AlphaFold2

Nature 21

Protein folding

- Consider folding as graph inference problem
- Pair representation: relation between residues
- Evoformer: MSA representation updates the pair representation through attention
- Structure module: predict relative positions and translations of backbone
- Iterative local refinement

Graph learning Evoformer

MSA Representation Attention

Highly accurate protein structure prediction with AlphaFold

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort, the structures of around 100,000 unique proteins have been determined, but this represents a small fraction of the billions of known protein sequences. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the 'protein folding problem'—has been an important open research problem for more than 50 years. Despite recent progress, existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14), demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

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.

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