

# 論文の内容の要旨

Edge expansion parallel cascade selection molecular dynamics

simulation (eePaCS-MD) for investigating protein dynamics

(エッジ拡張型並列カスケード選択分子動力学法 (eePaCS-MD)  
を用いたタンパク質の動的構造探索)

氏名 鷹羽 健一郎

Proteins are inherently dynamical molecules that undergo large-scale conformational changes to exert its functions. To investigate the high anisotropic nature of protein dynamics, Molecular dynamics (MD) simulation is an essential computational tool that can elucidate the conformational transitions of proteins, providing time-dependent information on protein fluctuation at atomic resolution. However, observing conformational changes relevant to biological functions remains a challenge because these events tend to occur stochastically in a time scale longer than feasible MD simulation time. To overcome this difficulty, many enhanced conformational sampling methods have been proposed. However, some of the methods require an external force to enhance the conformational transition, which does not necessarily guarantee that the obtained trajectories follow the lowest energy pathway. Other methods do not need such external forces but may require pre-test of simulations to determine the simulation parameters which can be cumbersome. Therefore, an enhanced sampling method that can simulate protein conformations relevant to biological functions without external forces and does not require cumbersome parameter setting is attractive. In addition, a method that can simulate protein conformations starting from a single structure without the prior knowledge of other conformational states can be valuable, for example situations where a novel protein structure is solved and its conformational transitions are unknown.

This thesis focuses on the development of a new enhanced conformational sampling method,

edge expansion parallel cascade selection molecular dynamics (eePaCS-MD), to investigate the large-amplitude collective motions of proteins with a focus on domain motions. eePaCS-MD is an efficient adaptive sampling method which does not require prior knowledge of the conformational transitions or external forces to enhance the conformational sampling. eePaCS-MD takes advantage of the fact that large-amplitude fluctuations of many proteins can be described in terms of only a few principal components (PCs). In this method, multiple independent MD simulations are iteratively conducted from initial structures randomly selected from the vertices of a multi-dimensional PC subspace with new initial velocities to help the simulated system to overcome the energy barrier. This subspace is defined by an ensemble of protein conformations sampled during previous cycles of eePaCS-MD. The edges and vertices of the conformational subspace are determined by solving the “convex hull problem”.

The conformational sampling efficiency of eePaCS-MD utilizing Cartesian coordinate PCA was assessed for the open-close transitions of glutamine binding protein, maltose/maltodextrin binding protein, and adenylate kinase. The free energy landscape of open-to-closed conformational transitions of glutamine binding protein was obtained by constructing a Markov state model from trajectories generated by eePaCS-MD. The obtained free energy landscape showed an energy barrier separating the open and closed states where the open state was suggested to be energetically more favorable than the closed state. To further enhance the conformational sampling efficiency, eePaCS-MD was combined with accelerated MD, where the total computational cost of observing the open-close transitions can be reduced at most 36% compared to the original eePaCS-MD method. eePaCS-MD is expected to offer 1–3 orders of magnitude shorter simulation time compared to conventional MD simulation.

To address how the conformational sampling efficiency of eePaCS-MD is affected by the choice of the PCA method, eePaCS-MD utilizing Cartesian coordinate PCA (cPCA) and distance-based PCA (dPCA) were compared for the open-close conformational transitions of adenylate kinase. This study suggested that considering all C $\alpha$  distance pairs in dPCA is essential to capture the intrinsic anharmonic nature of proteins during eePaCS-MD simulations to promote the conformational sampling. However, the sampling efficiency of eePaCS-MD utilizing dPCA was comparable to those obtained by cPCA. Considering the computational complexity of dPCA which scales quadratically with the number of atoms, eePaCS-MD utilizing cPCA is suggested as a first choice of the PCA method, although the optimal choice is expected to depend on the target.