Protein folding stability estimation with explicit consideration of unfolded states

Heechan Lee[1,2], Yugyeong Cho[3], Jeongwon Yun[3], Martin Steinegger[2,4,5,6],

Ho Min Kim[3,*], and Hahnbeom Park[1,7,*]

[1] *Biomedical Research Division, KIST*
[2] *School of Biological Sciences, SNU*
[3] *Department of Biological Sciences, KAIST*
[4] *Interdisciplinary Program in Bioinformatics, SNU*
[5] *Institute of Molecular Biology and Genetics, SNU*
[6] *Artificial Intelligence Institute, SNU*
[7] *Institute for Convergence, SKKU*
Corresponding author: hahnbeom@kist.re.kr, hm_kim@kaist.ac.kr

Folding stability is a critical requirement for the vast majority of proteins. Computational methods suggested to date for the absolute folding stability (ΔG) prediction – including those driven from protein structure prediction AIs – show clear limitations on reproducing quantitative experimental values. Here we present IFUM, a deep neural network that jointly estimates ΔG and the equilibrium ensemble of folded and unfolded states represented by their residue-pair distance probability distributions. This joint learning considerably enhances the ΔG prediction accuracy against the scenario where ΔG prediction was learned alone. To improve the model, we extend the dataset beyond previous related works to include the Mega-scale small proteins and disordered proteins for training as well as wild-type natural proteins with sizes up to 869 residues for validation. We show that IFUM is robust to various protein types and sizes, and is capable of accurately predicting more general types of mutational effects such as sequence insertions or deletions. The applicability of IFUM is demonstrated through two real-world design challenges. First, for blind-tested protein engineering scenarios containing many sequence substitutions and insertions, good correlation with experimental melting temperatures (T_m) is observed. Second, for the *de novo* design selection, IFUM shows considerably improved performance over broadly used AlphaFold-based metrics. IFUM is a free software available at github.com/HParklab/IFUM and also through Google Colab.