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

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

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

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

Abstract

The transition between the native and amyloid states of proteins can proceed via a deposition pathway through oligomeric intermediates or via a condensation pathway through liquid droplet intermediates generated through liquid-liquid phase separation. The maturation of these droplet intermediates into ordered assemblies has been associated with human disease, including in particular amyotrophic lateral sclerosis (ALS), although the mechanisms of toxicity have not been yet clarified. Here we investigate the processes by which ALS-related mutations give rise to cytotoxicity along the condensation pathway. Based on the sequence-determinants of the different types of interactions stabilising the droplet and amyloid states, we accurately predict the levels of toxicity of about 50,000 deep mutagenesis variants of TDP-43 prion-like domain. We find that condensation is not typically initiated by structural ordering, but rather through non-specific interactions, and that the cytotoxicity of ALS-related TDP-43 mutations stems from promiscuous interactions within the droplet intermediates, rather than from the mature aggregates. These results provide insights into the mechanisms by which condensates convert into amyloids and their links with human disease.

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

Significance

Protein liquid-liquid phase separation underlies the formation of functional protein condensates, which upon dysregulation can mature into cytotoxic amyloid-containing aggregates. The sequence-based principles governing this pathway, and the mechanisms giving rise to cytotoxicity, however, are still not known in detail. Here, based on the different amino acid codes leading to the droplet and amyloid states, we show how one can predict the toxicity of the intermediate states along the condensation pathway. Our results highlight that this toxicity originates from the interaction promiscuity of amyloid-containing intermediates, in particular those with ALS-related mutations, rather than from the mature amyloid state. These results contribute to our understanding of the mechanisms through which TDP-43 mutations are linked to ALS.

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 an intermediate condensation pathway, going through a liquid-liquid phase separation (**Figure 1**) (5, 6). The discovery of the phenomenon of liquid-liquid phase separation is revealing a variety of previously unsuspected biological processes (7). 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 (8, 9).

Upon dysregulation of the native-to-droplet transition, however, the droplet state can evolve into a transient gel form that can further proceed to the amyloid state (**Figure 1**). This gel state may give rise to cytotoxicity, as it can less readily revert to the native state (5, 10-13). In particular, proteins can remain trapped in the gel form because they recruit other cellular components (14). These observations prompt the question of whether there is an amino acid code determining the condensation pathway of amyloid formation (15, 16), in analogy to the amino acid code that has been already identified for the deposition pathway (17, 18).

In this work, we investigate protein interactions in different higher-order assemblies along the condensation pathway and in experimentally observed droplet-promoting and amyloid-promoting regions. We find that despite the distinct interaction behaviours in droplet-driver and amyloid forming proteins, the properties of droplet-promoting and amyloid-promoting regions overlap. Based on the sequence determinants enabling these interaction modes, we show that the condensation pathway is driven by conformationally heterogeneous elements with increased interaction diversity, which are often referred to as 'labile' structures (19, 20).

We then investigate the impact of mutations on the condensation pathway and how amyloid-containing intermediates give rise to cytotoxicity (5, 11, 12). Based on changes in preferences for the droplet and amyloid states and the impact of mutation on the interaction diversity, we accurately predict cellular toxicity using ~50,000 deep sequencing variants of the TDP-43 prion-like domain (21). We show, in particular in variants with ALS-related mutations, that promiscuous intermediates, and not the aggregates are cytotoxic. Our results give insights

into the mechanisms and sequence determinants governing the condensation pathway and the influence of the pathological mutations on this process.

Results

Datasets of higher-order assemblies 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) (22), PhaSePro (https://phasepro.elte.hu) (23), and LLPSDB (http://bio-comp.org.cn/llpsdb) (24) (Methods, **Table S1**). Droplet-client proteins (DROP-C), which require an interacting partner for phase separation, are derived from LLPSDB database (24). Regions of droplet-driver or droplet-client proteins that mediate droplet formation are assembled from the PhaSePro dataset (23) (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 (19).

The amyloid state is represented by both pathological and functional amyloids (AMYLOID), which are derived from the AmyPro database (http://amypro.net) (25) (Methods, **Table S1**). Proteins with inheritable traits in *Saccharomyces cerevisiae* (26), are defined as yeast prions (YPRIO, **Table S1**). These proteins represent functional amyloids, but have not been listed in the droplet and amyloid datasets (Methods, **Table S1**). Amyloid-promoting regions (APRs) are derived from the AmyPro database (25), which include amyloid cores forming cross-β spines and their flanking regions (Methods, **Table S1**). 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).

Taken together, the assembled datasets represent different stages along the condensation pathway, sampling a wide range of intrinsic solubility values (27) (**Figure 2A**).

Interaction behaviours in the amyloid and droplet states

Increasing experimental evidence demonstrates that protein assemblies can exhibit a wide range of interaction modes by sampling a continuum between ordered and disordered bound states (28). Ordered interactions are mediated by well-defined contact patterns (29), while disordered interactions are realized via different binding configurations (30). Proteins may also switch between ordered and disordered bound states by posttranslational modifications or according to the cellular context (31). It has been demonstrated that these interaction behaviours can be predicted from sequence (32) (Methods, eqs. 1-3).

We have demonstrated previously that the probability for spontaneous phase separation depends on the probability of disorder in the free state (p_D) and the probability of disordered binding (p_{DD}) (33). Based on these two properties, which capture the conformational entropy in the bound state, it is possible to predict droplet-driver proteins and droplet-promoting regions from sequence (Methods eqs. 4-5), in good agreement with the experiments (453 proteins deposited in public databases, **Table S1**) (33). Thus, we have compared these properties for different higher-order assemblies, and found that while droplet-driver proteins are disordered in both the free and bound forms, amyloid-forming proteins are dominantly ordered in both states (**Figures 2B, S1A**). Droplet-clients are ordered similarly to amyloid-forming proteins, while yeast prions and in particular droplet-amyloids have similar characteristics to droplet-driver proteins (**Figures 2B, S1A**).

Amyloid-forming proteins have the highest probability to sample different interaction modes based on the binding mode entropy (S_{bind}, Methods eq. 3) (32), and switch thus between disordered and ordered binding (**Figure 2C**). Functional amyloids and yeast amyloids exhibit a broader spectrum of interactions than pathological amyloids (**Figures 2C, S2**). In contrast, droplet-driver proteins and in particular droplet-amyloids exhibit a lower interaction diversity (low binding mode entropy, **Figure 2C**), indicating a strong preference for disordered interactions in the condensates.

Taken together, we observe markedly different interaction behaviours in the droplet and amyloid states, the latter is associated with high interaction diversity to switch from disordered to ordered binding modes.

Droplet-promoting and amyloid-promoting regions populate both states

The two extremes of the condensation pathway are represented by LARKs with high droplet-promoting (p_{DP}, (33)) and low amyloid-promoting propensities (p_{AP}, (27)); and amyloid-cores with low droplet-promoting and high amyloid-promoting propensities (**Figure S3**). While these structural elements exhibit contrasting interaction modes (**Figures 3A, S3**), experimentally observed droplet-promoting and amyloid-promoting regions do not exhibit a significant difference in their predicted interaction modes (**Figures 3A, S3**). Surprisingly, we observed only slightly lower droplet-promoting propensities in APRs than in DPRs, while amyloid-promoting propensities of these regions are similar (**Figures 3A, S3**). These results indicate that residues in droplet-promoting and amyloid-promoting regions populate both states (**Figure 3B,C**).

We found that half of the experimentally identified amyloid-promoting regions contained short DPRs (≥ 10 residues **Table S2**) as well as half of the droplet-promoting regions contained short APRs (≥ 6-residues **Table S2**). The droplet-promoting propensity increased with the distance from the amyloid core and maximized at ~25-35 residues from the core (**Figures 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 wide-spread co-occurrence of droplet-promoting and amyloid-promoting regions in different proteomes (*Homo sapiens* 87%, *Saccharomyces cerevisiae* 88%, in *Caenorhabditis elegans* 89% , **Table S2**) based on predictions by the CamSol (27), Tango (17) and FuzDrop methods (33).

Taken together, although droplet-promoting and amyloid-promoting regions have distinct interaction properties, regions outside the amyloid core or droplet-driving segment can sample both states.

Describing the likelihood of conversion between the droplet and amyloid states

Our results indicate that the conversion requires a switch from disordered to ordered interactions involving intermediates, which can sample both states. As illustrated on the prion-like domain of TDP-43 (residues 262-414), residues can interact in both disordered and ordered modes (**Figure 4A**). The frequency of disordered interactions, biasing for the droplet

state is higher than those of ordered modes biasing for the amyloid state, nevertheless both states are accessible (**Figure 4A**).

Based on these results we characterize the likelihood to convert to the amyloid state using a 'droplet landscape' representing the droplet-promoting propensities (p_{DP}) of the residues and their interaction mode diversities (S_{bind}). As illustrated in TDP-43 prion-like domain, residues, which likely remain in the droplet state are in the bottom right of the droplet landscape with high droplet-promoting propensities and low interaction diversities (**Figure 4B**). In contrast, residues with an increased likelihood to form the amyloid state are in the top left of the droplet landscape with lowered droplet-promoting propensities and high interaction diversities, such as the amyloid-core of TDP-32 (residues 321-331) (**Figure 4B**) (34). Most residues of the TDP-32 aggregation hot-spot region (residues 312-342) are located at top right of the droplet landscape with high droplet-promoting propensities and high interaction diversity indicating that these can populate both ordered and disordered interactions and facilitate the conversion to the amyloid state, while being dynamic (**Figure 4B**) (35).

Taken together, we propose that residues with a wide-range of interaction modes are capable to drive the conversion from droplets to amyloids, while those with strong preference for disordered interactions (**Figure S5**) prefer to remain in the droplet state, which both can be predicted from sequence. The 'droplet' landscape conveniently visualizes the contribution of residues to the condensation pathway.

Impact of mutations on the condensation pathway and cytotoxicity

Then we used these insights to assess how mutations influence the conversion towards cytotoxic aggregates. ~50,000 TDP-43 prion-like domain variants generated by deep mutagenesis showed only a weak correlation between the aggregation propensity and cellular toxicity (21). In accord, we found that highly aggregation-prone TDP-43 variants are less toxic than those, which tend to form liquid droplets (**Figure 5A**). Most toxic mutants also exhibit high interaction diversity and are located on the top right regime of the droplet landscape (**Figure 5B**). These results indicate that not the stable aggregates, but the intermediates of the condensation pathway contribute to cellular toxicity (as discussed below).

To probe this hypothesis, we aimed to predict the impact of TDP-43 mutations on the condensation pathway. We used the experimentally observed changes in toxicity (Δe_{tox}) as a measure of progression along the condensation pathway, which we attempted to estimate based on the corresponding changes in amyloid-promoting (Δp_{AP}) and droplet-promoting (Δp_{DP}) propensities and interaction diversity (ΔS_{bind}) (Methods, eq. 6).

To predict changes in cytotoxicity (Δp_{tox}) we trained a support-vector machine (SVM) using the differences between the mutant and wild-type sequences (Δp_{AP}, Δp_{DP}, ΔS_{bind}; Methods, Eq. 6) of single and double mutant datasets (Methods). We hypothesized that the hot-spot region (312-342 residues) is a major determinant in toxicity, and thus averaged the corresponding parameters for this segment (Methods). On single TDP-43 mutants, we obtained comparable performance of Pearson's correlation coefficient r=0.89 with hot-spot averaged and single-residue based SVM models (Methods, **Figure S6**). In case of double missense mutants we obtained an r=0.88 correlation coefficient on the training set (Methods, **Table S4**) and r=0.81 on the test set of 49,735 double missense mutants (**Figure 5C**). Cross-validation of the SVM models trained on single and double mutants also resulted in a comparable performance with r~0.87 (**Table S4**).

Taken together, these results demonstrate that the impact of mutations on the condensation pathway and the associated toxicity can be described by the tendency to form droplets, aggregate and the likelihood of conversion estimated based on the interaction diversity. Sequence determinants underlying these three features provide the amino acid code for aggregation and cytotoxicity.

ALS-related mutations lead to cytotoxic, promiscuous intermediates along the condensation pathway

We next aimed at understanding how TDP-43 mutations associated with amyotrophic lateral sclerosis (ALS) affect the cellular toxicity of TDP-43 assemblies. Since mutations in condensate-forming proteins may interfere with many biochemical processes, in particular, several steps

of RNA processing, it is not straightforward to determine their contribution to cytotoxicity via the condensation pathway (12, 13, 36).

Using TDP-43 deep-sequencing data (21), we found that 78% of ALS-related mutations (36) in single mutants, and 70% in double mutants appeared as toxic variants. Toxicity of ALS-related TDP-43 mutants was associated with decreased aggregation and slightly decreased droplet formation (**Figure 6A**). In contrast, the non-toxic ALS-related TDP-43 mutants were more prone to aggregate and exhibited considerably decreased droplet formation (**Figure 6A**). These results suggest that cytotoxicity of disease-related mutants does not originate in the amyloid state. We observed that ALS-related TDP-43 mutations increased interaction diversity (**Figure 6A**). This can promote the conversion to the amyloid state via homotypic interactions or increase interaction promiscuity via heterotypic interactions (**Figures 2C, S3B**). The latter scenario can also lead to cellular toxicity upon over-expression (37). These results suggest that cytotoxicity of disease-related mutants may originate in promiscuous interactions of the condensation intermediates.

Along these lines, we attempted to predict condensation-related toxicity of known ALS-related mutations in TDP-43, FUS and TAF15 (36) (Methods). In FUS, for example, glycine mutations are predicted to increase toxicity via the condensation pathway, while arginine mutations likely use other mechanisms (**Figure 6B**). This finding is in line with the observations that these mutations have distinct impacts on the condensate properties, while glycine mutations affect condensate fluidity, arginine mutations lead to impairment of RNA binding (38). The toxicity of FUS P525L mutation is predicted not to be related to the condensation pathway (**Figure 6B)**, in accord with the observation to form fibrils via the deposition pathway (39). We found that only about half of the ALS-related mutations is predicted to increase cytotoxicity via the condensation pathway, whereas other disease mutations use other mechanisms (**Figure 6B**).

Taken together, these results offer insights into how disease-related mutations give rise to cytotoxicity along the condensation pathway.

Discussion and Conclusions

Most proteins are capable of converting into the amyloid state when their concentration exceeds their solubility limits (40), which can take place via disordered oligomers through the deposition pathway (3, 4). Recent results have shown that the droplet state is also widespread and accessible to most proteins (33). Although the droplets formed through liquid-liquid phase separation can be functional, with new biological roles being discovered with rapid pace (7), they can also convert to amyloid state through the condensation pathway (5, 6, 41). The possibility for proteins of populating different states, however, creates a challenge for the protein homeostasis system, since dysregulated transitions into mis-functional assemblies can initiate pathological processes.

It is therefore important to understand the sequence determinants driving dynamic droplets towards irreversible aggregates. We found that droplets and amyloids exhibit contrasting (disordered vs ordered) interaction behaviours, while residues in experimentally observed regions promoting either state can sample both interaction modes. Along these lines, we showed that the condensation pathway is driven by residues, which are capable of sampling multiple interaction modes that are often associated with conformationally heterogeneous, 'labile' structural elements (20, 35).

To characterise the likelihood of residues to convert to aggregates via the condensation pathway, we then introduced the 'droplet landscape' approach. As illustrated on the TDP-43 prion-like domain, residues, which tend to form droplets (high droplet-promoting propensity), and have high interaction diversity are likely convert to amyloids, while those with strong preference for disordered interactions tend to remain in the droplet state. These results indicate that structural ordering towards the amyloid state is preceded by exploring multiple interaction modes. In case of homotypic interactions, the system will proceed towards the amyloid state, whereas in case of heterotypic interactions aggregation will be slowed down. Increased interaction diversity however, may also lead to promiscuity and cause cytotoxicity.

Indeed, we have demonstrated that the droplet-promoting and amyloid-promoting propensities together with the interaction diversity can describe the impact of mutations on the condensation pathway. Using these parameters, we accurately predicted the cellular

toxicity of ~50,000 deep-sequencing mutants of the prion-like domain of TDP-43 representing different condensation intermediates (21). We showed in particular for ALS-related TDP-43 mutations, that cytotoxicity is not associated with the amyloid state. Instead, toxic variants are the amyloid-containing, yet dynamic intermediates with increased interaction promiscuity. This is somewhat analogous to the deposition pathway, where toxicity is associated with misfolded oligomers instead of the final amyloid state. We also note that interaction promiscuity was also observed to cause dosage sensitivity (37), and that cytotoxicity may stem from a variety of other factors (36, 42).

Taken together, our results uncover the mechanisms, as well as the corresponding sequence codes, that drive the irreversible transition of condensates into amyloids, and rationalise the impact of disease-related mutations on aggregation and cytotoxicity.

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) (22), the PhaSePro database with regions mediating phase separation identified (https://phasepro.elte.hu) (23), and the LLPSDB database (http://bio-comp.org.cn/llpsdb) (24) 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**) (33). Droplet client proteins, which phase separate via interactions with other partners, were assembled from the LLPSDB dataset (http://bio-comp.org.cn/llpsdb) (24) (DROP-C, **Table S1**). 20 proteins with evidence for both droplet and amyloid formation based on the AmyPro database (http://amypro.net) (25) were defined as droplet-amyloids (DROP-AMYLOID, **Table S1**).

Droplet-promoting regions were assembled from the PhaSePro dataset (23), and were grouped based on the evidence for spontaneous, or partner-assisted phase separation in the LLPSDB dataset (24) (DPR, **Table S1**). Low-complexity aromatic-rich kinked segments (LARKs)

were used as a reference for droplet-promoting regions. The LARK dataset contained 6075 six-residue long segments in 400 proteins, as in the original reference (19).

Datasets of amyloid-forming proteins. 120 amyloid-forming proteins were assembled from the AmyPro database (http://amypro.net) (25) (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 (25) (**Table S1**). 46 yeast proteins with heritable traits in *Saccharomyces cerevisiae* were defined as yeast prions (YPRION, **Table S1**) (26). Most yeast prions had Hsp104 or Hsp70 dependence (26). 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 (APRs, **Table S1**) 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 (43)

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

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} is computed using FuzPred (43).

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 (32)

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

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

Droplet-promoting propensity profiles. The droplet-promoting propensity profiles p_{DP} quantify the probability of spontaneous liquid-liquid phase separation (33)

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

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

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

where $p_D(A_i)$ 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 (33). p_D was derived from the disorder score as computed using the Espritz NMR algorithm (44). The p_{DD} values were predicted by FuzPred (43). p_{DP} =0.60 is the threshold that a residue is involved in spontaneous phase separation (33).

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}) (27). p_{AP} = - p_{CS} and p_{AP} = 0.90 is the threshold, above which a protein likely aggregates (45).

Analysis of TDP-43 deep sequencing data. We analyzed 877 single, and 40,419 double TDP-43 missense mutants with toxicity > 1σ(21). In these datasets, 23 single mutants were ALS-related, and 4,857 double mutants ($\Delta \boldsymbol{e_{tox}} \geq \sigma$) contained 32 out of the reported 46 ALS-related mutations (36).

Droplet-promoting propensities (p_{DP}) were computed by FuzDrop (33) and amyloid-promoting propensities (p_{AP}) were obtained from the solubility scores by CamSol (27). We determined the differences in these quantities computed for the mutant and the wild-type (UniProt Q13148) residue. In case of double mutants, we averaged the p_{DP}, p_{AP}, and S_{bind} values for the hot-spot segment (312-342 residues) and computed the difference between the average values of the mutant and wild-type hot-spot segments.

Predicting changes in cellular toxicity of droplets upon mutation. We modelled the change in cellular toxicity (Δp_{tox}) upon mutation using a linear model

$$\Delta p_{tox} = a * \Delta p_{AP} + b * \Delta p_{DP} + c * \Delta S_{bind} + d \quad (6)$$

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. The parameters *a, b, c* and *d* were determined by the R program. For training, single TDP-43 mutants with $\Delta \boldsymbol{e_{tox}}$ > 3σ were used. For training on double mutations, a randomly chosen set of 1,190 TDP-43 mutants with $\Delta \boldsymbol{e_{tox}} \geq 3\sigma$ was used.

We hypothesized that the hot-spot region (312-342 residues) is a major determinant in toxicity, and thus averaged the corresponding parameters for this segment (Methods). For single TDP-43 mutants with $\Delta \boldsymbol{e_{tox}} \geq 3\sigma$ the performance of the SVM models using hot-spot averaged and single-residue based Δp_{AP}, Δp_{DP} , ΔS_{bind} values were comparable (Pearson's correlation coefficient r=0.89, **Figure S6**). On the training set of double missense mutants with

$\Delta \boldsymbol{e_{tox}} \geq 3\sigma$, we obtained a r=0.88 correlation between the calculated and experimental data (Methods, **Table S4**). We applied this model to predict the change in toxicity for 49,735 double missense mutants, which resulted in r=0.81 (**Figure 6C**). We have also cross-validated the SVM models trained on single and double mutants, which resulted in a comparable performance using mutants with $\Delta \boldsymbol{e_{tox}} \geq 3\sigma$ (r~0.87, **Table S4**). Coefficients and performance of the different models on single and double TDP-43 mutants are reported in **Table S4**. The performance was evaluated based on the Pearson's r values, computed by the R program.

Predicting cellular toxicity of ALS-related mutations due to the condensation pathway

We derived ALS-related mutations from the reference (36). We have determined the $\Delta p_{AP} = p_{AP}(A_m) - p_{AP}(A_{wt})$, $\Delta p_{DP} = p_{DP}(A_m) - p_{DP}(A_{wt})$, and $\Delta S_{bind} = S_{bind}(A_m) - S_{bind}(A_{wt})$ term of the mutated sequences as compared to the wild-type proteins. We have applied the SVM model, which was trained on 498 single TDP-43 mutants with $\Delta \boldsymbol{e_{tox}} \geq 3\sigma$. The performance of the model on TDP-43 mutants is shown in **Table S4,** the AUC values were computed by the R program. The SVM model was applied to predict cytotoxicity of Fus and TAF15 mutants using Eq. 6. Mutations with condensation pathway-related toxicity were defined as $\Delta \boldsymbol{p_{tox}} \geq 0$, mutations with toxicity not related to the condensation pathway were defined as $\Delta \boldsymbol{p_{tox}} < 0$.

References

1. Banani SF, Lee HO, Hyman AA, & Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18(5):285-298.
2. Hyman AA, Weber CA, & Julicher F (2014) Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* 30:39-58.
3. Knowles TP, Vendruscolo M, & Dobson CM (2014) The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* 15(6):384-396.
4. Michaels TCT*, et al.* (2020) Dynamics of oligomer populations formed during the aggregation of Alzheimer's Abeta42 peptide. *Nat Chem*.
5. Patel A*, et al.* (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162(5):1066-1077.

6. Molliex A*, et al.* (2015) Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* 163(1):123-133.
7. Boeynaems S*, et al.* (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28(6):420-435.
8. Patel A*, et al.* (2017) ATP as a biological hydrotrope. *Science* 356(6339):753-756.
9. Schaefer KN & Peifer M (2019) Wnt/Beta-Catenin Signaling Regulation and a Role for Biomolecular Condensates. *Developmental cell* 48(4):429-444.
10. Schmidt HB & Gorlich D (2015) Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* 4.
11. Murakami T*, et al.* (2015) ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* 88(4):678-690.
12. Alberti S & Dormann D (2019) Liquid-Liquid Phase Separation in Disease. *Annu Rev Genet* 53:171-194.
13. Mathieu C, Pappu RV, & Taylor JP (2020) Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* 370(6512):56-60.
14. Tan PS*, et al.* (2018) Two Differential Binding Mechanisms of FG-Nucleoporins and Nuclear Transport Receptors. *Cell reports* 22(13):3660-3671.
15. Vernon RM & Forman-Kay JD (2019) First-generation predictors of biological protein phase separation. *Curr Opin Struct Biol* 58:88-96.
16. Wang J*, et al.* (2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174(3):688-699 e616.
17. Linding R, Schymkowitz J, Rousseau F, Diella F, & Serrano L (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* 342(1):345-353.
18. Tartaglia GG & Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* 37(7):1395-1401.
19. Hughes MP*, et al.* (2018) Atomic structures of low-complexity protein segments reveal kinked beta sheets that assemble networks. *Science* 359(6376):698-701.
20. Bienz M (2020) Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. *Cell* 182(4):799-811.

21. Bolognesi B*, et al.* (2019) The mutational landscape of a prion-like domain. *Nature communications* 10(1):4162.
22. You K*, et al.* (2020) PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res* 48(D1):D354-D359.
23. Meszaros B*, et al.* (2020) PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res* 48(D1):D360-D367.
24. Li Q*, et al.* (2020) LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Res* 48(D1):D320-D327.
25. Varadi M, De Baets G, Vranken WF, Tompa P, & Pancsa R (2018) AmyPro: a database of proteins with validated amyloidogenic regions. *Nucleic Acids Res* 46(D1):D387-D392.
26. Chakrabortee S*, et al.* (2016) Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* 167(2):369-381 e312.
27. Sormanni P, Aprile FA, & Vendruscolo M (2015) The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol* 427(2):478-490.
28. Fuxreiter M (2018) Fold or not to fold upon binding - does it really matter? *Current Opinion in Structural Biology* 54:19-25.
29. Kussie PH*, et al.* (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274(5289):948-953.
30. Borgia A*, et al.* (2018) Extreme disorder in an ultrahigh-affinity protein complex. *Nature* 555(7694):61-66.
31. Stamos JL, Chu ML, Enos MD, Shah N, & Weis WI (2014) Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *eLife* 3:e01998.
32. Horvath A, Miskei M, Ambrus V, Vendruscolo M, & Fuxreiter M (2020) Sequence-based prediction of protein binding mode landscapes *PLoS Comp Biol* 16:e1007864.
33. Hardenberg M, Horvath A, Ambrus V, Fuxreiter M, & Vendruscolo M (2020) Widespread occurrence of the droplet state of proteins in the human proteome. *Proc Natl Acad Sci U S A*.
34. Guenther EL*, et al.* (2018) Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat Struct Mol Biol* 25(6):463-471.

35. Conicella AE, Zerze GH, Mittal J, & Fawzi NL (2016) ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* 24(9):1537-1549.

36. Kapeli K, Martinez FJ, & Yeo GW (2017) Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum Genet* 136(9):1193-1214.

37. Vavouri T, Semple J, Garcia-Verdugo R, & Lehner B (2009) Intrinsic Protein Disorder and Interaction promiscuity are widely associated with Dosage Sensitivity. *Cell* 138:198-208.

38. Niaki AG*, et al.* (2020) Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations. *Mol Cell* 77(1):82-94 e84.

39. Marrone L, Quamar S, Mannini B, St George-Hyslop P, & Vendruscolo M (2020) P525L promotes the aggregation of FUS by altering its biochemical and biophysical properties. *Science Matters*.

40. Vecchi G*, et al.* (2020) Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc Natl Acad Sci U S A* 117(2):1015-1020.

41. Wegmann S*, et al.* (2018) Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J* 37(7).

42. Newberry RW, Leong JT, Chow ED, Kampmann M, & DeGrado WF (2020) Deep mutational scanning reveals the structural basis for alpha-synuclein activity. *Nat Chem Biol* 16(6):653-659.

43. Miskei M, Horvath A, Vendruscolo M, & Fuxreiter M (2020) Sequence-Based Prediction of Fuzzy Protein Interactions. *J Mol Biol* 432:2289-2303.

44. Walsh I, Martin AJ, Di Domenico T, & Tosatto SC (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* 28(4):503-509.

45. Sormanni P & Vendruscolo M (2019) Protein Solubility Predictions Using the CamSol Method in the Study of Protein Homeostasis. *Cold Spring Harb Perspect Biol* 11(12).

46. Haass C & Selkoe DJ (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol* 8(2):101-112.

47. Benilova I, Karran E, & De Strooper B (2012) The toxic Abeta oligomer and Alzheimer's disease: an emperor in need of clothes. *Nat Neurosci* 15(3):349-357.

48. Brorsson AC*, et al.* (2010) Intrinsic determinants of neurotoxic aggregate formation by the amyloid beta peptide. *Biophys J* 98(8):1677-1684.

49. Luheshi LM*, et al.* (2007) Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. *PLoS Biol* 5(11):e290.

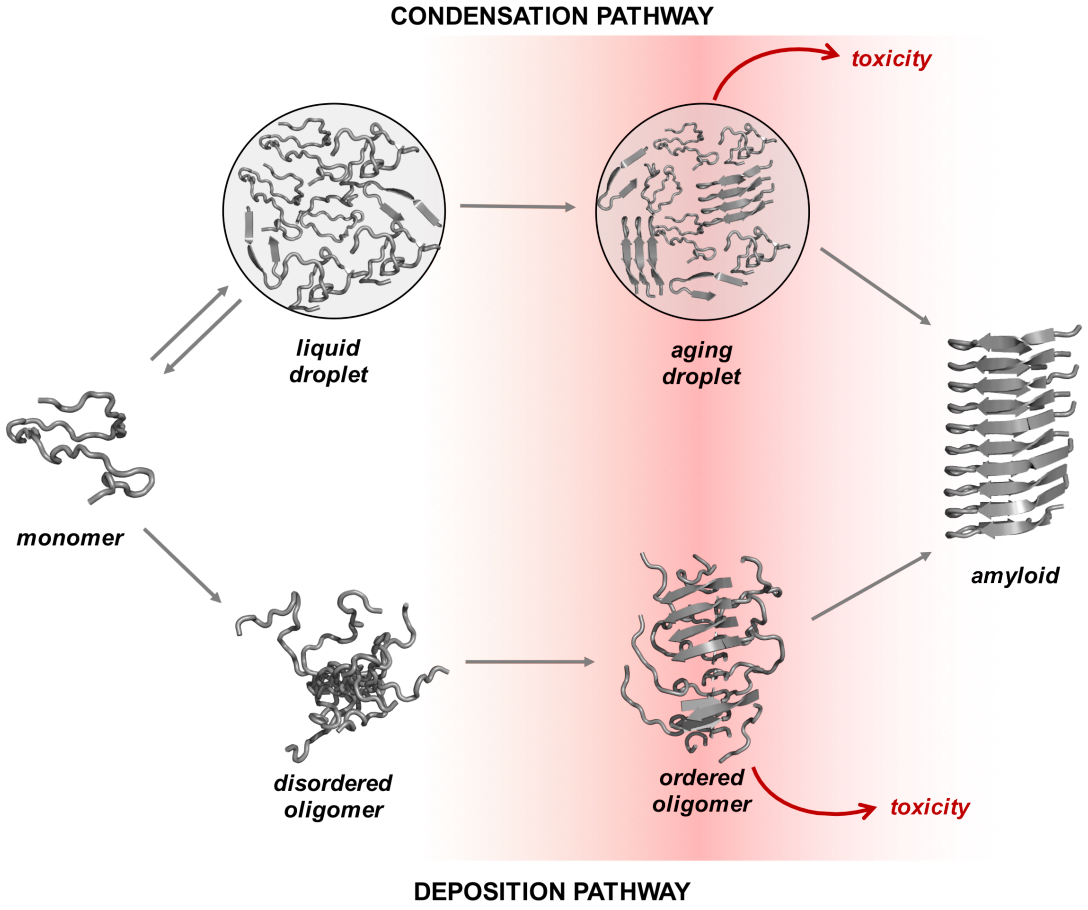


Figure 1. Cytotoxic protein aggregates appear as intermediates in both the condensation and the deposition pathways. Along the deposition pathway, proteins move from the native state to the amyloid state through increasingly ordered oligomer aggregates (3, 4). Some of these oligomers are highly cytotoxic (46, 47) according to an amino acid code that has been partly elucidated (48, 49). 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 (5, 10-12). In this work, we investigate the amino acid code driving the condensation pathway leading to cytotoxic intermediates.

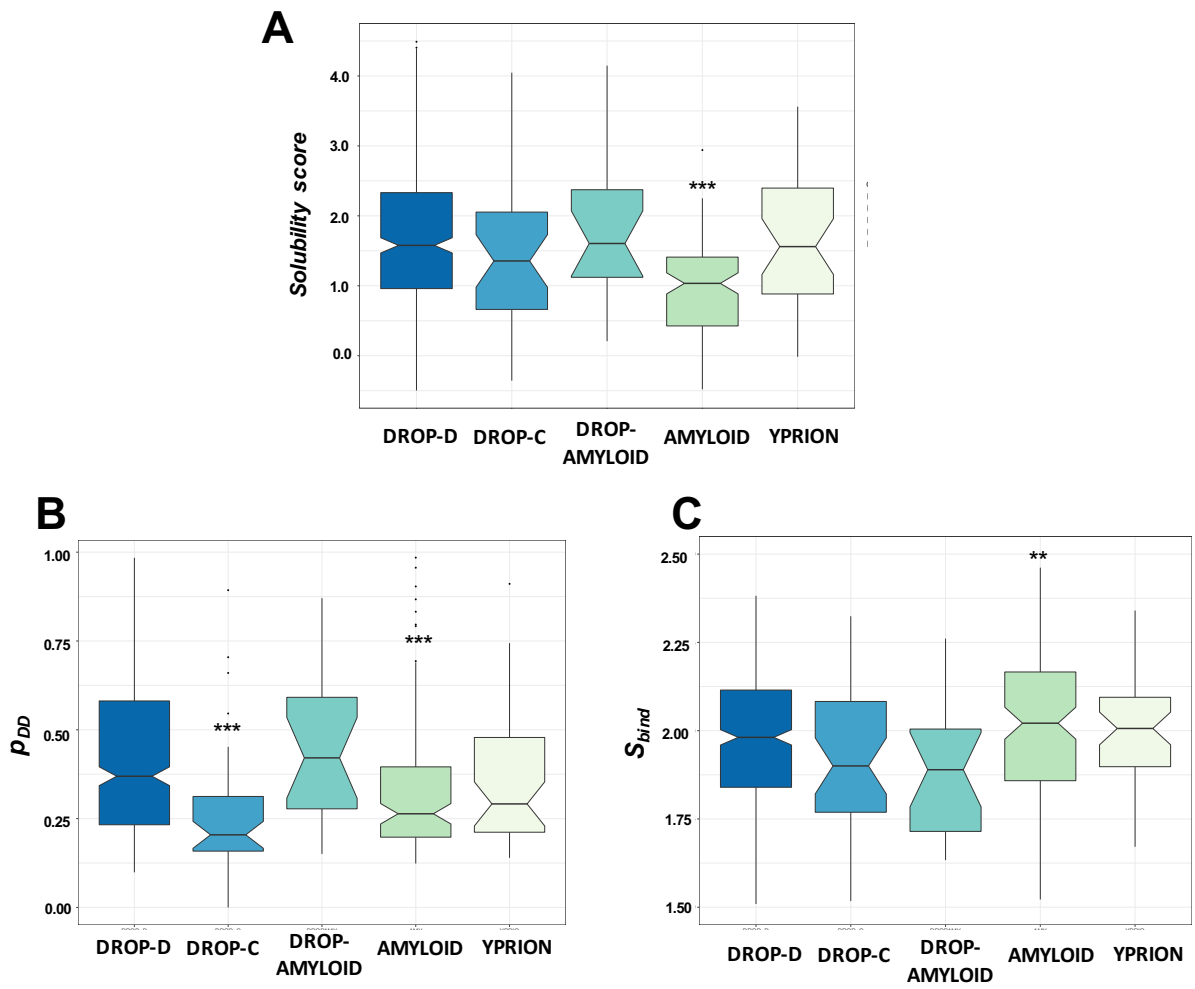


Figure 2. Comparison of solubility and interactions of amyloid-forming and droplet-forming proteins. (A) Comparison of the solubility scores of droplet and amyloid proteins. Droplet-driver proteins (DROP-D, blue, **Table S1**), which spontaneously undergo liquid-liquid phase separation (33) and droplet-amyloid proteins (DROP-AMYLOID, aquamarine, **Table S1**), which exhibit both droplet and amyloid behaviors, have the highest solubility scores by CamSol (27), while amyloid-forming proteins (AMYLOID, green) in the AmyPro database (http://amypro.net) (25) 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* (26), have intermediate solubility scores. **(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 (43) 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. **(C) Comparison of the interaction mode diversity** (S_{bind}) **of droplet and amyloid proteins.** The binding mode entropy S_{bind} values, which express the context-dependent variation of interactions, were computed by FuzPred (32) (Methods, Eq. 3). Amyloid-forming proteins exhibit the interaction diversity, as they can sample both disordered and ordered interactions. Droplet and droplet-amyloid proteins have a strong preference for disordered binding (low S_{bind}). Statistical significances as compared to droplet driver proteins were computed by Mann-Whitney test using the R program (** $p < 10^{-2}$, ***$p< 10^{-3}$).

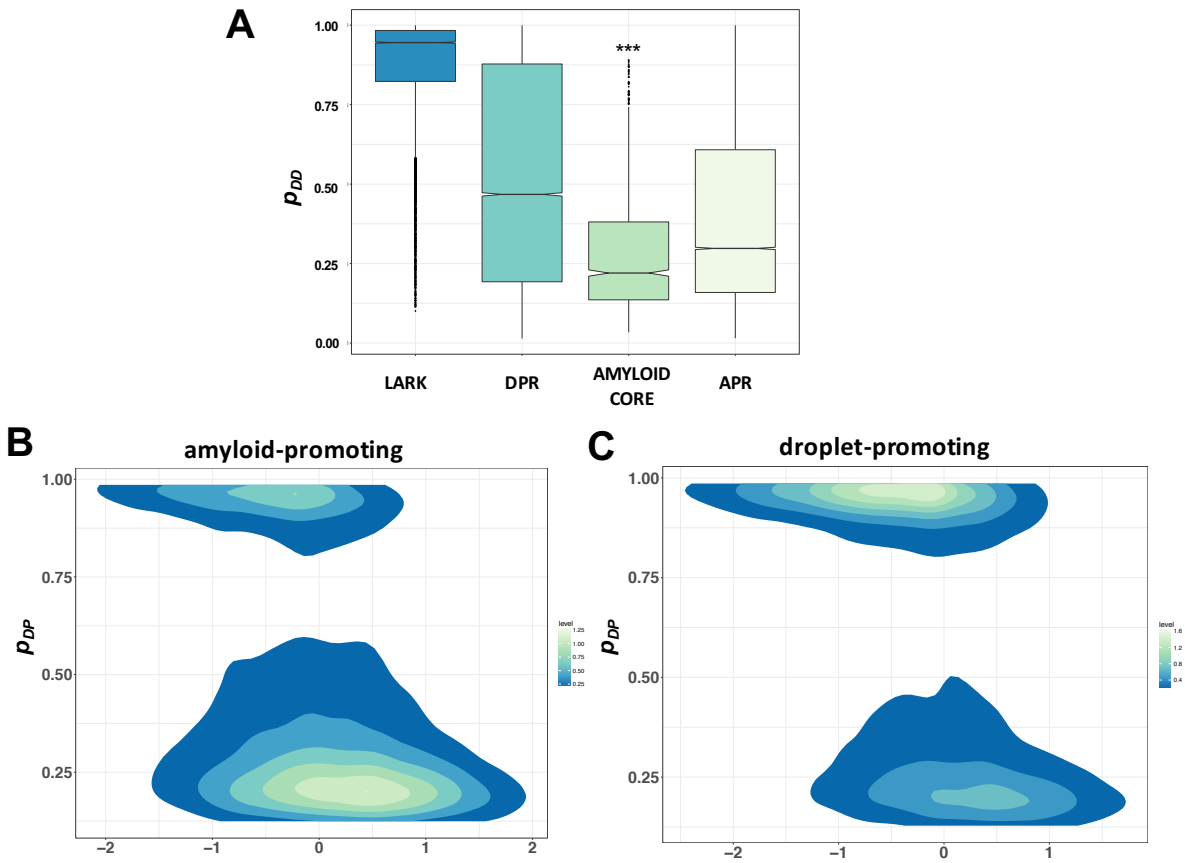


Figure 3. Experimentally observed droplet-promoting and amyloid-promoting regions can sample both states. (A) Comparison of the disordered binding modes (p_{DD}) of droplet-promoting (DPR, aquamarine) **and amyloid-promoting** (APR, light green) **regions.** Interactions of droplet-promoting and amyloid-promoting regions (**Table S1**) vary in wide range between these disordered and ordered binding modes, exemplified by LARKs (blue) and amyloid cores (green). Statistical significances in reference to droplet-promoting regions were computed by Mann-Whitney test using the R program (*** $p < 10^{-3}$). **(B,C) Droplet-promoting** (p_{DP}) **and amyloid-promoting** (p_{AP}) **propensities of amyloid-promoting (B) and droplet-promoting regions (C).** Droplet-promoting propensities were computed by FuzDrop (33) and amyloid-promoting propensities were derived from the solubility scores by CamSol (27). 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 of the amyloid-promoting regions also populate the droplet state and residues of the droplet-promoting regions also sample the amyloid state. These results indicate that the condensation pathway is driven by residues, which can populate both states.

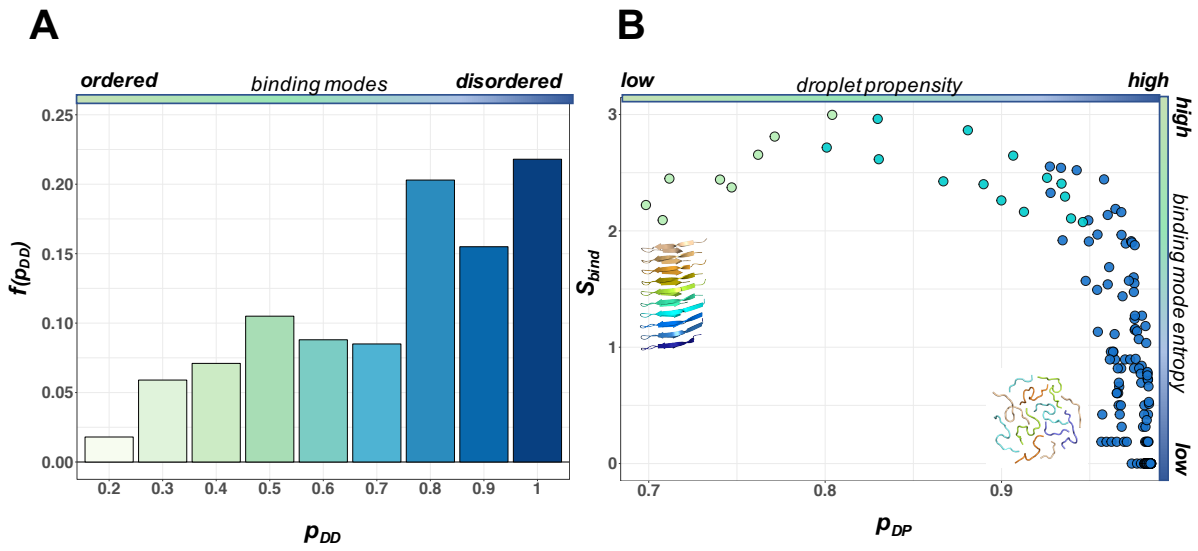


Figure 4. Droplet landscape of the TDP-43 prion-like domain. (A) Frequencies of different binding modes for the aggregation hot-spot (residues 312-342) of TDP-43. The frequencies of different interaction modes vary between disordered (high p_{DD} values, blue) and ordered (low p_{DD} values, light green) interactions. The TDP-43 prion-like domain preferentially samples disordered interactions corresponding to the droplet state (blue, right), but can also form ordered interactions corresponding to the amyloid state with lower frequency (green, left). **(B) Droplet landscape of the TDP-43 prion-like domain (residues 262-414).** The droplet-promoting propensities calculated by FuzDrop (33) are shown on the x-axis (Methods, eq. 4), and the interaction diversity (defined by binding mode entropy S_{bind}, Methods Eq 3), on the y axis, which is computed for each residue from the frequencies displayed in panel **(A)** (32). Residues, which tend to remain in the droplet state are in the bottom right regime (blue), they have high droplet-promoting (p_{DP}) propensities and low interaction diversity (S_{bind}), reflecting a strong preference for disordered interactions in the droplet state. Residues, which likely convert to amyloids, such as the amyloid-core of TDP-43 (321-330 residues, light green), are in the upper left regime (green) have decreased droplet-promoting (p_{DP}) propensities and high interaction diversity (S_{bind}), reflecting a possible change in interaction modes. Residues of the hot-spot region (aquamarine) have high droplet-promoting (p_{DP}) propensities and interaction diversity (S_{bind}), facilitate the switch between disordered and ordered interaction modes.

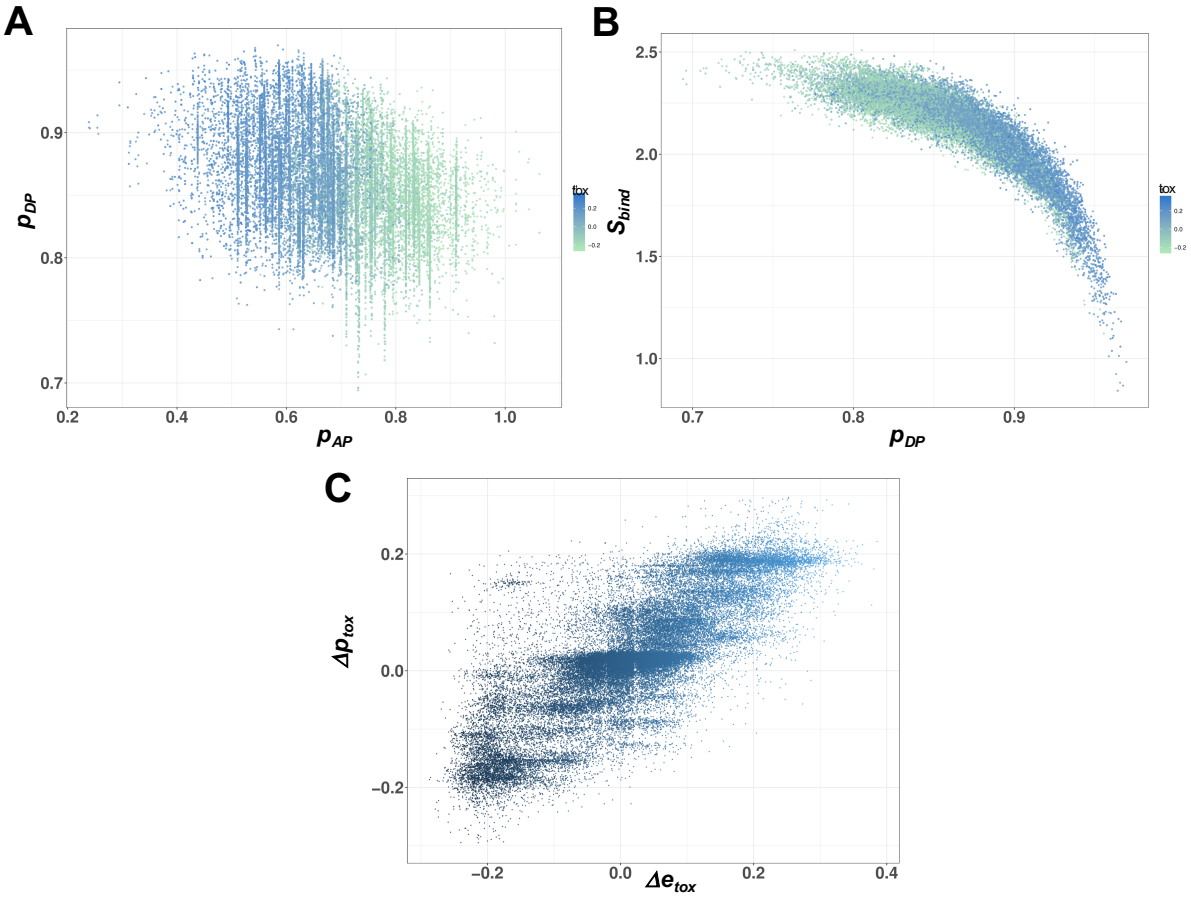
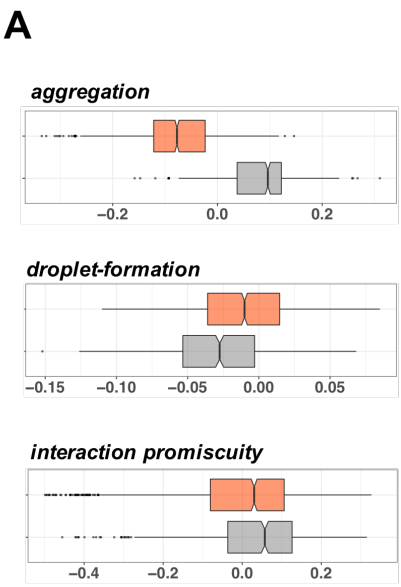
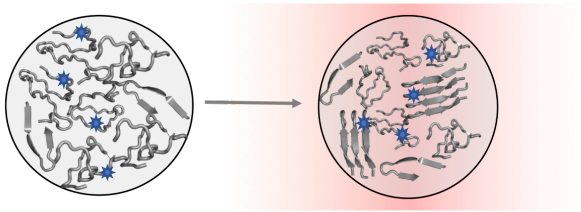


Figure 5. The impact of mutations on cytotoxicity of TDP-43. (A) The droplet-promoting (p_{DP}) **and amyloid-promoting** (p_{AP}) **propensities of double missense mutants of TDP-43 prion-like domain and their toxicity levels.** Droplet-promoting propensities by FuzDrop (33) and amyloid-promoting propensities derived from the solubility scores by CamSol (27) weakly correlate (Pearson's coefficient of correlation r=0.34). High amyloid-promoting propensities correspond to low cytotoxicity (green), while high droplet-promoting propensities correspond to high cytotoxicity (blue). The bar scale on the right shows the experimental toxicity values (Δe_{tox}) (21). **(B) Droplet-landscape of double missense mutants of TDP-43 prion-like domain.** The droplet-promoting propensities by FuzDrop (33) (x-axis), and the interaction diversity (S_{bind}) by FuzPred (32) (y axis) are shown. The droplet landscape indicates that the shift from disordered binding modes corresponding to the droplet-state towards the ordered binding modes of the amyloid-state is coupled to only a moderate decrease in droplet-formation. Mutants with higher droplet-promoting propensity and increases interaction diversity have higher cytotoxicity (blue) then mutants with increased propensity to convert to

the amyloid state (high S_{bind}, green). The bar scale on the right shows the experimental toxicity values (Δe_{tox}) (21). **(C) Correlation between experimental ($\Delta \boldsymbol{e_{tox}}$) and predicted ($\Delta \boldsymbol{p_{tox}}$) changes in cellular toxicity upon double missense TDP-43 mutations.** Δp_{tox} was computed based on the changes in droplet-promoting (Δp_{DP}), and amyloid-promoting (Δp_{AP}) propensities and interaction diversities (ΔS_{bind}) (Eq. 6), using an SVM model parametrised on a training set of double missense TDP-43 mutants with toxicity > 3σ (Methods) (21). The Pearson's coefficient of correlation between predicted and experimental toxicity values is r=0.81.





Protein	Non-condensation related	Condensation-related
TDP-43	A315T,A321V,G290A,G290V,G294A,G294V, G295C,G348C,G348V,I383V,M311V,M337V, M359V,N352S,N352T,N378S,P363A,R361T, S379C,S393L	A315E,A321G,A382P,A382T,G287S,G295R, G295S,G298S,G335D,G348R,G357R,G357S, G368S,G376D,G384R,I383T,N345K,N378D, N390D,Q331K,Q343R,R361S,S292N,S332N, S379P,W385G
Fus	E516V,G225V,G230C,G245V,G399V,K510R, M254V,M464I, N159Y,R216C,R234C,R234L, R244C, R383C,R487C,R491C, R514S,S462F, P525L	G156E,G187S,G191S,G206S,G226S,G507D, G509D,G515C,H517D,H517P,H517Q,K510E, N63S,P18S,Q210H,R514G,R522G,R524S,S115N, S142N,S513P
TAF15	A31T,R388H,R395Q,R408C	G391E,G452E,G473E,M368T

Figure 6. The impact of disease-related mutations on the condensation pathway and toxicity. (A) Impact of toxic (red) and non-toxic (grey) ALS-related mutations on the condensation pathway of TDP-43 prion-like domain. Changes in amyloid-promoting propensity (Δp_{AP}), droplet-promoting propensity (Δp_{DP}) and promiscuity-related interaction diversity (S_{bind}) of the hot-spot region (residues 312-342) in TDP-43 variants with ALS-related mutations (Methods). Differences were computed as compared to the wild-type hot-spot region. Toxicity was assigned based on experimental data using double mutants with experimental toxicity > 3σ (21). **(B) Predicted cytotoxicity of ALS-related mutations based on the condensation pathway**. Using on the model developed to predict cytotoxicity of TDP-43 (Method), we also estimated the toxicity of disease-related mutations in TDP-43, FUS and TAF15, which are likely to be linked to condensation. ALS-related mutations were assembled from (36), cytotoxicity was computed using Eq. (6). We found that condensation-related toxicity of half of the disease-mutations is linked with altered dynamics and interaction promiscuity; while the other mutations are toxic through the other mechanisms, such as altered RNA interactions, localisation, or translation.

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 (27) and Tango (17).

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 (17).

Table S4. Bioinformatics models to predict cellular toxicity and performance of binary toxicity predictions on different TDP-43 datasets. The performance of the SVM models trained on single and double mutants. Different datasets were defined based on the standard deviations of the experimental toxicity values (21). The terms in Eq. (6) have been computed for the hot-spot region. The models obtained by linear fit on the full set of single and double

mutants are also displayed. The coefficients of the different terms express that decreased aggregation and increased droplet-forming propensity and binding mode diversity contributes to increased cytotoxicity.

For TDP-43 ALS-related mutations, toxicity was estimated using Eq. (6), where the terms were averaged for the hot-spot region (residues 312-342) or the changes corresponding to the mutated residue were used. Binary classification of toxicity was based on $\Delta \boldsymbol{e_{tox}} < 0$ (non-toxic) and $\Delta \boldsymbol{e_{tox}} \geq 0$ (toxic). Area under the curve (AUC) values were computed by the R program.

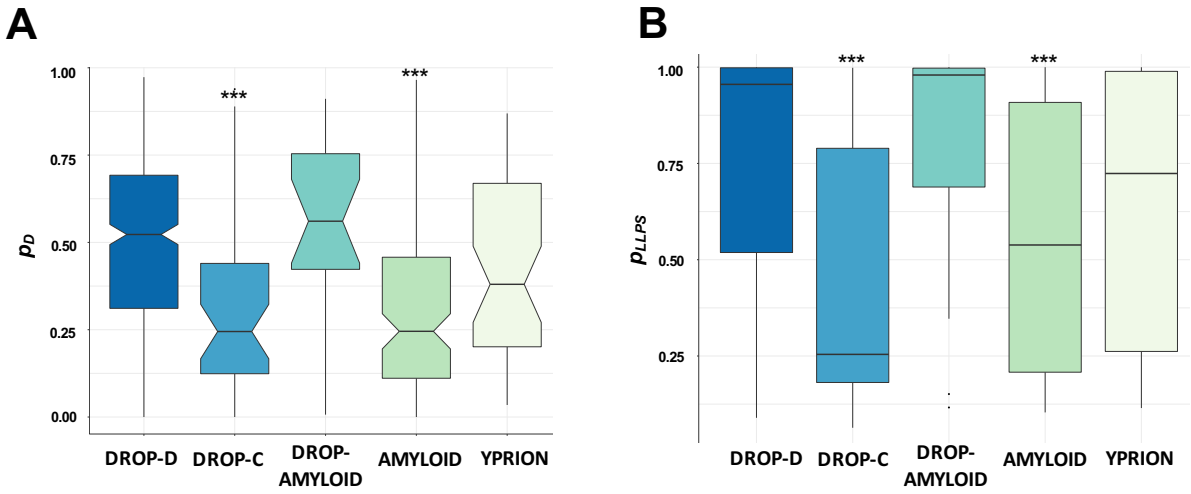


Figure S1. Free-state conformational entropy and droplet-forming propensities of amyloid- and droplet-forming proteins. (A) Free-state disorder (p_D) were estimated as the fraction of disordered residues by Espritz NMR (44). Droplet-driver proteins (DROP-D, blue, **Table S1**), which spontaneously undergo liquid-liquid phase separation (33) 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 (25)) 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* (26) have intermediate disorder values between the amyloid and droplet states. **(B) Droplet-promoting propensities (p_{LLPS})** were estimated by FuzDrop (33). 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 as compared to droplet-driver proteins were computed by Mann-Whitney test using the R program (*** $p< 10^{-3}$).

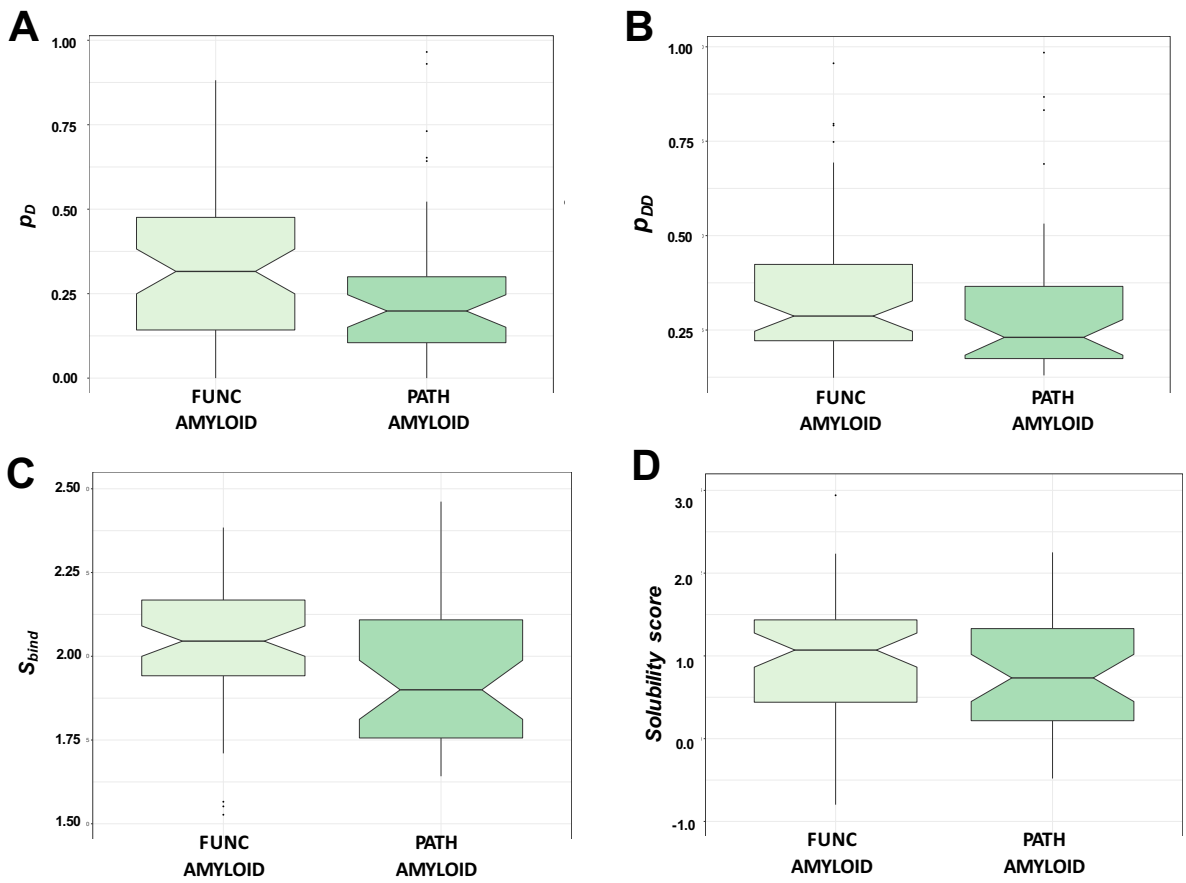


Figure S2. Conformational properties, interaction behavior and solubility scores of functional (light green) and pathological (green) amyloid-forming proteins. (A) Free-state disorder (p_D) was estimated by the fraction of disordered residues by Espritz NMR(44). Functional amyloids are more disordered in their unbound state than pathological amyloids. **(B) The propensity of disordered binding modes** (p_{DD}) was estimated by the probability of disorder-to-disorder transition by FuzPred (43) (Methods, Eq. 1). Functional and pathological amyloids both exhibit ordered binding modes. **(C) Interaction mode diversity** was defined by the entropy of possible binding modes (S_{bind}) as computed by FuzPred (32) (Methods, Eq. 3). Functional amyloids are more prone to change their binding modes in a context-dependent manner than pathological amyloids. **(D) Solubility scores** were computed by CamSol (27). Intrinsic solubility scores of functional and pathological amyloid-forming proteins do not differ significantly based on Mann-Whitney test using the R program.

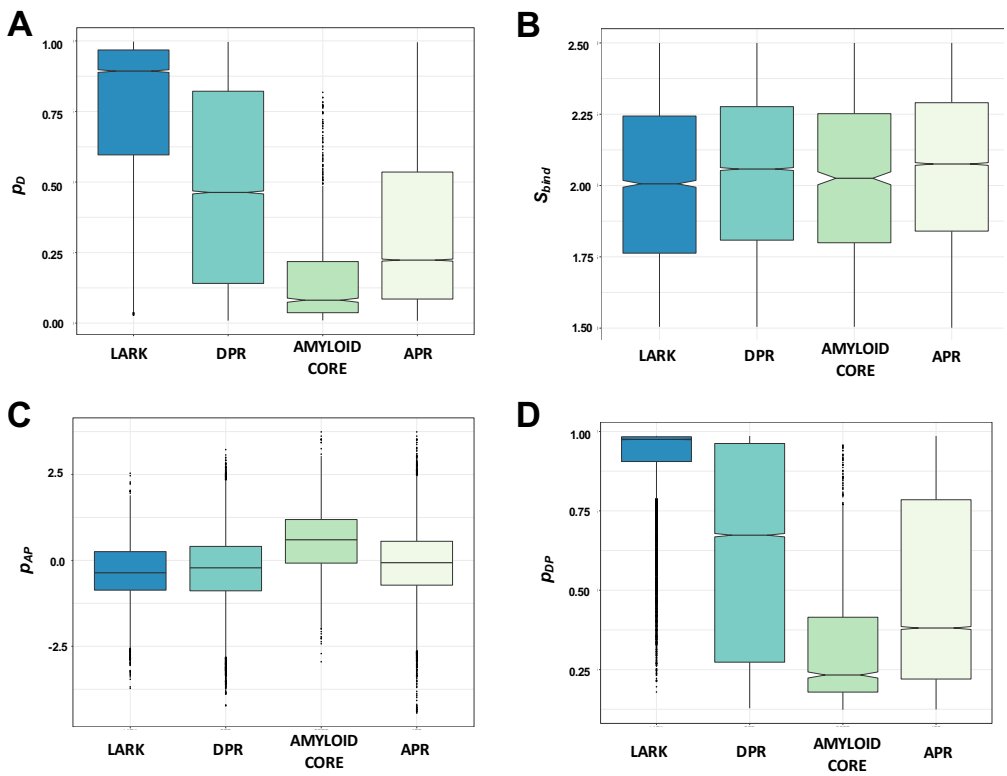


Figure S3. Analysis of droplet-promoting (DPR, aquamarine) **and amyloid-promoting** (APR, light green) **regions. (A) Free-state disorder** (p_D) was estimated by the fraction of disordered residues by Espritz NMR (44). Both droplet-promoting and amyloid-promoting regions sample a broad range of conformational states. Droplet-promoting regions are less disordered in the free state than LARKs (19) (blue), while amyloid-promoting regions are less ordered than amyloid cores (green). **(B) Interaction mode diversity** was defined by the entropy of possible binding modes (S_{bind}) as computed by FuzPred (32) (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 propensities** (p_{AP}) were derived from the solubility scores by CamSol (27). 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 propensities** (p_{DP}) by FuzDrop show 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.

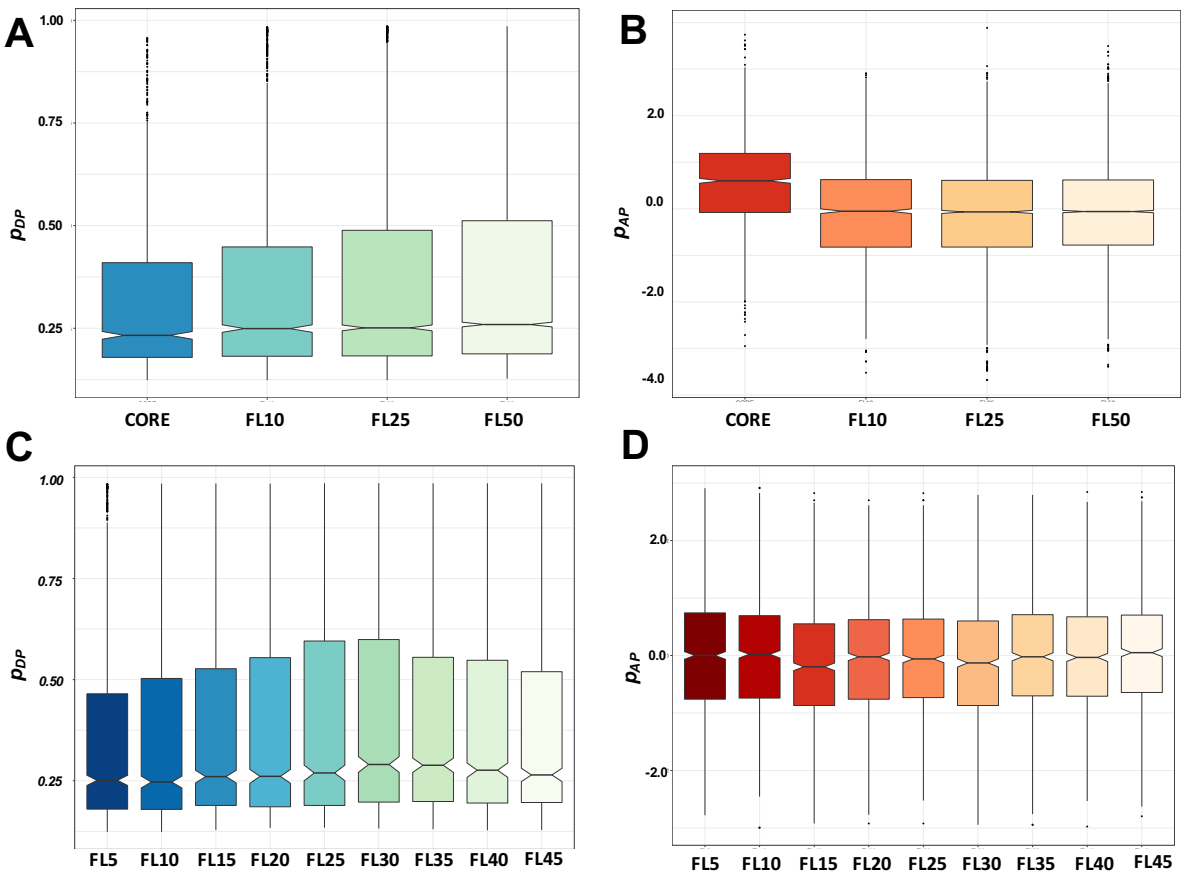


Figure S4. Droplet-promoting (p_{DP}) **and amyloid-promoting** (p_{AP}) **propensities of amyloid cores and their flanking regions. (A,B) Droplet-promoting** (p_{DP}) **and amyloid-promoting** (p_{AP}) **propensities as a function of flanking region length.** p_{DP} and p_{AP} values were evaluated in 10-residue (FL10), 25-residue (FL25) and 50-residue (FL50) flanking regions. Droplet-promoting propensities increase with increasing flanking region size. Amyloid-promoting propensities do not change considerably flanking region size. **(C,D) Droplet-promoting** (p_{DP}) **and amyloid-promoting** (p_{AP}) **propensities as a function of the distance from the amyloid core.** p_{DP} and p_{AP} values were evaluated in 5-residue segments, separation of which varied between 5 to 45 residues from the starting or terminating residue of the amyloid core. Residues, at 25-35 residue separation from the amyloid core exhibit the highest droplet-promoting propensities. Amyloid-promoting propensities exhibit negligible variations with increasing distance from the core. Amyloid-promoting propensities were derived from the solubility scores by CamSol (27), droplet-promoting propensities were computed by FuzDrop (33).

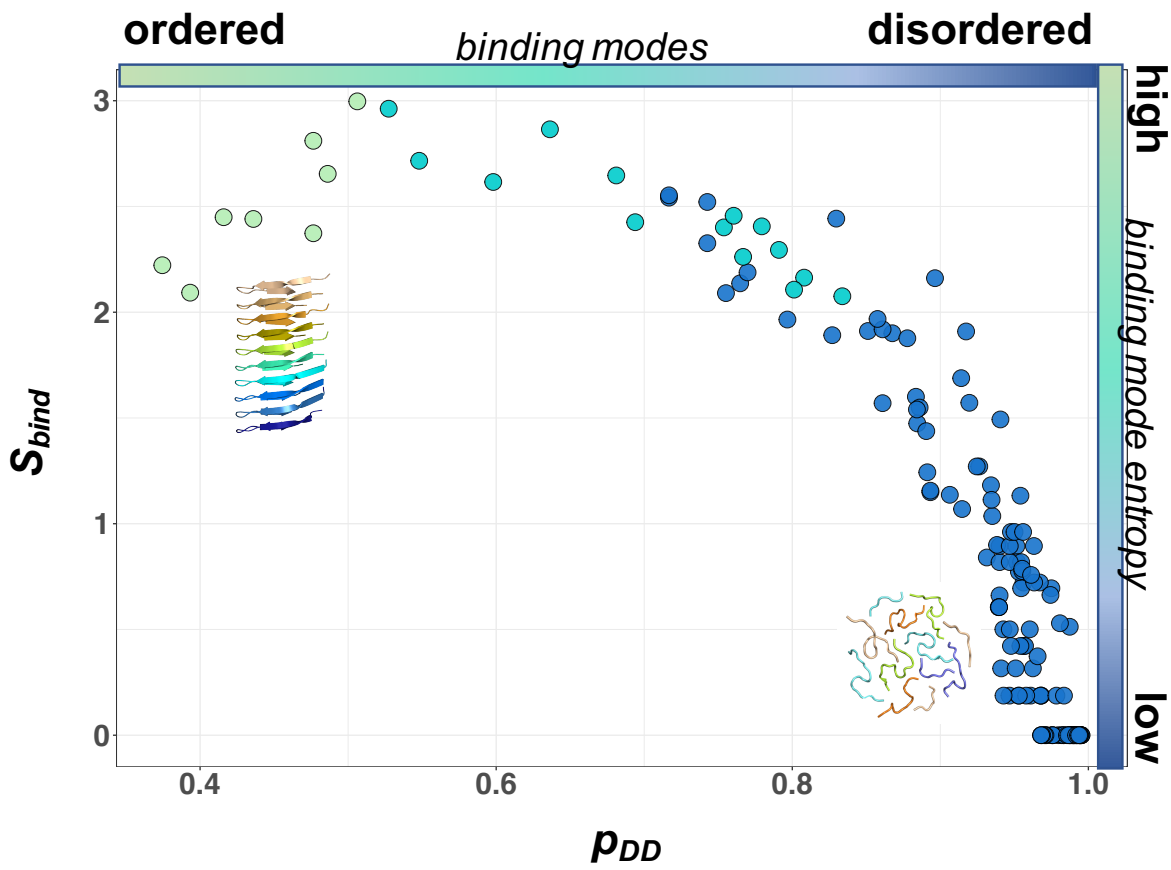


Figure S5. Binding landscape of the TDP-43 prion-like domain. The propensity of disordered binding modes (p_{DD}, Methods eq. 1) by FuzPred (43) is shown on the x-axis, and the binding mode entropy (S_{bind}, Methods Eq. 3), reflecting the multiplicity of interaction modes (32), on the y axis. Most residues 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 (321-330 residues, light green), exhibits more ordered binding modes with high interaction diversity (high S_{bind}) suggesting that it can sample both disordered and ordered interactions. The hot-spot region (aquamarine) also exhibits increased interaction diversity indicating an increased likelihood of conversion to the ordered interactions.

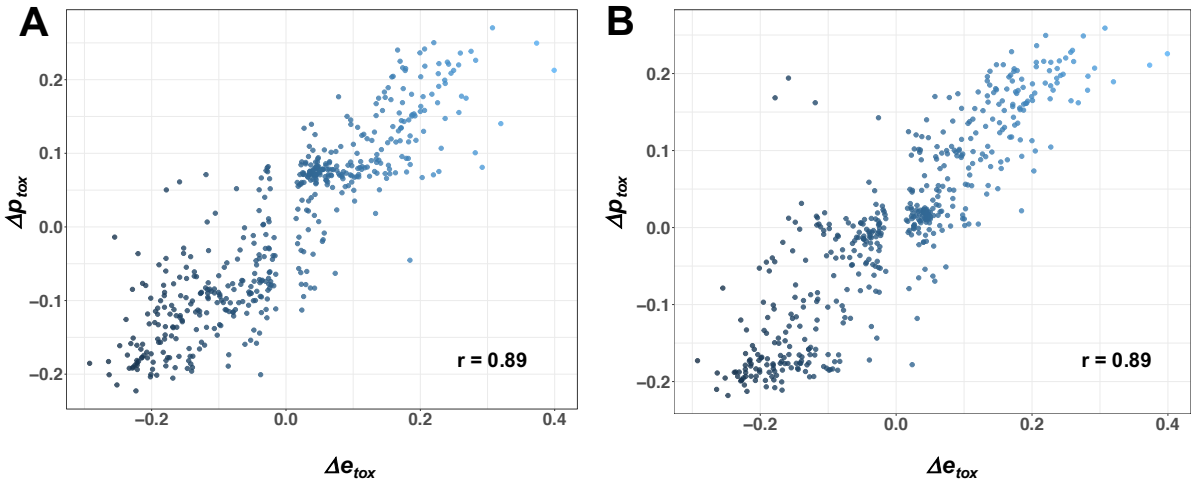


Figure S6. Correlation between experimental (Δe_{tox}) and predicted (Δp_{tox}) change in cellular toxicity upon TDP-43 single missense mutants based on values of individual amino acids (A) or the hot-spot region (312-342 residues). Toxicity was computed based on the changes in droplet-promoting (p_{DP}), and amyloid-promoting (p_{AP}) propensities and binding mode entropies (Methods, Eq. 6). The SVM model was developed using 498 single missense mutants with toxicity > 3σ (21). The Pearson's coefficients of correlation were computed using the R program.