

Conformine, a predictor of protein Conformational Variability from amino acid sequence

Jose Gavalda-García^{1,2,3}, Joel Roca-Martínez^{1,2,3}, Wim F. Vranken^{1,2,3}

1 Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium. 2 Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, 1050, Belgium.
3 VIB Structural Biology Research Centre, Brussels, 1050, Belgium.

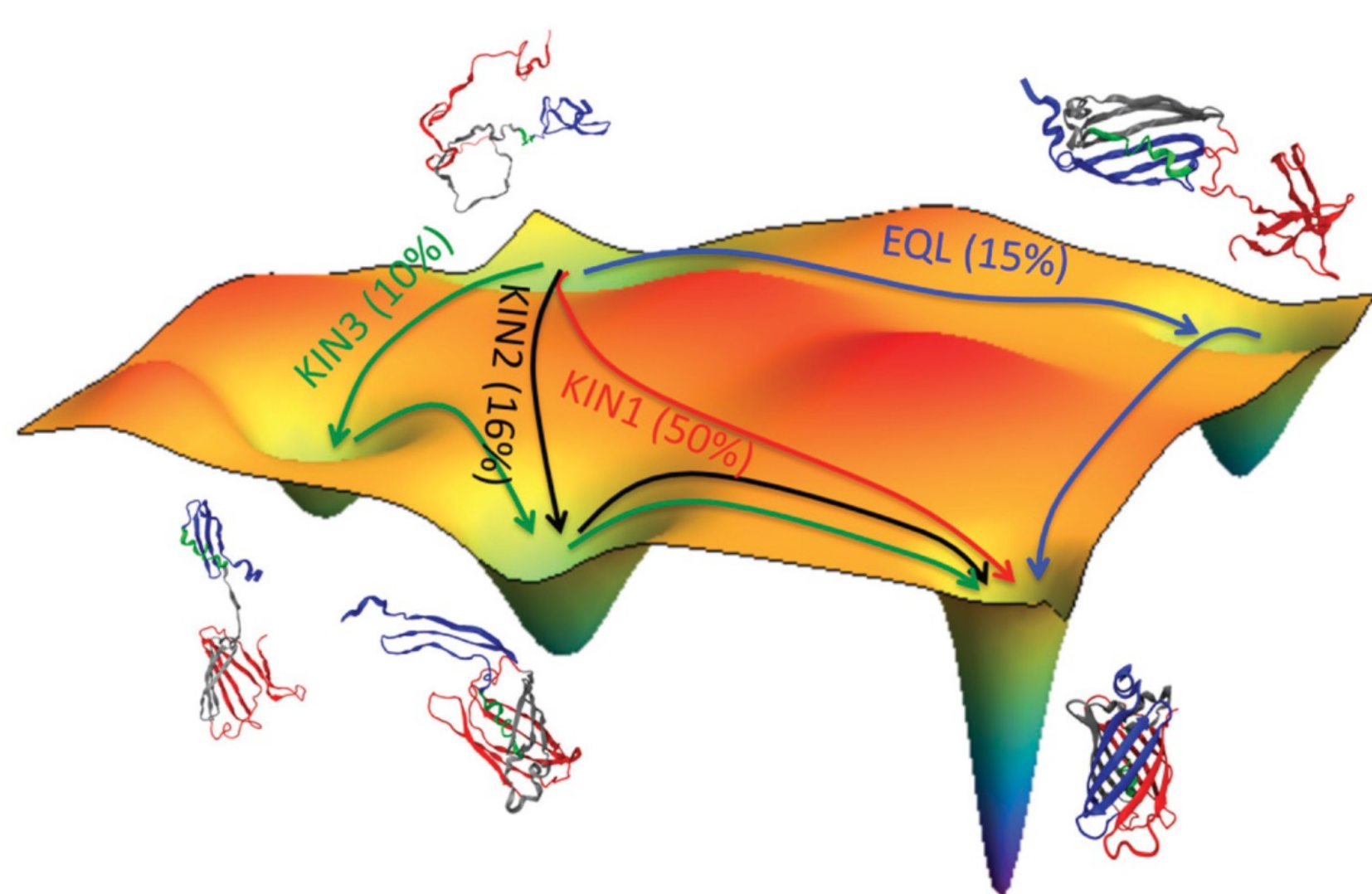
Motivation

Proteins are dynamic and can change conformation over time. We introduce a knowledge-based metric to describe conformational regions for the protein backbone at the residue level, including a means to quantify how often an amino acid residue moves between these regions. This metric provides complementary characterization of the protein in addition to other protein dynamics metrics, such as Random Coil Index. To calculate this metric (Conformational Variability), we performed Molecular Dynamics simulations of 100 proteins with diverse levels of disorder, for which each amino acid was assigned to one of five secondary structure categories for every sample time in the simulation. Then, the Conformational Variability was determined from how often changes in this secondary structure category occurred, calculated with the Frobenius matrix norm.

We then trained an estimator to predict Conformational Variability from amino acid sequence only. This estimator consists of a Long Short-Term Memory neural network which predicts Conformational Variability in multitask with synergetic problems: secondary structure propensity and ShiftCrypt index [1]. We trained on 118 sequences for the main task and secondary structure propensities and on 4500 sequences for the ShiftCrypt index. We performed a 5-fold cross-validation, which indicated a Pearson's correlation for the main task at approximately 0.7 and its corresponding p-value near 0.

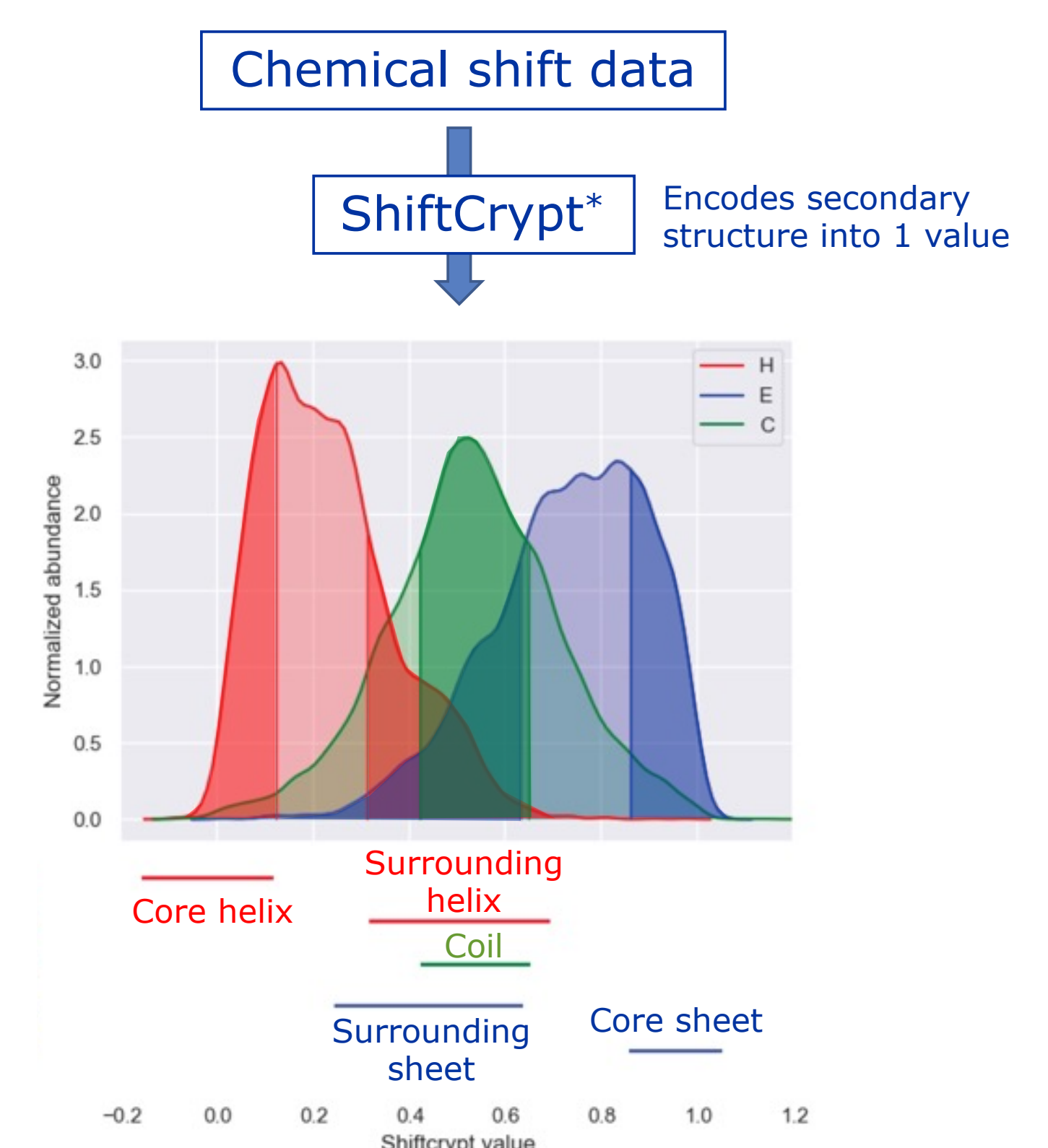
We present this estimator under the name ConforMine, a predictor of protein Conformational Variability from amino acid sequence. This tool will be incorporated in our tool suite "b2bTools", available in PyPI, Anaconda and Bioconda under the same name.

Proteins as non-static molecules

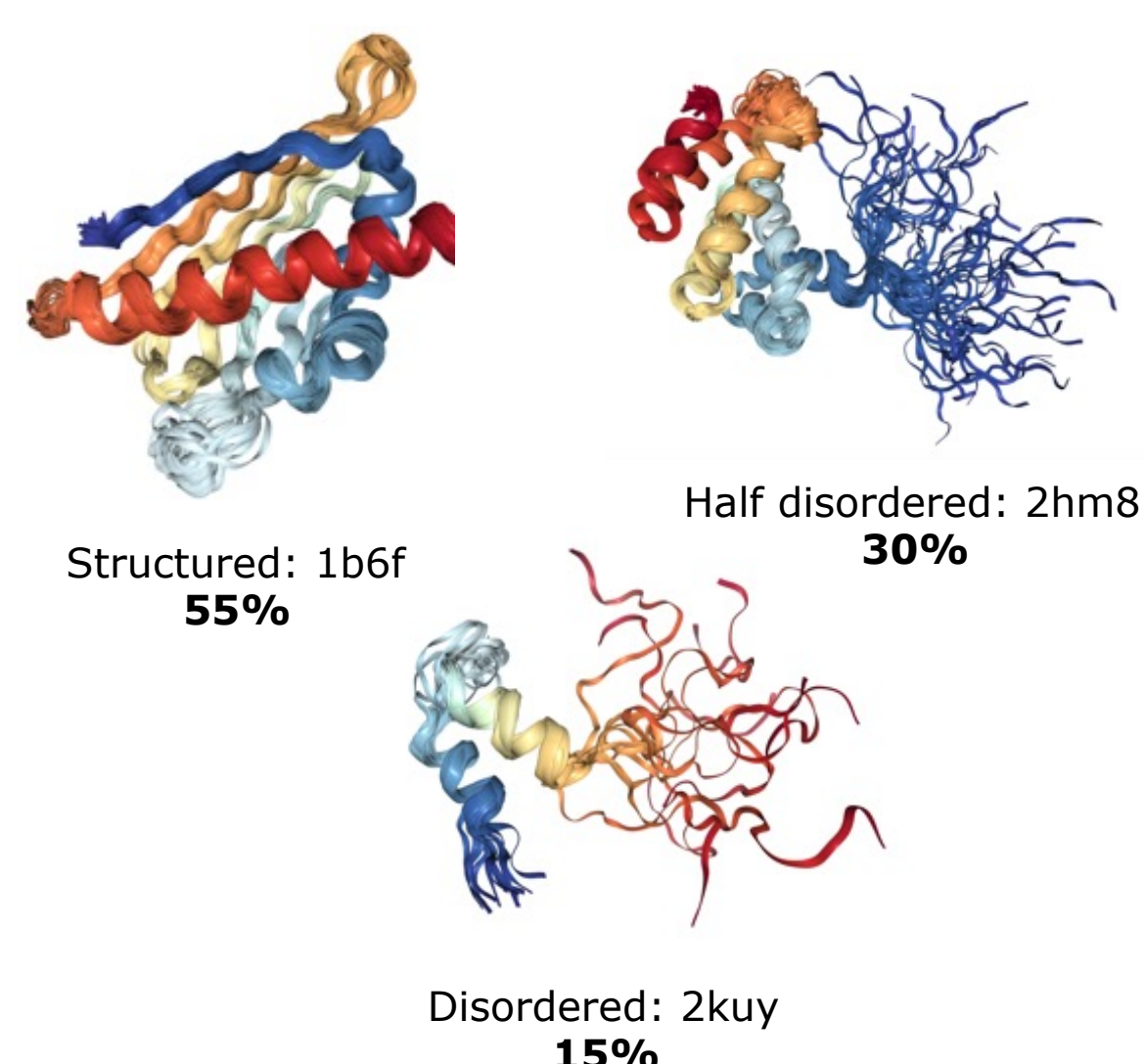


Source: ResearchGate, publication 290324826

- Fold is determined by a thermodynamic landscape
- Multiple stable conformational states
- Changes between conformational states are possible
- Some regions switch conformations often
- Some regions rarely switch conformations
- Metric: **Conformational Variability**



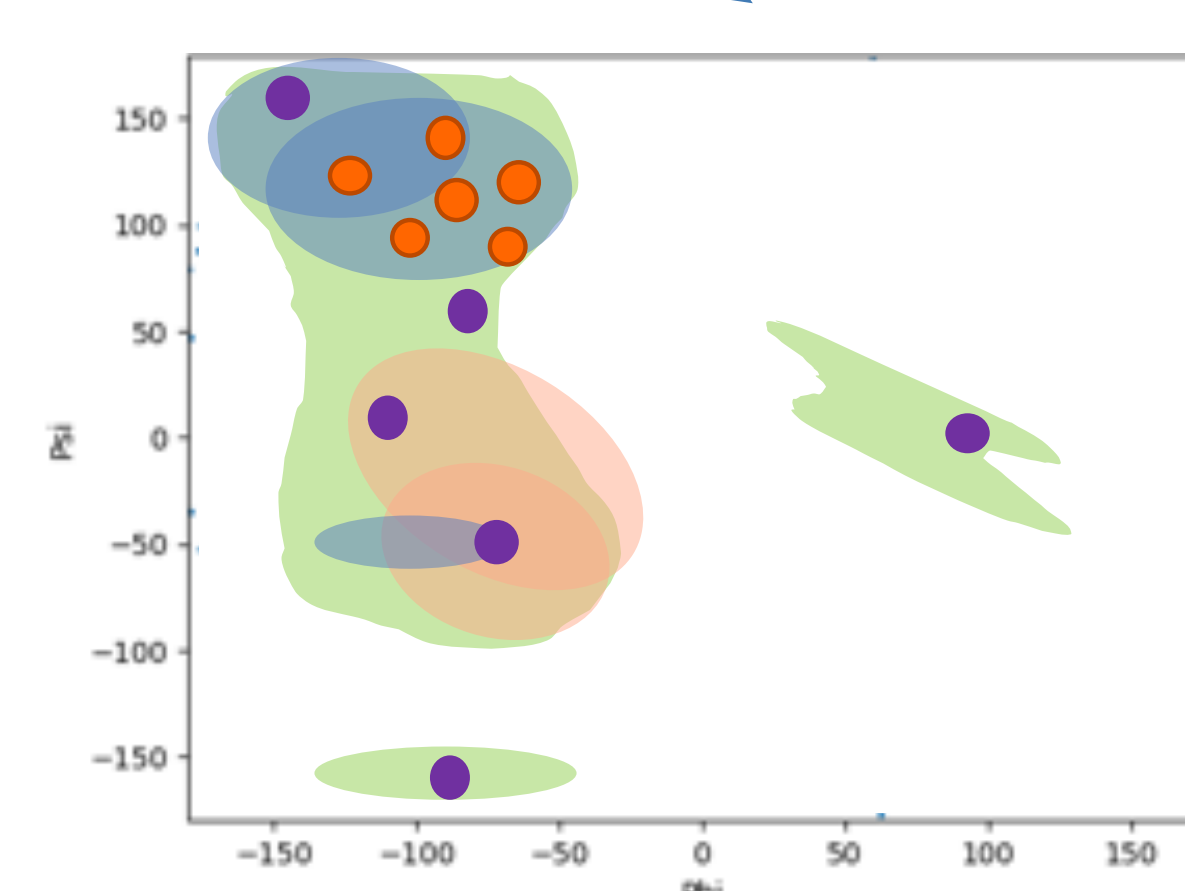
Definition of Conformational Variability



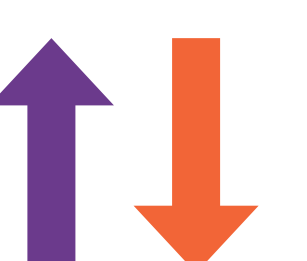
118 proteins
Simulation length: **100 ns**
Time-step: **1000 ps**
Force-field: **CHARMM36m**

Step	Phi (ϕ)	Psi (ψ)
1	-56.3	131.6
2	-30.3	122.1
3	-61.0	172.5
...
98	-60.1	128.9
99	-42.3	130.3
100	-51.7	161.6

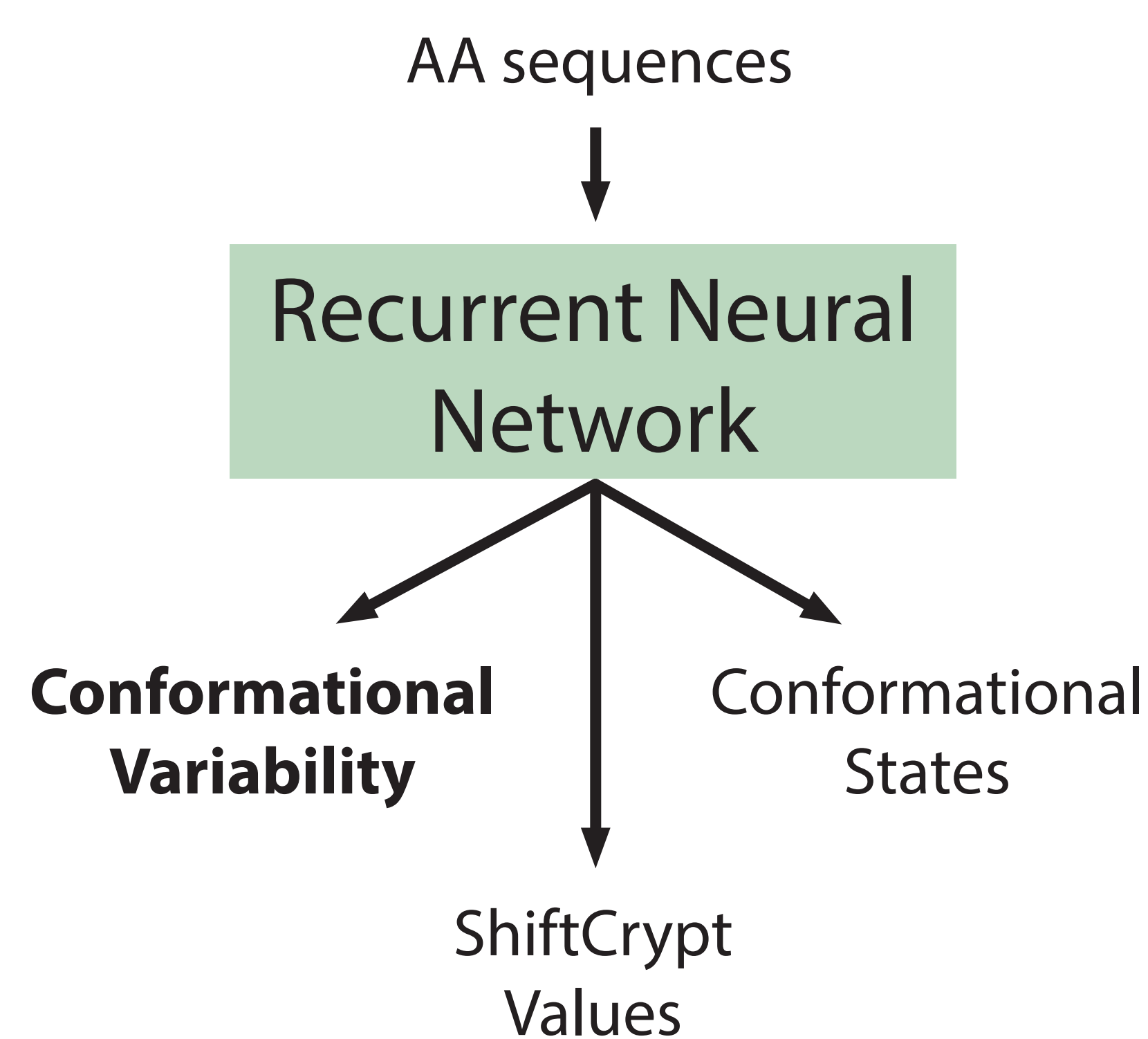
For every protein residue!



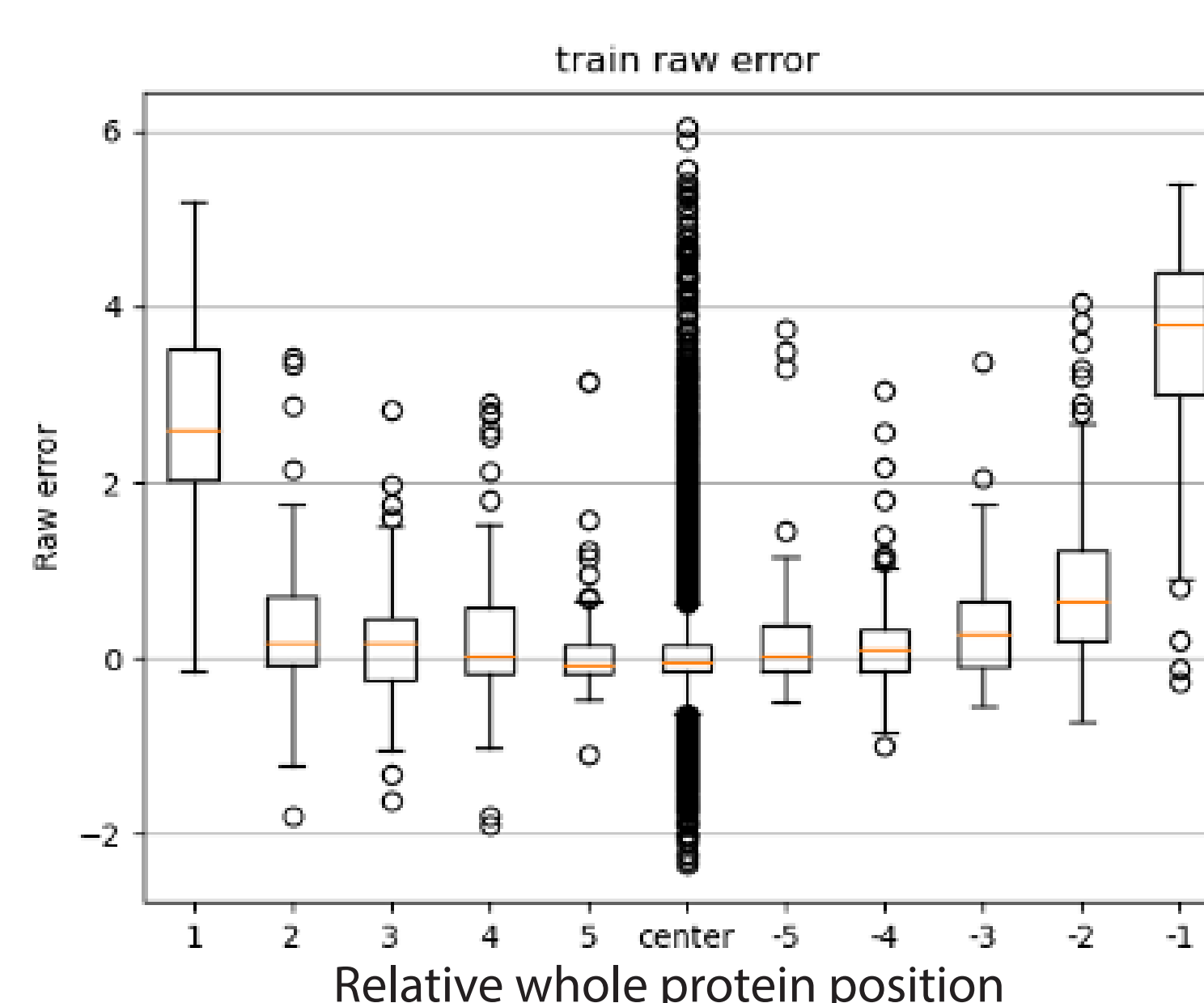
Conformational Variability



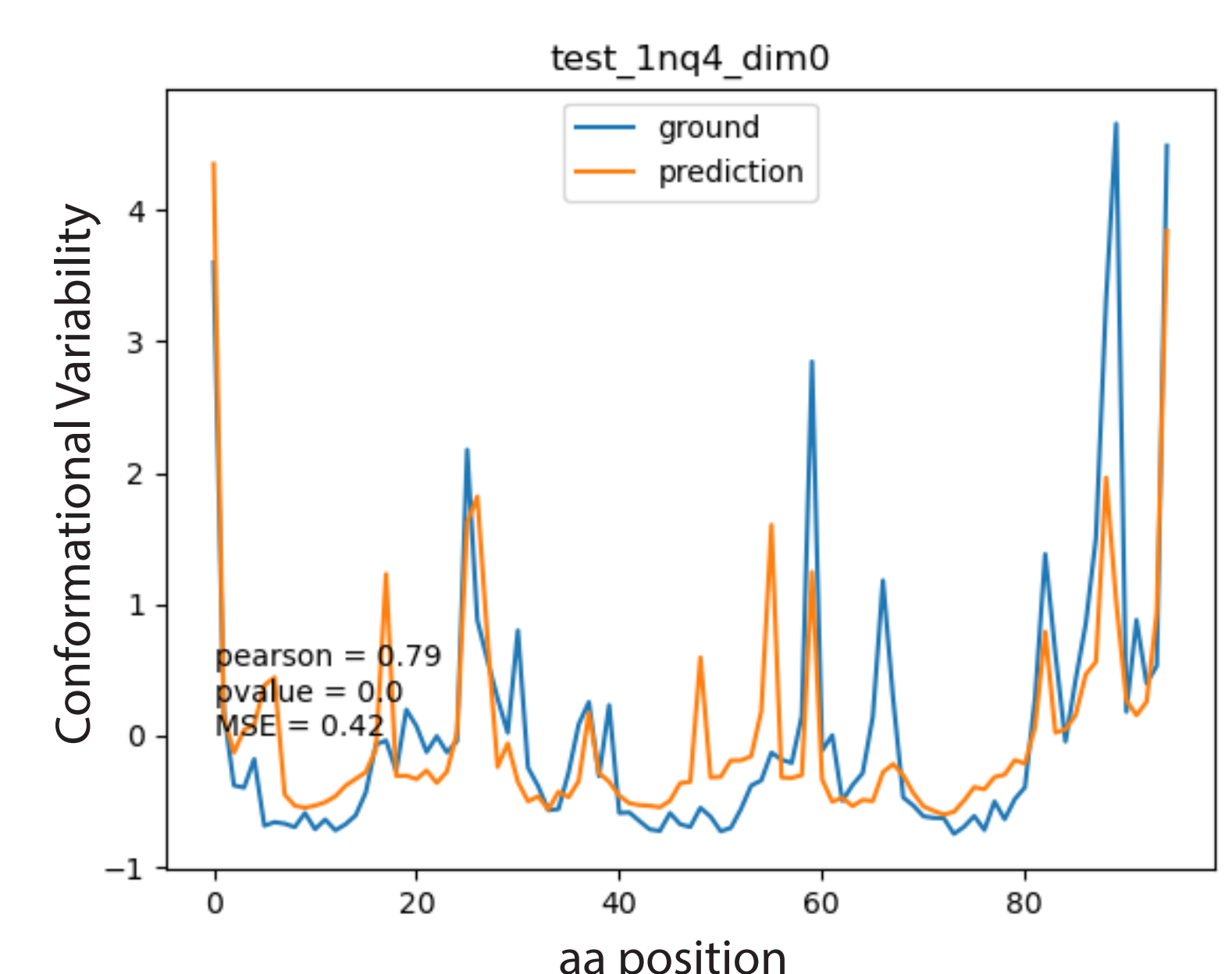
Training and results



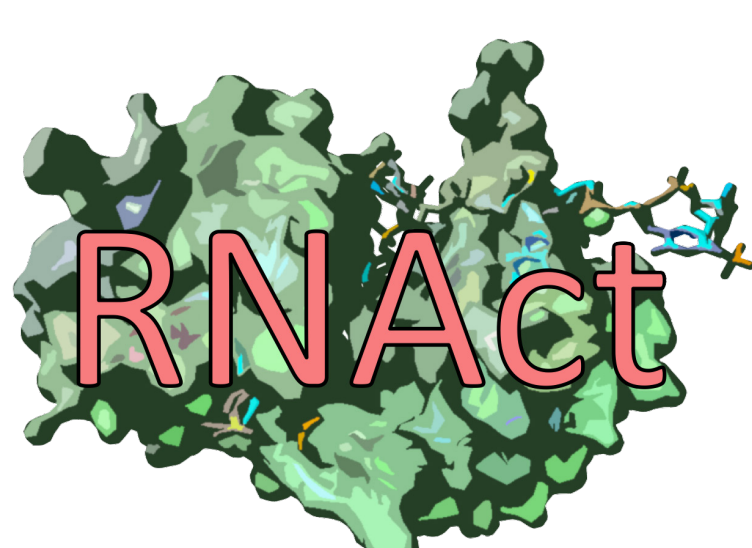
N- and C- extremes errors



After error correction example



References:
1- Orlando, G. et al. ShiftCrypt: a web server to understand and biophysically align proteins through their NMR chemical shift values. <https://doi.org/10.1093/nar/gkaa391>



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