Predicting Peptide Structural Feasibility with CNNs

Owen Krueger[1], Mohit Malu[3], Rebecca Sutcliffe[4], Nicholas Dunne[2], You Zhou[1] ,Giulia Pedrielli[1], Andreas Spanias[3]

1. SCAI at Arizona State University 2. Biodesign Europe at Dublin City University 3. School of ECEE at Arizona State University 4. Queens University Belfast

Insight SFI RESEARCH CENTRE FOR DATA ANALYTICS

A World Leading SFI Research Centre

Science Foundation Ireland sfi For what's next

Project Goal

Predict the structural feasibility of peptides utilizing CNNs

Motivation

Provide researchers with a predictive tool to accelerate the analysis of peptide model stability.

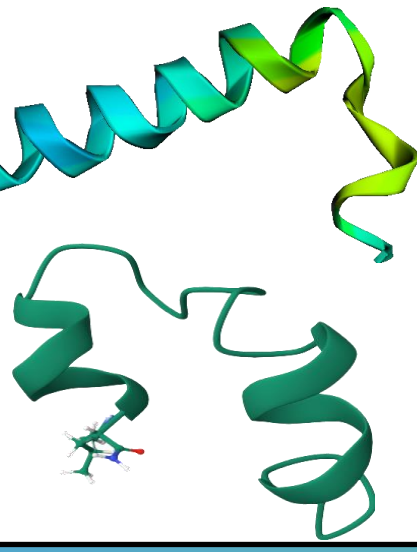
Task 1: Training Set Creation

Create a labeled training set by comparing the structural similarity of generated peptides to their real counterparts

Sequence:
SPLGEEMRDRARAHVDALRTHLAPYS
DELRQRLAARLEALKENGGA

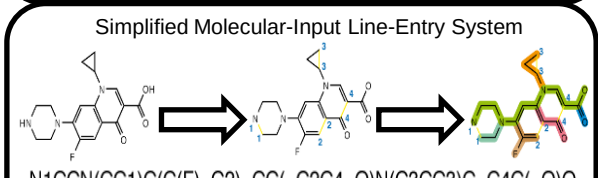
AlphaFold

RCSB PDB PROTEIN DATA BANK



S.M.I.L.E.S

Simplified Molecular-Input Line-Entry System



N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

Adapt Data

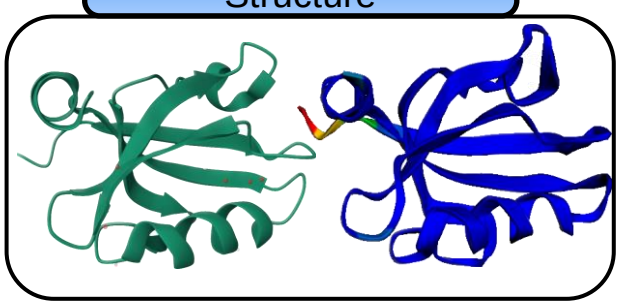
Focus:
- ❑ Reduce Dimensionality
- ❑ Nonadjacent bonding
- ❑ Structural components

.PDB Format

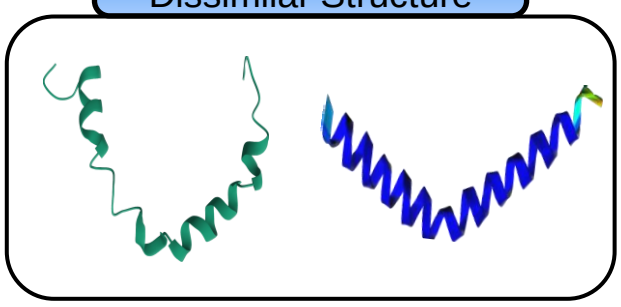
```
ATOM      1  N   HIS B  -1      75.822 -55.481  -8.354  1.00 47.79
ATOM      2  CA  HIS B  -1      76.394 -55.280  -7.029  1.00 57.91
ATOM      3  C   HIS B  -1      76.626 -53.799  -6.752  1.00 52.10
ATOM      4  O   HIS B  -1      77.492 -53.435  -5.956  1.00 47.84
ATOM      5  CB  HIS B  -1      77.707 -56.053  -6.892  1.00 54.37
```

Compare SMILES for Similarity

Positive Label: Similar Structure



Negative Label: Dissimilar Structure



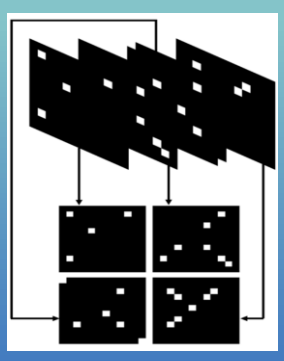
Preliminary Results

Peptide ID:	Cosine Similarity:	Label
Shu3	0.9974	Positive
N7l	0.9913	Positive
e0a	0.2074	Negative
M8l	0.2635	Negative
1CEU	0.3680	Negative
1Hzn	0.4200	Negative
Gw4	0.1675	Negative
1Loi	0.1848	Negative

Future Key Tasks

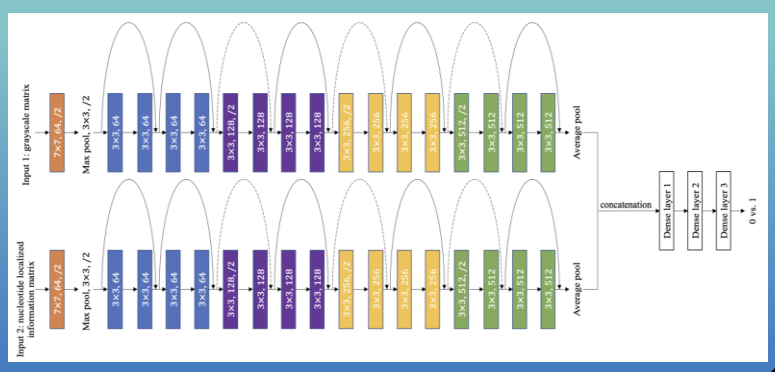
- ❑ **Task 2: Adapt NU-Resnet for Peptide Sequences**

Adapt current RNA based CNN to take in and predict peptide sequences



- ❑ **Task 3: Adapt NUMO-Resnet to utilize Peptide Motifs**

Implement motif input data for NUMO-Resnet CNN to utilize



Sources

1. Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
2. Y. Zhou, G. Pedrielli, F. Zhang, T. Wu "Deep learning to predict RNA Stability: NU-ResNet and NUMO-ResNet," unpublished.
3. O'Boyle, N.M., Banck, M., James, C.A. et al. Open Babel: An open chemical toolbox. J Cheminform 3, 33 (2011). https://doi.org/10.1186/1758-2946-3-33
4. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules David Weininger *Journal of Chemical Information and Computer Sciences* 1988 *28* (1), 31-36 DOI: 10.1021/ci00057a005

HOST INSTITUTIONS
DCU · Ollscoil na Gaillimhe University of Galway · UCC University College Cork · UCD Dublin

PARTNER INSTITUTIONS
Maynooth University · Tyndall · Trinity College Dublin · University of Limerick

FUNDED BY:
European Union European Regional Development Fund · Science Foundation Ireland sfi

ASU Biodesign Institute Arizona State University · DCU · Biodesign Europe

ASU Ira A. Fulton Schools of Engineering Arizona State University

SenSIP Sensor, Signal and Information Processing Center

NSF

IRES project sponsored by NSF Award 2107439