Design with ProteinMPNN Intro

This talk is focused on the publication:

PROTEIN DESIGN

Robust deep learning–based protein sequence design using ProteinMPNN

J. Dauparas[1,2], I. Anishchenko[1,2], N. Bennett[1,2,3], H. Bai[1,2,4], R. J. Ragotte[1,2], L. F. Milles[1,2], B. I. M. Wicky[1,2], A. Courbet[1,2,4], R. J. de Haas[5], N. Bethel[1,2,4], P. J. Y. Leung[1,2,3], T. F. Huddy[1,2], S. Pellock[1,2], D. Tischer[1,2], F. Chan[1,2], B. Koepnick[1,2], H. Nguyen[1,2], A. Kang[1,2], B. Sankaran[6], A. K. Bera[1,2], N. P. King[1,2], D. Baker[1,2,4]*

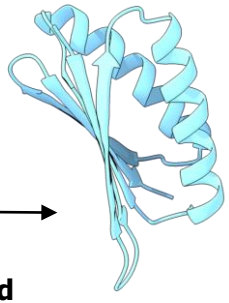
Sep 2022, Science

ProteinMPNN generates a sequence from a backbone

DIQVQVNIDDNG
KNFDYTYTVTT…

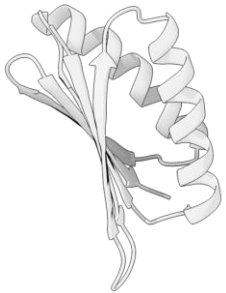
Input

Rosetta
AlphaFold
RosettaFold
…



Output

Structure prediction tools predict a protein structure from an input sequence, whereas…



Input

ProteinMPNN

DIQVQVNIDDNG
KNFDYTYTVTT…

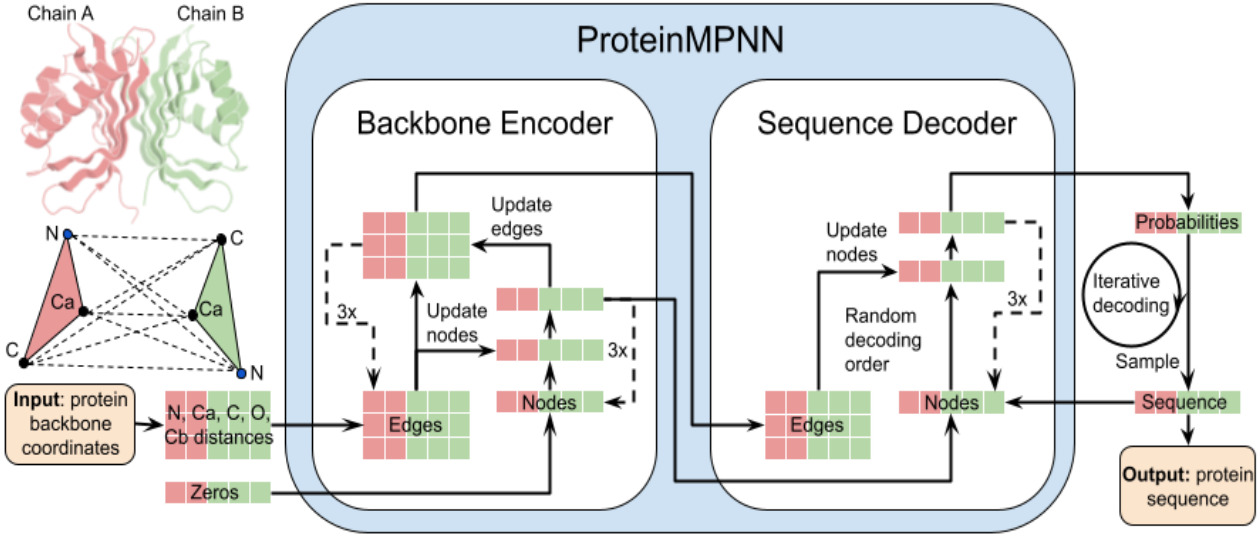
Output

ProteinMPNN tries to map a sequence onto a given protein structure. This is used for **protein redesign**.

Design with Rosetta vs ProteinMPNN

Rosetta Design (remode, fastdesign, packrotamers …)	ProteinMPNN
▪ Physics based; tries to optimize energies	▪ Deep learning model
▪ Slow (due to rotamer search and energy minimization steps)	▪ Fast (few seconds per sequence)
▪ Can handle monomers, multimers and symmetry	▪ Can also handle monomers, multimers and symmetry
▪ Excellent for special design challenges; e.g. sidechain modifications, pepide cyclyzation, ligand pocket design, antibody design, etc.)	▪ Excellent for rescuing de novo designs and improving expression

Design with Rosetta vs ProteinMPNN



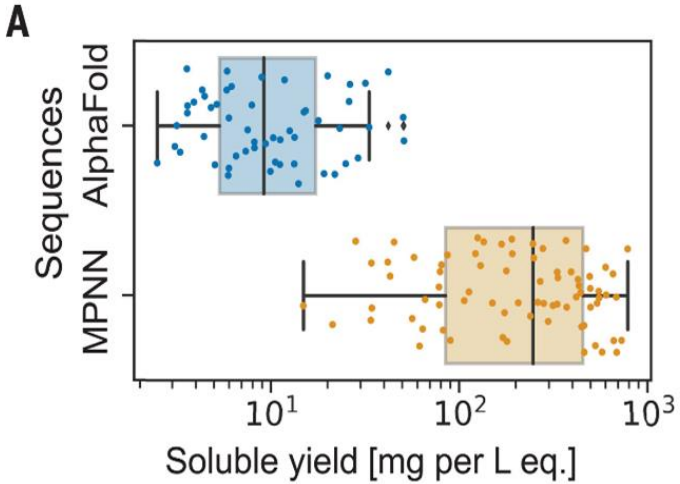
Input: encoded distances of backbone atoms

MPNN: message passing neural network

Output: Sequence and probability estimate (score)

from: Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).c

ProteinMPNN increases the soluble yield



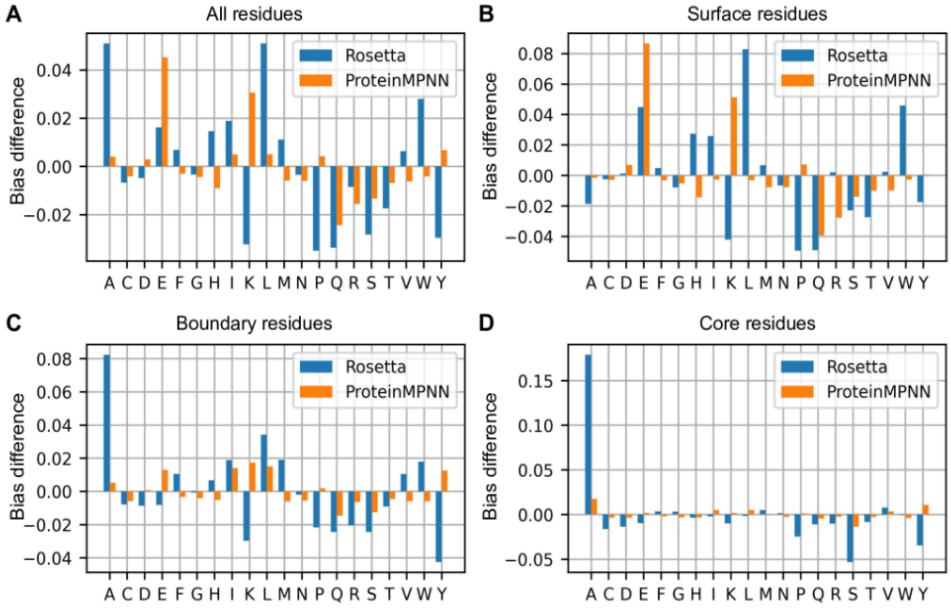
The Example:

Rescuing by AlphaFold hallucinated proteins with ProteinMPNN.

By redesigning the de novo designed proteins with ProteinMPNN the soluble yield could be significantly improved

from: Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).c

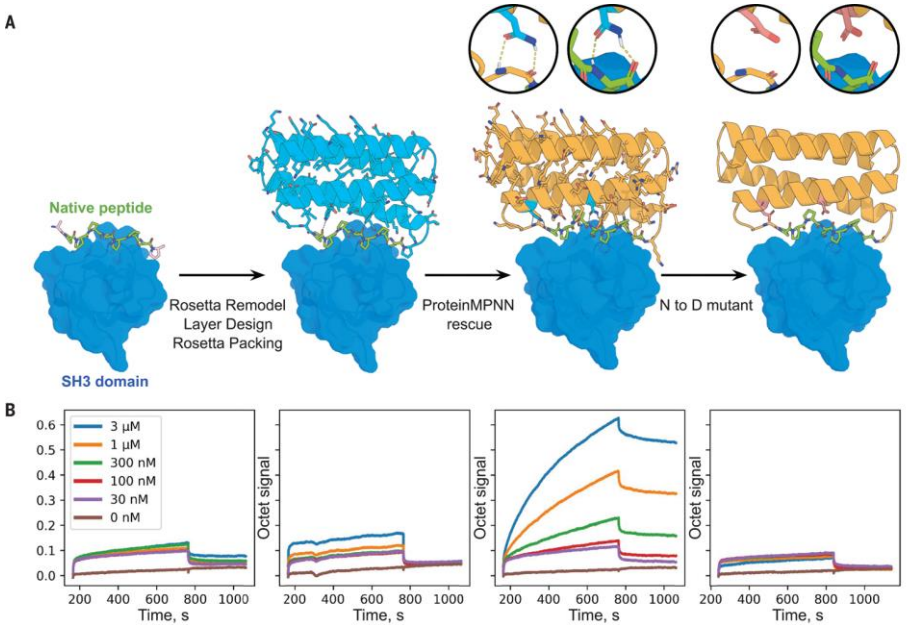
Amino acid bias Rosetta vs ProteinMPNN



With Rosetta design usually a slight bias towards hydrophobic residues can be observed, whereas ProteinMPNN tends to Glu and Lys on the surface of proteins.

from: Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).c

ProteinMPNN rescue example



The Example:

To stabilize the native peptide ligand, helical scaffolds were docked to the peptide-protein complex and connected to the peptide via Rosetta Remodel. But the de novo designs showed no binding to SH3.

After redesigning the helical bundle and loops with **ProteinMPNN** binding could be rescued.

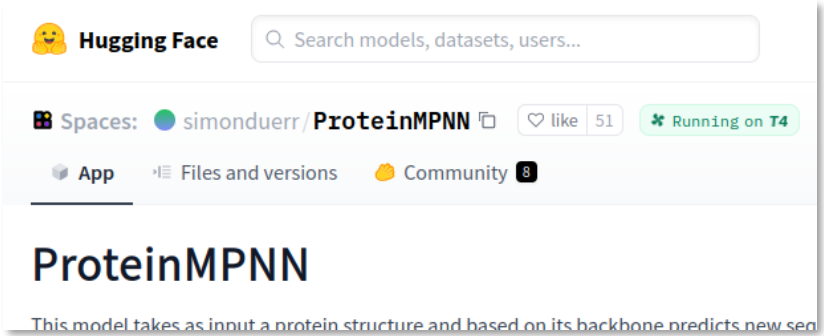
from: Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).c

How do I run ProteinMPNN

In the browser:

huggingface.co/spaces/simonduerr/ProteinMPNN

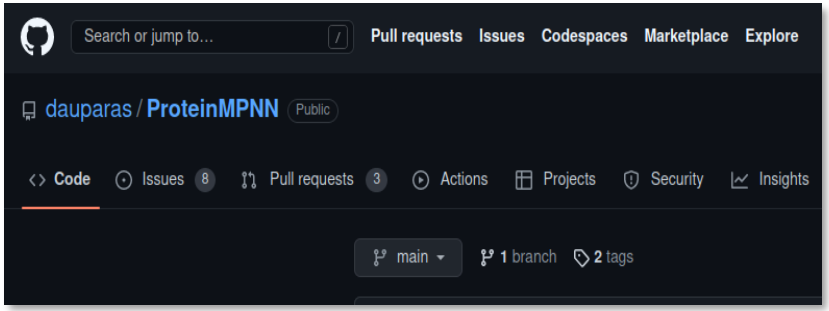
- Easy to use graphical interface
- Offers main functionalities and options
- Great for redesigning a small number of structures



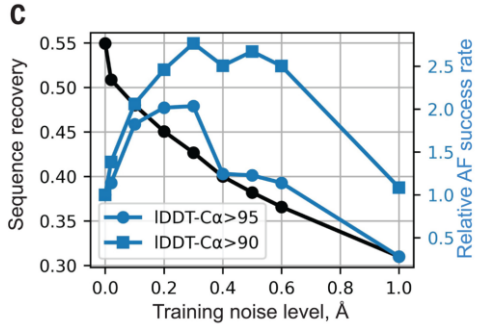
Via command line:

github.com/dauparas/ProteinMPNN

- Locally or on a HPC
- Offers some applications that are not available on huggingface
- Great for redesigning a large number of structures

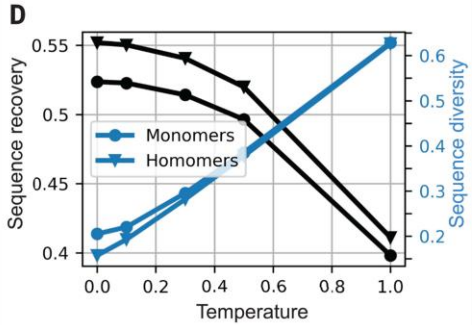


ProteinMPNN settings: Model and Temperature



Model (--model_name)

- default: v_48_020 (version with 48 edges and 0.2 A backbone noise)
- models trained with more noise are more robustly decoded into 3D coordinates by AlphaFold

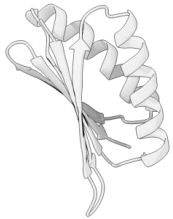


Temperature (--sampling_temp)

- default: 0.1; can be increased for more sequence diversity
- slight change in temperature gives more sequence diversity but only small decrease in sequence recovery

from: Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).c

Running ProteinMPNN



Input: one or multiple pdb files

protein_mpnn_run.py

Options

```
DIQVQVNIDDNG
KNFDYTYTVTT…
```

Output: Sequences with scores

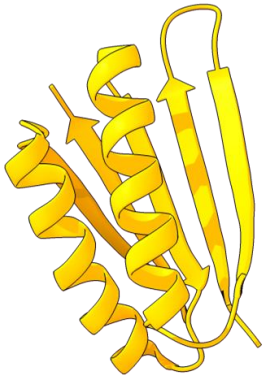
Important options:

- Model: selection of models with different noise levels
- Temperature: to increase sequence diversity
- Tying positions: for symmetric design of Homooligomers
- Fixing positions: to only design selected residues
- Omit amino acids: to not design with certain amino acids
- Bias design: to bias design to select certain amino acids with a higher or lower probability. Can be limited to selected positions.

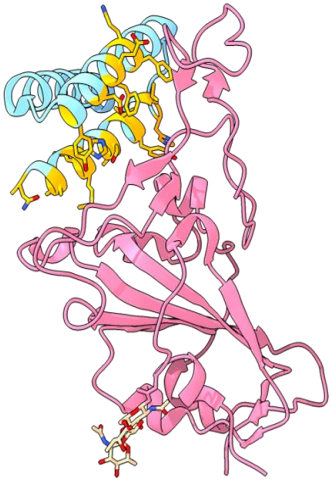
…more in the tutorial

ProteinMPNN Tutorial

How to run ProteinMPNN via command line



1. We will redesign the small de novo designed monomer Top7 (pdb:1QYS) and explore how to set an amino acid



2. We will redesign the interface of the small SARS-CoV-2 spike binder LCB3 and explore how to restrict the design to certain positions.