

Mapping interactions between disordered regions reveals promiscuity in biomolecular condensate formation

Atar Gilat¹, Benjamin Dubrueil¹, Emmanuel D. Levy^{1*}

Author Affiliations

¹ Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel

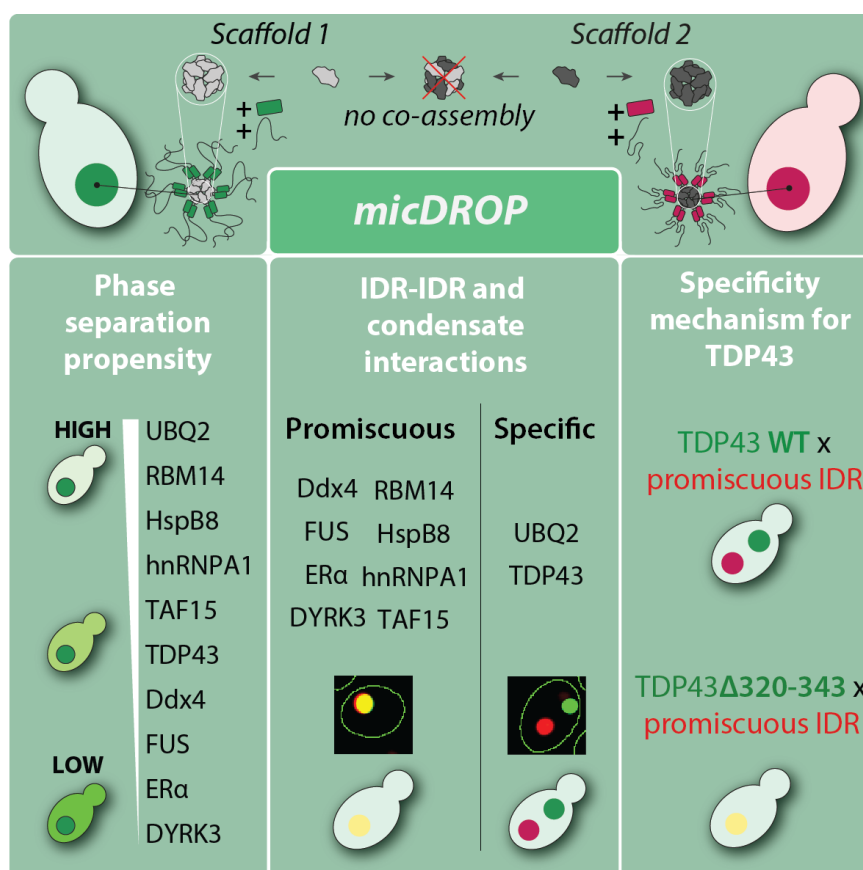
ORCID

E.D.L: 0000-0001-8959-7365

*Correspondence: emmanuel.levy@weizmann.ac.il

ABSTRACT

Intrinsically-disordered regions (IDRs) promote intracellular phase separation and the formation of biomolecular condensates through interactions encoded in their primary sequence. While these condensates form spatially distinct assemblies in cells, it is unclear whether such specificity can be conferred by IDRs alone. Indeed, IDRs exhibit high conformational flexibility whereas specificity in protein recognition is generally associated with well-defined 3D structures. To characterize IDR-IDR interactions and assess their ability to mediate self-specific partitioning, we developed a synthetic system of Multivalent IDRs forming Constitutive DROplets (*micDROP*). We investigated ten natural IDRs that underwent phase separation in *micDROP*. These IDRs exhibited a wide range of saturation concentrations *in vivo*, which correlated well with their total sequence stickiness. We then probed IDR-IDR specificity by co-expressing pairs of IDRs fused to homologous scaffolds that did not co-assemble. We observed a high degree of promiscuity, particularly among IDRs from the proteins Ddx4, DYRK3, ER α , FUS, hnRNPA1, HspB8, RBM14 and TAF15, whereas TDP43 and UBQ2 formed spatially distinct condensates regardless of their partner. Further investigation revealed the short and conserved α -helical segment of TDP43's IDR was governing its specific self-recognition. Our findings imply that IDRs can tune their phase separation propensity through sequence composition, while their formation of discrete condensates likely requires additional cellular or structural determinants.



INTRODUCTION

Intrinsically disordered regions (IDRs) represent a unique class of protein segments that lack a well-defined tertiary structure under physiological conditions. These regions possess inherent flexibility, enabling them to sample a variety of conformations and engage in diverse interactions¹. IDRs are prevalent in eukaryotic proteomes and play crucial roles in a wide range of biological processes².

One of the key emerging functions of IDRs is in mediating the formation of biomolecular condensates. Condensates are membraneless compartments within cells that facilitate the spatial organization of biomolecules and regulate a variety of cellular processes related to stress responses³, cellular mitosis⁴, DNA replication and repair⁵, and more⁶⁻⁹. Condensates may form through phase separation, whereby IDRs self-interact and demix from the surrounding bulk (cytoplasm/nucleoplasm), and the resulting assemblies often display liquid-like behavior⁹⁻¹¹. Experimental works conducted over the past decade have consistently shown that, for many protein components of biomolecular condensates, the IDR sequence alone can undergo phase separation through homotypic interactions both *in vitro* and *in vivo*¹²⁻¹⁴.

Each type of condensate has a unique milieu of components¹⁵⁻¹⁷. However, proteins such as FUS (FUsed in Sarcoma) have been identified in functionally and spatially distinct condensates including stress granules¹⁸, nucleoli¹⁹ and paraspeckles²⁰. The ability to partition proteins into discrete condensates without their intermixing implies some measure of specificity in the interactions driving their formation. Whether such specificity can be conferred through the IDR sequence alone, or whether additional cellular factors are required, remains unclear.

In the absence of a stable structure, the amino acid composition of an IDR sequence determines its ability to phase-separate, particularly through aromatic or charge-driven interactions²¹⁻²⁸. Unlike protein folding, in which the sequence itself is critical, overall compositional characteristics can explain an IDR's phase separation behavior. For example, the overall number of aromatic residues and their spacing modulate hnRNPA1's propensity to phase-separate²⁹. This notion is in line with the fact that only ~5% of amino acid positions in IDR sequences are conserved across orthologs³⁰, while other molecular features, such as sequence length or net charge, are under evolutionary constraint. At the same time, it is unclear whether amino acid composition alone can confer specificity in IDR-IDR interactions because interaction energies between amino acids show that no pair is exclusively favorable³¹.

To further our understanding of possible IDR-IDR specificity in intracellular phase separation, we present *micDROP* (Multivalent IDRs forming Constitutive DROplets), a synthetic system wherein two non-interacting dodecameric scaffolds present different IDRs and are co-expressed in yeast cells. Using confocal fluorescence microscopy, we quantitatively assessed the propensity of nineteen candidate IDRs to undergo phase separation *in vivo*. Our findings revealed that IDR sequences underwent phase separation with a wide range of apparent affinities, as inferred from their relative saturation concentration (C_{sat}). Despite displaying a wide range of C_{sat} and sequence characteristics, most IDRs interacted with high promiscuity and colocalized indiscriminately. Two IDRs, UBQ2 and TDP43, formed self-specific condensates. Examining TDP43's IDR revealed that a transient α -helical segment governed this self-specificity. Indeed, deletion of this short segment did not impact TDP43's ability to phase

separate, but abolished its specific self-recognition. Overall, *micDROP* provides a novel method for the quantitative study of phase separation *in vivo* and may elucidate additional features important both for phase separation propensity and specificity.

RESULTS

micDROP: A synthetic system for studying IDR-mediated phase separation *in vivo*

The genetic construct of *micDROP* consists of three domains connected by flexible linkers: a scaffold subunit, a fluorescent reporter, and an IDR of interest (**Figure 1A**). This design bears similarities to *Corelets*, an optogenetics-based system previously introduced by Bracha et al.³², with two significant distinctions: *micDROP* relies on constitutive oligomerization, and utilizes a different scaffold. Indeed, in our initial experiments, the 24-mer ‘maxi’ ferritins tested *in vivo* were prone to aggregation (**Figure S1**), prompting us to select 12-mer ‘mini’ ferritins³³, hereafter referred to as Dps (DNA Protection during Starvation) scaffolds. The resulting fusion protein (**Figure 1B**) forms a dodecameric complex (**Figure 1C**), thus increasing the effective valency of the IDR. This fusion strategy was selected because many IDRs, such as NPM1³⁴, FUS³⁵, and TDP43³⁶ exist in homo-oligomeric proteins. Moreover, a higher valence is known to enhance phase separation^{37–39}, which we expected would enable tracking the phase separation behavior of IDRs with weak propensities for homotypic interactions. *micDROP*-green, consisting of a Dps scaffold from *V. cholerae* (PDB code 3IQ1) fused to a yellow fluorescent protein (Venus⁴⁰), was first tested with the disordered region of hnRNPA1, an RNA-binding protein (RBP) whose phase separation propensity is well-documented^{22,29}. Both the monomeric hnRNPA1 IDR as well as the IDR-free *micDROP*-green were homogeneously dispersed when expressed in yeast cells. However, hnRNPA1 fused to *micDROP*-green formed bright puncta, indicating the formation of synthetic condensates (**Figure 1D**).

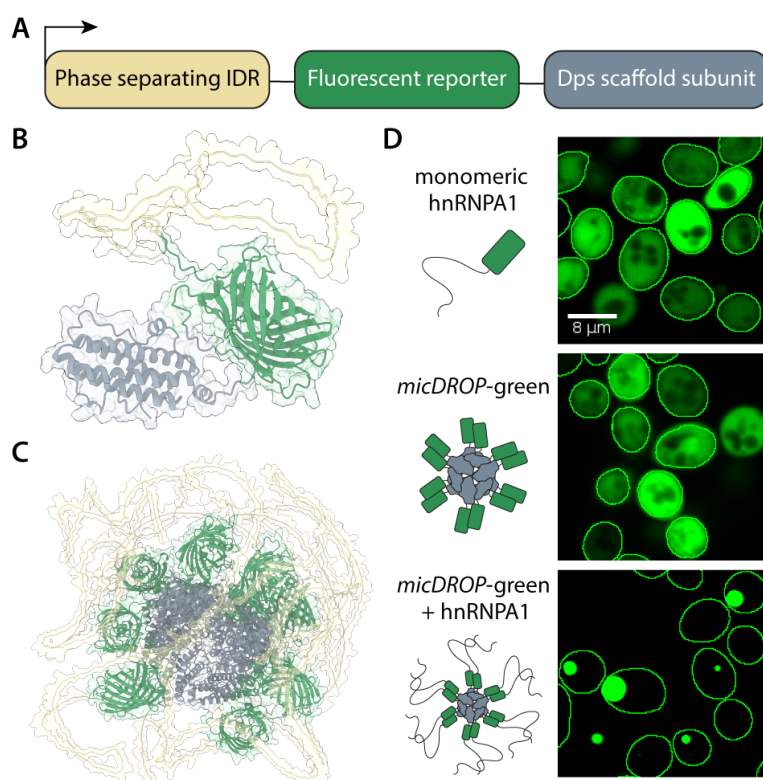


Figure 1. *micDROP*: A synthetic system for studying IDR-mediated phase separation *in vivo*.

A. Schematic representation of *micDROP*'s genetic construct. It comprises three domains connected by flexible linkers: an IDR of interest (yellow), a fluorescent reporter (green), and a Dps subunit as a homo-oligomeric scaffold (gray), which assembles into a dodecamer. **B.** Cartoon rendering of a single fusion protein harboring an IDR region. The structure model was generated with AlphaFold²⁴¹. **C.** Cartoon rendering of an assembled complex model generated with AlphaFold2. Two fluorescent proteins obstructing the view of the scaffold, as well as two subunits in the back of the structure, were hidden for clarity. **D.** Fluorescence microscopy images of yeast cells expressing three constructs: a monomeric construct encoding an IDR region from hnRNPA1 fused to Venus (top), *micDROP*-green - an assembled Dps complex from *V. cholerae* fused to Venus without an IDR region (center), and *micDROP*-green fused to the IDR region from hnRNPA1 (bottom).

***micDROP* captures IDR propensity for phase separation**

Phase separation is characterized by a phase boundary, which reflects the set of conditions at which demixing occurs (**Figure 2A**). In living cells, where temperature and pressure remain relatively constant, phase separation can be modulated by intracellular conditions such as crowding or pH, or component-related parameters, including multivalency, affinity, and concentration⁴². Because *micDROP* has a fixed valency and is expressed in similar conditions, the primary parameters modulating the phase boundary ought to be the intracellular concentration of the system and the homotypic affinity of the IDR. We sought to identify whether a phase boundary existed - i.e., a saturation concentration (C_{sat}) beyond which *micDROP* would phase separate into a dense phase, visible as a bright puncta, and a dilute phase, observable as homogeneous fluorescence in the cytosol. If a C_{sat} existed for an IDR in *micDROP*, we hypothesized its value should reflect the apparent affinity of homotypic interactions for that IDR.

The first criterion to determine IDR-mediated phase separation in *micDROP* was the presence of condensates in cells, which were identified and segmented using custom scripts for image analysis⁴³ (**Methods**). By plotting the median fluorescence intensity against the maximal intensity for each segmented cell, we predicted that cells without a condensate would spread along an approximately linear diagonal and display a constant ratio between median and maximal fluorescence intensity (**Figure 2B**). By contrast, cells with condensates were expected to display constant median and maximal fluorescence intensity values. Indeed, in accordance with phase separation theory, the dense and dilute phases should exhibit constant concentrations; an increase in expression of *micDROP* beyond the phase boundary should therefore increase the dense phase volume, but not its concentration. However, due to limitations of fluorescence microscopy, particularly in resolution⁴⁴, an accurate assessment of the fluorescence intensity within the dense phase was not possible, resulting in a wide range of maximal intensities that reflect condensate size (**Figure 2C**). As measuring the dilute phase (the cytoplasm) was not similarly limited by resolution, we reasoned that the measured median intensity at which condensates began appearing reflected C_{sat} . We expected that such inferred C_{sat} should be relatively constant, which we assessed from the slope of a linear regression of C_{sat} values among cells containing condensates. We classified an IDR as undergoing phase separation in *micDROP* when at least 20% of the cells were identified as containing a condensate ($p \geq 20\%$), and when the slope (s) of the linear regression fitted for those cells was ≤ 0.4 , indicating that beyond the phase boundary, the concentration of the dilute phase remained relatively constant.

To evaluate our ability to determine relative affinity from C_{sat} measurements, we examined how mutations within an IDR sequence affected its propensity to phase-separate. Previous work showed that radii of gyration of monomeric hnRNPA1 IDR and its critical temperature for phase separation could be modulated by the addition or removal of interaction 'stickers' in the form of aromatic residues (Tyr/Phe)²⁹. We used *micDROP*-green to express three hnRNPA1 mutants previously tested *in vitro*: Aro⁺, with five Ser→Phe mutations, Aro⁻, with six Phe→Ser mutations, and Aro^{-/-}, with thirteen Phe→Ser mutations (**Figure 2D**). hnRNPA1 Aro⁺ and hnRNPA1 WT showed a comparable propensity to undergo phase

separation, with a similar percentage of cells identified with condensates and C_{sat} values. While the calculated C_{sat} for Aro⁺ was slightly lower, the difference did not match previous *in vitro* results, in which a ~10-fold decrease in critical concentration was observed at 25°C. However, as C_{sat} values were close to autofluorescence levels (~20 A.U.), it is probable that the measurement of the Aro⁺ mutant C_{sat} was overestimated. By contrast, hnRNPA1 Aro⁻ showed a ~1.6-fold decrease in the percentage of cells containing condensates and a ~20-fold increase in C_{sat} compared to hnRNPA1 WT, while hnRNPA1 Aro^{-/-} did not undergo phase separation and no condensates were observed even at high expression levels ($p < 1\%$). These *in vivo* experiments recapitulated observations previously made *in vitro*, where Aro⁻ displayed a ~15-fold increase in C_{sat} at 25°C, and Aro^{-/-} was not observed to undergo phase separation at any concentration at 25°C. From these experiments, we concluded that *micDROP* captures differences in apparent homotypic affinities of IDR interactions that drive phase separation.

IDRs display a wide range of phase separation propensities that correlate with their overall stickiness

The validation of *micDROP* on hnRNPA1 mutants motivated us to carry out a broader analysis across various IDR families. We selected eighteen IDR candidates (**Table 1**) referenced in the PhaSePro database⁴⁵ and previously shown to undergo phase separation through homotypic interactions either *in vivo* or *in vitro*. In addition, we included as a negative control the disordered region of G3BP1, a scaffold protein in the assembly of stress granules⁴⁶ that binds RNA. While the C-terminal IDR of G3BP1 was found to be essential for phase separation of the full-length protein, it was unable to drive it without the nearby RNA-recognition motif (RRM) binding to RNA⁴⁷. As such, the G3BP1 IDR was not expected to form condensates in *micDROP*.

Importantly, the candidates covered a large space of physicochemical properties, with various sequence lengths, amino acid preferences, stickiness, hydrophobicity, and net charge, all of which may impact phase separation propensity and IDR-IDR interaction specificity. We compared the sequence diversity of the selected sequences to that of all IDRs across the human proteome (**Methods**). We visualized the diversity of the IDR properties on a two-dimensional space using a uniform manifold approximation and projection (UMAP)⁴⁸ (**Figure S2**). The selected IDRs were scattered across the map, highlighting that they cover a wide range of sequence characteristics on the IDR spectrum.

The nineteen IDR candidates were initially tested as monomeric chains fused to Venus and were expressed *in vivo*. However, only RBM14 and NUP98 formed condensates, with 12.5% and 25.2% of the cells identified with condensates, respectively (**Figure S3**). Only NUP98 fulfilled our criteria for determining phase separation as a monomeric construct ($p=25.2\%$, $s=0.07$).

To ascertain the phase separation behavior was not heavily influenced by the fluorescent reporter or scaffold, we tested the candidate IDRs using two constructs: *micDROP*-green, and *micDROP*-red, which consists of the same overall architecture, but employs a different Dps-like scaffold from *S. solfataricus* (PDB code 2CLB⁵⁰) and a red fluorescent protein (mScarlet⁵¹, **Methods**).

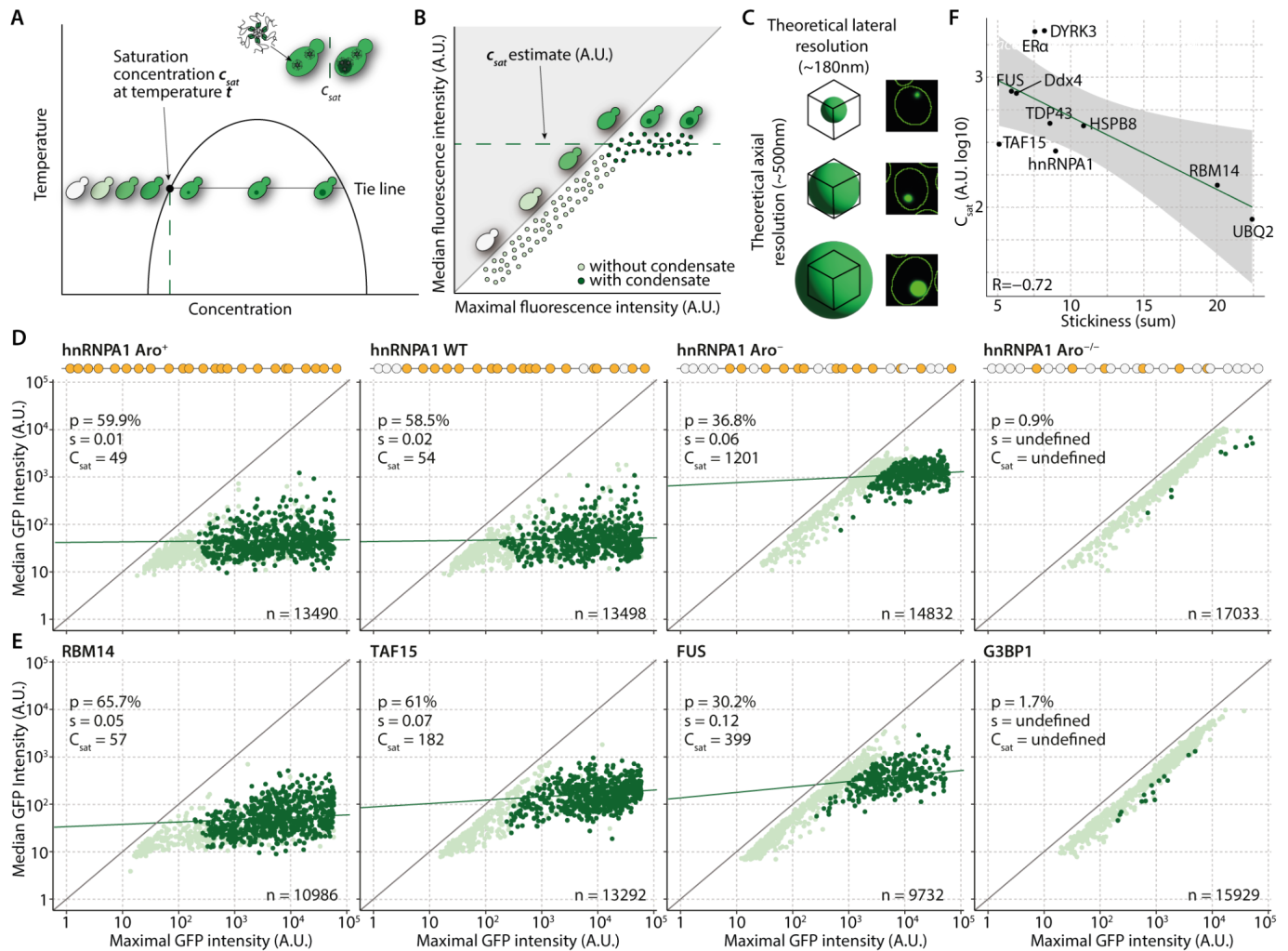


Figure 2. *In vivo* quantification of C_{sat} for model IDRs. **A**. Schematic representation of a phase diagram. At constant temperature (t), increasing the concentration to the point of saturation (C_{sat}) will result in phase separation into a condensate (visible as a bright puncta) and a dilute phase. Further increase of concentration will increase condensate volume without increasing its concentration. **B**. Schematic representation of a method to estimate C_{sat} from the fluorescence intensity of the dilute phase (cytoplasm) among cells with a detectable dense phase (condensate or puncta). Each segmented cell's median fluorescence intensity is recorded (y-axis) and reflects the dilute phase concentration. This quantity is analyzed as a function of the maximal fluorescence intensity in the same cell. Cells without a condensate (light green) show maximal intensities equal to ~ 1.5 -fold their cytosolic median intensity. By contrast, cells with condensates (dark green) show maximal fluorescence intensities 3- to 100-fold larger than the median. The dilute phase concentration among these condensate-containing cells reflects the C_{sat} . **C**. The size of condensates varies while the confocal volume is fixed, making it difficult to measure protein concentration in the dense phase: small condensates do not fill up the entire volume and their fluorescence intensity is underestimated, while condensates larger than the confocal volume will appear brighter due to fluorescence leakage. Consequently, cells with condensates display a wide range of maximal intensities associated with condensate size, but relatively constant median intensities (reflecting C_{sat}). **D**. Modulation of hnRNPA1 C_{sat} through changes in aromatic residue content. Three hnRNPA1 mutants previously characterized *in vitro*²⁹ were expressed with *micDROP*-green and their relative C_{sat} measured. Aromatic residues are represented by orange dots in the cartoon sequences above each scatterplot. The fraction of cells identified as containing a condensate is measured (p), along with C_{sat} and the slope (s) of the linear regression. To facilitate comparison we show the same number of randomly sampled data points ($n=1000$) for each scatterplot. Decreased aromatic residue content inhibits phase separation and increases C_{sat} . **E**. C_{sat} measurements of three model IDRs (RBM14, TAF15, and FUS) using *micDROP*-green. G3BP1's IDR serves as a negative control and does not form condensates even at high expression levels, in agreement with *in vitro* data⁴⁷. **F**. The total stickiness³¹ of IDRs correlates with the C_{sat} measured *in vivo*, with stickier IDRs exhibiting higher propensities to phase separate.

Table 1. Candidate IDRs for phase separation using *micDROD*. Disordered fraction of the protein was derived from MobiDB⁴⁹ consensus prediction. Biomolecular condensate, start and end positions in sequence were derived from annotations in the PhaSePro database⁴⁵.

PROTEIN	UNIPROT	IDR %	BIOMOLECULAR CONDENSATE	LENGTH	START POS.	END POS.
Cyclin T1	O60563	48	Nuclear Speckle	193	462	654
Ddx3	O00571	33	Stress Granule	168	1	168
Ddx4	Q9NQI0	39	Nuage	236	1	236
DYRK3	O43781	18	Stress Granule	188	1	188
ER α	P03372	61	Enhancer Condensate	96	79	174
FUS	P35637	84	Neuronal Granule, Paraspeckle, Stress Granule	214	1	214
G3BP1	Q13283	54	Stress Granule	325	142	466
GATA3	P23771	58	Enhancer Condensate	156	105	260
hnRNPA1	P09651	50	Nuclear Body, Stress Granule	135	186	372
HP1 α	P45973	53	Heterochromatin	177	1	177
HspB8	Q9UJY1	61	Stress Granule	92	1	92
Nephrin	O60500	11	Receptor Cluster	165	1077	1241
NPM1	P06748	51	Nucleolus	294	1	294
NUP98	P52948	23	Nuclear Pore Complex	500	1	500
RBM14	Q96PK6	32	Paraspeckle	320	350	669
Synapsin-1	P17600	53	Presynaptic active zone	285	421	705
TAF15	Q92804	89	Transcriptional Condensate	216	1	216
TDP43	Q13148	33	Nucleolus, Paraspeckle, Stress Granule	152	263	414
UBQ2	Q9UHD9	36	Stress Granule, UBQLN puncta	247	450	624

All nineteen candidates were tested in *micDROD*-red, and fifteen were additionally tested in *micDROD*-green. Thirteen candidates expressed well in both systems, and their phase separation behavior was consistent. Of those, seven did not meet our criteria for phase separation in both systems, while six (FUS, hnRNPA1, HspB8, RBM14, TAF15, and TDP43) exhibited a measurable C_{sat} (**Figure 2E**, **Figure S4**, **Figure S5**). The measured C_{sat} values were comparable between constructs ($R=0.76$, **Figure S2**), confirming that IDRs behaved similarly in both systems. Four additional candidates, Ddx4, DYRK3, ER α and UBQ2, showed phase separation propensity in *micDROD*-red but were not tested in *micDROD*-green (**Figure S5**).

The candidates not identified as phase separating fell into three categories. The first were lowly expressed (median intensity <100A.U.), and included CycT1, Ddx3x, and NUP98. Such low expression might be related to the toxicity and degradation of these particular constructs. Indeed, overexpression of disordered regions has been described as being frequently toxic⁵². The second category consisted of

candidates forming condensates that did not fulfill our criteria for determining phase separation, either because <20% of the cells contained condensates, or because the dilute phase concentration did not appear stable among cells containing a condensate ($s > 0.4$). This category included nuclear proteins (GATA3, NPM1 and HP1 α) and might reflect incompatibility with the cytosolic localization we imposed, or could be caused by interactions of the system with other cellular components. The final category consisted of candidates that did not form condensates, even at high expression levels. These included our negative control (G3BP1) as well as Syn1. Several reasons may explain Syn1's discrepancy. These include the solution conditions for Syn1's phase separation *in vitro* (e.g., the requirement for PEG⁵³), which may not be reflected *in vivo*, or the fusion of Syn1 to *micDROP*, which could inhibit its phase separation.

In summary, while nuclear and membrane-bound (NUP98, Nephlin) proteins were not suited for constitutive expression in the yeast cytosol, the majority of the remaining IDRs expressed well and were amenable to characterization with *micDROP*. In total, ten IDR candidates were identified as phase-separating and showed a range of C_{sat} values spanning two orders of magnitude. Such comparative data on a range of IDRs enabled investigating whether certain IDR properties were associated with self-recognition and phase separation propensity. To identify such properties, we compared C_{sat} values of IDRs to their respective features gathered for generating the UMAP. The strongest association appeared with the total stickiness of IDRs ($R = -0.72$, **Figure 2F**, **Figure S2**). This feature was previously identified as predictive of client recruitment into condensates³¹, which together with these observations implies that similar molecular interactions drive client-scaffold and scaffold-scaffold interactions.

Average particle valency modulates IDR saturation concentration

Multivalent interactions are a key driving force underlying phase separation^{37,54–56}. The avidity originating in multivalence can increase affinity and enables self-assembly while maintaining transient and dynamic interactions of individual sites, thus promoting a liquid-like state. Conversely, strong and stable interactions may not benefit from multivalency as much, as even a small number of interaction sites (e.g., four as opposed to twelve) could still promote the assembly of a network of physical interactions. In this case, however, such an assembly would not be dynamic and liquid-like.

We devised a simple strategy to modulate the valence of *micDROP*, guided by two primary motivations. First, we sought to test whether condensates with decreased valence were associated with increased C_{sat} . Such correlation would indicate that *micDROP* was behaving in accordance with classic theories of phase separation, where a lower molecular valence, due to fewer interaction points, is typically associated with a higher C_{sat} required to induce phase separation. Conversely, we reasoned that constructs driving the formation of aggregates (e.g., due to misfolding or mis-assembly) would not respond to the valence modulation strategy. Secondly, we were motivated to characterize the impact of valency changes across different IDR systems to identify general behaviors or invariant characteristics. To modulate the average valency of the IDRs presented, we co-expressed a *micDROP*-red subunit with a 'valence modulation' subunit, composed of the same Dps scaffold fused to a different fluorescent

reporter (Venus) with no IDR (**Figure 3A**). Upon co-expression, both subunits could co-assemble and form hybrid complexes, thereby modulating the number of IDRs presented per assembled particle. We harnessed the natural expression variability inherent to plasmids, whereby cells expressing both subunits on independent plasmids sampled an array of average particle valencies, which were estimated from the GFP and RFP fluorescence intensities in individual cells. For each strain expressing a specific IDR, the imaged cells were binned into five quintiles based on their GFP intensity, corresponding to the ‘valency modulator’ concentration (**Figure 3B, 3C**). The percentage of cells containing condensates in each quintile and their relative C_{sat} were then calculated as previously specified (**Figure 3D**).

Of the ten IDRs previously identified as undergoing phase separation in *micDROP*-red, seven displayed a marked sensitivity to valency changes. FUS, hnRNPA1, HspB8, TAF15, Ddx4, DYRK3 and ER α showed a 5-fold or greater decrease in the percentage of condensate-containing cells between the first and fourth quintiles (**Figure 3D**). A similar effect was observed for their C_{sat} values, which increased 3 to 5-fold for FUS, hnRNPA1, HspB8, and TAF15, or increased to a point where they could not be detected for Ddx4, DYRK3, and ER α .

Interestingly, three IDR constructs, RBM14, TDP43, and UBQ2, exhibited decreased sensitivity to valency changes, with either no change, 1.6 or 1.2-fold decrease observed in the percentage of cells with condensates, and 1.5, 1.2 or 1.9-fold increase in C_{sat} between the first and fourth quintiles, respectively. We carried out the mirror experiment using *micDROP*-green (**Figure S6**), which confirmed the decreased sensitivity to effective valency changes for RBM14 and TDP43. UBQ2 was not tested in *micDROP*-green.

Overall, valency modulation preserved the relative ranking of C_{sat} across the different constructs: those exhibiting the lowest C_{sat} at high valency also showed the lowest C_{sat} when valency decreased (**Figure 3E**). Constructs exhibiting the lowest initial C_{sat} were also least affected in their propensity to form condensates, as reflected by the fold change in the percentage of cells containing condensates (**Figure 3F**). It is possible that stronger homotypic interactions in these constructs compensate for the change in valency in the concentration regimes of our experiments. Interestingly, however, we noted that TDP43 was an outlier in that respect. Despite having a measured C_{sat} considerably higher than that of RBM14 and UBQ2 (and also higher than those of HspB8, TAF15, and hnRNPA1), TDP43 showed a lack of sensitivity to valency changes. This observation appears at odds with the notion of strong IDR-IDR interactions and may be related to a liquid-to-solid maturation previously reported *in vivo* and *in vitro*^{57–59}.

TDP43 and UBQ2 localize into discrete condensates while other IDRs exhibit pronounced promiscuity

Our experiments yielded ten IDR candidates that underwent phase separation in *micDROP*-red with a wide range of C_{sat} and sensitivity to valency changes. We next sought to determine whether these IDRs were capable of forming discrete condensates through sequence-specific interactions. To this aim, we first validated that homologous scaffolds from *micDROP*-green and *micDROP*-red did not co-assemble (**Figure 4A**). We purified each scaffold in two forms: one untagged and another fused to a GFP tag. The

untagged protein was incubated overnight with each of the GFP-tagged constructs, and the resulting particle size distribution was measured by mass photometry (**Methods**). Co-assembly was observed when the non-tagged scaffold was incubated with its identical GFP-tagged counterpart, but no co-assembly was detected between pairs of homologs (**Figure 4B, Figure S8**). In addition to these *in vitro* experiments, we ascertained that both homologs did not co-assemble *in vivo*. We co-expressed both scaffolds in cells, with the *micDROP*-green subunit fused to hnRNPA1 IDR, and the *micDROP*-red subunit without an IDR.

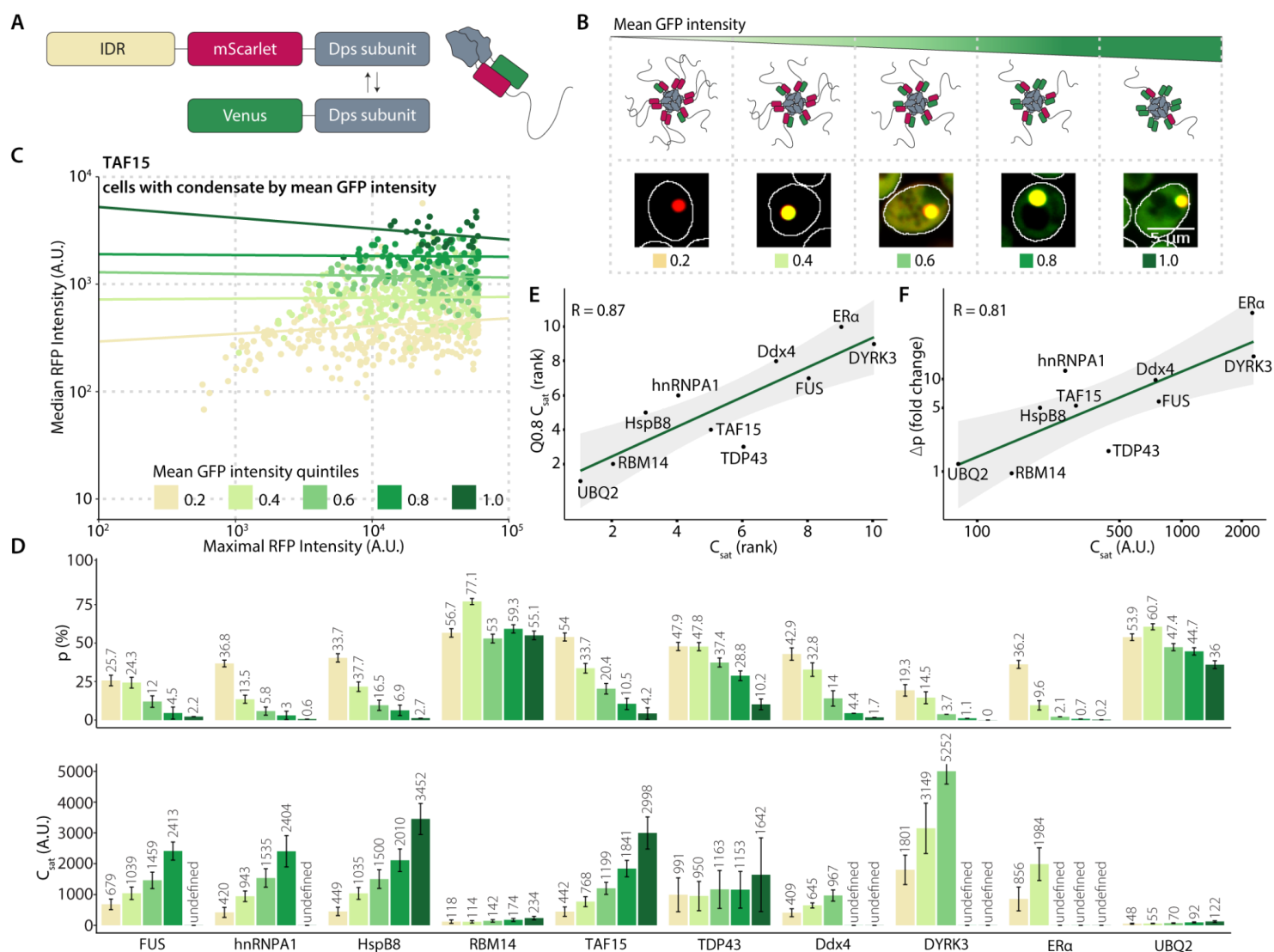


Figure 3. Average particle valency modulates IDR C_{sat} . **A.** *micDROP*-red is co-expressed with a valency modulator consisting of the same Dps scaffold subunit fused to Venus and lacking an IDR. **B.** Co-expression of both constructs *in vivo* results in their co-assembly into hybrid scaffolds, with average IDR valency depending on their relative stoichiometries. Cells containing condensates that co-express *micDROP*-red with the valency modulator are binned into five quintiles by GFP intensity. Cells in the first quintile exhibit the lowest GFP levels and highest average particle valency. Conversely, cells in the fifth quintile exhibit the highest GFP levels and lowest valency. **C.** C_{sat} measurements of TAF15 in *micDROP*-red with a valency modulator. Each dot represents a cell containing a condensate, and is colored by quintiles of GFP intensity. The coordinates of each cell are the maximal (x-axis) and median (y-axis) intensity of RFP fluorescence. Decreased particle valency (high quintile) is associated with increased C_{sat} . **D.** Barplots showing the percentage (p) of condensate-containing cells (top) and C_{sat} (bottom) per quintile of GFP expression, for ten IDR constructs. **E.** Rank-correlation of C_{sat} measurements of the first against fourth quintile. Missing values were ranked by data from other quintiles (Ddx4 < DYRK3 < ER α). **F.** Correlation of C_{sat} measurements against the fold-change in the percentage of condensate-containing cells between the first and fourth quintile. IDRs associated with high C_{sat} displayed a marked decrease in the percentage of condensate-containing cells, with the notable exception of TDP43, whereas IDRs with a low C_{sat} did not.

We did not detect the *micDROP*-red subunits enriched in condensates, which also indicated a lack of co-assembly (**Figure S8**). The mirror experiment, with the *micDROP*-red subunit fused to hnRNPA1 and the *micDROP*-green subunit with an identical scaffold and no IDR, also demonstrated a lack of co-assembly.

A matrix of strains was created wherein each strain co-expressed *micDROP*-green and *micDROP*-red fused to respective IDRs (**Figure 4C**). Considering cells in which condensates were identified both for *micDROP*-green and *micDROP*-red, we calculated the distance between the centers of the two nearest condensates. If the distance was smaller than $0.3\mu\text{m}$ (~ 3 pixels), we inferred that the two IDRs colocalized. We decided on this distance because pairs harboring the same IDR consistently showed distances below this value. Based on this cutoff, we determined that eight of the ten IDRs displayed pronounced promiscuity and colocalized indiscriminately (**Figure 4D**); these correspond to FUS, hnRNPA1, HspB8, RBM14, TAF15, Ddx4, DYRK3 and ER α .

Among these constructs, the pair with the greatest similarity in sequence features was FUS and TAF15, as seen in the UMAP (**Figure S2**). Both are members of the FET (FUS, EWSR1, TAF15) family, similarly structured RNA-binding proteins (RBPs) that undergo phase separation through a Prion-like domain (PrLD) harboring Tyr and Arg 'stickers'^{22,60}. As such, it could be expected that they would colocalize indiscriminately³¹, and their partitioning into different biomolecular condensates is likely regulated by cellular localization signals⁶¹. hnRNPA1 and RBM14 are also RBPs undergoing phase separation through PrLDs, though their disordered regions display greater diversity from those of FUS and TAF15, primarily in length, stickiness, and net charge, as seen in the UMAP. Despite this increased diversity, hnRNPA1 and RBM14 colocalized with FUS and TAF15 and with each other, showing that favorable interactions between these IDRs were permissive. Interestingly, RBM14 previously showed a low C_{sat} and a marked insensitivity to valency changes (**Figure 3D**). Nevertheless, its lack of specificity implies that strong homotypic affinity is insufficient to form discrete condensates that do not intermix.

The FUS-HspB8 pair is known to interact, with HspB8's folded α -crystallin domain preventing FUS maturation by chaperoning its RNA-binding domain^{62,63}, which is distinct from the IDR region. Interestingly, our results imply that interactions between the IDR regions of these two proteins are also favorable.

Among the co-localizing group of IDRs, the two furthest away from the RBP group (FUS, hnRNPA1, TAF15, RBM14) on the UMAP were Ddx4 and DYRK3, indicating they had the greatest diversity in their sequence features. Ddx4's phase separation is driven by electrostatic interactions, with repeating residue blocks of alternating net charge, but also contains an over-representation of Phe-Gly, Gly-Phe, Arg-Gly, and Gly-Arg motifs within the positively charged blocks¹². Our results indicate that these conserved motifs are sufficient to promote favorable interactions with PrLDs even in cases of strong homotypic affinity such as RBM14. The dual-specificity kinase DYRK3 partitions into stress granules via its intrinsically disordered N-terminal domain and its kinase activity is required for the dissolution of stress granules, possibly by phosphorylating multiple RBPs⁶⁴. The IDR construct (not including the kinase domain) colocalized with FUS and TAF15, but displayed only intermediate specificity when co-expressed with hnRNPA1, RBM14 and HspB8. This behavior was unexpected, as it implies some

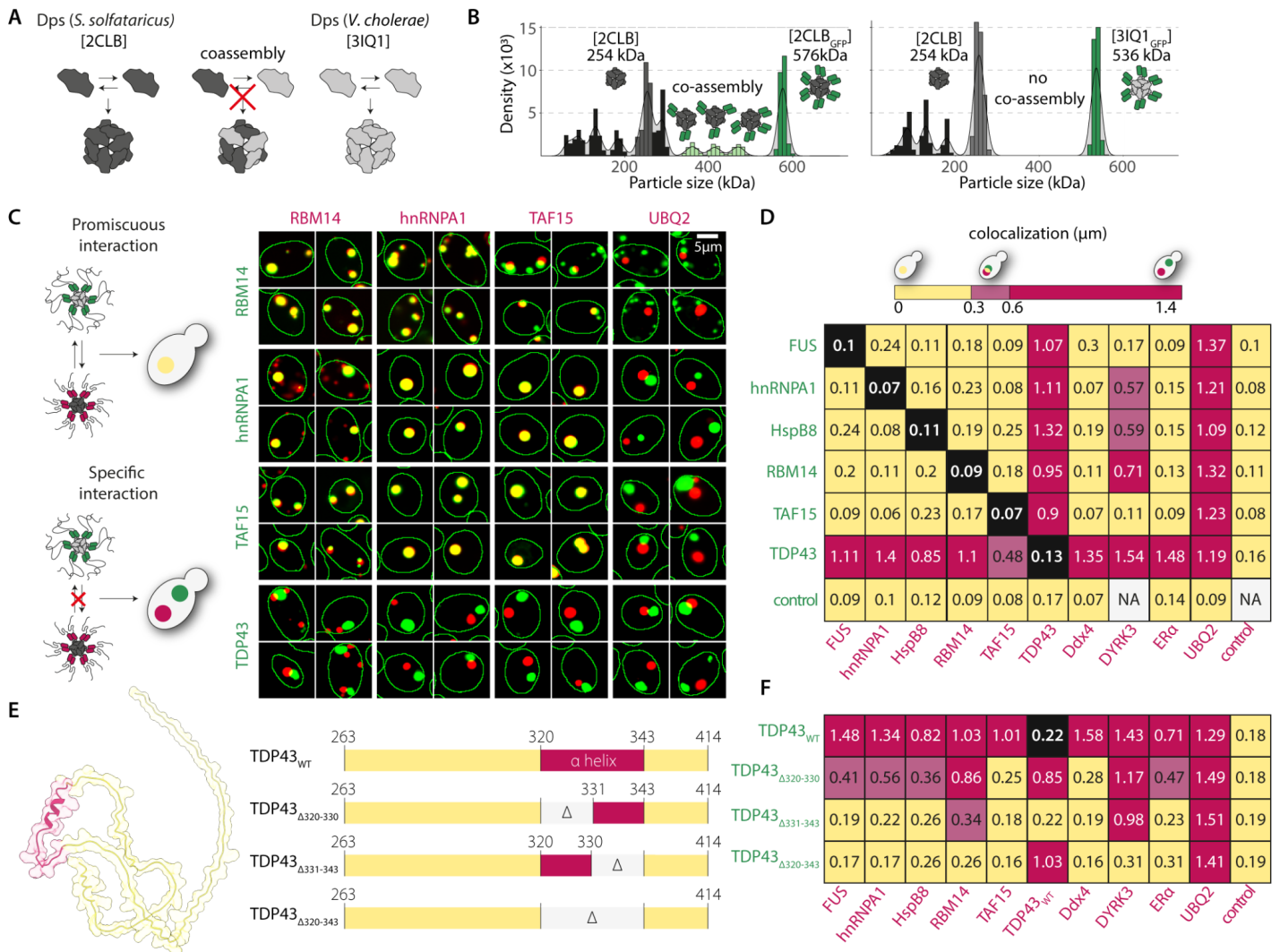


Figure 4. Most IDRs display high promiscuity while TDP43 forms discrete condensates through an α -helical segment.

A. Probing IDR-IDR interaction specificity with *micDROPs* necessitates two homologous scaffolds that do not co-assemble. We employed scaffolds from *S. solfataricus* (dark gray) and *V. cholerae* (light gray). **B.** We expressed and purified the scaffolds either tagged with GFP or untagged, and subsequently incubated both variants overnight to test their co-assembly by mass photometry. We observed hybrid complexes (left), but did not observe such co-assembly when incubating the homologous scaffolds (right). **C.** Schematic representation of specific vs. promiscuous IDR-IDR interactions (left). IDRs interacting promiscuously form condensates where both constructs colocalize, visible as a yellow puncta. Conversely, IDRs that interact in a self-specific manner yield spatially distinct condensates, as seen for TDP43 or UBQ2 in fluorescence microscopy images co-expressing *micDROPs*-green and *micDROPs*-red (right). **D.** Colocalization matrix showing the median distance (μ m) between condensates formed by *micDROPs*-green and *micDROPs*-red. **E.** Cartoon rendering of the TDP43 disordered region, based on a model generated with AlphaFold2 (left). TDP43 contains a transient α -helical region (red) predicted to undergo disorder-to-order transition upon homotypic binding⁶⁸. Three mutant constructs (Δ 320-330, Δ 331-343, and Δ 320-343) were designed to assess whether partial or complete deletion of the α helix impacted TDP43's specificity (left). **F.** Median distance (μ m) between condensates of TDP43 variants and condensates formed by other IDRs. The variants Δ 331-343 and Δ 320-343 lost their homotypic specificity and colocalized with all IDRs, with the exception of UBQ2. The variant Δ 320-330 showed intermediate specificity.

unique mechanism or conformation of the DYRK3 IDR, which can interact differently with members of the ‘promiscuous’ group.

The IDRs of TDP43 and UBQ2, which were located closer to the RBP group (TAF15, FUS, hnRNPA1 and RBM14) on the UMAP compared to DYRK3 or Ddx4, nevertheless formed discrete condensates regardless of their IDR partner. Such behavior implies these IDRs contain additional features driving their exclusive self-recognition. Upon examining the two sequences, we observed that both TDP43 and UBQ2 contain short structured segments in their disordered sequence, which might account for their specificity. First, UBQ2 has a ubiquitin-associated domain (UBA) at its C-terminus, which recognizes polyubiquitin chains on marked proteins^{65,66}. In the future, *micDROP* could serve to test mutant sequences of UBQ2 and examine the effect of the UBA segment on UBQ2’s phase separation propensity and self-recognition specificity. Second, TDP43 has an evolutionarily conserved transient α -helical segment known to modulate its phase separation propensity^{67,68}. This segment can undergo disorder-to-order transition upon homotypic binding, which we hypothesized could account for TDP43’s specificity.

A transient α -helix in TDP43’s disordered region governs its self-specific condensation

The disordered region of TDP43 is a PrLD that lacks extensive RGG repeats compared to other IDRs from RNA-binding proteins such as FUS or RBM14. However, it contains a conserved region (**Figure 4E**) that forms a transient α helix that promotes homotypic interactions⁵⁷. To test whether this region mediates the specific self-recognition of TDP43, we generated three mutant sequences with either partial (Δ 320-330 and Δ 331-343) or complete (Δ 320-343) deletions of this segment and tested their colocalization with other IDRs using *micDROP*. Variant Δ 320-330 showed intermediate specificity with other IDRs and surprisingly, lost the ability to colocalize with TDP43 WT. Variant Δ 331-343 lost all specificity and colocalized promiscuously with all IDRs, including TDP43 WT. Variant Δ 320-343 also colocalized promiscuously with all other IDRs, but lost the ability to colocalize with TDP43 WT (**Figure 4F, Figure S8**). We conclude that TDP43’s specificity is governed by the α helical segment, and hypothesize that the 320-330 region modulates homotypic recognition while the 331-343 segment is likely required to stabilize the interaction.

CONCLUSIONS

We developed a synthetic system called *micDROP* that proved powerful in studying the potential of IDRs to drive intracellular phase separation. Characterizing interaction properties of IDRs is notoriously difficult, particularly *in vivo*, owing to their conformational flexibility. Nevertheless, we showed that *micDROP* captured the phase separation propensity of various IDRs *in vivo* by quantifying their saturation concentrations. The simplicity of *micDROP*’s design enabled us to analyze a large number of IDRs, which revealed a correlation between an IDR’s C_{sat} and its total sequence stickiness. We also showed that the effective valency of *micDROP* could be easily modulated, which revealed both expected phase separation behavior for most IDRs, and a surprising one for TDP43. The characterization of two homologous scaffolds that did not co-assemble made *micDROP* amenable to studying IDR

self-recognition specificity. This feature is unique, as recent work has been employing different scaffold geometries to this effect⁶⁹, which adds confounding factors absent from *micDROP*.

Equally important, the modular nature of the system means it is readily generalizable to other IDR sequences of interest. Our findings indicate that IDRs capable of undergoing phase separation are generally promiscuous owing to their interactions through nonspecific sticky patches³¹. IDRs that formed discrete condensates contained structured segments, reinforcing the notion that fully disordered and flexible regions undergoing phase separation are inherently non-specific. These results imply that regulating the formation and composition of biomolecular condensates should generally require complex networks of IDRs functioning hand-in-hand with structured partners or latent structural elements encoded in their sequences.

METHODS

Ferritin scaffold candidate selection

The list of candidate ferritins was obtained from known structures found in the Protein Data Bank (PDB)⁷⁰, which we mined based on the 3DComplex database⁷¹. Specifically, we curated a list of 24-mer ferritin and 12-mer ferritin-like proteins non-redundant at a level of 40% sequence identity. Their UniProt sequences were aligned using Clustal Omega MSA tool⁷², and eight final candidates were selected (**Table S2**).

Ferritin scaffold cloning, expression and purification in *E. coli*

All UniProt sequences were optimized for bacterial expression, and synthesized into two pET T7 RNA polymerase-driven transcription vectors (Twist Bioscience) with the following sequence features:

Plasmid 1: [6xHIS] [TEV (tobacco etch virus)] [Ferritin scaffold] [Ampicillin resistance cassette]

Plasmid 2: [6xHIS] [TEV] [super-folder Green Fluorescent Protein (sfGFP)] [linker] [Ferritin scaffold] [Kanamycin resistance cassette]

Plasmids were transformed into *E. coli* BL21-CodonPlus (DE3)-RIPL cells. Transformant cells were grown in LB media at 37°C with appropriate antibiotics to OD₆₀₀ ~0.8. Expression was induced by the addition of 1mM isopropyl b-D-1-thiogalactopyranoside (IPTG) and allowed to proceed overnight (~7h) at 25°C. Cells were harvested using French press and centrifugation. Each protein was purified on a 5ml nickel nitrilotriacetic acid (Ni²⁺-NTA) resin (GE Healthcare). The column was washed stepwise with 20, 50, 100 and 250mM imidazole. Each protein was eluted with 500mM imidazole, and the N-terminal 6xHIS tag was removed by TEV protease cleavage. Each protein was then purified on a HiPrep™ 16/60 Sephacryl S-500 HR gel filtration column (GE Healthcare) equilibrated with 50mM HEPES (pH 7.5), 50mM KCl and 1mM dithiothreitol (DTT).

Assessing co-assembly of ferritin scaffolds using mass photometry

Purified protein samples were prepared for imaging by mixing each non-tagged scaffold with all GFP-tagged constructs at a 1:1 concentration ratio, followed by overnight incubation at 10°C.

Immediately prior to mass photometry measurements, samples were diluted using stock buffer (50mM HEPES (pH 7.5), 50mM KCl and 1mM dithiothreitol (DTT)). Measurements were performed using OneMP Refeyn Ltd. mass photometer. To find focus, a fresh stock buffer was first flown into the chamber, the focal position was identified and secured in place with an autofocus system based on total internal reflection for the entire measurement period. Typical working concentrations of protein mixtures were ~25–50nM, and each sample was measured in a new flow-chamber. For each acquisition, ~15µL of diluted protein mix was introduced into the flow-chamber and, following autofocus stabilization, movies of 60-90s were recorded. Each sample was measured three times independently. The molecular weight was obtained by contrast comparison with known mass standard calibrants measured on the same day.

UMAP and IDR candidate selection

A uniform manifold approximation and projection (UMAP) map was generated for the entire human disordered proteome (all protein segments of ≥ 35 residues predicted to be disordered according to consensus predictions reported in MobiDB⁷³ (**Table S1**, **Figure S2**). The number of neighbors was set to 20 or 40 to assess clustering stability across this range. Variables taken into account when generating the UMAP included:

- Amino acid enrichment compared to entire human disordered proteome
- Frequencies and over-representation of amino acids
- Net charge
- pI
- Hydrophobicity⁷⁴
- Stickiness³¹
- AGGRESCAN values⁷⁵

The entries of human IDRs listed in the PhaSePro database⁴⁵ and that we selected were highlighted onto the UMAP. The final candidate list (**Table 1**) was generated by considering UMAP distances, sequence length, overall net charge, and participation in functionally distinct biomolecular condensates.

IDR cloning and expression in *S. cerevisiae*

UniProt sequences were optimized for expression in *S. cerevisiae*. Fifteen candidates were synthesized (Twist Bioscience) in two custom expression vectors containing the following features:

micDROP-green: [TDH3 promoter] [Nuclear export signal (NES)] [IDR] [linker] [YFP Venus] [linker] [Ferritin scaffold] [CYC terminator] [Hygromycin resistance cassette]

micDROP-red: [TDH3 promoter] [NES] [IDR] [linker] [RFP mScarlet] [linker] [Ferritin scaffold] [CYC terminator] [Nourseothricin resistance cassette]

Six additional candidates were synthesized only into *micDROP-red*. *micDROP-red* plasmids were transformed into *S. cerevisiae* BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). *micDROP-green*

plasmids were transformed into *S. cerevisiae* BY4742 (MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0). Haploid strains were then mated on YPD agar plate and allowed to grow overnight (~12h). Diploid selection was repeated twice on YPD agar plate with Hygromycin and Nourseothricin selections.

Confocal fluorescence microscopy, image processing, and analysis

For imaging, 1 μ l of saturated cell suspension was transferred to an optical plate with 29 μ l fresh SD media with appropriate antibiotics and grown at 30°C for 3h to logarithmic growth. Cells were imaged with an Olympus IX83 microscope coupled to a Yokogawa CSU-W1 spinning-disc confocal scanner with dual prime-BSI sCMOS cameras (Photometrix). The 16-bit images were acquired with three illumination schemes of 50 ms exposure each: one for brightfield, one for green, and one for red fluorescence. For green fluorescence, we excited the sample with a 488 nm laser (Coherent 150 mW) and collected light through a bandpass emission (Em) filter (525/50 nm, Chroma ET/m). For red fluorescence, we excited at 561 nm (Coherent 150 mW) and used a bandpass emission filter (609/54 nm, Chroma ET/m). One multiband dichroic mirror was used for the three illuminations. Imaging was performed with a \times 60/1.42 numerical aperture (NA), oil-immersion objective (UPLSAPO60XO, Olympus).

Following acquisition, cells were identified, segmented and their fluorescent signal (median, average, minimum, maximum, 10th, 20th, ..., 90th percentile fluorescence) were identified using previously reported scripts in ImageJ⁴³. Condensates were identified in each cell independently, in a multistep process: (1) calculation of the median fluorescence intensity of pixels in a given cell; and (2) identification of the largest region composed of pixels with an intensity twofold (GFP) or threefold (RFP) above the median. If such a region existed and it showed a circularity over 0.4, the condensate properties (intensity, size, etc), and its coordinates were recorded. Tabulated data from image analyses were loaded and analyzed further with custom scripts in R.

Analysis of IDRs' phase separation propensity *in vivo*

We noticed that very high concentrations of RFP gave a noticeable (1-2%) cross-talk signal in the GFP channel. Therefore, we applied a GFP channel brightness correction to account for this crosstalk. We subtracted 2% of RFP brightness values from the GFP values. Additionally, we discarded all cells where the GFP intensity was saturated (value above 65536) as no correction could be reliably applied for those.

Condensates detected during image analysis were deemed valid if the mean brightness within the condensate was significantly higher than background fluorescence (>150 arbitrary fluorescence units, AFUs) after the correction for crosstalk was applied. If a strain had >20 segmented cells with valid condensates and at least 20% of the segmented cells were identified as containing a condensate, we analyzed the propensity of the corresponding IDR to phase separate by analyzing the slope of points defined by maximal (x-axis) and median (y-axis) fluorescence intensities within each cell.

Analysis of IDR colocalization *in vivo*

Colocalization calculations were done for IDR candidates previously identified as phase separating. For each strain, we ensured that at least 50 cells containing condensates in both fluorescent channels were

used to calculate a co-localization score. We used the minimal Euclidean distance between condensates, based on the largest two condensates per channel. Thus, the minimum of (at most four) pairwise distances was recorded for each cell. Subsequently, the median of these distances gave the colocalization score for the corresponding strain.

Generation of structure models

The structure models were generated with AlphaFold2 Multimer^{41,76} using a local implementation of Colabfold⁷⁷.

Data Availability

The code and data used to generate the UMAP is available at:

https://github.com/benjamin-elusers/human_idr.

Plasmids and strains are available upon request.

Authors contributions

AG and EDL designed and planned the research. AG carried out the experimental work and analyzed the microscopy data. AG and EDL interpreted the results. BD carried out the bioinformatics analyses on IDR regions to generate the UMAP. AG and EDL wrote the manuscript with input from BD.

Acknowledgments

We thank Prof. Sam Safran of the Department of Chemical and Biological Physics at the Weizmann Institute of Science for thoughtful ideas and suggestions regarding phase separation theory, Prof. Michael Elbaum and Dr. Alexandra Tayar for their insightful feedback on experimental design and analysis, Prof. Rina Rosenzweig of the Department of Chemical and Structural Biology at the Weizmann Institute of Science for her advice on chaperone selection, Dr. Adar Sonn-Segev for her expertise with mass photometry, and past and current members of the Levy lab for their invaluable support and assistance. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819318), by the Israel Science Foundation (grant no. 1452/18), by a research grant from A.-M. Boucher, by research grants from the Estelle Funk Foundation, the Estate of Fannie Sherr, the Estate of Albert Deligher, the Merle S. Cahn Foundation, Mrs. Mildred S. Gosden, the Estate of Elizabeth Wachsman, the Arnold Bortman Family Foundation.

REFERENCES

1. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18**, 756–764 (2008).
2. van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
3. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26**, 668–679 (2016).
4. Liu, X. *et al.* Phase separation drives decision making in cell division. *J. Biol. Chem.* **295**, 13419–13431 (2020).
5. Levone, B. R. *et al.* FUS-dependent liquid–liquid phase separation is important for DNA repair initiation. *J. Cell Biol.* **220**, (2021).
6. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
7. Mitrea, D. M. & Kriwacki, R. W. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal.* **14**, 1 (2016).
8. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
9. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
10. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell*

- Dev. Biol.* **30**, 39–58 (2014).
11. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604–609 (2000).
 12. Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
 13. Elbaum-Garfinkle, S. *et al.* 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 (2015).
 14. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
 15. Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out—Compositional Control of Biomolecular Condensates. *J. Mol. Biol.* **430**, 4666–4684 (2018).
 16. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).
 17. Marmor-Kollet, H. *et al.* Spatiotemporal Proteomic Analysis of Stress Granule Disassembly Using APEX Reveals Regulation by SUMOylation and Links to ALS Pathogenesis. *Mol. Cell* **80**, 876–891.e6 (2020).
 18. Sama, R. R. K. *et al.* FUS/TLS assembles into stress granules and is a prosurvival factor during hyperosmolar stress. *J. Cell. Physiol.* **228**, 2222–2231 (2013).
 19. Martinez-Macias, M. I. *et al.* FUS (fused in sarcoma) is a component of the cellular response to topoisomerase I-induced DNA breakage and transcriptional stress. *Life Sci. Alliance* **2**, e201800222 (2019).
 20. An, H. *et al.* ALS-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation of dysfunctional paraspeckles. *Acta Neuropathol Commun* **7**, 7 (2019).
 21. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
 22. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688–699.e16 (2018).
 23. Vernon, R. M. & Forman-Kay, J. D. First-generation predictors of biological protein phase separation. *Curr. Opin. Struct. Biol.* **58**, 88–96 (2019).
 24. Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**, (2018).
 25. Das, S., Lin, Y.-H., Vernon, R. M., Forman-Kay, J. D. & Chan, H. S. Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28795–28805 (2020).
 26. Bremer, A. *et al.* Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **14**, 196–207 (2022).
 27. Greig, J. A. *et al.* Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol. Cell* **77**, 1237–1250.e4 (2020).
 28. Hazra, M. K. & Levy, Y. Biophysics of Phase Separation of Disordered Proteins Is Governed by Balance between Short- And Long-Range Interactions. *J. Phys. Chem. B* **125**, 2202–2211 (2021).
 29. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
 30. Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife* **8**, (2019).
 31. Villegas, J. A. & Levy, E. D. A unified statistical potential reveals that amino acid stickiness governs nonspecific recruitment of client proteins into condensates. *Protein Sci.* **31**, e4361 (2022).
 32. Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* vol. 175 1467–1480.e13 Preprint at <https://doi.org/10.1016/j.cell.2018.10.048> (2018).
 33. Zhang, Y. & Orner, B. P. Self-assembly in the ferritin nano-cage protein superfamily. *Int. J. Mol. Sci.* **12**, 5406–5421 (2011).
 34. Mitrea, D. M. *et al.* Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. *Nat. Commun.* **9**, 1–13 (2018).
 35. Yang, L., Gal, J., Chen, J. & Zhu, H. Self-assembled FUS binds active chromatin and regulates gene transcription. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17809–17814 (2014).
 36. Afroz, T. *et al.* Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat. Commun.* **8**, 45 (2017).
 37. Nandi, S. K., Österle, D., Heidenreich, M., Levy, E. D. & Safran, S. A. Affinity and Valence Impact

- the Extent and Symmetry of Phase Separation of Multivalent Proteins. *Phys. Rev. Lett.* **129**, 128102 (2022).
38. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
 39. Flory, P. J. Thermodynamics of high polymer solutions. *J. Chem. Phys.* **10**, 51–61 (1942).
 40. Kremers, G.-J., Goedhart, J., van Munster, E. B. & Gadella, T. W. J., Jr. Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. *Biochemistry* **45**, 6570–6580 (2006).
 41. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
 42. Villegas, J. A., Heidenreich, M. & Levy, E. D. Molecular and environmental determinants of biomolecular condensate formation. *Nat. Chem. Biol.* **18**, 1319–1329 (2022).
 43. Heidenreich, M. *et al.* Designer protein assemblies with tunable phase diagrams in living cells. *Nat. Chem. Biol.* **16**, 939–945 (2020).
 44. Fouquet, C. *et al.* Improving axial resolution in confocal microscopy with new high refractive index mounting media. *PLoS One* **10**, e0121096 (2015).
 45. Mészáros, B. *et al.* PhaSePro: the database of proteins driving liquid–liquid phase separation. *Nucleic Acids Res.* **48**, D360–D367 (2019).
 46. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325–345.e28 (2020).
 47. Guillén-Boixet, J. *et al.* RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346–361.e17 (2020).
 48. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
 49. Piovesan, D. *et al.* MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res.* **49**, D361–D367 (2021).
 50. Gauss, G. H. *et al.* Structure of the DPS-like protein from *Sulfolobus solfataricus* reveals a bacterioferritin-like dimetal binding site within a DPS-like dodecameric assembly. *Biochemistry* **45**, 10815–10827 (2006).
 51. Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).
 52. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208 (2009).
 53. Milovanovic, D., Wu, Y., Bian, X. & De Camilli, P. A liquid phase of synapsin and lipid vesicles. *Science* **361**, 604–607 (2018).
 54. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife* **6**, (2017).
 55. Locatelli, E., Handle, P. H., Likos, C. N., Sciortino, F. & Rovigatti, L. Condensation and Demixing in Solutions of DNA Nanostars and Their Mixtures. *ACS Nano* **11**, 2094–2102 (2017).
 56. Sciortino, F. & Zaccarelli, E. Equilibrium gels of limited valence colloids. *Curr. Opin. Colloid Interface Sci.* **30**, 90–96 (2017).
 57. Portz, B., Lee, B. L. & Shorter, J. FUS and TDP-43 Phases in Health and Disease. *Trends Biochem. Sci.* **46**, 550–563 (2021).
 58. Gasset-Rosa, F. *et al.* Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. *Neuron* **102**, 339–357.e7 (2019).
 59. Mann, J. R. *et al.* RNA Binding Antagonizes Neurotoxic Phase Transitions of TDP-43. *Neuron* **102**, 321–338.e8 (2019).
 60. Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
 61. Marko, M., Vlassis, A., Guialis, A. & Leichter, M. Domains involved in TAF15 subcellular localisation: dependence on cell type and ongoing transcription. *Gene* **506**, 331–338 (2012).
 62. Boczek, E. E. *et al.* HspB8 prevents aberrant phase transitions of FUS by chaperoning its folded RNA-binding domain. *Elife* **10**, (2021).
 63. Ganassi, M. *et al.* A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism. *Mol. Cell* **63**, 796–810 (2016).
 64. Rai, A. K., Chen, J.-X., Selbach, M. & Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018).

65. Zhang, K. Y., Yang, S., Warraich, S. T. & Blair, I. P. Ubiquitin 2: A component of the ubiquitin–proteasome system with an emerging role in neurodegeneration. *Int. J. Biochem. Cell Biol.* **50**, 123–126 (2014).
66. Dao, T. P. *et al.* Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol. Cell* **69**, 965–978.e6 (2018).
67. Li, H.-R., Chiang, W.-C., Chou, P.-C., Wang, W.-J. & Huang, J.-R. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *J. Biol. Chem.* **293**, 6090–6098 (2018).
68. Conicella, A. E. *et al.* TDP-43 α -helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences* **117**, 5883–5894 (2020).
69. Rana, U. *et al.* Asymmetric oligomerization state and sequence patterning can tune multiphase condensate miscibility. *bioRxiv* 2023.03.11.532188 (2023) doi:10.1101/2023.03.11.532188.
70. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
71. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2**, e155 (2006).
72. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–600 (2013).
73. Piovesan, D. *et al.* MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res.* **51**, D438–D444 (2023).
74. Roseman, M. A. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200**, 513–522 (1988).
75. Conchillo-Solé, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of ‘hot spots’ of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).
76. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2021) doi:10.1101/2021.10.04.463034.
77. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
78. Macedo, S. *et al.* The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans*. *Nat. Struct. Biol.* **10**, 285–290 (2003).
79. Masuda, T., Morimoto, S.-I., Mikami, B. & Toyohara, H. The extension peptide of plant ferritin from sea lettuce contributes to shell stability and surface hydrophobicity. *Protein Sci.* **21**, 786–796 (2012).
80. Pfaffen, S., Abdulqadir, R., Le Brun, N. E. & Murphy, M. E. P. Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. *J. Biol. Chem.* **288**, 14917–14925 (2013).
81. Kauko, A., Haataja, S., Pulliainen, A. T., Finne, J. & Papageorgiou, A. C. Crystal structure of *Streptococcus suis* Dps-like peroxide resistance protein Dpr: implications for iron incorporation. *J. Mol. Biol.* **338**, 547–558 (2004).