

Background/objective

Machine learning (ML) has been increasingly used to address limitations in protein engineering strategies such as directed evolution. In ML-assisted directed evolution, a ML model is trained to learn the sequence-function relationship from gene sequencing and screening data. However, the sequence contexts that current models capture are not specific to the protein being engineered, which limits the accuracy of existing ML algorithms. We developed ECNet (Evolutionary Context-integrated neural Network), a deep-learning ML algorithm that exploits evolutionary contexts to predict functional fitness for protein engineering, which can more accurately predict the sequence-function relationship than existing algorithms. We used ECNet to guide the engineering of TEM-1 β -lactamase to identify variants with improved ampicillin resistance.

Approach

- ❖ Integrated the local evolutionary context from homologous sequences that explicitly model residue-residue epistasis for the protein of interest with the global evolutionary context that encodes rich semantic and structural features from the enormous protein sequence universe to develop the ECNet model.
- ❖ Compared the prediction accuracy of ECNet and other ML models for protein engineering.
- ❖ Experimentally validated the ECNet model using TEM-1 β -lactamase to identify variants with improved ampicillin resistance.

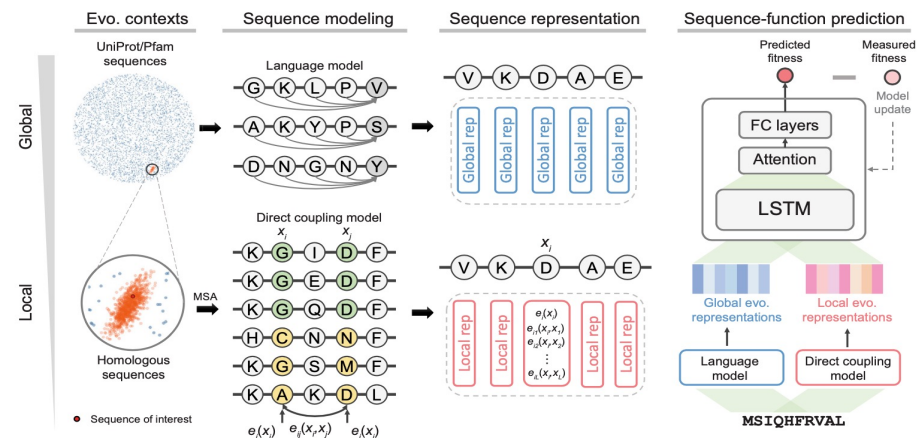
Results

- ❖ ECNet outperformed multiple existing ML models in prediction accuracy and generalizability for protein engineering when benchmarked against a large set of deep mutational scanning studies.
- ❖ ECNet can be generalized to predict higher-order mutants' functions by learning from the data of lower orders, which is a desired capability as higher-order mutants' functions are often costly and time-consuming to measure.

Significance

ECNet can predict the protein sequence-function relationship more accurately than existing ML algorithms and should be generally applicable to the engineering of proteins for bioenergy research.

ECNet is an Evolutionary Context-Integrated Deep Learning Framework for Protein Engineering



An overview of ECNet, an evolutionary context-integrated, deep learning model for protein engineering.