Sequence determinants of the aggregation of proteins within condensates generated by liquid-liquid phase separation

Michele Vendruscolo[1,*] and Monika Fuxreiter[2,3*]

[1]*Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, UK*

[2]*Department of Biomedical Sciences, University of Padova, Italy*

[3]*Department of Biochemistry and Molecular Biology, University of Debrecen, Hungary*

Abstract

The transition between the native and amyloid states of proteins can proceed via a deposition pathway via oligomeric intermediates or via a condensation pathway via liquid droplet intermediates generated through liquid-liquid phase separation. Here we investigate the sequence determinants of aggregation from within the droplet state based on generic interactions. We describe a model in which these sequence determinants can be captured by three features, the droplet-promoting propensity, the aggregation-promoting propensity and the binding mode entropy. By using this approach, we propose a formula to identify aggregation-promoting mutations in droplet-forming proteins. This analysis provides insights into the amino acid code for the conversion of proteins between liquid-like and solid-like condensates.

[*]Correspondence: mv245@cam.ac.uk, monika.fuxreiter@unipd.it

Introduction

Recent advances have revealed that proteins exhibit a complex phase behaviour, as they can populate the native, droplet and amyloid states [1, 2]. It has also become clear that the amyloid state can be reached either through a deposition pathway [3, 4], or through a condensation pathway, going through a liquid-liquid phase separation (**Figure 1**) [5-7]. The discovery of the phenomenon of liquid-liquid phase separation is highlighting a variety of previously unknown biological processes [8, 9]. Since the transition between the native and droplet states is usually reversible under cellular conditions, as opposed to that between the native and amyloid states, the droplet state can be rather safely exploited for functional processes in living systems [10, 11].

Upon dysregulation of the native-to-droplet transition, however, the droplet state can evolve into a transient gel form, which can then further proceed to the amyloid state (**Figure 1**) [12]. This gel state may give rise to cytotoxicity, as it can less readily revert to the native state [6, 13-16]. In particular, proteins can remain entangled in the gel form because they recruit other cellular components [17]. These observations prompt the question of whether there is an amino acid code determining the condensation pathway of amyloid formation [18, 19], in analogy to the amino acid code that has been identified for the deposition pathway [20, 21].

Structural studies indicate that characteristic signatures of the amyloid state, such as cross-β elements are also present in the droplet state [22-24]. Furthermore, labile β structures are enriched in droplet-forming sequences [25]. Here we focus on how the amino acid sequences encode distinct physico-chemical properties that drive the conversion between liquid-like and solid-like condensates.

We compare the nature of the interactions along the condensation pathway between droplet-promoting regions and amyloid-promoting regions. We propose a model in which the condensation pathway is driven by conformationally heterogeneous elements, which have been referred to as labile structures [25, 26]. Our analysis indicates the possibility of predicting aggregation-promoting mutations in droplet-forming proteins based on generic interactions and the ability of sequence elements to switch between different interactions modes. We thus

offer insights into the mechanisms and sequence determinants governing the condensation pathway between liquid-like and solid-like condensates and the modulation of this process by pathological mutations.

Results

A framework for the generic approach

Formation of condensed states, either the liquid-like droplet state or the solid-like amyloid state, is generic to proteins [3, 27]. Both states are stabilised by generic interactions, in accord with the wide variety of sequence motifs identified by experimental studies capable of driving liquid-liquid phase separation [27]. Here, we hypothesized that the condensation pathway can be described by sequence-encoded biophysical properties, and aimed at analysing these generic interaction behaviors.

Datasets of protein condensates representing the droplet and amyloid states

We assembled a database of droplet-forming proteins (DROP), termed 'droplet-drivers' (DROP-D), using three databases: PhaSepDB (http:/db.phasep.pro) [28], PhaSePro (https://phasepro.elte.hu) [29], and LLPSDB (http://bio-comp.org.cn/llpsdb) [30] (Methods, **Table S1**). Droplet-client proteins (DROP-C), which require an interacting partner for phase separation, are derived from LLPSDB database [31]. Regions of droplet-driver or droplet-client proteins that mediate droplet formation are assembled from the PhaSePro dataset [29] (Methods, **Table S1**). Droplet-promoting regions (DPRs) are defined based on an evidence for spontaneous phase separation (Methods, **Table S1**). In addition, low-complexity aromatic-rich kinked segments (LARKs) are used to represent structural hallmarks of droplet formation [25].

The amyloid state is populated by both pathological and functional amyloids (AMYLOID), which are derived from the AmyPro database (http://amypro.net) [32] (Methods, **Table S1**). Amyloid-promoting regions (APRs) are derived from the AmyPro database [32], which include amyloid core-forming cross-β structures and their flanking regions (Methods, **Table S1**).

In addition to the above categories, which represent proteins forming liquid-like and solid-like condensed states [33], we assembled proteins associated with both states. Proteins with an evidence for both droplet proteins and amyloid formation are defined as 'droplet-amyloids' (DROP-AMYLOID, Methods, **Table S1**). This category contains proteins with prion-like domains (e.g. Sup35) and proteins, which mature into fibrillar aggregates (e.g. tau). Proteins associated with inheritable traits in *Saccharomyces cerevisiae* [34], are defined as yeast prions (YPRIO, **Table S1**). These proteins represent functional amyloids, but we listed them separately from the droplet and amyloid datasets (Methods, **Table S1**).

The assembled datasets exhibit a wide range of solubility values [35] (**Figure 2A**).

Comparison of the binding modes of amyloid-forming and droplet-forming proteins

Increasing evidence demonstrates that proteins can exhibit a wide range of binding modes by sampling a continuum between ordered and disordered bound states [36, 37], and may switch between such states by post-translational modifications or according to the cellular context [38]. An ordered binding mode is mediated by specific contact patterns [39], which create a free energy landscape with a well-defined minimum corresponding to the bound state. By contrast, a disordered binding mode involves multimodal interactions [40], which create a heterogeneous ensemble of configurations with multiple minima in the free energy landscape of the bound state. We previously demonstrated that both ordered and disordered binding modes of proteins can be predicted from their amino acid sequences [41, 42]. We also showed that the propensity of proteins to spontaneously undergo liquid-liquid phase separation depends on the probability of being disordered in their free states (p_D) and the probability of remaining disordered upon binding (p_{DD}) [27].

In this approach, droplet-driver proteins tend to be disordered in both their free and bound forms [27], while amyloid-forming proteins tend to be ordered in both forms (**Figures 2B** and **S1A**). In addition, droplet-clients tend to resemble the ordered interactions of amyloid-forming proteins, while yeast prions, and in particular droplet-amyloid proteins exhibit disordered interactions, which do not deviate significantly from droplet-driver proteins (**Figures 2B** and **S1A**).

Amyloid-forming proteins exhibit the ability to switch between different binding modes, therefore having a high binding mode entropy (S_{bind}, Eq. 3) (**Figure 2C**). By contrast, droplet-driver and droplet-client proteins, and in particular droplet-amyloid proteins, exhibit low binding mode entropy (**Figure 2C**), reflecting a strong preference for forming disordered interactions in the condensates. Interestingly, the predictions indicate that yeast prions are less prone to switch from disordered to ordered interaction modes than amyloid-forming proteins.

Taken together, this analysis identifies different binding modes in droplet-forming and amyloid-forming proteins, the latter being associated with a greater propensity to switch between binding modes.

Droplet-promoting and amyloid-promoting regions can populate both the droplet and the amyloid states

Droplet-promoting regions are enriched in disorder-promoting residues (serine, glycine and proline) [43] as compared to amyloid-promoting regions, while depleted in order-promoting residues (valine and leucine) (**Figure S2**). This is in accord with previous observations that proline and glycine composition controls the properties of elastomeric and amyloid-like fibrils [44].

The two extreme scenarios leading to the condensation pathway are represented by low complexity regions (LARKs) [25], which have high droplet-promoting propensity (p_{DP}) [27] and low amyloid-promoting propensity (p_{AP}) [35], and by amyloid cores, which have low droplet-promoting propensity (**Figure S3D**) and high amyloid-promoting propensity (**Figure S3C**). These structural elements exhibit contrasting binding modes (p_{DD}) (**Figure 3A**) and binding mode entropy (S_{bind}) (**Figure S3B**). By contrast, droplet-promoting and amyloid-promoting regions do not exhibit a significant difference in these interaction behaviors (**Figures 3A** and **S3B**) neither in their amyloid-promoting (**Figure S3C**) and droplet-promoting propensities (**Figure S3D**). This analysis indicates that residues in amyloid-promoting regions (**Figure 3B**) and in droplet-promoting regions (**Figure 3C**) can populate both the droplet and the amyloid states. Indeed, residues in amyloid-promoting and droplet-promoting regions were predicted

in both amyloid and droplet states (**Figures 3B, 3C**). This analysis is in accord with structural studies reporting amyloid-like labile β elements in the droplet state [22, 24].

In addition, we found that about half of the amyloid-promoting regions contained short droplet-promoting regions (≥ 10 residues, **Table S2**), and that about half of the droplet-promoting regions contained short amyloid-promoting regions (≥ 6-residues, **Table S2**). The droplet-promoting propensity increased with the distance from the amyloid core to reach a maximum at ~25-35 residues from the core (**Figure S4A,C**), while the amyloid-promoting propensity did not depend considerably from the position of the flanking region (**Figures S4B,D**). Along these lines, we observed a widespread co-occurrence of droplet-promoting and amyloid-promoting regions in different proteomes (*Homo sapiens* 87%, *Saccharomyces cerevisiae* 88%, in *Caenorhabditis elegans* 89% , **Table S3**) based on predictions by the CamSol [35], Tango [20] [45] and FuzDrop methods [27].

Taken together, this computational analysis indicates that, although droplet-promoting regions and amyloid-promoting regions have distinct binding modes, these regions can sample both the droplet and the amyloid states.

A droplet landscape illustrates the propensity for the droplet and amyloid states

The computational analysis reported above indicates that the conversion between the droplet and amyloid states requires a switch from disordered to ordered binding modes involving sequence regions that can sample both states. As illustrated by the case of the prion-like domain of TDP-43 (residues 262-414), residues that drive the condensation pathway can interact in both disordered and ordered binding modes (**Figure 4A**). However, the frequency of disordered interactions, which favor the formation of the droplet state, is higher than those of ordered modes, which promote the stabilization of the amyloid state (**Figure 4A**).

Based on this analysis, we propose to characterize conversion from the droplet state to the amyloid state using a droplet landscape, which is defined by the droplet-promoting propensity (p_{DP}) of the residues and their binding mode entropy (S_{bind}). Residues that are likely remaining in the droplet state are found in the bottom right region of the droplet landscape, since they

have high droplet-promoting propensities and low binding mode entropy (**Figure 4B**). By contrast, residues with high propensity to form the amyloid state are found in the top left region of the droplet landscape, since they have low droplet-promoting propensity and high binding mode entropy (**Figure 4B**) [46]. Most residues of the TDP-43 aggregation hot-spot region (residues 312-342) are located in the top right region of the droplet landscape, indicating that these residues can exhibit both ordered and disordered binding modes, and can therefore facilitate the conversion from the droplet to the amyloid state (**Figure 4B**) [47].

Overall, the droplet landscape quantifies the contribution of different residues to the different components driving the condensation pathway (**Figure 4B**). An analysis of the droplet landscape suggests that residues with a wide range of binding modes can drive the conversion from the droplet state to the amyloid state, while those with strong preference for disordered binding modes tend to remain in the droplet state (**Figure S5**).

Impact of ALS-associated mutations on the condensation pathway

We used these insights to characterise ALS-associated mutations shown to promote aggregation of liquid droplets (**Table S4**): FUS G156E [6], G187S, G225V, G230C, G399V [48], P525L [15]; hnRNPA1 D262V [5], D214V [49]; hnRNPA2 D290V [50]; tau P301L, P301S, A152T [51]; TDP-43 A321V [47], G298S, M337V [52], A315T [53]; TIA1 P262L, A381T, E384K [54]; UBQL2 P506T [55]. As compared to the wild-type proteins, ALS-associated mutations shifted the droplet landscape towards lower droplet-promoting propensity (p_{DP}) and higher binding mode entropy (S_{bind}) (**Figure S6**). Thus, we expressed the change in aggregation propensity within a droplet, Δp_{AD}, as a function of Δp_{AP}, Δp_{DP}, ΔS_{bind} (see Methods), where $\Delta p_{AP} = p_{AP}(A_m) - p_{AP}(A_{wt})$ is the difference in amyloid-promoting propensities between the mutated (A_m) and wild-type (A_{wt}) residues, $\Delta p_{DP} = p_{DP}(A_m) - p_{DP}(A_{wt})$ is the corresponding difference in droplet-promoting propensities, and $\Delta S_{bind} = S_{bind}(A_m) - S_{bind}(A_{wt})$ is the corresponding difference in binding mode entropy (Methods). Using recovery rates of non-ALS associated FUS mutants (G154E, G156D, G156Q, G156P) and the ALS mutant G156E [56], we then trained a support vector model (SVM) for Δp_{AD} (see Methods) (**Figure 5A**). We note that some of these mutations, for example G156E, decrease the amyloid-promoting propensity Δp_{AP}, while considerably increasing the binding mode entropy ΔS_{bind} as compared to the wild-type protein. The non-ALS associated G154E mutation has a

considerably lower binding mode entropy than G156E and does not promote FUS aggregation. We then applied this model to predict the impact of the ALS-associated mutations on the aggregation propensity within droplets. For all the ALS-associated mutations we found high predicted aggregation propensity as compared to the non-ALS associated FUS mutations, which were experimentally shown not to promote aggregation (**Figure 5B**) [56].

Thus, we propose that using three sequence-based features (droplet-promoting propensity, aggregation-promoting propensity and binding mode entropy) it is possible to predict the impact of a mutation on the condensation pathway. Systematic studies on aggregation-promoting mutations under consistent experimental conditions would be required for robust parametrisation. In particular more mutations, which do not induce amyloid formation will need to be analysed.

Discussion and Conclusions

Most proteins are capable of converting into the amyloid state when their concentration exceeds their solubility limits [57], which can take place via disordered oligomers through the deposition pathway [3, 4]. It has been recently proposed that the droplet state is also widespread and likely to be accessible to most proteins [27]. Although the droplets formed through liquid-liquid phase separation can be functional, with new biological roles being discovered with rapid pace [8], they can also convert to amyloid state through the condensation pathway [5-7, 51].

The possibility for proteins of populating different states creates a challenge for the protein homeostasis system, since dysregulated transitions into dysfunctional assemblies can initiate pathological processes [3, 12-14]. It is therefore important to understand the sequence determinants driving liquid droplets towards solid-like aggregates. Our computational analysis indicates that that droplet-forming and amyloid-forming regions tend to exhibit different binding modes, while regions promoting both states tend to sample multiple binding modes. Along these lines, our study suggests that the condensation pathway is driven by residues that

are capable of multiple binding modes, which are often associated with conformationally-heterogeneous structural elements [26, 47, 58].

We characterised the aggregation-promoting propensity of residues via the condensation pathway through a droplet landscape approach. As illustrated in the case of the prion-like domain of TDP-43, residues that tend to form droplets as they have high droplet-promoting propensity, and have high binding mode entropy are likely convert to amyloids, while those with strong preference for disordered interactions tend to remain in the droplet state. This analysis indicates that the structural ordering towards the amyloid state depends on the binding modes of the proteins in the condensates. In case of homotypic interactions (low binding mode entropy), the condensates will mature into the amyloid state, whereas in case of heterotypic (high binding mode entropy) interactions aggregation will be slowed down. The latter interactions, however, may also lead to promiscuity and cause cytotoxicity [59].

Indeed, most ALS-associated mutations experimentally shown to promote the condensation pathway also tend to increase multimodal interactions. This analysis may explain why amyloid-like structural elements exhibiting ordered interactions are also observed in droplets [22, 24], where disordered interactions dominate [60, 61]. We have also discussed how the aggregation propensity within droplets of a protein can be estimated from its amino acid sequence based on the droplet landscape and amyloid-promoting propensity. Based on this approach, we have proposed that ALS-associated mutations, including FUS G156E, do not necessarily reduce solubility, but expand interaction multimodality and facilitate the conversion towards ordered interactions. This behavior is in contrast to FUS G154E, which is less prone to increase interaction multimodality and does not promote aggregation.

Recent experimental results indicate that biophysical properties of droplets are evolved for optimal physiological function, while disease-causing mutations lead to aberrant condensates [62], and the normal condensate properties can be restored by unrelated droplet-promoting sequences [62]. Non-specific interactions between FUS ALS mutants and nuclear import receptors were also capable of restoring normal FUS condensate properties [48]. Although droplet-related pathologies are due to different molecular mechanisms [12], these observations indicate that familial mutations interfere with generic interactions [63]. This is

consistent with earlier computational studies [27] and the analysis in this work indicates that the impact of a mutation on droplet aggregation can be estimated without specific sequence motifs.

Taken together, our computational analysis suggests an amino acid code based on three features (droplet-promoting propensity, aggregation-promoting propensity and binding mode entropy) that drive the transition of liquid-like condensates into amyloid deposits, and can be used to rationalise the impact of disease-related mutations on this transition.

Methods

Datasets of droplet-forming proteins. Spontaneously phase-separating proteins were assembled from three public databases (**Table S1**): the PhaSepDB Reviewed dataset based on curated literature search (http:/db.phasep.pro) [28], the PhaSePro database with regions mediating phase separation identified (https://phasepro.elte.hu) [29], and the LLPSDB database (http://bio-comp.org.cn/llpsdb) [30] with experimental conditions specified for *in vitro* liquid-liquid phase separation. Merging these datasets resulted in 426 droplet-driver proteins, which can spontaneously phase separate (DROP-D, **Table S1**) [64]. Droplet-client proteins, which phase separate via interactions with other partners, were assembled from the LLPSDB dataset (http://bio-comp.org.cn/llpsdb) (DROP-C, **Table S1**). 20 proteins with evidence for both droplet and amyloid formation based on the AmyPro database (http://amypro.net) [32] were defined as droplet-amyloids (DROP-AMYLOID, **Table S1**).

Droplet-promoting regions were assembled from the PhaSePro dataset, and were grouped based on the evidence for spontaneous, or partner-assisted phase separation in the LLPSDB dataset (DPR, **Table S1**). Low-complexity aromatic-rich kinked segments (LARKs) were used as a reference for droplet-promoting regions [25]. The LARK dataset contained 6075 6-residue long segments in 400 proteins, as in the original reference.

Datasets of amyloid-forming proteins. 120 amyloid-forming proteins were assembled from the AmyPro database (http://amypro.net) [32] (AMYLOID, **Table S1**). This dataset contained

only proteins with UniProt identifiers, with peptides excluded. Droplet-amyloid proteins were also eliminated from the amyloid dataset and were defined as a separate category (**Table S1**). Amyloid-forming proteins were annotated as pathological and functional based on the AmyPro database (**Table S1**). 46 yeast proteins with heritable traits in *Saccharomyces cerevisiae* were defined as yeast prions (YPRION, **Table S1**) [34]. Most yeast prions had Hsp104 or Hsp70 dependence [34]. Amyloid-promoting regions (APRs, **Table S1**) were defined based on the AmyPro database, including both the cross-β spine forming amyloid cores and the flanking regions. Amyloid-promoting regions were defined based on the AmyPro database, including both the cross-β spine forming amyloid cores and the flanking regions.

Disorder-to-disorder binding profiles. The disorder-to-disorder binding profile, p_{DD}, is estimated for each amino A_i as [42]

$$p_{DD}(A_i) = median[\pi_{DD}(R_i)]_N \quad (1)$$

where $\pi_{DD}(R_i)$ is the probability of disorder-to-disorder binding of R_i, a region of 5-9 residues around A_i, and N is the number of possible regions R_i. We refer to $\pi_{DD}(R_i)$ as the 'binding mode probability' of region R_i because a value of 0 indicates a binding from fully disordered to fully ordered states, and a value of 1 indicates a binding from fully disordered to fully disordered states. p_{DD} was computed using FuzPred [42].

Binding mode entropy profile. The binding mode entropy profile (S_{bind}) quantifies the variability of binding modes at the amino acid level. To define S_{bind} we start by defining the frequency f of different possible binding modes for an amino acid A_i as

$$f[\pi_{DD}(R_i)] = \frac{n_R[\pi_{DD}(R_i)]}{N} \quad (2)$$

To calculate f, the binding modes are divided into discrete bins (usually 10), and n_R is the number of binding regions within a given bin. The binding mode entropy profile (S_{bind}) is then defined as the entropy of the frequencies of f [41]

$$S_{bind}(A_i) = \sum f[\pi_{DD}(R_i)] \log_2 f[\pi_{DD}(R_i)] \quad (3)$$

where the summation is over the $[\pi_{DD}(R_i)]$ bins.

Droplet-promoting propensity profiles. The droplet-promoting propensity profile p_{DP} quantifies the probability of spontaneous liquid-liquid phase separation [27]

$$p_{DP}(A_i) = \frac{exp\ F_S(A_i)}{1 + exp\ F_S(A_i)} \quad (4)$$

where F_S (A_i) is the scoring function for residue A_i

$$F_S(A_i) = \lambda_1 p_D(A_i) + \lambda_2 p_{DD}(A_i) + \gamma \quad (5)$$

In this equation, $p_D(A_i)$ is the probability for the disorder in the free state, and $p_{DD}(A_i)$ is the probability for disordered binding. $p_D(A_i)$ approximates the conformational entropy in the unbound form, while $p_{DD}(A_i)$ estimates the conformational entropy of binding. λ_1 and λ_2 are the linear coefficients of the predictor variables and γ is a scalar constant (intercept), which were determined using the binary logistic model [27]. p_D was derived from the disorder score as computed using the Espritz NMR algorithm [65]. The p_{DD} values were predicted by FuzPred [42]. p_{DP} =0.60 is the threshold that a residue is involved in spontaneous phase separation.

Amyloid-promoting propensity profiles. The amyloid-promoting propensity profile p_{AP} of a protein expresses its sequence-dependent probability to aggregate. Amyloid-promoting propensity profiles were obtained by the solubility profiles obtained by CamSol (p_{CS}) [35]. p_{AP} = -p_{CS} and p_{AP} = 0.90 is the threshold, above which a protein likely aggregates [66]. The parameters *b* and *c* were determined by the R program.

Prediction of the propensity of aggregation within droplets (Δp_{AD}). To calculate Δp_{AD} we expressed it as a function of sequence-based features

$$\Delta p_{AD} = \mathrm{F}(\Delta p_{AP}, \Delta p_{DP}, \Delta S_{bind}) \quad (6)$$

where $\Delta p_{AP} = p_{AP}(A_m) - p_{AP}(A_{wt})$ is the difference in amyloid-promoting propensities, $\Delta p_{DP} = p_{DP}(A_m) - p_{DP}(A_{wt})$ is the difference in droplet-promoting propensities, and $\Delta S_{bind} = S_{bind}(A_m) - S_{bind}(A_{wt})$ is the difference in binding mode entropy between the mutated (A_m) and wild-type (A_{wt}) residues. To define the function F in Eq. (6), we trained a support vector machine (SVM) model to fit the aggregation rates derived from experimental data. We used the recovery rates (RR) of FUS mutations [56]: ALS-associated G156E, and non-ALS mutations G156D, G156Q, G156E, G156P and defined aggregation rates as 1-RR. We thus obtained a Pearson correlation coefficient r=0.943, p=0.016 (**Figure 5A**). Then we applied this SVM model on the predicted the Δp_{AP}, Δp_{DP}, ΔS_{bind} values of ALS-associated mutations of FUS, hnRNPA1, hnRNPA2, tau, TDP-43, TIA1 and UBQLN2 proteins (**Table S4**).

Acknowledgements

M.F. acknowledges the financial support of HAS-11015 and GINOP-2.3.2-15-2016-00044.

References

[1] Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017;18:285-98.

[2] Hyman AA, Weber CA, Julicher F. Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol. 2014;30:39-58.

[3] Knowles TP, Vendruscolo M, Dobson CM. The amyloid state and its association with protein misfolding diseases. Nat Rev Mol Cell Biol. 2014;15:384-96.

[4] Michaels TCT, Saric A, Curk S, Bernfur K, Arosio P, Meisl G, et al. Dynamics of oligomer populations formed during the aggregation of Alzheimer's Abeta42 peptide. Nat Chem. 2020.

[5] Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell. 2015;163:123-33.

[6] Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell. 2015;162:1066-77.

[7] Hardenberg MC, Sinnige T, Casford S, Dada S, Poudel C, Robinson EA, et al. Observation of an alpha-synuclein liquid droplet state and its maturation into Lewy body-like assemblies. J Mol Cell Biol. 2021.

[8] Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. 2018;28:420-35.

[9] Lyon AS, Peeples WB, Rosen MK. A framework for understanding the functions of biomolecular condensates across scales. Nat Rev Mol Cell Biol. 2020.

[10] Patel A, Malinovska L, Saha S, Wang J, Alberti S, Krishnan Y, et al. ATP as a biological hydrotrope. Science. 2017;356:753-6.

[11] Schaefer KN, Peifer M. Wnt/Beta-Catenin Signaling Regulation and a Role for Biomolecular Condensates. Developmental cell. 2019;48:429-44.

[12] Alberti S, Hyman AA. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. Nat Rev Mol Cell Biol. 2021.

[13] Alberti S, Dormann D. Liquid-Liquid Phase Separation in Disease. Annu Rev Genet. 2019;53:171-94.

[14] Mathieu C, Pappu RV, Taylor JP. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. Science. 2020;370:56-60.

[15] Murakami T, Qamar S, Lin JQ, Schierle GS, Rees E, Miyashita A, et al. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron. 2015;88:678-90.

[16] Schmidt HB, Gorlich D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. eLife. 2015;4.

[17] Tan PS, Aramburu IV, Mercadante D, Tyagi S, Chowdhury A, Spitz D, et al. Two Differential Binding Mechanisms of FG-Nucleoporins and Nuclear Transport Receptors. Cell reports. 2018;22:3660-71.

[18] Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. Curr Opin Struct Biol. 2019;58:88-96.

[19] Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018;174:688-99 e16.

[20] Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol. 2004;342:345-53.

[21] Tartaglia GG, Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities. Chem Soc Rev. 2008;37:1395-401.

[22] Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149:753-67.

[23] Lu J, Cao Q, Hughes MP, Sawaya MR, Boyer DR, Cascio D, et al. CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. Nature communications. 2020;11:4090.

[24] Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell. 2017;171:615-27 e16.

[25] Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, Cascio D, et al. Atomic structures of low-complexity protein segments reveal kinked beta sheets that assemble networks. Science. 2018;359:698-701.

[26] Bienz M. Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. Cell. 2020;182:799-811.

[27] Hardenberg M, Horvath A, Ambrus V, Fuxreiter M, Vendruscolo M. Widespread occurrence of the droplet state of proteins in the human proteome. Proc Natl Acad Sci U S A. 2020;117:33254-62.

[28] You K, Huang Q, Yu C, Shen B, Sevilla C, Shi M, et al. PhaSepDB: a database of liquid-liquid phase separation related proteins. Nucleic Acids Res. 2020;48:D354-D9.

[29] Meszaros B, Erdos G, Szabo B, Schad E, Tantos A, Abukhairan R, et al. PhaSePro: the database of proteins driving liquid-liquid phase separation. Nucleic Acids Res. 2020;48:D360-D7.

[30] Li Q, Peng X, Li Y, Tang W, Zhu J, Huang J, et al. LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. Nucleic Acids Res. 2020;48:D320-D7.

[31] Li Q, Peng X, Li Y, Tang W, Zhu Ja, Huang J, et al. LLPSDB: a database of proteins undergoing liquid–liquid phase separation in vitro. Nucleic Acids Res. 2020;48:D320-D7.

[32] Varadi M, De Baets G, Vranken WF, Tompa P, Pancsa R. AmyPro: a database of proteins with validated amyloidogenic regions. Nucleic Acids Res. 2018;46:D387-D92.

[33] Wu H, Fuxreiter M. The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. Cell. 2016;165:1055-66.

[34] Chakrabortee S, Byers JS, Jones S, Garcia DM, Bhullar B, Chang A, et al. Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. Cell. 2016;167:369-81 e12.

[35] Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol. 2015;427:478-90.

[36] Fuxreiter M. Classifying the Binding Modes of Disordered Proteins. Int J Mol Sci. 2020;21.

[37] Gianni S, Freiberger M, Jemth P, Ferreiro D, Wolynes P, Fuxreiter M. Fuzziness and Frustration in the Energy Landscape of Protein Folding, Function, and Assembly. Acc Chem Res. 2021.

[38] Stamos JL, Chu ML, Enos MD, Shah N, Weis WI. Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. eLife. 2014;3:e01998.

[39] Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274:948-53.

[40] Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. 2018;555:61-6.

[41] Horvath A, Miskei M, Ambrus V, Vendruscolo M, Fuxreiter M. Sequence-based prediction of protein binding mode landscapes PLoS Comp Biol. 2020;16:e1007864.

[42] Miskei M, Horvath A, Vendruscolo M, Fuxreiter M. Sequence-Based Prediction of Fuzzy Protein Interactions. J Mol Biol. 2020;432:2289-303.

[43] Hatos A, Hajdu-Soltesz B, Monzon AM, Palopoli N, Alvarez L, Aykac-Fas B, et al. DisProt: intrinsic protein disorder annotation in 2020. Nucleic Acids Res. 2020;48:D269-D76.

[44] Rauscher S, Baud S, Miao M, Keeley FW, Pomes R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. Structure. 2006;14:1667-76.

[45] Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins. J Mol Biol. 2004;342:345-53.

[46] Guenther EL, Cao Q, Trinh H, Lu J, Sawaya MR, Cascio D, et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. Nat Struct Mol Biol. 2018;25:463-71.

[47] Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure. 2016;24:1537-49.

[48] Niaki AG, Sarkar J, Cai X, Rhine K, Vidaurre V, Guy B, et al. Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations. Mol Cell. 2020;77:82-94 e4.

[49] Gui X, Luo F, Li Y, Zhou H, Qin Z, Liu Z, et al. Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. Nature communications. 2019;10:2006.

[50] Murray DT, Zhou X, Kato M, Xiang S, Tycko R, McKnight SL. Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. Proc Natl Acad Sci U S A. 2018;115:E9782-E91.

[51] Wegmann S, Eftekharzadeh B, Tepper K, Zoltowska KM, Bennett RE, Dujardin S, et al. Tau protein liquid-liquid phase separation can initiate tau aggregation. EMBO J. 2018;37.

[52] Gopal PP, Nirschl JJ, Klinman E, Holzbaur EL. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. Proc Natl Acad Sci U S A. 2017;114:E2466-E75.

[53] French RL, Grese ZR, Aligireddy H, Dhavale DD, Reeb AN, Kedia N, et al. Detection of TAR DNA-binding protein 43 (TDP-43) oligomers as initial intermediate species during aggregate formation. J Biol Chem. 2019;294:6696-709.

[54] Mackenzie IR, Nicholson AM, Sarkar M, Messing J, Purice MD, Pottier C, et al. TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. Neuron. 2017;95:808-16 e9.

[55] Sharkey LM, Sandoval-Pistorius SS, Moore SJ, Gerson JE, Komlo R, Fischer S, et al. Modeling UBQLN2-mediated neurodegenerative disease in mice: Shared and divergent properties of wild type and mutant UBQLN2 in phase separation, subcellular localization, altered proteostasis pathways, and selective cytotoxicity. Neurobiol Dis. 2020;143:105016.

[56] Rhine K, Makurath MA, Liu J, Skanchy S, Lopez C, Catalan KF, et al. ALS/FTLD-Linked Mutations in FUS Glycine Residues Cause Accelerated Gelation and Reduced Interactions with Wild-Type FUS. Mol Cell. 2020;80:666-81 e8.

[57] Vecchi G, Sormanni P, Mannini B, Vandelli A, Tartaglia GG, Dobson CM, et al. Proteome-wide observation of the phenomenon of life on the edge of solubility. Proc Natl Acad Sci U S A. 2020;117:1015-20.

[58] Kato M, McKnight SL. A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. Annu Rev Biochem. 2018;87:351-90.

[59] Vendruscolo M, Fuxreiter M. Sequence determinants for the aggregation and cytotoxicity of proteins within condensates generated by liquid-liquid phase separation. BioRxiv. 2020.

[60] Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J, et al. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. Nat Struct Mol Biol. 2019;26:637-48.

[61] Reichheld SE, Muiznieks LD, Keeley FW, Sharpe S. Direct observation of structure and dynamics during phase separation of an elastomeric protein. Proc Natl Acad Sci U S A. 2017;114:E4408-E15.

[62] Li W, Hu J, Shi B, Palomba F, Digman MA, Gratton E, et al. Biophysical properties of AKAP95 protein condensates regulate splicing and tumorigenesis. Nature cell biology. 2020;22:960-72.

[63] Fuxreiter M. Spot in a drop: mutations in aberrant condensates. Nat Rev Mol Cell Biol. 2021.

[64] Hardenberg M, Horvath A, Ambrus V, Fuxreiter M, Vendruscolo M. Widespread occurrence of the droplet state of proteins in the human proteome. Proc Natl Acad Sci USA. 2020;117:33254-62.

[65] Walsh I, Martin AJ, Di Domenico T, Tosatto SC. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics. 2012;28:503-9.

[66] Sormanni P, Vendruscolo M. Protein Solubility Predictions Using the CamSol Method in the Study of Protein Homeostasis. Cold Spring Harb Perspect Biol. 2019;11.

[67] Benilova I, Karran E, De Strooper B. The toxic Abeta oligomer and Alzheimer's disease: an emperor in need of clothes. Nat Neurosci. 2012;15:349-57.

[68] Haass C, Selkoe DJ. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nat Rev Mol Cell Biol. 2007;8:101-12.

[69] Luheshi LM, Tartaglia GG, Brorsson AC, Pawar AP, Watson IE, Chiti F, et al. Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. PLoS Biol. 2007;5:e290.

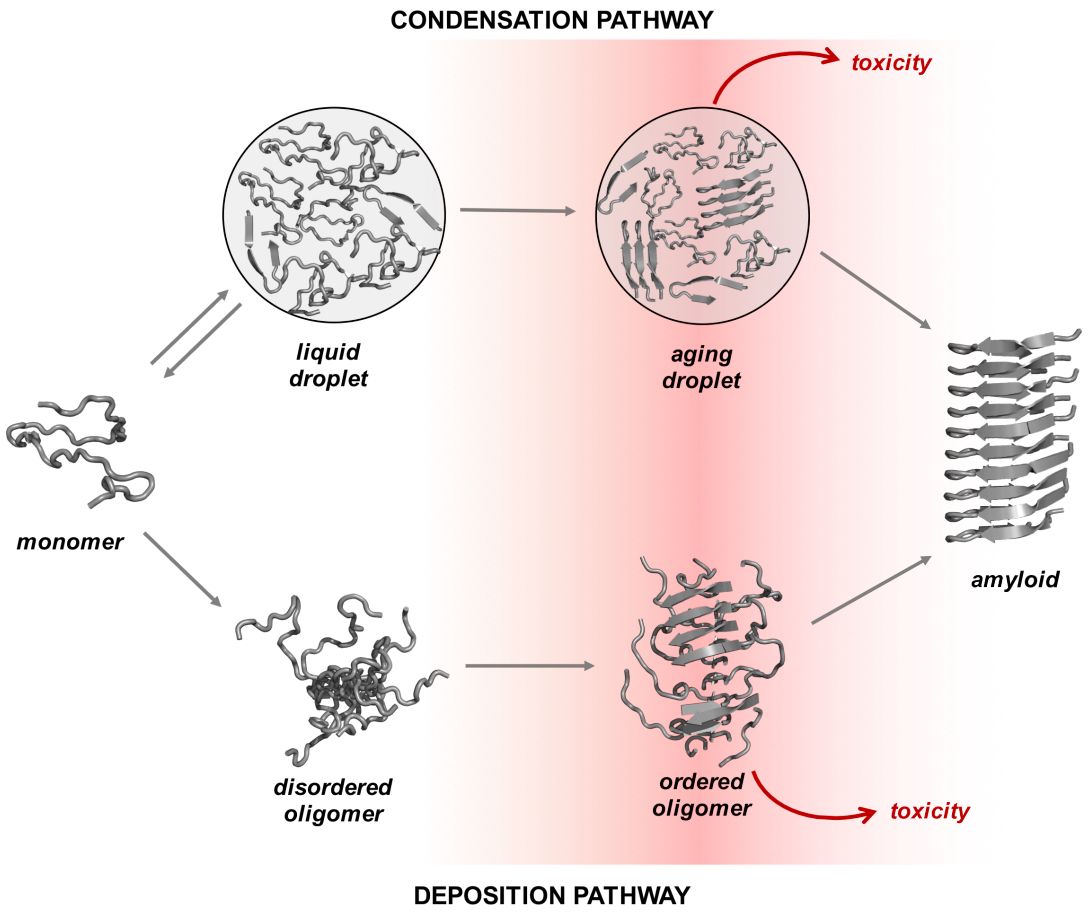

Figure 1. Overview of the condensation and the deposition pathways for amyloid formation. Along the deposition pathway, proteins move from the native state to the amyloid state through increasingly ordered oligomer aggregates [3, 4]. Some of these oligomers are highly cytotoxic [67, 68] according to an amino acid code that has been partly elucidated [69]. Along the condensation pathway, proteins that undergo condensation convert from the native state to the amyloid state through a dense liquid state (the droplet state). The stability of these different states, as well as the conversion rates between them, are different for different proteins. For many proteins under cellular conditions, the native and droplet states are metastable, and the conversion rates between the droplet and fibrillar states are fast. For certain proteins, the droplet state is functional, and it is stabilized by extrinsic factors, such as RNA and post-translational modifications. The maturation of droplets into the amyloid state creates cytotoxic intermediates [6, 13, 15, 16].

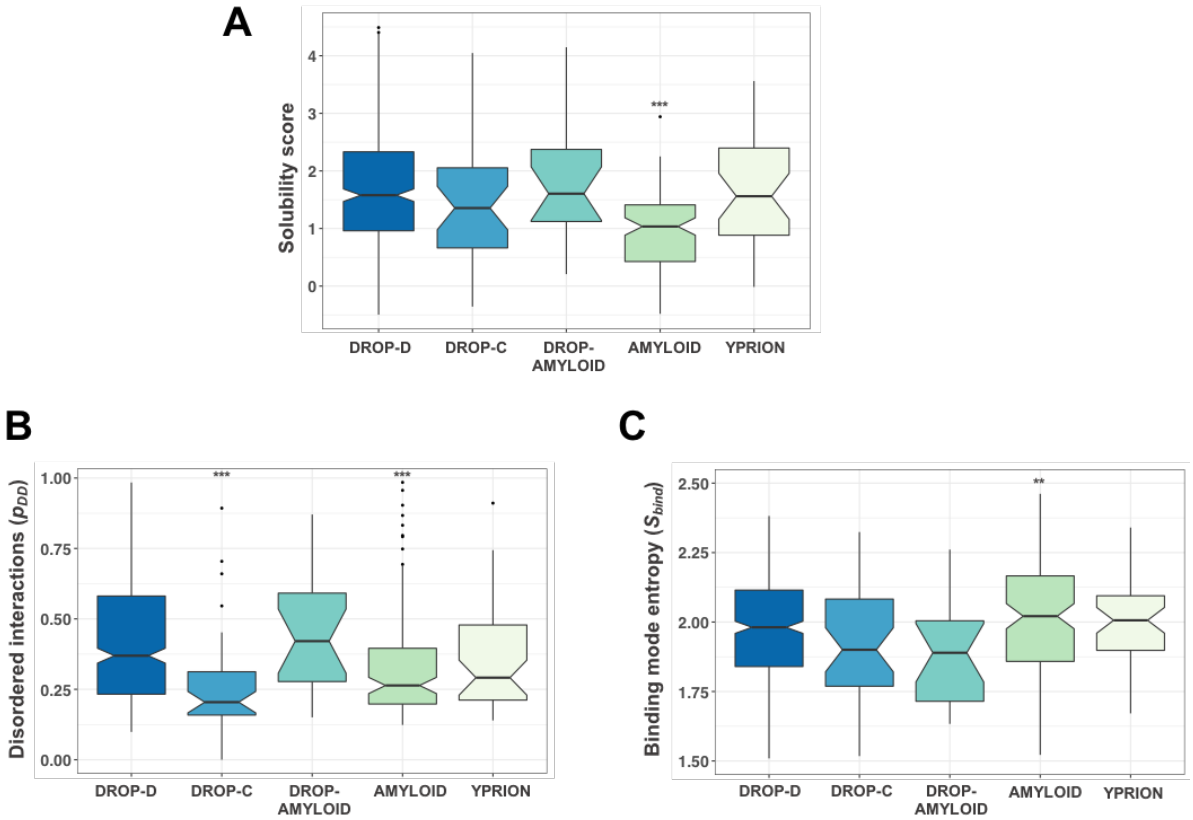

Figure 2. Comparison of solubility and interactions of amyloid-forming and droplet-forming proteins. (A) Comparison of the solubility scores of different datasets of droplet and amyloid proteins. Droplet-driver proteins (DROP-D, blue, **Table S1**), which spontaneously undergo liquid-liquid phase separation [27] and droplet-amyloid proteins (DROP-AMYLOID, aquamarine, **Table S1**), which exhibit both droplet and amyloid behaviors, have the highest solubility scores by CamSol [35], while amyloid-forming proteins (AMYLOID, green) in the AmyPro database (http://amypro.net) [32] have the lowest solubility scores. Droplet-client proteins (DROP-C, light blue, **Table S1**), which require partner-interactions for droplet formation, and yeast prions (YPRION, light green), with inheritable traits in *Saccharomyces cerevisiae* [34], have intermediate solubility scores. Statistical significances by Mann-Whitney test (***$p< 10^{-3}$) show that amyloid-forming proteins have significantly lower solubility scores than droplet-forming proteins, while other classes sample a continuum between these two extremes. **(B) Comparison of the disordered binding modes (p_{DD}) of droplet and amyloid proteins.** Droplet-driver and amyloid-forming proteins exhibit distinct interaction behaviors. FuzPred [42] predicts disordered binding modes (Methods, Eq. 1) for droplet-driver proteins, and ordered binding modes for amyloid proteins. Droplet-amyloid proteins have disordered

interaction behaviors as droplet-drivers, while yeast amyloids and droplet-clients have ordered binding modes as amyloids. Statistical significances by Mann-Whitney test (***$p < 10^{-3}$) show that amyloid-forming proteins and droplet-client proteins exhibit significantly more ordered interactions than droplet-forming proteins, while yeast prions and droplet amyloids sample both interaction behaviors. **(C) Comparison of the binding mode entropy** (S_{bind}) **of droplet-forming and amyloid-forming proteins.** Amyloid-forming proteins exhibit high binding mode entropy, as they can sample both disordered and ordered interactions. Droplet and droplet-amyloid proteins have a strong preference for disordered binding (low S_{bind}). S_{bind} values were computed by FuzPred [41] (Methods, Eq. 3). Statistical significances by Mann-Whitney test (** $p < 10^{-2}$) show that amyloid-forming proteins are more prone to multimodal interactions than droplet-forming proteins, while the other classes sample a similar range of binding entropies. Statistical significances as compared to droplet driver proteins were computed by Mann-Whitney test using the R program.

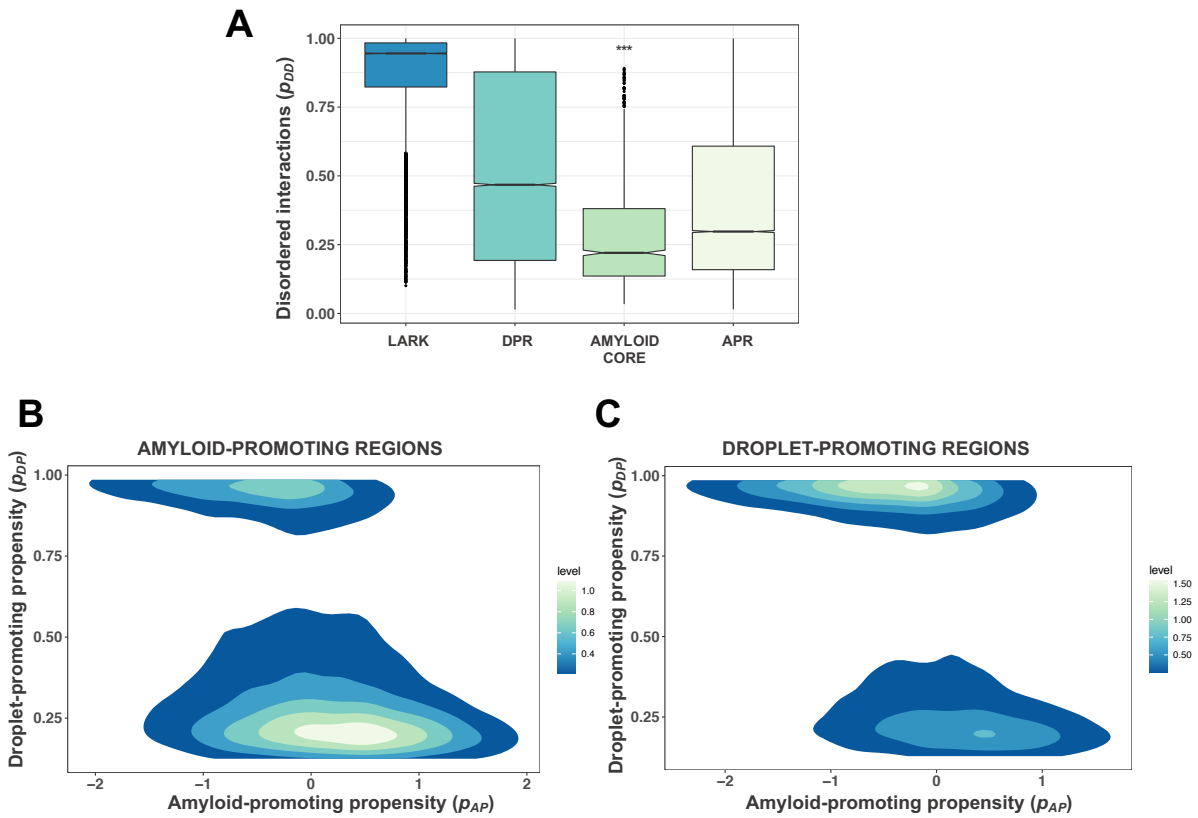

Figure 3. Experimentally-observed droplet-promoting and amyloid-promoting regions can sample both states. (A) Comparison of the disordered binding modes (p_{DD}) of droplet-promoting (DPR, aquamarine) **and amyloid-promoting** (APR, light green) **regions.** Interactions of droplet-promoting and amyloid-promoting regions (**Table S1**) vary in wide range between these disordered and ordered binding modes, exemplified by LARKs [25] (blue) and amyloid cores [32] (green). The statistical significance in reference to droplet-promoting regions, as computed by Mann-Whitney test using the R program (*** $p < 10^{-3}$), demonstrates that interaction properties of droplet-promoting and amyloid-promoting regions overlap. **(B,C) Comparison of the droplet-promoting propensity** (p_{DP}, y-axis) **and amyloid-promoting propensity** (p_{AP}, x-axis) **of amyloid-promoting (B) and droplet-promoting regions (C).** Droplet-promoting propensities were computed by FuzDrop [27] and amyloid-promoting propensities were derived from the solubility scores by CamSol [35]. The amyloid state is characterised by low p_{DP} and positive p_{AP} values, while the droplet state has high p_{DP} and negative p_{AP} values. Residues in amyloid-promoting regions can also populate the droplet state, and residues in droplet-promoting regions can also sample the amyloid state.

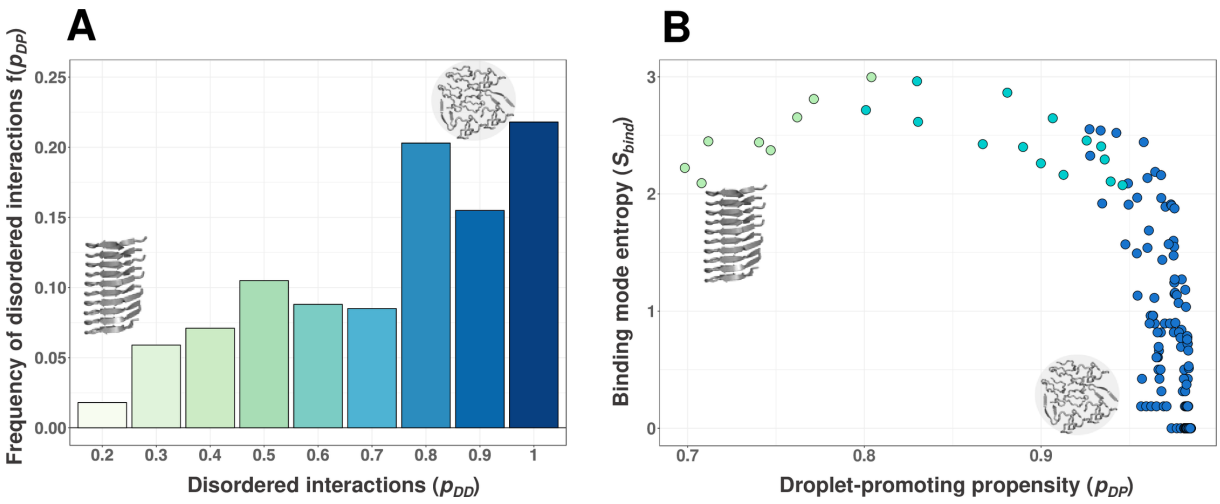

Figure 4. Droplet landscape of the prion-like domain of TDP-43. (A) Frequency of different binding modes, f(p_{DD}), for the residues in the aggregation hot-spot (residues 312-342) of TDP-43. The high values of f(p_{DD}) for large p_{DD} values indicate that the prion-like domain of TDP-43 preferentially samples disordered interactions corresponding to the droplet state (blue, right) [42], but also that it can form ordered interactions corresponding to the amyloid state (green, left). **(B) Droplet landscape of the TDP-43 prion-like domain (residues 262-414).** The droplet-promoting propensity (p_{DP}) is shown on the x-axis [27], and the binding mode entropy (S_{bind}) is shown on the y axis; S_{bind} is computed for each residue from the frequencies displayed in panel **(A)** (Eq. 3) [41]. Residues that tend to remain in the droplet state are found in the bottom right regions (blue), since they have high p_{DP} and low S_{bind} values, reflecting a strong preference for disordered interactions in the droplet state. Residues that are likely to convert to the amyloid stat, such as the amyloid-core of TDP-43 (321-330 residues, light green), are found in the upper left region (green), since they have low p_{DP} and high S_{bind} values, reflecting a tendency to change binding modes. Residues in the hot-spot region (aquamarine) have high p_{DP} and S_{bind} values, and therefore can facilitate the switch between disordered and ordered binding modes.

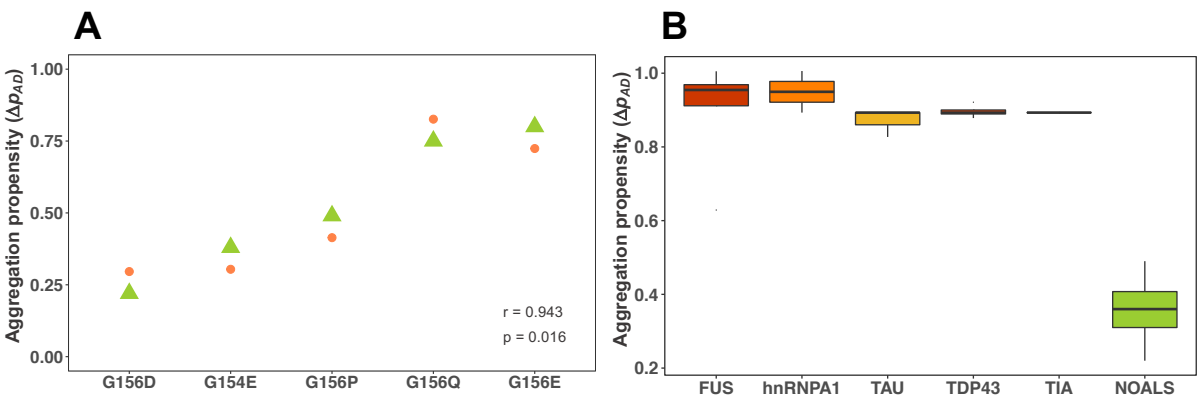

Figure 5. Prediction of the change of aggregation propensity within droplets (Δp_{AD}) for ALS-associated mutations. (A) Comparison of the change upon mutation of experimental (green) and predicted (orange) aggregation propensity within droplets of FUS mutants. The experimental change in aggregation propensity was derived from reference [56]. The predicted change in aggregation propensity (Δp_{AD}) was predicted using a SVM model with three features (change in droplet-promoting propensity, Δp_{AP}, change in aggregation-promoting propensity Δp_{DP}, and change binding mode entropy ΔS_{bind}) that can be calculated from the sequence of a protein (Methods). **(B) ALS-associated mutations of FUS, hnRNPA1, tau, TDP-43, TIA1 have higher predicted change in propensity of aggregation within droplets than mutations not associated with ALS.** ALS-associated single mutations (orange) were shown to promote aggregation (FUS G156E [6], G187S, G225V, G230C, G399V [48], P525L [15]; hnRNPA1 D262V [5], D214V [49]; hnRNPA2 D290V [50]; tau P301L, P301S, A152T [51]; TDP-43 A321V [47], G298S, M337V [52], A315T [53]; TIA1 P262L, A381T, E384K [54]; UBQL2 P506T [55]) (**Table S4**) are predicted to induce large changes in propensity of aggregation within droplets than non-ALS associated FUS mutations G145E, G156D, G156P, G156A (green). Only proteins with multiple mutations are shown for clarity.

Supplementary Information

Table S1. Datasets of droplet-driver, droplet-client, droplet-amyloid, amyloid, and yeast prion proteins and datasets of droplet-promoting and amyloid-promoting regions.
Columns for droplet-driver, droplet-client, droplet-amyloid, amyloid, and yeast prion proteins: UniProt ID (relevant isoforms listed), UniProt entry name, protein name, gene names, length, organism, subcellular location, function, involvement in disease.
Columns for droplet-promoting regions: UniProt ID (relevant isoforms listed), UniProt entry name, protein name, length, region, organism, subcellular location, function, involvement in disease, PDB structure codes. Driver and client roles are indicated when information is available. *Columns for amyloid-promoting regions*: UniProt ID (relevant isoforms listed), function category (pathogenic, functional, not known), core region, amyloid region, UniProt entry name, protein name, length, organism, subcellular location, function, involvement in disease, PDB structure codes.

Table S2. Prediction of droplet-promoting and amyloid-promoting regions in droplet-promoting and amyloid-promoting regions; and in droplet-driver, droplet-client, amyloid and droplet-amyloid proteins. The number and positions of short (≥ 10 residues) and long (≥ 25 residues) droplet-promoting regions are shown. Amyloid-promoting regions (≥ residues) were identified using CamSol [35] and Tango [20].

Table S3. Prediction of droplet-promoting and amyloid-promoting regions in the human, yeast and worm proteomes. The number and positions of short (≥ 10 residues) and long (≥ 25 residues) droplet regions are shown. Amyloid-promoting regions regions (≥ 6 residues) were identified using Tango [20].

Table S4. Single mutations with experimental evidence for promoting aggregation in the droplet state. We considered only mutations that have been directly related to aggregation and not through regulatory proteins, RNA interactions, misfolding or other indirect effects.

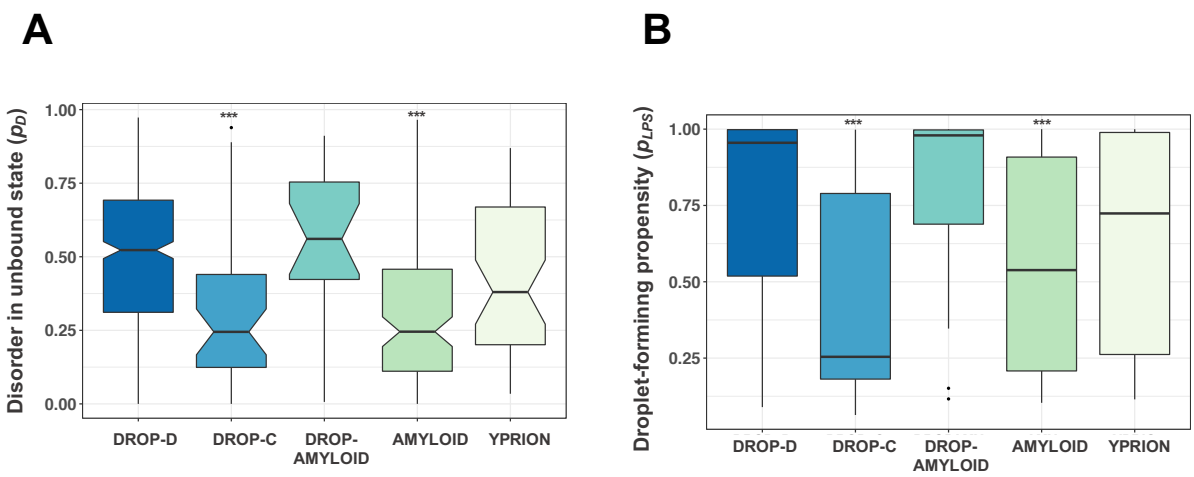

Figure S1. Free-state conformational entropy and droplet-forming propensities of amyloid- and droplet-forming proteins. (A) The free-state disorder propensity (p_D) was estimated as the fraction of disordered residues by Espritz NMR [65]. Droplet-driver proteins (DROP-D, blue, **Table S1**), which spontaneously undergo liquid-liquid phase separation [27] and droplet-amyloid proteins (DROP-AMYLOID, aquamarine, **Table S1**), which exhibit both droplet and amyloid behaviors are the most disordered in the free state. Amyloid-forming (AMYLOID, green [32]) and droplet-client proteins (DROP-C, light blue, **Table S1**) are the most ordered in the free state, while yeast prions (YPRION, light green), with inheritable traits in *Saccharomyces cerevisiae* [34] have intermediate disorder values between the amyloid and droplet states. Statistical significances by Mann-Whitney test as compared to droplet-driver proteins (*** $p< 10^{-3}$) demonstrate that amyloid-forming and droplet-client proteins are more ordered than droplet-forming proteins, while yeast prions and droplet-amyloid proteins have comparable degree of disorder to droplet-forming proteins. **(B)** The droplet-promoting propensity (p_{LLPS}) was estimated by FuzDrop [27]. Droplet-driver (blue) and droplet-amyloid (aquamarine) are predicted to form droplets, whereas droplet-promoting propensities of amyloid-forming (green) and yeast prion (light green) proteins vary in a wide range. Most droplet-client proteins (light blue) are predicted not to form droplets spontaneously. Statistical significances by Mann-Whitney test as compared to droplet-driver proteins (*** $p< 10^{-3}$) demonstrate that amyloid-forming and droplet-client proteins are less prone to form liquid droplets. Statistical significances were computed by the R program.

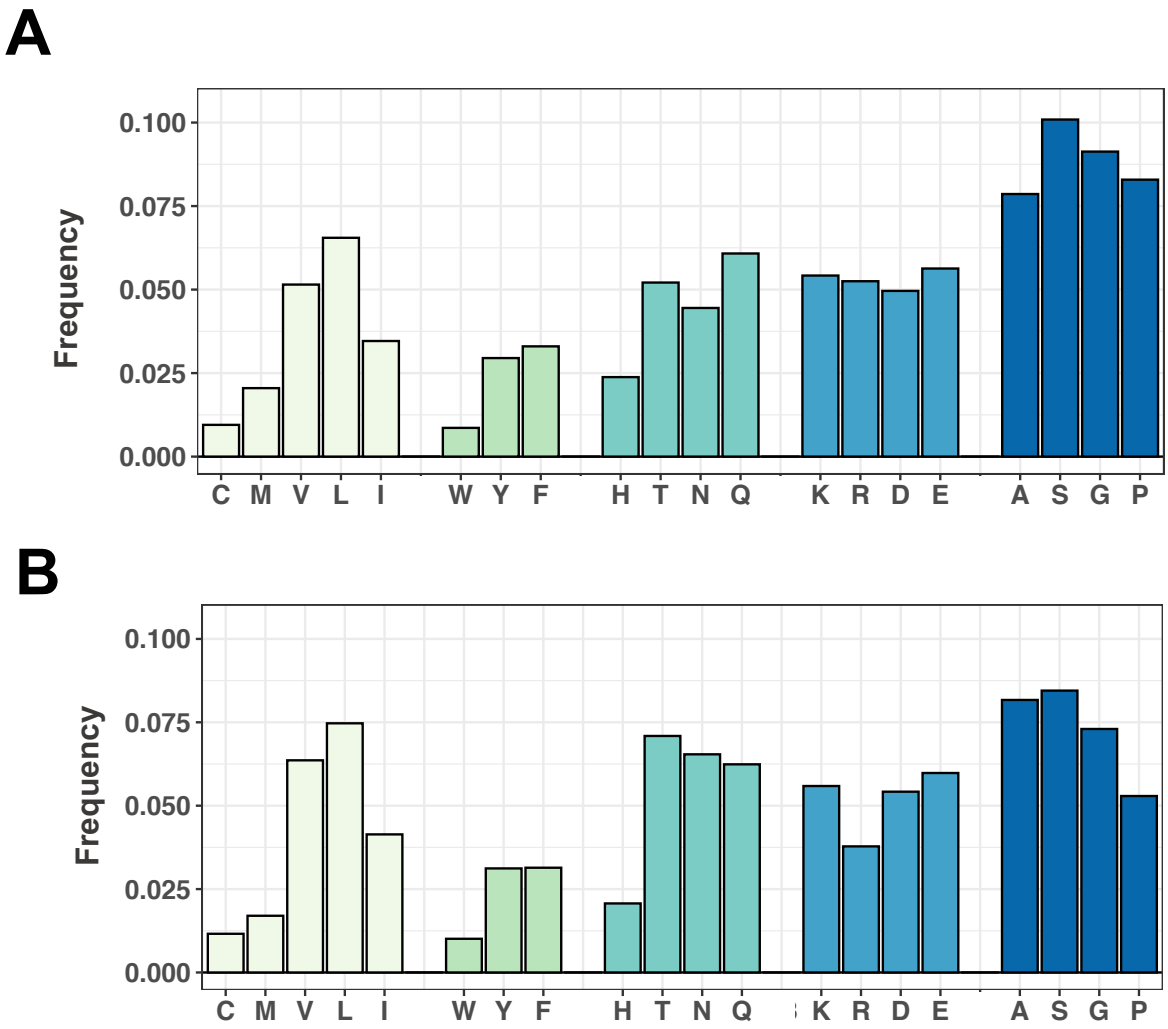

Figure S2. Amino acid composition of droplet-promoting (A) and amyloid-promoting (B) regions. Amino acids grouped as hydrophobic (light green), aromatic (green), hydrophilic (turquoise), charged (steel blue), disorder-promoting (dark blue). Droplet-promoting regions are enriched in disorder-promoting residues (dark blue, [43]) as compared to amyloid-promoting regions, in particular in glycine and proline in accord with previous results [44].

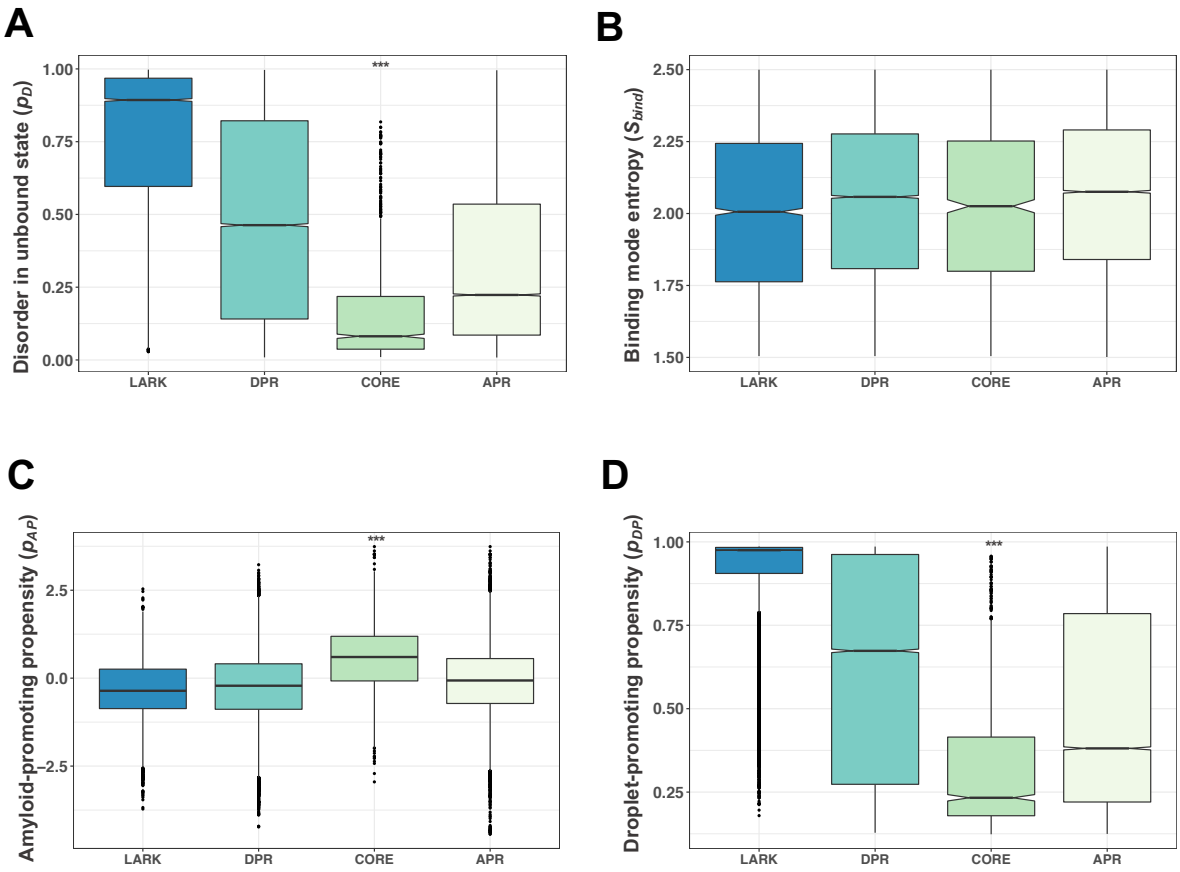

Figure S3. Analysis of the conformational and binding propensities of droplet-promoting regions (DPR, aquamarine) **and amyloid-promoting regions** (APR, light green)**. (A)** Disorder propensity in the free state (p_D). This propensity was estimated by the fraction of disordered residues by Espritz NMR [65]. Both droplet-promoting and amyloid-promoting regions sample a broad range of conformational states. Droplet-promoting regions are less disordered in the free state than LARKs [25] (blue), while amyloid-promoting regions are less ordered than amyloid cores (green). **(B)** Binding mode entropy (S_{bind})**.** This property was defined by the binding mode entropy (S_{bind}) as computed by FuzPred [41] (Methods, Eq. 3). Both droplet-promoting and amyloid-promoting regions exhibit larger variability of binding modes than LARKs and amyloid cores, indicating a possible conversion between ordered and disordered binding modes. **(C)** Amyloid-promoting propensity (p_{AP}). This propensity was derived from the solubility scores by CamSol [35]. Droplet-promoting and amyloid-promoting regions have similar aggregation propensities, which is comparable to LARKs. Amyloid cores have lower p_{AP} values and more prone to aggregate. **(D)** Droplet-promoting propensity (p_{DP}). This propensity was calculated by FuzDrop [27]. This propensity shows a sharp contrast between LARKs and

amyloid-cores, while spontaneous droplet-forming propensities of droplet-promoting and amyloid-promoting regions considerably overlap, indicating that many amyloid-promoting regions also tend to phase separate. Statistical significances by Mann-Whitney test as compared to LARKs (*** $p < 10^{-3}$) demonstrate that while amyloid cores deviate from labile beta elements, amyloid-promoting regions and droplet-promoting regions have comparable properties.

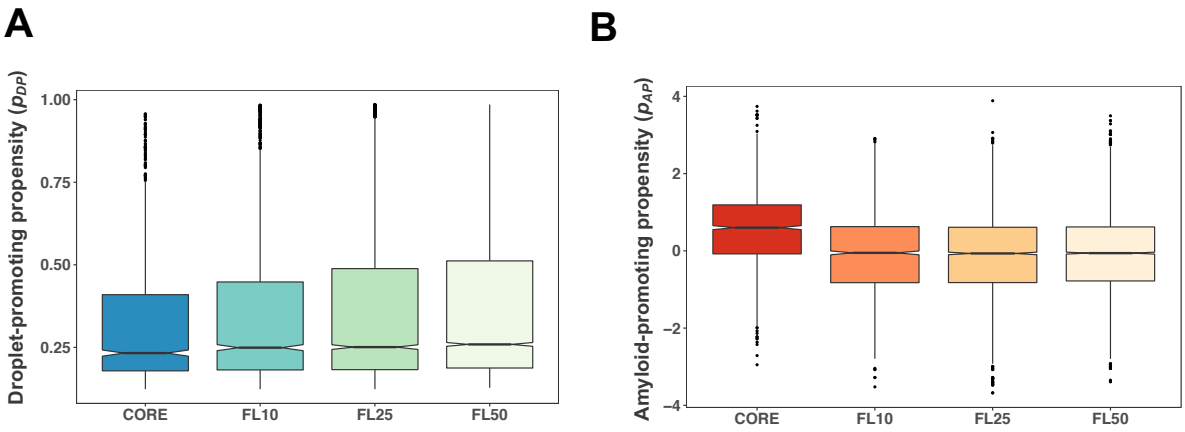

Figure S4. Comparison of the droplet-promoting propensity (p_{DP}) **and amyloid-promoting propensity** (p_{AP}) **of amyloid cores and their flanking regions. (A,B)** p_{DP} **and** p_{AP} **as a function of the length of the flanking region.** p_{DP} and p_{AP} values were evaluated in 10-residue (FL10), 25-residue (FL25) and 50-residue (FL50) flanking regions. The droplet-promoting propensity increases with the size of the flanking region. By contrast, the amyloid-promoting propensity does not change considerably with the size of the flanking region. Amyloid-promoting propensities were derived from the CamSol solubility scores [35], and droplet-promoting propensities were computed by FuzDrop [27].

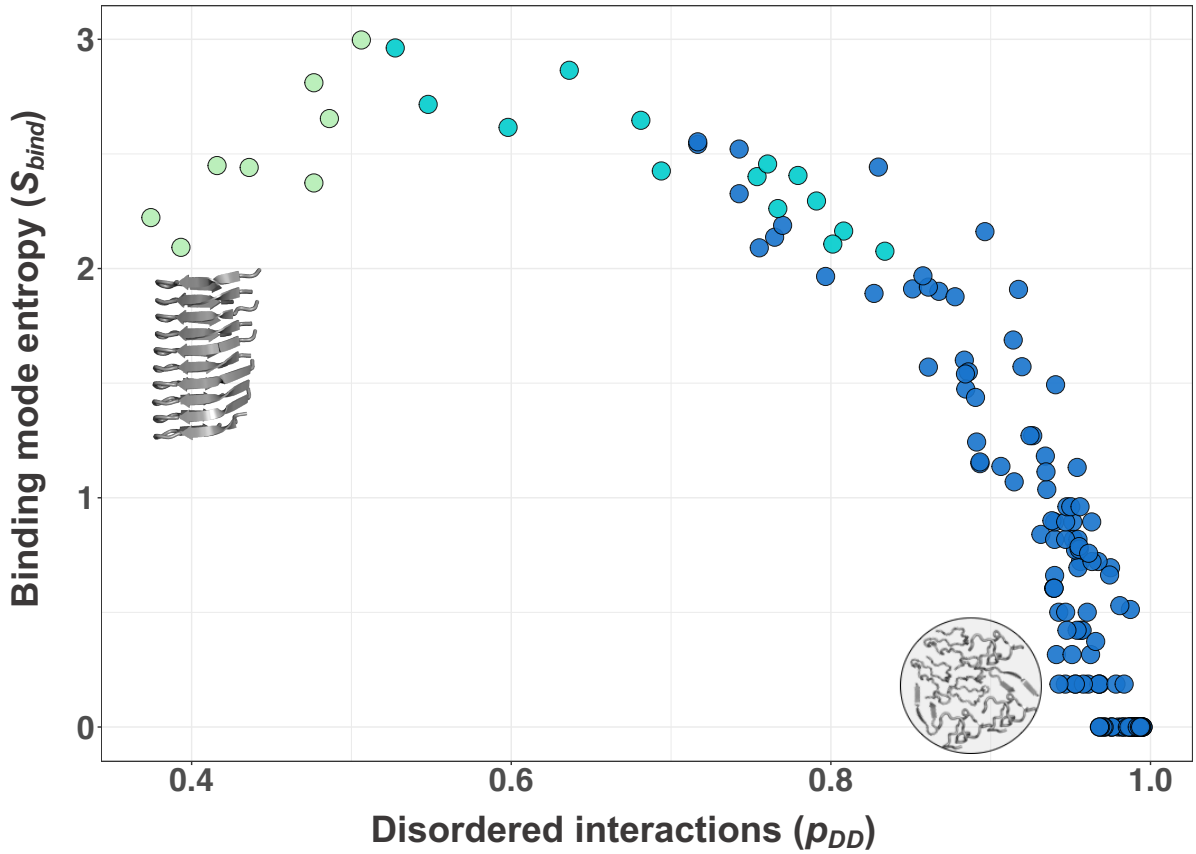

Figure S5. Binding landscape of the TDP-43 prion-like domain. The propensity of disordered binding modes (p_{DD}, Methods eq. 1) by FuzPred [42] is shown on the x-axis, and the binding mode entropy (S_{bind}, Methods Eq. 3), reflecting the multiplicity of interaction modes [41], on the y axis. Most residues of the TDP-43 prion-like domain sample disordered binding modes in the droplet-state (high p_{DD}) with low variability in their interaction behavior (low S_{bind}). The amyloid core region (residues 321-330, light green) exhibits more ordered binding modes with high S_{bind} values suggesting that it can sample both disordered and ordered interactions. The hot-spot region (aquamarine) also exhibits increased multimodal interactions indicating an increased likelihood of conversion to the ordered interactions.

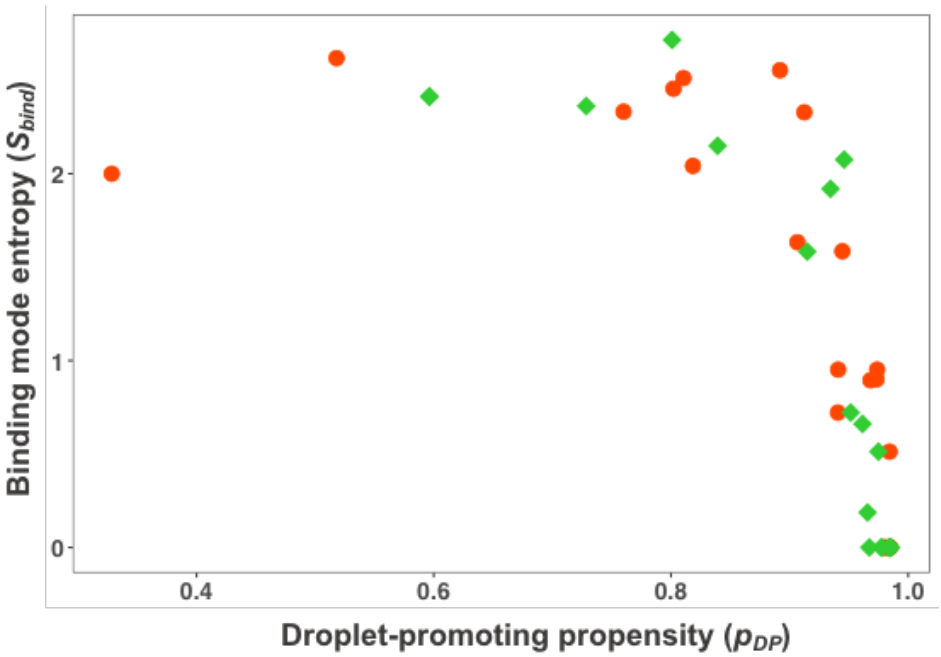

Figure S6. Change in the droplet landscape due to ALS-associated mutations of FUS, hnRNPA1, hnRNPA2, tau, TDP-43, TIA1 and UBQLN2. As compared to wild-type proteins (green diamond) single mutations (red circles), which were shown to promote aggregation (FUS G156E [6], G187S, G225V, G230C, G399V [48], P525L [15]; hnRNPA1 D262V [5], D214V [49]; hnRNPA2 D290V [50]; tau P301L, P301S, A152T [51]; TDP-43 A321V [47], G298S, M337V [52], A315T [53]; TIA1 P262L, A381T, E384K [54]; UBQL2 P506T [55]) (**Table S4**) shift the landscape towards lower droplet-promoting propensity (p_{DP}) [27] and higher binding mode entropy (S_{bind}) [41].