A108 · Fine-tuning Boltz-1 for protein-protein interaction prediction with positive and negative data

Presenter: Ruqi Liao — Broad Institute · MIT

Authors: Ruqi Liao, Sarah Gurev, Ashley Gin, Yo Akiyama, Jim Wells, Sergey Ovchinnikov, Hilary Finucane

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Predicting interactions between proteins is crucial to understanding the causes of disease and developing targeted therapeutic interventions. While highly accurate models such as AlphaFold exist for protein structure prediction, current structure-based approaches to protein interaction prediction rely on post-hoc analyses of these structural models or on small-scale models designed for proteome-wide predictions that lack pre-training advantages. To fill this gap, we fine-tuned the trunk of Boltz-1, an open-source reproduction of AF3, with a training data set of interacting and non-interacting domain pairs constructed to prevent the model from exploiting class-specific domain identities or paired MSA depth as shortcuts for interaction prediction.

We benchmarked our method B1-PPI with Boltz-1, Boltz-2, RF2-PPI, and three protein language model-based predictors. On a standard data set of human PPIs, all methods performed similarly, but restricting to subsets of proteins revealed heterogeneity: for example, RF2-PPI underperformed on longer proteins and PLM-interact underperformed on cell surface proteins. To compare methods on the more challenging task of distinguishing positive and negative pairs with high sequence similarity, we constructed a new benchmark and found that B1-PPI and RF2-PPI outperformed other methods. Consistent with this result, B1-PPI was a top-performing method for distinguishing among interacting and non-interacting paralogs in datasets of histidine kinase-response regulator interactions and MALG/MALK interactions. In addition, we found B1-PPI to be the top predictor of success in binder design, although differences among methods were non-significant. Our results demonstrate the power of modern structure-based modeling for protein interaction prediction.