# GVP

ICLR 21

Representation

- Dual representation: scalar and vector
- Augment graph neural networks with geometric reasoning layers
- Learning vector-valued and scalar-valued functions over geometric vectors and scalars
- Can approximate any rotation-invariant and reflection-invariant function of vectors

Graph learning

Dual representation

## Learning from protein structure with geometric vector perceptrons

Learning on 3D structures of large biomolecules is emerging as a distinct area in machine learning, but there has yet to emerge a unifying network architecture that simultaneously leverages the graph-structured and geometric aspects of the problem domain. To address this gap, we introduce geometric vector perceptrons, which extend standard dense layers to operate on collections of Euclidean vectors. Graph neural networks equipped with such layers are able to perform both geometric and relational reasoning on efficient and natural representations of macromolecular structure. We demonstrate our approach on two important problems in learning from protein structure: model quality assessment and computational protein design. Our approach improves over existing classes of architectures, including state-of-the-art graph-based and voxel-based methods.

# Hallucination

Nature 21

Engineering

- Structure prediction networks can be inverted to generate new protein sequences
- Random sequence → diffused distance map
- MCMC: repeat mutation and accept if it improved score → sharp distance map
- KL-divergence of trRosetta and background network, trRosetta without skip connection

trRosetta MCMC

Background network

## De novo protein design by deep network hallucination

There has been considerable recent progress in protein structure prediction using deep neural networks to predict inter-residue distances from amino acid sequences. Here we investigate whether the information captured by such networks is sufficiently rich to generate new folded proteins with sequences unrelated to those of the naturally occurring proteins used in training the models. We generate random amino acid sequences, and input them into the trRosetta structure prediction network to predict starting residue–residue distance maps, which, as expected, are quite featureless. We then carry out Monte Carlo sampling in amino acid sequence space, optimizing the contrast between the inter-residue distance distributions predicted by the network and background distributions averaged over all proteins. Optimization from different random starting points resulted in novel proteins spanning a wide range of sequences and predicted structures. We obtained synthetic genes encoding 129 of the network-hallucinated sequences, and expressed and purified the proteins in *E. coli;* 27 of the proteins yielded monodisperse species with circular dichroism spectra consistent with the hallucinated structures. We determined the three-dimensional structures of three of the hallucinated proteins, two by X-ray crystallography and one by NMR, and these closely matched the hallucinated models. Thus, deep networks trained to predict native protein structures from their sequences can be inverted to design new proteins, and such networks and methods should contribute alongside traditional physics-based models to the de novo design of proteins with new functions.

## 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