Widespread occurrence of the droplet state of proteins in the human proteome

Maarten Hardenberg[a] , Attila Horvath[b] , Viktor Ambrus[c], Monika Fuxreiter[c,d,1], and Michele Vendruscolo[a,1]

[a]Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; [b]The John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; [c]Laboratory of Protein Dynamics, Department of Biochemistry and Molecular Biology, University of Debrecen, H-4010 Debrecen, Hungary; and [d]Department of Biomedical Sciences, University of Padova, 35131 Padova, Italy

A wide range of proteins have been reported to condensate into a dense liquid phase, forming a reversible droplet state. Failure in the control of the droplet state can lead to the formation of the more stable amyloid state, which is often disease-related. These observations prompt the question of how many proteins can undergo liquid–liquid phase separation. Here, in order to address this problem, we discuss the biophysical principles underlying the droplet state of proteins by analyzing current evidence for droplet-driver and droplet-client proteins. Based on the concept that the droplet state is stabilized by the large conformational entropy associated with nonspecific side-chain interactions, we develop the FuzDrop method to predict droplet-promoting regions and proteins, which can spontaneously phase separate. We use this approach to carry out a proteome-level study to rank proteins according to their propensity to form the droplet state, spontaneously or via partner interactions. Our results lead to the conclusion that the droplet state could be, at least transiently, accessible to most proteins under conditions found in the cellular environment.

protein condensates | protein droplets | liquid–liquid phase separation

It has been recently observed that proteins can self-assemble through a liquid–liquid phase separation (LLPS) process into a dense liquid phase, while maintaining at least in part their functional native states (1–4). These liquid-like assemblies of complex compositions are often referred to as biomolecular condensates or membraneless organelles (1–4). Here, we refer to these dynamic and reversible condensates as droplets, in order to distinguish them from irreversible amyloids. Droplets can concentrate cellular components to perform efficiently a variety of different functions, with an increasing number of biological roles being discovered (1–4).

In this work, we investigate whether liquid–liquid phase separation can be expected to be a proteome-wide phenomenon. In this view, the condensation of proteins from the native state to the amyloid state may quite generally proceed through an intermediate dense liquid phase, which is typically metastable (5) (Fig. 1). Different proteins may have different propensities to remain in this metastable phase, depending in particular on the free energy barrier between the droplet and amyloid states (Fig. 1). This type of liquid–liquid phase separation is indeed typical of condensation phenomena (1, 6), and sometimes is referred to as the Ostwald step rule (7). One may think that for most proteins the free energy barrier between the droplet and fibrillar states is low, and therefore the droplet state cannot be readily observed (Fig. 1). Indeed, this state may be difficult to detect due to a variety of reasons, including because experimental methods to probe its formation, in particular high-throughput ones, are still under development (8). Furthermore, our current understanding of the interactions that stabilize the metastable dense liquid phase is still incomplete.

Native and amyloid states are stabilized by specific interactions including hydrogen bonds, ionic interactions, and van der Waals contacts typical of ordered states and enthalpic in nature (9, 10). By contrast, in droplets, transient short-range aromatic cation–π and π–π, dipole–dipole, and electrostatic and hydrophobic interactions

have been observed, providing low-specificity, weak-affinity contacts characteristic of disordered states (11–16). These observations have led to a series of prediction methods (11, 13, 17–19), which focused on specific side-chain interactions. The redundancy and multivalency of the interacting elements (20) suggest that conformational entropy is a driving force of the condensation (21), also including main-chain contributions. Indeed, proteins exhibiting many binding configurations with a specific partner are often capable of forming droplets (22).

Here, we exploit the observation that many proteins exhibit high conformational entropy upon binding, which can be predicted from their amino acid sequences (23). Based on this result, we develop the FuzDrop method to predict the droplet-promoting propensity of proteins and their droplet-promoting profiles based on the conformational entropy of their free states and the binding entropy. Using this method, we identify a list of "droplet-driving" proteins, which are predicted to undergo spontaneous liquid–liquid phase separation under physiological conditions, and estimate that they comprise about 40% of the human proteome. In addition, we also predict that about 80% of the proteins are "droplet clients," characterized by short droplet-promoting regions in their sequences, which facilitate condensation via interactions with suitable partners. Taken together, our results indicate that protein phase separation is a proteome-wide phenomenon.

Significance

Liquid–liquid phase separation of proteins results in biomolecular condensates, which contribute to the organization of cellular matter into membraneless organelles. It is still unclear, however, whether these condensates represent a common state of proteins. Here, based on biophysical principles driving phase separation, we report a proteome-wide ranking of proteins according to their propensity to condensate into a droplet state. We analyze two mechanisms for droplet formation— driver proteins can spontaneously phase separate, while client proteins require additional components. We conclude that the droplet state, as the native and amyloid states, is a fundamental state of proteins, with most proteins expected to be capable of undergoing liquid–liquid phase separation via either of these two mechanisms.

Fig. 1. Liquid–liquid phase separation could be expected to be a proteome-wide phenomenon. 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 (the minima in the free energy), as well as the conversion rates between them (the barriers in the free energy), is different for different proteins. For most proteins under cellular conditions, the native and droplet states could be expected to be metastable (56), being kinetically trapped by a free energy barrier (ΔG) between the droplet and fibrillar states. Proteins that can be observed in the droplet state tend to have a high free energy barrier (LLPS; green) while the other ones tend to have a low free energy barrier compared with the thermal energy (non-LLPS; orange). For certain proteins the droplet state is functional, and it is stabilized by extrinsic factors, such as RNA and posttranslational modifications.

Results

A Framework to Describe the Interactions Stabilizing the Droplet State.

The premise of this work is that the droplet state is characterized by low-specificity interactions and liquid-like conformational entropy. Thus, we hypothesized that proteins that are conformationally heterogeneous in their native states and maintain this property upon binding would be particularly prone to form the droplet state. In estimating the degree of conformational heterogeneity in both the native and bound states, we observe that proteins span a continuum between structural order and disorder (23, 24), which we will express by the probabilities of p_D (free state) and p_{DD} (bound state). We also note that interactions with high conformational entropy are realized via many different binding configurations, which can be achieved by both ordered and disordered domains (25). By contrast, ordered binding modes with low conformational entropy are mediated by well-defined interfaces, as exemplified by rigid docking or templated folding (26).

Ordered and disordered binding modes exhibit characteristic sequence signatures. Motifs mediating ordered binding modes have a strong compositional bias as compared with their embedding protein regions. In contrast, motifs mediating disordered binding modes are more similar to their flanking regions, which can be realized via a variety of sequence patterns and contact types, as their specificity stems from their distinct character as compared with their flanking regions (23). We have previously demonstrated (27) that by identifying such interaction elements based on compositional bias, it is possible to estimate structural order or disorder under cellular conditions in excellent agreement with in vivo proteomic studies (28).

### Properties of Proteins That Can Form the Droplet State.

Datasets of proteins representing the droplet state. We have analyzed three public datasets of proteins reported to undergo liquid–liquid phase separation (*Materials and Methods*). The first is the PhaSepDB dataset (http://db.phasep.pro/) (29), which assembles data from three resources (*Materials and Methods* and Dataset S1): 1) proteins from the literature with in vivo and in vitro experimental data on liquid–liquid phase separation (REV, 351 proteins; *Materials and Methods* and Dataset S1), 2) proteins from UniProt associated with known organelles (UNI, 378 proteins; *Materials*

and *Methods* and Dataset S1), and 3) proteins identified by high-throughput experiments of liquid–liquid phase separation (HTS, 2,572 proteins; *Materials and Methods* and Dataset S1). The second dataset is PhaSePro (https://phasepro.elte.hu) (30), which identifies protein regions associated with liquid–liquid phase separation (PSP, 121 proteins; *Materials and Methods* and Dataset S1). The third dataset is LLPSDB (http://bio-comp.org.cn/llpsdb) (31), which assembles proteins observed to undergo in vitro liquid–liquid phase separation with well-defined experimental conditions and phase diagrams (*Materials and Methods* and Dataset S1). LLPSDB distinguishes whether proteins can phase separate spontaneously as one component (droplet-driving proteins, LPS-D, 133 proteins; *Materials and Methods* and Dataset S1) or require a partner to undergo liquid–liquid phase separation (droplet-client proteins, LPS-C, 41 proteins; *Materials and Methods* and Dataset S1). In this dataset, 77 proteins exhibit both droplet-driving and droplet-client behaviors.

To create a dataset for liquid–liquid phase separation, we merged the proteins in the REV, PSP, and LPS-D datasets, which we consider as drivers of droplet formation (453 unique proteins, LLPS dataset; *Materials and Methods* and Dataset S1). We generated two negative control datasets, one with human proteins only and another with a mixture of organisms (Dataset S2). For the human negative set (hsnLLPS dataset, 18,108 proteins; *Materials and Methods*), we excluded from the Swiss-Prot human proteome all proteins that appeared in any of the liquid–liquid phase separation datasets (REV, UNI, HTS, PSP, LPS-D, LPS-C) (29–31) (Dataset S2). For the negative set corresponding to multiple organisms (nsLLPS; *Materials and Methods*), we derived the organism distribution from the LLPS dataset. To build a control dataset, we considered organisms populated more than 1% in the LLPS dataset and used their proteomes from UniProt (*Caenorhabditis elegans, Chlamydomonas reinhardtii, Drosophila melanogaster, Homo sapiens, Mus musculus, Rattus norvegicus, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Xenopus laevis; Materials and Methods*) and removed all proteins present in the LLPS or HTS datasets. We then randomly chose sequences according to the frequencies of these organisms in the LLPS dataset with 10 times enrichment (nsLLPS; *Materials and Methods* and Dataset S2).

Analysis of the amino acid compositions of droplet-driving and droplet-client proteins. Droplet-driving proteins are enriched in disorder-promoting residues (P, G, S) and depleted in order-promoting (F, I, V, C, W) residues as compared with non–phase-separating proteins (Fig. 2A). N and Q, which are distinguished in prion-like domains (32), are more abundant in droplet-driving proteins than those not reported to undergo LLPS. However, droplet-driving proteins are not significantly enriched in residues that mediate π–π and cation–π interactions (Y, R), as compared with non–phase-separating (nsLLPS) proteins (Fig. 2A and Dataset S3). These results indicate that droplet formation does not depend on a specific contact type but can rather be realized in many ways by low-specificity interactions. The composition of droplet-driving proteins is in between that of globular proteins (33) and disordered proteins in the DisProt database (33), more abundant in order-promoting residues (W, C, Y, F, I, V) as compared with disordered proteins (*SI Appendix*, Fig. S1B), and enriched in disorder-promoting residues (P, D, E) as compared with globular proteins (34) (*SI Appendix*, Fig. S1A). Aromatic residues observed in disordered regions, for example in nucleoporins, often mediate low-affinity interactions (35). These compositional properties reflect the preference of droplet-driving proteins for the disordered state in the bound form, which is comparable to protein complexes with disordered binding modes (23).

As compared with non–phase-separating proteins, droplet-client proteins are enriched in charged residues (D, K, E) and disorder-promoting prolines (Fig. 2B and Dataset S3). Droplet-client proteins exhibit characteristic differences from droplet-driving proteins, as they are enriched in charged residues (K,
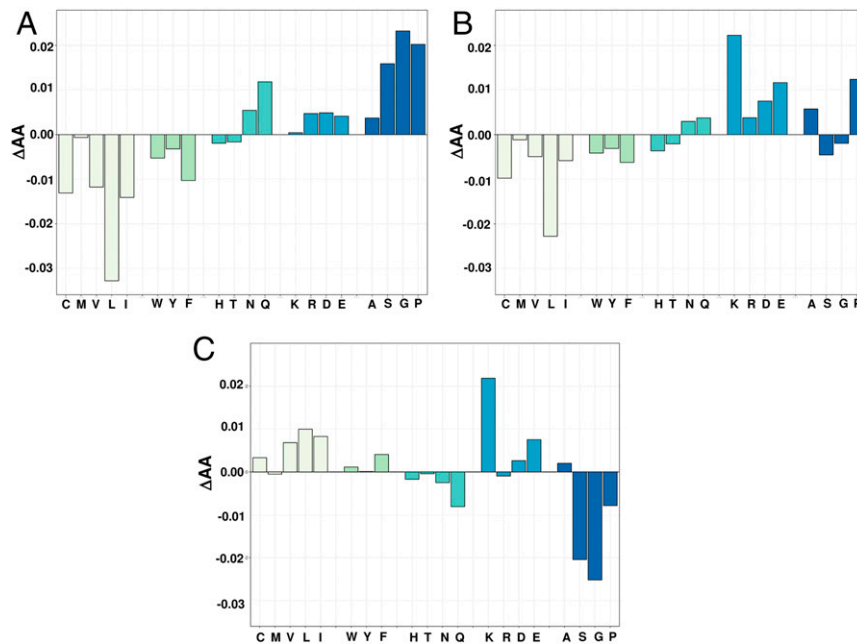
Fig. 2. Differential amino acid compositions of droplet-driver and droplet-client proteins. (*A*) Differences in amino acid compositions (ΔAA) of droplet-driver proteins in the LLPS dataset and of proteins not reported to phase separate (nsLLPS). (*B*) Differences in amino acid compositions of droplet-client proteins that require additional components for phase separation (LPS-C dataset) and of proteins that have not been reported to phase separate (nsLLPS). (*C*) Differences in amino acid compositions of droplet-client proteins (LPS-C) and droplet-driver proteins (LLPS). Amino acids grouped as hydrophobic (light green), aromatic (green), hydrophilic (turquoise), charged (steel blue), and disorder-promoting (dark blue) (34). The SEs and the significances of the differences by Kolmogorov–Smirnov test are shown in Dataset S3.

E) and hydrophobic motifs (L, V, I), while being depleted in amyloid-promoting (N, Q), phosphorylation-promoting (S), and disorder-promoting (G) residues (Fig. 2*C* and Dataset S3). The amino acid composition of droplet clients is thus more similar to structured than disordered proteins (*SI Appendix*, Fig. S1*B*).

Analysis of the conformational entropy of droplet-driving and droplet-client proteins. We observed that different protein datasets representing the droplet state have markedly different characteristics in their conformational entropy in the free state and its change upon binding. Drivers of droplet formation (LPS-D) have high levels of disorder in free (p_D) and bound states (p_{DD}), while droplet clients (LPS-C) are mostly ordered in both forms (Fig. 3 *A* and *B*). Proteins in the REV and PSP datasets exhibit disordered binding modes, which are comparable to droplet-driver proteins, so they likely phase separate spontaneously. Proteins associated with known membraneless organelles (UNI) or identified by high-throughput experiments (HTS) (29) have significantly lower conformational entropy in both free and bound states, and thus likely have components that form droplets via partner interactions. Comparison of spontaneously phase-separating and non–phase-separating proteins (Fig. 3 *C* and *D*) indicates that a high conformational entropy is a characteristic of the droplet state.

Sequence-Based Prediction of Droplet Propensity Profiles of Proteins. Based on the analysis reported above, in this section, we present a method of predicting the sequence-based profile of the propensity of proteins to form spontaneously the droplet state (p_{DP}). To achieve this result, we define the probability of residue A_i to be involved in spontaneous phase separation by $p_{DP}(A_i)$ using a binary logistic model as

$$p_{DP}(A_i) = \frac{exp F_S(A_i)}{1 + exp F_S(A_i)}, \qquad [1]$$

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

$$F_S(A_i) = \lambda_1 p_D(A_i) + \lambda_2 p_{DD}(A_i) + \gamma, \qquad [2]$$

where $p_D(A_i)$ is the probability for disorder in the free state and $p_{DD}(A_i)$ is the probability for disordered binding (23). $p_D(A_i)$ contains an estimate of the conformational entropy in the unbound form, while $p_{DD}(A_i)$ contains an estimate of the binding entropy. 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 (*Materials and Methods* and Dataset S4). p_D was derived from the disorder score as computed using the ESpritz NMR algorithm (36), with the best performance on disordered protein complexes (23). The p_{DD} values were predicted by the FuzPred method, which describes binding modes under cellular conditions (27). The p_D and p_{DD} values capture the balance between enthalpy and entropy that stabilizes the droplet state, which is associated with the nonspecific nature of a variety of side-chain interactions.

To train our model, we used a dataset of droplet-promoting regions, with evidence to mediate spontaneously phase separation (*Materials and Methods* and Dataset S1). As a negative set, we defined regions in non–phase-separating proteins with the same length distribution as in the positive set (*Materials and Methods*). The size of the negative set was 10 times that of the positive set and we applied stratified sampling in the training. We found that the linear coefficients were robust over many random selections of the positive and negative sets, as well as the training set size (Dataset S4). In the final parameterization, the linear coefficients of both disorder and binding modes were positive, reflecting the preference for a disordered bound state in the droplets. The threshold to mediate droplet formation was derived from the binary logistic model ($p_{DP} \geq 0.60$).

To estimate the performance of the method, we calculated an area under the curve (AUC) value of 87.0% on the training set and an AUC value of 85.9% on the test set (*Materials and Methods* and Dataset S4). We applied these coefficients to all
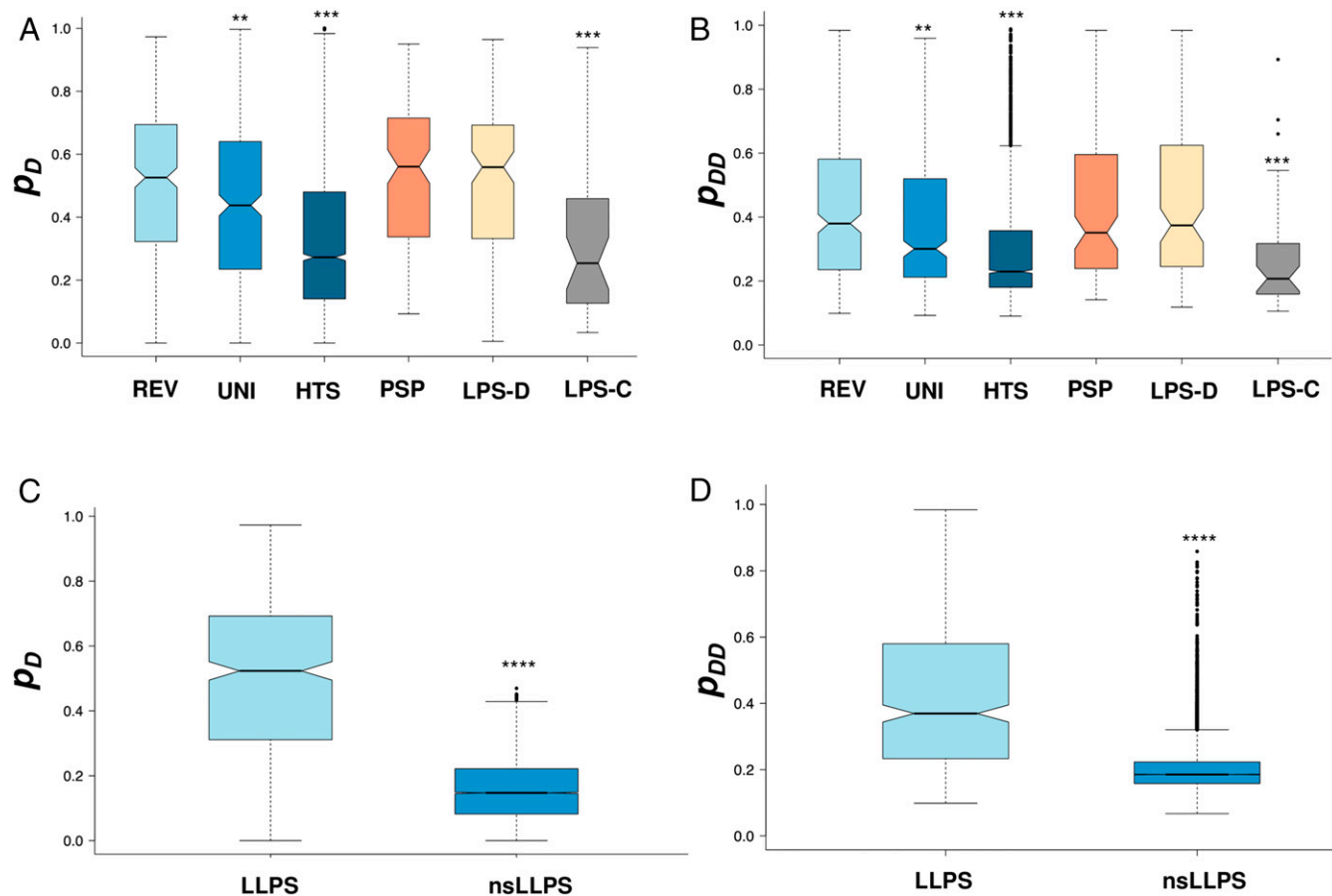
Fig. 3. Conformational properties in different datasets of LLPS proteins in the free and bound states. PhaSepDB literature reviewed (light blue), PhaSepDB human organelle-associated proteins from UniProt (steel blue), PhaSepDB proteins identified by high-throughput experiments (dark blue), PhaSePro (orange), LLPSDB one-component proteins (droplet drivers; wheat), and two-component (droplet clients; gray) phase-separating proteins. (A) The probability for the disordered state (p_D) in the free form was characterized by the fraction of disordered residues, as computed by the ESpritz NMR program (36). Residues are classified as disordered if they have an ID score ≥0.3089. The fraction of disordered residues was computed per protein as N_{ID}/N_{AA} and these values were averaged for each dataset. (B) Probability for disordered binding (p_{DD}) was computed by the FuzPred program (23). The median p_{DD} value was determined for each protein and these values were averaged for each dataset. (C and D) Comparison of p_D and p_{DD} in droplet-driving (LLPS; light blue) and non–phase-separating proteins (nsLLPS; dark blue). Statistical significances were computed by the Mann–Whitney U test using the R program (**$P < 10^{-3}$, ***$P < 10^{-5}$, ****$P < 10^{-10}$).

droplet regions and obtained an AUC value of 84.4%. These results illustrate that the parameters are robust across droplet regions from different organisms. We also note that droplet-promoting propensity profiles of proteins that were observed to form droplets under cellular conditions and those that were detected only by in vitro experiments are not significantly different (*SI Appendix*, Fig. S3).

We have thus developed the FuzDrop method to predict droplet-promoting propensities of residues from the primary sequence based on the conformational entropy of disordered binding modes in droplets.
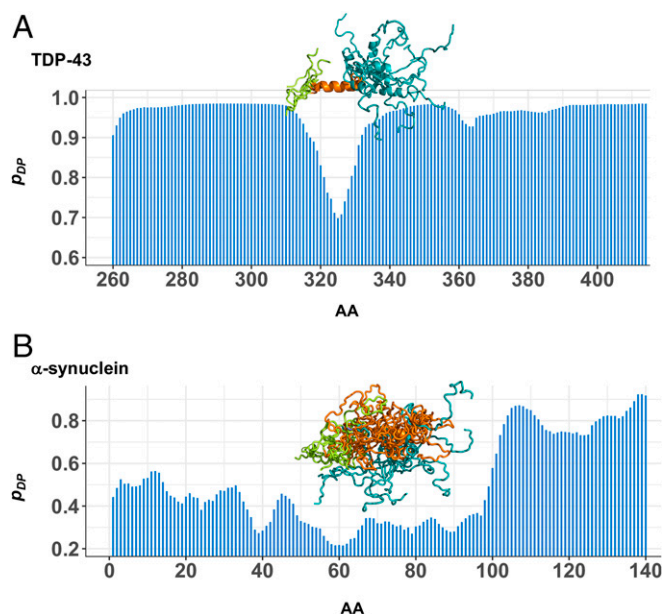
Droplet-Promoting Propensity Profiles of TDP-43 and α-Synuclein. We applied the FuzDrop method to predict the droplet-promoting propensities of two proteins reported to undergo liquid–liquid phase separation, TDP-43 (37) and α-synuclein (38, 39). Our results indicate that the low-complexity region of TDP-43 (residues 262 to 414) mediates spontaneous phase separation. We note that the α-helical segment (residues 320 to 331), which constitutes the amyloid core in TDP-43 fibrils (40) (Fig. 4A) and is predicted to undergo disorder-to-order transition upon binding, also has a high droplet-promoting propensity (Fig. 4A).

In the case of α-synuclein, the highly disordered C-terminal region (residues 98 to 140), which also remains disordered upon binding to lipid vesicles (41), is predicted to drive the formation of the droplet state (Fig. 4B). The central non-amyloid beta component (NAC) region has lower p_{DP} propensity to spontaneously phase separate, but may be involved in droplets via hydrophobic protein interactions, which are absent from β-synuclein and γ-synuclein (38).

Sequence-Based Prediction of Droplet-Driving Proteins. In this section, we present a method of ranking proteins according to their propensity to form the droplet state. In order to achieve this result, we estimate the probability of liquid–liquid phase separation (p_{LLPS}) using a binary logistic model (*Materials and Methods*) with a scoring function (F_{LLPS}) derived from residue droplet-promoting propensities and a term for hydrophobic interactions

$$F_{LLPS} = \lambda_1 * median\{p_{DP}(A_i)\} + \lambda_2 * n_{DPR} + \lambda_3 * H + \gamma, \quad [3]$$

where $median\{p_{DP}(A_i)\}$ is the median of the residue droplet-promoting propensities, n_{DPR} is the number of long droplet-promoting regions (DPRs; ≥25 consecutive residues with $p_{DP} \geq 0.6$),

Fig. 4. Droplet-promoting propensity profiles (p_{DP}) of the TDP-43 low-complexity (LC) domain and of α-synuclein. (*A*) The TDP-43 LC domain has overall high droplet-promoting propensities. The depletion in the droplet profile corresponds to the α-helical segment (orange), which is involved in the amyloid core. The N- (lime) and C- (blue) flanking regions are disordered in the NMR structure of the G335D mutant (PDB ID code 2n4g). (*B*) The disordered C-terminal region of α-synuclein (blue) is predicted to drive droplet formation. The N-terminal region (lime), which folds into an α-helix, has intermediate p_{DP} values. The ensemble is derived from the Protein Ensemble Database (PED9AAC). The p_{LLPS} threshold is indicated by a bold gray line.

and *H* is a hydrophobic term (≥6-residue hydrophobic motifs within disordered regions) (*Materials and Methods*). λ_1, λ_2, and λ_3 are the linear coefficients of the predictor variables and γ is a scalar constant (intercept), which we determined on the LLPS$_{train}$ and nsLLPS$_{train}$ datasets (*Materials and Methods* and Dataset S5). We found that the linear coefficients were robust over many random selections of the positive and negative sets,

as well as the training set size (Dataset S5). The threshold to mediate spontaneous liquid–liquid phase separation was derived from the binary logistic model ($p_{LLPS} \geq 0.61$). We propose that the p_{LLPS} value expresses the droplet-driving potential under physiological conditions, as droplet-promoting propensities of proteins that form droplets under physiological conditions and those that were detected to phase separate only in vitro do not deviate significantly (*SI Appendix*, Fig. S2). We also note that using nonphysiological conditions, such as high concentrations of protein and crowding agents, can induce liquid–liquid phase separation at p_{LLPS} values below the threshold, especially if droplet-promoting regions are present.

To estimate the performance of the method, we calculated an AUC value of 88.3% on the training set (0.75 of the LLPS dataset) and an AUC value of 90.7% on the test set, using stratified sampling (*Materials and Methods* and Dataset S5). As an attempt to further improve performance, we incorporated a π–π term (19) into the scoring function of the logistic model (*Materials and Methods*). Adding this term slightly increased the performance of the model (AUC 92.2%; Dataset S5) with a moderate contribution to the scoring function. These results are in accord with the presence of π–π interactions in many droplet proteins, but also show that these interactions are not prerequisites for droplet formation.

The performance and robustness of the model (Eq. **3** and Dataset S5) demonstrate that the droplet state can be predicted from sequence based on the estimated conformational entropy of binding and a nonspecific enthalpy term. We also note that our model by Eq. **3** serves as a general framework for predicting droplet-driver proteins. Accumulating data collected using more systematic and uniform experimental approaches (8) will enable further refinement of the parameters in our model and to predict the minimum concentration for phase separation, although this property can be expected to be highly dependent on the cellular conditions.

Region Specificity of the FuzDrop Method and Experimental Validation of the Predictions. We note that estimates of the overall propensity of a protein to form the droplet state cannot be readily obtained by a simple average of the values of the profiles of Eq. **2**. This overall propensity is also determined by specific regions, rather than only by the general properties of the entire sequence, including in particular droplet-promoting regions and
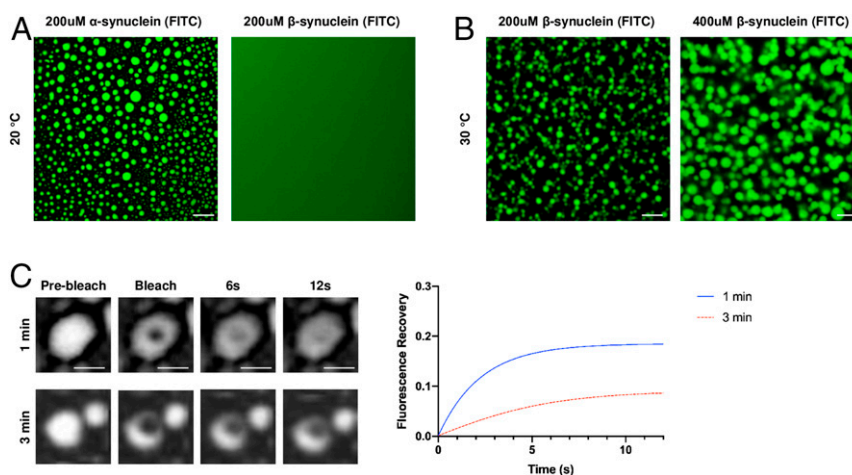
Fig. 5. Region-specific phase behavior of α-synuclein and β-synuclein. (*A*) FITC-labeled β-synuclein (p_{LLPS} 0.54), which lacks the characteristic NAC region found in α-synuclein, does not phase separate at high concentrations (200 μM) and under crowding conditions (10% [weight/volume] PEG), whereas FITC-labeled α-synuclein (p_{LLPS} 0.62) readily forms droplets under the same conditions. (*B*) Increasing the experimental temperature by 10 °C does lead to rapid coalescence of β-synuclein into micrometer-sized droplets. (*C*) Rapid FRAP of a small area within a droplet (*Top*) 1 min after phase separation; FRAP 3 min after phase separation (*Bottom*); and a nonlinear fit of fractional fluorescence recovery over time (*Right*). (Scale bars, 10 μm [*A* and *B*] and 5 μm [*C*].)

short motifs within disordered regions, which are prone to establish hydrophobic interactions (Eq. **3**). This point can be illustrated by distinct behaviors of α-synuclein and β-synuclein (38). The C-terminal region of both proteins possesses a droplet-promoting region, with a preference for disordered binding modes (Fig. 4 and *SI Appendix*, Fig. S3). In addition, the NAC region of α-synuclein contains eight hydrophobic residues, biased for disordered binding, which can exert a nonspecific driving force (resembling hydrophobic collapse) for droplet formation. Notably, however, β-synuclein and γ-synuclein, which lack these residues (*SI Appendix*, Fig. S3), were not observed to undergo liquid–liquid phase separation under physiological conditions (38).

The predicted p_{LLPS} values by the FuzDrop method (0.62 for α-synuclein, 0.54 for β-synuclein, and 0.40 for γ-synuclein) suggest that β-synuclein and γ-synuclein have lower propensity to adopt the droplet state as compared with α-synuclein. Indeed, γ-synuclein did not phase separate under any of the experimental conditions tested (38). To validate the predictions close to the prediction threshold, we explored β-synuclein phase behavior in a set of in vitro experiments (Fig. 5). In line with previous observations (38, 39), we did not observe any droplets after incubating high concentrations of fluorescein 5-isothiocyanate (FITC)-labeled β-synuclein on a glass surface, whereas we did observe droplets for FITC-labeled α-synuclein (Fig. 5 *A* and *B*). As hydrophobic effects are important for α-synuclein droplet formation and considering that β-synuclein lacks the predominantly hydrophobic segment in the NAC region, we reasoned that raising the experimental temperature would increase the strength of residual hydrophobic interactions, allowing the protein to cross the phase barrier. Indeed, β-synuclein formed micrometer-sized droplets when the temperature was raised by 10 °C and at high concentrations (Fig. 5C). Droplets formed by FITC–β-synuclein were initially liquid-like, as evidenced by fluorescence recovery after photobleaching (FRAP), but showed rapid conversion to a gel-like state (Fig. 5C). The phase separation behavior of β-synuclein illustrates that protein phase separation is highly dependent on the experimental conditions, that proteins with a predicted p_{LLPS} below the threshold ($p_{LLPS} \geq 0.61$) require more extreme conditions to adopt the droplet state, and that the droplet state of these proteins is generally short-lived.

As an additional test of our predictions, we ranked a set of proteins associated with Alzheimer's disease (42) based on their predicted FuzDrop scores (Dataset S6) and selected one of the top candidates, complexin-1, to experimentally test our predictions (Fig. 6). To assess whether complexin-1 can form droplets through liquid–liquid phase separation, we incubated Alexa 488-labeled complexin-1 on a glass surface under crowding conditions at physiological pH (*Materials and Methods*). After a brief lag phase (<1 min), complexin-1 formed micrometer-sized droplets in suspension (Fig. 6A). The droplets were characteristic of a liquid phase, as they showed distinct wetting behavior after prolonged incubation (>10 min) (Fig. 6A) and fused upon making contact (Fig. 6B). Furthermore, molecules within the droplets showed local rearrangement, as evidenced by rapid FRAP (Fig. 6C). We also predicted that the disordered N-terminal region of complexin-1 drives its liquid–liquid phase separation (*SI Appendix*, Fig. S4). This region cooperatively interacts with the SNARE complex and plasma membrane (43) to facilitate synaptic vesicle fusion (44). Phase separation may contribute to activation of complexin-1 by relieving its autoinhibition, which is a common mechanism by the droplet state (21).

Droplet-Driving and Droplet-Client Proteins in the Human Proteome. We applied the prediction method to estimate the proteins capable of undergoing spontaneous liquid–liquid phase separation (droplet-driving proteins) in the Swiss-Prot human proteome. We thus ranked the proteins in the human proteome according to their propensity to form the droplet state (Dataset S7), and estimated that about 40% of them are capable of spontaneous droplet formation.

This list contains only about 60% of the human proteins currently associated with membraneless organelles (UNI). This fraction is even lower for proteins identified by high-throughput experiments (HTS), including organelle purification (45, 46), affinity purification (47, 48), immunofluorescence image-based screen (49, 50), and proximity labeling (51, 52) (*SI Appendix*, Fig. S5). As the FuzDrop approach was developed for proteins that drive droplet formation, our results indicate that membraneless organelles contain also proteins that undergo phase separation by being driven by a partner (droplet-client proteins). We observed that droplet clients have a lower conformational disorder in both free and bound states (Fig. 3), suggesting the involvement of distinguished, local motifs. Thus, the droplet-client mechanism can provide a route for structured proteins to be engaged in condensates via specific droplet-promoting regions.

To investigate the properties underlying the droplet-client mechanism, we analyzed the presence of long and short droplet-promoting regions in the droplet-driver (LLPS) and droplet-client (LPS-C) datasets (Table 1). We found that ~90% of droplet-client proteins contain a short droplet-promoting region (≥10 residues), while only ~60% have long ones (≥25 residues). The frequency of short and long droplet-promoting regions in proteins, identified by high-throughput experiments, is comparable to droplet-client proteins (Table 1), indicating that they follow a partner-induced client mechanism. In contrast, the frequency of droplet-promoting regions in proteins associated with human membraneless organelles is comparable to droplet drivers (Table 1). Considering their lower droplet-promoting propensities (Dataset S7), these results indicate that proteins in membraneless organelles likely follow both driver and client mechanisms.
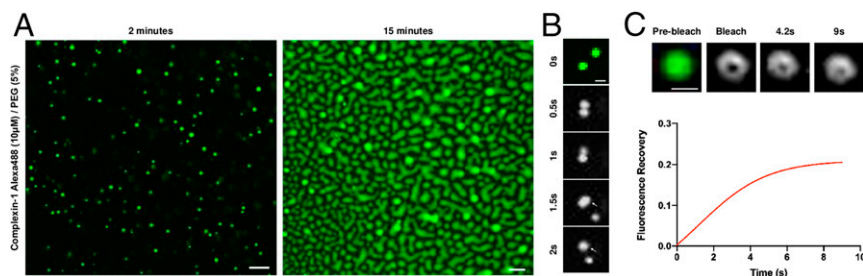
Fig. 6. Complexin-1 undergoes liquid–liquid phase separation. (*A*) Alexa 488-labeled complexin-1 (10 μM) coalesces into micrometer-sized droplets under crowding conditions (*Left*). Droplets exhibit a wetting phenotype when encountering a glass surface (*Right*). (*B*) Complexin-1 droplets readily fuse when in close proximity (<1 μm) and relax into a round structure after fusion, as noted by the arrows. (*C*) Rapid FRAP of a small area within a droplet (*Top*); nonlinear fit of fractional fluorescence recovery over time (*Bottom*). (Scale bars, 5 μm [*A*] and 1 μm [*B* and *C*].)

Overall, we thus estimate that over 80% of the proteins in the human proteome contain regions that can mediate droplet formation. Half of these proteins can condensate spontaneously, while the other half can do so by interacting with other components (Table 1). We have also observed that the number of droplet-promoting regions is comparable in proteins observed to form droplets under physiological conditions or detected by in vitro experiments (*SI Appendix*, Fig. S2), corroborating the relevance of the predictions under cellular conditions. We then extended these results to other organisms (Dataset S8), leading to the suggestion that the droplet state is a proteome-wide phenomenon.

Discussion and Conclusions

Increasing evidence indicates that a wide range of proteins unrelated in sequence, native structure, and function can form biomolecular condensates (1, 2, 4, 53). These observations suggest that the droplet state may have a generic nature and be accessible to most proteins. This possibility may not be immediately evident from the data currently available because the condensation of different proteins has been reported for experimental conditions often far from physiological ones. Moreover, a full understanding of the interactions driving droplet formation has not been achieved yet, owing to a wide variety of sequence motifs associated with the droplet state.

In this work, we have exploited that a large fraction of the proteins in the human proteome have favorable binding entropies by visiting an ensemble of bound states (54, 55), which is realized via disordered binding modes. We thus hypothesized that the high conformational entropy associated with nonspecific side-chain interactions contributes to the stabilization of the droplet state, and proposed a model to quantify it from its sequence. We have shown that droplet-promoting propensities can be predicted using such a generic model, even without the explicit incorporation of specific types of interactions. The specificity of our model originates from local compositional sequence biases, which are used to estimate the entropy in the bound state (23). That is, both hydrophobic and hydrophilic motifs can selectively mediate interactions if they are embedded in an environment of opposite character, explaining how selectivity can be achieved via a wide variety of interactions and contact types. We have shown earlier that this approach is capable of describing ordered and disordered binding under cellular conditions (27).

Using these general principles, we developed the FuzDrop method to predict droplet-promoting profiles and propensity of proteins to drive droplet formation. Applying this prediction method to different datasets of phase-separating proteins, we described two mechanisms of droplet formation: 1) the driver mechanism, which does not require additional components for phase separation, and depends on the overall conformational entropy of the protein, and 2) the client mechanism, which is induced by protein interactions, and is dependent on the presence of specific droplet-promoting regions in the sequence of the protein. Our results indicate that proteins may use the driver or the client mechanisms, or a combination of them, to form droplets.

Our proteome-wide analysis indicates that the presence of droplet-promoting regions is widespread in the sequences in the human proteome. Based on this analysis, we conclude that the droplet state is accessible, even if only transiently, for most proteins. In ~40% of the human proteome it is predicted to occur spontaneously, whereas an approximately equal fraction may require a variety of cellular components or nonphysiological conditions. Proteins in known membraneless organelles represent a combination of these mechanisms, whereas those identified by high-throughput studies mostly represent droplet clients.

Taken together, these results indicate that the droplet state is likely to be a fundamental state of proteins, alongside the native and amyloid states.

Materials and Methods

Datasets of Phase-Separating Proteins. All data in the present study were downloaded from public datasets without modifications (Dataset S1). The REV, UNI, and HTS datasets were assembled from the PhaSepDB dataset (http://db.phasep.pro/) (29). The 351 proteins in the REV dataset were collected based on a curated literature search; the 378 proteins in the UNI dataset were associated with human organelles in UniProt; and the 2,572 proteins in the HTS dataset were identified in high-throughput experiments. The PSP dataset contained 121 proteins from the PhaSePro database (https://phasepro.elte.hu) (30) with regions involved in LLPS identified. The 174 proteins in the LLPSDB dataset (http://bio-comp.org.cn/llpsdb) (31) were observed to undergo in vitro liquid–liquid phase separation for which the experimental conditions were also specified. All proteins observed to form droplets spontaneously were assigned to the LPS-D dataset and only those whose phase separation was dependent on interactions with a partner were in the LPS-C dataset. The LLPS dataset contained 453 nonredundant proteins, by merging the REV, PSP, and LPS-D datasets (Dataset S1). The 144 regions identified to mediate droplet formation were assembled from the PhaSePro dataset (30), and were grouped based on the evidence for spontaneous or partner-assisted phase separation in the LLPSDB dataset (31) (DPR; Dataset S1).

Datasets of Non–Droplet-Forming Proteins. All human proteins included in the phase separation datasets (LLPS and LPS-C) were removed from the Swiss-Prot human proteome, resulting in 18,108 sequences (hsnLLPS; Dataset S2). We also generated a negative set for phase separation (nsLLPS; Dataset S2), which reflected the composition of the LLPS dataset using organisms represented >1% in the LLPS dataset (*C. elegans, C. reinhardtii, D. melanogaster, H. sapiens, M. musculus, R. norvegicus, S. cerevisiae, S. pombe, X. laevis*). Only Swiss-Prot sequences were used except for *X. laevis*. Sequences were randomly chosen from these pools to match their frequency in LLPS. The size of the nsLLPS dataset was 10 times more than that of the LLPS dataset.

Analysis of amino acid compositions. The properties of LLPS proteins were compared with proteins with disordered regions in the DisProt v7 database (34) and the composition of globular proteins from the Protein Data Bank (PDB) (33). We used a bootstrap approach to compare the amino acid compositions in proteins of the LLPS, LPS-C (Dataset S1), and nsLLPS (Dataset S2) datasets and the statistical significance of the pairwise differences was determined by a two-sample Kolmogorov–Smirnov test of the R program (Dataset S3). We also computed the absolute maximum distances between

Table 1. Percentage in different datasets of proteins containing regions predicted to be droplet promoting

	$p_{DPR} \geq 25$ residues, %	$p_{DPR} \geq 10$ residues, %
H. sapiens	62	83
Membraneless organelles	84	94
High-throughput experiments	65	87
Droplet drivers	87	94
Droplet clients	60	88

Droplet-promoting regions were identified with the number of consecutive residues with $p_{DP} \geq 0.6$, either ≥ 25 (column 2) or ≥ 10 (column 3). The list presented includes membraneless organelles (UNI), high-throughput experiments (HTS), droplet drivers (LLPS; Dataset S1), and droplet clients (LPS-C; Dataset S1).

the cumulative distribution functions (Dataset S3). SE was calculated as $SE = SD/\sqrt{n}$, where SD represents the SD of the bootstrapped differences and n represents sample size.

Predicting residue droplet-promoting propensity.

Binary logistic regression model. Droplet-promoting propensity (p_{DP}) was defined as a probability of a binary response, whether a residue can promote spontaneous phase separation or not. We used two predictor variables (Eqs. **1** and **2**): 1) The probability of disorder in the free state (p_D) was predicted by the ESpritz NMR program (36), and 2) the probability of disordered binding (p_{DD}) and was computed by the FuzPred program (23). These two quantities approximated the conformational entropy in the free state and its change upon binding.

Training and parameterization. As a positive set, we used 67 droplet-promoting regions, with evidence for mediating spontaneous phase separation (Dataset S1). As a negative set, we randomly chose regions from proteins in nine representative organisms (*C. elegans*, *C. reinhardtii*, *D. melanogaster*, *H. sapiens*, *M. musculus*, *R. norvegicus*, *S. cerevisiae*, *S. pombe*, *X. laevis*) without evidence to spontaneously form droplets or serve as droplet clients. Frequencies of proteins were set according to the droplet dataset (Dataset S1) with a length distribution matching that of the positive DPR dataset. The size of the negative set was 10 times that of the positive set and we applied stratified sampling.

We used the R program to determine the coefficients of the independent variables (p_D and p_{DD}; Eqs. **1** and **2**) on the training set, which were chosen as 0.6 to 0.8 of the positive DPR set. The performance of the different models was evaluated based on AUC, specificity, sensitivity, and accuracy, which were computed by the R program (Dataset S4). Owing to the length dependence of the characteristics of the droplet-promoting regions, we used the coefficients obtained for regions <200 residues. The threshold for droplet-promoting propensity ($p_{DP} \geq 0.5994$) was determined based on the logistic model, and was in good agreement with the training and test sets.

Predicting the propensity of proteins to drive droplet formation.

A binary logistic model was used to estimate the probability of a binary response, whether a protein spontaneously forms droplets or not, based on three predictor variables (Eq. **3**): the median of the residue-based p_{DP} values, the number of droplet-promoting regions ($n_{DPR} \geq 25$ residues), and a factor representing weak hydrophobic interactions. To distinguish between hydrophobic interactions driving structure formation and those in droplets, we used hydrophobic motifs (≥ 6 consecutive residues), which were located in disordered regions. The threshold was set to -1.3 based on the Kyte–Doolittle hydrophobicity scale, to include S, T, and Y capable of undergoing phosphorylation. As the data regarding droplet-forming proteins are rapidly expanding, we aimed to use a general model, which can be reoptimized if more and more specific information will become available.

Training and parameterization. We divided the positive datasets LLPS and hsLLPS into training and test sets using various random selections, varying the training test size between 65 and 85%, and applied stratified sampling for the negative nsLLPS and hsnLLPS sets (Dataset S2). For parameterization, we removed "uncharacterized," "putative" proteins, and "coil-coiled" domains from the nsLLPS and hsnLLPS datasets. Owing to their repetitive sequences, coiled-coil domains still present a challenge for disorder predictions. We used the R program to determine the coefficients for the independent variables on the LLPS$_{train}$ and hsLLPS$_{train}$ datasets (Dataset S5). To decide the final coefficients, we aimed at high sensitivity, as we expected many false positives in the negative datasets (proteins not yet reported to form droplets) and we aimed to find coefficients consistent for many datasets.

The performance of the different models was evaluated based on AUC, specificity, sensitivity, and accuracy, which were computed by the R program (Dataset S5). The threshold for probability for droplet formation ($p_{LLPS} \geq 0.61$) was determined based on the logistic model, and was in good agreement

for the training and test sets. The π–π term was evaluated by the scripts given in ref. 19 using the same training and test sets (Dataset S5).

Predicting the droplet state in different proteomes. The UniProt Swiss-Prot (reviewed) sequences were downloaded for *C. elegans*, *D. melanogaster*, *H. sapiens*, *M. musculus*, *R. norvegicus*, *S. cerevisiae*, and *S. pombe* and TrEMBL for *X. laevis*. The degree of disorder was computed by the ESpritz NMR program (36), and the binding mode (p_{DD}) was predicted for each residue using the FuzPred program (23). The probability of droplet formation for each protein was determined based on Eq. **3**, with the coefficients given in Dataset S5. In each organism, we determined the frequency of proteins (including putative proteins), with $p_{LLPS} \geq 0.6$ (Dataset S8).

Observation of α-synuclein and β-synuclein liquid–liquid phase separation. Wild-type α-synuclein and β-synuclein were purified from *Escherichia coli* expressing plasmid pT7-7 encoding for the protein as previously described (38, 39). Following purification, the protein was concentrated using Amicon Ultra-15 centrifugal filter units (Merck Millipore) and buffer exchanged into phosphate-buffered saline (PBS) at pH 8.0. Protein was subsequently labeled with 10-fold molar excess of fluorescein 5-isothiocyanate (Sigma) for 3 h at room temperature, followed by an overnight incubation at 4 °C with constant mixing. The excess dye was removed on a Sephadex G-25 desalting column (Sigma) and used immediately for phase separation experiments.

To induce droplet formation, nonlabeled wild-type α-synuclein and β-synuclein were mixed with FITC-labeled proteins at a 10:1 molar ratio in PBS with 50 mM NaCl and 10% polyethylene glycol (PEG) (Thermo Fisher Scientific). The final mixture was pipetted onto a 35-mm glass-bottom dish (P35G-1.5-20-C; MatTek Life Sciences) and immediately imaged on a TCS SP5 confocal microscope using a 40×/1.3 HC PL Apo CS oil objective (Leica Microsystems) with the temperature controlled at either 20 or 30 °C. The excitation wavelength was 488 nm for all experiments. All images were processed and analyzed in ImageJ (NIH).

Complexin-1 phase separation. Recombinant human complexin-1 was obtained from Nkmax Bio (CPX0901). The C-terminal cysteine (C118) was labeled with a 1.5× molar excess of Alexa Fluor 488 C_5 maleimide (A10254; Life Technologies) overnight at 4 °C. The excess dye was removed on a Sephadex G-25 desalting column (G25150-100G; Sigma) and the protein was buffer exchanged into 50 mM Tris·HCl (pH 7.4).

For imaging, 10 μM nonlabeled complexin-1 was mixed with 10% (1 μM) Alexa Fluor 488-labeled protein in 50 mM Tris·HCl (pH 7.4), 100 mM NaCl, 1 mM dithiothreitol, and 5% PEG (B219555; Thermo Fisher Scientific) at 20 °C. The final mixture was pipetted onto a 35-mm glass-bottom dish (P35G-1.5-20-C; MatTek Life Sciences) and immediately imaged on a TCS SP5 using a 40×/1.3 HC PL Apo CS oil objective (Leica Microsystems). The excitation wavelength was 488 nm for all experiments. All images were analyzed with ImageJ (NIH).

Fluorescence recovery after photobleaching. FRAP was performed on the setup described above, under the same experimental conditions. Bleaching was done using a 488-nm laser at 50% intensity, to obtain ±50 to 60% photobleaching. Images were captured at 600-ms intervals, following a 1.8-s prebleach sequence and 1.2-s bleach. Intensity traces of the bleached area were background corrected and normalized. A nonlinear function of the recovery curve was fitted to obtain a relative recovery rate (Prism 8; GraphPad).

Data Availability. All study data are included in the article and supporting information.

1. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
2. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
3. L. P. Bergeron-Sandoval, N. Safaee, S. W. Michnick, Mechanisms and consequences of macromolecular phase separation. *Cell* **165**, 1067–1079 (2016).
4. S. Boeynaems et al., Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
5. C. Yuan et al., Nucleation and growth of amino acid and peptide supramolecular polymers through liquid-liquid phase separation. *Angew. Chem. Int. Ed. Engl.* **58**, 18116–18123 (2019).
6. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
7. P. R. ten Wolde, D. Frenkel, Homogeneous nucleation and the Ostwald step rule. *Phys. Chem. Chem. Phys.* **1**, 2191–2196 (1999).
8. S. Alberti, A. Gladfelter, T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
9. M. R. Sawaya et al., Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
10. T. P. Knowles et al., Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**, 1900–1903 (2007).
11. T. J. Nott et al., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
12. K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **60**, 231–241 (2015).
13. J. Wang et al., A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16 (2018).
14. M. P. Hughes et al., Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018).
15. G. L. Dignon, R. B. Best, J. Mittal, Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
16. G. Krainer et al., Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. bioRxiv:2020.05.04.076299 (7 May 2020).

17. R. M. Vernon, J. D. Forman-Kay, First-generation predictors of biological protein phase separation. *Curr. Opin. Struct. Biol.* **58**, 88–96 (2019).
18. B. Bolognesi *et al.*, A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* **16**, 222–231 (2016).
19. R. M. Vernon *et al.*, Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, e31486 (2018).
20. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
21. H. Wu, M. Fuxreiter, The structure and dynamics of higher-order assemblies: Amyloids, signalosomes, and granules. *Cell* **165**, 1055–1066 (2016).
22. S. Hahn, Phase separation, protein disorder, and enhancer function. *Cell* **175**, 1723–1725 (2018).
23. M. Miskei, A. Horvath, M. Vendruscolo, M. Fuxreiter, Sequence-based prediction of fuzzy protein interactions. *J. Mol. Biol.* **432**, 2289–2303 (2020).
24. P. Sormanni *et al.*, Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.* **13**, 339–342 (2017).
25. M. Fuxreiter, Fold or not to fold upon binding—Does it really matter? *Curr. Opin. Struct. Biol.* **54**, 19–25 (2019).
26. P. H. Kussie *et al.*, Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948–953 (1996).
27. A. Horvath, M. Miskei, V. Ambrus, M. Vendruscolo, M. Fuxreiter, Sequence-based prediction of protein binding mode landscapes. *PLoS Comput. Biol.* **16**, e1007864 (2020).
28. P. Leuenberger *et al.*, Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825 (2017).
29. K. You *et al.*, PhaSepDB: A database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* **48**, D354–D359 (2020).
30. B. Mészáros *et al.*, PhaSePro: The database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.* **48**, D360–D367 (2020).
31. Q. Li *et al.*, LLPSDB: A database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Res.* **48**, D320–D327 (2020).
32. S. Alberti, R. Halfmann, O. King, A. Kapila, S. Lindquist, A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
33. P. Tompa, Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533 (2002).
34. A. Hatos *et al.*, DisProt: Intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **48**, D269–D276 (2020).
35. S. Milles *et al.*, Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015).
36. I. Walsh, A. J. Martin, T. Di Domenico, S. C. Tosatto, ESpritz: Accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).
37. A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016).
38. S. Ray *et al.*, α-Synuclein aggregation nucleates through liquid-liquid phase separation. *Nat. Chem.* **12**, 705–716 (2020).
39. M. Hardenberg *et al.*, Observation of an α-synuclein liquid droplet state and its maturation into Lewy body-like assemblies. bioRxiv:2020.06.08.140798 (10 June 2020).
40. E. L. Guenther *et al.*, Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* **25**, 463–471 (2018).
41. G. Fusco *et al.*, Structural basis of synaptic vesicle assembly promoted by α-synuclein. *Nat. Commun.* **7**, 12563 (2016).
42. J. C. Lambert *et al.*; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
43. Y. Lai *et al.*, N-terminal domain of complexin independently activates calcium-triggered fusion. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4698–E4707 (2016).
44. M. Xue *et al.*, Binding of the complexin N terminus to the SNARE complex potentiates synaptic-vesicle fusogenicity. *Nat. Struct. Mol. Biol.* **17**, 568–575 (2010).
45. A. Hubstenberger *et al.*, P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157.e5 (2017).
46. J. S. Andersen *et al.*, Nucleolar proteome dynamics. *Nature* **433**, 77–83 (2005).
47. J. Ayache *et al.*, P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell* **26**, 2579–2595 (2015).
48. L. Jønson *et al.*, Molecular composition of IMP1 ribonucleoprotein granules. *Mol. Cell. Proteomics* **6**, 798–811 (2007).
49. K. W. Fong *et al.*, Whole-genome screening identifies proteins localized to distinct nuclear bodies. *J. Cell Biol.* **203**, 149–164 (2013).
50. D. Berchtold, N. Battich, L. Pelkmans, A systems-level study reveals regulators of membrane-less organelles in human cells. *Mol. Cell* **72**, 1035–1049.e5 (2018).
51. S. Markmiller *et al.*, Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell* **172**, 590–604.e13 (2018).
52. J. Y. Youn *et al.*, High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell* **69**, 517–532.e11 (2018).
53. M. Shi, P. Zhang, S. M. Vora, H. Wu, Higher-order assemblies in innate immune and inflammatory signaling: A general principle in cell biology. *Curr. Opin. Cell Biol.* **63**, 194–203 (2020).
54. M. Fuxreiter, Fuzziness in protein interactions—A historical perspective. *J. Mol. Biol.* **430**, 2278–2287 (2018).
55. G. T. Heller, P. Sormanni, M. Vendruscolo, Targeting disordered proteins with small molecules using entropy. *Trends Biochem. Sci.* **40**, 491–496 (2015).
56. G. Vecchi *et al.*, Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1015–1020 (2020).