D360–D367 *Nucleic Acids Research, 2020, Vol. 48, Database issue* *Published online 15 October 2019*
doi: 10.1093/nar/gkz848

PhaSePro: the database of proteins driving liquid–liquid phase separation

Bálint Mészáros[1,†], Gábor Erdős[1,†], Beáta Szabó[2], Éva Schád[2], Ágnes Tantos[2], Rawan Abukhairan[2], Tamás Horváth[2], Nikoletta Murvai[2], Orsolya P. Kovács[2], Márton Kovács[2], Silvio C.E. Tosatto[3], Péter Tompa[2,4], Zsuzsanna Dosztányi[1] and Rita Pancsa[2,*]

[1]MTA-ELTE Momentum Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, Budapest H-1117, Hungary, [2]Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest H-1117, Hungary, [3]Department of Biomedical Sciences, University of Padova CNR Institute of Neuroscience, Padova, Italy and [4]Structural Biology (CSB), Brussels, Belgium; Structural Biology Brussels (SBB), Vrije Universiteit Brussel (VUB), Brussels 1050, Belgium

Received August 07, 2019; Revised September 11, 2019; Editorial Decision September 18, 2019; Accepted October 07, 2019

ABSTRACT

Membraneless organelles (MOs) are dynamic liquid condensates that host a variety of specific cellular processes, such as ribosome biogenesis or RNA degradation. MOs form through liquid–liquid phase separation (LLPS), a process that relies on multivalent weak interactions of the constituent proteins and other macromolecules. Since the first discoveries of certain proteins being able to drive LLPS, it emerged as a general mechanism for the effective organization of cellular space that is exploited in all kingdoms of life. While numerous experimental studies report novel cases, the computational identification of LLPS drivers is lagging behind, and many open questions remain about the sequence determinants, composition, regulation and biological relevance of the resulting condensates. Our limited ability to overcome these issues is largely due to the lack of a dedicated LLPS database. Therefore, here we introduce PhaSePro (https://phasepro.elte.hu), an openly accessible, comprehensive, manually curated database of experimentally validated LLPS driver proteins/protein regions. It not only provides a wealth of information on such systems, but improves the standardization of data by introducing novel LLPS-specific controlled vocabularies. PhaSePro can be accessed through an appealing, user-friendly interface and thus has definite potential to become the central resource in this dynamically developing field.

INTRODUCTION

One of the most exciting recent developments in the field of molecular cell biology is the discovery that certain proteins can undergo liquid–liquid phase separation (LLPS) inside the cell, driving the formation of diverse membraneless organelles/biological condensates, such as stress granules, P-bodies, the nucleolus and postsynaptic densities (1–3). These dynamic, non-stoichiometric supramolecular assemblies represent a unique functional and structural level of cellular organization. Their functions often cannot be derived from the functions of individual proteins, but emerge from the collective behaviour of their constituent macromolecules (4). They confer a wide range of functional advantages on cells (5) due to their unique material properties (5,6), and rapid responses to environmental triggers (8). It was also proposed that multiple non-specific weak interactions can more readily emerge and be maintained through evolution than specific strong interactions, and thus liquid condensates exhibit a favourable cost-benefit ratio (7). Ever since the discovery and first analysis of liquid droplets in Drosophila embryos (9), an increasing number of cellular functions have been ascribed to such liquid condensates, including the regulation of most stages of the life cycle of RNAs (10–13), transcriptional regulation and silencing (14,15) and the signal transduction networks of membrane receptors (16,17). LLPS has emerged as a general mechanism of cellular organization, exploited not only by eukaryotic cells, but also by bacteria and viruses (18,19). Besides diverse physiological roles of these dense liquid condensates (1–3), their fundamental roles are also highlighted by the fact that mutations affecting their regulation are often implicated in devastating neurological disorders, such as amy-

*To whom correspondence should be addressed. Tel: +36 1 382 6705; Email: pancsa.rita@ttk.mta.hu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

otrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), and are also associated with cancers and muscular atrophies (20).

The ability to drive liquid–liquid phase separation (LLPS) is encoded in protein sequences, but it can be achieved by diverse functional modules, including disordered regions of low sequence complexity, multivalent domain – motif interactions, RNA binding domains, oligomerization domains and various combinations of these modules (21). As a unifying feature, LLPS is typically driven by multivalent weak interactions, which ensure the dynamic, liquid-like properties of the resulting condensates. Liquid condensates show a broad range of morphologies, shapes, sizes and compositions (2). Some of them are constituted of a single dense phase, while others show intricate core-shell structures, in which multiple immiscible phases are embedded into each other as dictated by their surface tension and viscosity properties (13,22,23). Many have a rounded shape as true liquid droplets (24,25), or form more irregular structures along the surface of membranes (16,17,26). While some only contain a few types of macromolecules (27,28), others, such as P-bodies and stress granules, host hundreds of proteins and thousands of RNA molecules (29–32). It has been proposed that regardless of the size and compositional richness of the condensates, usually only one or a few proteins, referred to as 'scaffolds', drive their formation. The other constituents, the so-called 'client' proteins, do not substantially contribute to the formation of the condensates. Through interactions with the scaffold, clients are readily compartmentalized by the condensate and often directly promote its characteristic functions (33,34).

While in the last years an avalanche of high-impact publications reported on novel cases of proteins involved in LLPS with the numbers still fast increasing, many open questions remain about the composition of the various condensates, their exact biological roles and modes of regulation, and the sequence characteristics of protein regions that drive LLPS. To a large extent, our limited insight comes from the lack of a comprehensive, curated database of experimentally validated LLPS driver proteins/protein regions. To fill this gap, here we introduce the PhaSePro database that is a comprehensive, carefully curated resource of proteins that have been experimentally demonstrated to drive LLPS.

INFORMATION AVAILABLE IN PhaSePro

PhaSePro (https://phasepro.elte.hu) is a novel database that provides a wealth of information on LLPS in a structured way. It is a queryable, public database currently containing 121 entries (as of September 2019) collected through comprehensive literature curation. Entries in the database describe proteins/protein regions that were demonstrated to drive LLPS together with the supporting literature references. Proteins can drive phase separation on their own or as part of well-defined multicomponent systems, the two scenarios being clearly distinguished in PhaSePro. PhaSePro currently includes 109 eukaryotic, 5 bacterial and 7 viral entries, well illustrating that the formation of liquid condensates through LLPS is a universal mechanism for creating dynamic subcompartments in the cell that has been demonstrated across different domains of life, as well as viruses.

We have collected proteins experimentally verified to drive phase separation *in vivo* and/or *in vitro* from the literature; cases inferred from computational prediction or homology were not included. The core data represented in PhaSePro were derived from manual curation, which are then extended with data automatically retrieved from diverse resources (Figure 1).

The bulk of information from manual annotation is encoded in a structured way through the use of references to existing databases, ontologies and custom-built controlled vocabularies (CVs). The protein sequences investigated in the experiments addressing LLPS are defined using canonical UniProt sequences, while for three entries specific isoforms needed to be used. The sequence boundaries and characteristics of the experimentally validated LLPS driver region(s) are also provided for each entry protein. The exact sequence regions driving LLPS were confirmed for most of the entry proteins, while in some cases only the full-length protein has been subjected to experiments. These scenarios are clearly distinguished in PhaSePro. For the proteins that are parts of multi-component systems being able to drive LLPS only together, PhaSePro defines the accessory proteins needed for LLPS. For all LLPS proteins/systems PhaSePro provides information on the membraneless organelle(s) (MO) formed, expressed via GeneOntology cellular component terms.

Position specific information known or assumed to directly affect LLPS based on the processed literature is specified using the UniProt sequence as reference. These information include post-translational modifications, disease mutations (defined through dbSNP (35) if possible and connected to OMIM (36)), and alternative splicing events. In addition, all other isoforms that contain any sequence changes within the LLPS driver region were also added from UniProt. Functional and molecular aspects of the organelle formation derived from the literature are encoded via purpose-built LLPS-specific CVs (see Supplementary Tables S1–S4 and next section for details).

The biological function of the formed organelle, the molecular determinants and the required molecular partners of LLPS are provided in the form of free-text descriptions. To decide whether a given case of LLPS was partner and/or modification dependent, only experimental parameters close to physiological conditions and protein concentration were considered. Experimental procedures used for the investigation of LLPS are condensed from the relevant papers by the annotators, and are detailed as free text enriched with ontology links (see next section).

Automated annotations also enrich the information provided by PhaSePro. All post-translational modifications (PTMs) that have been described for the full protein including phosphorylation, methylation, acetylation and ubiquitination were taken from PhosphoSitePlus (37); protein disorder predictions from the IUPred2A server (38), and conserved protein region assignations from Pfam (39). Sequence variants are imported through ProtVista (40), with predicted deleterious variants filtered. In addition, all structures that overlap with the LLPS driving region are imported from PDBe (41).

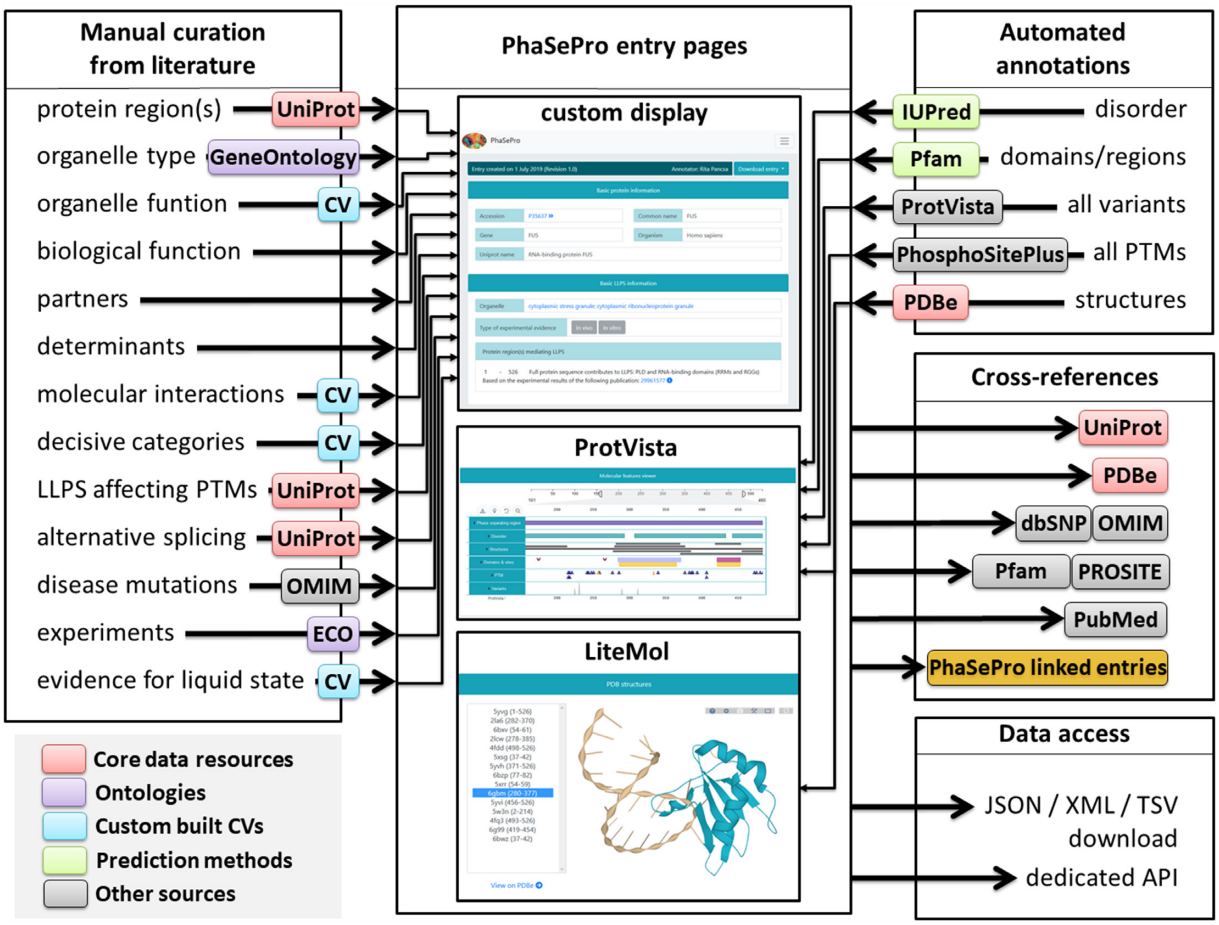

Figure 1. Data integrated into PhaSePro. Manual curation (left) is based on the literature. The biological function of the organelle, the partners required for and other determinants of LLPS are expressed as free text descriptions, while all other annotations are encoded using references to UniProt sequences or as terms in various ontologies and controlled vocabularies (CVs) (see Supplementary Tables). Automated annotations (top right) are added from the outputs of various sequence-based prediction methods (IUPred (38) for disorder prediction and Pfam (39) for conserved protein regions), from ProtVista (40) (for sequence variants omitting predicted variants), from PhosphoSitePlus (37) (for phosphorylation, methylation, acetylation and ubiquitination sites) and from PDBe (41) (for structures overlapping the LLPS driver region). Entries are cross-linked to various data resources and the data contained in PhaSePro can be accessed via download or the dedicated API.

For each protein we have a cross-reference to UniProt entries to include core information, including accessions, gene/protein names and source species. Other featured information are linked to the respective database entries (PDBe, Pfam/PROSITE, OMIM/dbSNP). All information taken from the literature through manual curation is linked to corresponding manuscripts via EuropePMC (42).

INTRODUCTION OF LLPS-SPECIFIC CONTROLLED VOCABULARIES AND EXTENSIONS TO AVAILABLE ONTOLOGIES

The majority of information in PhaSePro is expressed using controlled vocabularies (CVs) and ontologies to aid the findability of the data and the interoperability of the database, advancing adherence to FAIR principles (43). Using CVs instead of free text for data representation helps the interpretation of data, aids computational analyses, and reduces redundancy of information in databases. Correspondingly, several fields of biology established their respective CVs and ontologies, including biomolecular interactions (44) of post-translational modifications (45). However, CVs have not been developed for the rapidly expanding field of LLPS, and in parallel with the development of PhaSePro, we also laid down the foundations of data standardization by the development of 4 distinct CVs that describe the following four aspects of LLPS and the membraneless organelles formed.

(i) The functional roles of membraneless organelles (MOs)/granules in the cell are defined using eight classes already defined in several dedicated reviews (3–6) (Supplementary Table S1). These include terms such as 'protective storage/reservoir' for MOs that store molecules in an inactive state, or 'activation/nucleation/signal amplification/bioreactor' for MOs that bring together components of a reaction. (ii) We also introduced a CV of 19 terms for the different molecular interaction types that could contribute to LLPS based on (21), including terms such as 'multivalent domain-motif interactions' or 'coiled-coil formation' (Supplementary Table S2). (iii) A dedicated CV with 6 terms was defined to describe the molecular determinants and mechanisms that are

pertinent to LLPS, such as the ability of the proteins to form membrane clusters or if PTMs are required for the LLPS (Supplementary Table S3). (iv) Finally, the diverse experimental observations that could support the liquid state of the condensates, such as temperature-dependence or the observed dynamic exchange of molecules within the droplet, were grouped in a separate CV of seven terms that was partly built based on the review by Mitrea D *et al.* (46) (Supplementary Table S4). The terms of these CVs are explained in more detail on the About/Help page of PhaSePro.

To ensure the best integration with frequently used ontologies, we reviewed the existing classification of membraneless organelles available in Gene Ontology (GO) (47) (see Supplementary Table S5). Several cellular component GO terms, falling under the 'non-membrane bounded organelle' parent term already describe membraneless organelles. However, to make PhaSePro annotations more precise, we created new terms when it was necessary. Experimental procedures used for studying liquid–liquid phase separation were similarly reviewed, connecting them to terms of the Evidence and Conclusion Ontology (48) (ECO—see Supplementary Table S6). The use of GO and ECO are fully in line with the practices of core data resources, such as UniProt, and will enable future integration and standardization efforts.

IMPLEMENTATION AND SERVER FEATURES

PhaSePro is presented through a DJANGO (version 2.1.1) based web interface, fueled by a multi-layer SQL database, which allows a vast amount of parallel queries to be completed in a fraction of a second. The SQL database contains all the information collected from the literature alongside with protein information derived from UniProt. Each record represents a single protein and is linked to a unique UniProt accession. To maintain the best possible compatibility through the various devices and browsing options users have, the front-end of PhaSePro is represented solely as a combination of bootstrap (version 4.3.1) and JQuery (version 2.1.4).

In addition to access the data through the online interface, data can also be accessed via downloading the data in JSON, XML or TSV formats, or by the RESTful API serving standard JSON format (e.g. https://phasepro.elte.hu/rest/P35637.json). New data currently missing from PhaSePro can be submitted through a dedicated interface found on the 'Annotate' page.

THE INTERFACE

The online interface of PhaSePro offers convenient approaches for users to find relevant data. The web-server opens with a Home page containing general information on LLPS and specific information of PhaSePro, with links to two example entry pages, and a search bar, which can be found on the top of the Browse/Search page as well. The search bars use a completion based method, offering the best matches for the queries. The browse function encompasses a table containing all entries, where users are able to filter the database by various options alongside with the ability to search by keywords or regular expressions. Clicking on any row in either the search bar results or inside the Browse table directs the user to the relevant entry page. Any set of entries compiled through the use of filters can be further customized and downloaded in any of the three available formats (JSON, XML and TSV).

A comprehensive online documentation about the usage and functionalities of PhaSePro are available on the 'About/Help' page.

PhaSePro also provides information on candidate proteins that are likely to drive LLPS but their status cannot be fully ascertained by the available experimental data. These entries are collected on the 'Candidates' page (https://phasepro.elte.hu/candidates), which is presented in a similar fashion as the 'Browse' page, with the same functionalities, including the option of downloading them. The 'Statistics' page provides various statistics about the database, including the fraction of entries with *in vivo*/*in vitro* support, taxonomic distributions of proteins, or the frequencies of various terms in the four developed CVs.

ENTRY PAGES

Each entry in PhaSePro correspond to a single protein, and has a dedicated entry page detailing all relevant information collected either manually from the literature or in an automated way from source databases (Figure 2).

The first section of the entry page contains the basic information on the protein undergoing LLPS and the formed organelle. The top bar contains information on the release of the given entry and the annotator(s), and provides the option of downloading the annotation of the given entry in JSON, TSV and XML formats. Also, it is indicated here when the given entry is not a self-sufficient LLPS driver, but is part of a multi-component system in which the collective behaviour of several proteins are required for LLPS to occur. This bar is followed by core information on the given entry protein from UniProt (49) (UniProt accession, common name, gene name, Ensembl transcript ID, species name, NCBI taxon ID, and protein name). On the right side, the PhaSePro core information is provided, namely the specific membraneless organelle connected to the GO cellular components system, the specification of the type of experimental evidence (*in vitro*, *in vivo* or both), the joined entries (if the given entry is part of a multicomponent LLPS system) and specific sequence region(s) that drive LLPS, including UniProt sequence boundaries and a description of the domain/compositional/disorder properties of the given sequence region supported by literature reference.

These core sections are followed by a section for the graphical representation of sequence- and structure-level information connected to the protein. For sequence-level annotations PhaSePro applies the recently released smart visualization tool of UniProt, ProtVista (40) that enables mapping of UniProt-annotated features onto the entry proteins, including domain and site information, variations and disease-associated mutations, and many more. To fit with the requirements specific for our database, the molecular features viewer was extended with our annotated LLPS regions. As many regions driving LLPS are intrinsically disordered and are regulated by posttranslational modifica-

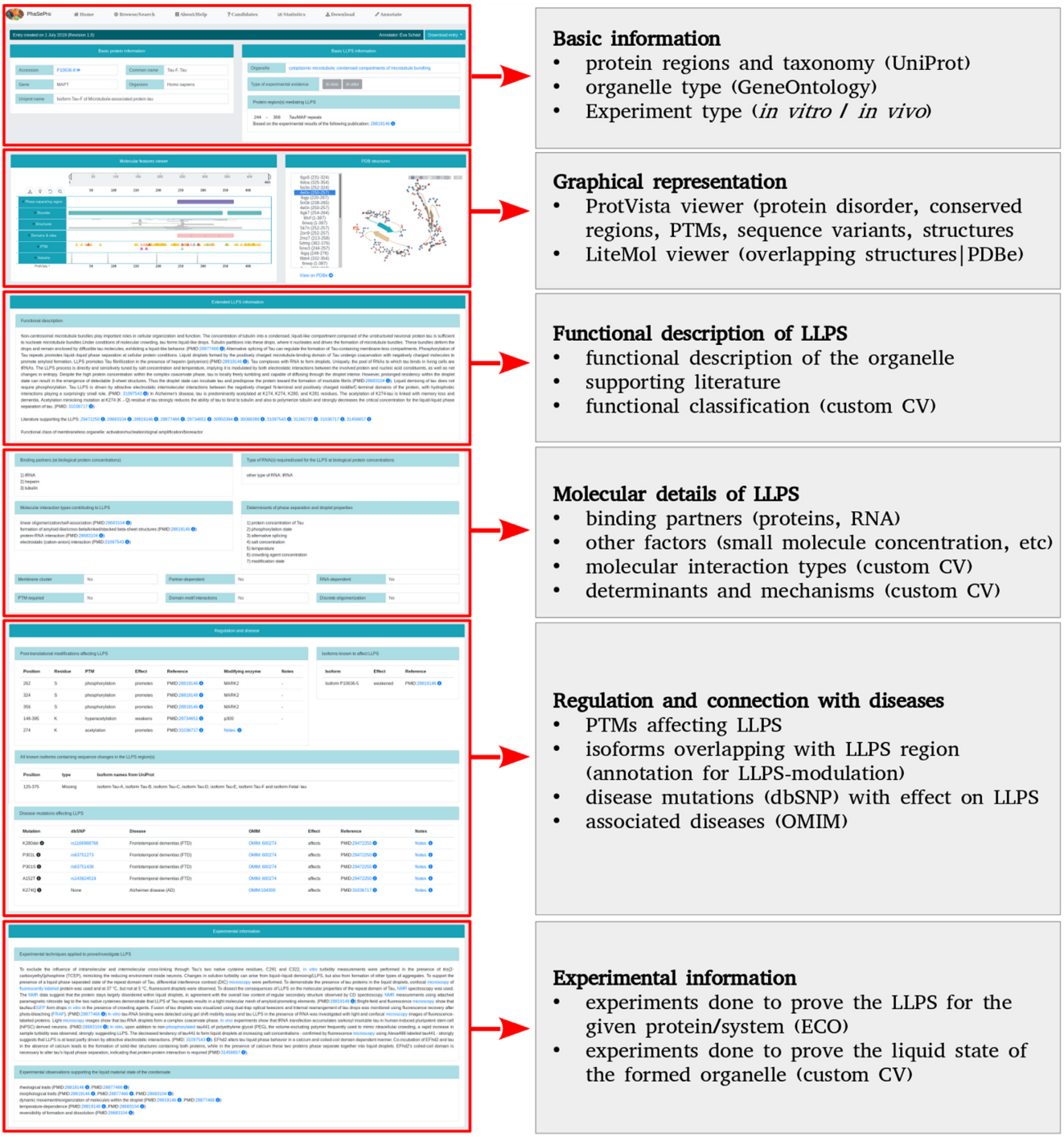

Figure 2. Entry pages in PhaSePro. Each entry page is structured to display information pertaining to specific aspects of the liquid–liquid phase separation (LLPS) or the formed membraneless organelles. The figure uses the entry of human Tau protein (https://phasepro.elte.hu/entry/P10636--8).

tions (PTMs), disorder prediction by IUPred (38), domain predictions by PfamScan (50) and PTMs from PhosphoSitePlus (37) were also incorporated into the graphical representation. We provide an overview of the PDB structures overlapping with the corresponding LLPS protein regions, if available, together with a LiteMol structure viewer (51) for their visualization. The viewer provides an interactive cartoon-style structural view for the selected overlapping PDB structure in beige color with the region(s) overlapping the annotated LLPS protein region(s) highlighted in turquoise in one of the chains belonging to the entry protein.

The visualization tools are followed by a larger block of 'Extended LLPS information' incorporating two separate

sections. The first section gives a description of the functional relevance and distinctive features of the given LLPS system using a free-text description distilled by the annotators based on the relevant literature. Together with the description, a list of related articles is also provided. To represent the functional class(es) of the formed MO in a more structured way, it is also defined using the custom built CV (Supplementary Table S1).

The second part of the extended LLPS section provides information on the partners and other determinants known to be required for, or to promote or negatively regulate LLPS. As several known LLPS events are known to depend on the presence of RNA, a dedicated block specifies what type of RNA(s) (if any) are involved. If known, the molecular interaction types playing a role in driving the given LLPS process are also listed using the CV described in Supplementary Table S2. At the end of this section, categorical classifications of the given LLPS system are provided using the CV described in Supplementary Table S3.

The 'Regulation and disease' section provides detailed information on the PTMs and alternative splicing events demonstrated to affect LLPS by experiments. Both types of data are organized into a strict format, defining the sequence change together with other supporting information defining the effect the change has on LLPS and cross-references to supporting literature. Apart from isoforms directly affecting LLPS, this section also provides alternative splicing-derived isoforms that have not yet been shown to have a direct effect on LLPS, but do contain sequence changes within the annotated LLPS region, and thus may show an altered LLPS behaviour compared to the canonical isoform. Disease mutations, whose effects on LLPS have been experimentally investigated (if any) are also listed using the sequence variant nomenclature (52), crosslinked to dbSNP, together with the related diseases crosslinked to OMIM.

Furthermore, each entry page contains a block of 'Experimental information' that contains an extensive textual description of the LLPS-specific experiments performed that is enriched with links to associated ECO ontology terms. Finally, a list of evidence that support the liquid state of the given condensate backed by literature references is given using the CV detailed in Supplementary Table S4.

CONCLUSIONS

PhaSePro is a database of proteins driving liquid–liquid phase separation that aims to provide an up-to-date view on the variety of biological condensates that rely on LLPS, including their major architectural properties, functions and regulation. It is carefully curated and incorporates experimentally validated LLPS drivers from all kingdoms of life. We are convinced that PhaSePro will greatly benefit the scientific community by providing (i) a freely accessible, easy-to-use, organized resource with all relevant data on LLPS proteins, (ii) the basis for standardization of experimental approaches and functional characterization, (iii) crucial data for furthering the elucidation of the sequence determinants and molecular mechanisms enabling liquid–liquid phase separation and (iv) a high-quality training set for the development of new methods targeting the computational identification of novel LLPS proteins.

By storing the most comprehensive list of phase separation driver proteins published so far, supplied with detailed annotation on the biological relevance and regulation of condensates, PhaSePro has a definite potential to become the central resource in this fast-expanding field. To achieve this goal we are dedicated to ensure the long-term availability of the database.

DATA AVAILABILITY

Our aim is to maintain a regularly updated online resource with periodic releases at least twice a year, by incorporating previously published LLPS literature. To successfully accomplish this goal, we also kindly encourage the scientific community to submit newly identified phase separation driver proteins to PhaSePro using our detailed downloadable annotation guidelines and sample document or the more simple online submission form (https://phasepro.elte.hu/annotate). We also encourage authors to contact us if they have published new information on the already existing entries. As PhaSePro is located in the EU, data collection from users and submitters through the server is executed via a secure interface using HTTPS, and fully adheres to the General Data Protection Regulation (GDPR).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Hungarian National Research, Development, and Innovation Office (NKFIH) [FK-128133 to R.P., K-124670 to P.T. and K-125340 to A.T.]; Hungarian Academy of Sciences [LP2014-18 to Z.D., PREMIUM-2017-48 to R.P.]; VUB [SRP51, 2019–24 to P.T.]; New National Excellence Programme (UNKP) [ELTE/12653/314(2018) to M.K.]; National Research Council of Science and Technology (NST) of Korea [NTM2231611 to P.T.]; European Union's Horizon 2020 research and innovation programme [778247 (IDPfun) to S.T., P.T. and Z.D.]. Funding for open access charge: Hungarian Academy of Sciences.
Conflict of interest statement. None declared.

REFERENCES

1. Shin,Y. and Brangwynne,C.P. (2017) Liquid phase condensation in cell physiology and disease. *Science*, **357**, eaaf4382.
2. Banani,S.F., Lee,H.O., Hyman,A.A. and Rosen,M.K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, **18**, 285–298.
3. Alberti,S., Gladfelter,A. and Mittag,T. (2019) Considerations and challenges in studying liquid–liquid phase separation and biomolecular condensates. *Cell*, **176**, 419–434.
4. Pancsa,R., Schad,E., Tantos,A. and Tompa,P. (2019) Emergent functions of proteins in non-stoichiometric supramolecular assemblies. *Biochim. Biophys. Acta: Proteins Proteomics*, **1867**, 970–979.
5. Alberti,S. (2017) The wisdom of crowds: regulating cell function through condensed states of living matter. *J. Cell Sci.*, **130**, 2789–2796.

6. Kaganovich,D. (2017) There is an inclusion for that: Material properties of protein granules provide a platform for building diverse cellular functions. *Trends Biochem. Sci.*, **42**, 765–776.
7. Li,X.-H., Chavali,P.L., Pancsa,R., Chavali,S. and Babu,M.M. (2018) Function and regulation of phase-separated biological condensates. *Biochemistry*, **57**, 2452–2461.
8. Yoo,H., Triandafillou,C. and Drummond,D.A. (2019) Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.*, **294**, 7151–7159.
9. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Jülicher,F. and Hyman,A.A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, **324**, 1729–1732.
10. Luo,Y., Na,Z. and Slavoff,S.A. (2018) P-bodies: composition, properties, and functions. *Biochemistry*, **57**, 2424–2431.
11. Sheu-Gruttadauria,J. and MacRae,I.J. (2018) Phase transitions in the assembly and function of human miRISC. *Cell*, **173**, 946–957.
12. Sawyer,I.A., Hager,G.L. and Dundr,M. (2017) Specific genomic cues regulate Cajal body assembly. *RNA Biol.*, **14**, 791–803.
13. Feric,M., Vaidya,N., Harmon,T.S., Mitrea,D.M., Zhu,L., Richardson,T.M., Kriwacki,R.W., Pappu,R.V. and Brangwynne,C.P. (2016) Coexisting liquid phases underlie nucleolar subcompartments. *Cell*, **165**, 1686–1697.
14. Larson,A.G., Elnatan,D., Keenen,M.M., Trnka,M.J., Johnston,J.B., Burlingame,A.L., Agard,D.A., Redding,S. and Narlikar,G.J. (2017) Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, **547**, 236–240.
15. Sabari,B.R., Dall'Agnese,A., Boija,A., Klein,I.A., Coffey,E.L., Shrinivas,K., Abraham,B.J., Hannett,N.M., Zamudio,A.V., Manteiga,J.C. *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, **361**, eaar3958.
16. Su,X., Ditlev,J.A., Hui,E., Xing,W., Banjade,S., Okrut,J., King,D.S., Taunton,J., Rosen,M.K. and Vale,R.D. (2016) Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*, **352**, 595–599.
17. Li,P., Banjade,S., Cheng,H.-C., Kim,S., Chen,B., Guo,L., Llaguno,M., Hollingsworth,J.V., King,D.S., Banani,S.F. *et al.* (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature*, **483**, 336–340.
18. Al-Husini,N., Tomares,D.T., Bitar,O., Childers,W.S. and Schrader,J.M. (2018) α-Proteobacterial RNA degradosomes assemble liquid–liquid phase-separated RNP bodies. *Mol. Cell*, **71**, 1027–1039.
19. Nikolic,J., Le Bars,R., Lama,Z., Scrima,N., Lagaudrière-Gesbert,C., Gaudin,Y. and Blondel,D. (2017) Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.*, **8**, 58.
20. Mackenzie,I.R., Nicholson,A.M., Sarkar,M., Messing,J., Purice,M.D., Pottier,C., Annu,K., Baker,M., Perkerson,R.B., Kurti,A. *et al.* (2017) TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron*, **95**, 808–816.
21. Mittag,T. and Parker,R. (2018) Multiple modes of protein-protein interactions promote RNP granule assembly. *J. Mol. Biol.*, **430**, 4636–4649.
22. Boeynaems,S., Alberti,S., Fawzi,N.L., Mittag,T., Polymenidou,M., Rousseau,F., Schymkowitz,J., Shorter,J., Wolozin,B., Van Den Bosch,L. *et al.* (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.*, **28**, 420–435.
23. Boeynaems,S., Holehouse,A.S., Weinhardt,V., Kovacs,D., Van Lindt,J., Larabell,C., Van Den Bosch,L., Das,R., Tompa,P.S., Pappu,R.V. *et al.* (2019) Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 7889–7898.
24. Berry,J., Weber,S.C., Vaidya,N., Haataja,M. and Brangwynne,C.P. (2015) RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5237–E5245.
25. Roovers,E.F., Kaaij,L.J.T., Redl,S., Bronkhorst,A.W., Wiebrands,K., de Jesus Domingues,A.M., Huang,H.-Y., Han,C.-T., Riemer,S., Dosch,R. *et al.* (2018) Tdrd6a regulates the aggregation of buc into functional subcellular compartments that drive germ cell specification. *Dev. Cell*, **46**, 285–301.
26. Ma,W. and Mayr,C. (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3′UTR-mediated protein-protein interactions. *Cell*, **175**, 1492–1506.
27. Zacharogianni,M., Aguilera-Gomez,A., Veenendaal,T., Smout,J. and Rabouille,C. (2014) A stress assembly that confers cell viability by preserving ERES components during amino-acid starvation. *Elife*, **3**, e04132.
28. Guillén-Boixet,J., Buzon,V., Salvatella,X. and Méndez,R. (2016) CPEB4 is regulated during cell cycle by ERK2/Cdk1-mediated phosphorylation and its assembly into liquid-like droplets. *Elife*, **5**, e19298.
29. Hubstenberger,A., Courel,M., Bénard,M., Souquere,S., Ernoult-Lange,M., Chouaib,R., Yi,Z., Morlot,J.-B., Munier,A., Fradet,M. *et al.* (2017) P-Body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell*, **68**, 144–157.
30. Jain,S., Wheeler,J.R., Walters,R.W., Agrawal,A., Barsic,A. and Parker,R. (2016) ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell*, **164**, 487–498.
31. Khong,A., Matheny,T., Jain,S., Mitchell,S.F., Wheeler,J.R. and Parker,R. (2017) The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell*, **68**, 808–820.
32. Youn,J.-Y., Dunham,W.H., Hong,S.J., Knight,J.D.R., Bashkurov,M., Chen,G.I., Bagci,H., Rathod,B., MacLeod,G., Eng,S.W.M. *et al.* (2018) High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell*, **69**, 517–532.
33. Ditlev,J.A., Case,L.B. and Rosen,M.K. (2018) Who's in and who's out-compositional control of biomolecular condensates. *J. Mol. Biol.*, **430**, 4666–4684.
34. Banani,S.F., Rice,A.M., Peeples,W.B., Lin,Y., Jain,S., Parker,R. and Rosen,M.K. (2016) Compositional control of phase-separated cellular bodies. *Cell*, **166**, 651–663.
35. Sherry,S.T. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
36. Amberger,J.S., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2019) OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.*, **47**, D1038–D1043.
37. Hornbeck,P.V., Kornhauser,J.M., Latham,V., Murray,B., Nandhikonda,V., Nord,A., Skrzypek,E., Wheeler,T., Zhang,B. and Gnad,F. (2019) 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.*, **47**, D433–D441.
38. Mészáros,B., Erdos,G. and Dosztányi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
39. El-Gebali,S., Mistry,J., Bateman,A., Eddy,S.R., Luciani,A., Potter,S.C., Qureshi,M., Richardson,L.J., Salazar,G.A., Smart,A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.
40. Watkins,X., Garcia,L.J., Pundir,S., Martin,M.J. and UniProt Consortium (2017) ProtVista: visualization of protein sequence annotations. *Bioinformatics*, **33**, 2040–2041.
41. Mir,S., Alhroub,Y., Anyango,S., Armstrong,D.R., Berrisford,J.M., Clark,A.R., Conroy,M.J., Dana,J.M., Deshpande,M., Gupta,D. *et al.* (2018) PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, **46**, D486–D492.
42. Europe PMC Consortium (2015) Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Res.*, **43**, D1042–D1048.
43. Wilkinson,M.D., Dumontier,M., Aalbersberg,I.J.J., Appleton,G., Axton,M., Baak,A., Blomberg,N., Boiten,J.-W., da Silva Santos,L.B., Bourne,P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.
44. Sivade Dumousseau,M., Alonso-López,D., Ammari,M., Bradley,G., Campbell,N.H., Ceol,A., Cesareni,G., Combe,C., De Las Rivas,J., Del-Toro,N. *et al.* (2018) Encompassing new use cases - level 3.0 of the HUPO-PSI format for molecular interactions. *BMC Bioinformatics*, **19**, 134.
45. Montecchi-Palazzi,L., Beavis,R., Binz,P.-A., Chalkley,R.J., Cottrell,J., Creasy,D., Shofstahl,J., Seymour,S.L. and Garavelli,J.S. (2008) The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.*, **26**, 864–866.
46. Mitrea,D.M., Chandra,B., Ferrolino,M.C., Gibbs,E.B., Tolbert,M., White,M.R. and Kriwacki,R.W. (2018) Methods for physical characterization of phase-separated bodies and membrane-less organelles. *J. Mol. Biol.*, **430**, 4773–4805.

47. The Gene Ontology Consortium and The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
48. Giglio,M., Tauber,R., Nadendla,S., Munro,J., Olley,D., Ball,S., Mitraka,E., Schriml,L.M., Gaudet,P., Hobbs,E.T. *et al.* (2019) ECO, the Evidence & Conclusion Ontology: community standard for evidence information. *Nucleic Acids Res.*, **47**, D1186–D1194.
49. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
50. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
51. Sehnal,D., Deshpande,M., Vařeková,R.S., Mir,S., Berka,K., Midlik,A., Pravda,L., Velankar,S. and Koča,J. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods*, **14**, 1121–1122.
52. den Dunnen,J.T., Dalgleish,R., Maglott,D.R., Hart,R.K., Greenblatt,M.S., McGowan-Jordan,J., Roux,A.-F., Smith,T., Antonarakis,S.E. and Taschner,P.E.M. (2016) HGVS recommendations for the description of sequence variants: 2016 update. *Hum. Mutat.*, **37**, 564–569.