[BIOINFO 2025]

LoGo-BERT: Transformer-Based Integration of Local and Global Features for Scalable Protein–Protein Interaction Prediction

Hae Been Lee[1], and Insuk Lee[1, *]

[1] *Department of Biotechnology, Yonsei University*
**Corresponding author.* insuklee@yonsei.ac.kr

Large-scale genome projects, including the Earth BioGenome Project (EBP), have generated vast protein sequence resources, yet functional annotation remains a major challenge. Protein–protein interaction (PPI) networks are central to linking genome to phenome, but data remain skewed toward a few species, and even state-of-the-art computational models still trade accuracy for scalability. We present LoGo-BERT, a lightweight and accurate model for sequence-based PPI prediction. Built on pretrained ESM-2 embeddings, LoGo-BERT integrates global sequence descriptors with local residue-level similarity via a MaxSim late-interaction mechanism. This design preserves the efficiency of bi-encoders while approaching the accuracy of cross-encoders. Trained on human PPIs, LoGo-BERT generalized to mouse, fly, worm, *E. coli*, and yeast datasets, achieving accuracy comparable to state-of-the-art models while reducing inference time by as much as 300-fold. LoGo-BERT thus enables large-scale cross-species PPI network construction, providing a foundation for functional and biomedical research.