

From Molecular Dynamics to Machine Learning: A Local-to-Global Strategy for Predicting Protein–RNA Binding Affinity

Protein–RNA interactions play a central role in numerous biological processes, and the accurate prediction of their binding affinity has become increasingly important for the rational design of RNA-based therapeutics, including RNA aptamers and other emerging RNA-targeting drugs. However, reliable estimation of protein–RNA binding free energies remains a major computational challenge due to the complexity of these interactions and the limited availability of experimental data. We present PANTHER Score (Protein-Affinity for Nucleic Target-binding, Hybridization, and Energy Regression), a novel machine learning framework that combines molecular dynamics (MD)-derived energetic information with a local-to-global scoring strategy to predict protein–RNA binding affinity. The approach is based on the prediction of residue–nucleotide pairwise interaction energies, which are subsequently integrated into a global score strongly correlated with experimental binding free energies. The workflow was developed using more than 87,000 MD-derived local interaction energies extracted from curated protein–RNA complexes and evaluated on independent test and external validation datasets. Among the machine learning models investigated, Random Forest Regression achieved the best performance, reaching a Pearson correlation coefficient of 0.80 on the test set and maintaining robust predictive accuracy on an external dataset of 110 protein–RNA complexes, outperforming currently available methods. By combining physics-based simulations with data-driven learning, PANTHER Score provides an interpretable, computationally efficient, and scalable framework for estimating protein–RNA binding affinity. To facilitate its adoption by the scientific community, the PANTHER scoring pipeline is made freely accessible through a dedicated web service.