Research Outline

Title: *In silico Statistical Mechanics of Protein Conformational Landscape*

Soichiro DEGUCHI

Department of Energy Science and Technology
Graduate School of Energy Science, Kyoto University

Proteins are flexible macromolecules, and their biochemical functions are regulated by a variety of conformations. Therefore, understanding the accessible conformational landscape of a protein help us to identify how the protein behaves *in vivo*. Also, in response to various external stimuli such as allosteric ligand binding, environmental temperature and pH, and post-translational modifications, proteins have specific conformational dynamics that are potentially encoded in their 3D structures. The conformational dynamics of proteins vary from local (residue-wise) to global (domain-wise), however, global conformational changes play a central role in major biochemical reactions, such as enzymatic reactions, signal transduction, and membrane transport, at the micro-, meso-, and macro- scales.

Understanding the complex conformational dynamics of proteins is also important for biomaterial applications. In particular, inorganic nanostructures can strongly affect cell behaviors such as cell viability, proliferation, and differentiation. As a typical example, nanoporous gold (NPG) affects various cellular functions such as adhesion, polarity, and apoptosis of adherent cells. Since the mechanical properties of substrates such as elasticity and nanostructures are transmitted into the cell via transmembrane proteins called integrins, a general method to analyze protein conformational dynamics is essential to manually control material-cell interactions.

In silico computational techniques have made it possible to analyze the behavior of statistical systems with higher resolution than conventional experimental methods. In particular, molecular dynamics (MD) is a useful tool for studying a wide variety of systems, from metals to proteins, at the atomic and molecular level. However, even with the power of modern computers, the typical time scale that MD can handle is limited to ~1 μs, and therefore it cannot capture the slow dynamics of macromolecules. In addition, the electronic interaction between organic and inorganic atoms cannot be considered strictly for organic-inorganic complex systems.

To break the limitations of traditional *in silico* computational techniques, the main policies of this study are set as follows:

1. Understand the electronic correlations between proteins and nanometallic materials from a computational biology perspective.
2. Develop a deep learning-based sampling method to explore wide conformational landscapes of macroscopic proteins which has various metastable basins.

Chapters 2-4 are mainly content that follows policy 1, while Chapters 5 and 6 are studies based on policy 2.

In Chapter 2, the adsorption behavior of collagen molecules on the NPG surface is analyzed by MD simulation with the introduction of virtual electron model. This chapter describes the mechanism by which the surface strain of NPG facilitates the adsorption dynamics of collagen, resulting in lower adsorption energy to NPG substrate.

In Chapter 3, the interaction between integrin and gold substrate is analyzed by combining MD simulation and first-principles calculations. The electronic correlation between NPG and RGD (Arg-Gly-Asp) ligands greatly distorted the structure of RGD,

resulting in binding discordance with integrins. The reduced affinity between RGD and integrin limits conformational activation of integrins and inhibits subsequent outside-in signaling.

Chapter 4 investigates the mechanosensing mechanism of integrins using MD simulations with cyclic tensile forces. A heterogeneous role of integrins is demonstrated, i.e., whether integrins contain an αI domain critically determines their inherent mechanosensitivity.

Chapter 5 proposes a protein structure generator using deep learning instead of the traditional MD simulation. The proposed method pushes the number of atoms that previous methods can handle, and successfully predicts a wide range of conformational landscapes of macroscopic proteins by mapping the configurations to the latent variables that are easy to sample

Chapter 6 describes Metadynamical Q-network (MetaQNet), a novel extended ensemble method that combines deep reinforcement learning and Monte Carlo sampling. MetaQNet provides a powerful framework for efficiently exploring the rare dynamics inherent in many-body systems that would otherwise require prohibitive computational time in conventional Monte Carlo algorithms.

Overall, this study has succeeded in analyzing systems that are unmanageable by conventional in silico techniques with high efficiency, which will open new doors in the statistical mechanics of protein conformational dynamics.