Artificial Intelligence for Molecular Mechanism Discovery

Hendrik Jung[1,2], Roberto Covino[3], Arjun Wadhawan[4], Christian Leitold[5], Peter Bolhuis[4], Christoph Dellago[5], and Gerhard Hummer[1,2]

[1] *Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt, Germany*
[2] *Department of Physics, Goethe University Frankfurt, Frankfurt, Germany*
[3] *Frankfurt Institute for Advanced Studies, Frankfurt, Germany*
[4] *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands*
[5] *Computational and Soft Matter Physics, Universität Wien, Vienna, Austria*

We develop an AI algorithm to extract the mechanism of collective molecular phenomena from computer simulations. Our algorithm combines transition path sampling (TPS), deep learning, and statistical inference to simulate the dynamics of complex molecular reorganizations while simultaneously learning how to predict their outcome. TPS is a Markov Chain Monte Carlo method to sample the rare transition trajectories connecting metastable states. We iteratively train a deep learning model to predict the outcomes of the shooting moves used in TPS to generate new trial transitions. In this way, we simultaneously increase the efficiency of the rare-event sampling while gradually revealing the underlying mechanism of the transition dynamics. The AI can learn from and steer multiple TPS simulations simultaneously, and, therefore, the algorithm can make full use of highly parallel computing infrastructure and becomes increasingly effective in learning the transition dynamics with an increasing degree of parallelization. In a second step, we then distill the knowledge about the reaction encoded in the deep learning model into a simplified mathematical model. The simplified model describes only the most concise features of the transition ensemble in a human-understandable manner.

With this algorithm, we study the oligomerization of a transmembrane alpha-helix involved in membrane sensing using a coarse-grained MARTINI simulation model. In less than 20 days of walltime with minimal human intervention, the AI distributed approximately 50000 node hours of computation on a highly parallel compute cluster. In this way, the AI sampled 5 ms simulation time distributed over 10000 trajectories, collecting approximately 4000 transition paths with almost optimal efficiency. We estimated the dissociation rate as approximately 1/s, making it unlikely to observe even a single dissociation event in much longer equilibrium simulations. This achievement highlights the great potential of automatic and informed initialization of simulations compared to a manual setup of the simulations. Additionally, the simplified mathematical model helps to understand the mechanism by highlighting the presence of two distinct reaction channels. In conclusion, our algorithm enables researchers to make full use of highly parallel computing resources by autonomously driving many parallel simulations and subsequently aiding the interpretation of the collected data.

We believe that approaches like ours will empower researchers to efficiently use parallel computing infrastructures to reveal the detailed mechanisms of molecular transitions with waiting times on the order of seconds or above. This knowledge will be instrumental in developing quantitative models of complex molecular self-organization phenomena that define many systems of central importance in biophysics and material science.