DI
GA
ML
BI

DISCOVERY OF SLOW KINETIC MODES FROM MOLECULAR SIMULATION TRAJECTORIES

Allocation: Exploratory/30 Knh
PI: Andrew Ferguson[1]

[1]University of Chicago

EXECUTIVE SUMMARY

Slow modes discovery is an important topic in molecular simulation because it can help extract meaningful kinetic information for the analysis of dynamic processes and can serve as good collective variables for enhanced sampling. In this project, the researcher developed a new deep learning-based method called "state-free reversible VAMPnets" that is able to discover hierarchical nonlinear slow modes accurately with much lower computational cost than its predecessors. Moreover, this method can be well integrated with the powerful Markov model to improve its kinetic resolution significantly. This method is expected to be very helpful for understanding dominant kinetic transitions in biomolecular processes and to be useful for drug discovery.

RESEARCH CHALLENGE

Identifying the collective motions governing the longtime behaviors of biomolecules such as DNA and proteins is vital in understanding and engineering the behavior of these molecules of relevance to industrial catalysis, human health, and clean energy. This work establishes new basic science techniques combining applied mathematics and deep learning to "harness the data revolution" and perform data-driven inference of collective motions from molecular dynamics simulation trajectories. Existing methods for slow mode recovery such as time-lagged independent component analysis, or TICA, are only able to discover linear slow modes. Further, they require expert knowledge and hyperparameter tuning (*e.g.*, kernel TICA) or fail to discover multiple hierarchical slow modes (*e.g.*, time-lagged autoencoders or variational dynamics encoders). It is an outstanding challenge to develop a simple, robust, flexible, accurate, and efficient method to extract nonlinear hierarchical slow modes from simulation data.

METHODS & CODES

Recent advances in deep learning have made it a powerful tool to solve problems in many different fields. A variational principle developed for slow mode recovery has proved to be successful in discovering slow modes given good basis sets. The researcher's work is founded on a combination of these two powerful ideas: Deep-learning models can naturally estimate excellent basis sets for the variational principle, whereas a variational principle defines a natural objective function with which to train deep neural networks to achieve this goal. Orthogonality and hierarchy of the modes are naturally imposed in the application of the variational principle. The new method surpasses previous methods since it is simple, robust, accurate, efficient, and insensitive to feature selection and scaling. An open source package with extensive documentation and examples has been developed to make the method freely available to the machine learning and molecular simulation community (https://github.com/hsidky/srv).

RESULTS & IMPACT

The method has been tested on four different systems, including two toy models and two realistic molecular systems. Where available, the results show excellent agreement with theoretical analysis or previous calculations. This approach led to the development of kinetic models for protein folding at unprecedented time resolution and the establishment of highly efficient molecular simulators that, once trained, can perform molecular simulations at a six orders of magnitude lower cost than conventional approaches. These advances are valuable for better understanding and engineering of molecular machines for clean energy production as well as drugs and vaccines to improve human health.

WHY BLUE WATERS

This research requires running many long, large simulations; performing computationally intensive data processing; and training large numbers of machine-learning models with different training parameters. Access to Blue Waters allowed these computations to be performed at the scale and parallelism necessary to support this research. The Blue Waters staff were also invaluable in helping to improve the efficiency, performance, and workflow of the computations.

PUBLICATIONS & DATA SETS

W. Chen, H. Sidky, and A. L. Ferguson, "Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets," *J. Chem. Phys.*, vol. 150, p. 214114, 2019.

State-free reversible VAMPnets, GitHub, May 2019. [Online]. Available: https://github.com/hsidky/srv

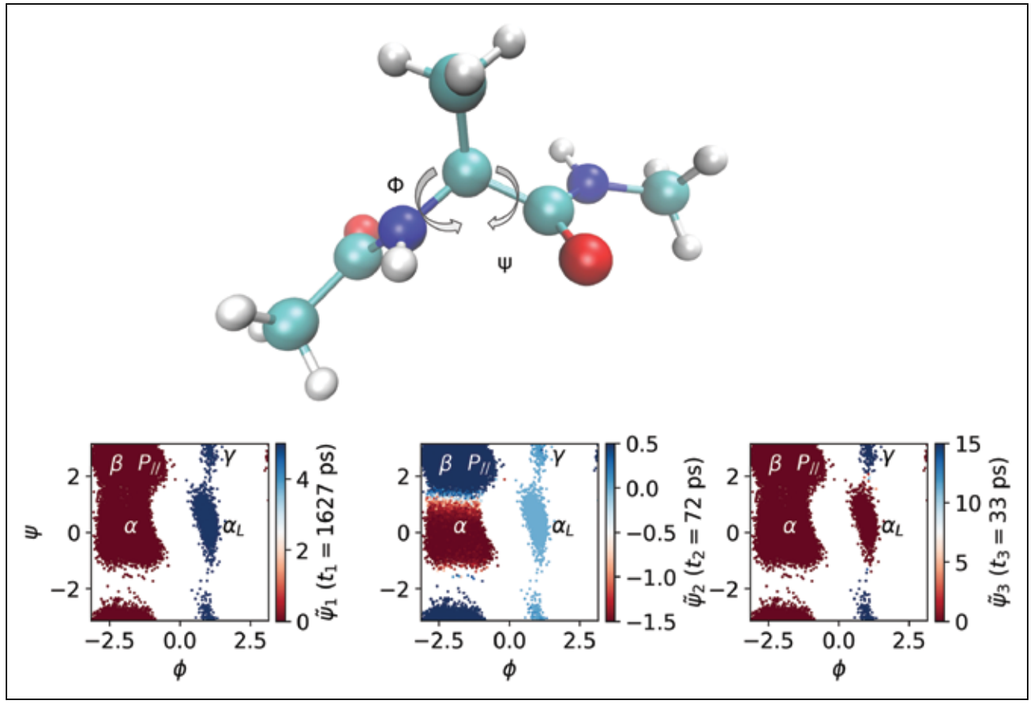

Figure 1: The molecular structure of alanine dipeptide and its three leading slowest kinetic modes: These three collective variables capture dominant transitions between the metastable basins marked in the image to providing molecular understanding of folding. The timescale associated with each transition is listed alongside each panel.