Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Databases of liquid–liquid phase separation proteins and condensates

Carlos Pintado-Grima[1], Michał Burdukiewicz[2,3,4] and Salvador Ventura[1,5]

Protein liquid–liquid phase separation (LLPS) and the formation of biomolecular condensates are increasingly recognized as important in both cellular physiology and pathology. However, the computational databases developed to catalog LLPS-associated proteins differ significantly in scope, curation strategies, evidence models, and annotation granularity. Here, we compare the main LLPS databases, discuss what each resource captures well, where role definitions and evidence models diverge, which metadata remain limited, and how these differences affect data interpretation and reuse. We argue that future resources should prioritize interoperable vocabularies, explicit data origin, condition-aware annotations, and stronger integration with region-level, structural, and mutational information.

Addresses
[1] Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain
[2] Clinical Research Centre, Medical University of Białystok, Kilińskiego 1, Białystok 15-369, Poland
[3] Institute of Biotechnology, Vilnius University, Saulėtekio al. 7, 10257 Vilniaus, Lithuania
[4] Faculty of Life Sciences, University of Bradford, BD7 1DP Bradford, United Kingdom
[5] Hospital Universitari Parc Taulí, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Universitat Autònoma de Barcelona, Sabadell, Spain

Corresponding author: Ventura, Salvador (salvador.ventura@uab.cat)

Current Opinion in Structural Biology 2026, **100**:103341

This review comes from a themed issue on **Catalysis and Regulation (2026): Molecular Condensates**

Edited by **Francesco Aprile** and **Salvador Ventura**

For complete overview of the section, please refer the article collection - Catalysis and Regulation (2026): Molecular Condensates

https://doi.org/10.1016/j.sbi.2026.103341

0959-440/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Introduction

Intracellular environments are crowded and highly organized milieus in which diverse molecular species must be coordinated in space and time [1,2]. Biomolecular condensates provide such organization by compartmentalizing biochemical reactions without delimiting membranes, often through liquid–liquid phase separation (LLPS), in which macromolecules reversibly demix into dense and dilute phases [3,4]. In contrast to classical membrane-bound organelles, such as mitochondria or the nucleus, condensates' formation is governed by material properties, so they are commonly described as membraneless organelles. They may contain proteins, nucleic acids, and occasionally lipids, with different contributions to condensate assembly and function [5,6]. Proteins are central to this behavior because they can undergo reversible rearrangements in response to changing cellular conditions [7,8]. Condensate formation and regulation are strongly influenced by the molecular properties of their protein components, including sequence grammar, residue distribution, multivalency, and conformational plasticity, particularly in disordered proteins, together with partner-dependent interactions and post-translational modifications.

Biomolecular condensate formation relies on weak multivalent interactions, including electrostatic interactions, π–π contacts between aromatic residues, cation–π interactions, and interactions mediated by short motifs or folded domains [9]. Intrinsically disordered regions (IDRs) often make major contributions to these assemblies because their sequence composition and conformational heterogeneity favor reversible, multivalent interactions [9]. Yet, LLPS is not restricted to disordered proteins: structured interaction modules, oligomerization interfaces, and surfaces of globular domains can also provide the multivalency required for condensate formation [10]. In this context, the stickers-and-spacers framework helps rationalize how sequence patterning influences LLPS propensity, where the distribution and patterning of specific sticky residues (e.g., Y, R, and F) in between spacers (e.g., G, S, Q, and N) have been proposed as a general model to explain the regulation of LLPS [11].

The biological relevance of these processes is illustrated by the wide range of condensates, including the nucleolus, stress granules, P bodies, Cajal bodies, and synaptic

densities, described across organisms [3]. In yeast, proteins containing prion-like domains can be recruited to stress granules upon heat shock and return to solution once stress is relieved [12,13]. In human cells, RNA-binding proteins such as FUS and TDP-43 undergo LLPS under physiological conditions, yet under pathological circumstances, often driven by mutations or altered cellular environments, their material properties can shift toward aberrant solid-like assemblies associated with amyotrophic lateral sclerosis and frontotemporal dementia [14,15].

As the field has expanded, so too has the number of proteins reported to participate in condensates [16–18]. However, the label 'LLPS protein' is applied to entities supported by very different kinds of evidence: some phase separate in purified systems, some do so only in the presence of specific partners, some merely partition into pre-existing condensates, and others regulate assembly without being stable constituents themselves [19,20]. Lumping all LLPS-associated proteins into a single category obscures their distinct roles in condensate assembly, partitioning, and regulation. Therefore, repositories of these proteins have evolved into sophisticated databases annotating proteins with precisely defined LLPS-related functional classes. The growing standardization of terminology and improved interoperability have increased our understanding of LLPS mechanisms and the performance of LLPS-prediction models. Here, we review the main databases of LLPS proteins and condensates, focusing on the biological meaning of their entries, the evidence they encode, and their utility for structural and computational studies.

Defining liquid–liquid phase separation protein roles: Controlled vocabulary and experimental context

Because proteins can play distinct roles in condensates, and because the same protein may adopt different roles across systems and conditions, a controlled vocabulary is essential for LLPS databases. However, different databases have distinct definitions and nomenclatures for proteins involved in LLPS, including drivers, clients, and regulators (Table 1). Broadly, driver proteins form condensates on their own or with other partners, such as other proteins or nucleic acids (co-drivers). By contrast, clients partition into pre-existing condensates without being required for their formation. The only proteins that are not physically found in condensates are regulators (which can also be nucleic acids or small molecules), which instead modulate their assembly, composition, or material properties.

At the same time, these categories are not mutually exclusive. A given protein may behave as a driver *in vitro*, as a client in a different cellular condensate, or as a regulator through enzyme-mediated modification or changes in its interaction partners. For this reason, databases should not only assign a role, but also record the experimental context in which that role was inferred. This context should also include assay type, molecular partners, concentration range, buffer composition, pH, salt, crowding agents, temperature, nucleic-acid identity, and whether the evidence derives from purified reconstitution, cell-based perturbation, or localization alone.

Considering that *in vitro* droplet formation can be induced by sufficiently strong, non-physiological, experimental perturbations, the mere observation of droplets does not necessarily establish that a protein undergoes biologically relevant phase separation in cells. In LLPS databases, the evidence required to support a specific protein role may range from *in silico* predictions to *in vivo* functional validation. Consequently, the same protein may be assigned a driver, client, scaffold, or regulatory role, depending not only on its molecular behavior but also on the assay type, experimental conditions, and interpretative criteria adopted by a given resource. In this review, we therefore examine the main repositories of LLPS proteins and condensates, with particular attention to how they assign LLPS-related roles and to the types of evidence that underpin these annotations.

The current database landscape

The rapid expansion of the LLPS field, particularly following the seminal 2009 study by Brangwynne et al. [21], which established phase separation as a mechanism underlying membraneless organelles, led to a steady increase in the number of proteins reported to be associated with biomolecular condensates.

This growth created a clear need for dedicated computational resources able to collect these proteins and their associated annotations. Most of the first generation of LLPS databases appeared around 2020 and were built based on different premises, with distinct curation strategies, evidence criteria, and annotation schemes [22]. Therefore, a number of LLPS databases with different contents were released over the last few years, trying to address distinct biological questions (Table 2). Some databases primarily catalog proteins reported to undergo phase separation [23,24], whereas others place greater emphasis on experimental evidence [25], protein roles [26], or condensate association [27]. Although these resources have greatly accelerated access to LLPS-related information, their heterogeneity and limited interoperability have also hampered database comparisons. More recent efforts aim to reconcile these differences by standardizing terminology, harmonizing annotations and evidence types, and integrating data across varied sources [28,29]. In the following sections, we summarize the main LLPS databases, the types of

Table 1

List of synonyms and definitions of LLPS participants provided by LLPS databases. Please note that definitions do not always imply that the databases collect these types of proteins (they might simply describe them). To check the curated content of each resource, please check Table 2.

Database	Driver definition	Co-driver definition	Client definition	Regulator definition
PhaSePro	Scaffold (one or a few proteins that drive condensate formation)	Scaffold (one or a few proteins that drive condensate formation)	Client (protein that 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.)	_
PhaSepDB	PS-self (the protein forms condensates *in vitro* without requiring additional biomolecules)	PS-other (evidence does not demonstrate that the protein separates alone *in vitro*, or only cellular data are provided)	_	_
LLPSDB	Main components (the type of main components in each entry, which is recorded in the form of 'Protein(n) + RNA' or 'Protein(n) + DNA' or 'Protein(n)', where the 'n' indicates the number of proteins)	Main components (the type of main components in each entry, which is recorded in the form of 'Protein(n) + RNA' or 'Protein(n) + DNA' or 'Protein(n)', where the 'n' indicates the number of proteins)	_	_
DrLLPS	Scaffold (driver of LLPS essential for the structural integrity of MLOs, and the major components which, alone or with co-scaffolds, undergo LLPS)	Scaffold (drivers of LLPS essential for the structural integrity of MLOs, and the major components which, alone or with co-scaffolds, undergo LLPS)	Client (proteins identified to be co-complexed with known MLO scaffolds by conventional biochemical assays or mass spectrometry, or co-localized with MLOs by immunofluorescence)	Regulator (the LLPS of scaffold proteins and the stability of MLOs can be modulated by PTMs and other proteins. None of these proteins have been characterized to undergo LLPS, and here we classified all these proteins as MLO regulators.)
CD-CODE	Driver (protein which fulfill at least one of the following criteria: Induce the formation of a condensate. Are essential for the integrity of a condensate.)	Driver (protein which fulfill at least one of the following criteria: Induce the formation of a condensate. Are essential for the integrity of a condensate.)	Member (proteins which is part of a condensate but their role is unknown. It is important to note that a particular protein may be a driver in one condensate and a member in another one)	_
BAV-LLPS	LLPS protein (protein undergoing LLPS)	_	_	_
MLOsMetaDB	Driver (protein or nucleic acid that play a direct role in the formation and/or the stability of the MLOs through the LLPS process)	Driver (protein or nucleic acids that play a direct role in the formation and/or the stability of the MLOs through the LLPS process)	Client (component that do not actively participate in the formation but are present in the MLO under specific conditions)	Regulator (protein that can alter the formation, composition, dissolution, or stability of MLOs)
LLPS-datasets	Driver (protein can undergo LLPS on their own, without any partner-either protein, DNA, or RNA)	Co-driver (protein that can undergo LLPS but require a partner-either protein, DNA, or RNA)	Client (protein recruited into pre-existing condensates and is not essential for their integrity)	Regulator (protein that influence the behavior of drivers and clients, but they are not physically part of condensates)

LLPS, liquid–liquid phase separation; PS, phase separation.

Table 2

Summary of LLPS databases, their curated content and availability (May 2026).

Database	Year (first; last)	Type	Protein types	N proteins (drivers, clients, regulators)	N condensates	Nucleic acids	c-mods (condensate modulators)	Disease-associations	Mutations	Link	Reference (last)
PhaSePro	2020	Original	Drivers	121	61 (*in vitro*/*in vivo*)	Yes	No	Yes	Yes	https://phasepro.elte.hu/	[24]
PhaSepDB	2020; 2025	Original	Drivers	1873	194 (*in vitro*/*in vivo*)	Yes	No	Yes	Yes	https://db.phasep.pro/	[33]
LLPSDB	2020; 2022	Original	Drivers	586	0	Yes	No	No	Yes	http://bio-comp.org.cn/llpsdbv2/home.html	[34]
DrLLPS	2020	Original	Drivers, clients, regulators	9285[a]	40 (*in vitro*/*in vivo*)	Yes	No	Yes	Yes	https://llps.biocuckoo.cn/	[26]
CD-CODE	2023; 2025	Original	Drivers, clients	11133	798 (*in vitro*/*in cellulo*/*in vivo*)	Yes	Yes	Yes	No	https://cd-code.org/	[35]
BAV-LLPS	2025	Original	Drivers	94[a]	14 (*in vitro*/*in vivo*)	No	No	No	No	https://bav-llps-db.bioinformatica.org/	[36]
MLOsMetaDB	2024	Unified	Drivers, clients, regulators	12038	158 (*in vitro*/*in vivo*)	No	No	No	No	http://mlos.leloir.org.ar	[28]
LLPS-datasets	2025	Unified	Drivers, clients, negative proteins	755	0	No	No	No	No	https://llpsdatasets.ppmclab.com/	[29]

[a] Number of manually-curated LLPS proteins. These resources contain a larger number derived from homology. LLPS, liquid–liquid phase separation.

information they contain, and ongoing efforts to unify the field.

Protein-centric resources

PhaSePro is one of the earliest and most stringent LLPS resources [24]. It remains particularly valuable because it was built around explicit manual curation and a controlled vocabulary for proteins that drive phase separation. Its primary focus is on experimentally supported driver proteins and on the sequence regions linked to that behavior, making it especially useful for mechanistic studies and for developing predictors aimed at identifying scaffold regions or proteins rather than generic condensate-associated proteins [30,31]. PhaSePro collects manually curated information on driver proteins from different organisms, their associated biomolecular condensates, and the *in vitro* and *in vivo* evidence supporting LLPS. Although the database is centered on drivers, it also records relevant dependencies on their molecular partners, including additional proteins, RNA, and post-translational modifications, which helps capture the context in which phase separation occurs. The main strength of PhaSePro is therefore its conceptual clarity and curation rigor. Its main limitations are its more restricted coverage than broader resources and less consistent updates.

PhaSepDB adopts a broader view, including proteins that phase-separate on their own and those supported in more complex contexts [23,32,33]. In its current organization, entries are divided into proteins that undergo phase separation alone *in vitro* (PS-self) and proteins that require partners or have only been observed *in vivo* (PS-other) [33]. This scope makes PhaSepDB particularly useful to distinguish the ability of a certain protein to undergo LLPS. Importantly, 'PS-other' and 'PS-self' should not be viewed as mutually exclusive categories. A protein may phase-separate independently under specific conditions, but require molecular partners in other contexts, for instance, to assemble a distinct condensate. A major strength of the database is the richness of its annotations, which include associated biomolecular condensates, experimental assays, material states, intrinsic determinants, molecular partners, regulatory features, and functional implications. Since its first release in 2020 [23], PhaSepDB has undergone substantial development, with major updates in 2022 that expanded annotations on droplet states (liquid, hydrogel or solid nature of condensates), co-phase separation partners, and experimental details [32], and a more recent 2025 release, PhaSepDB 3.0, that further increased coverage, introduced artificial intelligence (AI)-assisted curation and additional annotations such as functional relevance, experimental evidence, and the intrinsic and extrinsic regulations of phase separation [33].

DrLLPS explicitly classifies eukaryotic proteins according to their proposed roles in condensates, distinguishing scaffolds or drivers, clients, and regulators [26]. This role-aware design is particularly useful because regulators are often underrepresented in other databases. In addition to experimentally supported proteins, DrLLPS includes unreviewed predictions and ortholog-based transfers across eukaryotes, thereby expanding coverage beyond the curated literature. Each entry is richly annotated with information on intrinsically disordered regions, domains, post-translational modifications, mutations, disease associations, physicochemical properties, and condensate localization, making the resource valuable for exploratory analyses that connect LLPS to sequence features, function, and disease. Since its initial release in 2020, no major updates have been reported on the website.

LLPS-DB focuses on proteins and macromolecular systems shown to undergo LLPS *in vitro*, including both natural and designed sequences and systems that phase separate either alone or with additional components [25,34]. Its main strength lies in the detailed capture of experimental context; entries are manually curated and record parameters such as salt concentration, pH, crowding agents, incubation time, and detection method, making the database especially valuable for comparing assay conditions and for analyzing how physicochemical environments shape phase behavior. This level of condition-resolved annotation is particularly useful for mechanistic studies and for benchmarking or training predictors against defined *in vitro* evidence. Since its original release in 2020 [25], LLPS-DB has been substantially expanded, with version 2.0 in 2022 doubling the curated content and introducing a class of 'ambiguous systems' to accommodate complex mixtures whose phase-separating behavior is poorly resolved [34]. Its main limitation is that, despite its experimental depth, it provides less biological context than other resources, including limited annotation of condensate type or cellular role. As a result, LLPS-DB is especially suitable for interrogating the biochemical basis of phase separation, but it is less informative when the question concerns condensate identity, physiological relevance, or the protein role within a cellular assembly.

Condensate-centric resources

CD-CODE shifts the focus of LLPS resources from individual proteins to biomolecular condensates [27,35]. This perspective is important because many key questions concern the composition, function, regulation, and disease relevance of a condensate rather than the behavior of a single protein in isolation. In contrast to earlier protein-centered databases, CD-CODE is organized around condensates and their components and includes only experimentally validated entries. It assigns different confidence scores based on the type and strength of supporting evidence, including PubMed-based annotations, high-throughput studies, and

in vitro, *in cellulo*, or *in vivo* observations. A major strength of the resource is its use of a controlled vocabulary, made publicly available through the CD-CODE encyclopedia, which helps standardize annotations and improve consistency across entries. Its crowdsourcing model is also a notable advantage, as it enables dynamic, community-driven addition of and review of data in a field that is evolving quickly for static resources to remain comprehensive. The more recent CD-CODE 2.0 expands this framework by incorporating nucleic acid components, infectious condensates, condensate-modulating compounds, disease-linked biomolecular condensates (condensatopathies or condensate phenotypes associated with disease pathophysiology), and diseases in which condensates are implicated but are not necessarily the primary pathological hallmark [35].

Lineage-specific resources

BAV-LLPS is a database specifically designed to collect LLPS-associated proteins from bacteria, archaea, and viruses [36]. Although bacterial and viral condensates had already been incorporated into previous resources such as CD-CODE [27], this expansion is timely, as increasing evidence points to relevant roles for condensate formation across diverse prokaryotic and viral systems [37,38], which remain far less characterized than their eukaryotic counterparts. To organize this broader landscape, BAV-LLPS is divided into three datasets: manually curated proteins with experimental evidence for LLPS or biomolecular condensate formation, homologous proteins identified through sequence-similarity searches, and additional homologous proteins imported from external LLPS resources, all of which can be independently searched, filtered, and downloaded. Entries are enriched with functional annotations, subcellular localization, taxonomy, associated condensates, structural features, and predicted disorder and LLPS profiles, making the resource useful both for comparative analyses and for exploring underrepresented taxa. At the same time, evidence in these systems is often more heterogeneous than that in better-characterized eukaryotic examples and may sometimes reflect broader condensate-like behavior. For this reason, users should carefully distinguish between entries supported by direct evidence and those based on transferred associations. This distinction is especially relevant for archaeal proteins, as all of their annotations are homology-based, while the dataset with actual experimental evidence contains only bacterial and viral proteins.

Integrative and harmonizing resources

MLOsMetaDB and LLPS-datasets exemplify a second generation of LLPS database development, centered on integration, reconciliation, and benchmarking rather than primary literature collection alone [28,29]. MLOsMetaDB aims to centralize information on LLPS proteins and membraneless organelles by aggregating proteins assigned to well-defined functional roles, including drivers, clients, and regulators, together with their associated biomolecular condensates [28]. Each entry is cross-linked to the original LLPS resources, making the database particularly useful for navigating a comprehensive annotation landscape. In addition, MLOsMetaDB enriches entries with molecular features such as intrinsic disorder, low-complexity regions, domains, LLPS propensity, and ortholog information, thereby adding structural and evolutionary context.

LLPS-datasets takes a more selective and computationally oriented approach [29]. Rather than serving primarily as a browsing resource, it assembles curated datasets of drivers, clients, and non-LLPS proteins through a stringent biocuration protocol designed to reconcile information from multiple LLPS databases. Its main contribution is the generation of standardized positive classes together with operational negative datasets comprising candidates that should not undergo LLPS, including both structured and disordered proteins not currently associated with LLPS. The generation of these negative datasets is based on stringent filtering criteria that include i) their absence from any of the original LLPS source databases, ii) no previous associations with condensates, and iii) the lack of LLPS interactors [29]. Importantly, by including both globular and disordered proteins as separate negative subsets, this design reduces the risk that machine-learning models trained on these data simply equate intrinsic disorder with LLPS propensity. Each entry is accompanied by annotations, including the fraction of ordered and disordered residues, localization, predicted prion-like domains, LLPS specificity, and overlap with source databases. This design makes LLPS-datasets especially valuable for interoperability, for training multi-label machine-learning models, and for fairer benchmarking of predictive tools. At the same time, its negative classes should be interpreted with caution: in LLPS, true negatives are difficult to define, so these datasets are best understood as candidate non-LLPS proteins rather than certain negatives.

Related resources beyond core liquid–liquid phase separation catalogs

Beyond the core LLPS databases, several adjacent resources provide information that is highly relevant to condensate biology. DisPhaseDB and PhaSeDis focus on disease-associated variants in LLPS proteins and on how sequence changes may alter condensate behavior [39,40]. FuzDB catalogs fuzzy interactions in disordered protein complexes, a class of interactions that often overlaps mechanistically with condensate formation [41]. In parallel, RPS complements the predominantly protein-centered LLPS resources by collecting RNAs involved in phase separation [42]. DisProt, the gold-

standard database of manually curated and experimentally validated intrinsically disordered protein regions [43], incorporates a dedicated thematic dataset of '*condensates-related proteins*'. Moreover, it provides functional annotation specific to LLPS in both the Gene Ontology and Intrinsically Disordered Proteins Ontology (condensate transition and LLPS). Although these related databases are not substitutes for dedicated LLPS repositories, they complement them by capturing additional layers of information, including sequence variation, disorder-mediated binding, disease associations, and RNA specificity. Together, they highlight that a full understanding of condensate biology requires linking core LLPS annotations to broader molecular and functional contexts.

Merits and limitations of liquid–liquid phase separation databases

LLPS databases have been pivotal in storing and organizing proteins reported to undergo LLPS or to associate with biomolecular condensates, and many have also served as training sets for sequence-based LLPS predictors [31,44–47]. At the same time, no single resource captures the protein's role, evidence strength, condensate identity, region-level determinants, and the full experimental context reported for a protein in a condensate. Instead, existing resources typically excel in capturing specific layers of annotation (Table 3). For instance, LLPS-DB provides specific experimental conditions (e.g., pH, ionic strength, or temperature) for *in vitro* systems, whereas CD-CODE distinguishes biomolecular and synthetic condensates and links components to supporting experimental evidence.

Among the early resources, PhaSePro is recognized for introducing an explicit controlled vocabulary and for focusing on experimentally supported drivers, which has made it particularly useful for training models aimed at scaffold prediction. PhaSepDB has a broader scope and has undergone repeated updates, incorporating proteins that phase separate alone as well as proteins supported in more complex contexts. LLPS-DB is distinguished by its detailed annotation of experimental conditions, a feature that has already enabled the development of predictors trained on *in vitro* phase-separation behavior under defined physicochemical settings [48]. DrLLPS is particularly valuable because it explicitly includes not only drivers and clients, but also regulators, a class often neglected elsewhere. More recent databases like CD-CODE shift the focus from proteins to condensates, treating biomolecular condensates as primary biological entities and combining a controlled vocabulary with evidence scoring. BAV-LLPS aims to broaden our understanding of LLPS mechanisms beyond eukaryotic systems by extending curation to bacterial, archaeal, and viral systems.

Inconsistent annotation across resources has driven the emergence of derivative repositories focused on reconciliation and standardization. MLOsMetaDB integrates annotations for drivers, clients, regulators, and condensates and is useful as a navigation layer across a

Table 3

Main merits of LLPS databases. The criteria behind the selection of LLPS databases fulfilling the specific requirements are described below.

Database	Coverage[a]	LLPS diversity[b]	Continuous updates[c]	Extended filtering[d]	Varied exports[e]	Expanded condensates[f]	User-friendly website[g]	API access[h]
PhaSePro				X	X		X	X
PhaSepDB	X		X		X	X	X	
LLPSDB								
DrLLPS	X	X			X	X		
CD-CODE	X	X	X		X	X	X	X
BAV-LLPS			X					
MLOsMetaDB	X	X			X	X	X	X
LLPS-datasets		X		X	X		X	

[a] n LLPS proteins $\geq$ 1000.

[b] Specifically collects clients and/or regulators beyond drivers.

[c] Incorporates new submission option and is updated within the last year.

[d] Allows for multi-column filtering and sorting options.

[e] Allows for customizable exports or diverse file formats.

[f] Includes specific tabs/pages about condensates.

[g] Allows direct browsing, searching and filtering on a single page.

[h] Allows for specific access and documentation to programmatic access. LLPS, liquid–liquid phase separation.

fragmented ecosystem. LLPS-datasets takes a different approach, applying a common biocuration framework to derive confident sets of drivers, clients, and operational negative proteins not currently associated with LLPS, which makes it especially valuable for benchmarking and machine-learning applications. Together, these integrative resources illustrate the field's shift from primary collection toward cross-database harmonization.

Several gaps now define the next stage of database development. The most important is condition-aware metadata. Proteins' condensation goes beyond sequence alone; concentration, stoichiometry, buffer composition, salt, pH, crowding agents, temperature, redox state, RNA identity, post-translational modifications, and incubation history can all alter phase behavior [49–52]. Such variables are key for structural biology, computational simulations, and predictive models. LLPS-DB captures part of this space, but the field still lacks a common, minimal metadata standard that spans all resource types. This topic has been covered by the MIADE guidelines (Minimum Information About Disorder Experiments) [53], which define minimal reporting requirements for experiments on intrinsically disordered protein regions and are implemented in DisProt as a current best practice [43]. Building on the established MIADE guidelines, a similar community consensus on the minimum metadata needed to describe LLPS experiments would be extremely valuable for interpreting phase separation in its proper context. A second major gap is the level of evidence for LLPS. Terms such as (co)localization, partitioning, recruitment, and purified reconstitution are often grouped under the same broad label of LLPS association, although they may reflect distinct biological situations, including condensate driving, client recruitment, regulatory activity, or simple molecular interactions.

The need for standardized contextual annotation is illustrated by the MIADE guidelines, which define minimal reporting requirements for experiments on intrinsically disordered protein regions and are implemented in DisProt as a current best practice [44,54]. An analogous community-agreed framework for LLPS experiments would be highly valuable for interpreting phase separation within its proper experimental context.

Future resources should therefore expose evidence classes explicitly and preserve provenance by linking it to specific experiments, figures, or claims in the source article. AI-assisted manual curation may help scale this effort, but only if original prompts remain transparent [33]. A third gap is structural granularity. Most databases still annotate proteins primarily at the full-length level, whereas the mechanism often depends on specific regions, motifs, folded domains, repeat architectures, or mutation–sensitive interaction surfaces. Stronger integration with databases of disorder, structure, domains, variants, and disease-associated mutations would make LLPS resources far more informative for structural and mechanistic studies [39,40,43].

Conclusions and future perspectives

In this review, we have examined the current landscape of LLPS databases, the types of information they capture, and the strengths and limitations that shape their use. Existing resources collect complementary but not equivalent classes of data, covering drivers, co-drivers, clients, regulators, condensates, and, in some cases, experimental conditions or related molecular features. Their value is therefore substantial, but so is their heterogeneity. Differences in controlled vocabularies, role definitions, and evidence models still limit interoperability and complicate comprehensive integration across resources. Therefore, the field is shifting towards interoperable, versioned, and evidence-aware databases that distinguish not only protein's role in condensates but also preserve the experimental context.

A central requirement for next-generation LLPS resources the standardized, machine-readable annotation of experimental context. Because phase separation depends strongly on assay conditions [49], entries for *in vitro* assays or synthetic condensates should capture variables such as concentration, stoichiometry, buffer composition, salt, pH, crowding agents, temperature, redox state, nucleic-acid identity, post-translational modifications, and incubation history. These are not accessory metadata; they are relevant for structural biology, simulations, and predictive tools. At present, such information is often buried in source papers, incompletely reported, or encoded inconsistently, making cross-database comparisons difficult and introducing uncertainty into downstream analyses. For *in cellulo* condensates, the relevant context differs and should include the cellular model, endogenous or ectopic expression, construct design and tags, mutations or perturbations, treatment conditions, assay timing, subcellular localization, and the type of evidence supporting condensate association. Recording these parameters would help distinguish drivers, clients, regulators, and simple interactors, improving the reproducibility and interpretation of LLPS annotations. Consensus guidelines for annotating both evidence classes and experimental conditions would therefore be highly valuable. They would enable predictive models and molecular simulations that learn not only from sequence features associated with LLPS, but also from the physicochemical contexts in which phase separation occurs [54]. Human-validated, AI-assisted curation may help make this level of annotation scalable [33].

Future databases should also support fairer benchmarking and more mechanistically informative modeling. This will require curated positive classes together with carefully defined operational negative datasets for proteins not currently associated with LLPS [29]. Such negatives should include both structured and intrinsically disordered proteins, as well as low-complexity and prion-like sequences, so that predictive performance is not overlapped by disorder-related features. In parallel, the identification, characterization, and annotation of disease-associated conformational transitions of LLPS-prone regions would make LLPS resources more informative for biomedical applications. In practical terms, next-generation databases should combine controlled vocabularies for protein roles, explicit evidence classes, standardized condition metadata, links to region-level structural and mutational information, and stable export formats or application programming interface options for downstream analysis. If the field converges on such shared standards, LLPS databases could become long-term resources for the community, in a manner analogous to what the Protein Data Bank did for structural coordinates [55] and DisProt has done for intrinsic disorder [43].

Funding sources

CP-G was supported by the Secretariat of Universities and Research of the Catalan Government, the European Social Fund (2023 FI_3 00018), and by the UAB (2025PILIFRUA37). MB was supported by the Maria Zambrano grant funded by the European Union-NextGenerationEU. SV was supported by the Spanish Ministry of Science and Innovation (PID2022-137963OB-I00), IICREA-Academia 2020, Academia d'exel·lencia 2025 (AGAUR), and 2021-SGR-00635 AGAUR, Generalitat de Catalunya) and CERCA Programme (Generalitat de Catalunya). This publication is based upon work from COST Action ML4NGP, CA21160, supported by COST (European Cooperation in Science and Technology).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Salvador Ventura reports financial support was provided by Autonomous University of Barcelona. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Editorial disclosure statement

Given the role as Guest Editor, Salvador Ventura had no involvement in the peer review of the article and has no access to information regarding its peer-review. Full responsibility for the editorial process of this article was delegated to Francesco Aprile.

Data availability

No data was used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Feig M, Yu I, Wang PH, Nawrocki G, Sugita Y: **Crowding in cellular environments at an atomistic level from computer simulations**. *J Phys Chem B* 2017, **121**:8009–8025.
2. Gnutt D, Ebbinghaus S: **The macromolecular crowding effect–from in vitro into the cell**. *Biol Chem* 2016, **397**:37–44.
3. Banani SF, Lee HO, Hyman AA, Rosen MK: **Biomolecular condensates: organizers of cellular biochemistry**. *Nat Rev Mol Cell Biol* 2017, **18**:285–298.
4. Wheeler RJ, Hyman AA: **Controlling compartmentalization by non-membrane-bound organelles**. *Philos Trans R Soc Lond B Biol Sci* 2018, **373**.
5. Chin Sang C, Upadhyay S, Nosella ML, Forman-Kay JD, Lee HO: **The dynamic and heterogeneous composition of biomolecular condensates and its functional relevance**. *Nat Rev Mol Cell Biol* 2026, **27**:278–296.
6. Dumelie JG, Chen Q, Miller D, Attarwala N, Gross SS, Jaffrey SR: **Biomolecular condensates create phospholipid-enriched microenvironments**. *Nat Chem Biol* 2024, **20**:302–313.
7. Wang M, Song X, Chen J, Chen X, Zhang X, Yang Y, Liu Z, Yao L: **Intracellular environment can change protein conformational dynamics in cells through weak interactions**. *Sci Adv* 2023, **9**:eadg9141.
8. Pintado-Grima C, Iglesias V, Santos J, Uversky VN, Ventura S: **DispHScan: a multi-sequence web tool for predicting protein disorder as a function of pH**. *Biomolecules* 2021, **11**.
9. Martin EW, Holehouse AS: **Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof**. *Emerg Top Life Sci* 2020, **4**:307–329.
10. Choi JM, Holehouse AS, Pappu RV: **Physical principles underlying the complex biology of intracellular phase transitions**. *Annu Rev Biophys* 2020, **49**:107–133.
11. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, *et al.*: **A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins**. *Cell* 2018, **174**:688–699.e616.
12. Boncella AE, Shattuck JE, Cascarina SM, Paul KR, Baer MH, Fomicheva A, Lamb AK, Ross ED: **Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules**. *Proc Natl Acad Sci U S A* 2020, **117**:5826–5835.
13. Iglesias V, Santos J, Santos-Suárez J, Pintado-Grima C, Ventura S: **SGnn: a web server for the prediction of prion-like domains recruitment to stress granules upon heat stress**. *Front Mol Biosci* 2021, **8**, 718301.
14. Vendruscolo M, Fuxreiter M: **Protein condensation diseases: therapeutic opportunities**. *Nat Commun* 2022, **13**:5550.
15. Portz B, Lee BL, Shorter J: **FUS and TDP-43 phases in health and disease**. *Trends Biochem Sci* 2021, **46**:550–563.
16. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, *et al.*: **Phase transitions**

in the assembly of multivalent signalling proteins. *Nature* 2012, **483**:336–340.

17. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoynov S, Mahamid J, Saha S, Franzmann TM, *et al*.: **A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation**. *Cell* 2015, **162**:1066–1077.

18. Su X, Ditlev JA, Hui E, Xing W, Banjade S, Okrut J, King DS, Taunton J, Rosen MK, Vale RD: **Phase separation of signaling molecules promotes T cell receptor signal transduction**. *Science* 2016, **352**:595–599.

19. Ganser LR, Myong S: **Methods to study phase-separated condensates and the underlying molecular interactions**. *Trends Biochem Sci* 2020, **45**:1004–1005.

20. Mitrea DM, Chandra B, Ferrolino MC, Gibbs EB, Tolbert M, White MR, Kriwacki RW: **Methods for physical characterization of phase-separated bodies and membrane-less organelles**. *J Mol Biol* 2018, **430**:4773–4805.

21. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA: **Germline P granules are liquid droplets that localize by controlled dissolution/condensation**. *Science* 2009, **324**:1729–1732.

22. Farahi N, Lazar T, Wodak SJ, Tompa P, Pancsa R: **Integration of data from liquid-liquid phase separation databases highlights concentration and dosage sensitivity of LLPS drivers**. *Int J Mol Sci* 2021, **22**.

23. You K, Huang Q, Yu C, Shen B, Sevilla C, Shi M, Hermjakob H, Chen Y, Li T: **PhaSepDB: a database of liquid-liquid phase separation related proteins**. *Nucleic Acids Res* 2020, **48**: D354–D359.

24. ** Mészáros B, Erdős G, Szabó B, Schád É, Tantos Á, Abukhairan R, Horváth T, Murvai N, Kovács OP, Kovács M, *et al*.: **PhaSePro: the database of proteins driving liquid-liquid phase separation**. *Nucleic Acids Res* 2020, **48**: D360–D367.
This work presents one of the first databases of LLPS proteins. In particular, the authors focus on the collection of proteins driving LLPS. The database provides information about condenstaes, evidence, dependencies, and modifications influencing the process.

25. Li Q, Peng X, Li Y, Tang W, Zhu J, Huang J, Qi Y, Zhang Z: **LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro**. *Nucleic Acids Res* 2020, **48**: D320–D327.

26. Ning W, Guo Y, Lin S, Mei B, Wu Y, Jiang P, Tan X, Zhang W, Chen G, Peng D, *et al*.: **DrLLPS: a data resource of liquid-liquid phase separation in eukaryotes**. *Nucleic Acids Res* 2020, **48**:D288–D295.

27. Rostam N, Ghosh S, Chow CFW, Hadarovich A, Landerer C, Ghosh R, Moon H, Hersemann L, Mitrea DM, Klein IA, *et al*.: **CD-CODE: crowdsourcing condensate database and encyclopedia**. *Nat Methods* 2023, **20**:673–676.

28. * Orti F, Fernández ML, Marino-Buslje C: **MLOsMetaDB, a meta-database to centralize the information on liquid-liquid phase separation proteins and membraneless organelles**. *Protein Sci* 2024, **33**, e4858.
This work creates the first centralized LLPS database that integrates data from distinct resources. It specifies the role of each protein in LLPS, provides condensate annotations and evolutionary information with orthologs.

29. ** Pintado-Grima C, Bárcenas O, Arribas-Ruiz E, Iglesias V, Burdukiewicz M, Ventura S: **Comprehensive protein datasets and benchmarking for liquid-liquid phase separation studies**. *Genome Biol* 2025, **26**:198.
This work presents an integrated biocuration protocol to obtain confident drivers and clients involved in LLPS as well as negative datasets of both disordered and globular proteins. The authors find relevant physicochemical differences between the datasets and provide the most comprehensive benchmark of LLPS predictors to date.

30. Monti M, Fiorentino J, Miltiadis-Vrachnos D, Bini G, Cotrufo T, Sanchez de Groot N, Armaos A, Tartaglia GG: **catGRANULE 2.0: accurate predictions of liquid-liquid phase separating proteins at single amino acid resolution**. *Genome Biol* 2025, **26**:33.

31. Sun J, Qu J, Zhao C, Zhang X, Liu X, Wang J, Wei C, Wang M, Zeng P, Tang X, *et al*.: **Precise prediction of phase-separation key residues by machine learning**. *Nat Commun* 2024, **15**: 2662.

32. Hou C, Wang X, Xie H, Chen T, Zhu P, Xu X, You K, Li T: **PhaSepDB in 2022: annotating phase separation-related proteins with droplet states, co-phase separation partners and other experimental information**. *Nucleic Acids Res* 2023, **51**:D460–D465.

33. * You K, Li R, Lian R, Li Y, Yang H, Zhou Y, Chen Y, Wang L, Fan Z, Ma L, *et al*.: **PhaSepDB 3.0: a comprehensive knowledgebase of phase separation-related proteins from AI-assisted curation**. *Nucleic Acids Res* 2026, **54**:D445–D450.
This work introduces the latest version of the PhaSepDB database, with a notable expansion of the resource, enriched annotations and an enhanced, innovative AI-assisted curation workflow.

34. Wang X, Zhou X, Yan Q, Liao S, Tang W, Xu P, Gao Y, Li Q, Dou Z, Yang W, *et al*.: **LLPSDB v2.0: an updated database of proteins undergoing liquid-liquid phase separation in vitro**. *Bioinformatics* 2022, **38**:2010–2014.

35. ** Kuznetsova K, Scheremetjew M, Yin J, Moon H, Vargas DA, Hadarovich A, Lewis N, Hoege C, Chow CFW, Kuster D, *et al*.: **CD-CODE 2.0: an enhanced condensate knowledgebase integrating pathobiology, condensate modulating drugs, and host-pathogen interactions**. *Nucleic Acids Res* 2026, **54**: D375–D382.
This article introduces the second version of CD-CODE, the only LLPS database that shifts the focus towards condensates rather than single proteins. The database offers information about condensate's components and their roles, disease-related condensates, pathogen interactions and modulating compounds.

36. * Rodriguez CB, Tunque Cahui RR, Demitroff N, Hirsh L, Devos DP, Boccaccio G, Parisi G: **BAV-LLPS: a database of bacterial, archaea, and virus liquid-liquid phase separation proteins**. *Bioinformatics* 2025, **41**.
This article presents a dedicated LLPS database that goes beyond eukaryotic organisms and focuses on bacterial, archaeal, and viral LLPS proteins. Proteins are manually curated and the dataset is extended with homologous entries.

37. Azaldegui CA, Vecchiarelli AG, Biteen JS: **The emergence of phase separation as an organizing principle in bacteria**. *Biophys J* 2021, **120**:1123–1138.

38. Lopez N, Camporeale G, Salgueiro M, Borkosky SS, Visentín A, Peralta-Martinez R, Loureiro ME, de Prat-Gay G: **Deconstructing virus condensation**. *PLoS Pathog* 2021, **17**, e1009926.

39. Navarro AM, Orti F, Martínez-Pérez E, Alonso M, Simonetti FL, Iserte JA, Marino-Buslje C: **DisPhaseDB: an integrative database of diseases related variations in liquid-liquid phase separation proteins**. *Comput Struct Biotechnol J* 2022, **20**: 2551–2557.

40. Chen T, Tang G, Li T, Yanghong Z, Hou C, Du Z, You K, Ma L: **PhaSeDis: a manually curated database of phase separation-disease associations and corresponding small molecules**. *Genom Proteom Bioinform* 2025, **23**.

41. Hatos A, Monzon AM, Tosatto SCE, Piovesan D, Fuxreiter M: **FuzDB: a new phase in understanding fuzzy interactions**. *Nucleic Acids Res* 2022, **50**:D509–D517.

42. * He Y, Bao X, Chen T, Jiang Q, Zhang L, He LN, Zheng J, Zhao A, Ren J, Zuo Z: **Rps 2.0: an updated database of RNAs involved in liquid-liquid phase separation**. *Nucleic Acids Res* 2025, **53**: D299–D309.
This article introduces the second version of the RPS database, which, in contrast to most resources, collects RNAs involved in LLPS instead of proteins. The new release includes several new RNAs and offers extensive annotations, including interactions with other biomolecules.

43. Nugnes MV, Bouhraoua KEA, Zoubiri M, Pancsa R, Fichó E, Tompa P, Piovesan D, Tosatto SCE, Aspromonte MC, Consortium D: **DisProt in 2026: enhancing intrinsically**

disordered proteins accessibility, deposition, and annotation. *Nucleic Acids Res* 2026, **54**:D383–D392.

44. Chen Z, Hou C, Wang L, Yu C, Chen T, Shen B, Hou Y, Li P, Li T: **Screening membraneless organelle participants with machine-learning models that integrate multimodal features**. *Proc Natl Acad Sci U S A* 2022, **119**, e2115369119.

45. Hadarovich A, Scheremetjew M, Singh HR, Moon H, Hersemann L, Toth-Petroczy A: **PICNIC web server for predicting proteins involved in biomolecular condensates**. *Bioinformatics* 2026, **42**.

46. Kilgore HR, Chinn I, Mikhael PG, Mitnikov I, Van Dongen C, Zylberberg G, Afeyan L, Banani SF, Wilson-Hawken S, Lee TI, *et al.*: **Protein codes promote selective subcellular compartmentalization**. *Science* 2025, **387**:1095–1101.

47. Saar KL, Morgunov AS, Qi R, Arter WE, Krainer G, Lee AA, Knowles TPJ: **Learning the molecular grammar of protein condensates from sequence determinants and embeddings**. *Proc Natl Acad Sci U S A* 2021:118.

48. Raimondi D, Orlando G, Michiels E, Pakravan D, Bratek-Skicki A, Van Den Bosch L, Moreau Y, Rousseau F, Schymkowitz J: **In silico prediction of in vitro protein liquid-liquid phase separation experiments outcomes with multi-head neural attention**. *Bioinformatics* 2021, **37**:3473–3479.

49. Pintado-Grima C, Bárcenas O, Ventura S: **In-Silico analysis of pH-Dependent liquid-liquid phase separation in intrinsically disordered proteins**. *Biomolecules* 2022, **12**.

50. Franzmann TM, Jahnel M, Pozniakovsky A, Mahamid J, Holehouse AS, Nüske E, Richter D, Baumeister W, Grill SW, Pappu RV, *et al.*: **Phase separation of a yeast prion protein promotes cellular fitness**. *Science* 2018, **359**.

51. Hofweber M, Dormann D: **Friend or foe-post-translational modifications as regulators of phase separation and RNP granule dynamics**. *J Biol Chem* 2019, **294**:7137–7150.

52. André AAM, Yewdall NA, Spruijt E: **Crowding-induced phase separation and gelling by co-condensation of PEG in NPM1-rRNA condensates**. *Biophys J* 2023, **122**:397–407.

53. ** Mészáros B, Hatos A, Palopoli N, Quaglia F, Salladini E, Van Roey K, Arthanari H, Dosztányi Z, Felli IC, Fischer PD, *et al.*: **Minimum information guidelines for experiments structurally characterizing intrinsically disordered protein regions**. *Nat Methods* 2023, **20**:1291–1303.
This work covers the minimum information required about disorder experiments (MIADE) to fully characterize intrinsically disordered proteins in their context. It highlights community-driven consensus efforts to generate guidelines for the annotation of experimental conditions that should also apply to condensate assays.

54. Rauh AS, Tesei G, Lindorff-Larsen K: **A coarse-grained model for disordered proteins under crowded conditions**. *Protein Sci* 2025, **34**, e70232.

55. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The protein data bank**. *Nucleic Acids Res* 2000, **28**:235–242.