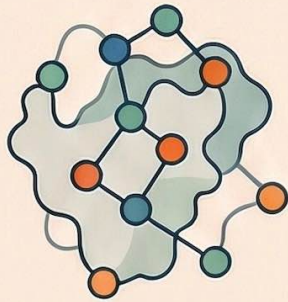


# RankFlow: Transforming Protein Language Models for Accurate Fitness Prediction

## CURRENT PLM CHALLENGES



### Property-Agnostic Representations

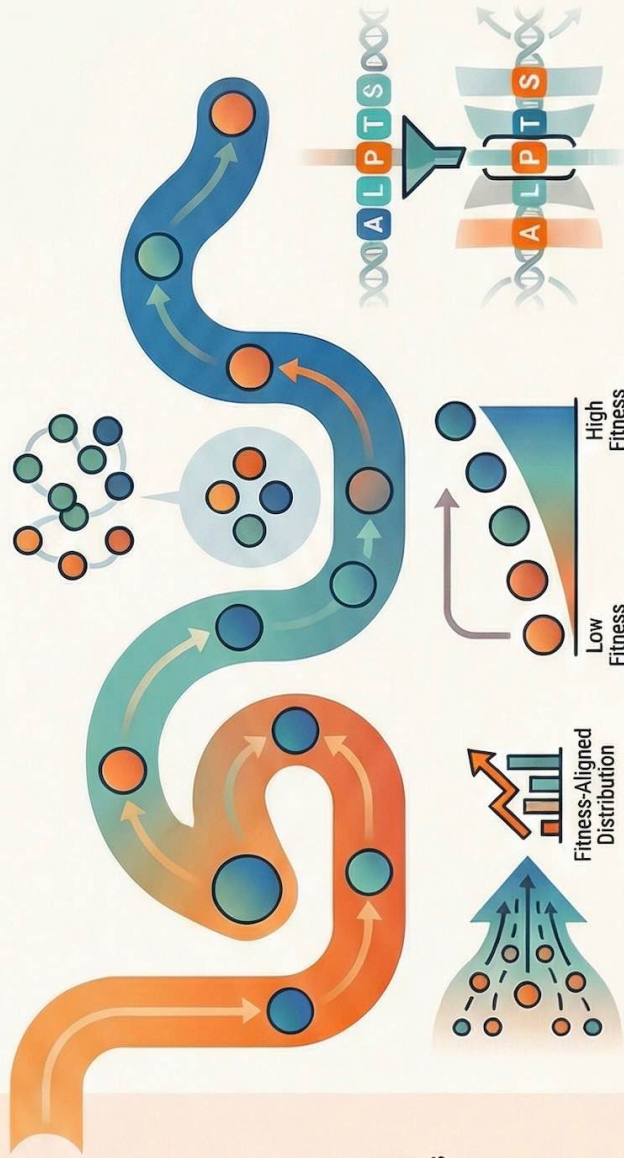
Broad Evolutionary Data, Often Ignoring Specific Target Traits

Misses Complex Multi-Mutation Interactions (Epistasis)

Inaccurate Fitness Optimization

## THE RANKFLOW FRAMEWORK

Refining representations into property-aligned distributions via conditional flow.



### PROPERTY-AWARE CONDITIONAL FLOW

Learns a velocity field to transport PLM representations toward a fitness-aligned distribution via an energy function.

### RANK-CONSISTENT (RC2) LOSS

A differentiable objective that enforces the correct ordering of mutants, significantly improving out-of-distribution generalization.

### PROPERTY-GUIDED STEERING GATE (PSG)

Concentrates learning on specific sequence positions carrying functional signals while suppressing unrelated evolutionary biases.

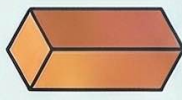


### STATE-OF-THE-ART (SOTA) ACCURACY

Consistently outperforms supervised methods across ProteinGym, PEER, and FLIP benchmarks in ranking accuracy.

### 17x MORE PARAMETER EFFICIENT

Achieves SOTA results with only 37.1M trainable parameters compared to 650M in traditional PLM fine-tuning.



Traditional PLM  
650M  
Parameters



RankFlow  
37.1M  
Parameters



### RAPID TRAINING & INFERENCE

Enables complete model training in approximately one hour on a single A100 GPU.



### MODEL PERFORMANCE COMPARISON

Model	Avg. ProteinGym Score	Trainable Parameters
RankFlow (Ours)	0.911	37.1M
DePLM (ESM2)	0.897	42.2M
SaProt (Fine-Tuned)	0.882	650M