GA

TN

REFINING THE CONFORMATIONAL ENSEMBLES OF FLEXIBLE PROTEINS USING SIMULATION-GUIDED SPECTROSCOPY

Jennifer Hays, University of Virginia
2017-2018 Graduate Fellow

EXECUTIVE SUMMARY

Flexible molecular recognition is a common paradigm in immunity and infection; many pathogens have proteins that are structurally flexible and/or highly tolerant of mutations but still effectively bind to human cells. Determining the structural basis of this recognition experimentally is challenging because of the multitude of structures involved. We have developed a computational methodology that provides a way to select the best experiments to measure these structures and subsequently combine them in an integrated model. The conformations of many important flexible proteins are still not well understood; this methodology will allow researchers to characterize systems that are currently too difficult to understand with existing refinement techniques. A systematic approach to refining these flexible receptor–ligand complexes would help elucidate the fundamental physical principles of receptor–ligand binding and promote better drug design for infectious disease.

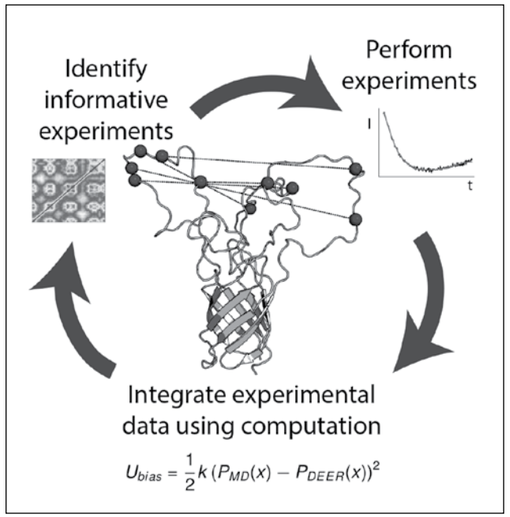

Figure 1: Iterative refinement of flexible conformational ensemble using simulation-guided spectroscopy. We measured experimentally derived distance distributions and then integrated them into molecular dynamics simulations using restrained-ensemble or bias-resampling ensemble approaches. The resulting hybrid ensemble was then analyzed using mRMR to prospectively identify a set of optimal experiments to perform.

RESEARCH CHALLENGE

Multistructured proteins play critical roles in infectious disease but are often difficult to characterize; experimental techniques often capture either a subset of the structures at high resolution or a more complete set of structures at low resolution. Double electron–electron resonance (DEER) spectroscopy is a powerful tool for measuring multiple structures, but these experiments are low-throughput, which means that it is critical to select only the very best, most informative experiments. We have developed a model-free simulation-based approach for selecting a set of optimal DEER experiments and integrating the resulting data to estimate the full set of structures at high resolution (Fig. 1). This combined experimental and computational approach will allow us to rapidly study and even redesign flexible proteins and thus accelerate the study and treatment of infectious disease.

METHODS & CODES

DEER allows measurement of the distances between nearby pairs of chemically labeled amino acids in a protein. An ideal set of these pairs would have two properties: each selected pair should resolve as many other distances in the system as possible, and each selected pair should resolve distances that are distinct from those determined by each of the other measured pairs. These criteria are optimally satisfied by selecting pairs using the information–theoretic criteria of maximum relevance and minimum redundancy (mRMR) [1].

We incorporated the analyzed DEER data from sets of optimal experiments into our scheme as distance distributions. These distributions are used to drive restrained-ensemble Molecular Dynamics (MD) simulations in which the simulation distance distribution is biased toward the experimental distribution. We performed simulations using a modified version of the restrained-ensemble method described in [2]; code to perform these simulations is available at https://github.com/kassonlab/restrained-ensemble.

For DEER-derived distributions with well-separated probability modes, such as would happen for a system with distinct open and closed states, current state-of-the-art methods of incorporation fail to fully integrate the data. As a result, we have also developed a new methodology for integrating the spectroscopic data into MD simulations (Fig. 2).

RESULTS & IMPACT

We tested our iterative refinement methodology on the Neisserial virulence-associated protein Opa60, selecting a set of optimal experiments using the mRMR criteria. We performed those experiments and then incorporated the resulting experimental data into an MD simulation. In comparison to experiments selected according to current structure-guided methods, simulation-guided spectroscopic measurements are significantly more informative. By integrating the data in restrained-ensemble MD, we were able to obtain not only a refined ensemble of apo Opa60 but also structural insight into how Opa60 engages cellular receptors. After two rounds of mRMR-guided refinement, we were able to identify specific residue–residue interaction patterns that were not determined in a structure-guided refinement approach. Further experiments revealed the subset of conformations responsible for binding Opa's target receptor.

We are currently expanding this iterative methodology to include more robust ways of incorporating distance distributions into MD simulation. In preliminary studies of the open/closed conformational equilibria of syntaxin-1a, a protein involved in SNARE complex formation in synaptic exocytosis, we have successfully and robustly incorporated DEER experimental data that resisted previous methods of incorporation. Specifically, these bias-resampling simulations facilitate the study of transitions between the open/closed conformations, whereas alternative methods prohibit sampling of transitions and, in the case of syntaxin-1a, sampling of the open state.

Many bacterial pathogens improve their ability to evade the human immune response by having proteins that are structurally flexible and/or highly tolerant of mutations, thus preventing recognition by immune cells. However, these proteins are also able to engage receptors on nonimmune cells and trigger infection. By selecting sets of optimal experiments and incorporating these informative results into an estimate of the conformational ensemble, experimentalists can now study biological systems that were once prohibitively complex or expensive. This systematic approach to refining flexible receptor–ligand complexes helps elucidate the fundamental physical principles of receptor–ligand binding and promotes better drug design for infectious disease.

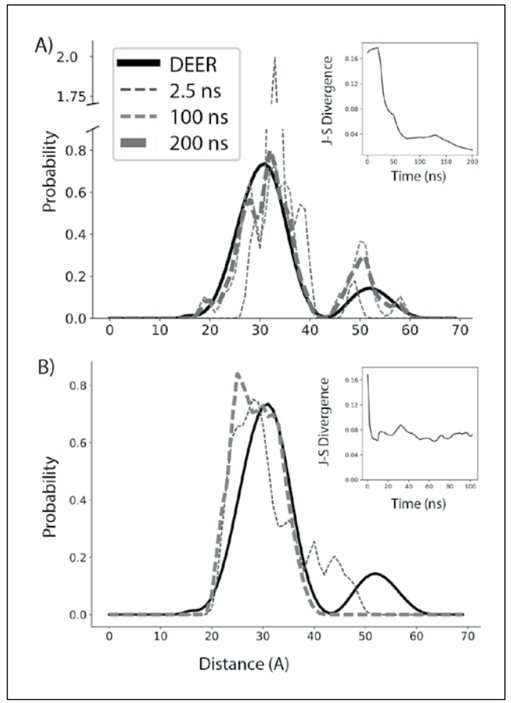

Figure 2: Novel methods of incorporating DEER data facilitate sampling of both the open and closed states of syntaxin-1a. Bias resampling of DEER distributions in ensemble simulations samples both peaks of the distribution (A), while current state-of-art methods for incorporating DEER data do not permit sampling of the far peak (B). The difference between the MD and DEER distributions is quantified as Jensen–Shannon divergence in the insets of (A) and (B).

WHY BLUE WATERS

Access to Blue Waters has greatly accelerated the time-to-completion of this project. An enormous amount of molecular dynamics sampling is required to capture the full set of structures of a flexible system; that sampling would not have been possible without an petascale system like Blue Waters. We have run massively parallel MD simulations, scaling to multiple nodes for a single ensemble member and to many ensemble members.

Jennifer Hays is in the fourth year of a PhD program in biomedical engineering at the University of Virginia (UVA). She is working under the direction of Peter Kasson of UVA and Uppsala Universitet and expects to receive her degree in July 2019.