

## Databases and ontologies

# BAV-LLPS: a database of bacterial, archaea, and virus liquid–liquid phase separation proteins

Cecilia B. Rodriguez<sup>1,†</sup>, Ronaldo Romario Tunque Cahui<sup>2,†</sup>, Nicolas Demitroff<sup>1,†</sup>, Layla Hirsh<sup>2</sup>, Damien P. Devos<sup>3</sup>, Graciela Boccaccio<sup>4,5,\*</sup>, Gustavo Parisi<sup>1,\*</sup>

<sup>1</sup>Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal B1876BXD, Argentina

<sup>2</sup>Department of Engineering, Faculty of Science and Engineering, Pontifical Catholic University of Peru, Lima 15088, Peru

<sup>3</sup>Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019—UMR 9017—CIII—Center for Infection and Immunity of Lille, Lille F-59000, France

<sup>4</sup>Fundación Instituto Leloir (FIL)—Instituto de Investigaciones Bioquímicas de Buenos Aires (IIBBA)—Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires C1405BWE, Argentina

<sup>5</sup>Department of Molecular and Cellular Biology and Physiology (FBMyC), Facultad de Ciencias Exactas y Naturales (FCEN), University of Buenos Aires (UBA), Buenos Aires C1428EHA, Argentina

\*Corresponding authors. Gustavo Parisi, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal B1876BXD, Argentina. E-mail: gusparisi@gmail.com; Graciela Boccaccio, Fundación Instituto Leloir (FIL)—Instituto de Investigaciones Bioquímicas de Buenos Aires (IIBBA)—Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires C1405BWE, Argentina. E-mail: gboccaccio@leloir.org.ar

†= equal contribution.

Associate Editor: Inanc Biroi

## Abstract

**Motivation:** Liquid–liquid phase separation (LLPS) is a key process underlying the formation of biomolecular condensates, such as membrane-less organelles, that compartmentalize biochemical processes inside the cells. While LLPS has been extensively studied in eukaryotes, its role in bacteria, archaea, and viruses remains far less characterized. Recent studies in bacteria have revealed that LLPS-driven condensates play critical roles in RNA processing, stress response, and pathogenicity. Similarly, many viruses exploit LLPS to facilitate crucial steps in their infection cycles, including viral entry, genome replication, assembly, and host immune evasion.

**Results:** In this work, we introduce a hand-curated database of LLPS proteins from bacteria, archaea, and viruses (BAV-LLPS Database). This resource, extended through sequence similarity searches, comprises over 5000 proteins and integrates diverse data including biological annotations, sequence features, predicted disordered regions, LLPS per site probability, and AlphaFold2-based structural models. Additionally, our web server enables users to explore both the curated and homologous derived datasets, providing a platform to uncover evolutionary relationships and intrinsic and differential properties of LLPS proteins across various taxonomic groups. This work seeks to deepen our understanding of LLPS mechanisms beyond eukaryotic organisms, emphasizing their significance across diverse life forms. It also aims to foster the development of specialized predictive tools that will facilitate the exploration and characterization of LLPS processes in a wide array of living organisms, thereby contributing to advancements in both fundamental biological research and applied biomedical sciences.

**Availability and implementation:** BAV-LLPS DB is freely accessible at <https://bav-llps-db.bioinformatica.org/>. The data can be retrieved from the website. The source code of the database can be downloaded from <https://bav-llps-db.bioinformatica.org/download>.

## 1 Introduction

Liquid–liquid phase separation (LLPS) involves the transition from an homogeneous system containing given macromolecular components into two separated phases, one enriched and another depleted in the involved macromolecules (Alberti and Dormann 2019). LLPS is greatly influenced by intrinsically disordered protein regions (IDRs) (Tesei *et al.* 2021, Fonin *et al.* 2022) and generates the so-called biomolecular condensates (BMCs) or membrane-less organelles (MLOs) (Darling and Uversky 2023, Fernández-Alvarez *et al.* 2023). Unlike other eukaryotic organelles such as mitochondria or Golgi apparatus, BMCs and MLOs lack a delimiting lipid membrane. Due to their reversible and dynamic formation, BMCs and MLOs are transient microcompartments that can selectively concentrate certain macromolecules while excluding others. In doing

so, MLOs generate microenvironments that can accelerate or inhibit biochemical reactions, they can buffer protein and nucleic acid concentrations, participate in the sensing of environmental changes, control signaling pathways, and in the nucleation of cytoskeletal structures among other functions. Dozens of BMCs and MLOs have been identified in mammalian cells, a list which is continually expanding (Darling and Uversky 2023, Fernández-Alvarez *et al.* 2023, Thomas *et al.* 2025). Proteins undergoing LLPS and their derived condensates have been found in different organisms across eukaryotes, prokaryotes, and viruses reflecting an evolutionary conservation of these mechanisms to control metabolic processes using spatial and temporal compartmentalization (Rangachari 2023). In addition to their broad taxonomic distribution, LLPS-related proteins in bacteria, archaea, and viruses are significantly less characterized than their eukaryotic counterparts.

In fact, established databases dedicated to LLPS proteins [i.e. PhaSepDB (You *et al.* 2020), LLPSDB (Li *et al.* 2020) and MLOsMetaDB (Orti *et al.* 2024)] are predominantly populated with eukaryotic examples.

In contrast to the well characterized eukaryotic MLOs and BMCs, interest in prokaryotic BMCs has risen abruptly in recent years. The use of *in vitro* reconstitution, fluoroscopic assays, differential interference contrast microscopy and super-resolution microscopy applied to *in vivo* studies, have significantly advanced the detection of proteins associated with LLPS in bacteria, facilitating the exploration of the importance of BMCs in prokaryotes (Azaldegui *et al.* 2021). One of the systems better characterized in prokaryotic condensates are the bacterial ribonucleoprotein bodies (BR-bodies), which control mRNA decay by locally concentrating the RNA degradosome proteins and their substrates (Al-Husini *et al.* 2018, Al-Husini *et al.* 2020). Different proteins participating in LLPS in bacteria were also related to different biological processes involved with the pathogenicity, resistance to antibiotics or antimicrobial agents and bacterial survival (Liu *et al.* 2024). In the human commensal *Bacteroides thetaiotaomicron*, the activity of the transcription termination factor Rho largely depends on its LLPS, with consequences in gene expression and bacteroid competence (Kryptou *et al.* 2023). Another well characterized bacterial condensate involves the proteins FtsZ and SlmA which are part of the divisome complex (Mahone and Goley 2020). A particular bacterial condensate termed aggresome is formed by a diverse collection of endogenous proteins under different stress conditions. They have been associated with the bacterial fitness (Jin *et al.* 2021), protection during desiccation (Kuczyńska-Wiśnik *et al.* 2023), cell dormancy and regrowth regulation, and with drug tolerance (Pu *et al.* 2019). In particular, regarding LLPS in Archaea, although some proteins, such as the thermostable nucleoid protein Cren7 (Banerjee *et al.* 2024), have been proposed to form liquid droplets, further studies are required.

Similarly, many viruses harness phase separation processes to accomplish nearly every step in their infection cycles—including entry, immune evasion, genome replication, new particle assembly, and particle release. As a result, LLPS is now widely recognized as a critical mechanism that facilitates the compartmentalization of viral replication within host cells. While most research has focused on the role of BMCs in replication processes (Lopez *et al.* 2021), fewer studies have examined how viral proteins disrupt host functions (Boccaccio *et al.* 2023). Furthermore, an emerging field in the role of LLPS in viruses is related with proteins adopting condensed states such as LLPS and amyloid fibrils and their roles in pathogenesis and antiviral therapeutic strategies (Gondelaud *et al.* 2023).

Most viral proteins undergoing LLPS are associated with the so-called viral factories (VF), also known as inclusion bodies, replication organelles, viral replication compartments, transcription–replication complexes, virosomes, or viroplasms (Lopez *et al.* 2021). These VF are mostly cytoplasmic spots which concentrate viral proteins and nucleic acids where viral replication and assembly is in this way optimized. Cytoplasmic location shows some exceptions such as those members of the Orthomyxoviridae and Bunyvirales order which establish replication organelles in the nucleus or in association with the Golgi complex. These VF not only serve as crucial centers for optimizing viral replication—through selective uptake or exclusion of host cellular components—but also contribute to immune evasion and escape antiviral defense pathways by trapping host proteins involved in innate immune responses. This process has been described for

several viruses including Nipah virus, respiratory syncytial virus, measles virus, and SARS-CoV-2 (Li *et al.* 2022). Other viral proteins undergoing LLPS have been associated with host cell entry taking control of the host protein disaggregation pathway (Banerjee *et al.* 2014).

In addition, virus infection frequently implicates the inactivation or alteration of cellular condensates to facilitate immune escape. Among other viral components, LLPS proteins are key factors that disrupt the assembly and function of protective BMCs. An important antiviral response is the production of interferon downstream of the assembly of specific BMCs containing immunostimulatory DNA and cyclic GMP-AMP synthase (cGAS), termed cGAS foci. The formation of cGAS foci can be efficiently interrupted by viral LLPS proteins, which phase-separate and sequester activating DNA. Kaposi sarcoma-associated herpesvirus (KSHV) ORF52, also known as KSHV inhibitor of cGAS (KicGAS) and VP22 from several other viruses are emerging examples [reviewed in Boccaccio *et al.* (2023)]. The identification of novel viral LLPS will contribute to understanding viral mechanisms to escape the immune surveillance mediated by cellular BMCs.

In this work, we present a hand-curated database of proteins undergoing LLPS from bacteria, archaea, and viruses. To expand this dataset, we performed sequence similarity searches to identify closely related homologues, resulting in a comprehensive collection of over 5000 proteins—51% from Bacteria, 1% from Archaea, 28% from Viruses, and 20% from Eukaryotes. These datasets are available to download from the web server.

## 2 Dataset construction

BAV-LLPS relies on three hand-curated datasets of LLPS proteins coming from bacteria, archaea and viruses. These sets contain proteins with experimental evidence of LLPS process and/or evidence of MLOs and biocondensate formation. Each entry of this curated dataset offers biological, sequence and structural information for each protein (see below). Each of these curated datasets have been extended using sequence based similarity searches using BLAST to recruit close homologues (e-values < 1.10<sup>-5</sup>). BLAST searches have been restricted to find homologous proteins for each taxonomic group of the curated dataset and in Eukaryotes. Moreover, we gathered together sequences coming from main available LLPS databases [PhaSepDB (You *et al.* 2020), LLPSDB (Wang *et al.* 2022), DrLLPS (Ning *et al.* 2020), PhasePro (Mészáros *et al.* 2020), and CD-Code (Rostam *et al.* 2023)] with different curation levels provided from the corresponding databases. This set of curated LLPS mainly contains proteins from eukaryotic organisms (98%) and was used to further search for homologous proteins using our curated dataset coming from bacteria, archaea and viruses. As disorder content has been reported to be critical to LLPS process (Fonin *et al.* 2022, Tesei *et al.* 2021), we used Aiupred (Erdős and Dosztányi 2024) and AF-RSA (Piovesan *et al.* 2022) to predict disordered regions, Fuzdrop (Hardenberg *et al.* 2020), ParSev2 (Wilson *et al.* 2023) and catGRANULE 2.0 (Monti *et al.* 2025) to predict LLPS per site probability and AlphaFold 2 models from AlphaFold database (<https://alphafold.ebi.ac.uk/>) to predict 3D models of each curated protein. When curated proteins were missing in AlphaFold database we obtain the Alpha Fold 2 model using ColabFold (Mirdita *et al.* 2022).



Nacional de Promoción Científica y Tecnológica [ANPCyT, Argentina; PICT 2018-03190 and PICT 2020-2195 to G.L. B.]; and the European Union [Horizon Europe MSCA SE Grant N.101182949].

## References

- Al-Husini N, Tomares DT, Pfaffengerger ZJ *et al.* BR-bodies provide selectively permeable condensates that stimulate mRNA decay and prevent release of decay intermediates. *Mol Cell* 2020;**78**:670–82.e8.
- Al-Husini N, Tomares DT, Bitar O *et al.*  $\alpha$ -proteobacterial RNA degradosomes assemble liquid–liquid phase-separated RNP bodies. *Mol Cell* 2018;**71**:1027–39.e14.
- Alberti S, Dormann D. Liquid–liquid phase separation in disease. *Annu Rev Genet* 2019;**53**:171–94.
- Azaldegui CA, Vecchiarelli AG, Biteen JS *et al.* The emergence of phase separation as an organizing principle in bacteria. *Biophys J* 2021;**120**:1123–38.
- Banerjee I, Miyake Y, Nobs SP *et al.* Influenza a virus uses the aggressive processing machinery for host cell entry. *Science* 2014;**346**:473–7.
- Banerjee T, Geethika K, Kanbayashi S *et al.* Thermostable nucleoid protein Cren7 slides along DNA and rapidly dissociates from DNA while not inhibiting the sliding of other DNA-binding protein. *J Mol Biol* 2024;**436**:168803.
- Boccaccio GL, Thomas MG, García CC *et al.* Membraneless organelles and condensates orchestrate innate immunity against viruses. *J Mol Biol* 2023;**435**:167976.
- Darling AL, Uversky VN. Known types of membrane-less organelles and biomolecular condensates. In: Uversky VN (ed.), *Droplets of Life*. University of South Florida, Tampa, FL, United States: Elsevier, 2023, 271–335.
- Erdős G, Dosztányi Z. AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Res* 2024;**52**:W176–81.
- Fernández-Alvarez AJ, Thomas MG, Giudice J *et al.* Active regulation mechanisms of LLPS and MLOs biogenesis. In: Uversky VN (ed.), *Droplets of Life*. University of South Florida, Tampa, FL, United States: Elsevier, 2023, 337–73.
- Fonin AV, Antifeeva IA, Kuznetsova IM *et al.* Biological soft matter: intrinsically disordered proteins in liquid–liquid phase separation and biomolecular condensates. *Essays Biochem* 2022;**66**:831–47.
- Gondelaud F, Lozach P-Y, Longhi S *et al.* Viral amyloids: new opportunities for antiviral therapeutic strategies. *Curr Opin Struct Biol* 2023;**83**:102706.
- Hardenberg M, Horvath A, Ambrus V *et al.* Widespread occurrence of the droplet state of proteins in the human proteome. *Proc Natl Acad Sci USA* 2020;**117**:33254–62.
- Jin X, Lee J-E, Schaefer C *et al.* Membraneless organelles formed by liquid–liquid phase separation increase bacterial fitness. *Sci Adv* 2021;**7**:eabh2929.
- Kryptou E, Townsend GE, Gao X *et al.* Bacteria require phase separation for fitness in the mammalian gut. *Science* 2023;**379**:1149–56.
- Kuczyńska-Wiśnik D, Stojowska-Swędryńska K, Laskowska E *et al.* Liquid–liquid phase separation and protective protein aggregates in bacteria. *Molecules* 2023;**28**:6582.
- Liu D, Yang J, Cristea IM *et al.* Liquid–liquid phase separation in innate immunity. *Trends Immunol* 2024;**45**:454–69.
- Li H, Ernst C, Kolonko-Adamska M *et al.* Phase separation in viral infections. *Trends Microbiol* 2022;**30**:1217–31.
- Li Q, Peng X, Li Y *et al.* LLPSDB: a database of proteins undergoing liquid–liquid phase separation in vitro. *Nucleic Acids Res* 2020;**48**:D320–7.
- Lopez N, Camporeale G, Salgueiro M *et al.* Deconstructing virus condensation. *PLoS Pathog* 2021;**17**:e1009926.
- Mahone CR, Goley ED. Bacterial cell division at a glance. *J Cell Sci* 2020;**133**:jcs237057.
- Mészáros B, Erdős G, Szabó B *et al.* PhaSePro: the database of proteins driving liquid–liquid phase separation. *Nucleic Acids Res* 2020;**48**:D360–7.
- Mirdita M, Schütze K, Moriwaki Y *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–82.
- Monti M, Fiorentino J, Miltiadis-Vrachnos D *et al.* catGRANULE 2.0: accurate predictions of liquid–liquid phase separating proteins at single amino acid resolution. *Genome Biol* 2025;**26**:33.
- Ning W, Guo Y, Lin S *et al.* DrLLPS: a data resource of liquid–liquid phase separation in eukaryotes. *Nucleic Acids Res* 2020;**48**:D288–95.
- Orti F, Fernández ML, Marino-Buslje C *et al.* MLOsMetaDB, a meta-database to centralize the information on liquid–liquid phase separation proteins and membraneless organelles. *Protein Sci* 2024;**33**:e4858.
- Piovesan D, Monzon AM, Tosatto SCE *et al.* Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci* 2022;**31**:e4466.
- Pu Y, Li Y, Jin X *et al.* ATP-dependent dynamic protein aggregation regulates bacterial dormancy depth critical for antibiotic tolerance. *Mol Cell* 2019;**73**:143–56.e4.
- Rangachari V. Biomolecular condensates—extant relics or evolving microcompartments? *Commun Biol* 2023;**6**:656.
- Rostam N, Ghosh S, Chow CFW *et al.* CD-CODE: crowdsourcing condensate database and encyclopedia. *Nat Methods* 2023;**20**:673–6.
- Tesei G, Schulze TK, Crehuet R *et al.* Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc Natl Acad Sci USA* 2021;**118**:1–e2111696118.
- Thomas MG, Fernández-Alvarez AN, Giménez M *et al.* The non-canonical Smoothed-AMPK axis regulates Smaug1 biomolecular condensates. *J Cell Sci* 2025;**138**:JCS263433.
- Wang X, Zhou X, Yan Q *et al.* LLPSDB v2.0: an updated database of proteins undergoing liquid–liquid phase separation in vitro. *Bioinformatics* 2022;**38**:2010–4.
- Wilson C, Lewis KA, Fitzkee NC *et al.* ParSe 2.0: a web tool to identify drivers of protein phase separation at the proteome level. *Protein Sci* 2023;**32**:e4756.
- You K, Huang Q, Yu C *et al.* PhaSepDB: a database of liquid–liquid phase separation related proteins. *Nucleic Acids Res* 2020;**48**:D354–9.