学 位 論 文 内 容 の 要 旨
Abstract of Doctoral Dissertation

博士の専攻分野の名称 Degree requested: 博士（生命科学） Doctor of Life Science
氏 名 Applicant name: Tahmina Sultana

学 位 論 文 題 名
Title of Doctoral Dissertation

A Theoretical Study on Multiscale Reaction Network Extracted from Single Molecule Time Series
（一分子時系列から彫りおこされた階層的反応ネットワークに関する理論的研究）

Ensemble experiments can access only average characteristics of biomolecular system, thus the individual behavior of single molecules cannot be distinguished. On contrast, single molecule (SM) experiments manifest the detailed complexity in kinetics and dynamics of these multiscale biological processes at the single molecule level. The complex dynamics of single molecules, such as the On/Off blinking of nano-particles, the opened/closed gating of ion channels, the bound/unbound kinetics in cell signaling processes, are often probed experimentally in the form of time series with finite discrete levels. The main goal of SM experiments is to extract kinetics and dynamical information from the observable time series. However, how one can extract such information from an observed one-dimensional time series is still an unresolved question.

In the analyses of SM time series, hidden Markov models (HMM) are often used to provide insights on the complex mechanism of molecular machinery. To derive HMM from SM time series that incorporates non-exponential kinetics and molecular memory is one of the most contemporary and intriguing subjects in the analysis of single molecule biology. In this thesis, I reviewed a recently developed mathematical approach which provides not only interpretation of the complex kinetics but also new insights into biological function buried in ensemble measurements. This method, based on information theory, is free from any *a priori* assumption such as local equilibrium, detailed balance, or number of states. Mathematically, it is proved that its scheme is the simplest representation that can predict the future outcome maximally. This approach is known as State Space Network (SSN). The SSN is a type of HMM that takes into account both multiple states buried in the measurement and memory effects in the process of the observable whenever they exist. The states of SSN depend not only on the present value of the observable but also on the past values along the course of time evolution so that the state-to-state transitions are Markovian even though dynamical correlation may exist in the time series.

The scheme of SSN was based on the assumption that a given time series is stationary and long enough. Time series obtained from SM experiments always suffer from as an insufficient number of data points and sometimes may be non-stationary. To overcome these difficulties, a generalization of SSN was developed by using a multiscale decomposition scheme based on discrete wavelet transforms in our laboratory (Li *et al. PNAS*, 105, 536–541, 2008). The main drawback to apply this decomposition scheme to time series is that it depends on the choice of wavelet basis function. Instead of wavelet decomposition, I introduced a simple but efficient skipping step method (SSM) to decompose the original time series into a set of time series at different timescales. The SSM can deal with the nature of multiscale nature and nonstationarity for a discrete time series in the framework of SSN.

Previously, in an *in vitro* reconstituted system of the recognition process between epidermal growth factor receptor (EGFR) on the plasma membrane and its adaptor protein Ash/Grb2 (Morimatsu *et al.*, *PNAS*, 2007, 104, 18013–1801 non-exponential kinetics in the association and dissociation processes were found. In addition, the association kinetics abnormally depends on the concentration of Grb2 owing to reaction memory concealed in the conformation of EGFR. Two extreme reaction schemes for the association process were proposed (Takagi *et al. Adv. Chem. Phys*, 2012, 146,195–215); one of them is called defect diffusion model in which, the system visits many dissociated states with different rate constants to reach a specific target state for

association. The other is called multiple-reaction channel model where all (dissociated) states with different association rate constants directly connected to the associated state independently. However, they could not identify the actual, most plausible kinetic scheme that reflects the observed kinetics objectively.

To demonstrate the versatility of my scheme, I investigated the time series of the EGFR-Grb2 system. It was found that a newly-derived analytical expression of autocorrelation function from SSN combined with SSM successfully reproduce the autocorrelation for a wide range of timescale up to the timescale that loses the memory, 3 s. It was found that the underlying SSNs are in between the defect diffusion model and multiple-reaction channel model and change their topographical structure as a function of the timescale: while the corresponding SSN is simple at the short timescale (0.033 s to 0.1 s), the SSN at the longer timescales (0.1 s to 3 s) becomes rather complex in order to capture multiscale kinetics emerging at longer timescales.

The Y1068F mutant of EGFR replaced tyrosine (Y) 1068 in EGFR (whose phosphorylation has been reported to construct the primary strong Grb2 binding site) for analyzing the exponential properties and memory effects in association and dissociation kinetics. The SSNs for the mutant (Y1068F) EGFR are simpler than the wild type EGFR, indicating the existence of non-Markovian nature in the wild type more than the mutant (Y1068F) EGFR. In addition, the SSN can capture the heterogeneity of the memory in the process depending on each state. By looking into splitting patterns as an increase of the history length for the shortest time scale SSN of the wild type EGFR at low concentration of Grb2, It was found that visiting the unbound form of the wild EGFR-Grb2 system approximately resets all information of history or memory of the process.

To implement the direct relationship between the states in the multiscale SSNs and the underlying high-dimensional conformation of the system, it is necessary to construct a series of SSNs in terms of a set of different single-molecule time series data from systematic mutations of amino-residues. I expect that one may identify which amino residues perform to yield non-Markovian kinetics by monitoring the morphological changes of the network. This may be regarded as Φ value analysis of protein folding kinetics to infer the energy landscape in the framework of SM biology.