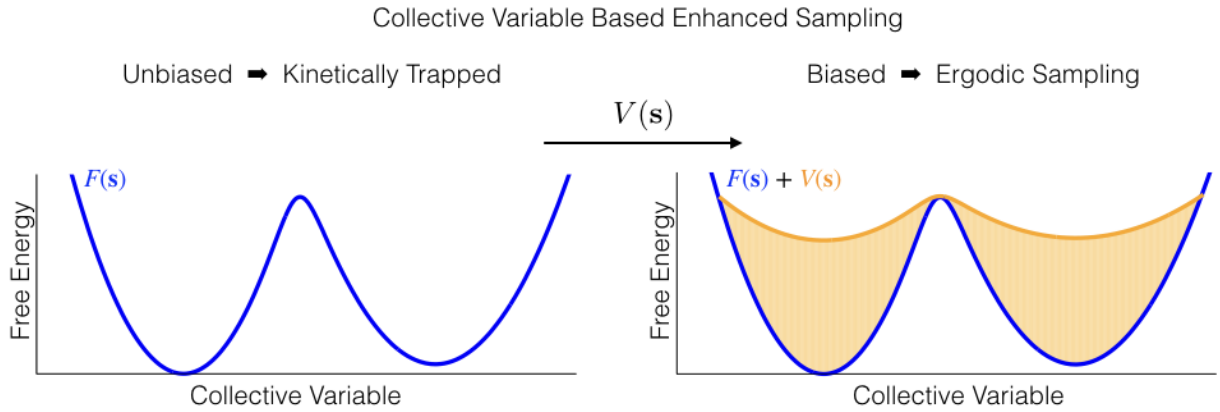
Modeling the long-timescale behavior of complex dynamical systems is a fundamental task in the chemical sciences. In principle, molecular dynamics (MD) simulations allow to probe the spatiotemporal details of molecular processes. Still, the so-called Boltzmann sampling problem severely limits their usefulness in practice. The Boltzmann sampling problem comes from the fact that a typical free energy landscape consists of many metastable states separated by free energy barriers much higher than thermal energy. On the timescale one can simulate, barrier crossings are rare events, and the system remains kinetically trapped in a metastable state. Although a slow reaction pathway can characterize the rare transition between two metastable states, driving the system out of equilibrium along the path is infeasible with unbiased MD methods.

One way to alleviate the Boltzmann sampling problem is to employ enhanced sampling simulations in which fluctuations of few critical degrees of freedom, called collective variables (CVs), are enhanced by an external bias potential. The performance of such methods depends on the quality of the CVs. Effective CVs should discriminate between the relevant metastable states and include essential degrees of freedom. Typically, the CVs are selected using intuition. However, it may not be trivial to find CVs that quantify the crucial characteristics of a rare event. Methods attempting to construct CVs directly from MD data are limited to learning only from unbiased MD simulations. Thus, if the MD data cannot describe the long timescale events, the CVs will lack important information. Such methods are mainly tested on pre-existing data sets or simple models for which we already know the slow CVs.



Currently, no method can learn from biased MD data, limiting the applicability of the existing methods only to short timescales. Therefore, for more complex systems in which long timescales are involved, we need to generate the data based on enhanced sampling techniques to reach the interesting long timescales. In this project, we consider crucial problems related to the methodology of estimating CVs: How to construct the CVs without resorting to system-specific expert knowledge? Is it possible to construct the slow CVs directly from biased data from enhanced sampling simulations?

By improving our recent method, multiscale reweighted stochastic embedding (MRSE) [*J. Phys. Chem. A* **2021**, 125], we will be able to provide statistical estimates of the slow CVs inferred from the biased data from enhanced sampling simulations. We will be able to use our learning method iteratively at different times of biased simulations to estimate the slow CVs and then drive the sampling. Overall, we plan to devise a method that will be able to learn the slow CVs from the biased data in a near-blind manner, making it accessible to many users without detailed knowledge about enhanced sampling theory. We expect the method will have a significant impact on the current state of the MD community and be applicable to long-timescale processes in chemistry, physics, and biology.