Omni-Prot: Unifying Protein Function Prediction With PLM-Centric Fusion

Rezaei, Ryan (School: Canyon Crest Academy)

Computational annotation of protein function remains severely constrained by the vast gap between available sequences and experimentally validated characterizations. Current methods often fail to fully exploit protein language models by compressing representations too aggressively, relying on a single model, requiring unnecessary auxiliary evolutionary information that limits applicability, or employing architectures that cannot fully capture feature expressiveness. I introduce Omni-Prot, a unified framework for diverse protein prediction challenges that strategically combines representations from four complementary protein language models: ESM-C, ProtT5, Ankh3, and PGLM. The architecture employs reciprocal cross-attention with adaptive gating to enable communication between model streams, producing integrated representations capturing both fine-grained sequence motifs and broader organizational patterns. Evaluated on major benchmarks including Gene Ontology annotation, enzyme classification, and stability regression, Omni-Prot sets new performance standards. Systematic ablations reveal that commonly relied-upon inputs such as 3D coordinates and multiple sequence alignments contribute negligibly when powerful language model representations are properly leveraged, suggesting that large-scale self-supervised pretraining effectively encodes the biological signals these supplementary features provide. Furthermore, by depending exclusively on amino acid sequences, Omni-Prot provides more robust generalization while simplifying deployment and eliminating error propagation from upstream pipelines. Open-source implementations and a web interface are provided to facilitate adoption in applications including therapeutic development, protein design, functional genomics, and more.

Awards Won: