**AdaptiveMD and PySFD: Large-Scale Computations and Analyses of Molecular Dynamics Simulations**
John Ossyra, Sebastian Stolzenberg, Eugen Hruska, Cecilia Clementi, Frank Noe, Jeremy Smith

Oak Ridge National Laboratory

ossyrajr@ornl.gov

**Abstract**

It is a great challenge for regular Molecular Dynamics (MD) simulations to reach the length and time scales of biologically relevant processes of biomolecular machines (proteins) because their relevant conformational states are typically separated by large free energy barriers. To overcome this challenge with unbiased, adaptive sampling MD, we have developed AdaptiveMD, a novel, supercomputer platform-independent Python package. AdaptiveMD guides swarms of trajectories by restarting in low-population conformational states identified, e.g., with continually updated Markov State Models (MSMs). To this end, the software runs an asynchronous workflow of uncoupled trajectory tasks and is thus highly scalable. In contrast to regular, long trajectory MD, AdaptiveMD brings a strong scaling property to the exploration of relevant conformational states and their transitions accelerated by adaptive sampling. Together with a supercomputing resource and MD simulation software, AdaptiveMD generates large amounts of simulation data with many conformational states, which are challenging to characterize systematically and in molecular, i.e. experimentally verifiable detail. To overcome this broad analysis challenge, we have developed PySFD (Significant Feature Difference analyzer for Python), a novel multi-processing software toolkit that efficiently selects significantly different user-defined molecular features among potentially many different simulated conformational states. We demonstrate how this software can extract and visualize valuable mechanistic information from big MD simulation data by applying PySFD on MSMs generated from an aggregate of 300 microseconds of MD simulations recently performed on the major histocompatibility complex class II (MHCII) protein.