GA

TN

USING SPECTROSCOPIC DATA AND MOLECULAR SIMULATIONS TO ESTIMATE HETEROGENEOUS ENSEMBLES: HOW TO STUDY COMPLICATED, FLEXIBLE PROTEINS WHEN EXPERIMENTAL DATA ARE LIMITED

Jennifer M. Hays, University of Virginia
2017–2018 Graduate Fellow

EXECUTIVE SUMMARY

Flexible proteins play critical roles in cellular transport but are extremely challenging to model at high resolution. Experimental techniques such as double electron–electron resonance (DEER) report on conformational heterogeneity but are sparse over atom coordinates. The PI developed a methodology, Bias-Resampling Ensemble Refinement (BRER), to incorporate multimodal DEER data into molecular dynamics (MD) simulations to obtain high-resolution, experimentally validated models of flexible proteins. The results from ensemble simulations of unbound syntaxin-1a, a protein involved in the formation of SNARE complexes, which drive neuronal vesicle fusion, show that the PI's method better reproduces experimental data than current state-of-the-art methods. Specifically, the methodology promotes sampling of significant backbone conformational change, unlike any other existing methods. In addition, BRER simulations of the soluble domain of syntaxin-1a revealed a previously unresolved open conformation of syntaxin-1a.

RESEARCH CHALLENGE

It is difficult to study flexible proteins that play critical roles in infectious disease and cellular transport because so many states contribute to their conformational ensembles. High-resolution models of these systems are important for innovation in drug development and for answering fundamental questions in biophysics. It is challenging to develop atomic-resolution models using experiments alone because experimental techniques that report on heterogeneity often do so for only a few atomic degrees of freedom. Therefore, new hybrid methods are needed that include experimental and computational approaches to understand these systems.

METHODS & CODES

The PI developed a new method for incorporating distributional data into MD simulations (Hays, Cafiso, and Kasson, 2019; see Publications & Data Sets below). This was done using the software package gmxapi, a Python interface for the GROMACS MD engine. The Python package for BRER simulations is freely available at https://github.com/jmhays/run_brer and the gmxapi code is at https://github.com/kassonlab/gmxapi.

RESULTS & IMPACT

The PI developed both a method and an open source software package to integrate sparse experimental data using MD simulation to better understand flexible proteins. This will enable scientists, specifically spectroscopists, to study systems that would be too heterogeneous and complicated for standard refinement methods.

WHY BLUE WATERS

The researcher has run multiple sets of ensemble simulations on Blue Waters to test the novel method and accompanying software package. A petascale, multi-GPU resource like Blue Waters was absolutely essential for completing both testing and production of all-atom ensemble simulations, which demanded over 200K node-hours over the course of the Graduate Fellowship.

The Blue Waters staff were also critical both in terms of the software development and in professional development. The PI learned to compile and run complicated software packages on CRAY systems. Further, she learned a great deal about general high-performance computing through the online and NCSA Symposium workshops, including how to utilize singularity containers, which have become an essential part of the researcher's laboratory workflow development.

PUBLICATIONS & DATA SETS

J. M. Hays, D. S. Cafiso, and P. M. Kasson, "Hybrid refinement of heterogeneous conformational ensembles using spectroscopic data," *J. Phys. Chem. Lett.*, vol. 10, no. 12, pp. 3410–3414, 2019, doi: 10.1021/acs.jpclett.9b01407.

Jennifer M. Hays obtained her Ph.D. in biomedical engineering from the University of Virginia in November 2019 and worked under the direction of Peter M. Kasson.

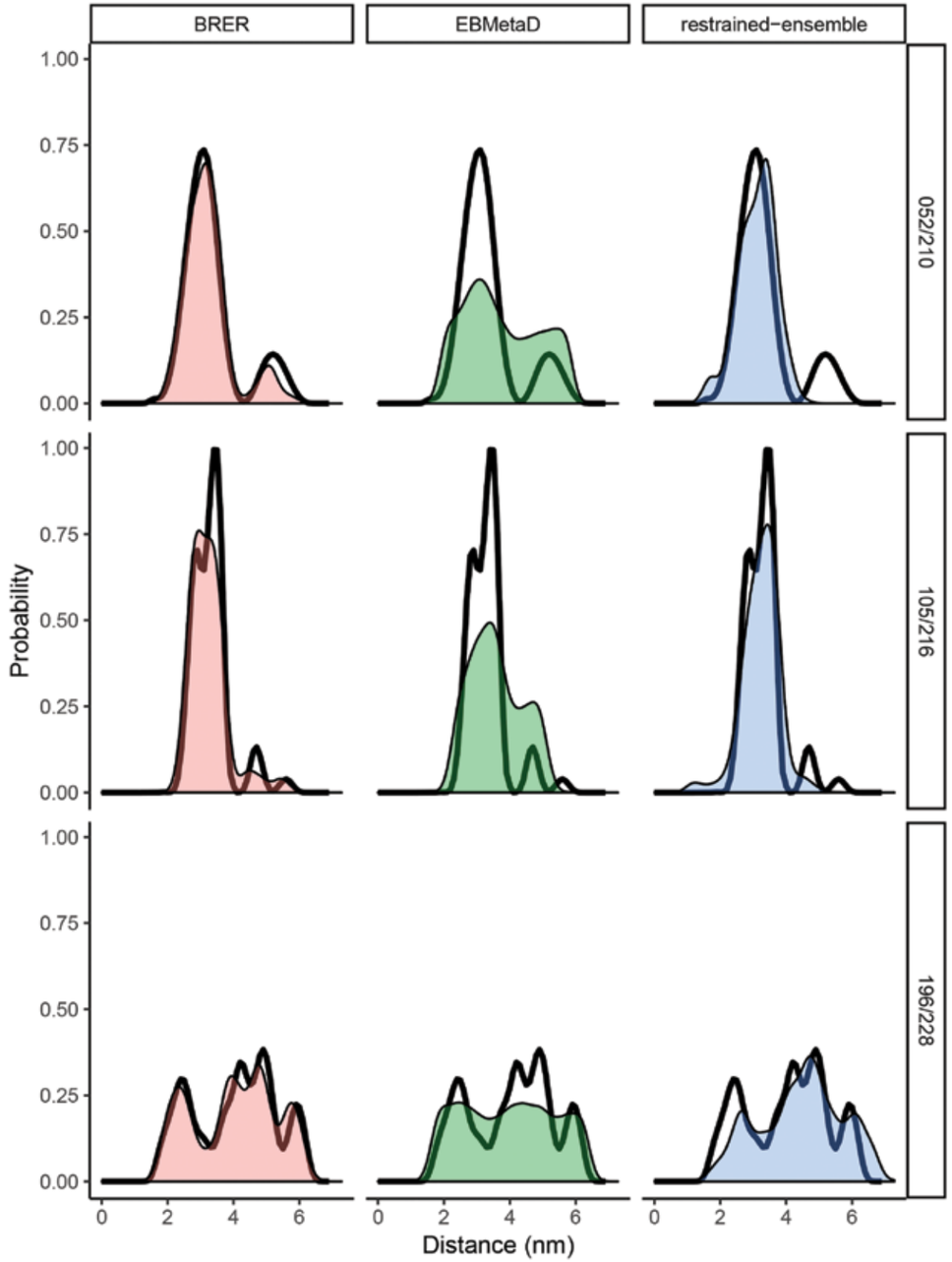

Figure 1: Ensemble-MD simulations refined using BRER better reproduce three experimental distributions (52/210, 105/216, 196/228) than two other preexisting refinement methods: EBMetaD and restrained-ensemble MD.