Enhancing Biomolecular Simulations With Hybrid Potentials Incorporating NMR Data

Guowei Qi[1], Michail D. Vrettas[2], Carmen Biancaniello[2], Maximo Sanz-Hernandez[3], Conor T. Cafolla[1], John W. R. Morgan[1], Yifei Wang[1], Alfonso De Simone*[2] and David J. Wales*[1]

1) Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
2) Department of Pharmacy, University of Naples Federico II, 80131 Naples, Italy
3) Department of Life Sciences, Imperial College Longon, South Kensington, London SW7 2AZ, UK
* Corresponding authors: dw34@cam.ac.uk or alfonso.desimone@unina.it

INTRODUCTION

Biomolecular simulations are very useful methods to overcome the inherent limitations of experimental structural elucidation techniques in the study of dynamics of flexible proteins. Previous works have shown that the performance of these computational approaches can be improved through the definition of hybrid restraint potentials, where molecular mechanics force fields are combined with experimental data. Using NMR chemical shifts (CSs), we developed a hybrid restraint potential based on NapShift, a machine learning model trained to predict chemical shifts from protein sequence and structure. The functioning of our method was evaluated in basin-hopping (BH) global optimization, energy landscape (EL) sampling and molecular dynamics (MD) simulations with promising results.

METHODS

We developed **Napshift,** a program in python based on simple feed-forward, fully-connected, artificial neural networks (ANNs) to predict CS values **for each backbone atom (N, C, Cα, Cβ, H and Hα)** using protein sequence and dihedral angles of tripeptides as inputs.

NapShift™

RCSB PDB PROTEIN DATA BANK ⋎ BMRB Biological Magnetic Resonance Data Bank

2987 proteins for training and validation

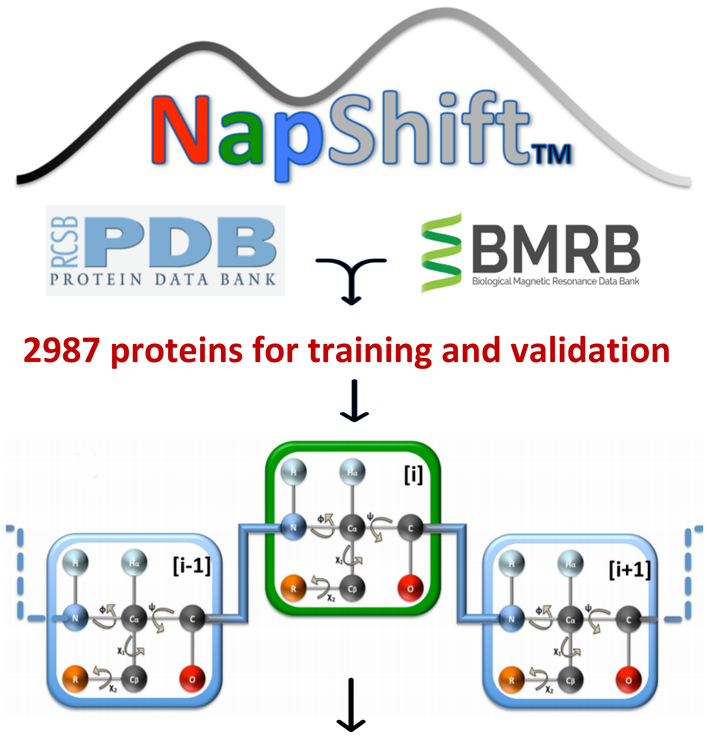

FIG.1: Schematic representation of input data and Napshift ANNs architecture

Our method express CSs as differentiable functions of cartesian coordinates and allow us to compute an energy penalty based on the difference between the predicted values and the experimental CSs and the corresponding restraint forces to pair with empirical force fields.

$$V^{\mathrm{CS}} = K \sum_{i}^{N_{\mathrm{res}}} \sum_{j}^{6} (\delta_{ij}^{\mathrm{experimental}} - \delta_{ij}^{\mathrm{back\text{-}calculated}})^2 \quad (1)$$

In Eq. (1) V^{CS} is the Napshift-based hybrid restraint potential energy

CONCLUSIONS

- Napshift is an accurate CS prediction method.
- NapShift ANN functions are fully differentiable with respect to atomic coordinates, allowing to use this method for structural refinement.
- The implementation of Napshift hybrid potential into a molecular mechanics force field improves the efficiency of BH global optimization, EL analysis and MD simulations.

RESULTS

In comparison with other CS predicting methods, Napshift shows improved accuracy on 250 PDB files used as the test set.

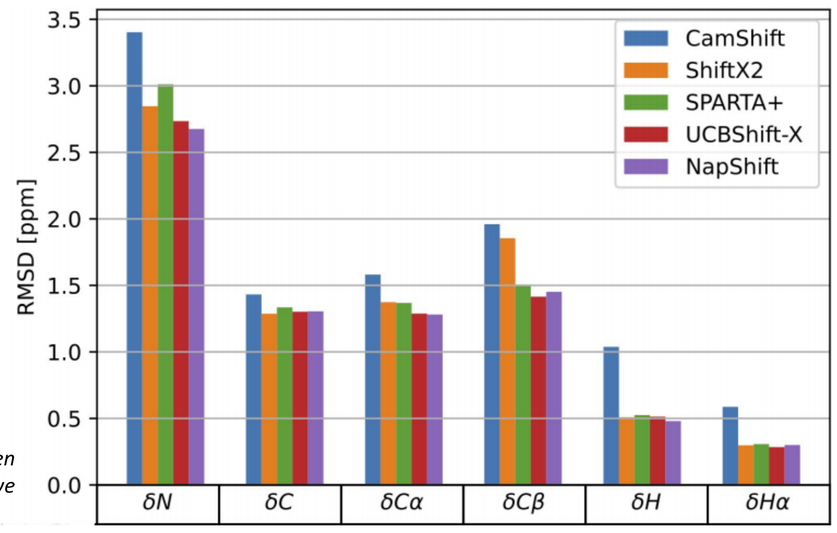

FIG.2: Comparison of the RMSD between predicted and experimental CS for five different prediction methods

The performance of the NapShift hybrid restraint potential was tested in BH global optimization, EL analysis and in MD simulations. The following results come from one of different systems explored: tryptophan zipper 1 (trpzip), a 12-residue peptide that folds into a β-hairpin with specific packing of tryptophan sidechains. Using trpzip distorted conformations as the starting point, the NapShift CS restraints implementation was able to lead to better structure prediction via BH and increased local stability within MD simulations.

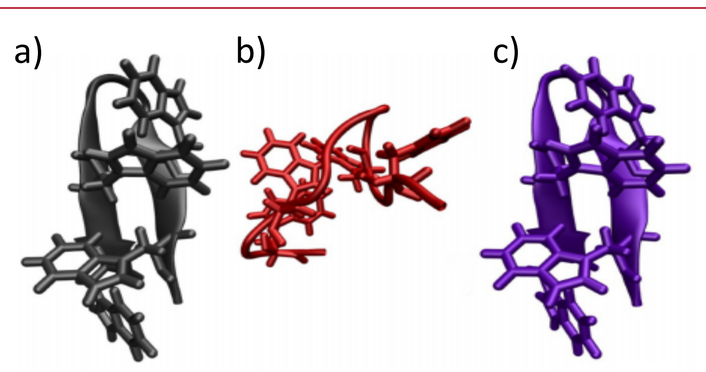

FIG.3: a) The experimental structure of trpzip (PDB code 1LE0), followed by minimum energy structures located by BH global optimization using the GMIN program b) with no experimental restraints and c) with Napshift hybrid potential addition .

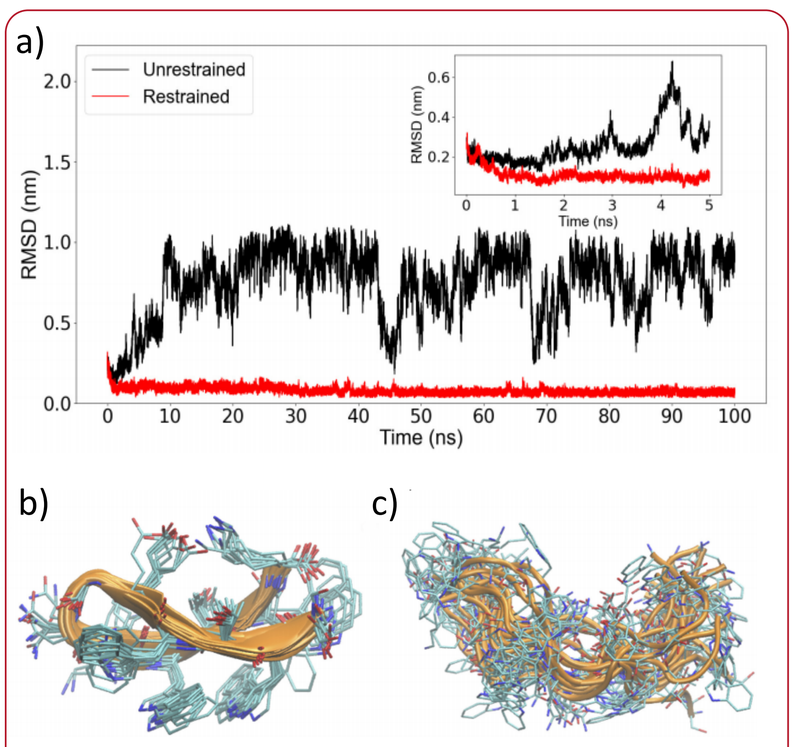

FIG.5: MD simulation results for trpzip with the GROMACS software. a) RMSD to experiment structure over restrained and unrestrained 100 ns MD simulations. Ensembles of conformations sampled in b) restrained and c) unrestrained simulations..

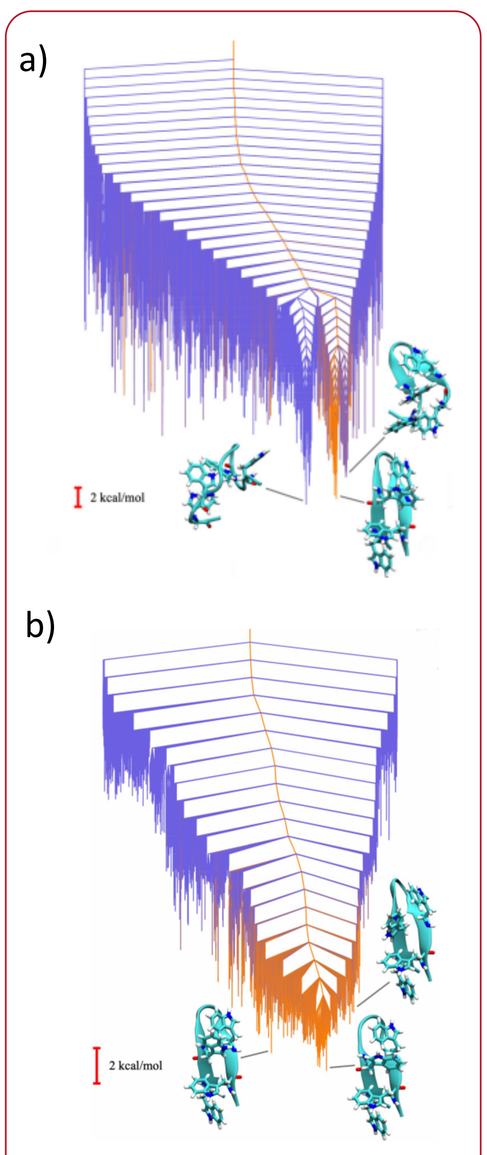

FIG.4: Disconnectivity graphs representing the energy landscape of trpzip a) without restraints and b) with Napshift hybrid potential. The visible structures are selected minima from the corresponding energy landscape. Results derived from the discrete path sampling scheme with the OPTIM2 and PATHSAMPLE3 programs

REFERENCES

Qi Guowei et al., (2022). "Enhancing Biomolecular Simulations With Hybrid Potentials Incorporating NMR Data." Submitted for publication.

The NapShift method is freely available at https://github.com/vrettasm/NapShift