trRosetta

PNAS 20

Protein folding

- Consider folding as classification problem
- Classify 4 features: distance, dihedral angles
- Residual network with 2D convolution, InstanceNorm, ELU, and Dropout
- Loss: categorical cross entropy
- Used Rosetta functions for 3D structure prediction

ResNet Classification

MSA Representation

Improved protein structure prediction using predicted interresidue orientations

The prediction of interresidue contacts and distances from coevolutionary data using deep learning has considerably advanced protein structure prediction. Here, we build on these advances by developing a deep residual network for predicting interresidue orientations, in addition to distances, and a Rosetta-constrained energy-minimization protocol for rapidly and accurately generating structure models guided by these restraints. In benchmark tests on 13th Community-Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP13)- and Continuous Automated Model Evaluation (CAMEO)-derived sets, the method outperforms all previously described structure-prediction methods. Although trained entirely on native proteins, the network consistently assigns higher probability to de novo-designed proteins, identifying the key fold-determining residues and providing an independent quantitative measure of the “ideality” of a protein structure. The method promises to be useful for a broad range of protein structure prediction and design problems.

UniRep

Nature Methods 19

Representation

- Multiplicative LSTM
- Embedding: hidden units of LSTM
- Top model: can be applied to diverse tasks
- Rich information: evolutionary, functional and chemical property, secondary structure
- Tested on ~10 benchmarks of structural and functional properties of protein

mLSTM Top model

Unified rational protein engineering with sequence-based deep representation learning

Rational protein engineering requires a holistic understanding of protein function. Here, we apply deep learning to unlabeled amino-acid sequences to distill the fundamental features of a protein into a statistical representation that is semantically rich and structurally, evolutionarily and biophysically grounded. We show that the simplest models built on top of this unified representation (UniRep) are broadly applicable and generalize to unseen regions of sequence space. Our data-driven approach predicts the stability of natural and de novo designed proteins, and the quantitative function of molecularly diverse mutants, competitively with the state-of-the-art methods. UniRep further enables two orders of magnitude efficiency improvement in a protein engineering task. UniRep is a versatile summary of fundamental protein features that can be applied across protein engineering informatics.

Design and Engineering

ReLeaSE
#RL #SMILES

GEGL
#RL #GA #SMILES

DyNA-PPO
#RL #Binding #PO

Deep learning and architecture

SAC
#RL

PPO
#RL #PO

Annotation

Color box
Reviewed!
Click to see the summary

Grey box
Not yet reviewed

Representation

eUniRep
#LSTM

UniRep
#LSTM

GVP
#Graph

Hallucination
#MCMC

Property prediction

ProteinBERT
#BERT

LM-GVP
#Graph #LM

DeepFRI
#GCN

Augmentation

AlphaFold2
#Graph #Attention

ProteinQA
#Graph #QA

trRosetta
#ResNet

ProteinSolver
#Graph #Inverse

ProteinGAN
#GAN #Sequence

GVP+AF
#Inverse #Structure

ProteinGCN
#GCN #QA

AR-VAE
#VAE #Sequence

DeepGO
#CNN

DeepGOPlus
#CNN

DeepGOZero
#CNN #Zero-shot

TAPE
#Benchmark #Transformer

GO
#Benchmark

Folding / Inverse folding