





Overview

A one-line drop-in replacement ESM2 that incorporates tor enriched structural information

How? Learn to predict tokens derived from structure model(s)

Training Data

- 6M Alphafold proteins in Uniclust30

Structure-Tuning Loss



Distilling Structural Representations into Protein Sequence Models

Jeffrey Ouyang-Zhang Chengyue Gong Yue Zhao Philipp Krähenbühl Adam Klivans Daniel J. Diaz

Training Implicit Structure Model (ISM)



*ISM is structure-tuned with 2 models - Atomic Autoencoder, MutRank

Results

Long Range Contact

tasks, preserves performance on others