# DynaMine v2, an updated version of the sequence-to-dynamics predictor

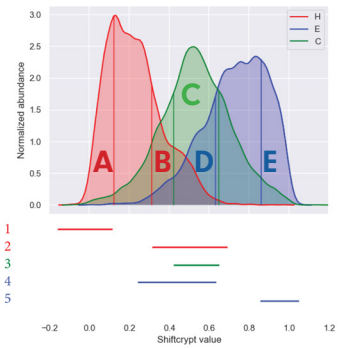
**Jose Gavaldá-García[1,*], Joel Roca-Martínez[1,*], Wim F. Vranken[1]**
* Presenting authors, contributed equally to this work.
1 Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium

## Abstract

We present DynaMine v2, the successor of DynaMine[1], a predictive tool for protein backbone dynamics from protein sequence. For DynaMine v2 we processed nuclear magnetic resonance (NMR) chemical shift data into their ShiftCrypt index[2] to define 5 conformational states, a richer representation to better assess the protein conformation and dynamics. The molecular dynamics (MD) data was then tagged according to those predefined conformational states, and our definition of general dynamics was calculated from it. This was used as the training data for a neural network (NN) dynamics regressor. For now, only protein dynamics have been predicted with promising results, and the 5 conformational states defined in this work will be predicted in the next steps for the project. Our network will also internally predict the ShiftCrypt values for each amino acid, although they will not be presented to the user. We expect to improve generalization and therefore improve our predictions when the network is forced to predict conformational states and ShiftCrypt values

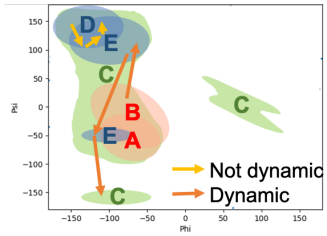
## Conformational states definition

- We use chemical shift data from NMR experiments to gather conformational information




- The ShiftCrypt index is calculated for the available entries and it is used to define 5 conformational states while minimizing the overlap between classes
- The angles from the 5 states are extracted to generate the kernel density estimators (KDE) to classify the MD data

## Data generation




- The KDE regions for the 5 conformational states are used to classify the angles extracted from the MD simulations
- Residues moving from one conformational state to another will have a higher dynamic value

Selection of 100 proteins with different degree of disorder

↓

MD simulations

↓

Sim. Length = 100 ns
Time-step = 1000 ps
Force field = CHARMM36m

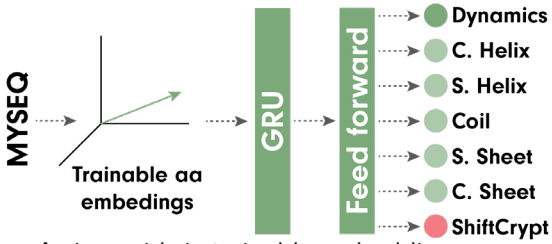
↓

Time-dependent phi and psi angles extraction from each residue

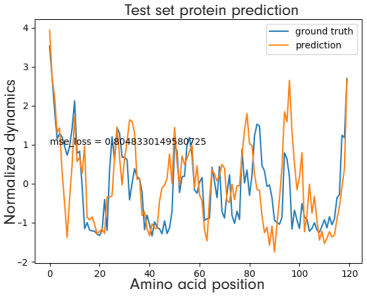
↓

Calculation of the conformational states probabilities and our definition of general dynamics

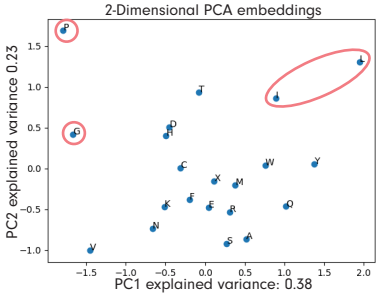
## Neural Network




- Amino acids in trainable embedding
- Gated Recurent Unit (GRU) captures surrounding information
- Feed forward calculates predictions
- Dynamics currently used for training
- Conformational states and ShiftCrypt will be added for generalization




- Test predictions follow general trend
- Working on improved generalization
- Multi-objective training is expected to improve test predictions




- Trained embeddings seem to show biolophysical signal
- Should improve with test set predictions

## Discussion

- Improved dynamics definition reflects protein biophysical behaviour
- Surrounding helix and sheet might enrich conformational definition
- MD-based ground truth should result in better dynamics data
- Neural network is more versatile than the original linear regression and provides more training strategies than the original DynaMine

## Future work

An estimator from Chemical Shift data to Dynamics

- DynaMine v2 sets the building blocks for this estimator

[1] Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., & Vranken, W. F. (2013). From protein sequence to dynamics and disorder with DynaMine. Nature Communications, 4(1), 2741. https://doi.org/10.1038/ncomms3741. http://dynamine.ibsquare.be/
[2] Orlando, G., Raimondi, D., & F. Vranken, W. (2019). Auto-encoding NMR chemical shifts from their native vector space to a residue-level biophysical index. Nature Communications, 10(1), 2511. https://doi.org/10.1038/s41467-019-10322-w

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813239