monitored the standard deviations of displacements of each of the molecules in the simulation. In accord with unhindered motions, we observed free sampling of individual polymers across the entire simulation volume. This confirms that the dynamics remain unhindered by the formation of dense phases.

With the exception of the volume parameterizing simulations and the valence dependency simulations, each simulation had $2 \times 10^3$ poly-PRM proteins, $10^4$ [poly-SH3]$_3$ proteins, and $10^3$ [poly-SH3]$_3$ proteins for a total of $4 \times 10^3$ proteins. Simulations that were designed to test the formation of spatially organized droplets were performed on lattices of length 91 lattice units. These dimensions are large enough to allow formation of phase-separated droplet with minimal boundary condition artifacts.
For simulations examining the role of valence on spatial organization, poly-PRM had a valence of eight with 2 × 10^3 proteins, poly-SH3 had a valence of 8 with 1 × 10^3 proteins, and poly-SH3_F had a valence of 4 with 2 × 10^3 proteins. This ensures that while the valence of the different poly-SH3 proteins is varied, the absolute concentration of SH3 domain remains the same. For the simulations in which the bulk solvent phase is removed, we calculated the expected protein concentration needed to ensure the lattice dimensions would yield a one-phase system where the protein density matched the droplet density in the center of the droplet. This was accomplished by simulating 2 × 10^3 poly-PRM proteins and 10^3 poly-SH3 proteins as well as 10^3 poly-SH3_F proteins on a lattice of length 154 lattice units with periodic boundary conditions and calculating the radial density function for the protein from the center of mass of the droplet. This yields a sigmoidal density, which provides a direct readout of the densities in the droplet phases. Accordingly, the high-density simulations that lack a bulk phase are then simulated with a volume and protein concentration such that the densities of the two phases will fill the entire simulation box. These volumes range from 44 to 61 lattice units.

### Results

In most of the simulations, each multivalent protein is comprised of five domains connected by flexible linkers. The poly-PRM proteins are shared scaffold components that can bind either of the other two proteins ([poly-SH3]_S or [poly-SH3]_F). Poly-PRM will phase separate in the presence of either poly-SH3 proteins, but does not do so in isolation. Similarly, both types of poly-SH3 protein do not undergo phase separation when in isolation or when mixed with one another. A schematic of the system components and the types of droplet organization is shown in figure 2.

**Differences in effective solvation volumes of linkers drive droplet organization**

We first performed simulations with 2 × 10^3 poly-PRM proteins with FRC linkers, 10^3 poly-SH3 proteins with FRC linkers and 10^3 poly-SH3 proteins with SARC linkers. All proteins were valence 5. In previous work, we showed that a single uniform droplet forms in simulations that contain poly-SH3 proteins with only FRC linkers [16]. Here, despite having identical binding domains, the poly-SH3 proteins with FRC linkers were sequestered into the core of the droplet while the poly-SH3 proteins with SARC linkers were predominantly found in the periphery, leading to a core-and-shell architecture. Figure 3(a) quantifies the protein-specific density as a function of radial position, while figures 3(b) and (c) show snapshots of the droplet, illustrating the resulting core–shell architecture.

We next examined the physical origins of this spatial organization. To do so, we performed simulations at large total protein volume fractions such that there is no longer a bulk solvent phase, as described in the methods.
section. Under these conditions, despite exactly the same sets of components, we no longer observed spatial organization, and instead a single uniform phase is observed. This demonstrates that droplet organization is driven by an effective interaction between the SARC-linker containing poly-SH3 proteins and the bulk solvent. The absence of an interface between poly-SH3 S proteins and the bulk solvent leads to the loss of a thermodynamic driving force for the two poly-SH3 proteins to demix. These results are shown in figures 3(e), (f).

These results suggest that linker-mediated droplet organization is not driven by interactions between the two poly-SH3 proteins. Instead, by creating a layer of poly-SH3 proteins at the droplet periphery, the surface tension of the droplet is reduced. These results imply that this level of spatial organization should be limited to a small number of polymer lengths thick at the droplet-bulk interface. As the system size (and consequently droplet radius) is increased, the thickness of the outer shell that is enriched in the poly-SH3 proteins should stay roughly constant, while the remaining material will be split between the core of the droplet and the bulk solution. This effect should be manifest on the order of several tens of nanometers. Because of volume scaling, we expect this effect to become realized on the order of tens of thousands of proteins, which is outside of our current computational capabilities.

We sought to assess the dependence of our results on the model parameters. We varied the interaction strengths between PRM and SH3 binding domains and the effective solvation volume of the SARC linkers. Figure 4 demonstrates that the interaction strength between binding domains has almost no impact on the
observed droplet organization (compare results along rows). As expected, the effective solvation volume has an impact (compare along columns), with spatial organization being more robust and well defined in the limit of the full SARC linker (5 bead) for \([\text{poly-SH3}]_5\) proteins. As the effective solvation volumes of linkers are reduced by replacing explicit beads with implicit square well potentials, the distinction between the \([\text{poly-SH3}],[5]_5\) and \([\text{poly-SH3}],[5]_5\) proteins becomes less pronounced. However, even for linkers with a relatively weak effective solvation volume, spatial separation is still achieved. To further explore these results, we repeated the same set of simulations at protein concentrations that are consistent with the droplet interior (i.e., in the absence of a bulk solvent phase). Figure 5 shows the equivalent radial density analysis, illustrating that spatial organization depends on the presence of a bulk solvent phase to organize the droplet interior.

Valence of multivalent proteins with equivalent binding domains does not drive droplet organization

Having established that the effective solvation volume can play a role in driving the formation of spatially organized droplets, we investigated if multivalent proteins with distinct valence could also drive spatial organization. To do this, we performed simulations with \(2 \times 10^3\) valence 8 poly-PRM proteins, \(10^3\) valence 8 \([\text{poly-SH3}],[5]_5\) proteins, and \(10^3\) valence 4 \([\text{poly-SH3}],[5]_5\) proteins. All three proteins had identical FRC linkers, so the only difference between the two types of poly-SH3 proteins in this system was the valence. To ensure that our conclusions were not impacted by a specific combination of linker lengths and interaction strengths, we performed a series of simulations in which we systematically varied the linker lengths between the stickers and interaction strengths of the stickers. For all combinations of parameters, we observed only well-mixed droplets (figure 6). These results suggest that, to a first approximation, valence of otherwise equivalent linear multivalent proteins does not provide a mechanism to drive spatial organization of a phase-separated droplet.

Competition between multivalent proteins with different binding strengths for a common scaffold does not drive droplet organization

Finally, we asked if differences in interaction strengths could drive spatial organization within a droplet. We performed simulations with \(2 \times 10^3\) valence 5 poly-PRM proteins, \(10^3\) valence 5 \([\text{poly-SH3}],[5]_5\) proteins where the PRM-SH3 interaction strength was varied, and \(10^3\) valence 5 \([\text{poly-SH3}],[5]_5\) proteins with a fixed interaction strength of \(-2kT\). All three proteins had FRC linkers of length 6. As before, to ensure that our conclusions are not

---

Figure 4. Impact of model parameters, specifically the linker effective solvation volume and SH3:PRM binding strength for low density simulations in which a droplet is able to form. Two-dimensional titration of solvation volume and binding strength illustrates that spatial organization is robust to binding strength, but is systematically weakened as the SARC linker approaches the FRC limit.
simply a consequence of a specific parameter set, we performed a series of simulations in which we systematically varied the linker lengths as well as the interaction strength between the SH3 domains and PRMs. As with our results exploring the impact of valence, a single homogenous droplet formed for all simulations (Figure 7). These results suggest that competition between binding partners for a common scaffold is insufficient for driving spatial organization in phase-separated droplets.

Discussion

Membraneless organelles typically consist of more than $\sim 10^2$ scaffold and client components [10, 11, 28–30, 40–43]. Accordingly, the droplets that form need not conform to the behaviors expected for unitary liquids, which would be a uniform distribution of components across the droplet. In accord with this expectation, recent studies have shown that key membraneless organelles demonstrate spatial organization, with certain components preferentially localizing to distinct regions in the droplet [28, 29, 32]. Despite the inherent chemical complexity associated with membraneless organelles in vivo, significant insights regarding the molecular driving forces that control biological phase separation have been obtained through the study of reconstituted droplets that contain only one or two scaffold molecules [44–55]. This suggests that while simplified systems will presumably fail to capture the true complexity of these organelles, they can provide a tractable platform upon which specific hypotheses can be tested. In keeping with this ansatz, we have implemented a minimalist abstraction of a four-component system of linear multivalent proteins to uncover the physical determinants of spatial organization. Four-component systems characterized by differences in binding affinity and valance failed to generate core–shell architectures. However, four-component systems, characterized by differences in effective solvation volumes of disordered linkers between domains readily give rise to spatially organized droplets. The key ingredient is the differential solvation preferences of linkers whereby proteins with SARC or SARC-like linkers are preferentially solvated, whereas proteins with FRC or FRC-like linkers partition to the core because the zero effective volume of FRC linkers implies that they are agnostic about being solvated/
Figure 6. Parameter titration of linker length and SH3:PRM binding strength for low density simulations in which the poly-SH3 proteins have a different valence. Two-dimensional titration of linker length and binding strength for systems in which the two poly-SH3 proteins are of different valence. No combination of parameters examined yields spatially organized droplets.

Figure 7. Parameter titration of linker length where the PRM:SH3 binding interaction is allowed to vary for one poly-SH3 protein but not the other. Two-dimensional titration of solvation volume and binding strength illustrates that spatial organization is not obtained if multivalent proteins are a different valence but with identical linkers. In all cases, a uniform droplet (that contains all three protein components) is obtained.
desolvated. The equivalent interactions between the SH3 domains and PRMs lead to a sharing of ligands and a wetting of the core by the shell.

Our model is a simplified representation of linear multivalent proteins. The most obvious mapping between our ‘multivalent proteins’ and a real biological system is the system of folded domains connected by disordered linkers as depicted in figure 8(a) [15, 56]. An alternative mapping would be one in which the binding domains represent SLiMs that can drive interaction between fully disordered regions, as depicted in figure 8(b). Such a model is more consistent with observations of fully disordered low complexity domains that can drive phase separation through homotypic or heterotypic interactions [13, 54, 57]. Regardless of the specific mapping between our model and the biological system of interest, the relevant take away from our work remains the same: modulating the effective solvation volumes for regions of proteins that would not appear to act as key drivers of phase separation, i.e., disordered linkers, can give rise to spatially organized protein droplets. The effective solvation volume associated with a linker is not a fixed parameter, but it is determined by the interplay amongst linker–linker, linker–solvent, and solvent–solvent interactions. Consequently, the effective solvation volumes can be modulated in different ways. Post-translational modifications could increase or decrease the effective solvation volume. Multi-site phosphorylation of disordered linkers could increase the effective solvation volume through a combination of charge repulsion, preferential solvation, and changes to secondary structure, as illustrated in figure 8(c) [58–60]. In a similar vein, other types of post-translational modifications such as acetylation or methylation and the presence or absence of binding partners may have direct or indirect impacts on the effective solvation volume. Finally, distinct isoforms could in principle give rise to proteins with identical binding domains but distinct linkers, providing a route for post-transcriptional regulation of spatial organization. As an example, the protein PML, a key component of PML–nuclear bodies, has 12 distinct isoforms, while the nuclear speckle scaffolding protein SON has 10 [29, 56].

In our previous work, we demonstrated that the effective solvation volume determines whether gelation occurs with or without phase separation [16]. We found that systematically increasing the effective solvation volume of linkers in multivalent proteins could increase the saturation concentration, effectively suppressing phase separation (figure 10 in [16]). Here we report that in a multi-component droplet, multivalent proteins with a high effective solvation volume can drive spatial organization and preferentially localize to the droplet periphery forming a layer that wets the core of the spatially organized droplet. Taken together, our results show how disordered linkers can tune both the appearance and spatial organization of membraneless organelles even when the linkers do not engage in the associative interactions that drive phase transitions.

The majority of membraneless organelles studied in vivo also contain RNA, although this is not necessarily a requirement for phase separation and biomolecular condensate formation, at least in vitro [15]. The framework described in this work considers two components that we nominally designate as proteins. However, the same framework would also be directly applicable to a system comprising of protein and RNA molecules. In several cases, well-defined binding sites have been identified on RNA molecules, and these binding sites are necessary for the recruitment of additional protein [61]. We speculate that the sequence-intrinsic effective solvation volumes of RNA stretches that lie between binding sites could influence the overall phase behavior, as has been demonstrated in recent studies [62]. Alternatively, given the fact that many multivalent proteins that drive phase separation contain multiple folded or disordered RNA recognition motifs (RRMs) that bind RNA, our results are consistent with a model in which RNA–protein interactions provide a key driving force for phase separation, and the effective solvation volumes of disordered linkers that connect RRMs can enhance, suppress, and

Figure 8. Routes to mapping biologically relevant scenarios for linear multivalent proteins systems to the model used in this work. Panel (a) describes one possible mapping, in which the binding domains in our system are folded protein domains and the linkers are intrinsically disordered regions. Panel (b) describes an alternative in which instead of folded binding domains, the binding domains are simply linear motifs that engage in attractive interactions, embedded within a disordered protein scaffold. Panel (c) illustrates one possible manner in which the same proteins could partition into the core and periphery based on the phosphorylation state of the connecting linkers. Highly phosphorylated linkers have a larger effective solvation volume, driving them to the periphery.
organize droplets according to additional regulatory factors. Such a model is in agreement with recent results from Protter et al [63].

Spatial organization appears to be prevalent across a majority of multicomponent membraneless organelles [64]. The observation of spatially organized droplets as opposed to unitary liquids might be viewed as negating the case for non-stoichiometric assemblies and/or phase transitions as a way to regulate the formation of such droplets in living cells. In such a scenario, spatial organization must require an order of operations and hence arise as a consequence of energy-dependent processes that are out of equilibrium in living cells. The case for such non-equilibrium transitions has been buttressed by the observation that ATP-dependent machines are required to dissolve droplets such as stress granules [30]. It is noteworthy that energy dependent processes might provide a way to nucleate droplet formation if the system lies outside the spinodal regime [65]. Similarly, dissolution of droplets might require energy dependent to ensure that the dynamics of dissolution do not become a limiting consideration for cellular dynamics. Our results suggest that the spontaneous formation of spatial organization can come about through the sequence-intrinsic interactions encoded by linear multivalent proteins/polymer. This framework has been applied to two archetypal systems that demonstrate spatial organization in cells and upon reconstitution. In the case of the nucleolus, the distinct co-existing phases are large, well-defined assemblies that had been previously identified via conventional microscopy [28, 66]. Nuclear speckles and paraspeckles, which were studied using advanced super-resolution approaches coupled with statistical analysis, were found to form distinct regions [29, 32]. Whether or not spatial organization plays a key functional role is a separate and arguably more challenging question to answer. We anticipate that our contributions to the mechanisms of spatial organization will aid in designing experiments that directly test its biological importance, or lack thereof.

Acknowledgments

We thank members of the Pappu lab and our colleagues/collaborators including Clifford Brangwynne, Anthony Hyman, Michael Rosen, and Geraldine Seydoux for helpful discussions. We acknowledge financial support through funds from the St. Jude Research Collaborative on Membraneless Organelles (RVP), the Human Frontier Science Program RGP0034/2017 (RVP), the US National Science Foundation through grant MCB1614766 (RVP), Washington University’s Center for Biological Systems Engineering (TSH), and the Max Planck Society (TSH).

ORCID iDs

Tyler S Harmon https://orcid.org/0000-0002-2822-540X
Alex S Holehouse https://orcid.org/0000-0002-4155-5729
Rohit V Pappu https://orcid.org/0000-0003-2568-1378

References

[1] Zhou H-X, Rivas G and Minton A P 2008 Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences Annu. Rev. Biophys. 37 375–97
[2] Chen W W, Freinkman E, Wang T, Birsoy K and Sabatini D M 2016 Absolute quantification of matrix metabolites reveals the dynamics of mitochondrial metabolism Cell 166 1324–37 e11
[3] Milo R 2013 What is the total number of protein molecules per cell volume? A call to rethink some published values Bioessays 35 1050–5
[4] Mao Y S, Zhang B and Spector D L 2011 Biogenesis and function of nuclear bodies Trends Genet. 27 295–306
[5] Mintz P J and Spector D L 2000 Compartmentalization of RNA processing factors within nuclear speckles J. Struct. Biol. 129 241–51
[6] Muratani M, Gerlich D J, Janicki S M, Gebhard M, Eils R and Spector D L 2002 Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus Nat. Cell Biol. 4 106–10
[7] Janicki S M and Spector D L 2003 Nuclear choreography: interpretations from living cells Curr. Opin. Cell Biol. 15 149–57
[8] Lamond A I and Spector D L 2003 Nuclear speckles: a model for nuclear organelles Nat. Rev. Mol. Cell Biol. 4 605–12
[9] Gall J G, Bellini M, Wu Z and Murphy C 1999 Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes Mol. Biol. Cell 10 4385–402
[10] Shin Y and Brangwynne C P 2017 Liquid phase condensation in cell physiology and disease Science 357 eaaf4382
[11] Banani S F, Lee H O, Hyman A A and Rosen M K 2017 Biomolecular condensates: organizers of cellular biochemistry Nat. Rev. Mol. Cell Biol. 18 285–98
[12] Brangwynne C P et al 2009 Germline P granules are liquid droplets that localize by controlled dissolution/condensation Science 324 1729–32
[13] Brangwynne C P, Tompa P and Pappu R V 2015 Polymer physics of intracellular phase transitions Nat. Phys. 11 899–904
[14] Hyman A A, Weber C A and Jülicher F 2014 Liquid–liquid phase separation in biology Annu. Rev. Cell Dev. Biol. 30 39–58
[15] Li P et al 2012 Phase transitions in the assembly of multivalent signalling proteins Nature 483 536–40
[16] Harmon T S, Holehouse A S, Rosen M K and Pappu R V 2017 Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins elife 6 e30294
Lin Y, Currie S L and Rosen M K 2017 Intrinsically disordered sequences enable modulation of protein phase separation through

Burke K A, Janke A M, Rhine C L and Fawzi N L 2015 Residue-by-residue view of

Zhang H

Lyle N, Das R K and Pappu R V 2013 A quantitative measure for protein conformational heterogeneity J. Chem. Phys. 139 121907

Saitoh N, Spah C S, Patterson S D, Bubulya P, Neuwald A F and Spector D L 2004 Proteomic analysis of interchromatin granule clusters Mo. Biol. Cell 15 3876–90

Hubstenberger A et al 2017 P-body purification reveals the condensation of repressed mRNA regulons Mol. Cell 68 144–57.e5

Khong A, Matheny T, Jain S, Mitchell S F, Wheeler J R and Parker R 2017 The stress granule transcriptome reveals principles of mRNA accumulation in stress granules Mol. Cell 68 808–20.e5

Jain S, Wheeler Joshua R, Walters Robert W, Agrawal A, Barsic A and Parker R 2016 ATPase-modulated stress granules contain a diverse proteome and substructure Cell 164 487–98

Jacobs W M and Frenkel D 2017 Phase transitions in biological systems with many components Biophys. J. 112 683–91

Lin Y-H, Brady J P, Forman-Kay J D and Chan H S 2017 Charge pattern matching as a fundamental phase separations in intrinsically disordered proteins New J. Phys. 19 115003

Feric M et al 2016 Coexisting liquid phases underlie nuclear subcompartments Cell 165 1686–97

Fei J et al 2017 Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution J. Cell Sci. 130 4180–92

Jain S, Wheeler J R, Walters R W, Agrawal A, Barsic A and Parker R 2016 ATPase-modulated stress granules contain a diverse proteome and substructure Cell 164 487–98

Lang M et al 2010 Three-dimensional organization of promyelocytic leukemia nuclear bodies J. Cell Sci. 123 392–409

West A et al 2016 Structural super-resolution microscopy analysis of parasepucle nuclear body organization J Cell Biol. 214 817–30

Harmon T S, Holehouse A S and Pappu R V 2017 To mix, or to demix, that is the question J. Chem. Phys. 10 51–61

Flory P 1942 Thermodynamics of high polymer solutions J. Chem. Phys. 10 51–61

Flory P 1969 Statistical Mechanics of Chain Molecules (New York: Oxford University Press)

Madras N and Slade G 1996 The Self-Avoiding Walk ed C N Thomas Liggett and P. Lorenz (Boston, MA: Birkhauser)

Holehouse A S, Garai K, Lyle N, Vitalis A and Pappu R V 2015 Quantitative assessments of the distinct contributions of peptolyptide backbone amides versus side chain groups to chain expansion via chemical denaturation J. Am. Chem. Soc. 137 2984–95

Mao A H, Lyle N and Pappu R V 2013 Describing sequence–ensemble relationships for intrinsically disordered proteins Biochem J. 449 307–18

Tran H T and Pappu R V 2016 Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions Biophys. J. 91 1868–86

Smith J, Calidas D, Schmidt H, Lu T, Rasoloson D and Seydoux G 2016 Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3 eLife 5

Wallace E W J et al 2015 Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress Cell 162 1286–98

Decker C J and Parker R 2012 P-bodies and stress granules: possible roles in the control of translation and mRNA degradation Cold Spring Harbor Perspect. Biol. 4 a012286

Saha S et al 2016 Polar positioning of phase-separated liquid compartments in cells regulated by an mRNA competition mechanism Cell 166 1572–84.e16

Wei M T et al 2017 Phase behaviour of disordered proteins underlining low density and high permeability of liquid organelles Nat. Chem. 9 1118–25

Zhang H et al 2015 RNA controls PolyQ protein phase transitions Mol. Cell 60 220–30

Elbaum-Gar M et al 2016 Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity EMBO J. 36 2951–67

Burke K A, Janke A M, Rhine C J and Fawzi N L 2015 Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II Mol. Cell 60 231–41

Han T W et al 2012 Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies Cell 149 768–79

Kato M et al 2012 Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels Cell 149 753–67

Boeynaems S et al 2017 Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics Mol. Cell 65 1044–55.e5

Monahan Z et al 2017 Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity EMBO J. 36 2951–67

Burke K A, Janke A M, Rhine C J and Fawzi N L 2015 Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II Mol. Cell 60 231–41

Han T W et al 2012 Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies Cell 149 768–79

Kato M et al 2012 Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels Cell 149 753–67

Boeynaems S et al 2017 Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics Mol. Cell 65 1044–55.e5

Monahan Z et al 2017 Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity EMBO J. 36 2951–67

Burke K A, Janke A M, Rhine C J and Fawzi N L 2015 Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II Mol. Cell 60 231–41

Han T W et al 2012 Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies Cell 149 768–79

Kato M et al 2012 Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels Cell 149 753–67

Boeynaems S et al 2017 Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics Mol. Cell 65 1044–55.e5

Monahan Z et al 2017 Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity EMBO J. 36 2951–67

Burke K A, Janke A M, Rhine C J and Fawzi N L 2015 Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II Mol. Cell 60 231–41

Han T W et al 2012 Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies Cell 149 768–79
[62] Langdon E M et al 2017 mRNA structure determines specificity of a polyQ-driven phase separation bioRxiv https://www.biorxiv.org/content/early/2017/12/13/233817
[63] Protter D S W et al 2018 Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly Cell Rep. 22 1401–12
[64] Trcek T, Grosch M, York A, Shroff H, Lionnet T and Lehmann R 2015 Drosophila germ granules are structured and contain homotypic mRNA clusters Nat. Commun. 6 7962
[65] Brangwynne C P, Mitchison T J and Hyman A A 2011 Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes Proc. Natl. Acad. Sci. USA 108 4334–9
[66] Boisvert F-M, van Koningsbruggen S, Navascués J and Lamond A I 2007 The multifunctional nucleolus Nat. Rev. Mol. Cell Biol. 8 574–85