



Published in final edited form as:

Biochemistry. 2018 May 01; 57(17): 2478–2487. doi:10.1021/acs.biochem.8b00008.

The relationship of sequence and phase separation in protein low-complexity regions

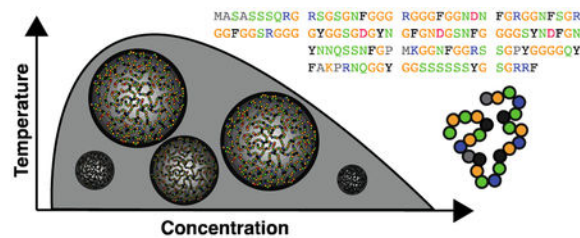
Erik W. Martin* and Tanja Mittag*

Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN

Abstract

Liquid-liquid phase separation seems to play critical roles in the compartmentalization of cells through the formation of biomolecular condensates. Many proteins with low-complexity regions are found in these condensates, and they can undergo phase separation *in vitro* in response to changes in temperature, pH and ion concentration. Low-complexity regions are thus likely important players in mediating compartmentalization in response to stress. However, how the phase behavior is encoded in their amino acid composition and patterning is only poorly understood. We discuss here that polymer physics provides a powerful framework for our understanding of the thermodynamics of mixing and demixing and for how the phase behavior is encoded in the primary sequence. We propose to classify low-complexity regions further into sub-categories based on their sequence properties and phase behavior. Ongoing research promises to improve our ability to link the primary sequence of low-complexity regions to their phase behavior as well as the emerging miscibility and material properties of the resulting biomolecular condensates, providing mechanistic insight into this fundamental biological process across length scales.

Graphical Abstract



Introduction

The spatial organization of proteins and nucleic acids in cells is critical to maintaining the biochemical reactions essential for life. Evidence is accumulating that the condensation of biomolecules into liquid-like, non-membrane-bound cellular bodies is mediated by liquid-liquid phase separation, in which solution components spontaneously demix forming dense and light phases.¹⁻¹⁴ Phase separation is typically sensitive to solution conditions such as ionic strength, pH, redox potential, osmotic pressure and temperature, many of which are

*To whom correspondence should be addressed: erik.martin@stjude.org and tanja.mittag@stjude.org.

typical biological stress factors. Phase separation thus provides an attractive explanation for the change in cellular compartmentalization and the formation of biomolecular condensates in response to stress, although some biomolecular condensates are permanent structures formed under non-stress conditions.

In vitro, phase separation can be observed macroscopically as the spontaneous formation of droplets from a miscible solution in response to a change in solution conditions (Fig. 1A), resulting in the coexistence of dense and light liquid phases. Many of the biomolecules thus far identified as essential components to biomolecular condensates can conveniently be categorized as ‘bio-polymers’, such as intrinsically disordered proteins, DNA or RNA, making analysis of these phase separation phenomena amenable to the analytic framework of polymer physics. Theory predicts that a polymer solution can demix in response to either an increase or decrease in temperature, resulting in phase transitions with lower critical solution temperature or upper critical solution temperature, respectively (Fig. 1B). Indeed, depending on primary amino acid sequence, bio-polymers can separate into distinct liquid phases at both high and low temperatures.

Notably, biomolecular condensates are enriched in proteins containing so-called ‘low-complexity regions’ (LCRs)^{15, 16}, suggesting that these proteins may play critical roles in their formation, may tune the material properties of the formed structures, or both. Within the limited subset of LCRs whose phase transitions have so far been studied, a wide array of properties has been observed, prompting exploration of how the primary amino acid sequence encodes the phase behavior. Clearly, the categorization as a ‘low-complexity’ region is woefully inadequate as a descriptor of proteins capable of demixing. In this perspective, we will focus on hydrophobicity, aromatic character and charge in low-complexity sequences, and how the modulation of these sequence features can change phase separation properties. The topic has been reviewed in the past^{17, 18}, but new primary data allows us to propose a general framework for phase separation of LCRs.

Low-complexity regions in the proteome

Protein segments containing significant enrichment in specific amino acid types or sequence repeats, i.e. LCRs in the broadest possible sense, occur frequently in the eukaryotic proteome. Soon after researchers began scrutinizing sequenced genomes, the existence of such LCRs became obvious.¹⁹ Whether these sequences were functional or purely an evolutionary artifact resulting from gene duplication or expansion became an active question.^{20, 21} As one might guess from the diversity in function, LCRs come in many ‘flavors’, i.e. in different sequence compositions and with different biophysical properties. Any attempt to define what a LCR is must begin with a statistical treatment of amino acid composition. A particularly sequence can be defined as low complexity if the composition within a given window is statistically biased to a small fraction of amino acid types relative to a random distribution.²² Using the probabilistic definition of low-complexity as implemented in the SEG algorithm^{22, 23}, the sum of all LCRs in the human proteome have roughly the same amino acid frequencies as the bulk proteome (Fig. 2A,B) highlighting the fact that LCRs do not have one specific sequence flavor, but encompass a variety of sequence types - an observation also made by Wootton and Federhen.²² Of particular note is

the overall enrichment of the amino acid leucine, which is frequently found in leucine-rich repeat proteins, which fold into a defined arc-like shape (Fig. 2C). While intrinsically disordered regions (IDR) of proteins, i.e. regions that do not adopt a unique folded conformation, often have a limited sequence complexity, LCRs are not necessarily disordered, and structured proteins are frequently enriched in a small fraction of amino acid types.²⁴ Furthermore, intrinsically disordered LCRs vary substantially with respect to the types of enriched amino acids (Fig. 2D). Careful consideration of the meaning of low complexity draws attention to the fact that the term LCR on its own possesses only statistical meaning and hence does not, a priori, speak to any specific function or property of a protein sequence. Therefore, any discussion of protein sequences that demix in solution necessarily must seek more detailed description of the sequence space.

LCRs and phase separation

Significant progress has been made in our understanding of LCR function, and LCRs have been implicated in transcription, stress response, cellular localization, and myriad other processes.¹⁵ Curiously, it is now being suggested that protein and nucleic acid phase separation contributes to many of these processes.²⁶⁻²⁸ The subset of LCR-containing proteins that have been observed to phase separate can be demarcated by the specific types of enriched amino acids and a general qualitative framework can be set into place to correlate LCR types and their phase behavior. Intrinsically disordered LCRs that are rich in charged amino acids can demix into complex coacervates with oppositely charged proteins or nucleic acids.²⁹⁻³² Polar LCRs containing a high fraction of serine, glutamine, asparagine and glycine (where glycine is counted with polar residues because its properties are dominated by its polar backbone in the context of a protein), have been observed to phase separate homotypically *in vitro* when the sequence is punctuated by aromatic and charged amino acids.^{5, 6, 8-11, 25, 28, 33, 34} The RNA binding protein hnRNPA1, which partitions into stress granules under stress conditions, is a typical example (Fig. 2D). Both the sequence composition and patterning play important roles in encoding phase behavior. The influence of patterning, i.e. how specific amino acid types are distributed in the primary sequence, has so far only been studied in detail for the phase behavior of ELPs. However, its influence on the global dimensions and function of IDPs is well appreciated,³⁵⁻³⁷ and we thus expect similar influence from patterning on the phase behavior of LCRs. Slight variations to the aromatic, aliphatic, charge content and sequence patterning significantly influences the phase coexistence temperature and concentration or can alter the material properties of the protein dense phase.^{5, 18, 32, 38, 39} As an extreme example, elastin-like peptides (ELP), which contain a high fraction of hydrophobic amino acids, show an inverse temperature dependence compared to polar LCRs, i.e. they demix at increasing temperature.⁴⁰ The absence of any unifying feature beyond statistical low-complexity highlights the need for caution in using the term 'low-complexity regions'. Furthermore, in the limited subset of LCRs observed to phase separate, the type of enriched amino acids can encode many types of phase behavior, leading to further complexity. For now, we will use the terms polar, hydrophobic or charged LCR to further categorize LCR sequences that are able to demix – sequence features that in the broader context of protein phase transitions have been observed to confer different properties.^{17, 18, 39}

While many disordered LCRs have been demonstrated to undergo liquid demixing and in cells are frequently localized to biomolecular condensates, the conditions under which they demix, and the material properties of the dense phases they form, vary widely and are encoded in their primary sequence. These material properties encompass the density of the droplets, viscoelasticity, and the surface tension. These properties determine whether two dense phases are miscible with each other or form separate droplets, what the concentration within a biomolecular condensate is, and where it is on the spectrum of liquid to solid. How the primary sequence of LCRs encodes viscoelastic properties has been reviewed previously.¹⁸ Here, we will concentrate on the coexistence concentration; how high is the propensity of a particular LCRs to demix (i.e. determining the low-concentration arm of the coexistence line), how is this propensity modulated by conditions such as temperature and ionic strength, and what is the concentration of the resulting droplets (i.e. determining the high-concentration arm of the coexistence line)? Many of the differences in the solution behavior of different flavors of LCRs are apparent, at least at a qualitative level, from considerations of the basic thermodynamics of mixing.

Thermodynamics of liquid-liquid phase separation

Phase separation *in vivo* is a multi-component process which inherently occurs away from equilibrium, complicating its theoretical treatment. However, *in vitro*, the equilibrium phase transitions of isolated components can be measured precisely and analyzed through the combined expertise of a century of polymer chemistry and physics.⁴¹ A phase transition is driven by the minimization of the global free energy of the system. The relevant free energy term by which the solution properties of a given polymer are described, is the free energy of mixing, ΔG_m , which is given by the familiar equation:

$$\Delta G_m = \Delta H_m - T\Delta S_m, \quad (1)$$

where ΔH_m and ΔS_m are the enthalpy and entropy of mixing, respectively. Minimizing the global free energy involves contributions from enthalpy, containing the interaction potentials between protein and solvent, and the entropy, which encompasses the available degrees of freedom of protein and solvent molecules.

The combinatoric entropy of mixing of an ideal polymer solution was described by Flory and Huggins^{42, 43} in terms of volume fractions to account for the size difference of solvent and polymer. This mixing entropy only accounts for the entropic cost associated with confining a polymer in a dense phase.

$$\Delta S_m = -kV \left[\frac{\phi_s}{v_s} \ln \phi_s + \frac{\phi_p}{v_p} \ln \phi_p \right] \quad (2)$$

The volume fractions of solvent and protein are the fraction of the total system volume (V) occupied by solvent or protein ($\phi_{s,p}$). The terms $v_{s,p}$ correct for the volume of a single

molecule of the solvent and protein. In general, the solvent parameter v_s is taken as 1, while the volume of an individual protein molecule is a function of the sequence length (Fig. 3A). Increasing the length of the protein effectively decreases the entropic cost of confining the protein into a dense phase and lowers the concentration at which a protein phase separates.

In the simplest case, the enthalpic contribution to the free energy is modeled as a mean field consisting of pairwise self- and cross-interactions involving solvent and protein. This is written as a function of the volume fractions of solvent and protein, ϕ_s and ϕ_p , and an interaction parameter, χ .

$$\Delta H_m = \phi_s \phi_p \chi \frac{kTV}{v_r} \quad \chi = \frac{z}{kT} \left[\frac{1}{2}(\epsilon_{ss} + \epsilon_{pp}) - \epsilon_{ps} \right] \quad (3)$$

The χ term is essentially a fitting parameter that is related to the balance of interaction energies between solvent molecules (ϵ_{ss}), protein molecules (ϵ_{pp}), and protein molecules with solvent molecules (ϵ_{ps}). The additional terms v_r and z are related to the lattice, where v_r is the volume of an individual monomer (often taken as the square root of the product of solvent and polymer monomer volumes) and z is the number of potential interactions on the lattice. Combining equations 2 and 3 results in the Flory-Huggins expression:

$$\Delta G_m = \phi_s \phi_p \chi \frac{kTV}{v_r} + kTV \left[\frac{\phi_s}{v_s} \ln \phi_s + \frac{\phi_p}{v_p} \ln \phi_p \right] \quad (4)$$

While this description neglects more nuanced interactions and completely ignores the sequence-dependent interactions of heteropolymeric proteins – even that of low-complexity proteins – the framework has proven itself useful as a tool for considering the balance of energies contributing to phase separation.

The clear implication of Flory-Huggins Theory is that the entropy of mixing is always positive; however, as protein-protein interactions become more favorable or solvent protein interactions become less favorable, multiple energy minima exist on the volume fraction coordinate (Fig. 3B), which define the composition of protein dense and light phases. The conditions under which the solution is unstable and can spontaneously demix are defined by the inflection points on the free energy curve:

$$\frac{\partial^2 \Delta G_m}{\partial \phi^2} = 0 \quad (5)$$

Classic Flory-Huggins theory has been used to determine the critical solution temperature of the germ granule protein DDX4.⁵ The addition of a more complex description of the enthalpy component is allowing a more quantitative description of the sequence-specific interactions driving phase separation. The Overbeek and Voorn extension of the Flory-

Huggins free energy takes into account electrostatic interactions that can mediate complex coacervation.⁴⁴ Recently, the random phase approximation has been effectively used to model the patterning of charged residues.⁴⁵ Furthermore, adding a three body correction to the enthalpy was used to successfully explain the phase behavior of the germ granule protein Laf1, which forms semi-dilute droplets, i.e. a dense phase of surprisingly low concentration.⁴⁶ The heteropolymeric nature of proteins complicates the systematic analysis of the enthalpic contributions to demixing, but the ongoing development of analytic models to describe experimental data holds great promise towards creating a general framework that relates the sequence of LCRs to their phase behavior.

Regardless of how detailed the enthalpy term, basic Flory-Huggins Theory neglects more complicated entropic effects. The combinatoric entropy inherently favors mixing, and the entropy gain from mixing is more favorable at higher temperatures. However, if the content of hydrophobic residues and their patterning in a sequence favors the compaction of individual chains, the opposite is true; the condensation of many protein molecules of this type into a dense phase is also driven by an increase in entropy due to the release of water molecules from the hydration shell of hydrophobic sidechains. Given that the cost of confining solvent becomes higher with increasing temperature, hydrophobic compaction of individual protein molecules has an inverse temperature dependence, i.e. it becomes more favorable as the temperature increases. Entropically driven release of solvent is the origin of phase transitions with lower critical solution temperatures (LCST) (Fig. 1). LCST phase transitions have been well characterized experimentally using synthetic polymers; the theory to describe these transitions typically involves increasing the complexity of the lattice and allowing compressibility in the system.⁴⁷

The driving forces for phase separation of biopolymers can be generalized thus: a phase transition upon increasing the temperature, also called a LCST transition, is entropic in nature and is driven by release of water molecules from the hydration shell and the accompanying gain of water entropy; a UCST transition upon decreasing the temperature is driven by molecular interactions that increase in strength relative to the entropic contributions at lower temperature. In apparent conflict with these driving forces, aromatic and hydrophobic amino acids have been observed to be required for phase separation in proteins with LCST as well as UCST phase transitions. In the next sections, we will review the evidence and propose a reconciliation of the observations with the underlying thermodynamics of mixing/demixing.

Elastin-like polypeptides and LCST transitions

The majority of phase separating systems in synthetic polymer chemistry are block-type polymers and exhibit LCST phase behavior. The closest pseudo-biological analog of these synthetic polymers is the elastin-like polypeptide (ELP) with the typical hydrophobic repeat sequence (Val-Pro-Gly-Xaa-Gly)_n from tropoelastin, the soluble protein that forms elastin, the main elastic material in mammals. Xaa represents a 'guest' amino acid⁴⁸, and greater hydrophobicity of Xaa results in a lower transition temperature, i.e. a stronger propensity for demixing. In fact, a phenomenological hydrophobicity scale has been derived based on the effect of the guest residue on the demixing temperature.⁴⁹

Increasing the propensity of ELPs for phase separation by increasing their hydrophobicity has an interesting corollary; recent work with block-repeat peptides has demonstrated that, inspired by resilin, an insect protein that forms elastic materials in insect wings, titrating down hydrophobicity and adding arginine residues can generate synthetic sequences that transition with a UCST instead of a LCST.³⁹

The physical basis for the phase separation of ELPs on the molecular level is likely influenced also by enthalpic effects, i.e. collapsed conformations and protein dense phases are stabilized by hydrogen bonding and other interactions.⁵⁰ Further, by mixing ELPs with different properties, multiphasic systems can be generated that form droplets within droplets.⁵¹ Since multiphasic droplets have been observed also in vivo^{12, 52, 53} and seem to be the basis for the formation of biomolecular condensates with internal structure, understanding their molecular basis is of particular interest.

These complexities aside, the fact that the transition temperatures of ELPs follow predictable trends based on the hydrophobicity of guest amino acids and chain length are informative for understanding the impact of the hydrophobic effect in the context of chain collapse transitions and phase separation. Generically, these effects are described by quantifying the loss of configurational entropy of water due to ordering in solvation shells surrounding the hydrophobic solute. Regardless of whether chain collapse or collective condensation into a dense phase occurs with increasing or decreasing temperature, i.e. whether the sequence exhibits a UCST or LCST transition, the entropy increases concomitantly due to the release of solvent. A point which emerges from the pioneering work of Dan Urry is that whether or not a LCST phase transition is observed within the window of physiologically accessible temperatures, the driving forces underlying such transitions are critical to the conformational dynamics and stability of all systems.⁴⁸

These principles are on display in studies of poly-A binding protein (Pab1), a yeast heat shock granule-associated protein with a particularly interesting LCST behavior.²⁸ The folded RNA binding domains of Pab1 mediate demixing at physiologically accessible temperatures, but the demixing temperature is tuned by the hydrophobicity of the disordered P-domain in a manner predicted by the hydrophobicity scale from work on ELPs. The Pab1 phase behavior illustrates how the driving forces of hydrophobic assembly can be at play behind the scenes regardless of the molecular interactions that dominate the global phase behavior.

Polar low-complexity regions and UCST transitions

In contrast to most synthetic polymers, many of the disordered protein segments that have recently been reported to phase separate under near-physiological conditions have a UCST phase behavior. However, it is noteworthy that in most cases, the phase space has been insufficiently sampled to rule out the possibility of a coexisting LCST. These systems can loosely be categorized as polar LCRs wherein the majority of amino acids are serine, glycine, asparagine and glutamine (Fig. 4). Direct experimental evidence is sparse, but it is typically assumed that purely polar sequences have the potential to collapse in solution and aggregate.⁵⁴⁻⁵⁶ Despite contributing a significant fraction of residues in LCRs, there is no available information on the conformational properties of poly-serine. It is possible that

serine increases solubility or collapse in a context dependent manner. The remaining sequence space of polar LCRs likely serves to modulate solution properties of the protein by increasing excluded volume, introducing electrostatic intrachain repulsion and decreasing the solvation free energy and thus potentially making condensation either more or less favorable and dependent on solution conditions. A number of amino acid types can fill the remaining sequence space of polar LCRs, but the frequent punctuation of sequences with a low fraction of regularly spaced charged or aromatic residues stands out due to its ubiquity.

In several cases, experimental evidence demonstrates the importance of hydrophobic amino acids in UCST phase transitions in apparent contradiction to the classic notion of hydrophobic assembly as an entropically driven LCST transition. In the LCR of fused in sarcoma (FUS), tyrosine stood out conspicuously from the background of small polar residues and it has thus been the subject of the few currently available mutational studies of the relationship of the primary sequence to phase behavior in LCRs. The LCR of FUS is incapable of forming hydrogels when even a fraction of tyrosine residues are removed.¹⁵ While hydrogel formation and liquid-liquid phase separation are not necessarily mediated by the same interactions, the role of tyrosine residues for phase separation was confirmed when it was shown that replacing them with phenylalanines reduces the driving force for demixing. Interestingly, aromatic and aliphatic hydrophobic residues had opposing effects on phase separation. The experiments made use of a multivalent SH3 domain module that was able to undergo phase separation with a multivalent binding partner; different FUS LCR variants were fused to this module and their ability to potentiate or attenuate the phase separation propensity of the SH3 module was tested. Tyrosines promoted phase separation more than phenylalanines, and leucines attenuated phase separation.³⁸ These results show that the aromatic character of tyrosine and phenylalanine, not simply their hydrophobic character, is critical for the UCST transition of FUS.³⁸

In the germ granule protein DDX4 (Fig. 3C), the removal of phenylalanine residues eliminates phase separation.^{5, 57} When the phenylalanine residues were fluorinated to withdraw electrons from the π system, phase separation was also hindered, suggesting that there is a specific enthalpic contribution from $\pi - \pi$, cation - π or, likely, both interactions. Indeed, in NMR measurements on DDX4 in the dense phase, contacts between aromatic residues, and between aromatic residues and arginines were observed.⁵⁷

Many proteins studied for their phase behavior so far, such as the RNA-binding proteins hnRNPA1, hnRNPA2, TDP-43 and EWS, have a similar, polar LCR sequence architecture (Fig. 4A,B), and it is likely that aromatic residues are essential for their UCST phase behavior.^{9, 11, 58-60} This observation is completely in line with the previously mentioned work on resilin-like repeat peptides. Often when a UCST phase transition was observed after the addition of arginine, an aromatic amino acid was also present in the repeat peptide,³⁹ potentially arguing for the existence of enthalpic interactions involving their conjugated π systems.

Polar low-complexity regions are present in many organisms. Plant glycine-rich proteins are particularly intriguing targets for studying condensation of proteins and nucleic acids. Proteins such as atGRP7 are remarkably similar to polar LCR proteins like FUS and

hnRNPA1 (Fig. 3E), including the punctuation with aromatic residues. Plant glycine-rich proteins typically bind RNA, have been implicated in circadian response, flowering and a wide array of stress responses including cold stress.⁶¹ It is rational to hypothesize that RNA-binding proteins such as atGRP7 could form biomolecular condensates in response to stress, and that modulation of this behavior could have wide reaching impact on crop yield and adaptation to a changing environment.

Modulation of UCST phase behavior and material properties

Increasing the content of charged amino acids in a polar LCR while maintaining the polar background and aromatic punctuation modulates the solvation properties of the LCR. In DDX4, the importance of aromatic residues has been demonstrated and experimental evidence points to their role in mediating protein-protein interactions, not only confining water molecules in a hydration shell. Additionally, in the DDX4 system, interactions between patches of charged residues appear to work in tandem with aromatic interactions. Native DDX4 has multiple patches of positive and negative charge along the sequence, which are necessary for its phase behavior and modulate the DDX4/solvent interactions.^{5, 57} In fact, modeling the enthalpic component of ΔG_m via a mean field electrostatic component, which can account for how evenly distributed positive and negative charges are interacting within a neutral polymer in a sequence-dependent manner, predicts a lower critical concentration when like-charged residues are grouped into patches.^{45, 62} The oppositely charged patches therefore contribute enthalpically via electrostatic interactions to the UCST phase transition of DDX4.

In apparent contrast, the *C. elegans* DDX3X-type protein Laf-1 has similar charge content, but with positively and negatively charged residues well mixed in the sequence, and shows dramatically different solvation properties. The dense phase of Laf-1 is up to two orders of magnitude more dilute than that of DDX4.^{16, 46, 57} Atomistic simulations revealed that charge-mediated expansion of the sequence was a likely cause because it lead to the expansion of the individual chain, presumably without interfering with its stickiness. It is important to note that the role of aromatic residues in the phase behavior of Laf-1 has not been experimentally demonstrated. However, it is reasonable to speculate that the interplay between content and patterning of charged and aromatic residues tunes the saturation concentration for phase separation and dramatically impacts the material properties, i.e. the density, of the resulting dense phase. Considering that DDX4 has several unique features – enrichment of asparagine versus glutamine and the presence of a small fraction of ELP-like hydrophobic residues such as valine – it is likely that multiple effects contribute to the different behaviors.

A case in which the influence of charged residues on the phase behavior of a polar LCR is even more pronounced is the nephrin intracellular domain (NICD), which undergoes complex coacervation with positively charged counterions.³² Nephrin is a membrane protein necessary for proper filtration by the kidney. NICD contains substantial patches of negatively charged residues along its sequence. When NICD is alone in solution, charge – charge repulsion prevents assembly. In the presence of counterions, e.g. in the form of positive supercharged GFPs, the complex phase separates. When the amino acid dependence

of demixing was carefully tested, tyrosine was strongly correlated with demixing, suggesting that after charge neutralization, NICD phase separates via similar interactions as polar LCRs do. The charge neutralization paradigm might be pervasive in many LCRs where ionic strength dramatically increases the UCST demixing temperature such as in the condensation of mussel adhesive proteins in seawater.⁶³ This mechanism is likely at play in the phase behavior of many polar LCRs.¹⁰ The content and patterning of charged residues in polar LCRs influences the coexistence temperature, concentration, the material properties and response to ionic strength. Punctuation with other residue types also has the potential to modulate these properties and future studies are needed to address this.

Structural basis for phase separation of LCRs

From our discussion, we have seen that many LCRs likely have the ability to demix from solution. A major outstanding question is what, precisely, mediates the favorable enthalpic interactions, i.e. what is the structural basis of the phase separation of LCRs? In synthetic polymers, repetitive polar interaction motifs with the ability to form strong hydrogen bonds can mediate UCST phase transitions.⁶⁴ While it is convenient to imagine that similar interactions also mediate UCST transitions of LCRs, evidence clearly shows that a small subset of aromatic amino acids are required for phase separation. These residues may participate in multibody interactions. If protein-protein contacts result in structuring of motifs of several amino acid length, the interaction energy has the potential to be much greater than that of individual $\pi - \pi$, cation - π , charge-charge or polar interactions.⁶⁵ There is substantial evidence to support the hypothesis that FUS can form extended cross-beta structures within solid assemblies.⁶⁶ Recent crystal structures of hexa-peptides with (G/S)-Y-(G/S)-type sequences from LCRs show a non-ideal beta sheet structure.⁶⁷ It was proposed that this imperfection would allow these motifs to be dynamic enough to be compatible with the liquid-like character of biomolecular condensates.⁶⁷ The degree of disorder in LCR dense phases remains an open question. While entropically-mediated LCST transitions do not conceptually require structuring, dense phases of some designed ELPs are stabilized by structure. Some polar LCRs undergo UCST transitions without the appearance of stable structure,^{10, 57} suggesting the possibility that if structuring occurs it is local and transient.

LCRs – UCST or LCST transition?

The potential need for multiple weak interaction motifs in phase separating proteins can be satisfied by imperfect sequence repeats, distributed along a largely disordered sequence with solvation properties that are responsive to the solution conditions. This may be an explanation for the fact that low-complexity sequences are enriched in biomolecular condensates, and that Nature may have settled on them as one archetype of proteins mediating phase separation. LCRs with ELP-like or polar character fall under this umbrella. Whether these LCRs phase separate at physiologically relevant conditions depends on their sequence composition and patterning. LCST transitions require hydrophobicity, which has to be balanced with solubility to avoid collapse and aggregation. Sequences that undergo UCST transitions, in contrast, must be largely devoid of hydrophobicity to prevent a concomitant LCST transition. A large fraction of charged residues would result in high solubility and might prevent phase transitions. These sequence constraints are likely what

results in the largely polar backbones of UCST-type LCRs. However, the published data show the importance of aromatic residues for driving UCST phase behavior^{5, 32, 66, 68}; enthalpic interactions of their π electron systems may allow for a temperature dependence of their phase behavior in the physiological range.

Outlook

Liquid – liquid phase separation has emerged from being a biophysical curiosity restrained to extreme conditions, e.g. frequently observed in attempts to crystallize proteins, to biological relevance. It is far from certain whether phase separation is the mechanism by which biomolecular condensates form, but evidence clearly supports it for some condensates such as nucleoli.^{7, 13} However, a detailed understanding of the sequence heuristics and physical chemistry of molecular interactions driving phase separation *in vitro* will also inform on the specific interactions of LCRs in membraneless organelles in cells. The growing experimental evidence on LCRs suggests that a complicated balance of free energy contributions determine the solvation properties of an individual sequence. The relative strength of intra-protein, inter-protein, protein-solvent and solvent-solvent interactions along with entropic contributions poise an LCR at a defined position along the continuum from extended to collapsed conformations. These conformational and thermodynamic properties can be correlated with demixing conditions and finally the material properties of dense protein states. The ability to fine tune the sequence of LCRs has potentially been an evolutionary tool to generate a mosaic of biomolecular condensates with varied material properties as a function of environmental conditions optimized for specific cellular functions.

Acknowledgements

This work was funded by R01GM112846, the American Lebanese Syrian Associated Charities and St. Jude Children's Research Hospital (to T.M.). We are grateful to Julie Forman-Kay, Alexander Holehouse, Rohit Pappu, Tyler Harmon, Thomas Boothby, Ivan Peran, Josh Riback, Ben Schuler and many other colleagues for many discussions and insights. TM acknowledges the 2016 and 2017 participants of the Bellairs workshop on the Physical Basis for Cellular Memory and Adaptation for helpful discussions.

References

- [1]. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoeghe C, Gharakhani J, Julicher F, and Hyman AA (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation, *Science* 324, 1729–1732. [PubMed: 19460965]
- [2]. Hyman AA, and Brangwynne CP (2011) Beyond stereospecificity: liquids and mesoscale organization of cytoplasm, *Dev Cell* 21, 14–16. [PubMed: 21763600]
- [3]. Brangwynne CP, Mitchison TJ, and Hyman AA (2011) Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes, *Proc Natl Acad Sci U S A* 108, 4334–4339. [PubMed: 21368180]
- [4]. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, Russo PS, Jiang QX, Nixon BT, and Rosen MK (2012) Phase transitions in the assembly of multivalent signalling proteins, *Nature* 483, 336–340. [PubMed: 22398450]
- [5]. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowitz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, and Baldwin AJ (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles, *Mol Cell* 57, 936–947. [PubMed: 25747659]

- [6]. Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC, Eckmann CR, Myong S, and Brangwynne CP (2015) The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics, *Proc Natl Acad Sci U S A* 112, 7189–7194. [PubMed: 26015579]
- [7]. Berry J, Weber SC, Vaidya N, Haataja M, and Brangwynne CP (2015) RNA transcription modulates phase transition-driven nuclear body assembly, *Proc Natl Acad Sci U S A* 112, E5237–5245. [PubMed: 26351690]
- [8]. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoynov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, and Alberti S (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation, *Cell* 162, 1066–1077. [PubMed: 26317470]
- [9]. Mollieux A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, and Taylor JP (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization, *Cell* 163, 123–133. [PubMed: 26406374]
- [10]. Burke KA, Janke AM, Rhine CL, and Fawzi NL (2015) Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II, *Mol Cell* 60, 231–241. [PubMed: 26455390]
- [11]. Lin Y, Protter DS, Rosen MK, and Parker R (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins, *Mol Cell* 60, 208–219. [PubMed: 26412307]
- [12]. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, and Brangwynne CP (2016) Coexisting Liquid Phases Underlie Nucleolar Subcompartments, *Cell* 165, 1686–1697. [PubMed: 27212236]
- [13]. Weber SC, and Brangwynne CP (2015) Inverse size scaling of the nucleolus by a concentration-dependent phase transition, *Curr Biol* 25, 641–646. [PubMed: 25702583]
- [14]. Shin Y, Berry J, Pannucci N, Haataja MP, Toettcher JE, and Brangwynne CP (2017) Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets, *Cell* 168, 159–171 e114. [PubMed: 28041848]
- [15]. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, and McKnight SL (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels, *Cell* 149, 753–767. [PubMed: 22579281]
- [16]. Hennig S, Kong G, Mannen T, Sadowska A, Kobelke S, Blythe A, Knott GJ, Iyer KS, Ho D, Newcombe EA, Hosoki K, Goshima N, Kawaguchi T, Hatters D, Trinkle-Mulcahy L, Hirose T, Bond CS, and Fox AH (2015) Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles, *J Cell Biol* 210, 529–539. [PubMed: 26283796]
- [17]. Brangwynne Clifford P., Tompa P, and Pappu Rohit V. (2015) Polymer physics of intracellular phase transitions, *Nature Physics* 11, 899.
- [18]. Weber SC (2017) Sequence-encoded material properties dictate the structure and function of nuclear bodies, *Current Opinion in Cell Biology* 46, 62–71. [PubMed: 28343140]
- [19]. Golding GB (1999) Simple sequence is abundant in eukaryotic proteins, *Protein Sci* 8, 1358–1361. [PubMed: 10386886]
- [20]. Alba MM, Tompa P, and Veitia RA (2007) Amino acid repeats and the structure and evolution of proteins, *Genome Dyn* 3, 119–130. [PubMed: 18753788]
- [21]. Toll-Riera M, Rado-Trilla N, Martys F, and Alba MM (2012) Role of low-complexity sequences in the formation of novel protein coding sequences, *Mol Biol Evol* 29, 883–886. [PubMed: 22045997]
- [22]. Wootton JC, and Federhen S (1993) Statistics of local complexity in amino acid sequences and sequence databases, *Computers & Chemistry* 17, 149–163.
- [23]. Wootton JC, and Federhen S (1996) Analysis of compositionally biased regions in sequence databases, *Methods Enzymol* 266, 554–571. [PubMed: 8743706]
- [24]. Kumari B, Kumar R, and Kumar M (2015) Low complexity and disordered regions of proteins have different structural and amino acid preferences, *Mol Biosyst* 11, 585–594. [PubMed: 25468592]

- [25]. Mitrea DM, Cika JA, Guy CS, Ban D, Banerjee PR, Stanley CB, Nourse A, Deniz AA, and Kriwacki RW (2016) Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA, *Elife* 5.
- [26]. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, and Sharp PA A Phase Separation Model for Transcriptional Control, *Cell* 169, 13–23. [PubMed: 28340338]
- [27]. Franzmann TM, Janel M, Pozniakovsky A, Mahamid J, Holehouse AS, Nuske E, Richter D, Baumeister W, Grill SW, Pappu RV, Hyman AA, and Alberti S (2018) Phase separation of a yeast prion protein promotes cellular fitness, *Science* 359.
- [28]. Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, Sosnick TR, and Drummond DA (2017) Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response, *Cell* 168, 1028–1040 e1019. [PubMed: 28283059]
- [29]. Banerjee PR, Milin AN, Moosa MM, Onuchic PL, and Deniz AA (2017) Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets, *Angew Chem Int Ed Engl* 56, 11354–11359. [PubMed: 28556382]
- [30]. Mitrea DM, Grace CR, Buljan M, Yun MK, Pytel NJ, Satumba J, Nourse A, Park CG, Madan Babu M, White SW, and Kriwacki RW (2014) Structural polymorphism in the N-terminal oligomerization domain of NPM1, *Proc Natl Acad Sci U S A* 111, 4466–4471. [PubMed: 24616519]
- [31]. Aumiller WM Jr., and Keating CD (2016) Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles, *Nat Chem* 8, 129–137. [PubMed: 26791895]
- [32]. Pak CW, Kosno M, Holehouse AS, Padrick SB, Mittal A, Ali R, Yunus AA, Liu DR, Pappu RV, and Rosen MK (2016) Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein, *Mol Cell* 63, 72–85. [PubMed: 27392146]
- [33]. Murakami T, Qamar S, Lin JQ, Schierle GS, Rees E, Miyashita A, Costa AR, Dodd RB, Chan FT, Michel CH, Kronenberg-Versteeg D, Li Y, Yang SP, Wakutani Y, Meadows W, Ferry RR, Dong L, Tartaglia GG, Favrin G, Lin WL, Dickson DW, Zhen M, Ron D, Schmitt-Ulms G, Fraser PE, Shneider NA, Holt C, Vendruscolo M, Kaminski CF, and St George-Hyslop P (2015) ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function, *Neuron* 88, 678–690. [PubMed: 26526393]
- [34]. 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.
- [35]. Das RK, Ruff KM, and Pappu RV (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins, *Curr Opin Struct Biol* 32, 102–112. [PubMed: 25863585]
- [36]. Martin EW, Holehouse AS, Grace CR, Hughes A, Pappu RV, and Mittag T (2016) Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation, *J Am Chem Soc* 138, 15323–15335. [PubMed: 27807972]
- [37]. Das RK, and Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proc Natl Acad Sci U S A* 110, 13392–13397. [PubMed: 23901099]
- [38]. Lin Y, Currie SL, and Rosen MK (2017) Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs, *J Biol Chem*.
- [39]. Quiroz FG, and Chilkoti A (2015) Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers, *Nat Mater* 14, 1164–1171. [PubMed: 26390327]
- [40]. Urry DW, Trapane TL, and Prasad KU (1985) Phase-structure transitions of the elastin polypentapeptide–water system within the framework of composition–temperature studies, *Biopolymers* 24, 2345–2356. [PubMed: 4092092]
- [41]. Flory PJ (1953) *Principles of polymer chemistry*, Cornell University Press, Ithaca,.
- [42]. Flory P (1942) *Thermodynamics of High Polymer Solutions*, *The Journal of Chemical Physics* 10, 51–61.

- [43]. Huggins M (1942) Some properties of solutions of long-chain compounds., *The Journal of Physical Chemistry* 46, 151–158.
- [44]. Overbeek JTG, and Voorn MJ (1957) Phase separation in polyelectrolyte solutions. Theory of complex coacervation, *Journal of Cellular and Comparative Physiology* 49, 7–26.
- [45]. Lin YH, Forman-Kay JD, and Chan HS (2016) Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles, *Phys Rev Lett* 117, 178101. [PubMed: 27824447]
- [46]. Wei MT, Elbaum-Garfinkle S, Holehouse AS, Chen CC, Feric M, Arnold CB, Priestley RD, Pappu RV, and Brangwynne CP (2017) Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles, *Nat Chem* 9, 1118–1125. [PubMed: 29064502]
- [47]. Fenkel D (1992) Order through disorder: Entropy-driven phase transitions, In *Complex Fluids: Proceedings of the XII Sites Conference, Barcelona Spain* (Garrido L, Ed.), pp 137–148, Springer, Berlin.
- [48]. Urry DW (1992) Free energy transduction in polypeptides and proteins based on inverse temperature transitions, *Prog Biophys Mol Biol* 57, 23–57. [PubMed: 1549698]
- [49]. Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, Pattanaik A, and Harris RD (1992) Hydrophobicity scale for proteins based on inverse temperature transitions, *Biopolymers* 32, 1243–1250. [PubMed: 1420991]
- [50]. Cho Y, Sagle LB, Iimura S, Zhang Y, Kherb J, Chilkoti A, Scholtz JM, and Cremer PS. (2009) Hydrogen bonding of beta-turn structure is stabilized in D(2)O, *J Am Chem Soc* 131, 15188–15193. [PubMed: 19919159]
- [51]. Simon JR, Carroll NJ, Rubinstein M, Chilkoti A, and Lopez GP (2017) Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity, *Nat Chem* 9, 509–515. [PubMed: 28537592]
- [52]. West JA, Mito M, Kurosaka S, Takumi T, Tanegashima C, Chujo T, Yanaka K, Kingston RE, Hirose T, Bond C, Fox A, and Nakagawa S (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization, *The Journal of Cell Biology*.
- [53]. Fei J, Jadhaliha M, Harmon TS, Li ITS, Hua B, Hao Q, Holehouse AS, Reyer M, Sun Q, Freier SM, Pappu RV, Prasanth KV, and Ha T (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution, *J Cell Sci*.
- [54]. Vitalis A, Wang X, and Pappu RV (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization, *J Mol Biol* 384, 279–297. [PubMed: 18824003]
- [55]. Holehouse AS, Garai K, Lyle N, Vitalis A, and Pappu RV (2015) Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation, *J Am Chem Soc* 137, 2984–2995. [PubMed: 25664638]
- [56]. Asthagiri D, Karandur D, Tomar DS, and Pettitt BM (2017) Intramolecular Interactions Overcome Hydration to Drive the Collapse Transition of Gly15, *J Phys Chem B* 121, 8078–8084. [PubMed: 28774177]
- [57]. Brady JP, Farber PJ, Sekhar A, Lin YH, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, and Kay LE (2017) Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation, *Proc Natl Acad Sci U S A* 114, E8194–E8203. [PubMed: 28894006]
- [58]. Wang A, Conicella AE, Schmidt HB, Martin EW, Rhoads SN, Reeb AN, Nourse A, Ramirez Montero D, Ryan VH, Rohatgi R, Shewmaker F, Naik MT, Mittag T, Ayala YM, and Fawzi NL (2018) A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing, *EMBO J*.
- [59]. Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, Amaya J, Burke KA, Mittal J, and Fawzi NL (2018) Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation, *Mol Cell* 69, 465–479 e467. [PubMed: 29358076]
- [60]. Altmeyer M, Neelsen KJ, Teloni F, Pozdnyakova I, Pellegrino S, Grofte M, Rask MB, Streicher W, Jungmichel S, Nielsen ML, and Lukas J (2015) Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose), *Nat Commun* 6, 8088. [PubMed: 26286827]

- [61]. Mousavi A, and Hotta Y (2005) Glycine-rich proteins: a class of novel proteins, *Appl Biochem Biotechnol* 120, 169–174. [PubMed: 15767691]
- [62]. Lin YH, and Chan HS (2017) Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated, *Biophys J* 112, 2043–2046. [PubMed: 28483149]
- [63]. Kim S, Yoo HY, Huang J, Lee Y, Park S, Park Y, Jin S, Jung YM, Zeng H, Hwang DS, and Jho Y (2017) Salt Triggers the Simple Coacervation of an Underwater Adhesive When Cations Meet Aromatic pi Electrons in Seawater, *ACS Nano* 11, 6764–6772. [PubMed: 28614666]
- [64]. Seuring J, and Agarwal S (2012) Polymers with upper critical solution temperature in aqueous solution, *Macromol Rapid Commun* 33, 1898–1920. [PubMed: 22961764]
- [65]. Gallivan JP, and Dougherty DA (1999) Cation-pi interactions in structural biology, *Proc Natl Acad Sci U S A* 96, 9459–9464. [PubMed: 10449714]
- [66]. Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, and Tycko R (2017) Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains, *Cell* 171, 615–627 e616. [PubMed: 28942918]
- [67]. Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, Cascio D, Chong L, Gonen T, and Eisenberg DS (2018) Atomic structures of low-complexity protein segments reveal kinked beta sheets that assemble networks, *Science* 359, 698–701. [PubMed: 29439243]
- [68]. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, and McKnight SL (2012) Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies, *Cell* 149, 768–779. [PubMed: 22579282]

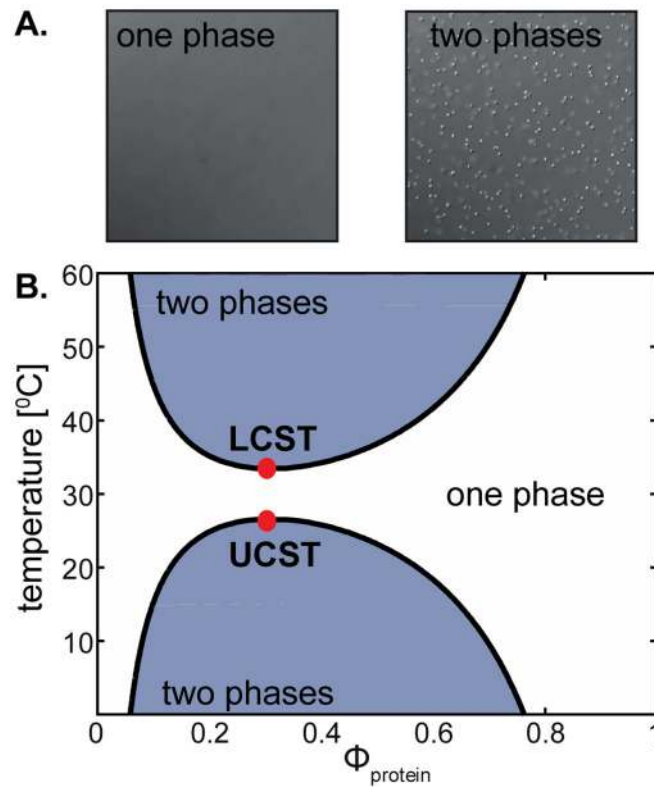


Figure 1. Co-existence lines with lower and upper critical solution temperatures.

A) Phase separation of biopolymers can be observed macroscopically using differential interference contrast (DIC) microscopy and is manifest as the formation of droplets. B) Biopolymers in solution can be in the one-phase or two-phase regime; in the latter, a dense and a light phase coexist. Whether the phase transitions have an upper or lower critical solution temperature (UCST or LCST, respectively) is defined by their critical temperatures for mixing; if the protein mixes when the temperature is lowered, the phase transition has a LCST; if mixing occurs at increasing temperature, the phase transition has a UCST.

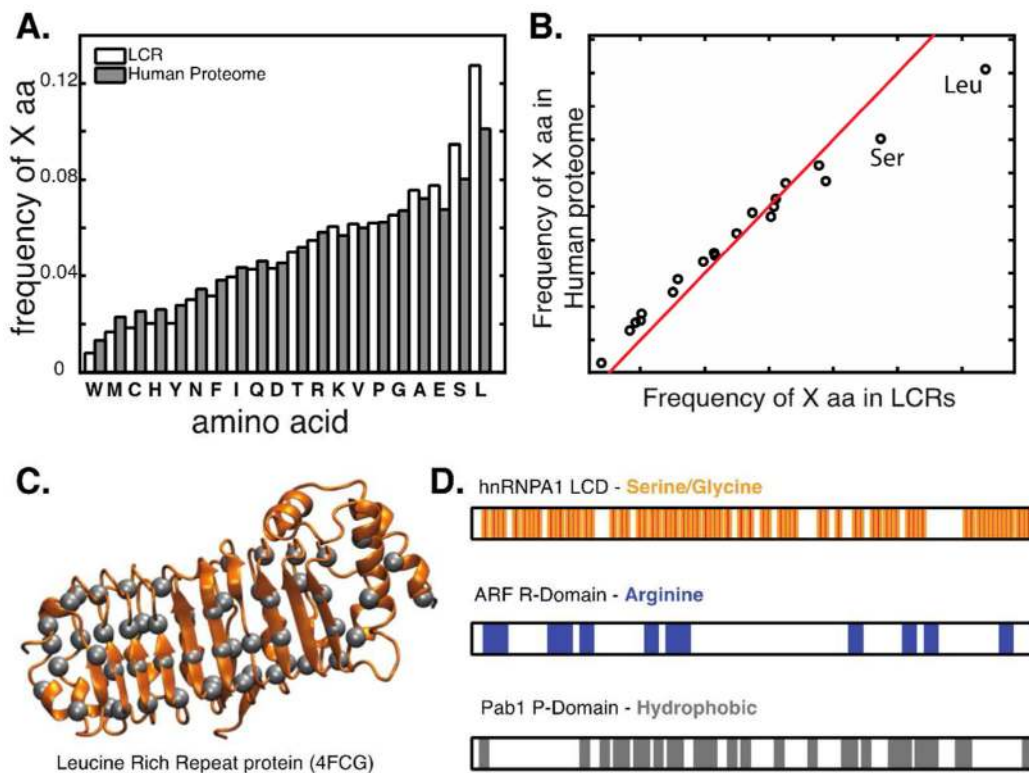


Figure 2. Low-complexity regions are highly variable in composition and not all undergo phase transitions under physiologically relevant conditions.

A) The frequency of amino acid types in the entire human proteome (grey) is compared with the frequency of amino acid types in statistical low-complexity regions (white) found using the SEG algorithm²³. The SEG algorithm scores the amino acid frequency within a sliding window (10 residues) versus the probability of that sequence occurring randomly. Regions with complexity scores below a threshold value were extracted as LCRs and the amino acid composition compared to the complete human proteome. B) The frequency of amino acids in statistical low-complexity regions correlates well ($R = 0.989$) with the frequency in the bulk proteome. The red line represents the ideal correlation. Rare amino acids (W / M / C / H / Y) are underrepresented in LCRs while common amino acids (L / S) appear enriched in LCRs. C) The structure of a leucine-rich repeat (PDB: 4FCG), which is a common LCR and a folded domain, but does not undergo liquid-liquid phase separation under physiological conditions and does not preferentially localize to biomolecular condensates. Leucine residues are represented by grey spheres. D) A schematic representation of the LCRs of hnRNPA1, Arf and Pab1 with their respective enriched amino acid types indicated, i.e. serine/glycine, arginine and hydrophobic residues, respectively. These LCRs mediate phase separation homotypically⁹ or in case of Arf with its binding partner Npm1.²⁵

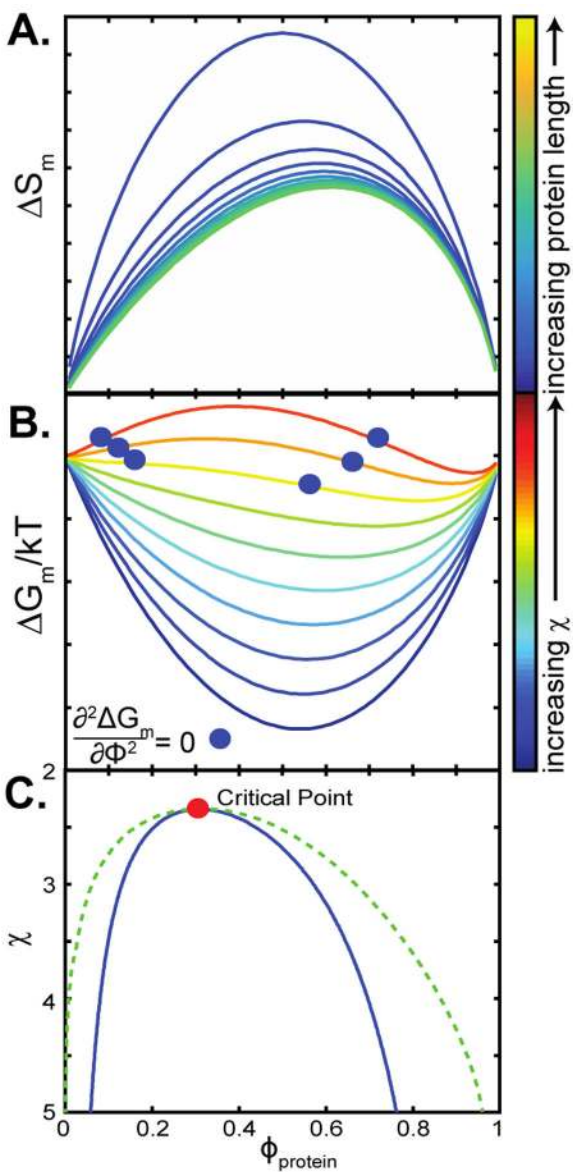


Figure 3. Protein length and interchain interaction potential potentiate phase separation.

A) The combinatoric entropy of mixing is shown as a function of chain length. Increasing chain length decreases the entropic cost of demixing, particularly at lower volume fractions. B) The free energy as a function of the interaction parameter χ . Increasing the strength of protein-protein interactions versus protein-solvent interactions can drive phase separation. The conditions for spinodal demixing occur where the second derivative of the free energy is zero, as noted by blue circles. Under conditions for the green to blue curves, spinodal decomposition does not occur. C) The binodal curve (dashed green line) defines the coexistence of light and dense protein phases, the spinodal curve defines the unstable region below which phase separation must occur. Between the binodal and spinodal curves is the region of metastability, in which phase separation occurs if it is nucleated.

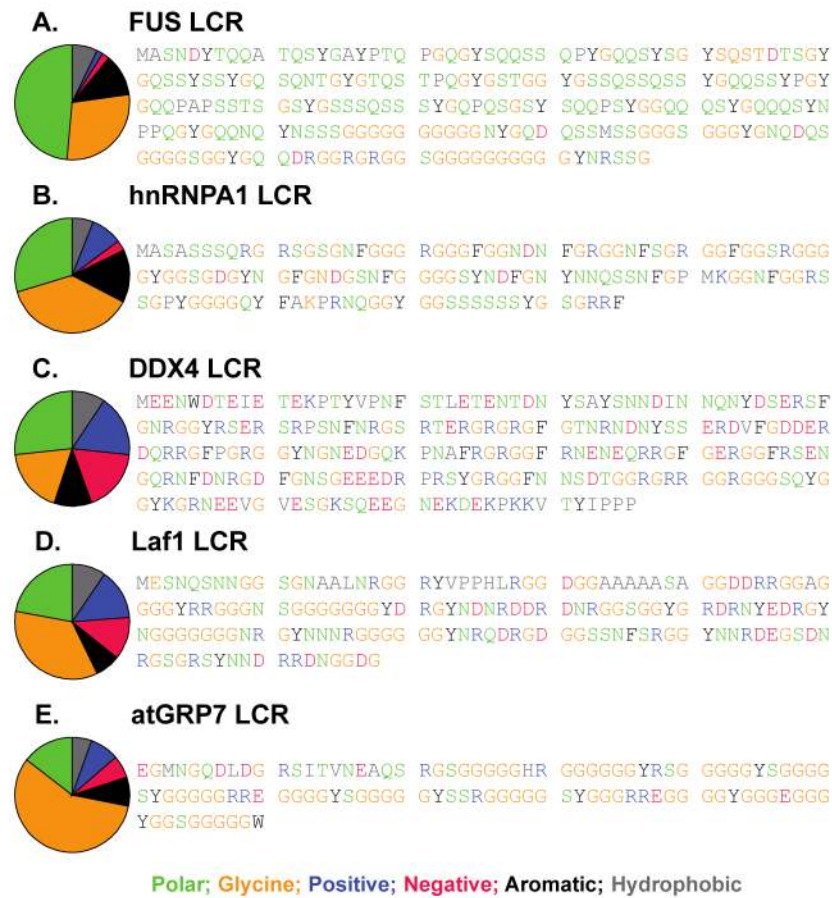


Figure 4. Polar LCRs are enriched in glycine and polar residues. FUS, hnRNPA1, DDX4, Laf-1 and atGRP7 (A-E) LCR sequences are shown with polar, glycine, aromatic, hydrophobic, negatively charged and positively charged residues colored green, orange, black, grey, red and blue, respectively. The amino acid fractions for each sequence are displayed as a pie chart with similar coloring.

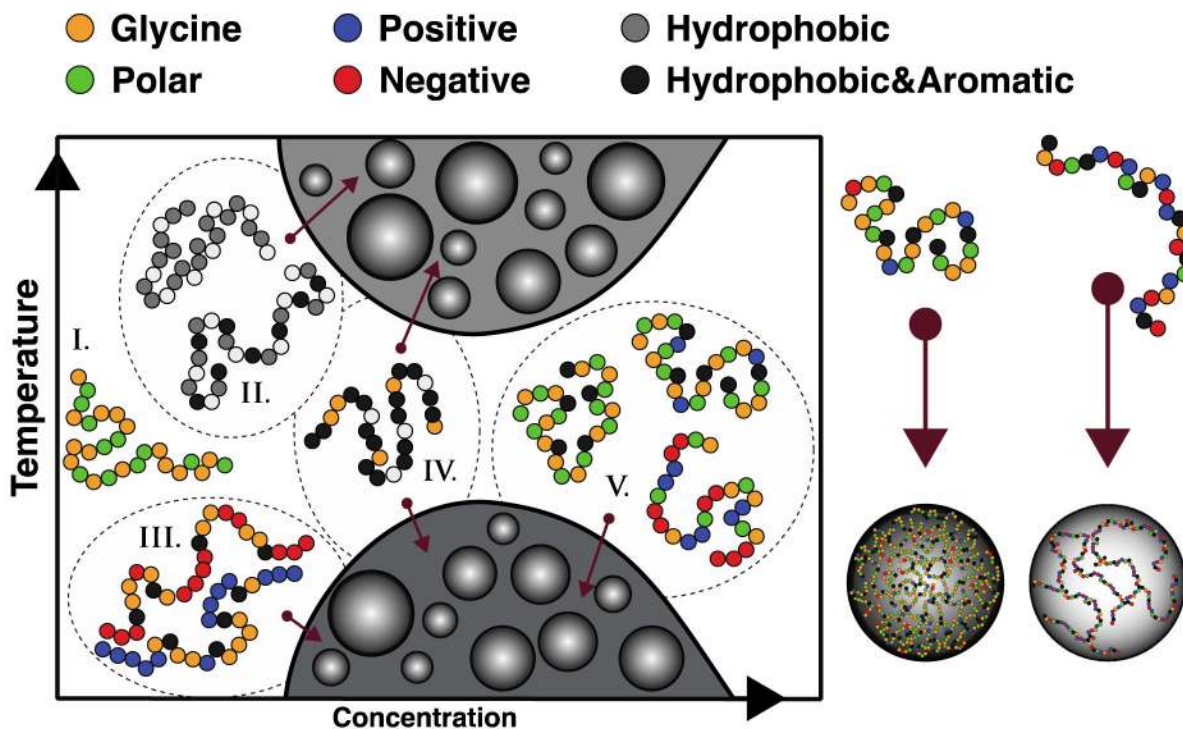


Figure 5. The phase behavior of LCRs is encoded in their primary amino acid sequence. Existing experimental data suggests that whether a LCR is soluble at typical solution conditions or demixes with UCST or LCST behavior depends on the balance of hydrophobic versus polar / charged amino acids. A purely polar LCR (I) is likely soluble over a wide temperature range. The addition of hydrophobic amino acids (II) results in LCST behavior. A mixture of oppositely charged polymers has UCST behavior. If the hydrophobic amino acids are predominately aromatic (IV) the LCR could have either a UCST or LCST transition. Finally, polar LCRs with aromatic amino acids and increasing charge (V) has UCST behavior. Further, the intervening sequence space modulates the properties of the dense protein phase. For example, increasing the fraction of charged residues could result in a dense protein phase that has a higher fraction of solvent and is therefore less dense.