Genomics Proteomics Bioinformatics 19 (2021) 13–24

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

REVIEW

Computational Screening of Phase-separating Proteins

Boyan Shen [1,#], Zhaoming Chen [1,#], Chunyu Yu [1,2], Taoyu Chen [1], Minglei Shi [3] Tingting Li [1,2,*]

[1] *Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China*
[2] *Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China*
[3] *MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic & Systems Biology, BNRist, School of Medicine, Tsinghua University, Beijing 100084, China*

Received 11 December 2019; revised 17 November 2020; accepted 10 December 2020
Available online 19 February 2021

Handled by Yu Xue

KEYWORDS

Phase separation;
Prediction;
Protein–protein interaction;
Protein post-translational modification;
Immunofluorescence image

Abstract **Phase separation** is an important mechanism that mediates the compartmentalization of proteins in cells. Proteins that can undergo phase separation in cells share certain typical sequence features, like intrinsically disordered regions (IDRs) and multiple modular domains. Sequence-based analysis tools are commonly used in the screening of these proteins. However, current phase separation predictors are mostly designed for IDR-containing proteins, thus inevitably overlook the phase-separating proteins with relatively low IDR content. Features other than amino acid sequence could provide crucial information for identifying possible phase-separating proteins: **protein–protein interaction (PPI)** networks show multivalent interactions that underlie phase separation process; **post-translational modifications (PTMs)** are crucial in the regulation of phase separation behavior; spherical structures revealed in **immunofluorescence (IF) images** indicate condensed droplets formed by phase-separating proteins, distinguishing these proteins from non-phase-separating proteins. Here, we summarize the sequence-based tools for predicting phase-separating proteins and highlight the importance of incorporating PPIs, PTMs, and IF images into phase separation **prediction** in future studies.

Introduction

Cellular organelles can be categorized into two classes, membrane-bound organelles and membraneless organelles. Membrane-bound organelles include classic organelles such as the Golgi apparatus, mitochondrion, and lysosome. These

* Corresponding author.
E-mail: litt@hsc.pku.edu.cn (Li T).
# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2020.11.003

cellular compartments enclosed by lipid bilayers have been well studied in the last century. However, in living cells, many biochemical reactions take place in membraneless organelles [1,2]. The formation mechanisms and functions of membraneless organelles had remained perplexing until ten years ago. In 2009, Brangwynne et al. reported liquid-like behaviors of P granules, which are protein-rich membraneless organelles in the cytoplasm of cells from *Caenorhabditis elegans* [3]. These granules can flow, deform, and undergo fission freely, just like liquid droplets. Proteins within P granules are also highly mobile and can exchange rapidly with the surrounding cytoplasm. These findings suggest that liquid–liquid phase separation (LLPS, also called liquid–liquid demixing) could be one of the mechanisms underlying membraneless organelle formation.

In a phase separation process, a set of macromolecules such as proteins and nucleic acids are separated from their surrounding environments and form an independent phase. The separated phase shares a similar molecular composition with the surrounding environment, yet at different concentrations [4]. In cells, proteins or nucleic acids form separated phases via intra- or intermolecular interactions [1,4,5], thereby allowing the formation of phase-separated compartments, which are also named membraneless organelles or biomolecular condensates. Besides P granules [3], the nucleoli [1,2], centrosomes [6], stress granules [7,8], and processing bodies (P-bodies) [8,9] are also membraneless organelles formed through phase separation. In addition, phase separation underlies many biological processes such as translation regulation [10,11], mRNA deadenylation [10,12], heterochromatin formation [13,14], and the control of signal transduction [15–17].

Phase separation is a complex biophysical process. Changes to any property of the system, *e.g.*, molecular composition, temperature, electrostatic property, and viscoelasticity of the solution, may affect the phase separation process [1,5,18]. Being able to undergo phase separation under specific conditions may be a universal property of proteins. However, only a few proteins with specific sequence-dependent features have the potential to undergo phase separation in living cells [19]. In this review, we name these proteins phase-separating proteins. Scientists have found that certain sequence features may correlate with phase separation behaviors, which brings out a range of useful bioinformatics tools to predict phase-separating proteins. Herein, we summarize the sequence-based predicting tools for phase-separating proteins and integrate the available phase-separating protein data to evaluate their performances. Furthermore, we propose protein–protein interaction (PPI) networks, post-translational modifications (PTMs), and immunofluorescence (IF) images as three promising features to be incorporated into phase separation prediction in future studies.

Driving force of phase separation

Phase separation is a conditional process. Proteins indispensable to the formation of one condensate can be alternative in another or do not participate in phase separation under some conditions [19]. Multivalent interactions between condensate components are the driving force of phase separation [17,20,21].

Proteins able to form multivalent interactions that promote phase separation can be classified into two types: one characterized by multiple modular domains and the other characterized by intrinsically disordered regions (IDRs). The first type of proteins often carry several folded interaction domains. The driving force of phase separation of these proteins is the multivalent interactions between their interaction domains [4]. An example is the interaction between small ubiquitin-like modifier (SUMO) and SUMO-interacting motif (SIM), which plays an essential role in the overall architecture of promyelocytic leukemia (PML) body. The PML protein contains a SUMO-interacting motif and multiple SUMOylation sites [22]. As the scaffold protein of the PML body, PML can not only self-assemble via interactions between its tripartite motifs (TRIMs) but also interact with itself and other proteins via SUMO–SIM interactions [23,24] (**Figure 1**A). Another example is the multivalent interactions of the nephrin–non-catalytic region of tyrosine kinase (NCK)–neuronal Wiskott–Aldrich syndrome protein (N-WASP) system in the actin regulatory signaling pathway. Nephrin is a transmembrane protein, the cytoplasmic tail of which harbors three tyrosine phosphorylation (pTyr) sites. Each of these three pTyr sites can bind to a Src homology 2 (SH2) domain on NCK. N-WASP contains six proline-rich motifs (PRMs), which can also bind to three of the SH3 domains on NCK. These interactions further stimulate actin assembly and phase separation [17] (Figure 1B).

The other type of phase-separating proteins are characterized by the presence of IDRs [25,26]. Instead of having a fixed three-dimensional structure, IDRs can interconvert between a range of slightly different low-energy states with discrepant conformations [27–29]. Proteins with high proportions of IDRs are hot targets in phase separation studies, *e.g.*, the neurodegenerative disorder-related protein fused in sarcoma (FUS) [30–33] (Figure 1C), the BUB3-interacting and GLEBS motif-containing protein BuGZ [34,35] (Figure 1D), and the microtubule-associated protein Tau [36–38]. However, not all IDRs facilitate phase separation. Phase-separating IDRs often share some sequence features. One type of IDRs that often undergo phase separation is the IDRs encompassing low-complexity regions (LCRs), which are regions with biased amino acid composition, often containing repeated segments. Moreover, phase-separating IDRs are often enriched with specific amino acid residues, *e.g.*, arginine and glycine for the RGG/RG domains [39,40] or polar amino acid residues like serine, tyrosine, glutamine, and asparagine for the prion-like domains [41]. These amino acids do not appear randomly, but are often found as functional sites like short/eukaryotic linear motifs and alternating charge blocks [25,42].

IDR content analysis in the prediction of phase-separating proteins

IDR-containing proteins account for a large proportion of phase-separating proteins, although our understanding of how IDRs are involved in the phase separation process and what features these IDRs share is still limited [43]. Therefore, IDR content analysis is often utilized in bioinformatics screening of potential phase-separating proteins (**Table 1**). DisProt is one of the earliest and most influential disorder databases

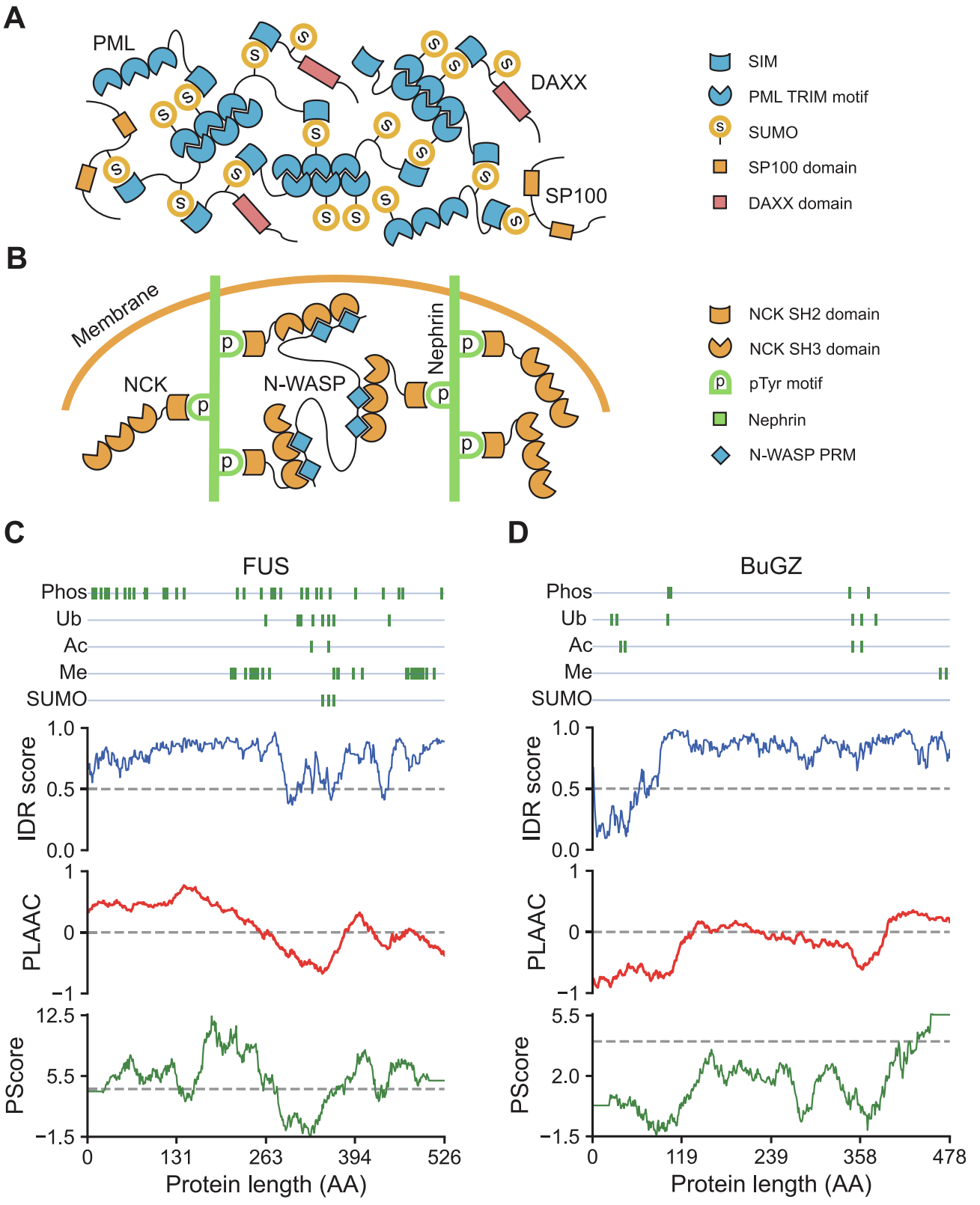

A
PML
DAXX
SP100
SIM
PML TRIM motif
SUMO
SP100 domain
DAXX domain
B
Membrane
NCK
N-WASP
Nephrin
NCK SH2 domain
NCK SH3 domain
pTyr motif
Nephrin
N-WASP PRM
C
FUS
D
BuGZ
Phos
Ub
Ac
Me
SUMO
IDR score
PLAAC
PScore
Protein length (AA)

Table 1 IDR content analysis tools

Tool	Type	Focus	Availability	Refs.
DisProt	Database	Intrinsically disordered proteins	www.disprot.org	[44,45]
ESpritz	Predictor	IDRs within proteins	protein.bio.unipd.it/espritz	[46]
IUPred2	Predictor	IDRs within proteins	iupred2a.elte.hu	[47,48]
D2P2	Database	Comparison of protein disorder prediction methods	d2p2.pro	[49]
MobiDB	Database	Annotations of intrinsically disordered proteins	mobidb.bio.unipd.it	[50,51]

Note: IDR, intrinsically disordered region.

[44,45]. Two long-tested IDR predictors, ESpritz [46] and IUPred [47,48], are frequently used in the IDR content analysis of potential phase-separating proteins. In recent years, meta-IDR predictors like D2P2 [49] and MobiDB [50,51], which integrate predictions from several predictors currently available and provide comprehensive results, are widely used as well. For example, in a previous study on how intrinsically disordered linkers influence the interplay between phase separation and gelation, researchers used the meta-IDR predictor D2P2 to identify proteins containing disordered regions from the human proteome [52].

Phase separation predictors

Although multivalent, transient interactions that drive phase separation are often facilitated by IDRs [21], only proteins with certain types of IDRs facilitate phase separation. Non-IDR interacting elements like the coiled-coil domain can also promote these interactions [6]. Bioinformaticians thus aim to develop prediction tools that are based explicitly on phase separation-specific sequence features.

By studying the sequences of several well-known phase-separating proteins, *e.g.*, FUS, DEAD box protein 4 (DDX4), TATA-binding protein-associated factor 15 (TAF15), Ewing sarcoma protein (EWS), TAR DNA-binding protein 43 (TDP43), and heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1), two common types of LCRs, prion-like domains (PLDs) and RGG domains, were found to promote weak interactions and thereby promote phase separation [33,39,53,54]. Furthermore, low-complexity aromatic-rich kinked segments (LARKSs) and steric zipper motifs have been found to promote the transition between different physical properties of biomolecular condensates [55,56]. Such features have inspired bioinformaticians to design algorithms to predict the phase separation propensities of proteins [57]. These include PScore [58], prion-like amino acid composition (PLAAC) [59], PSPer [60], catGRANULE [61], R + Y [41], LARKS [55], and ZipperDB [56] (**Table 2**).

PScore

Pi–pi interactions can occur between protein sequences enriched in pi-orbital-containing residues. In contrast to the conventional view, pi–pi interaction involves not only amino acids with an aromatic ring (tyrosine, phenylalanine, tryptophan, and histidine) [62], but also non-aromatic amino acids with pi bonds on their side chains (glutamine, asparagine, glutamic acid, aspartic acid, and arginine) and small amino acids with exposed backbone peptide bonds (glycine, serine, threonine, and proline). Face-to-face interactions formed by these pi-containing groups are called planar pi–pi contacts. In 2018, the Forman-Kay group reported that planar pi–pi contact represents a predominant interaction type and is highly relevant to self-association and phase separation of proteins [58]. Then a planar pi–pi contact predictor named PScore was developed for screening potential phase-separating proteins.

PLAAC

PLAAC is an application initially designed to screen PLDs [59], which utilizes a hidden Markov model (HMM) for the

Figure 1 Schematic view of multivalent interactions that promote phase separation

A. PML protein can not only self-assemble via interactions between its TRIMs but also interact via its SIM with SUMOs of itself and other proteins, such as DAXX and SP100. **B.** Schematic view of nephrin–NCK–N-WASP system. The cytoplasmic tail of nephrin contains three pTyr sites, each of which can bind to an SH2 domain on NCK. The three SH3 domains on NCK can also bind to PRMs within N-WASP. **C.** Residue-wise plot of scaffold protein FUS. Information of PTM sites was collected from the PhosphoSitePlus database; IDR scores were predicted by IUPred for long disorder; the prion-like domain was identified with PLAAC score greater than zero; the pi-contact was identified with PScore greater than four, as indicated by dashed line. **D.** Residue-wise plot of scaffold protein BuGZ. PML, promyelocytic leukemia; TRIM, tripartite motif; SIM, SUMO-interacting motif; SUMO, small ubiquitin-like modifier; DAXX, death domain-associated protein; SP100, nuclear autoantigen Sp-100; NCK, non-catalytic region of tyrosine kinase; N-WASP, neuronal Wiskott–Aldrich syndrome protein; pTyr, tyrosine phosphorylation; SH, Src homology; PRM, proline-rich motif; FUS, fused in sarcoma; PTM, post-translational modification; IDR, intrinsically disordered region; BuGZ, BUB3-interacting and GLEBS motif-containing protein ZNF207; Phos, phosphorylation; Ub, ubiquitination; Ac, acetylation; Me, methylation; AA, amino acid.

Table 2 Phase separation analysis tools

Tool	Type	Focus	Information used for prediction	Availability	Refs.
PScore	Predictor	Phase-separating proteins	Pi–pi contact frequencies	elifesciences.org/articles/31486	[58]
PLAAC	Predictor	PLDs of proteins	Amino acid frequencies in PLDs of *S. cerevisiae*	plaac.wi.mit.edu	[59]
PSPer	Predictor	Prion-like RNA-binding phase-separating proteins	Rules of domain arrangements in FUS-like proteins	bio2byte.com/psp	[60]
catGRANULE	Predictor	Phase-separating proteins	Nucleic acid binding propensities, structural disorder, sequence length, R/G/F content	www.tartaglialab.com	[61]
R+Y	Predictor	Saturation concentration of phase separation for FUS family-like proteins	Protein sequence	List of proteins with predicted saturation concentration	[41]
LARKS	Predictor	LARKSs within proteins	Protein sequence and 3D structure information	Information of top 400 human proteins rich in LARKSs	[55]
ZipperDB	Database	Fibril-forming segments within proteins	Protein sequence and 3D structure information	services.mbi.ucla.edu/zipperdb	[56]

Note: PLD, prion-like domain; FUS, fused in sarcoma; R, arginine; G, glycine; F, phenylalanine; Y, tyrosine; LARKS, low-complexity aromatic-rich kinked segment.

discrimination of PLDs and non-PLDs based on amino acid composition. PLAAC was originally trained on yeast proteome but later extended to screen human proteins [63,64]. It supports both single-protein and proteome scanning.

PSPer

PSPer is a rule-based model developed for screening prion-like RNA-binding phase-separating proteins. Expected properties of the FUS-like phase-separating regions are used to build an HMM-like model [41], which identifies PLDs, RNA-recognition motifs, and disordered, arginine-rich regions within a protein [60].

catGRANULE

catGRANULE is a phase separation prediction algorithm with good performance for predicting dosage-sensitive proteins [61]. This tool is based on the discovery that cellular toxicity mechanisms of some dosage-sensitive proteins in yeast are well-explained by LLPS theory, as these proteins take part in the formation of cytoplasmic foci in a concentration-related manner. Further studies reveal that these proteins have an increased nucleic acid binding propensity. catGRANULE was therefore developed to screen these proteins by combining nucleic acid binding propensities, structural disorder, sequence length, and content of arginine, glycine, and phenylalanine. Although initially trained against the yeast proteome, catGRANULE has been successfully applied to mammalian and even human proteomes [37].

R + Y

R + Y is a predictor built upon the analysis of the molecular grammar of FET family proteins, including FUS, EWS, and TAF15 [65]. In 2018, the Hyman, Alberti, and Pappu groups reported that phase separation behaviors of FET family proteins are determined by interactions between the tyrosine-rich PLDs and the arginine-rich RNA-binding domains [41]. Further analyses indicate that the numbers of tyrosine and arginine residues are inversely correlated with the measured saturation concentrations for phase separation. Extrapolating these findings to the prediction of non-FET proteins, researchers developed the R + Y predictor and utilized it for a human proteome-wide analysis, which has remarkable prediction performance on DNA- and RNA-binding proteins (RBPs).

ZipperDB

ZipperDB is a database that includes predicted fibril-forming segments from more than 20,000 putative amyloid-forming protein sequences [56]. Fibrils are highly ordered aggregates characterized by a "steric zipper" structure, whose formation is an essential step in amyloidosis. Amyloid deposition is occasionally observed in cells. Recent evidence indicates that the formation of such deposits could be attributed to a liquid-to-solid phase transition or an atypical phase separation process that forms solid-state compartments. ZipperDB utilizes an algorithm named 3D profiling to analyze the probability of forming a steric zipper structure for every hexapeptide segment.

LARKS

LARKSs share similar structural characteristics with steric zippers yet have lower binding energy, as aromatic residues predominate the kinks and affect LARKS stability through weak interactions. LARKSs are characterized by the stacking kinked β sheet pairs, which promote the formation of amyloid fibrils and hydrogels during phase transition. Similar to ZipperDB, 3D profiling was utilized to identify potential LARKSs from inquiry protein sequences. After querying the human proteome of 20,120 proteins, a list of 400 proteins with the most enriched LARKS was provided [55].

In summary, all current phase separation prediction tools are developed based on sequence-dependent features. PScore calculates the pi-contact propensities of each residue in a given protein sequence. PLAAC and PSPer identify specific domains

like PLDs based on HMM. catGRANULE calculates the granule propensity of each residue in a given sequence. R + Y calculates the number of tyrosine and arginine residues within disordered regions of a given sequence. LARKS and ZipperDB adopt an algorithm named 3D profiling to measure the probability of a given sequence to fold into a LARKS or a steric zipper.

Performance evaluation of phase separation predictors

Vernon et al. have recently reviewed several phase separation predictors and compared their prediction performance comprehensively [57]. They find that since each algorithm predicts different kinds of interactions and sequence features, very different protein categories are covered by these predictors. The only exceptions are RBPs, as high prediction confidence is obtained from all predictors for RBPs. However, due to the insufficient phase-separating protein data available, evaluations could only be made on a set of 30 human proteins [57].

At least four LLPS protein databases were released until 2020, including LLPSDB [66], PhaSePro [67], PhaSepDB [68], and DrLLPS [69] (**Table 3**). LLPSDB collects *in vitro* data on LLPS-related proteins. The current version of LLPSDB includes 295 independent proteins, which are integrated into 1192 entries with corresponding experimental phase separation conditions. PhaSePro provides the experimental data on 121 proteins driving phase separation in living cells, which is less focused on *in vitro* phase separation conditions and contains a broader array of information than LLPSDB. PhaSepDB contains less detailed information than either LLPSDB or PhaSePro, but provides a larger set of data with 2914 non-redundant proteins localized in more than 30 different organelles, which comprise PubMed-reviewed data, UniProt-reviewed data, and high-throughput data. DrLLPS contains 9285 curated proteins that are known to be associated with LLPS, including 150 scaffold proteins, 987 regulator proteins, and 8148 potential client proteins. All four databases extensively reference the original literature, allowing the user to verify information or conduct further scrutinization.

With the availability of high-quality data provided by the four databases mentioned above, a more comprehensive comparison of phase separation predictors is possible. We constructed a non-redundant positive set (set P) of 278 human proteins involving 90, 59, 233, and 86 proteins from LLPSDB, PhaSePro, the PubMed-reviewed part of PhaSepDB, and the scaffold part of DrLLPS, respectively (Table S1). The remaining majority of human proteome was collected from UniProt and defined as the negative set (set N), which includes 20,227 proteins (Table S1).

To compare the prediction performance of available phase separation predictors, we scored the proteins in both set P and set N using PScore, PLAAC, catGRANULE, and PSPer, which provide online/offline batch prediction (Table S2). For R + Y and LARKS without prediction tools, R + Y scores of 2657 human proteins and 400 human proteins enriched with LARKSs were downloaded from the resepctive websites (Table S2). ZipperDB provides neither batch prediction tools nor bulk download. As a result, we collected prediction scores from six phase separation predictors (Table S2).

Firstly, we adopted a comparison between those six tools on human proteome by plotting the receiver operating characteristic curve (ROC). As shown in **Figure 2**A, catGRANULE, PScore, PLAAC, and PSPer have nearly the same values for area under the curve (AUC), which are > 0.7, while the performance of R + Y and LARKS is not as good as that of the other four tools.

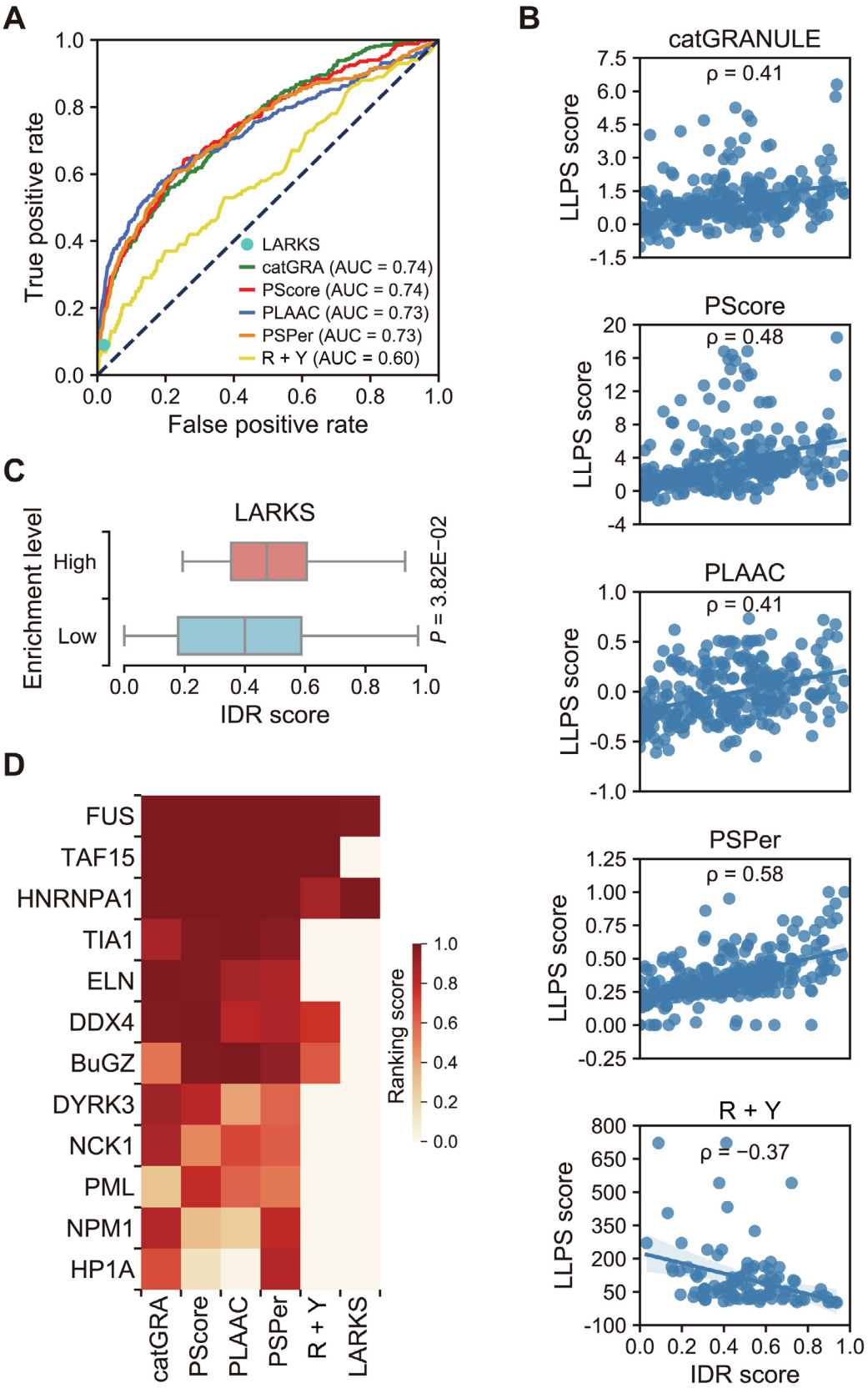
To test the influence of IDR contents on the scores of these phase separation predictors, we scored the proteins in set P by IUPred with default setting (Table S3). As shown in the scatter plots, phase separation scores of the first five predictors are significantly correlated with IDR scores (Figure 2B). In R + Y, a lower score means a higher phase separation potential, and therefore its prediction scores are significantly negatively-correlated with IDR scores. For LARKS that did not provide prediction scores, proteins in set P were divided into LARKS High and LARKS Low groups, according to whether the protein is in the list of LARKS enriched proteins. The IDR scores of proteins in the LARKS High group were significantly higher than those in the LARKS Low group (Figure 2C). The aforementioned comparison processes were also performed for all human proteins, scores of all phase separation predictors are significantly correlated with IDR scores (Figure S1). These results demonstrate that all phase separation predictors prefer proteins with high IDR contents.

Discriminating the scaffold proteins from the client proteins remains challenging. Therefore, it is possible that some low IDR proteins in set P are client proteins. However, taking the 12 phase-separating proteins for example, heterochromatin protein 1 homolog alpha (HP1A), nucleophosmin 1 (NPM1), PML, and NCK1, which are all scaffold proteins, were ranked

Table 3 Information of four LLPS databases

Database	No. of entries of curated proteins	Website	Ref.
LLPSDB	295	bio-comp.org.cn/llpsdb	[66]
PhaSePro	121	phasepro.elte.hu	[67]
PhaSepDB	2914	db.phasep.pro	[68]
DrLLPS	9285	llps.biocuckoo.cn	[69]

Note: LLPS, liquid–liquid phase separation.


A
True positive rate
False positive rate
LARKS
catGRA (AUC = 0.74)
PScore (AUC = 0.74)
PLAAC (AUC = 0.73)
PSPer (AUC = 0.73)
R + Y (AUC = 0.60)
B
catGRANULE
ρ = 0.41
PScore
ρ = 0.48
PLAAC
ρ = 0.41
PSPer
ρ = 0.58
R + Y
ρ = −0.37
LLPS score
IDR score
C
LARKS
Enrichment level
High
Low
P = 3.82E−02
IDR score
D
FUS
TAF15
HNRNPA1
TIA1
ELN
DDX4
BuGZ
DYRK3
NCK1
PML
NPM1
HP1A
catGRA
PScore
PLAAC
PSPer
R + Y
LARKS
Ranking score

outside of the top 10% of the human proteome by all the six predictors, while PML was ranked outside of the top 20% (Figure 2D).

In summary, a range of useful bioinformatics tools have been developed to predict phase-separating proteins, however, most of which are designed for screening IDR-containing proteins. Proteins with modular domains account for a considerable part of phase-separating proteins as well, yet a corresponding computational tool to identify such proteins is not available. It may not be so challenging to identify proteins with multiple modular interaction domains. However, some proteins with few modular domains can also participate in phase separation via assembly, like HP1A [13]. More comprehensive features besides sequence composition might be required for phase separation predictors.

Other features potentially used to predict phase-separating proteins

Features other than sequence composition could provide crucial information for identifying possible phase-separating proteins. This section will discuss three new features that could be utilized in phase separation prediction: PPI networks, PTMs, and IF images.

PPI networks

As described in the preceding section, the driving force of phase separation is the multivalent interactions between molecules. In some cases, especially for those with low IDR contents, proteins cannot undergo phase separation alone. Take the nephrin–NCK–N-WASP system in Figure 1B for example. The pTyr–SH2 and SH3–PRM interactions among cooperated proteins are the driving force of the phase separation process, and NCK cannot phase separate under the conditions with nephrin or N-WASP abscent [17]. Another example is the clustering of T cell receptor (TCR) signaling pathway molecules. Upon TCR activation, the tyrosine kinase ZAP70 phosphorylates the transmembrane protein LAT, and phospho-tyrosines on LAT can bind to the SH2 domain on Grb2. Two SH3 domains within Grb2 further interact with Sos1, and the phosphorylated LAT, Grb2, and Sos1 together form the LAT complex, which can coalesce into T cell microclusters that show phase separation behaviors [15]. Similar to the aforementioned example, components in this TCR pathway cannot phase separate without interactions among cooperated proteins.

One issue of the available phase separation predictors is that they are based on sequence-dependent features of individual proteins. These features may not be sufficient, as phase separation driven by complex modular interaction domains or motifs and other multivalent interactions might be missed. A more comprehensive approach that also considers PPI network information could be helpful. It is hard to integrate PPI networks into phase separation predictors directly. One possible approach is network embedding method such as node2vec [70]. Taking the adjacency matrix of the PPI network as input, node2vec encodes each node in the network as a vector, which can be used for various downstream machine learning tasks. The distance between vectors reflects the similarity of interaction networks of corresponding proteins. However, it should be noted that PPI networks based on experimental evidence usually bias to well-studied proteins. BioPlex database that provides an unbiased mapping of the human PPIs by affinity-purification mass spectrometry might be more appropriate to be incorporated into phase separation prediction [71,72].

PTMs

The interactions required for phase separation can be weak interactions or strong interactions reversible on a short timescale [21]. The reversibility can often be regulated by PTMs, like the interaction between the SH2 domain and the pTyr residue in nephrin–NCK–N-WASP and TCR pathways [15,17]. In these examples, PTMs can regulate the reversibility of a binding event by generating/degenerating a modular interaction domain recognition site [21].

PTMs can also regulate phase separation processes by changing protein physical properties, like the charge state, bulkiness, solubility, hydrophobicity, or binding affinity [10,73–75]. Citrullination of RG/RGG motifs has been reported to increase the solubility of proteins like FUS, EWS, and TAF15, inhibiting the arginine methylation, aggregation, and stress granule formation of these proteins [76]. Another example is the PTMs of fragile X mental retardation protein (FMRP) and cytoplasmic activation- and proliferation-associated protein 1 (CAPRIN1). The two proteins do not co-phase separate without phosphorylation. However, co-phase separation occurs when Tyr of CAPRIN1 aromatic-rich regions is phosphorylated, or the C-terminal LCR of FMRP is phosphorylated [10,74].

Given the widespread regulatory roles of PTMs in phase separation, we tested whether PTM levels could be used as features to discriminate phase-separating proteins. We collected PTM datasets from the PhosphoSitePlus database [77], including phosphorylation, ubiquitination, acetylation, and methylation (Table S4). As shown in **Figure 3**A, the Venn diagram

Figure 2 Comparison of phase separation predictors on human proteome

A. ROC curve for each predictor. Since PScore, PLAAC, and PSPer have restrictions on the length of protein sequence, only catGRANULE returned scores for all human proteins. Except for catGRANULE, the AUC scores of the remaining tools were calculated on the subsets of P and N sets. Prediction of LARKS was shown as a point since it did not provide scores for each protein. **B.** Scatter plots of predicted values for proteins in set P, with one axis being the IDR score and the other axis being the phase separation score. Spearman correlation coefficient with $P < 0.05$ indicates significant correlation. **C.** Proteins in set P were divided into LARKS High and LARKS Low groups, according to whether the protein is in the list of LARKS-enriched proteins. P value is calculated through the two-sided Mann–Whitney U test. **D.** Ranking scores of 12 specific LLPS proteins by six phase separation predictors. ROC, receiver operating characteristic; AUC, area under the curve; LARKS, low-complexity aromatic-rich kinked segment; LLPS, liquid–liquid phase separation.

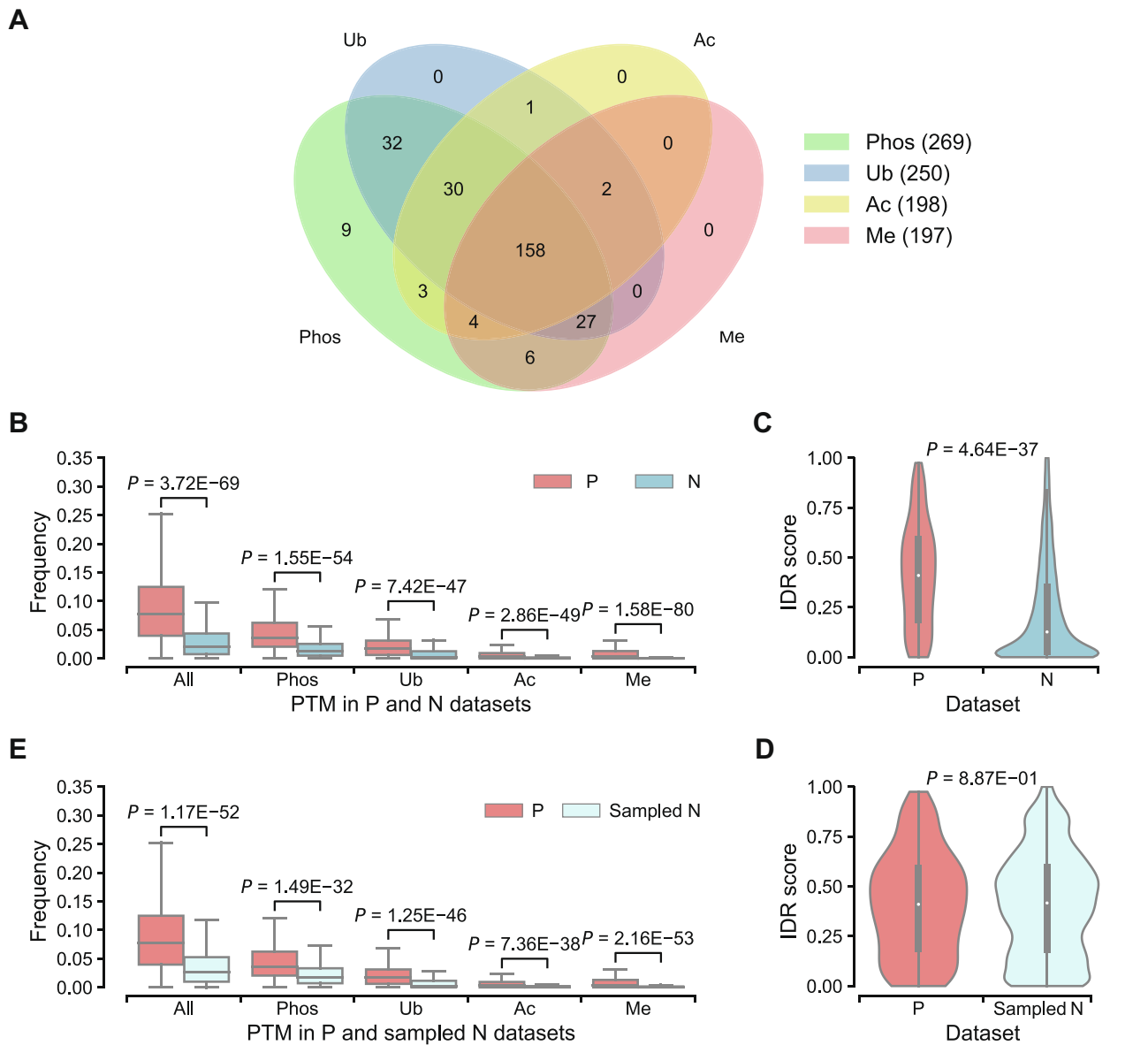


Figure 3 Comparison of PTM frequencies between different groups on human proteome

A. Venn diagram displaying the overlap of phase-separating proteins with different PTM types. The PTM datasets were collected from the PhosphoSitePlus database. **B.** For all PTM types, frequencies for proteins in set P are significantly higher than those for set N. *P* values are calculated through the two-sided Mann–Whitney U test. **C.** Distribution of IDR contents of proteins in set P and set N. Most proteins in set N have low proportions of IDRs. **D.** Set N was resampled into a subset according to IDR content distribution of set P. *P* value indicates that IDR contents of proteins in sampled set N have a similar distribution with that in set P. **E.** PTM frequencies of proteins in set P are still higher than those in resampled set N.

displays the overlap of phase-separating proteins with different PTM types. For 278 phase-separating proteins in set P, 221 of them contain more than three types of PTMs, and 158 of them possess all four types of PTMs considered. The overlap demonstrates that most phase-separating proteins are modified by multiple types of PTMs. Then we defined the PTM frequency of a protein as the number of modification sites on a sequence divided by the length of the sequence, and calculated PTM frequencies of proteins in set P and set N. For each protein, frequency of all PTM types and frequencies of each specific PTM types with more than 10,000 recorded sites in the database were calculated (Table S5). For all PTM types, frequencies for proteins in set P were significantly higher than those in set N (Figure 3B). However, it is well established that PTM sites are enriched in IDRs, and proteins with high proportions of IDRs tend to have higher PTM frequencies (Figure 3C), which might cause bias. To control the impact of different IDR contents, the set N was resampled into a subset, whose distribution of IDR contents was similar to that of the set P (Figure 3D; see details of the resampling process in Table S6). The PTM frequencies of the set P were still significantly higher than those of the resampled set N (Figure 3E),

which indicates that PTM frequency can be regarded as a feature to discriminate phase-separating proteins. Previous studies reported that an increased number of PTMs was correlated with an increased LLPS propensity in predicted phase-separating proteins [10,58], and our results show the similar observations using the actual LLPS proteins from more datasets.

It should be noted that although a considerable number of PTM sites have been collected into the database, most of them are phosphorylation sites. Poor data quality and low coverage of the PTM dataset might affect the performance of incorporating PTM frequency as a feature. A high-confidence dataset, including 119,809 phosphorylation sites, has been reported recently [78], which could provide a high-quality PTM feature for identifying possible phase-separating proteins.

IF images

Identification of spherical droplet structures through IF images represents the most common approach for validation of phase-separating proteins. Phase-separating proteins usually appear as spherical droplet-like structures in IF images, which allows them to be distinguished from non-phase-separating proteins. Thus phase-separating proteins can be identified if we screen out the proteins which appear as spherical structures in IF images. The Cell Atlas (part of the Human Protein Atlas) database provides antibody-based profiling by IF confocal microscopy for 12,073 proteins [79], allowing the screening of phase-separating proteins based on IF images.

It should be noted that the formation of phase separation is condition-dependent. Phase separation proteins that do not appear as droplets in the inquiry IF images cannot be identified. In addition, some phase-separated condensates are too small to be detected in IF images, such as the ''transcription hubs'' in transcriptional activation [80], and some membrane-bound organelles like vesicles have similar spherical appearance in IF images. Therefore, IF images alone are insufficient to determine whether a specific protein can undergo phase separation or not. However, IF images with droplet-like structures could provide evidence that the labeled proteins aggregate in cells, which allows us to screen out the potential phase-separating proteins. In a previous study, we built a convolutional neural network (CNN) classifier to identify IF images with droplet-like structures and found that the aggregation evidence extracted from IF images is useful in screening phase-separating proteins with low IDR contents [81]. Besides deep learning methods, CellProfiler [82] that can segment droplets and cells in the IF images should also be useful in extracting aggregation evidence from IF images. The outputs of both the CNN classifier and the CellProfiler can measure the aggregation state of the labeled proteins in IF images, and these outputs can be used as features for various downstream machine learning tasks.

Conclusion and perspectives

Experimental studies have enabled significant progress in improving our understanding of phase separation. Researchers have also noticed that some sequence features are closely related to the phase separation behavior, and several sequence-based computational tools have been developed accordingly. These computational methods facilitate studies on the phase separation phenomenon by providing predictions and proteome-scale screening of phase separation candidates. Furthermore, to examine the roles of different domains of specific proteins in phase separation, truncation mutants are usually constructed to detect segments crucial in forming liquid-like droplets. Computational methods that provide residue-specific predictions also assist phase separation studies in screening critical fragments. However, a meta-predictor that integrates predictions from existing phase separation predictors is in need to reduce the complexity of sequence analysis, like D2P2 and MobiDB in IDR analysis.

Furthermore, current phase separation predictors are mostly designed for IDR-containing proteins and are based on sequence-dependent features of individual proteins. More comprehensive features should be incorporated into phase separation predictors. As multivalent interactions are critical to phase separation, PPI networks can be useful in phase separation computational analysis. PTMs and IF images are also available features that can be utilized in the prediction of phase-separating proteins. PTMs are predominant regulators of phase separation behavior. Accordingly, in this review, we also find that phase-separating proteins tend to have high PTM frequencies. Components of phase-separating condensates usually appear as spherical droplets in IF images, allowing the screening of phase-separating proteins. We expect that incorporating PPI networks, PTMs, and IF images into prediction algorithms will lead to more effective and unbiased phase separation analytic tools.

CRediT author statement

Boyan Shen: Methodology, Investigation, Writing - original draft, Writing - review & editing. **Zhaoming Chen:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Chunyu Yu:** Methodology. **Taoyu Chen:** Writing - review & editing. **Minglei Shi:** Writing - review & editing. **Tingting Li:** Conceptualization, Methodology, Supervision, Writing - review & editing. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2018YFA0507504) and the National Natural Science Foundation of China (Grant Nos. 61773025 and 32070666).

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2020.11.003.

ORCID

0000-0002-1574-1958 (Boyan Shen)
0000-0002-6810-5445 (Zhaoming Chen)
0000-0002-3519-1332 (Chunyu Yu)
0000-0002-8966-1264 (Taoyu Chen)
0000-0002-1108-4233 (Minglei Shi)
0000-0003-4266-0317 (Tingting Li)

References

[1] Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol 2017;18:285–98.
[2] Hyman AA, Weber CA, Julicher F. Liquid–liquid phase separation in biology. Annu Rev Cell Dev Biol 2014;30:39–58.
[3] Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Jöbin G, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 2009;324:1729–32.
[4] Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein phase separation: a new phase in cell biology. Trends Cell Biol 2018;28:420–35.
[5] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science 2017;357:eaaf4382.
[6] Woodruff JB, Gomes BF, Widlund PO, Mahamid J, Honigmann A, Hyman AA. The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. Cell 2017;169:1066–77.
[7] Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. ATPase-modulated stress granules contain a diverse proteome and substructure. Cell 2016;164:487–98.
[8] Aulas A, Fay MM, Lyons SM, Achorn CA, Kedersha N, Anderson P, et al. Stress-specific differences in assembly and composition of stress granules and related foci. J Cell Sci 2017;130:927–37.
[9] Hubstenberger A, Courel M, Bénard M, Souquere S, Ernoult-Lange M, Chouaib R, et al. P-body purification reveals the condensation of repressed mRNA regulons. Mol Cell 2017;68:144–57.
[10] Kim TH, Tsang B, Vernon RM, Sonenberg N, Kay LE, Forman-Kay JD. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science 2019;365:825–9.
[11] Aguilera-Gomez A, Rabouille C. Membrane-bound organelles versus membrane-less compartments and their control of anabolic pathways in drosophila. Dev Biol 2017;428:310–7.
[12] Aizer A, Kalo A, Kafri P, Shraga A, Ben-Yishay R, Jacob A, et al. Quantifying mRNA targeting to P-bodies in living human cells reveals their dual role in mRNA decay and storage. J Cell Sci 2014;127:4443–56.
[13] Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature 2017;547:236–40.
[14] Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. Nature 2017;547:241–5.
[15] Su X, Ditlev JA, Hui E, Xing W, Banjade S, Okrut J, et al. Phase separation of signaling molecules promotes T cell receptor signal transduction. Science 2016;352:595–9.
[16] Wippich F, Bodenmiller B, Trajkovska MG, Wanka S, Aebersold R, Pelkmans L. Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. Cell 2013;152:791–805.
[17] Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, et al. Phase transitions in the assembly of multivalent signalling proteins. Nature 2012;483:336–40.
[18] Alberti S, Saha S, Woodruff JB, Franzmann TM, Wang J, Hyman AA. A user's guide for phase separation assays with purified proteins. J Mol Biol 2018;430:4806–20.
[19] Alberti S, Gladfelter A, Mittag T. Considerations and challenges in studying liquid–liquid phase separation and biomolecular condensates. Cell 2019;176:419–34.
[20] Forman-Kay JD, Mittag T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure 2013;21:1492–9.
[21] Simon Alberti DD. Liquid–liquid phase separation in disease. Annu Rev Genet 2019;53:171–94.
[22] Nisole S, Maroui MA, Mascle XH, Aubry M, Chelbi-Alix MK. Differential roles of PML isoforms. Front Oncol 2013;3:125.
[23] Banani SF, Rice AM, Peeples WB, Lin Y, Jain S, Parker R, et al. Compositional control of phase-separated cellular bodies. Cell 2016;166:651–63.
[24] Zhong S, Müller S, Ronchetti S, Freemont PS, Dejean A, Pandolfi pp.. Role of SUMO-1–modified PML in nuclear body formation. Blood 2000;95:2748–53.
[25] Brangwynne CP, Tompa P, Pappu Rohit V. Polymer physics of intracellular phase transitions. Nat Phys 2015;11:899–904.
[26] Protter DSW, Rao BS, Van Treeck B, Lin Y, Mizoue L, Rosen MK, et al. Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. Cell Rep 2018;22:1401–12.
[27] Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 2007;17:3–14.
[28] Jensen MR, Ruigrok RWH, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 2013;23:426–35.
[29] Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol 2015;32:102–12.
[30] Lin Y, Currie SL, Rosen MK. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J Biol Chem 2017;292:19110–20.
[31] Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, et al. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. EMBO J 2017;36:2951–67.
[32] Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. Cell 2017;171:615–27.
[33] Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell 2015;162:1066–77.
[34] Woodruff JB. Phase separation of BuGZ promotes aurora a activation and spindle assembly. J Cell Biol 2018;217:9–10.
[35] Huang Y, Li T, Ems-McClung SC, Walczak CE, Prigent C, Zhu X, et al. Aurora a activation in mitosis promoted by BuGZ. J Cell Biol 2018;217:107–16.
[36] Zhang X, Lin Y, Eschmann NA, Zhou H, Rauch JN, Hernandez I, et al. RNA stores tau reversibly in complex coacervates. PLoS Biol 2017;15:e2002183.
[37] Ambadipudi S, Biernat J, Riedel D, Mandelkow E, Zweckstetter M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. Nat Commun 2017;8:275.
[38] Wegmann S, Eftekharzadeh B, Tepper K, Zoltowska KM, Bennett RE, Dujardin S, et al. Tau protein liquid–liquid phase separation can initiate tau aggregation. EMBO J 2018;37:e98049.
[39] Nott Timothy J, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase transition of a disordered nuage

protein generates environmentally responsive membraneless organelles. Mol Cell 2015;57:936–47.
[40] Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. Nucleic Acids Res 2017;45:7984–96.
[41] Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell 2018;174:688–99.
[42] Tompa P, Davey NE, Gibson TJ, Babu MM. A million peptide motifs for the molecular biologist. Mol Cell 2014;55:161–9.
[43] Franzmann TM, Alberti S. Prion-like low-complexity sequences: key regulators of protein solubility and phase behavior. J Biol Chem 2019;294:7128–36.
[44] Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, et al. Disprot: a database of protein disorder. Bioinformatics 2004;21:137–40.
[45] Hatos A, Hajdu-Soltesz B, Monzon AM, Palopoli N, Alvarez L, Aykac-Fas B, et al. Disprot: intrinsic protein disorder annotation in 2020. Nucleic Acids Res 2019;48:D269–76.
[46] Walsh I, Martin AJ, Di Domenico T, Tosatto SC. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 2012;28:503–9.
[47] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21:3433–4.
[48] Dosztányi Z. Prediction of protein disorder based on IUPred. Protein Sci 2018;27:331–40.
[49] Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, et al. D2P2: database of disordered protein predictions. Nucleic Acids Res 2012;41:D508–16.
[50] Di Domenico T, Walsh I, Martin AJM, Tosatto SCE. MobiDB: a comprehensive database of intrinsic protein disorder annotations. Bioinformatics 2012;28:2080–1.
[51] Piovesan D, Tabaro F, Paladin L, Necci M, Mičetić I, Camilloni C, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res 2018;46:D471–6.
[52] Harmon TS, Holehouse AS, Rosen MK, Pappu RV. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. Elife 2017;6:e30294.
[53] March ZM, King OD, Shorter J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. Brain Res 2016;1647:9–18.
[54] Shorter J. Prion-like domains program Ewing's sarcoma. Cell 2017;171:30–1.
[55] Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, Cascio D, et al. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. Science 2018;359:698–701.
[56] Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A 2010;107:3487–92.
[57] Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. Curr Opin Struct Biol 2019;58:88–96.
[58] Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, et al. Pi-pi contacts are an overlooked protein feature relevant to phase separation. Elife 2018;7:e31486.
[59] Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics 2014;30:2501–2.
[60] Orlando G, Raimondi D, Tabaro F, Codicè F, Moreau Y, Vranken WF, et al. Computational identification of prion-like RNA-binding proteins that form liquid phase-separated condensates. Bioinformatics 2019;35:4617–23.
[61] Bolognesi B, Lorenzo Gotor N, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, et al. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. Cell Rep 2016;16:222–31.
[62] Sherrill CD. Energy component analysis of π interactions. Acc Chem Res 2013;46:1020–8.
[63] King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res 2012;1462:61–80.
[64] Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. Biochem J 2017;474:1417–38.
[65] Schwartz JC, Cech TR, Parker RR. Biochemical properties and biological functions of FET proteins. Annu Rev Biochem 2015;84:355–79.
[66] Li Q, Peng X, Li Y, Tang W, Zhu Ja, Huang J, et al. LLPSDB: a database of proteins undergoing liquid–liquid phase separation *in vitro*. Nucleic Acids Res 2019;48:D320–7.
[67] Meszaros B, Erdos G, Szabo B, Schad E, Tantos A, Abukhairan R, et al. PhaSePro: the database of proteins driving liquid–liquid phase separation. Nucleic Acids Res 2019;48:D360–7.
[68] You K, Huang Q, Yu C, Shen B, Sevilla C, Shi M, et al. PhaSepDB: a database of liquid–liquid phase separation related proteins. Nucleic Acids Res 2019;48:D354–9.
[69] Ning W, Guo Y, Lin S, Mei B, Wu Y, Jiang P, et al. DrLLPS: a data resource of liquid–liquid phase separation in eukaryotes. Nucleic Acids Res 2020;48:D288–95.
[70] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks. KDD 2016;2016:855–64.
[71] Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, et al. The BioPlex network: a systematic exploration of the human interactome. Cell 2015;162:425–40.
[72] Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. Nature 2017;545:505–9.
[73] Saito M, Hess D, Eglinger J, Fritsch AW, Kreysing M, Weinert BT, et al. Acetylation of intrinsically disordered regions regulates phase separation. Nat Chem Biol 2019;15:51–61.
[74] Tsang B, Arsenault J, Vernon RM, Lin H, Sonenberg N, Wang LY, et al. Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. Proc Natl Acad Sci U S A 2019;116:4218–27.
[75] Hofweber M, Dormann D. Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics. J Biol Chem 2019;294:7137–50.
[76] Tanikawa C, Ueda K, Suzuki A, Iida A, Nakamura R, Atsuta N, et al. Citrullination of RGG motifs in FET proteins by PAD4 regulates protein aggregation and ALS susceptibility. Cell Rep 2018;22:1473–83.
[77] Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. Nucleic Acids Res 2015;43:D512–20.
[78] Ochoa D, Jarnuczak AF, Viéitez C, Gehre M, Soucheray M, Mateus A, et al. The functional landscape of the human phosphoproteome. Nat Biotechnol 2019;38:365–73.
[79] Thul PJ, kesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science 2017;356:eaal3321.
[80] Yokoshi M, Fukaya T. Dynamics of transcriptional enhancers and chromosome topology in gene regulation. Dev Growth Differ 2019;61:343–52.
[81] Yu C, Shen B, You K, Huang Q, Shi M, Wu C, et al. Proteome-scale analysis of phase-separated proteins in immunofluorescence images. Brief Bioinform 2020;22:bbaa187.
[82] McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: next-generation image processing for biology. PLoS Biol 2018;16:e2005970.