A Unified Protein Embedding Model With Local and Global Structural Sensitivity

Xu, Jerry (School: Lexington High School)

Structural comparison between proteins is key to many research tasks, including evolutionary analysis, peptidomimetics, and functional annotation. Traditional structure alignment tools such as TM-Align, DALI, or ProBiS, which directly superimpose one protein onto another, are accurate but computationally expensive, and impractical at scale. Existing protein language models (PLMs), such as TM-Vec, improve computational efficiency but only capture global structural similarity, overlooking important motif-level structural details. In this paper, we propose a novel PLM consisting of a Siamese neural network, enabling efficient embedding-based structural comparison while also capturing both global and local structural similarity. Our model was trained on a dual loss function combining TM-score, a global similarity metric, and a variation of lDDT scores, a per-residue similarity metric. We tested against two datasets: a varied TM-score dataset from TM-Vec, and a high TM-score mutant dataset from VIPUR. Against these sets, our model achieved a TM-score MAE of 0.0741 and 0.0583, respectively, and a lDDT-score MAE of 0.0788 and 0.0038, respectively. Our model fulfills two key roles: first, it rapidly detects global structural differences. Second, it supports fine-grained structural assessments, improving sensitivity to subtle but functionally important structural changes.

Awards Won:
Second Award of $2,400