Automized dataset collection for PES interpolation for pigment-protein complexes

Kwang Hyun Cho

Center for AI and Natural Sciences, KIAS

khcho@kias.re.kr

Nature has evolved various species to capture light energy and convert it into chemical energy during photosynthesis. Light harvesting complexes, composed of spatially arranged pigments and surrounding proteins, play an essential role in this process by delivering light energy through excitation energy transfer (EET) process. The high efficiency of the EET process is a key factor in the overall energy yield of photosynthesis. Understanding the molecular origins of this efficient process requires a detailed investigation of how molecular motions affect pigment quantum states, which can be achieved through molecular dynamics simulations. A key challenge in these simulations is providing an accurate potential energy surface (PES) while maintaining computational efficiency. Previously, we have demonstrated that an interpolation-based approach, with properly sampled datasets, can capture the realistic dynamics and environmental noise of the pigment molecules in light-harvesting complexes, such as the FMO and LH2 complex. However, more systematic approaches are still required, as any other complex requires its own PES due to its unique protein environment. Here, I propose utilizing machine learning techniques to automize dataset collection by reusing pre-obtained data for PES interpolation. This approach aims to systematically construct PESs with minimal effort and to allow for the comparisons of the molecular environments across various pigment-protein complexes. It can provide insights into the fundamental molecular mechanisms underlying efficient energy transfer, which may be ubiquitous in nature.