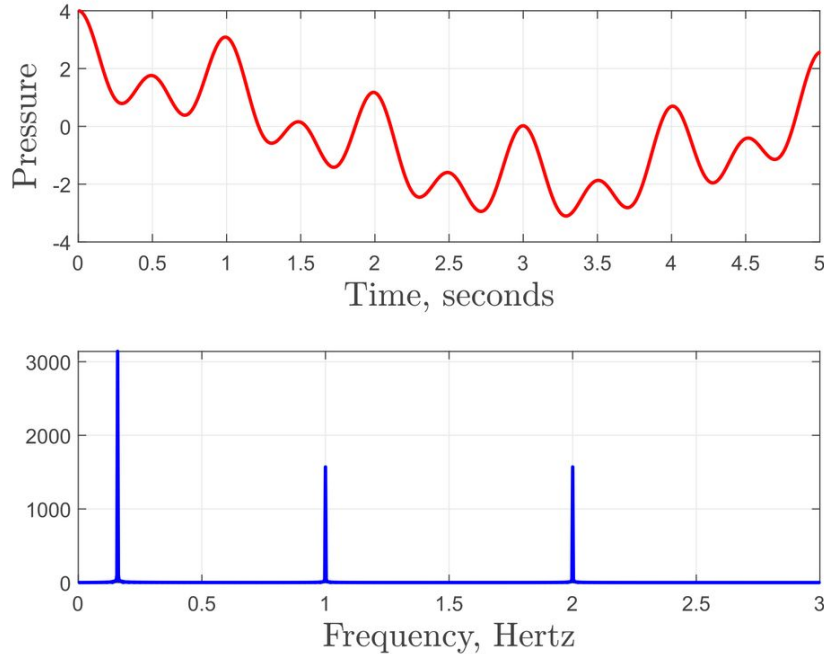


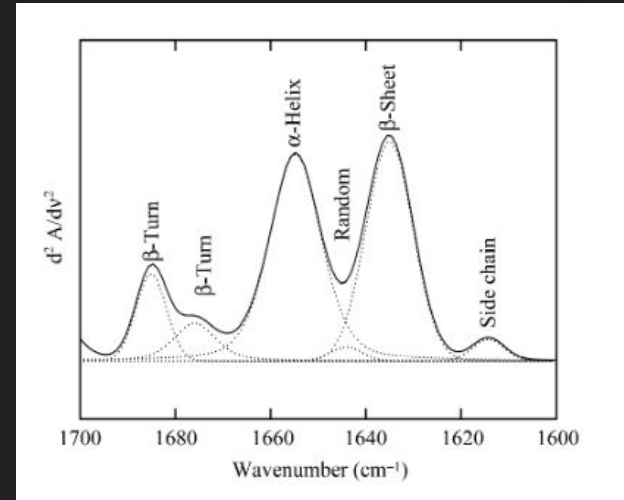
Can We Learn From Protein Latent Space

Emerson Goss, Long Tran

What is Fourier Transform



Protein structure contains
frequency-domain relationships!

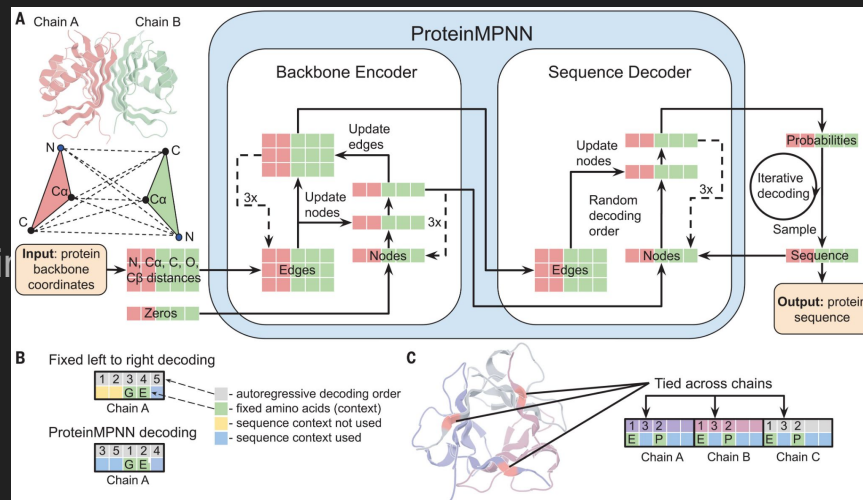


Problem

Sequence design is a critical step in protein design

Currently, protein's structure is represented as 3D xyz coordinate.

Can we learn or improve protein design from protein latent space?



FTT in Other Studies

Random Noise Generation Using Fourier Series

Jared A. Grauer

Published Online: 22 Feb 2018 • <https://doi.org/10.2514/1.C034616>

Sections

Read Now

Tools

SimVAE: Simulator-Assisted Training for Interpretable Generative Models

Akash Srivastava*
MIT-IBM Watson AI Lab
IBM Research
Cambridge, MA
akash.srivastava@ibm.com

Jessie Rosenberg*
MIT-IBM Watson AI Lab
IBM Research
Cambridge, MA
jcrosemb@us.ibm.com

Variational Autoencoders: A Harmonic Perspective

Alexander Camuto
University of Oxford

Matthew Willetts
UCL & The Alan Turing Institute

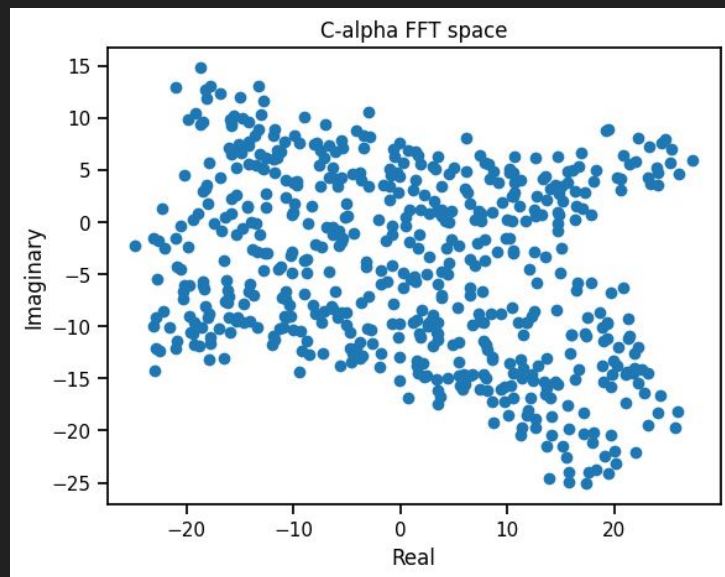
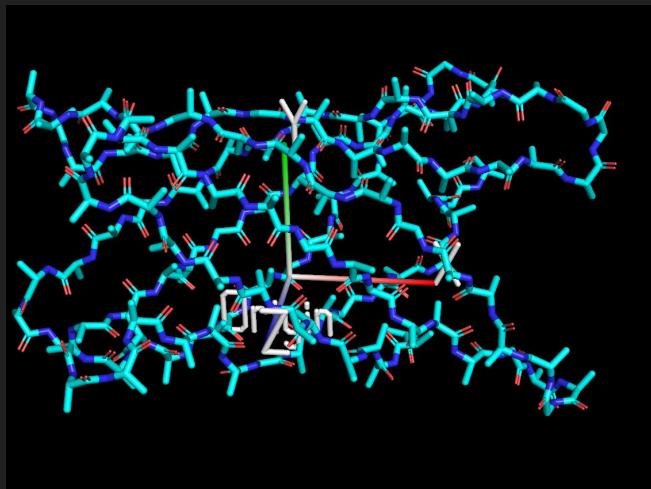
Propose POC Solution

Transform xyz to FTT → sequence design → Fold to validate

Current implementation Transform xyz to FTT → noise to get FTT characteristics
→ inverse FTT for formatting

Case Study

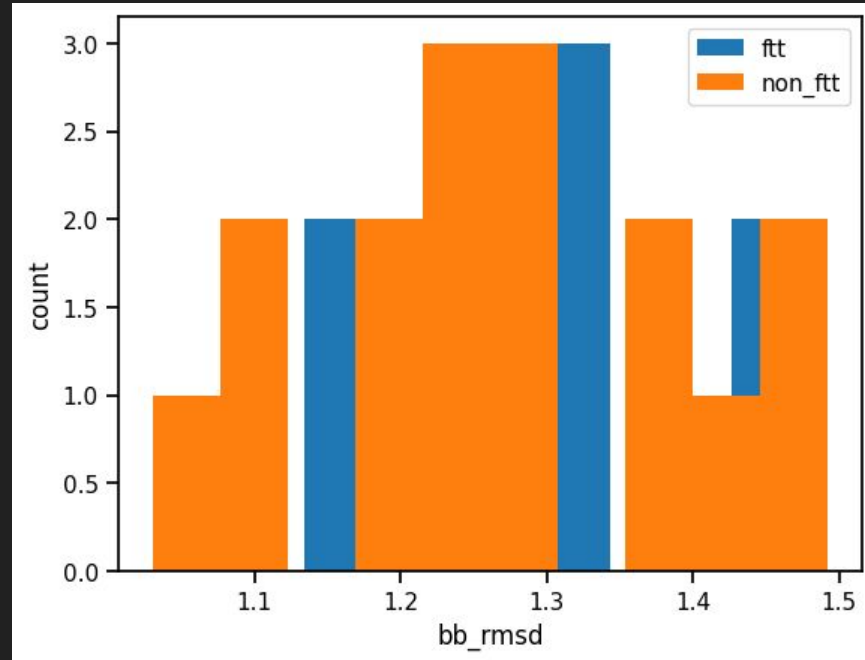
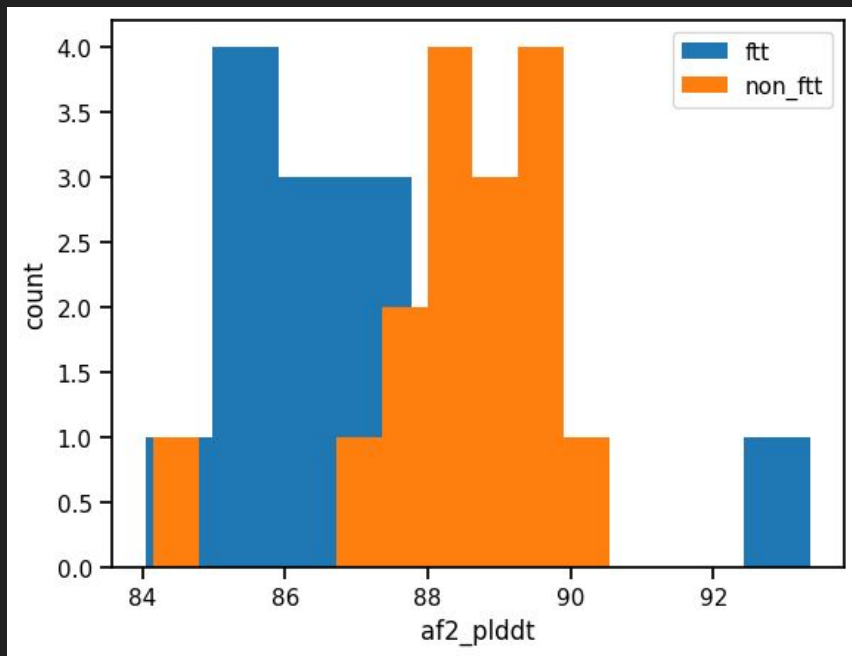
C4.pdb



POC Pipeline for Sequence Design

- Feed MPNN the FTT-noised xyz
- get sequence
- threaded sequence on protein backbone for folding
- extract folding metrics

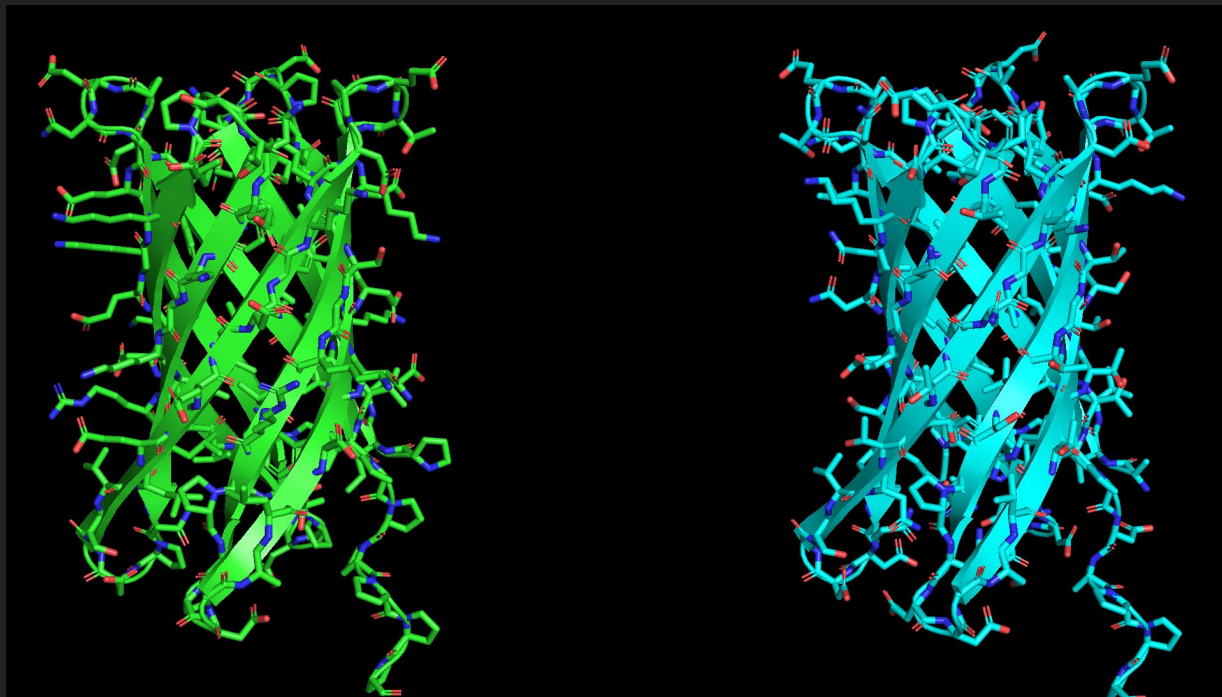
POC Pipeline for Sequence Design - Results



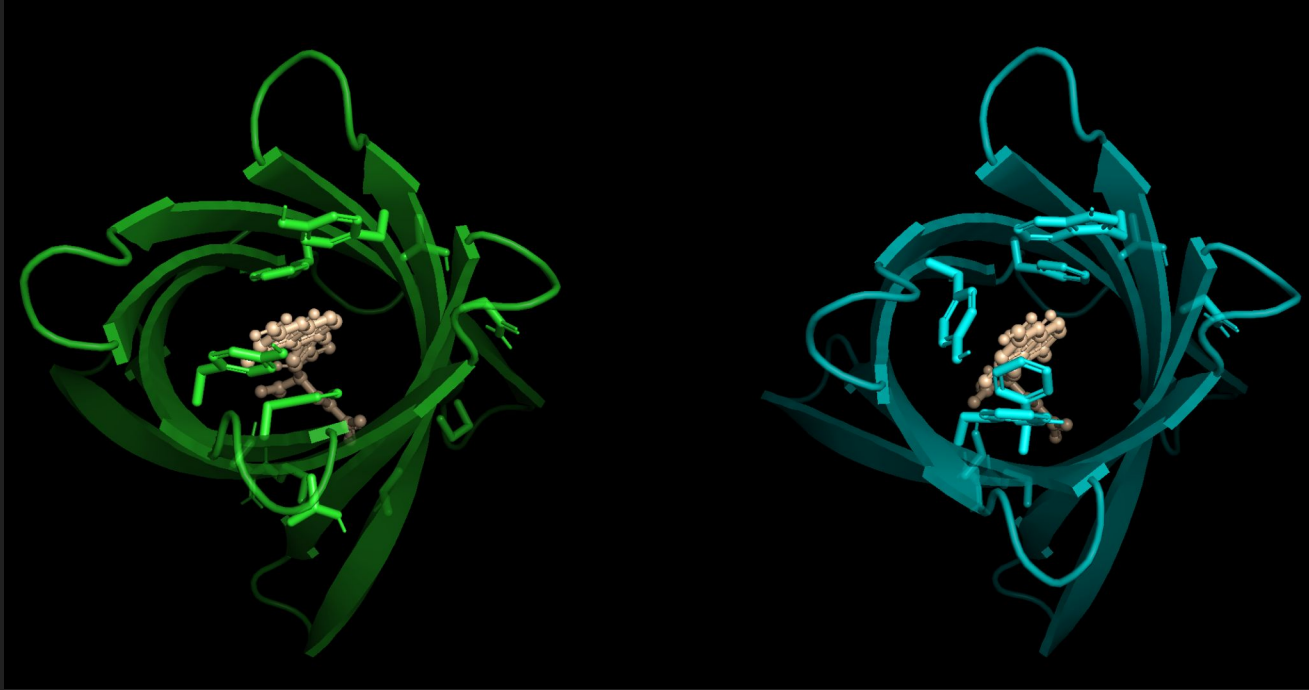
Folding Structure

Green = FTT

Cyan = non_FTT



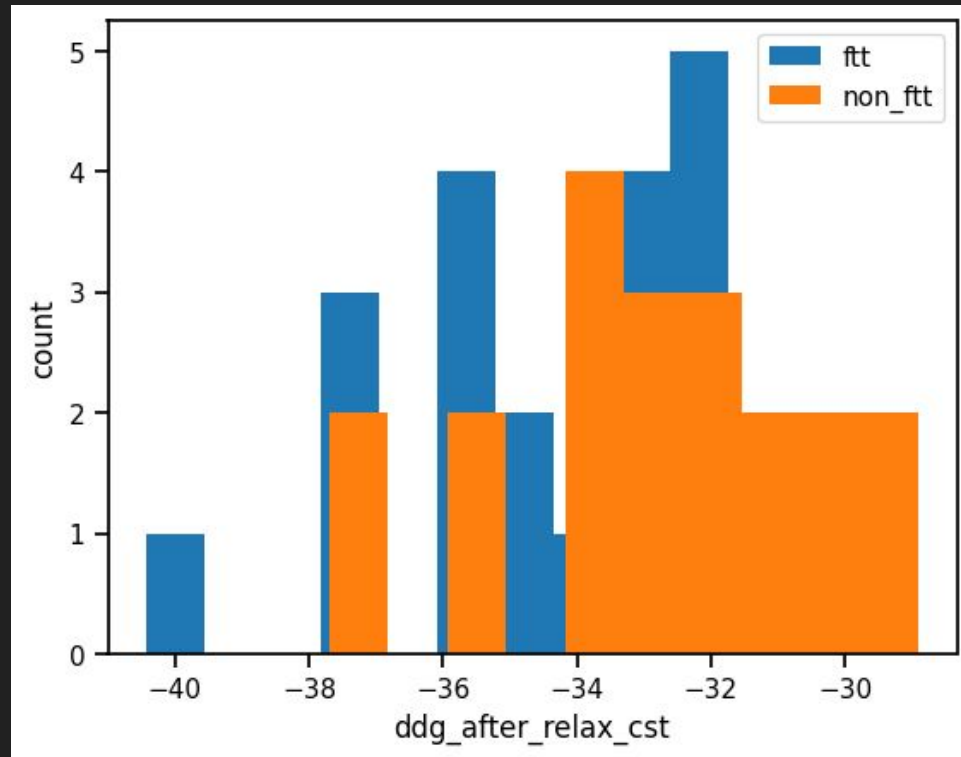
What about Protein Design at Higher Resolution



Protein Design at Higher Resolution - Results

ddg is similar to Gibbs energy

Predicts the likelihood of
reaction equilibrium



Impacts from Funding



Improve sequence generation algorithms

More sensitive protein for more applications

- Model small molecule interactions
- Model secondary structures more precisely

References

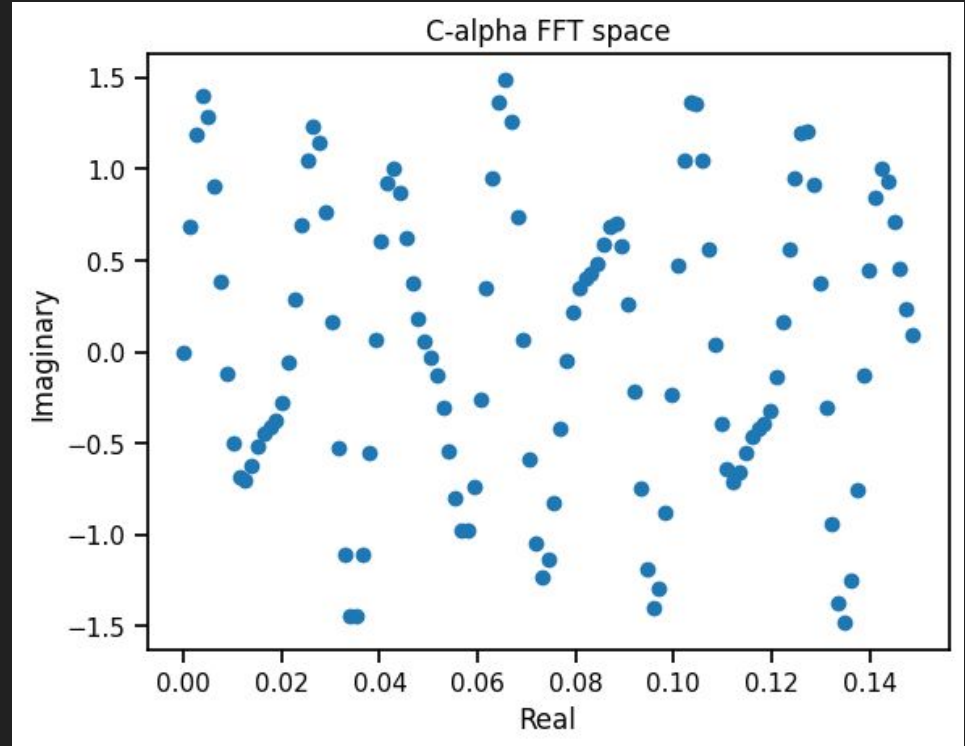
1. Grauer, J. A. (2018b). Random Noise Generation Using Fourier Series. *Journal of Aircraft*, 55(4), 1754–1760. <https://doi.org/10.2514/1.c034616>
2. SIMVAE: Simulator-assisted training for interpretable Generative Model - arxiv.org, <https://arxiv.org/pdf/1911.08051v1.pdf> (accessed May 10, 2023)
3. Camuto, A., & Willetts, M. (2021). Variational Autoencoders: A Harmonic Perspective. ArXiv. /abs/2105.14866

https://github.com/iamlongtran/cheme_577_demos

SI

Case Study

3NIR xyz after coordinate transformation



Noise in FTT space



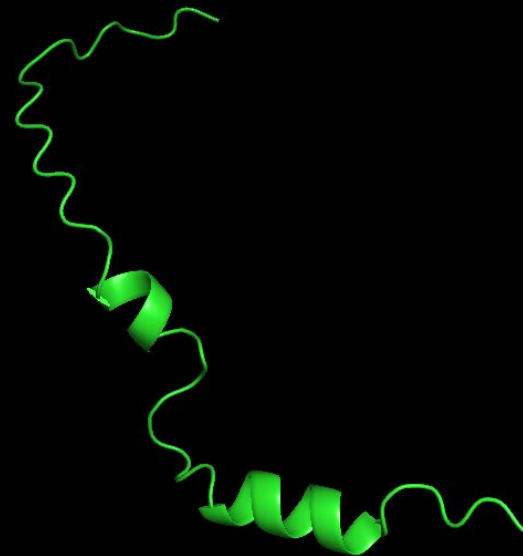
3NIR before FTT

3NIR after FTT

AF2 Folding - FTT backbone

Zero shot testing run

```
/3nir_bb_FTT_dldesign_pbias_0//A/1 6 11 16 21 26 31 36 41 46  
MVAGPSPEAREQYEKNLEPGTPKSEAAKRTGLIEIGKEQPGDYPY
```



MPNN Sequence Design



```
1 pdirs = [f'{home_dir}/3nir_bb.pdb', f'{home_dir}/3nir_bb_FTT.pdb']
2 cmds = []
3 for pdb in pdirs:
4
5     pdb_name = pdb.split('/')[-1].split('.')[::-4]
6     cmd = '/software/conda/envs/PPI_design/bin/python /home/lhtran/sc/mpnn_generic_tied_design2.py
7         --design_these_chains A --keep_these_chains NA --tie_chainsa NA --tie_chainsb NA --tie_repeats 0 --fix_a NA'
8     cmd += ' --fix_b NA --rmsd_cutoff 1.0 --max_designs_per_backbone 5 --num_seq_per_target 8 --sampling_temp "
9     cmd += ' 0.2 0.3" --plddt_cutoff 0.9 --ptm_cutoff 0.65 --pdb_path '
10    cmd += ' '
11    cmd += f' --out_name resampling --suffix _pbias --pdb_out_dir /home/lhtran/class/cheme_577/mpnn_af2/pdirs'
12    cmd += '\n'
13    cmds.append(cmd)
14 # with open(f'{home_dir}/mpnn_af2/{pdb}/array_task_mpnn_af2', 'w') as fp:
15 #     for j in cmds:
16 #         fp.write(j)
```

✓ 0.0s

Python