Molecular Simulation Methods for Traversing Barriers In Time and Complexity

Overcoming the barrier presented by the difference between biologically relevant timescales and timescales that can be readily simulated using molecular dynamics (MD) simulations is the motivation behind many innovations in both atomistic enhanced sampling methodologies and various coarse-graining techniques. A less commonly addressed barrier inherent to molecular simulations is the gap between the degree of complexity that can be captured using molecular simulations and human limitations for interpreting high dimensional data. For example, simulations of molecular systems with many thousands of degrees of freedom are commonplace, yet humans are notoriously bad at estimating simple 3 dimensional quantities like volume. Although the calculation of known experimental observables from molecular simulations is crucial, realizing the full potential of molecular simulations must also involve their role as a source for novel mechanistic insights that motivate further experimental exploration. Machine learning (ML) methods provide an ideal framework for developing methods that identify previously unknown high dimensional mechanistic patterns in molecular simulations. This talk will encompass some of the work that I have done toward surmounting barriers in both time and complexity, including: a coarse-grained model for amyloid fibril formation capable of simulating fibril formation on the timescale of hours for thousands of monomers, multiple advances in extending the reach of the Milestoning methodology for extracting long-time global kinetics from shorter MD simulations, and using ML methods like support vector machines (SVMs) for both prediction of reaction products and agnostic identification of the most characteristic molecular degrees of freedom leading to different reaction products in multiple pathway reaction dynamics (with application to acetaldehyde photodissociation).