12 October 2026

AI platform and database map protein phase separation across species



Scientists at Nanyang Technological University, Singapore (NTU Singapore) have developed an AI-powered database and a platform that could enable researchers to more accurately predict which proteins undergo phase separation—a process through which proteins in cells separate like oil droplets in water.

The tools are the result of a systematic analysis of predicted phase-separating proteins from various living things, including animals, plants, fungi and single-celled organisms such as bacteria. From their analyses, the scientists also uncovered fundamental insights about the phase separation process.

The research was led by Professor Miao Yansong from the School of Biological Sciences and the Institute for Digital Molecular Analytics and Science (IDMxS) at NTU, in collaboration with Professor Gao Weibo from NTU's School of Electrical and Electronic Engineering. It could be applied to advance the understanding of diseases and aging, as well as to improve crops.

Both AI tools and the research findings have been reported in the journal *Molecular Plant*.

Predicting which proteins separate

Phase-separating proteins separate from their surroundings and concentrate into droplet-like compartments, enabling cells to bring molecules together and accelerate or suppress reactions to respond rapidly to stress.

When phase separation is altered, these proteins can form abnormal condensates or aggregates, which are associated with diseases, including cancer and neurodegenerative disorders such as Alzheimer's disease. In plants, changes in phase separation are linked to impaired stress responses and lowered immunity against microbial pathogens.

Identifying proteins that undergo phase separation and determining how this behavior is regulated are important for applications ranging from biomedical research on diseases and aging to devising approaches for sustainable agriculture.

However, accurate prediction remains challenging because phase-separating proteins assemble dynamically into complex structures and interact with numerous cellular components. Addressing these challenges requires comprehensive experimental data sets and systems-level information across diverse organisms.

In addition, most collections focus predominantly on full-length proteins while underrepresenting validated negative experimental results, truncated proteins and sequence variants like mutations. This imbalance often causes predictive models to overestimate a protein's tendency to condense.

To address this gap, the NTU team used large language AI models to systematically mine scientific papers indexed in PubMed. The researchers extracted information about experimentally tested proteins and their phase behavior before manually reviewing the results to ensure scientific accuracy.

The resulting database, PLATO, contains 5,118 unique protein sequences, including more than 3,600 proteins that undergo phase separation and 1,300 that do not. It captures shortened versions of proteins, mutations and engineered variants that are rare in existing phase-separation data sets. This makes PLATO a valuable resource for developing more accurate AI tools that predict phase-separating proteins.

Predicting protein partners

Many different molecules, such as nucleic acids, also interact with phase-separating proteins to organize them into droplets.

To investigate these interactions, the scientists also created PhaseHub. They first used MolPhase, a machine-learning algorithm that they previously developed, to predict phase-separating proteins from 1,106 species. They then combined MolPhase results with experimental data on protein abundance and interactions to identify these multicomponent compartments in various biological processes.

This approach allows researchers to search for a protein or submit a sequence, examine its predicted phase-separation properties, visualize potential interactions and explore features that influence molecular interactions, such as continuous stretches of repeated amino acids known as homorepeats.

The scientists found that eukaryotes—organisms whose cells contain a nucleus—generally have a substantially higher proportion of proteins that may undergo phase separation than organisms like bacteria that do not have a nucleus. Among organisms with smaller genomes, the proportion of phase-separating proteins increased with genome size before reaching a plateau.

The findings suggest a balance between genomic DNA content and phase-separation capacity. Larger genomes code for more phase-separating proteins, which could enhance cellular processes. However, an excessive tendency to condense may increase the risk of harmful aggregation.

"Our work provides a comprehensive framework for understanding phase-separating proteins, which will serve as a valuable resource for researchers in the field," said Dr. Liang Qiyu, a research fellow at IDMxS, who was the first author of both papers.

"By leveraging AI and systems biology to integrate experimental data with findings from existing literature, our platforms offer accurate, user-friendly tools for predicting and identifying proteins that may undergo phase separation. These tools will advance investigations into the roles of these proteins in agriculture as well as human diseases," said Miao.

https://phys.org/news/2026-10-ai-platform-database-protein-phase.html