STRUCTURE-BASED SELF-SUPERVISED LEARNING ENABLES ULTRAFAST PREDICTION OF STABILITY CHANGES UPON MUTATION

Jinyuan Sun, AIM center, Institute of Microbiology, Chinese Academy of Sciences, China
jysun@im.ac.cn
Tong Zhu, AIM center, Institute of Microbiology, Chinese Academy of Sciences, China
Yinglu Cui, AIM center, Institute of Microbiology, Chinese Academy of Sciences, China
Bian Wu, AIM center, Institute of Microbiology, Chinese Academy of Sciences, China

Key Words: Protein mutation prediction, Self-supervised learning, Thermostability, Protein engineering

Predicting free energy changes (ΔΔG) is of paramount significance in advancing our comprehension of protein evolution and holds profound implications for protein engineering and pharmaceutical development. Traditional methods, often suffer from limitations such as sluggish computational speed or heavy reliance on biased training datasets. We present Pythia[1], a self-supervised learning (SSL) based graph neural network tailored for zero-shot ΔΔG predictions. In comparative benchmarks with other SSL-based models and force field-based methods, Pythia also exhibits a remarkable acceleration in computational speed, up to 10^5-fold on a single CPU core. The efficacy of Pythia is corroborated through its application in predicting thermostable mutations of limonene epoxide hydrolase with significant higher experimental success rates. Comparing with previous used methods includes energy function calculations (EFC) and molecular dynamic simulations (MDs), the success ratio is more than two-fold higher when considering ΔT_m higher than 1˚C.This efficiency propels the exploration of 26 million high-quality protein structures. Such a grand-scale application signifies a leap forward in our capacity to traverse the protein sequence space and potentially enrich our insights into the intricacies of protein genotype-phenotype relationships. We provided a web app at https://pythia.wulab.xyz for users to conveniently execute predictions and the source code is available at https://github.com/Wublab/Pythia.

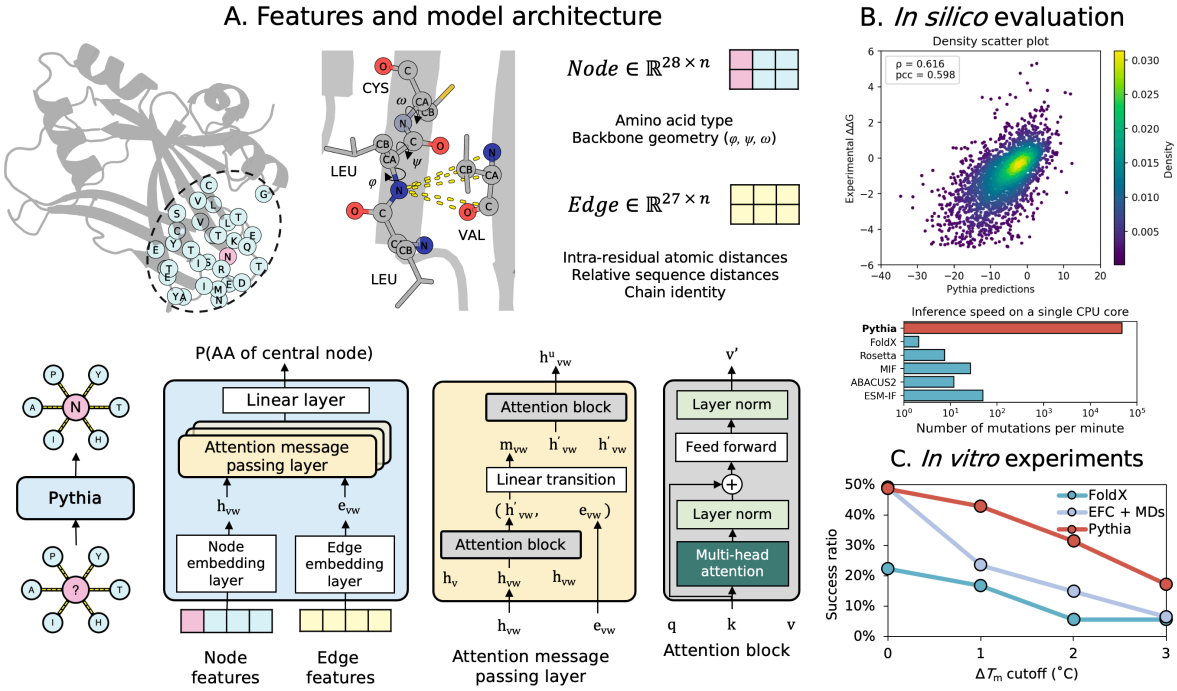

Figure 1 A. Features and model architecture of Pythia. The graph neural network (GNN) is trained to predict the masked out central node of an abstracted protein structure graph. B. Correlations and inference speed of Pythia at the in silico evaluation. C. The success ratio of characterized mutations versus different ΔTm cutoff values across three strategies.

1. Sun, Jinyuan, et al. "Structure-based self-supervised learning enables ultrafast prediction of stability changes upon mutation at the protein universe scale." bioRxiv (2023): 2023-08.