Sequence grammar underlying unfolding and phase separation of globular proteins

Kiersten M. Ruff[1,§], Yoon Hee Choi[2,§], Dezerae Cox[2], Angelique R. Ormsby[2], Yoochan Myung[3,4,5], David B. Ascher[3,4,5], Sheena E. Radford[6], Rohit V. Pappu[1,†], Danny M. Hatters[2,†]

[1]Department of Biomedical Engineering, Center for Science & Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO 63130, USA

[2]Department of Biochemistry and Pharmacology; and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, VIC 3010. Australia

3Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

[4] Structural Biology and Bioinformatics, Department of Biochemistry and Pharmacology, University of Melbourne, Melbourne, Victoria, Australia

[5] Systems and Computational Biology, Bio21 Institute, University of Melbourne, Melbourne, Victoria, Australia

[6]Astbury Centre for Structural and Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK

§Equal contributions

†Corresponding authors:

R.V. Pappu (pappu@wustl.edu)

D.M. Hatters (dhatters@unimelb.edu.au)

Summary

Protein homeostasis involves regulation of the concentrations of unfolded states of globular proteins. Dysregulation can cause phase separation leading to protein-rich deposits. Here, we uncover the sequence-grammar that influences the triad of folding, binding, and phase equilibria of unfolded proteins in cells. We find that the interactions that drive deposit formation of ALS-associated superoxide dismutase 1 mutations are akin to those that drive phase separation and deposit formation in variants of a model protein, barnase. We examined a set of barnase variants to uncover the molecular interactions that drive phase separation of unfolded proteins and formation of unfolded protein deposits (UPODs). The formation of UPODs requires protein destabilization, to increase the concentration of unfolded states, and a requisite sequence grammar to enable cohesive interactions among unfolded proteins. We further find that molecular chaperones, Hsp40 and Hsp70, destabilize UPODs by binding preferentially to and processing unfolded proteins in the dilute phase.

Highlights

- Unfolded states of globular proteins undergo aggregation-mediated phase separation to form unfolded protein deposits (UPODs) in cells.
- The threshold concentration of unfolded proteins required to drive phase separation is governed by a combination of protein stability and the requisite grammar of cohesive motifs known as stickers that provide cohesive intermolecular interactions.

- The sequence grammar that contributes to phase separation of unfolded proteins shares features with those of proteins that drive the formation of functional biomolecular condensates.
- Molecular chaperones regulate the concentrations of free unfolded proteins in cells by binding preferentially to unfolded states in dilute phases thereby destabilizing UPOD formation mainly via binding-mediated polyphasic linkage.

Keywords

Intrinsically Foldable Proteins; Molecular grammar; Phase separation; Protein Folding Equilibrium; Proteostasis; Unfolded Protein Deposits.

Introduction

In cells, protein homeostasis (proteostasis) is achieved by linked processes that control protein production, folding, trafficking, and degradation (Balch et al., 2008; Powers et al., 2009). One of the major functions of proteostasis is to facilitate the correct folding of globular proteins that have a stable fold (Bobori et al., 2017; Reinle et al., 2021; Sontag et al., 2017). We refer to these proteins as intrinsically foldable proteins. It is well established that unfolded or misfolded forms of intrinsically foldable proteins can undergo concentration-dependent, aggregation-mediated phase separation to form aberrant deposits that are also referred to as inclusions, tangles, and plaques (Dobson, 2004; Iadanza et al., 2018; Nedelsky and Taylor, 2019; Vendruscolo et al., 2011). Aberrant deposits are pathological hallmarks of neurodegenerative disorders such as Alzheimer's disease (Arendt et al., 1984; Jacobs et al., 1985; Kidd et al., 1985; Perl and Brody, 1980), Huntington's disease (Becher et al., 1998; Cooper et al., 1998; Davies et al., 1998), Parkinson's disease (Masliah et al., 2000), and amyotrophic lateral sclerosis (ALS) (Alberti et al., 2017; Dao et al., 2018; Mathieu et al., 2020; Wood et al., 2003). The formation of aberrant deposits is mediated by numerous factors including mutations to specific proteins, cellular aging, post-translational modifications, infections, and maladapted stress responses (Hofweber and Dormann, 2019; Sinnige et al., 2020). Collectively, these parameters impact the interconnected equilibria of folding, binding, and phase separation – a triad that requires regulation by the proteostasis machinery (Alberti and Hyman, 2021; Chen et al., 2020; Darling and Shorter, 2021; McCray and Taylor, 2008; Zbinden et al., 2020).

Phase separation also has functional relevance in living cells (Banani et al., 2017; Shin and Brangwynne, 2017). Multivalent protein and nucleic acid molecules drive the formation of membraneless biomolecular condensates, which are implicated in a variety of cellular functions (Mitrea and Kriwacki, 2016). As micron- or sub-micron sized structures, condensates can function as viscous liquids, viscoelastic fluids, gels, liquid crystals, or solids (Alshareedah et al., 2021; Choi et al., 2020; Franzmann et al., 2018; Kaur et al., 2019; Kroschwald et al., 2018; Shen et al., 2020; Shin and Brangwynne, 2017; Yu et al., 2021). Among the molecular drivers of phase separation are distinct types of intrinsically disordered proteins (IDPs) (Posey et al., 2018a) with the requisite sequence grammar (Greig et al., 2020; Kim et al., 2019; Martin et al., 2020; Martin and Mittag, 2018; Vernon et al., 2018; Wang et al., 2018). This underlying grammar has been described in terms of a stickers-and-spacers framework (Choi et al., 2019; Choi et al., 2020). Here, stickers are cohesive motifs that drive phase separation by forming reversible physical crosslinks. Spacers are interspersed between stickers and determine the synergy between phase separation and gelation (Harmon et al., 2018; Harmon et al., 2017). A range of sticker-sticker interactions have been uncovered and these include interactions between: aromatic residues (Chong et al., 2018; Kim et al., 2019; Lin et al., 2017; Martin et al., 2020; Vernon et al., 2018), oppositely charged residues

(Greig et al., 2020; Pak et al., 2016), cation-π pairs (Nott et al., 2015; Wang et al., 2018), distributed hydrophobic residues (Riback et al., 2017), and those that have specific structural features (Conicella et al., 2020).

We reason, based on significant precedence in the literature, that many aberrant deposits are the result of concentration-dependent interactions among unfolded or misfolded proteins (Balchin et al., 2020; Clark, 2004; Hartl, 2016; McMillan et al., 2005; Ryno et al., 2013; Solomon et al., 2012; Song, 2018). For example, in the context of some familial forms of ALS, mutations to superoxide dismutase 1 (SOD1) affect the stability of the SOD1 dimer and promote the formation of amorphous protein deposits that accumulate through interactions among unfolded monomeric states (Gomez and Germain, 2019; Meiering, 2008).

Aggregation-mediated phase separation is a density transition whereby a protein plus solvent system separates into a dilute, protein-deficient phase and a coexisting dense, protein-rich phase (Mathieu et al., 2020; Pappu et al., 2008; Posey et al., 2018a) (**Figure 1A**). For a given set of solution conditions, the strengths of driving forces for phase separation driven by homotypic interactions can be quantified by a saturation concentration, c_{sat}, which is the threshold concentration of the protein above which it separates into coexisting dilute and dense phases (Wang et al., 2018). Thus, lower values of c_{sat} imply stronger driving forces for phase separation. The material states of dense, protein-rich phases can be liquid-, gel-, or solid-like. An example of a dense phase is the gel-like deposits of mutant SOD1 in the spinal cord tissue of individuals with ALS (Farrawell et al., 2015; Hwang et al., 2010). While soluble wild-type (WT) SODI exists as a homodimer, stabilized by metal binding and an intra-subunit disulfide bond, evidence suggests that aberrant phase separation and formation of SOD1 deposits is driven by interactions among unfolded immature states of SOD1 (Nordlund et al., 2009). With this in mind, understanding how unfolded SOD1 – and unfolded states of globular proteins in general – engage in intermolecular interactions is central to understanding the molecular determinants of how deposits are formed by intrinsically foldable proteins.

Are all unfolded states of intrinsically foldable proteins equivalent as drivers of phase separation and the formation of aberrant, *de novo* deposits in cells, or must the unfolded states expose distinctive stickers that can drive phase separation? Here we answer this question by utilizing the model protein barnase, whose structure, stability and folding *in vitro* (Dalby et al., 1998; Matthews and Fersht, 1995) and *in vivo* (Wood et al., 2018) have been extensively studied. The answers to this question are directly relevant to the targeting and processing of aberrant deposits, either via proteostatic regulation or by suppressing the process of phase separation. Our work here focuses on uncovering the mechanisms of phase separation linked to the folding-unfolding reactions of intrinsically foldable proteins under cellular conditions.

Results

Destabilizing mutations drive aberrant phase separation of SOD1

We first investigated the impact of ALS-related mutations on the phase behavior of apo-states of SOD1. To avoid confounding effects of SOD1 dimerization, we focused on an obligate monomeric derivative called $SOD1_{barrel}$ that maintains the folding characteristics of SOD1 without binding to metals or chaperone cofactors (Danielsson et al., 2011). We focused on mutations in $SOD1_{barrel}$ at sequence positions A4, G93, and I35 because these were shown to destabilize the

folded state (Danielsson et al., 2011; Danielsson et al., 2015; Gnutt et al., 2019). We characterized destabilizing single mutations A4V and G93A (Danielsson et al., 2011; Danielsson et al., 2015; Gnutt et al., 2019), and the double mutant (A4V, G93C) (**Figure S1A**).

Using pulse shape analysis (PulSA) flow cytometry (Ramdzan et al., 2012) we determined which cells contained deposits (**Figure S1B**) and quantified the concentration thresholds, c_{sat}, at which the deposits formed (**Figures 1B** and **S1C-H**). The WT $SOD1_{barrel}$ did not form deposits at any expression level (**Figures 1B** and **S1D**). In contrast, each mutational variant of $SOD1_{barrel}$ formed deposits (**Figure 1B** and **S1B, E-H)**. The formation of deposits by destabilized variants points to phase separation being driven by an increased abundance of unfolded proteins. Thus, it would appear that above a threshold concentration (Ciryam et al., 2015), cohesive interactions among unfolded proteins drive phase separation. We hypothesize that some of the residues important for these intermolecular interactions are those that are normally sequestered in the core of the folded state (**Figure 1C**). The question is if the only requirement is that the concentration of unfolded proteins cross a generic threshold or if there are additional considerations that arise? We answer this question by comparing the phase behavior of SOD1 to a model protein, barnase, and using the latter as a model system to uncover the interplay between protein stability and sequence grammar as drivers of phase separation.

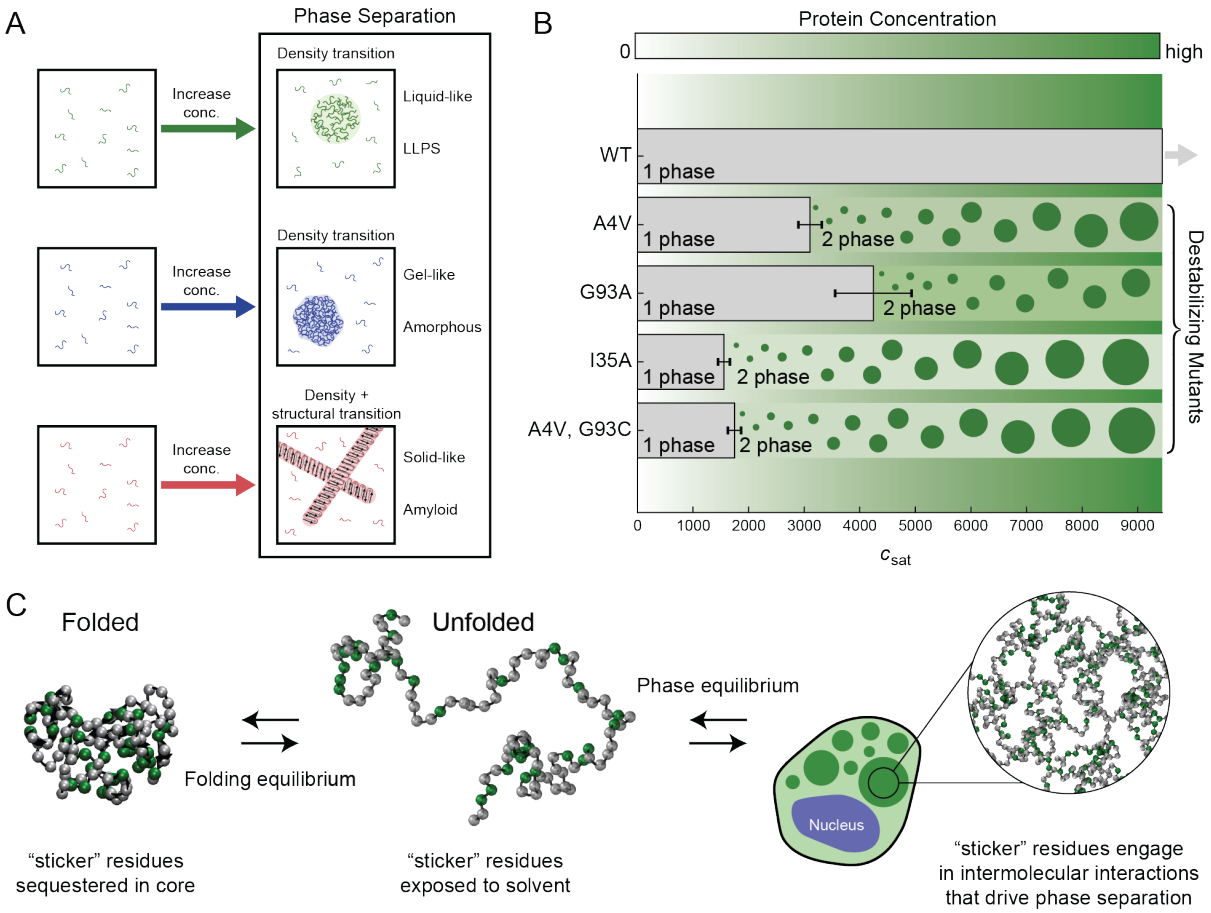

Figure 1. Destabilizing mutations promote phase separation. (A) The schematic shows how crossing a distinct concentration threshold gives rise to three categories of protein phase separation, distinguished by the material properties of the dense phase. (B) The concentration threshold above which the system separates into two distinct phases. This is designated as c_{sat} and is given in units of fluorescence intensity (referred to as arbitrary units). We quantify c_{sat} for WT $SOD1_{barrel}$ and four other $SOD1_{barrel}$ mutational variants. Error bars indicate standard deviations across three replicates. Note that $SOD1_{barrel}$ does not phase separate across the wide range of expression levels examined (denoted by the gray arrow). The dark green circles represent the dense phase, and circle sizes indicate stabilization of the dense phase beyond c_{sat}. The dilute phase is represented by a single light green color corresponding to the c_{sat}. The one-phase regime lies below c_{sat} and is depicted as a gray box. (C) Illustration of the hypothesis that interactions among exposed hydrophobic "sticker" residues within unfolded proteins drive phase separation and are linked to the folding-unfolding equilibrium. Here, green spheres correspond to sticker residues and gray spheres correspond to spacer residues. When the total protein concentration $c_{tot} > c_{sat}$, the homogeneous well-mixed phase is unstable, and the system separates into two coexisting phases. See also **Figure S1**.

The sequence grammar that drives phase separation of unfolded states appears to be similar between SOD1 and a model protein barnase

While SOD1 is physiologically important, the complexities associated with the multi-way linkage of dimerization, metal binding, and its normal cellular functions in mitigating oxidative damage make it ill-suited as a model system for systematic characterization of the molecular interactions that are likely to be shared across unfolded states of generic intrinsically foldable proteins. In contrast, the small bacterial ribonuclease barnase is a model monomeric globular protein that has been studied extensively in the context of protein folding (Dalby et al., 1998; Matthews and Fersht, 1995). Prior work has measured the free energy of unfolding (ΔG°_{U}) of a wide variety of barnase variants, such that the relative concentrations of folded and unfolded states can be easily estimated. Importantly, the catalytically dead H102A variant (referred to as the WT form for the purposes of this study), is benign as far as mammalian cell physiology and viability are concerned (Wood et al., 2018).

WT barnase does not form visible deposits in mammalian cells, when expressed fused to mTFP1 at the N-terminus and Venus at the C-terminus (Wood et al., 2018). In contrast, variants with decreased values of ΔG°_{U}, which increases the concentration of unfolded proteins, form visible deposits. Unfolded states of barnase are also known to engage with components of the quality control machinery implying that we can interrogate the three-way interplay of protein stability, phase separation, and engagement with the quality control machinery (Wood et al., 2018).

We assessed whether the deposits formed by a barnase destabilizing variant colocalized with deposits formed by SOD1 A4V. Colocalization would imply that that phase separation of unfolded states of generic intrinsically foldable proteins and unfolded SOD1 are governed by similar driving forces. When co-expressed in HEK293T cells, the destabilized double mutant of barnase (I25A, I96G, ΔG°_{U}=-0.8 kJ/mol) formed deposits that colocalized with those of SOD1 A4V (**Figure 2A**). Conversely, WT barnase (ΔG°_{U}=24.9 kJ/mol) remained diffuse and did not colocalize with SOD1 A4V deposits. We interpret these results to imply that phase separation of unfolded barnase and SOD1 are driven by similar physical interactions.

To further interrogate the underlying physical chemistry that dictates phase separation of unfolded barnase, we examined the colocalization of barnase deposits with those of polyglutamine (polyQ)-expanded Huntingtin exon 1 (Httex1), a protein that forms amyloid-like solids in cells (Bäuerlein et al., 2017). Httex1 is an IDP with an overall tadpole-like architecture (Newcombe et al., 2018; Warner et al., 2017). Its phase behavior is driven mainly by amide-amide interactions involving the polyQ domain (Crick et al., 2013; Posey et al., 2018b). These interactions are distinct

from hydrophobic interactions anticipated to be responsible for driving phase separation of unfolded barnase or SOD1. As with previous work on SOD1 destabilizing variants, the barnase I25A, I96G deposits did not co-localize with Httex1-72Q deposits (**Figure 2B**) (Farrawell et al., 2015; Polling et al., 2014). This lack of colocalization supports the hypothesis that distinctive interactions underlie the phase behavior of SOD1 and barnase variants when compared to Httex1.

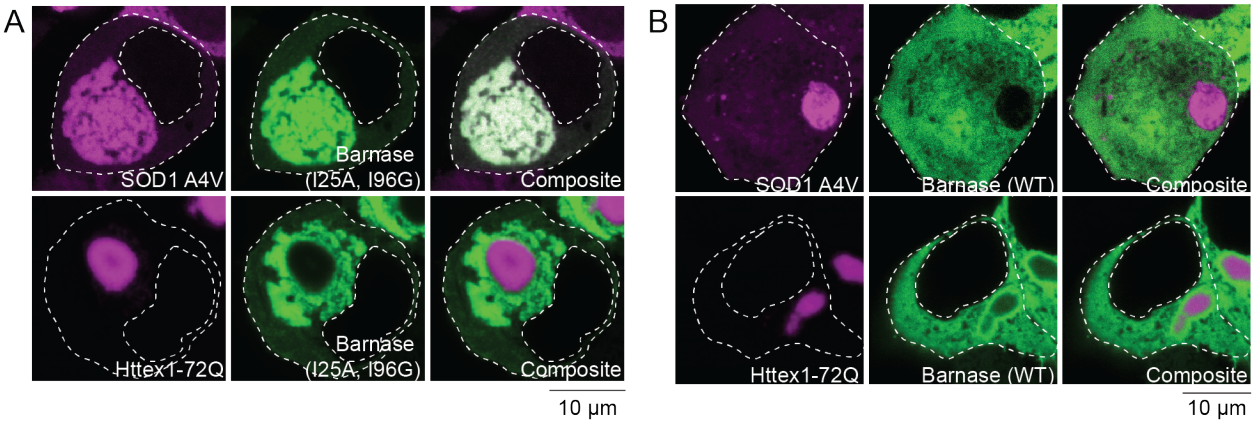

Figure 2. Destabilizing variants of barnase and SOD1 share a common phase separation grammar that differs from that of Huntingtin exon 1 with an expanded polyQ tract. (A) The fluorescence micrographs show deposits formed by a destabilized variant of barnase (I25A, I96G) flanked with fluorescent proteins (mTFP1 and Venus) (Wood et al., 2018) along with deposits formed by mutant SOD1 (SOD1 A4V) or mutant Httex1 containing a glutamine tract of 72 residues (Httex1-72Q) fused to mCherry. The constructs were co-transfected in HEK293T cells. (B) Images show similar experiments but with WT barnase flanked by two fluorescent proteins and the SOD1 A4V or Httex1-72Q fusions to mCherry. The outlines of the cells and nuclei are shown with dashed lines.

Phase separation is driven by interactions among unfolded barnase molecules

We focused on uncovering the sequence grammar that underlies the phase separation of mutant variants of barnase. For this, an optoDroplet system was deployed to study phase separation of unfolded barnase molecules in live cells. The optoDroplet system was developed as a tool to study the phase separation of multivalent proteins in cells using a precise and controllable reaction triggered by blue light (Shin et al., 2017). The system involves a fusion of the protein of interest to the photoactivatable Cry2 domain (Hsu et al., 1996; Lin et al., 1998) and a fluorescent protein reporter (**Figure 3A**). Cry2 forms sub-microscopic oligomers upon blue light illumination and is insufficient to drive phase separation upon light activation (Lin et al., 1998). However, when fused to a domain that can undergo phase separation, the oligomerization of Cry2 reduces c_{sat} for the test protein allowing quantitative and inducible comparison of apparent c_{sat} values to be measured (Shin et al., 2017).

The IDR of DDX4, which forms condensates, was used as a positive control (**Figure 3B**) because it has been shown to undergo reversible phase separation in the optoDroplet setup (Brady et al., 2017; Nott et al., 2015; Shin et al., 2017). WT barnase did not form droplets when Cry2 was light activated (**Figure 3B**). However, the (I25A, I96G) double mutant, which has a finite probability of accessing unfolded states under physiological conditions, rapidly formed droplets in a blue-light dependent manner (**Figure 3B, 3C**). These results suggest that phase separation of unfolded states of barnase can be assessed in a controlled manner without the confounding effects of slow kinetics that typically characterize the formation of deposits in cells.

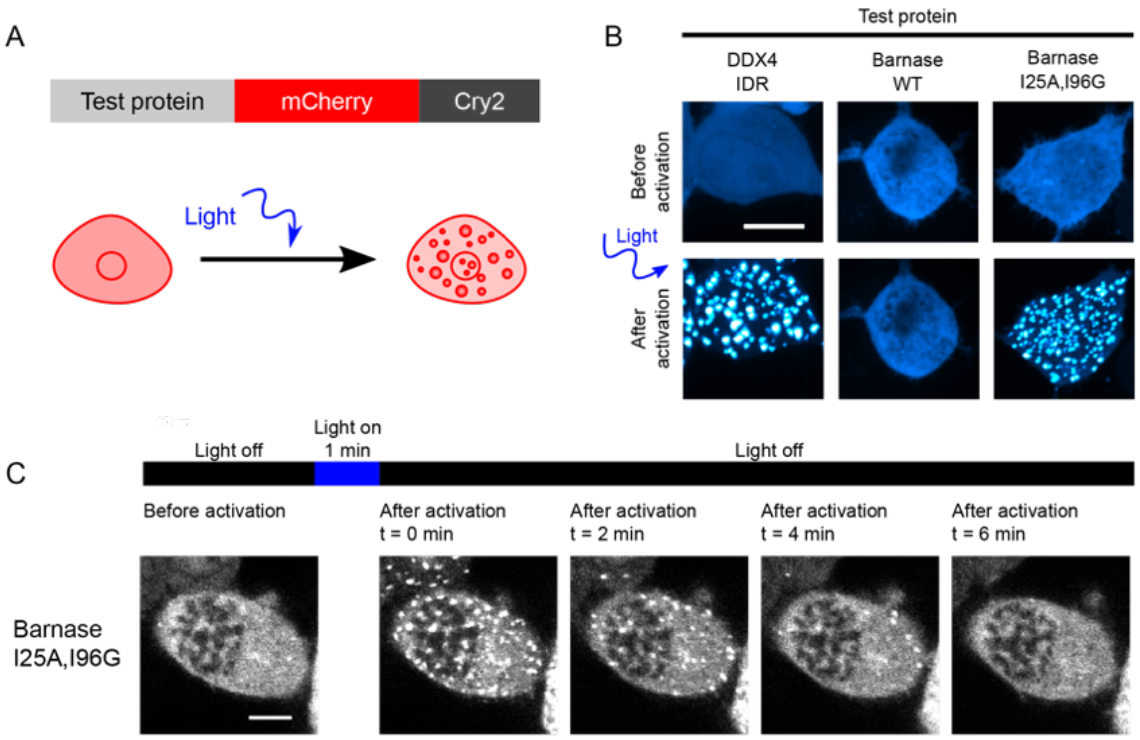

Figure 3. Phase separation is driven by interactions among unfolded barnase molecules. (A) Schematic of constructs used for the optoDroplet assay. (B) Representative confocal micrograph images of Neuro2a cells transfected with DDX4 IDR (positive control), WT barnase, and the destabilizing barnase variant (I25A, I96G) optoDroplet constructs before and after light activation. (C) Time lapsed confocal imaging of live Neuro2a cells expressing the I25A, I96G barnase optoDroplet construct. Scale bars in panels B and C correspond to 10 µm.

Protein destabilization and a distinct sequence grammar are required for phase separation

We examined a range of mutations in barnase to titrate the impact of ΔG°_{U} on phase separation. The ΔG°_{U} values for these variants ranged from 18.7 kJ/mol (highly stable) to –0.8 kJ/mol (highly unstable) (Wood et al., 2018). Mutations with ΔG°_{U} values above 13.0 kJ/mol were resistant to droplet formation, whereas those with ΔG°_{U} values below this threshold readily formed droplets (**Figure 4A**). If abundance of unfolded proteins, dictated by ΔG°_{U}, is the sole determinant for phase separation, then there should be a threshold concentration of unfolded proteins, c^*, above which the system separates into dilute and dense phases (**Figure 4B**). This concentration quantifies the saturation threshold of the unfolded species. The value of $c^* = p_U \times c_{sat}$ is a product of the apparent saturation concentration, c_{sat}, with respect to total barnase (folded and unfolded proteins), and p_U, which is the fraction of molecules in the unfolded state. Positive ΔG°_{U} values imply that the fraction of unfolded proteins is less than the fraction of folded molecules, and thus a larger c_{sat} value is needed to reach c^*. Note that $c^* \approx c_{sat}$ as p_U approaches 1. Therefore, if phase separation is driven exclusively by the concentration of unfolded proteins, we expect that variants with the lowest fraction of unfolded proteins will have the highest c_{sat} values.

Previous studies have shown that many intrinsically foldable proteins, including barnase, tend to be more unfolded in cells than would be predicted based on estimates of ΔG°_{U} from *in vitro* measurements (Danielsson et al., 2015; Gnutt et al., 2019; Wood et al., 2018). Also, barnase has

two appendages, Cry2 and mCherry, in the optoDroplet system, which are likely to alter the ΔG°_{U} values. To account for this possibility, the fitting procedure we deployed estimates both the value of c^* and the constant offset for ΔG°_{U} in relation to the ΔG°_{U} measured *in vitro* (see Methods). For each barnase variant, we estimated c_{sat} using the dilute phase fluorescence intensity because it corresponds to the threshold concentration for the appearance of droplets (**Figures 4C** and **S2A-B**). The estimated c_{sat} values are well described by a model with $c^* \approx 10{,}830$ fluorescence intensity units (AU) and an offset of -12.9 kJ/mol (-3.1 kcal/mol) for all ΔG°_{U} values (dashed line **Figure 4C**). This analysis is therefore consistent with the hypothesis that the concentration of unfolded states is a major determinant of phase separation.

Mutations that destabilize the folded state of globular proteins often do so by weakening the hydrophobic core. Accordingly, if the residues that drive chain collapse and phase separation are equivalent (Zeng et al., 2020), then destabilizing mutations would be expected to weaken the driving forces for phase separation of unfolded proteins. Therefore, we propose that even if the protein were completely unfolded ($p_U \approx 1$), phase separation will only occur if the requisite sticker residues are present and accessible. To test our hypothesis, we first examined the linear patterning of hydrophobic versus hydrophilic regions (which we refer to as "blobs") to discern whether mutations that remove or preserve hydrophobic blobs affect phase separation. For this, we investigated the sequence features of WT barnase and different variants to identify blobs by adapting a method introduced by Lohia et al. (Lohia et al., 2019). This analysis revealed that the single (I96G) and double (I55G, L89G and I25A, I96G) mutations remove a hydrophobic blob. Additionally, the single (I96G) and double (I25A, I96G) mutations insert a hydrophilic blob. In line with our hypothesis, mutations that remove hydrophobic blobs and / or insert hydrophilic blobs have a higher c_{sat} value than would be expected from the fit to the model that predicts phase behavior strictly based on the fraction of unfolded proteins derived from the values of ΔG°_{U} (**Figures 4C, D and F**). These results suggest that c^* is governed both by p_U and the presence / accessibility of stickers that contribute to cohesive interactions among unfolded proteins.

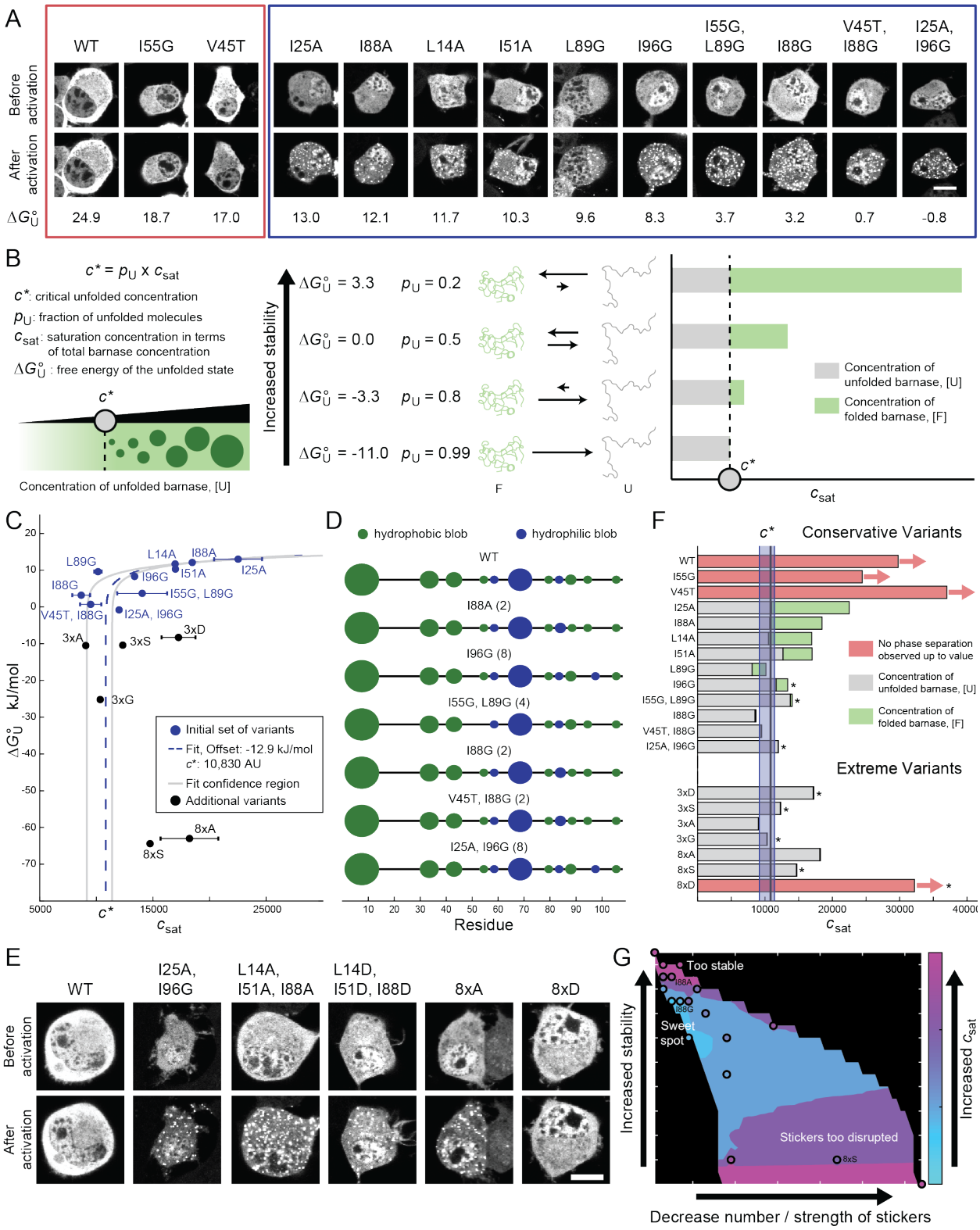

Figure 4. Protein destabilization as well as a distinct sequence grammar are required for phase separation. (A) Representative confocal micrograph images of Neuro2a cells transfected with the barnase-optoDroplet constructs as shown before and after light activation. Red box indicates constructs that do not undergo phase separation, whereas the blue box denotes constructs that do. Scale bar corresponds to 10 µm. (B) Model for how c_{sat} should change if

stability, and thus a critical concentration of unfolded proteins, c^*, is all that matters for phase separation. (C) c_{sat} for optoDroplet formation, in arbitrary units (AU), versus ΔG°_U. Error bars indicate standard deviations across 50 bootstrapped trials. Dashed line corresponds to a fit of the initial barnase construct data (blue circles) to a model that assumes only crossing a critical concentration of unfolded proteins is necessary for phase separation. Fit assumes that barnase is more unfolded in the cell than predicted by ΔG°_U = -12.9 kJ/mol with c^* = 10830 AU. Gray lines denote the fitted confidence interval determined by bootstrapped trials of the variants, where the variants were picked based on the degree to which they modulated the hydrophobic and hydrophilic blobs from WT (Methods). Black points show results for additional, more extreme barnase variants in which effectively all molecules should be in the unfolded state. (D) Sequence blob diagrams of WT and the initial barnase variants in which mutations modulated the hydrophobic and hydrophilic blobs compared to the WT sequence (Methods) (Lohia et al., 2019). Circle sizes denote the number of residues in a blob. Numbers in parentheses quantify the degree of change from WT, i.e., |decrease in size of hydrophobic blobs| + increase in size of hydrophilic blobs. (E) Representative confocal micrograph images of Neuro2a cells transfected with the additional barnase-optoDroplet constructs (black circles in C) with negative ΔG°_U values. Scale bar indicates 10 µm. (F) c_{sat} of all barnase-optoDroplet constructs examined, in AU, ordered by ΔG°_U. The blue region corresponds to the gray c^* boundary in (C). Bars with asterisks indicate constructs in which the blob diagrams changed by at least four from WT. (G) Schematic for how modulating stability and the valence / strength of stickers changes c_{sat}. Black regions indicate inaccessible regions based on mutations designed to reduce stability of the hydrophobic core and decrease the hydrophobicity of the sequence. Circles correspond to the barnase-optoDroplet constructs examined, where the sticker axis is determined by x=|decrease in size of hydrophobic blobs| + increase in size of hydrophilic blobs + 2*(number of mutations). The background c_{sat} histogram was determined by randomly sampling c_{sat} values based on the limits of the constructs examined. See also **Figure S2**.

To further explore the importance of hydrophobicity and the requisite number of stickers being present and available for interactions among unfolded proteins, we introduced additional mutations into barnase that effectively ablated the folded state ($\Delta G^\circ_U \leq -10$ kJ/mol). This enabled the quantification of driving forces for phase separation based solely on the properties of unfolded states (**Figures 4C, E, F**, and **S2C**). The mutations were chosen to alter the chemical environment of key bulky hydrophobic residues that would normally be in the core of the folded state. This includes triple mutations L14X, I51X, and I88X with X being A, G, S or D. These mutations are hereafter referred to as the 3×X variants. We also introduced octuple mutations L14X, L42X, I51X, L63X, I76X, I88X, L89X, I96X with X as A, S, or D, referred to as 8×X variants. In terms of hydrophobicity, the substitutions should follow the trend A > G > S > D (Kyte and Doolittle, 1982).

The measured values of c_{sat} for 3×A and 3×G fall within the confidence interval for c^* as defined purely by the concentration of unfolded proteins (**Figures 4C, F** and **S2C**). These mutations are predicted to minimally decrease the hydrophobicity of the unfolded state in comparison to the other triple and octuple mutational variants. By contrast the 3×D and 8×D mutations, which substantially decreased the hydrophobicity of the unfolded state, either significantly increased c_{sat} or abolished phase separation altogether (**Figures 4C, E, F,** and **S2C**). The octuple variants had larger c_{sat} values than the corresponding triple variants. The takeaway from these results is that there is a "sweet spot" for phase separation of intrinsically foldable proteins (**Figure 4G**). Specifically, phase separation requires that the unfolded state be favorably populated and that sticker-mediated interactions be minimally diminished by mutations that destabilize the folded state.

Phe and Tyr function as stickers that drive phase separation of unfolded barnase

To identify residues that function as stickers, we performed atomistic simulations of seven different barnase variants and identified contacts between residues within unfolded states. Residues predicted to be optimal stickers were required to satisfy two criteria: (i) they *should not* engage in contacts that are present in the folded state, which we define as a pair of residues that

are within 4.5 Å of one another; (ii) they should have a high probability of engaging in non-native contacts in unfolded states. Only thirteen sequence positions satisfied these criteria, thereby implicating them as stickers for phase separation (**Figure 5A**). These residues were either hydrophobic-aliphatic (Ile, Leu, and Val) or hydrophobic-aromatic (Phe, Trp, and Tyr). Of particular interest was the enrichment of Tyr and Phe among the set of residues that engage in non-native interactions, given its role in the phase separation of intrinsically disordered prion-like low complexity domains (Bremer et al., 2021; Lin et al., 2017; Martin et al., 2020; Wang et al., 2018) and in forming the selectivity filter of nuclear pore complexes (Frey et al., 2006).

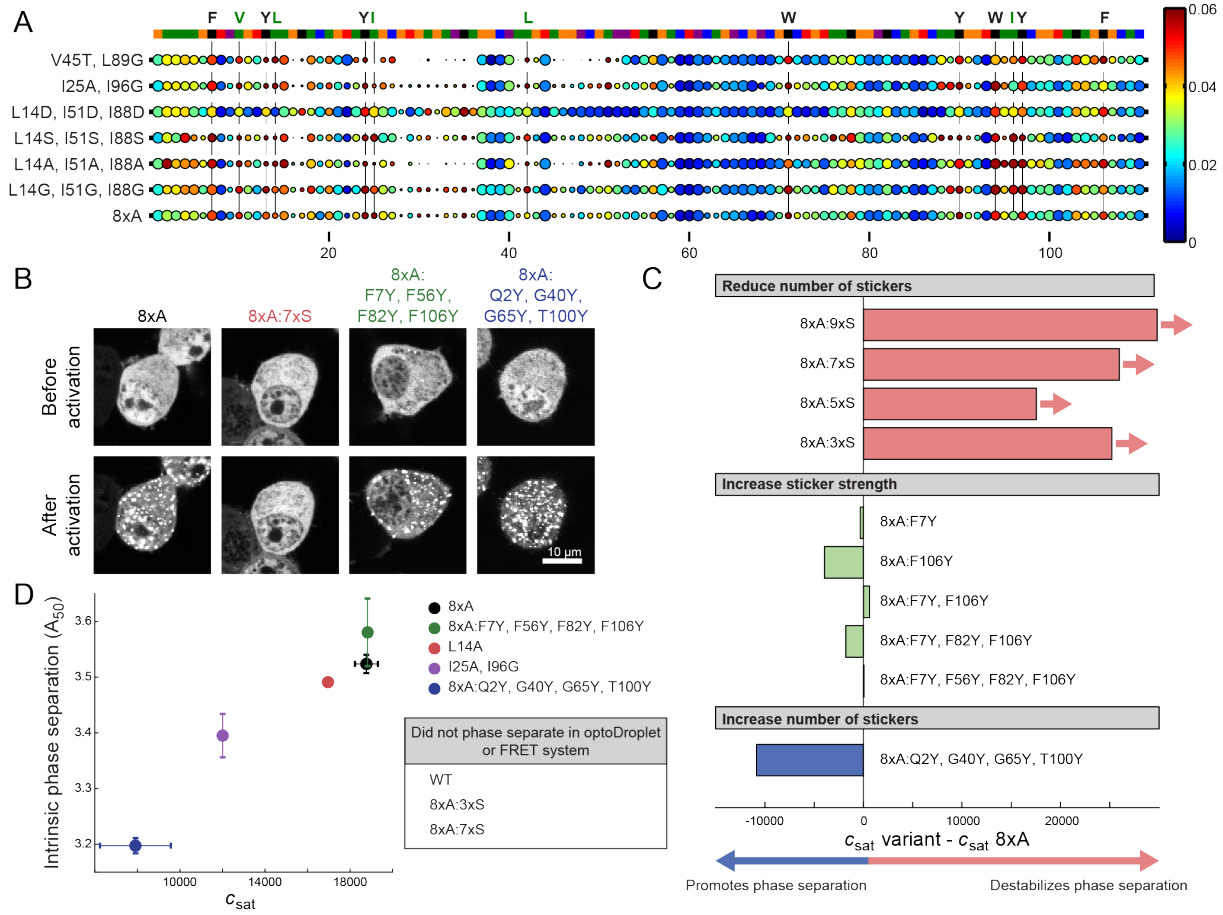

Figure 5. Phe and Tyr function as stickers that drive phase separation of unfolded barnase. (A) The size of each circle quantifies 1 – Q, where Q is the fraction of native contacts involving the residue in question. The color of each circle quantifies the contact order, which refers to the probability that a residue is in contact with any other residue. For each residue, nearest and second nearest neighbour contacts are excluded from the calculation of the contact order. The WT barnase sequence is color-coded across the top: positively charged residues are blue, negatively charged residues are red, aromatics are black, glycine is purple, strongly interacting residues (V, M, I, L, Q, N, H) are green, and weakly interacting residues are orange (A, P, S, T, C) (Ruff et al., 2015). Residues with mean contact order > 0.05 and 1-Q > 0.5 over the seven different variants are indicated by black dashed lines and labelled at the top of the figure. (B) Representative confocal micrograph images of Neuro2a cells transfected with the sticker barnase-optoDroplet variant constructs. (C) Comparison of the c_{sat} values of each sticker barnase-optoDroplet variant construct with the c_{sat} of the 8xA construct, in arbitrary units. Bars with arrows indicate that a c_{sat} value could not be extracted for these constructs and must be at least above the value of the bar. (D) Comparison of the intrinsic phase separation of barnase

as fusions to fluorescent proteins mTFP1 and Venus, using the A_{50} analysis, to c_{sat} values of barnase in the optoDroplet format (R^2 = 0.95 for linear regression). Error bars indicate standard deviations. Also, see **Figure S3**.

We tested the importance of the hydrophobic-aromatic residues as stickers. To avoid confounding factors arising from the folded state, we introduced mutations in the 8×A variant – a variant that is completely unfolded but still drives phase separation (**Figure 4C, E**). Three categories of mutations were examined. First was the replacement of aromatic residues with Ser, which was predicted to reduce the number of stickers. Second was the replacement of Phe with Tyr, which was predicted to increase the sticker strength (Bremer et al., 2021). Third was the replacement of polar residues with Tyr, which was predicted to increase the number of stickers (Bremer et al., 2021; Martin et al., 2020). The effects of these mutations were assessed using the optoDroplet assay (**Figures 5B**, **5C and S3**). These measurements showed that decreasing the number of stickers with Ser mutations weakened the driving forces for droplet formation, whereas polar to Tyr mutations enhanced the driving forces for phase separation. Substituting one or more Phe residues with Tyr had minimal impact on c_{sat}. This suggested that Phe and Tyr have equivalent efficacy as stickers when phase separation is driven by interactions among unfolded barnase molecules.

Interactions that drive phase separation of unfolded states have an equivalent impact on deposit formation

We next asked how our optoDroplet measurements related to the formation of protein deposits in cells. To do this we examined deposit formation using a FRET assay involving the barnase constructs fused to mTFP1 and Venus, where acceptor (Venus) fluorescence provides a readout on the local assembly of barnase molecules (Wood et al., 2018). Specifically, we derived estimates of the concentration of barnase in cells at which 50% of the cells contain deposits (A_{50} value). The lower this concentration, the stronger the driving force for deposit formation. We found a strong positive correlation between the A_{50} and c_{sat} values (R^2 = 0.95 for linear regression) (**Figure 5D**). These measurements, made for five variants, suggest that the driving forces for droplet and deposit formation are equivalent. Furthermore, two variants that did not form deposits also did not form droplets.

Molecular chaperones suppress phase separation of unfolded barnase

Proteostasis involves components of the cellular machinery such as chaperones that bind to unfolded and / or partially unfolded states of proteins and mediate their processing for folding or delivery to other cellular machinery (Hartl, 2016). Here, we sought to explore how chaperones influence phase separation driven by interactions among unfolded barnase molecules.

Upon phase separation, unfolded proteins exist either in the dilute phase or in the dense phase, defined by protein deposits. Components of the chaperone system can bind to unfolded barnase either in the dense (**Figure S4**) or dilute phase (Wood et al., 2018). If binding to unfolded proteins in the dilute phase is stronger than in the dense phase, then c_{sat} in the presence of the chaperone, designated as $c_{sat}^{chaperone}$, will be greater than c_{sat} in the absence of the chaperone. Conversely, if binding to unfolded proteins in the dilute phase is weaker than in the dense phase, then $c_{sat}^{chaperone}$ will be lower than c_{sat} in the absence of the chaperone. This type of preferential binding, which would be true of chaperones that function independently of ATP hydrolysis, is referred to as polyphasic linkage (Ruff et al., 2021b; Wyman and Gill, 1980).

Upon binding of unfolded proteins by chaperones there should be three states of barnase in the dilute phase. These are folded barnase, unfolded barnase, and unfolded barnase bound to chaperones (**Figure 6A**). Previous studies showed that members of the Hsp70 and Hsp40 families can bind to barnase in the dilute phase and suppress deposit formation of unfolded proteins (Wood et al., 2018). Accordingly, we propose that while the total concentration of unfolded proteins (free + bound) is higher in the presence of chaperones, the fraction of molecules capable of phase separation is lowered (**Figure 6A**).

To investigate how chaperones modulate the phase behavior of unfolded proteins, we co-expressed the optoDroplet construct containing the L14A barnase variant with the Hsp70 protein HSPA1A and / or its cofactor, the Hsp40 protein DNAJB1. The canonical model is that Hsp40 binds substrates, and then forms a ternary complex with Hsp70 in the ATP-bound state (Alderson et al., 2016; Jiang et al., 2019) (**Figure 6B**). ATP hydrolysis correlates with release of Hsp40 and the formation of a high affinity complex between substrate and Hsp70. Because the ternary complex is needed for Hsp70 to stimulate ATP hydrolysis and form a high affinity complex with unfolded proteins, overexpression of Hsp70 alone should result in fewer unfolded proteins being bound by chaperones when compared to Hsp40 alone or Hsp70 overexpressed with Hsp40. This expectation is consistent with the effect of chaperones on the phase behavior of L14A barnase. We found that overexpressing chaperones suppressed droplet formation of L14A barnase (**Figure 6C**). Overexpression of Hsp70 alone had the smallest effect on suppression of droplet formation, whereas droplets were not observed when Hsp40 was overexpressed, or when Hsp70 was jointly overexpressed with Hsp40. Additionally, we found that suppression of droplet formation was more pronounced in the cytoplasm than in the nucleus. This finding is consistent with overexpressed Hsp70 and Hsp40 accumulating predominantly in the cytoplasm, as confirmed by immunostaining (**Figure 6D**).

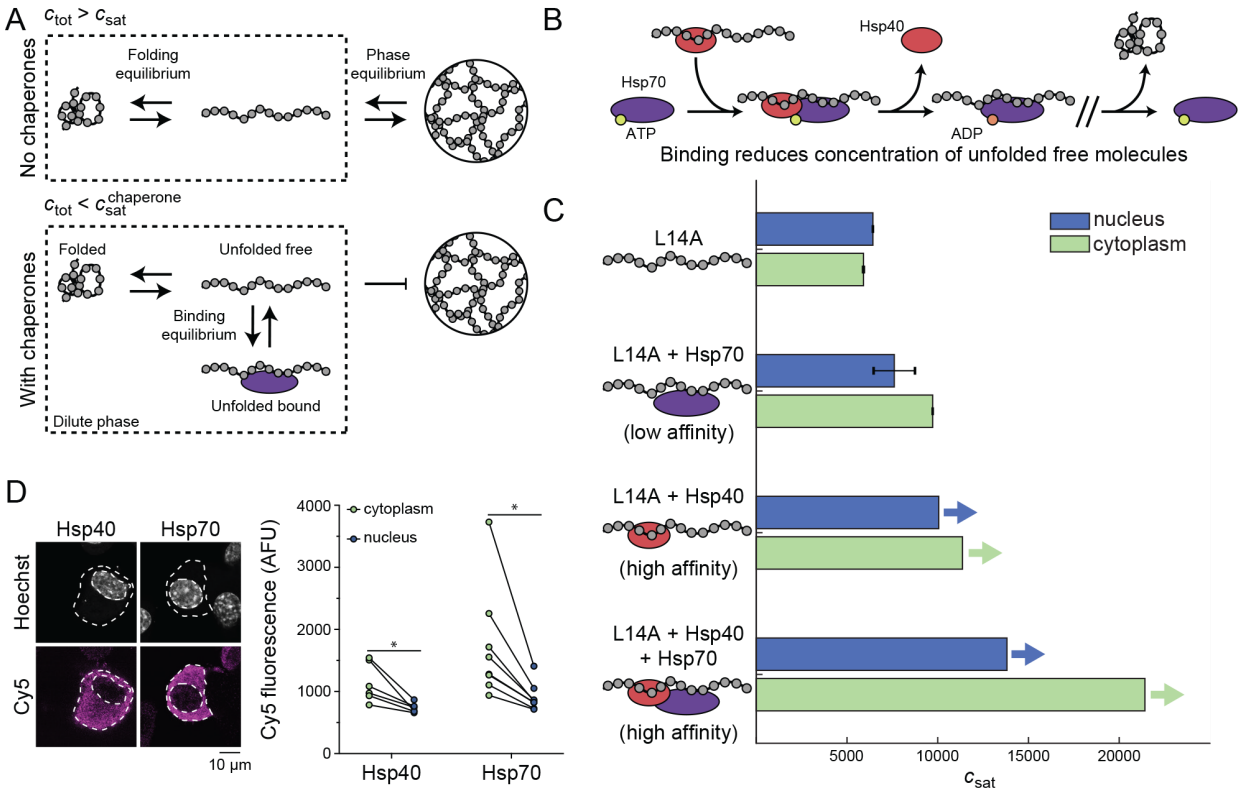

Figure 6. Molecular chaperones suppress phase separation of unfolded barnase. (A) In the absence of chaperones, the dilute phase consists of two dominant states, folded and unfolded barnase. There exists a phase equilibrium when the total concentration of barnase, c_{tot}, is greater than c_{sat} and thus phase separation occurs. In the presence of chaperones, barnase in the dilute phase consists of three dominant states: folded, unfolded free, and unfolded bound to chaperones. At the same total concentration of barnase as in the absence of chaperones, barnase cannot phase separate because c_{tot} is less than the saturation concentration needed in the presence of chaperones, $c_{sat}^{chaperone}$. This results from the fact that chaperone binding reduces the concentration of free unfolded barnase. (B) Basic model for chaperone function. Hsp40 binds the unfolded molecule and forms a ternary complex with Hsp70 in the ATP-bound state. ATP hydrolysis leads to the release of Hsp40 and the formation of a high affinity complex between Hsp70 and the unfolded substrates. (C) c_{sat} for the barnase-optoDroplet variant construct in the absence or presence of overexpressed chaperones. Bars with arrows indicate a c_{sat} value could not be extracted for these systems and must be at least above the value of the bar. Error bars denote the standard deviation from 50 bootstrapped trials. (D) Confocal images of Neuro2a cells transfected with V5-tagged DNAJB1 (Hsp40) or HSPA1A (Hsp70). Cells were stained by immunofluorescence for the V5-tag (Cy5) and the nucleus was stained with Hoechst 33342. Graphs show quantitation of immunofluorescence. Data shown as paired samples from individual cells. Paired t-test results shown; * $p<0.05$. See also **Figure S4**.

Discussion

In this work, we show that interactions among unfolded states of intrinsically foldable proteins drive intracellular phase separation leading to the formation of *de novo* deposits. Kaganovich et al., previously proposed a model for two types of cellular deposits they named the insoluble protein deposit (IPOD) and juxtanuclear quality control compartment (JUNQ) (Kaganovich et al., 2008). Following this nomenclature, we refer to the deposits that we observe as Unfolded Protein Deposits or UPODs. Interestingly, polyglutamine containing proteins were shown to form IPOD structures (Kaganovich et al., 2008), whereas other misfolded proteins, such as SOD1 destabilizing variants, formed JUNQs (Polling et al., 2014). Reversibly misfolded proteins were proposed to be delivered to the JUNQ compartment for refolding, whereas

terminally misfolded proteins go into the IPOD compartment for sequestration. Compositional profiling and details of the engagement of unfolded proteins with various components of the proteostasis machinery will be required to understand if UPODs are the same as or distinct from JUNQ compartments.

The formation of UPODs is influenced by two sequence features. Phase separation is thermodynamically favored if the protein (i) has a large enough concentration of unfolded proteins and (ii) has the requisite valence (number) and strength of stickers (**Figure 7A**). The concentration of unfolded proteins is dictated by the free energy of unfolding, ΔG°_{U}, whereas the sticker valence and strength are dictated by the composition, accessibility, and sequence contexts of hydrophobic motifs in the unfolded state. Both features jointly control the threshold concentration, c^* for phase separation.

We found that hydrophobic-aliphatic and hydrophobic-aromatic residues act as stickers to drive UPOD formation (**Figure 7B**). Interestingly, hydrophobic-aromatic residues in intrinsically disordered domains also drive the formation of distinct biomolecular condensates (Frey et al., 2006; Martin et al., 2020). However, unlike their intrinsically disordered counterparts that often lack hydrophobic-aliphatic residues, intrinsically foldable proteins can engage in aliphatic-aliphatic and aliphatic-aromatic interactions within and between their unfolded states. The prevalence of these additional interactions might explain why Tyr and Phe act as stickers of equivalent strength in intrinsically foldable proteins but not in intrinsically disordered prion-like low complexity domains (Bremer et al., 2021). The computational approach we developed to identify stickers (**Figure 5A**) is, in principle, deployable in conjunction with advances in machine learning (Russ et al., 2020) across the unfolded proteome to make quantitative predictions and identify residues that act as stickers to drive the formation of UPODs.

We showed that the driving forces for forming UPODs can be modulated by molecular chaperones. Specifically, we find that preferential binding of chaperones to unfolded proteins in the dilute phase leads to a destabilization of UPODs (**Figure 7C**). Our findings suggest that the modulation of phase separation by preferential binding of chaperones to the dilute phase represents a thermodynamic route, known as polyphasic linkage (Ruff et al., 2021a, b; Wyman and Gill, 1980), to the regulation of the concentrations of free unfolded proteins. It is worth noting that while the action of Hsp70 is likely to involve a combination of preferential binding and ATP hydrolysis, Hsp40 functions purely through preferential binding. Interestingly, overexpression of Hsp40 has a stronger effect than Hsp70 alone, and the combination of the two has the strongest destabilizing effect on phase separation.

Overall, our work shows how the triad of folding, binding, and phase equilibria might be relevant for processing of unfolded proteins in cells (**Figure 7D**). In disease contexts, mutations to the protein of interest and / or perturbations to extrinsic factors can shift the equilibria away from the folded state, thus promoting formation of UPODs. The persistent presence of UPODs in disease might point to the proteostasis machinery being overwhelmed. Additionally, there is likely a synergy between protein refolding and dissolution of UPODs (Nillegoda et al., 2015). If true, then the material properties of UPODs might influence the overall yield of refolded species vs. the generation of inclusions and amorphous aggregates (Wentink et al., 2020; Yoo et al., 2021). These and related topics are the foci for ongoing investigations.

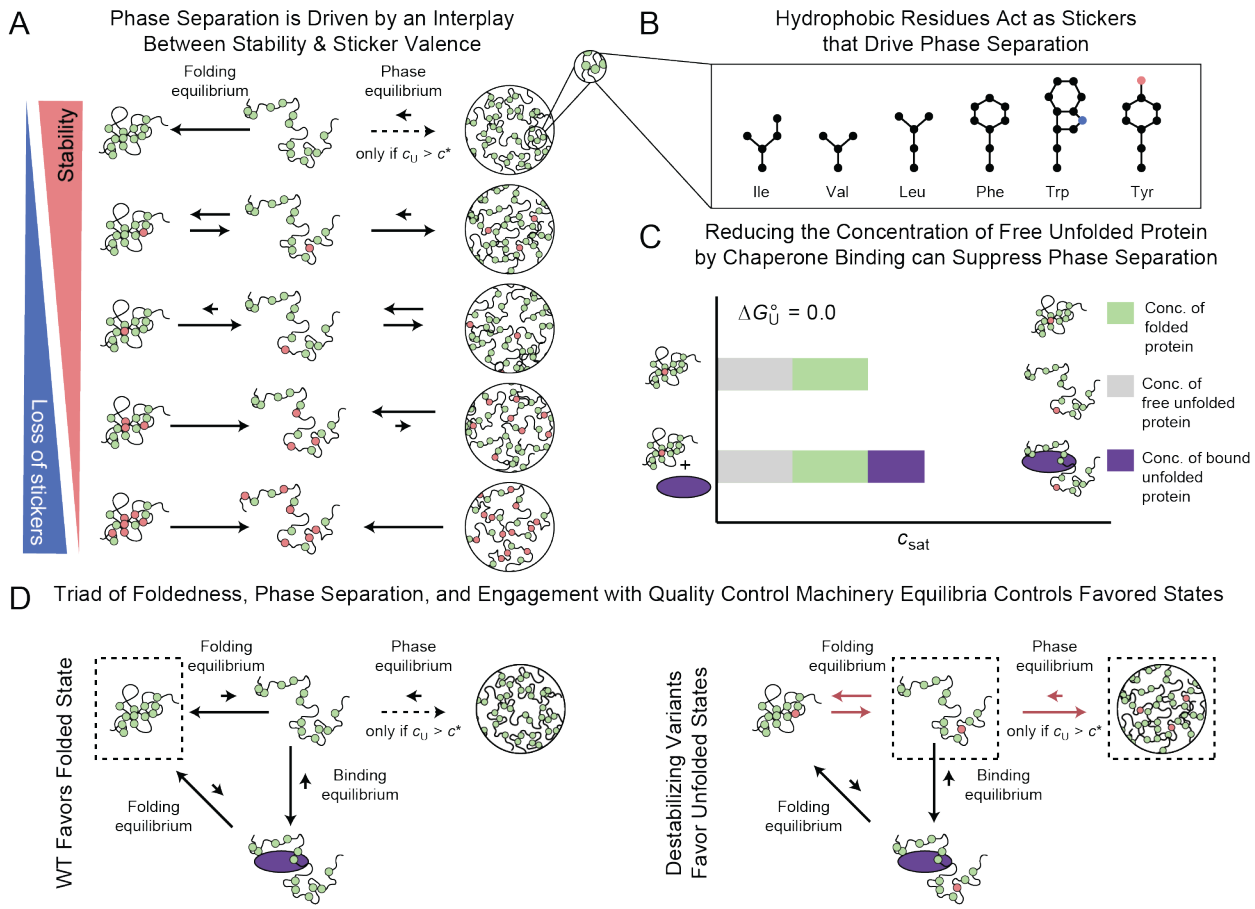

Figure 7. Schematic of features that drive and modulate phase separation of unfolded proteins. (A) The relationship between stability and sticker valence for the propensity to phase separate. Green beads correspond to stickers that drive phase separation. Mutations to stickers that destabilize the folded state are denoted as red beads. The left set of arrows corresponds to the folding equilibrium and the right set corresponds to the phase equilibrium. The latter is only relevant when the concentration of unfolded protein, c_U, is greater than c^*. Dashed arrows indicate the phase equilibrium is unlikely because c_U is unlikely to be greater than c^*. Phase separation is suppressed when the fraction of the folded state approaches one and when too many stickers are lost such that the protein is no longer sticky for itself. (B) The stickers in unfolded proteins appear to be hydrophobic residues. (C) Introducing a binding equilibrium with preferential binding to unfolded proteins in the dilute phase can suppress phase separation by proportioning molecules into a less competent phase separation driving state. (D) Schematic of the triad of foldedness, phase separation, and engagement with quality control machinery equilibria for WT and a destabilizing variant. Here, dashed boxes denote which states are most favorable due to the interplay of the triad. The purple oval denotes chaperone binding.

ACKNOWLEDGMENTS

We are grateful to our colleagues and collaborators Chloe Gerak, Mina Farag, Beatriz Ferreira-Gomes, Marta Frigole-Vivas, Paul Gooley, Alex Holehouse, Matthew King, Tanja Mittag, Ammon Posey, and Min Kyung Shinn for helpful discussions. The contributions of Y H.C., A.R.O., D.C., and D.M.H., were supported by grants APP1161803 and APP1154352 from the National Health and Medical Research Council of Australia, and DP170103093 from the Australian Research Council awarded to D.M.H. The contributions of K.M.R. and R.V.P were supported by grants from the US National Institutes of Health (5R01NS056114) and the US Air

Force Office of Scientific Research (FA9550-20-1-0241) awarded to R.V.P. S.E.R. is a Royal Society Research Professor (RSRP\R1\211057) and is funded by the Wellcome Trust (204963).

AUTHOR CONTRIBUTIONS

D.M.H. and R.V.P conceived of the project. Y H.C, D.C., and A.R.O. performed in the *in vitro* and in cell measurements. K.M.R., R.V. P., and D.M.H., designed the analysis. K.M.R. performed the simulations and computational analysis. Y.M. and D.A.B. helped generate estimates of stabilities for the barnase variants. S.E.R. and D.C. helped with critical thinking and variant design. K.M.R. designed the figures, and K.M.R., Y H.C, D.C., and A.R.O. made the figures. K.M.R., D.M.H., and R.V.P. wrote and edited the first full drafts of the manuscript. K.M.R., Y H.C., D.C., A.R.O., R.V.P., and D.M.H., collaborated extensively on the final rounds of editing and all authors were involved in revising the manuscript. R.V.P. and D.M.H. secured funding.

STAR Methods

Cell culture and imaging

Mouse Neuro2a and human HEK293T cells were used in this study. Neuro2a and HEK293T cells were maintained in opti-MEM and Dulbecco's Modified Eagle Medium (DMEM) respectively, supplemented with 10% v/v foetal bovine serum and 2 mM L-glutamine (Thermo Fisher Scientific) in a humidified incubator at 37 °C and 5% v/v atmospheric CO_2. For all imaging experiments, cells were plated at $3x10^4$ cells per well in 8-well μ-slides (Ibidi) and transfected using Lipofectamine 3000 (Thermo Fisher Scientific) as per the manufacturer's protocol. In the case of HEK293T cells, plates were pre-coated with poly-L-lysine to aid adhesion. Imaging was conducted on a Leica TCS SP5 Confocal microscope using a HCX APO CS 63 × 1.40 Oil objective lens.

For optoDroplet experiments, cells were stained 24 h post-transfection, with Hoechst 33342 at 20 μM for 20 min at 37 °C, washed and imaged in Hank's Balanced Salt Solution (HBSS). mCherry fluorescence was imaged (561 nm excitation, 600-650 nm emission) prior to optoDroplet activation, followed by photoactivation with the 488 nm laser for 60 s at a laser intensity of 30%. mCherry and Hoechst fluorescence (excitation 405 nm, emission 420-540 nm) were then imaged immediately after activation. Droplet disassembly was observed post-activation by time-lapse imaging of mCherry fluorescence every 60 s for 15 min.

For chaperone optoDroplet experiments, Neuro2A cells coexpressed either opto-barnase with HSPA1 and DNAJB1, opto-barnase with HSPA1 or DNAJB1 and emerald (Y66L) or opto-barnase with emerald (Y66L). The cells expressing three constructs were transfected at a concentration ratio of 1:1:1 and cells expressing the opto-barnase with emerald (Y66L) were transfected at a 1:2 ratio. Emerald (Y66L) was used as an inert control protein to ensure the same amount of opto-barnase DNA was being added to the cells while maintaining the recommended DNA amount for lipofectamine transfection. Imaging was carried out as described above for optoDroplet experiments.

For immunofluorescence, cells were fixed 24 h post-transfection in 4% w/v paraformaldehyde for 15 min at room temperature. Cells were then permeabilized with 0.5% v/v Triton X-100 in phosphate buffered saline (PBS) for 20 mins at room temperature. Samples were blocked in 5% w/v bovine serum albumin in PBS for 1 hour at room temperature followed by staining with anti-V5 antibody (1:250 dilution, Abcam cat# ab27671) diluted in PBS containing 1% w/v bovine serum albumin and 0.3% v/v Triton X-100 overnight at 4°C. Samples were then incubated in goat anti-mouse Cyanine5 (1:500) (Life technologies cat# A10524) diluted in PBS

for 30 mins at room temperature. Finally, cell nuclei were stained with Hoechst 33342 at 20 μM for 20 min at 37 °C. Cyanine5 fluorescence was imaged using 633 nm excitation and 695-765 nm emission and Hoechst using 405 nm excitation and 410-450 nm emission.

Constructs

The sequence for the DDX4-mCherry-Cry2 optoDroplet construct, which was based on the work of Shin et al. (Shin et al., 2017), was synthesised (Thermo Fisher Scientific) and cloned into the pTriEx4 expression vector by restriction cloning using BamHI and XhoI restriction enzymes. Barnase optoDroplet constructs were generated by PCR amplification, restriction digestion using BamHI and SacI restriction enzymes, and ligation to replace the DDX4 with barnase variants. Barnase sticker variants were synthesised (Genscript) in the pTriEx4 optoDroplet expression vector. Additional barnase variants were synthesized as cassettes (GenScript) and cloned into the pTriEx4 optoDroplet expression vector using BamHI and SacI restriction enzymes. Barnase and SOD1 were cloned into the pTriEx4 FRET vectors using the FastCloning strategy (Li et al., 2011) where the inserts and vector were PCR amplified with overlapping primers, template plasmids were digested with the methylation-sensitive restriction enzyme DpnI, and the product was directly transformed in chemically competent DH5α cells. Hsp40 and Hsp70 constructs were prepared as described previously (Ormsby et al., 2013). V5-tagged chaperone proteins were overexpressed from pcDNA5/FRT/TO V5 DNAJB1 and pcDNA5/FRT/TO V5 HSPA1A provided as gifts from Harm Kampinga (Hageman and Kampinga, 2009) via Addgene, Watertown, Massachusetts. All sequences were verified by sequencing. Httex1-72Q fused to mCherry in the pGW1 vector were prepared as previously described (Arrasate et al., 2004) and kindly provided by Steven Finkbeiner.

Image Analysis

Representative confocal micrographs including cell outlines were manually produced using FiJi (Schindelin et al., 2012). The brightness and contrast of individual images were adjusted to maximise the visible range of fluorescence intensity across constructs with different ranges of expression. Additional quantitative analyses on unmodified images were carried out using custom scripts written in the python programming language. In essence, cells and nuclei were first automatically segmented using the Cellpose package (Stringer et al., 2021), and segmentation was manually inspected for quality control using Napari (Tyson et al., 2021). Cells on the image boundary, those that did not contain a nucleus, or those that were associated with more than one nucleus, were removed from subsequent analyses. Pixel coordinates were then extracted for the individual whole-cell and nuclei segmentation masks, and the nuclei coordinates excluded from whole-cell coordinates to yield cytoplasmic pixels. For immunofluorescence experiments, compartment fluorescence was calculated as the mean intensity of pixels in the nucleus or cytoplasm respectively. Otherwise, pixel intensities for individual cells were saved as csv files for further analysis as indicated below.

Urea denaturation and stability measurements

For unfolding curve measurements, HEK293T cells were plated in poly-L-lysine-coated 12-well plates at a density of 1.3×10^5 cells per well and transfected with lipofectamine 3000 as per manufacturer's protocol. 24 h post-transfection, HEK293T cells were harvested and lysed in native lysis buffer. Stabilities of $SOD1_{barrel}$ variants were measured as described previously (Wood

et al., 2018). Fluorescence measurements were read 24 h after incubation at room temperature to equilibrate samples.

Flow cytometry

HEK293T cells were plated on a poly-L-lysine-coated 24-well plate (Falcon) at a density of 7.5×10^4 cells per well and transfected with lipofectamine 3000 as per manufacturer's protocol. Following 48 h after transfection, HEK293T cells were washed once with PBS and detached by gentle pipetting and transferred into a U-bottom microplate. Flow cytometry was performed as described previously (Wood et al., 2018). Flow cytometry data were processed with FlowJo (Tree Star Inc.) to exclude un-transfected cells and cell debris and compensate the Venus channel to remove bleed-through from the mTFP1 and FRET channels. The mTFP1, Venus and FRET data were exported as csv files for further analysis (Wood et al., 2018).

Extraction of SOD1 c_{sat} values

To determine c_{sat} values for WT and variant $SOD1_{barrel}$ constructs the $\log_{10}$ fluorescence area (A) and width (W) parameters were extracted from flow cytometry data. The area parameter can be used as an approximation for expression level, whereas the width parameter can be used as an approximation for the presence / absence of phase separation. Specifically, the transition to smaller widths as the area increases suggests that deposits have formed. To determine the expression level, A, at which phase separation occurred, cells were binned by A. The W parameters of all cells that fall into a given area bin were converted into a histogram and fit to a Gaussian distribution. For each area bin the centroid of the Gaussian was collected. To identify where the centroid stops increasing with increasing A, the slope of every four centroid values is calculated. The area bin value that is two bins higher than the first occurrence of the slope less than -0.01 was taken to be the c_{sat}.

Extraction of c_{sat} values for optoDroplet formation of barnase

From the imaging analysis results of the confocal fluorescence micrographs, the pixel intensities of all cells were converted to natural log space. Cells in which greater than 25% of the pixels have the max intensity were then removed. For the remaining cells, pixel intensity histograms were generated using the data obtained prior to activation to identify the dominant peak. Since cells should have approximately uniform intensity before activation, the histograms were fit to a Gaussian distribution to filter out pixels whose intensities were not numerically similar to the mean intensity. Specifically, the Gaussian fit was used to identify the maximum frequency of the histogram and the mean intensity. The width of the distribution was then determined by finding the first instances of 20% of the maximum frequency on either side of the mean intensity. All pixels that did not fall within the intensity bins bound by this filter histogram were removed, as well as all pixels that were already at the maximum intensity before activation. This filtering process accounts for the idea that before activation cells should have relatively uniform intensities. The positions of the filtered pixels were then used to extract the relevant pixels from the data obtained after activation. Raw intensity histograms of the before and after activation data were then created using the filtered pixels. We further removed cells in which histograms had data in less than or equal to five bins and had fewer than 100-pixel positions. These filters make sure that there is enough data for a Gaussian fit of the histograms to be reasonable. Then, the before activation histogram was fit to a single Gaussian and the mean intensity before activation ($I_{dil,before}$) was collected. This mean intensity should be proportional to the total concentration of barnase. The after-activation histogram was then fit to a model that is a mixture of two Gaussians. The

premise is that if phase separation occurs there should be low intensity and high intensity peak, where the mean of the low intensity peak ($I_{dil,after}$) is proportional to the concentration of barnase in the dilute phase and the mean of the high intensity peak ($I_{den,after}$) is proportional to the concentration of barnase in the dense phase. We restrict the fit such that the $I_{dil,after} \leq I_{dil,before}$, since the concentration of barnase in the dilute phase should not be greater than the total barnase concentration. Any cells in which $I_{dil,after}$ or $I_{dil,before}$ were less than zero or the R^2 value for the after-activation fit was less than 0.85 were removed. For the remaining cells, $I_{dil,after}$ was then collected.

Next, if there is no phase separation $I_{dil,before}$ and $I_{dil,after}$ should follow a one-to-one correspondence, i.e., $I_{dil,before} \approx I_{dil,after}$. It follows that the c_{sat} for phase separation should correspond to the intensity at which this one-to-one correspondence no longer holds. To extract the c_{sat}, we first removed outliers in the $I_{dil,before}$ versus $I_{dil,after}$ plots using three steps. Before any step was performed two threshold values were set. The value x_{1to1} = 1000 corresponds to $I_{dil,before}$ - $I_{dil,after}$ threshold for a cell to still be considered within the one-to-one regime. Then, for all cells outside of this regime, m=0.9*mean($I_{dil,before}$ - $I_{dil,after}$) was calculated. Here, m is used to distinguish between cells slightly outside the one-to-one regime or largely outside this regime. The first step to remove outliers consisted of removing cells that fell off the diagonal even though other cells in this intensity regime showed one-to-one behaviour. Next additional outliers were removed using Cook's distance (Cook, 1977). Specifically, any cells that corresponded to $I_{dil,before}$ - $I_{dil,after} < m$ were fit using a linear regression model and any cells that were 5*mean(Cook's distance for all points) were filtered out. Finally, cells well outside the one-to-one regime ($I_{dil,before}$ - $I_{dil,after} \geq m$) were fit using a linear regression model and any cells that were 5*mean(Cook's distance for all points) were filtered out.

Once outliers were removed, each barnase variant was checked for whether they had at least three off diagonal cells ($I_{dil,before}$ - $I_{dil,after} \geq m$) so a fit for c_{sat} could be performed. Barnase variants that did not satisfy this cut-off were taken as not undergoing phase separation. For the remaining barnase variants, 50 bootstrapping trials were performed with the sample number corresponding to 0.9 times the number of cells corresponding to $I_{dil,before}$ - $I_{dil,after} \geq m$. Then these cells and the cells corresponding to the one-to-one regime were systematically split into two data sets to extract the optimal fit of c_{sat}. Specifically, all cells corresponding to $I_{dil,before} \geq splitVal$ where fit using a linear regression that crossed the one-to-one line at $splitVal$. The $splitVal$ that minimized both the sum of squares due to error and 1-R^2 was taken to be the c_{sat} for phase separation.

Fitting of c_{sat} versus ΔG°_{U}

The fraction of unfolded proteins for a given ΔG of unfolding is given by:

$$p_U = \frac{e^{-\Delta G^{\circ}_U/RT}}{1 + e^{-\Delta G^{\circ}_U/RT}}$$

where ΔG°_{U} is the standard state free energy of unfolding, R is the gas constant (8.131 J/mol*K) and T is the temperature (293 K). However, even variants with ΔG°_{U} = 13000 J/mol, which equates to >99% folded molecules, undergo phase separation. This suggests barnase variants may be more unstable in cells than their calculated *in vitro* ΔG°_{U} values imply. Thus, we define the fraction of unfolded proteins for the shifted ΔG°_{U} as follows:

$$p_U = \frac{e^{-(\Delta G^{\circ}_U + \Delta G^{\circ}_S)/RT}}{1 + e^{-(\Delta G^{\circ}_U + \Delta G^{\circ}_S)/RT}}.$$

We then fit the extracted c_{sat} values assuming phase separation occurs at a critical unfolded concentration, c^* using $c^* = c_{sat} \times p_U$.

Blob diagrams

To identify hydrophobic and hydrophilic blobs in the barnase sequences we utilized the method of Lohia et al. (Lohia et al., 2019). Briefly, the average scaled Kyte-Doolittle hydrophobicity score (0 to 1) was calculated over three residue windows. Four or more contiguous windows with a score > 0.37 is considered a hydrophobic blob, whereas four or more contiguous windows with a score of ≤ 0.37 is considered a hydrophilic blob. We define the change in blobs from wild type as the sum of the magnitude of the decrease in size of hydrophobic blobs and the increase in size of hydrophilic blobs.

***c** confidence interval**

To determine the confidence interval for c^* a picking weight for each barnase variant sequence was determined by (10-(|decrease in size of hydrophobic blobs| + increase in size of hydrophilic blobs))/10, to ensure that sequences that had limited change in blobs were picked more often. Then 1000 bootstrapping trials were performed selecting 10 sequences each time based on these weights. Each trial was fit as described above to extract ΔG°_S and c^*. Then, the mean and standard deviations of these values were calculated. The interval corresponds to plotting $c_{sat} = c^*/p_U$ with $\{\text{mean}(c^*)+\text{std}(c^*), \text{mean}(\Delta G^\circ_S)-\text{std}(\Delta G^\circ_S)\}$ and $\{\text{mean}(c^*)-\text{std}(c^*), \text{mean}(\Delta G^\circ_S)+\text{std}(\Delta G^\circ_S)\}$.

Computational mutagenesis study predicting ΔG°_U

To assess the effect of mutations on the stability of barnase, the X-ray structure of barnase (PDB ID: 1A2P, resolution of 1.5 Å) was obtained from Protein Data Bank. We further processed the 3D structure by removing redundant chains, ions, water molecules and alternative conformations of residues 28, 31, 38, 85 and 96. The stability changes upon mutations, measured as the change in Gibbs Free Energy (ΔΔG in kcal/mol), were predicted using "Calculate Mutation Energy (Stability)" in Discovery Studio 2018 (https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/) with preliminary minimization of wild-type structure. The results of single and double mutations were used to build a transformation model to adjust the predicted ΔΔG of high multiple mutations (up to 8 mutations per case).

Atomistic simulations

For each barnase variant simulated, SWISS-MODEL was used to create a starting PDB structure using homology modelling (PDB ID: 1BGS or 1BSE) (Guex et al., 2009; Waterhouse et al., 2018). Atomistic simulations were performed using the ABSINTH implicit solvation model and forcefield paradigm (Vitalis and Pappu, 2009) and the CAMPARI simulation engine. Simulations were performed using a parameter set based on abs3.2_opls.prm. Each simulation was performed in a spherical droplet of radius 150 Å. Additionally, counterions and an excess of 5 mM NaCl were modelled explicitly. Starting from the PDB structures a single simulation step was performed at 20 K to add any missing atoms from the PDB file and make sure bond lengths were consistent with the ABSINTH model. Then the end PDB structures from this step were further equilibrated by running 20,000 steps at 100 K. The end PDB structures from this step were then used as the input structures for full simulations ran at 335 K. Each Metropolis Monte Carlo simulation comprised of 1×10^7 equilibration steps and 5.15×10^7 production steps. For each barnase variant, we performed five independent simulations.

Identifying stickers from atomistic simulations of unfolded states

For each atomistic simulation, only frames in which the root-mean-squared deviation from the starting PDB structure was greater than or equal to 10 were considered. These frames should represent conformations that are at least partially unfolded. For these frames, the contact order as well as the probability of native contacts per residue, Q, were calculated using the SOURSOP analysis package (https://soursop.readthedocs.io/en/latest/). Here, contact order refers to the probability a residue is in contact with another residue, excluding the nearest and second nearest neighbour contacts. For Q the cut-off for maintaining a native contact was 4.5 Å, whereas for the contact order the cut-off for a contact was 5 Å. Residues that have high non-native contact probabilities (1–Q) and high contact order are likely to be stickers. Thus, we identified residues with 1–Q > 0.5 and contact order > 0.05 as potential stickers.

REFERENCES

Alberti, S., and Hyman, A.A. (2021). Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. Nature Reviews Molecular Cell Biology *22*, 196-213.

Alberti, S., Mateju, D., Mediani, L., and Carra, S. (2017). Granulostasis: Protein Quality Control of RNP Granules. Frontiers in Molecular Neuroscience *10*.

Alderson, Thomas R., Kim, Jin H., and Markley, John L. (2016). Dynamical Structures of Hsp70 and Hsp70-Hsp40 Complexes. Structure *24*, 1014-1030.

Alshareedah, I., Kaur, T., and Banerjee, P.R. (2021). Chapter Six - Methods for characterizing the material properties of biomolecular condensates. In Methods in Enzymology, C.D. Keating, ed. (Academic Press), pp. 143-183.

Arendt, T., Bigl, V., Tennstedt, A., and Arendt, A. (1984). Correlation between cortical plaque count and neuronal loss in the nucleus basalis in Alzheimer's disease. Neuroscience Letters *48*, 81-85.

Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature *431*, 805-810.

Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting Proteostasis for Disease Intervention. Science *319*, 916.

Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2020). Recent advances in understanding catalysis of protein folding by molecular chaperones. FEBS Letters *594*, 2770-2781.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology *18*, 285-298.

Bäuerlein, F.J.B., Saha, I., Mishra, A., Kalemanov, M., Martínez-Sánchez, A., Klein, R., Dudanova, I., Hipp, M.S., Hartl, F.U., Baumeister, W.*, et al.* (2017). In Situ Architecture and Cellular Interactions of PolyQ Inclusions. Cell *171*, 179-187.e110.

Becher, M.W., Kotzuk, J.A., Sharp, A.H., Davies, S.W., Bates, G.P., Price, D.L., and Ross, C.A. (1998). Intranuclear Neuronal Inclusions in Huntington's Disease and Dentatorubral and Pallidoluysian Atrophy: Correlation between the Density of Inclusions andIT15CAG Triplet Repeat Length. Neurobiology of Disease *4*, 387-397.

Bobori, C., Theocharopoulou, G., and Vlamos, P. (2017). Molecular Chaperones in Neurodegenerative Diseases: A Short Review. (Cham, Springer International Publishing), pp. 219-231.

Brady, J.P., Farber, P.J., Sekhar, A., Lin, Y.-H., Huang, R., Bah, A., Nott, T.J., Chan, H.S., Baldwin, A.J., Forman-Kay, J.D.*, et al.* (2017). Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. Proceedings of the National Academy of Sciences *114*, E8194.

Bremer, A., Farag, M., Borcherds, W.M., Peran, I., Martin, E.W., Pappu, R.V., and Mittag, T. (2021). Deciphering how naturally occurring sequence features impact the phase behaviors of disordered prion-like domains. bioRxiv, 2021.2001.2001.425046.

Chen, X., Wu, X., Wu, H., and Zhang, M. (2020). Phase separation at the synapse. Nature Neuroscience *23*, 301-310.

Choi, J.-M., Dar, F., and Pappu, R.V. (2019). LASSI: A lattice model for simulating phase transitions of multivalent proteins. PLoS computational biology *15*.

Choi, J.-M., Holehouse, A.S., and Pappu, R.V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annual Review of Biophysics *49*, 107-133.

Chong, P.A., Vernon, R.M., and Forman-Kay, J.D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. Journal of molecular biology *430*, 4650-4665.

Ciryam, P., Kundra, R., Morimoto, R.I., Dobson, C.M., and Vendruscolo, M. (2015). Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. Trends in Pharmacological Sciences *36*, 72-77.

Clark, P.L. (2004). Protein folding in the cell: reshaping the folding funnel. Trends in Biochemical Sciences *29*, 527-534.

Conicella, A.E., Dignon, G.L., Zerze, G.H., Schmidt, H.B., D'Ordine, A.M., Kim, Y.C., Rohatgi, R., Ayala, Y.M., Mittal, J., and Fawzi, N.L. (2020). TDP-43 α-helical structure tunes liquid–liquid phase separation and function. Proceedings of the National Academy of Sciences *117*, 5883.

Cook, R.D. (1977). Detection of Influential Observation in Linear Regression. Technometrics *19*, 15-18.

Cooper, J.K., Schilling, G., Peters, M.F., Herring, W.J., Sharp, A.H., Kaminsky, Z., Masone, J., Khan, F.A., Delanoy, M., Borchelt, D.R.*, et al.* (1998). Truncated N-Terminal Fragments of

Huntingtin with Expanded Glutamine Repeats form Nuclear and Cytoplasmic Aggregates in Cell Culture. Human Molecular Genetics *7*, 783-790.

Crick, S.L., Ruff, K.M., Garai, K., Frieden, C., and Pappu, R.V. (2013). Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. Proceedings of the National Academy of Sciences *110*, 20075.

Dalby, P.A., Oliveberg, M., and Fersht, A.R. (1998). Movement of the Intermediate and Rate Determining Transition State of Barnase on the Energy Landscape with Changing Temperature. Biochemistry *37*, 4674-4679.

Danielsson, J., Kurnik, M., Lang, L., and Oliveberg, M. (2011). Cutting Off Functional Loops from Homodimeric Enzyme Superoxide Dismutase 1 (SOD1) Leaves Monomeric β-Barrels. Journal of Biological Chemistry *286*, 33070-33083.

Danielsson, J., Mu, X., Lang, L., Wang, H., Binolfi, A., Theillet, F.-X., Bekei, B., Logan, D.T., Selenko, P., Wennerström, H.*, et al.* (2015). Thermodynamics of protein destabilization in live cells. Proceedings of the National Academy of Sciences, 201511308.

Dao, T.P., Kolaitis, R.-M., Kim, H.J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P., and Castañeda, C.A. (2018). Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. Molecular Cell *69*, 965-978.e966.

Darling, A.L., and Shorter, J. (2021). Combating deleterious phase transitions in neurodegenerative disease. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1868*, 118984.

Davies, S.W., Beardsall, K., Turmaine, M., DiFiglia, M., Aronin, N., and Bates, G.P. (1998). Are neuronal intranuclear inclusions the common neuropathology of triplet-repeat disorders with polyglutamine-repeat expansions? The Lancet *351*, 131-133.

Dobson, C.M. (2004). Principles of protein folding, misfolding and aggregation. Seminars in Cell & Developmental Biology *15*, 3-16.

Farrawell, N.E., Lambert-Smith, I.A., Warraich, S.T., Blair, I.P., Saunders, D.N., Hatters, D.M., and Yerbury, J.J. (2015). Distinct partitioning of ALS associated TDP-43, FUS and SOD1 mutants into cellular inclusions. Scientific Reports *5*, 13416.

Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nüske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R.V.*, et al.* (2018). Phase separation of a yeast prion protein promotes cellular fitness. Science *359*, eaao5654.

Frey, S., Richter, R.P., and Görlich, D. (2006). FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. Science *314*, 815.

Gnutt, D., Timr, S., Ahlers, J., König, B., Manderfeld, E., Heyden, M., Sterpone, F., and Ebbinghaus, S. (2019). Stability Effect of Quinary Interactions Reversed by Single Point Mutations. Journal of the American Chemical Society *141*, 4660-4669.

Gomez, M., and Germain, D. (2019). Cross talk between SOD1 and the mitochondrial UPR in cancer and neurodegeneration. Molecular and Cellular Neuroscience *98*, 12-18.

Greig, J.A., Nguyen, T.A., Lee, M., Holehouse, A.S., Posey, A.E., Pappu, R.V., and Jedd, G. (2020). Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. Molecular Cell *77*, 1237-1250.e1234.

Guex, N., Peitsch, M.C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. ELECTROPHORESIS *30*, S162-S173.

Hageman, J., and Kampinga, H.H. (2009). Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. Cell Stress and Chaperones *14*, 1-21.

Harmon, T.S., Holehouse, A.S., and Pappu, R.V. (2018). Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. New Journal of Physics *20*, 045002.

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.

Hartl, F.U. (2016). Cellular Homeostasis and Aging. Annual Review of Biochemistry *85*, 1-4.

Hofweber, M., and Dormann, D. (2019). Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics. Journal of Biological Chemistry *294*, 7137-7150.

Hsu, D.S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R.-P., Todo, T., Wei, Y.-F., and Sancar, A. (1996). Putative Human Blue-Light Photoreceptors hCRY1 and hCRY2 Are Flavoproteins. Biochemistry *35*, 13871-13877.

Hwang, Y.-M., Stathopulos, P.B., Dimmick, K., Yang, H., Badiei, H.R., Tong, M.S., Rumfeldt, J.A.O., Chen, P., Karanassios, V., and Meiering, E.M. (2010). Nonamyloid Aggregates Arising from Mature Copper/Zinc Superoxide Dismutases Resemble Those Observed in Amyotrophic Lateral Sclerosis*. Journal of Biological Chemistry *285*, 41701-41711.

Iadanza, M.G., Jackson, M.P., Hewitt, E.W., Ranson, N.A., and Radford, S.E. (2018). A new era for understanding amyloid structures and disease. Nature Reviews Molecular Cell Biology *19*, 755-773.

Jacobs, R.W., Farivar, N., and Butcher, L.L. (1985). Plaque-like lesions in the basal forebrain in Alzheimer's disease. Neuroscience Letters *56*, 347-351.

Jiang, Y., Rossi, P., and Kalodimos, C.G. (2019). Structural basis for client recognition and activity of Hsp40 chaperones. Science *365*, 1313.

Kaganovich, D., Kopito, R., and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. Nature *454*, 1088-1095.

Kaur, T., Alshareedah, I., Wang, W., Ngo, J., Moosa, M.M., and Banerjee, P.R. (2019). Molecular Crowding Tunes Material States of Ribonucleoprotein Condensates. Biomolecules *9*, 71.

Kidd, M., Allsop, D., and Landon, M. (1985). SENILE PLAQUE AMYLOID, PAIRED HELICAL FILAMENTS, AND CEREBROVASCULAR AMYLOID IN ALZHEIMER'S DISEASE ARE ALL DEPOSITS OF THE SAME PROTEIN. The Lancet *325*, 278.

Kim, T.H., Tsang, B., Vernon, R.M., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2019). Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science *365*, 825.

Kroschwald, S., Munder, M.C., Maharana, S., Franzmann, T.M., Richter, D., Ruer, M., Hyman, A.A., and Alberti, S. (2018). Different Material States of Pub1 Condensates Define Distinct Modes of Stress Adaptation and Recovery. Cell Reports *23*, 3327-3339.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of molecular biology *157*, 105-132.

Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., and Chang, Y. (2011). FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnology *11*, 92.

Lin, C., Yang, H., Guo, H., Mockler, T., Chen, J., and Cashmore, A.R. (1998). Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2. Proceedings of the National Academy of Sciences *95*, 2686.

Lin, Y., Currie, S.L., and Rosen, M.K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. Journal of Biological Chemistry *292*, 19110-19120.

Lohia, R., Salari, R., and Brannigan, G. (2019). Sequence specificity despite intrinsic disorder: How a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein. PLOS Computational Biology *15*, e1007390.

Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science *367*, 694-699.

Martin, E.W., and Mittag, T. (2018). Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. Biochemistry *57*, 2478-2487.

Masliah, E., Rockenstein, E., Veinbergs, I., Mallory, M., Hashimoto, M., Takeda, A., Sagara, Y., Sisk, A., and Mucke, L. (2000). Dopaminergic Loss and Inclusion Body Formation in α-Synuclein Mice: Implications for Neurodegenerative Disorders. Science *287*, 1265.

Mathieu, C., Pappu, R.V., and Taylor, J.P. (2020). Beyond aggregation: Pathological phase transitions in neurodegenerative disease. Science *370*, 56.

Matthews, J.M., and Fersht, A.R. (1995). Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: Observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. Biochemistry *34*, 6805-6814.

McCray, B.A., and Taylor, J.P. (2008). The Role of Autophagy in Age-Related Neurodegeneration. Neurosignals *16*, 75-84.

McMillan, P.F., Clary, D.C., Vendruscolo, M., and Dobson, C.M. (2005). Towards complete descriptions of the free energy landscapes of proteins. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences *363*, 433-452.

Meiering, E.M. (2008). The Threat of Instability: Neurodegeneration Predicted by Protein Destabilization and Aggregation Propensity. PLOS Biology *6*, e193.

Mitrea, D.M., and Kriwacki, R.W. (2016). Phase separation in biology; functional organization of a higher order. Cell Communication and Signaling *14*, 1.

Nedelsky, N.B., and Taylor, J.P. (2019). Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. Nature Reviews Neurology *15*, 272-286.

Newcombe, E.A., Ruff, K.M., Sethi, A., Ormsby, A.R., Ramdzan, Y.M., Fox, A., Purcell, A.W., Gooley, P.R., Pappu, R.V., and Hatters, D.M. (2018). Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity that Persists regardless of Polyglutamine Length. Journal of molecular biology *430*, 1442-1458.

Nillegoda, N.B., Kirstein, J., Szlachcic, A., Berynskyy, M., Stank, A., Stengel, F., Arnsburg, K., Gao, X., Scior, A., Aebersold, R.*, et al.* (2015). Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation. Nature *524*, 247-251.

Nordlund, A., Leinartaitė, L., Saraboji, K., Aisenbrey, C., Gröbner, G., Zetterström, P., Danielsson, J., Logan, D.T., and Oliveberg, M. (2009). Functional features cause misfolding of the ALS-provoking enzyme SOD1. Proceedings of the National Academy of Sciences *106*, 9667.

Nott, Timothy J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, David P., Pawson, T., Forman-Kay, Julie D.*, et al.* (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Molecular Cell *57*, 936-947.

Ormsby, A.R., Ramdzan, Y.M., Mok, Y.-F., Jovanoski, K.D., and Hatters, D.M. (2013). A Platform to View Huntingtin Exon 1 Aggregation Flux in the Cell Reveals Divergent Influences from Chaperones hsp40 and hsp70*. Journal of Biological Chemistry *288*, 37192-37203.

Pak, Chi W., Kosno, M., Holehouse, Alex S., Padrick, Shae B., Mittal, A., Ali, R., Yunus, Ali A., Liu, David R., Pappu, Rohit V., and Rosen, Michael K. (2016). Sequence Determinants

of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Molecular Cell *63*, 72-85.

Pappu, R.V., Wang, X., Vitalis, A., and Crick, S.L. (2008). A polymer physics perspective on driving forces and mechanisms for protein aggregation. Archives of Biochemistry and Biophysics *469*, 132-141.

Perl, D.P., and Brody, A.R. (1980). Alzheimer's disease: X-ray spectrometric evidence of aluminum accumulation in neurofibrillary tangle-bearing neurons. Science *208*, 297.

Polling, S., Mok, Y.-F., Ramdzan, Y.M., Turner, B.J., Yerbury, J.J., Hill, A.F., and Hatters, D.M. (2014). Misfolded Polyglutamine, Polyalanine, and Superoxide Dismutase 1 Aggregate via Distinct Pathways in the Cell*. Journal of Biological Chemistry *289*, 6669-6680.

Posey, A.E., Holehouse, A.S., and Pappu, R.V. (2018a). Chapter One - Phase Separation of Intrinsically Disordered Proteins. In Methods in Enzymology, E. Rhoades, ed. (Academic Press), pp. 1-30.

Posey, A.E., Ruff, K.M., Harmon, T.S., Crick, S.L., Li, A., Diamond, M.I., and Pappu, R.V. (2018b). Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers. Journal of Biological Chemistry *293*, 3734-3746.

Powers, E.T., Morimoto, R.I., Dillin, A., Kelly, J.W., and Balch, W.E. (2009). Biological and Chemical Approaches to Diseases of Proteostasis Deficiency. Annual Review of Biochemistry *78*, 959-991.

Ramdzan, Y.M., Polling, S., Chia, C.P.Z., Ng, I.H.W., Ormsby, A.R., Croft, N.P., Purcell, A.W., Bogoyevitch, M.A., Ng, D.C.H., Gleeson, P.A.*, et al.* (2012). Tracking protein aggregation and mislocalization in cells with flow cytometry. Nature Methods *9*, 467-470.

Reinle, K., Mogk, A., and Bukau, B. (2021). The Diverse Functions of Small Heat Shock Proteins in the Proteostasis Network. Journal of molecular biology, 167157.

Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E.V., Rojek, A.E., Sosnick, T.R., and Drummond, D.A. (2017). Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. Cell *168*, 1028-1040.e1019.

Ruff, K.M., Dar, F., and Pappu, R.V. (2021a). Ligand effects on phase separation of multivalent macromolecules. Proceedings of the National Academy of Sciences *118*, e2017184118.

Ruff, K.M., Dar, F., and Pappu, R.V. (2021b). Polyphasic linkage and the impact of ligand binding on the regulation of biomolecular condensates. Biophysics Reviews *2*, 021302.

Ruff, K.M., Harmon, T.S., and Pappu, R.V. (2015). CAMELOT: A machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences. The Journal of Chemical Physics *143*, 243123.

Russ, W.P., Figliuzzi, M., Stocker, C., Barrat-Charlaix, P., Socolich, M., Kast, P., Hilvert, D., Monasson, R., Cocco, S., Weigt, M., *et al.* (2020). An evolution-based model for designing chorismate mutase enzymes. Science *369*, 440.

Ryno, L.M., Wiseman, R.L., and Kelly, J.W. (2013). Targeting unfolded protein response signaling pathways to ameliorate protein misfolding diseases. Current Opinion in Chemical Biology *17*, 346-352.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676-682.

Shen, Y., Ruggeri, F.S., Vigolo, D., Kamada, A., Qamar, S., Levin, A., Iserman, C., Alberti, S., George-Hyslop, P.S., and Knowles, T.P.J. (2020). Biomolecular condensates undergo a generic shear-mediated liquid-to-solid transition. Nature Nanotechnology *15*, 841-847.

Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E., and Brangwynne, C.P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. Cell *168*, 159-171.e114.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*, eaaf4382.

Sinnige, T., Yu, A., and Morimoto, R.I. (2020). Challenging Proteostasis: Role of the Chaperone Network to Control Aggregation-Prone Proteins in Human Disease. In HSF1 and Molecular Chaperones in Biology and Cancer, M.L. Mendillo, D. Pincus, and R. Scherz-Shouval, eds. (Cham: Springer International Publishing), pp. 53-68.

Solomon, J.P., Page, L.J., Balch, W.E., and Kelly, J.W. (2012). Gelsolin amyloidosis: genetics, biochemistry, pathology and possible strategies for therapeutic intervention. Critical Reviews in Biochemistry and Molecular Biology *47*, 282-296.

Song, J. (2018). Environment-transformable sequence–structure relationship: a general mechanism for proteotoxicity. Biophysical Reviews *10*, 503-516.

Sontag, E.M., Samant, R.S., and Frydman, J. (2017). Mechanisms and Functions of Spatial Protein Quality Control. Annual Review of Biochemistry *86*, 97-122.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nature Methods *18*, 100-106.

Tyson, A.L., Rousseau, C.V., Niedworok, C.J., Keshavarzi, S., Tsitoura, C., Cossell, L., Strom, M., and Margrie, T.W. (2021). A deep learning algorithm for 3D cell detection in whole mouse brain image datasets. PLOS Computational Biology *17*, e1009074.

Vendruscolo, M., Knowles, T.P.J., and Dobson, C.M. (2011). Protein Solubility and Protein Homeostasis: A Generic View of Protein Misfolding Disorders. Cold Spring Harbor Perspectives in Biology *3*.

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

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

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D.*, et al.* (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell *174*, 688-699. e616.

Warner, J.B., Ruff, K.M., Tan, P.S., Lemke, E.A., Pappu, R.V., and Lashuel, H.A. (2017). Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. Journal of the American Chemical Society *139*, 14456-14469.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A P., Rempfer, C., Bordoli, L.*, et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research *46*, W296-W303.

Wentink, A.S., Nillegoda, N.B., Feufel, J., Ubartaitė, G., Schneider, C.P., De Los Rios, P., Hennig, J., Barducci, A., and Bukau, B. (2020). Molecular dissection of amyloid disaggregation by human HSP70. Nature *587*, 483-488.

Wood, J.D., Beaujeux, T.P., and Shaw, P.J. (2003). Protein aggregation in motor neurone disorders. Neuropathology and Applied Neurobiology *29*, 529-545.

Wood, R.J., Ormsby, A.R., Radwan, M., Cox, D., Sharma, A., Vöpel, T., Ebbinghaus, S., Oliveberg, M., Reid, G.E., Dickson, A.*, et al.* (2018). A biosensor-based framework to measure latent proteostasis capacity. Nature Communications *9*, 287.

Wyman, J., and Gill, S.J. (1980). Ligand-linked phase changes in a biological system: applications to sickle cell hemoglobin. Proceedings of the National Academy of Sciences *77*, 5239.

Yoo, H., Bard, J.A.M., Pilipenko, E., and Drummond, D.A. (2021). Chaperones directly and efficiently disperse stress-triggered biomolecular condensates. bioRxiv, 2021.2005.2013.444070.

Yu, H., Lu, S., Gasior, K., Singh, D., Vazquez-Sanchez, S., Tapia, O., Toprani, D., Beccari, M.S., Yates, J.R., Da Cruz, S.*, et al.* (2021). HSP70 chaperones RNA-free TDP-43 into anisotropic intranuclear liquid spherical shells. Science *371*, eabb4309.

Zbinden, A., Pérez-Berlanga, M., De Rossi, P., and Polymenidou, M. (2020). Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. Developmental Cell *55*, 45-68.

Zeng, X., Holehouse, A.S., Chilkoti, A., Mittag, T., and Pappu, R.V. (2020). Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins. Biophysical Journal *119*, 402-418.