Impact on Nucleosome Dynamics via Histone Variants and Post-Translational Modifications

Hidetoshi Kono

Group Leader, Molecular Modeling and Simulation Group
Quantum Beam Science Center, Japan Atomic Energy Agency
JAPAN

Dynamics of the nucleosome and accessibility of nucleosomal DNA play critical roles in many nuclear processes. The nucleosome is the fundamental structural unit of chromatin, comprising a histone octamer formed by two copies each of the four core histones H3, H4, H2A, and H2B and approximately 150bp of DNA that is wrapped around the octamer. There are variants of these histone proteins. In humans, the variants include H3.1 (canonical), H3.2, H3.3, H3T, H3.5, H3.X, H3.Y, and CENP-A, each of which is likely to have a specific role. However, their crystal structures look similar to each other and the dynamics of them has not been fully understood. In addition, post-translational modifications of histones have been known to affect gene regulation, but it is still unknown how they affect changes of the structure and dynamics of chromatin. Using supercomputers including K computer, we are currently investigating the following topics to understand the molecular mechanism: 1) how nucleosomal DNA is unwrapped from the histone core, 2) if there are differences in DNA unwrapping process among nucleosome types composed of histone variants and 3) how acetylation on histone tails impacts nucleosome structure.

We have found that different nucleosomes showed different free energy profiles for DNA unwrapping. The canonical nucleosome which has H3 histones showed that a state where nucleosomal DNA was fully wrapped around the histone core was the most stable while in the nucleosome having a H3 variant, CENP-A, was in a state where DNA at entry and exit regions were partially unwrapped. We also observed the asymmetric DNA unwrapping: only either of ends of DNA was unwrapped from the histone core after about 35bp of DNA on both sides were dissociated from the core. In addition, our simulation indicated that an acetylation of H3 tail induced α-helix formation and enhanced DNA flexibility, which may be the first step of the disruption of nucleosome.

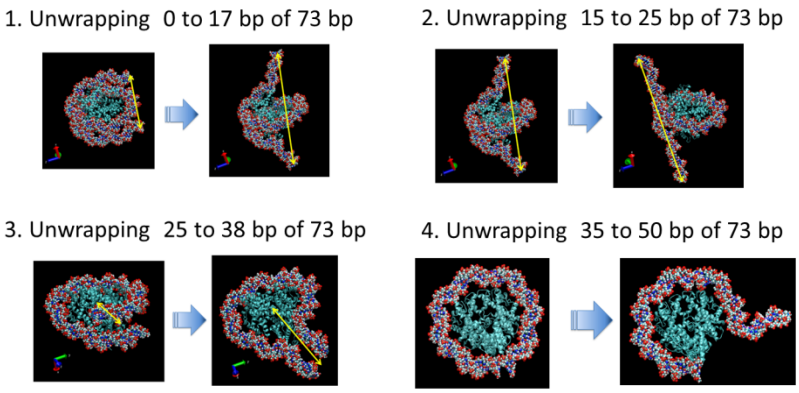

Figure 1. Nucleosomal DNA unwrapping process simulated with K computer. For clarity, only DNA and the histone core (blue) are shown.

Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution

Xuhui Huang
Associate Professor, Department of Chemistry
Division of Biomedical Engineering
The Hong Kong University of Science and Technology
HONG KONG

Transcription, the synthesis of RNA from a complementary DNA template, plays a crucial role in cellular regulation, including differentiation, development, and other fundamental processes. We performed molecular dynamics (MD) simulations to understand the thermodynamics of the major states of transcription elongation and the kinetics of transitioning between them for RNA polymerase II (pol II). In this talk, I will discuss our recent results on simulating pol II Translocation and other functional conformational changes of this enzyme. The main challenge for MD simulations is to reach biologically relevant timescales, which are orders of magnitude longer than most atomistic simulations. In order to overcome this timescale gap, we have developed a novel algorithm, Hierarchical Nystrom Extension Graph method, to construct Markov State Models to extract long timescale dynamics from short simulations. We reveal that RNA polymerase II translocation is driven purely by thermal energy and does not require the input of any additional chemical energy. Our simulations show an important role for the bridge helix: Large thermal oscillations of this structural element facilitate the translocation by specific interactions that lower the free-energy barriers between four metastable states. Among these states, we identify two previously unidentified intermediates that have not been previously captured by crystallography. The dynamic view of translocation presented in our study represents a substantial advance over the current understanding based on the static snapshots provided by X-ray structures of transcribing complexes.

Molecular Simulation for Antibody Drug Development

Keiko Shinoda

Project Assistant Professor, LSBM, RCAST
The Univ. of Tokyo
JAPAN

The use of antibodies that bind specifically to antigens has achieved considerable success in cancer therapy in recent years. Epiregulin (EPR) is one such promising therapeutic target protein for advanced cancer. EPR is a member of the epidermal growth factor family, and is overexpressed on various cancer cells such as in colon, ovarian, breast and prostate cancer as well as in pancreatic cancer, which is difficult to detect in its early stages. Here we developed an anti-EPR murine antibody, 9E5. In advanced cancer, since the cancer spreads inside the cancer tissue, it is important that the molecular weight of antibody, while including the recognition site, is small to penetrate the cancer tissue. However, a small size will result in low affinity between the antibody and antigen. Humanization is also important to reduce murine antibody immunogenicity. In order to humanize an antibody and to increase affinity, it is necessary to modify the antibody. To this end, we first analyzed the structures of apo 9E5 (EPR-free 9E5) and its complex with EPR (this complex is hereafter referred to as the “trans-Complex”) by X-ray crystallography. It was found that a proline at residue 103 in the third complementarity-determining region (CDR) of the heavy chain changed its conformation from cis to trans by binding to EPR. The cis-trans isomerization is known to occur slowly, with the duration ranging from several seconds to hours. We then aimed to determine whether EPR forms a stable complex with apo 9E5 in which Pro103 is in the cis conformation before the isomerization occurs. In order to identify an EPR-binding structure with apo 9E5, we conducted extensive binding MD simulations using K computer. From the simulations, an EPR-binding structure similar to the “trans-Complex” was found. We then conducted additional long MD simulations to determine the stability of this structure. This structure changed to a more similar structure to the “trans-Complex” (this complex is referred to as the “cis-Complex”) and remained stable during total 16 μs simulation. However, there were several major differences between the “cis-Complex” and the “trans-Complex” in the short-interaction energy, fluctuation of the CDR H3 loop, and dynamics of water at the EPR-9E5 interface.

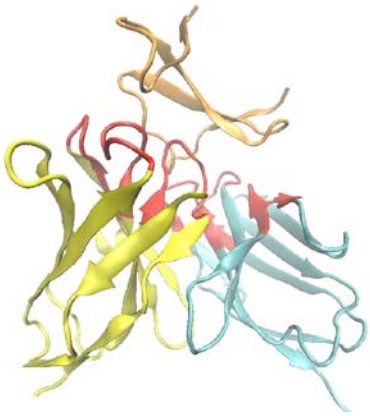

Figure 1. Epiregulin (orange) and the recognition site of antibody 9E5 consisting of heavy chains (yellow) and light chains (light blue). CDRs are indicated in red.

The Path to Exascale Molecular Simulation: Parallelization, Acceleration, and Ensemble Simulation Techniques to Understand Biological Macromolecules

Erik R. Lindahl
Professor of Theoretical Chemistry, Department of Biochemistry & Biophysics, Stockholm University, SWEDEN
Professor of Theoretical Biophysics, Science for Life Laboratory, KTH Royal Institute of Technology, SWEDEN

Supercomputers have been a revolution to modern simulation research, and yet they also lead to gigantic challenges. As computers have grown in size, power, and parallelism, researchers have resorted to simulating larger and larger systems simply to achieve good scaling to justify access to the resources. This works well in many engineering disciplines, but in biomolecular life sciences we are typically constrainted by the size of the target molecule: A membrane protein system might need 200,000 atoms, but very few (if any) systems need atomic detail with millions or billions of particles. This makes classical molecular simulation an extremely difficult problem for supercomputers, since improved performance depends on decreasing the iteration time for an entire step to hundreds of microseconds. Here, I will describe some of the efforts we have done to achieve this in GROMACS over the last decade, and how we are increasingly relying on large-scale fully automated ensembles of several simulations, even for complex systems such as membrane proteins. I will present our new COPERNICUS framework that interacts with GROMACS to enable usage of a number of important adaptive sampling algorithms such as free energy calculations, markov state models, and the string method using swarms. This combines the strongest points of traditional supercomputers with throughput-focused techniques such as distributed computing, and will make it possible to use tens to hundreds of thousands processors efficiently for biomolecular simulation. Finally, I will discuss how we apply these methods to studies of ion channels, and how the simulations make it possible to probe both structure and dynamics that are simply out of reach to current experimental methods.

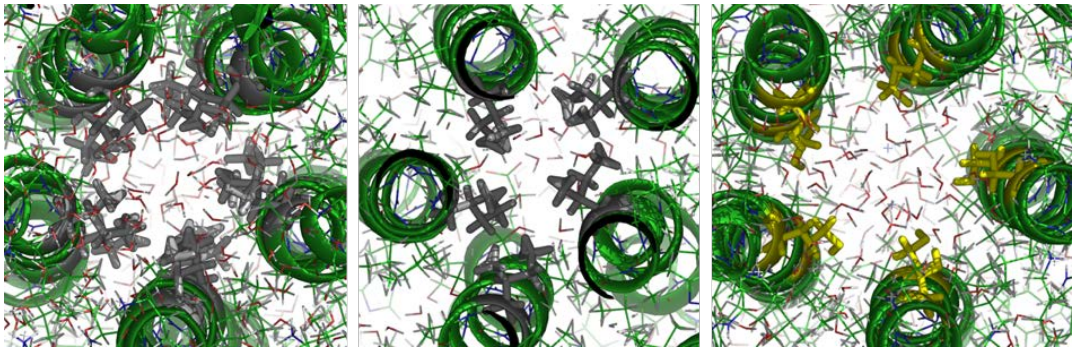

Figure 1. The ion-conducting pore in a ligand-gated ion channel. Ensemble molecular simulations make it possible to predict how the response will be left- or right-shifted (as a function of the voltage), depending on the pH, mutations, and binding of allosteric modulators.

Can an Integrated Model of the Basal Ganglia-Thalamo-Cortico-Spinal Network and the Musculoskeletal System Reproduce Healthy and Pathological Motor Behaviors?

Kenji Doya

Professor, Neural Computation Unit
Okinawa Institute of Science and Technology Graduate University
JAPAN

Under the Strategic Program in Supercomputational Life Science, we perform hierarchical integrated simulation of the brain, the spinal cord, muscles and the skeletal system for understanding the mechanisms of movement disorders and for identifying therapeutic targets. Our team at OIST focuses on constructing realistic spiking neural network models of the basal ganglia, the thalamus, and the motor cortex for elucidating the mechanism of Parkinsonian tremors. The specific aim is to clarify the dynamic mechanism for oscillations in different frequencies observed in the brain and the musculoskeletal system. The present network model includes five types of neurons in the basal ganglia, four types in the thalamus, and eight types in five layers of the motor cortex. They form a topographic basal ganglia-thalamo-cortical loop circuit and the size of the network is matched to the entire basal ganglia and 1.6x1.6 mm^2 cortical surface of the rat brain. Each sub-network is implemented by NEST (www.nest-initiative.org) and the entire network, consisting of more than 3 million neurons, is connected by MUSIC (software.incf.org/software/music) to run in parallel on K Supercomputer. The model reproduces 8 to 15 Hz oscillation in the basal ganglia and sub-harmonic responses in the thalamus and the motor cortex.

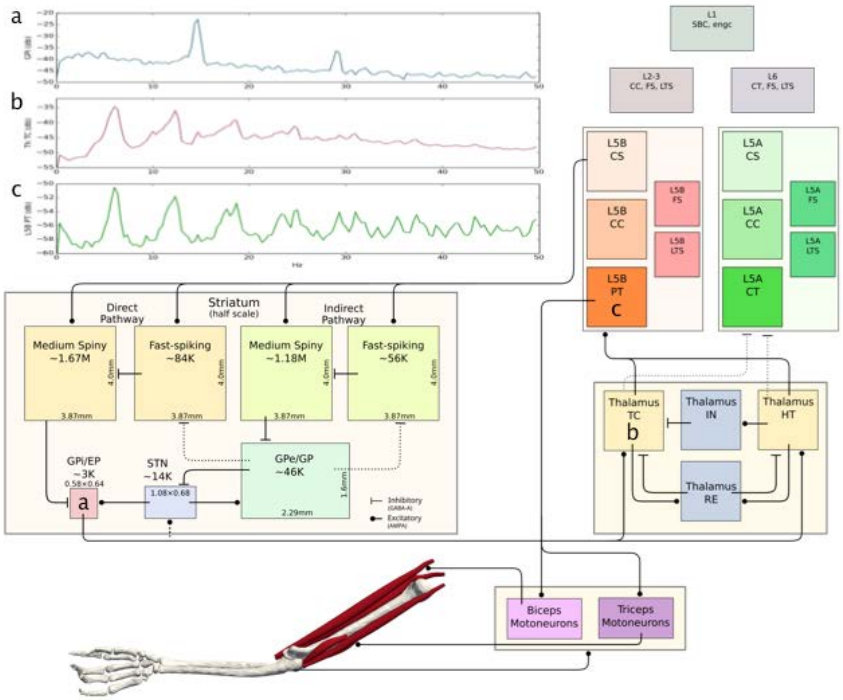

Figure 1. The basal ganglia-thalamo-cortical network model combined with the spinal and musculoskeletal models. Insets: power spectra of activities in **a:** globus pallidus internal segment (GPi), **b:** thalamo-cortical (TC) neurons, and **c:** cortical pyramidal tract (PT) neurons.

Sensing Positive Versus Negative Reward Related Signals in the Direct and Indirect Basal Ganglia Pathways - A Computational Investigation of Read-Out Mechanisms

Jeanette Hellgren Kotaleski
Professor, Science for Life Laboratory
School of Computer Science and Communication
KTH Royal Institute of Technology
SWEDEN

To understand the inner workings of the brain, it is necessary to integrate data from the molecular level up to the systems level. Here computational modeling and simulations provide important tools. Some recent insights resulting from kinetic modeling on the subcellular level are reviewed, and it will be demonstrated how modeling may help when extrapolating to *in vivo* like conditions in the search for the cellular mechanisms involved in basal ganglia dependent learning and decision making. Transient positive and negative changes in the striatal dopamine (DA) concentration are considered to encode a reward prediction error (RPE) in reinforcement learning tasks. This phasic DA change occurs concomitantly with a transient dip in striatal acetylcholine (ACh) while other neuromodulators like adenosine (Adn) changes more slowly. There are abundant G protein coupled receptors (GPCR) linked to adenylyl cyclase for these neuromodulators in striatal medium spiny neurons (MSNs), but little is known about the way in which these fast transients might affect the cAMP/PKA signaling axis which is activated mainly by G_{olf} and inhibited by $G_{i/o}$ in MSNs. The goal of our computational study was to investigate how transient changes in these neuromodulators are detected and read out by MSNs. We have found that in direct pathway MSNs expressing D1 receptors the ACh dip, operating via M4R/$G_{i/o}$, could function as a gate for the activation of PKA triggered by the D1R/G_{olf}. On the other hand, in indirect pathway D2R expressing MSNs the DA dip, associated with reward omission and operating via D2R/$G_{i/o}$, produced significant PKA activation due to the presence of a high Adn tone acting through A2aR/G_{olf}. Our study highlights the common pattern of $G_{i/o}$ and G_{olf} interactions in both the MSN types, where a dip in $G_{i/o}$ acts as a gate for the G_{olf} activation. These results together with earlier experimental and computational studies also corroborate the functional dichotomy between D1R- and D2R expressing MSNs in terms of the RPE sign (positive or negative) which could elicit significant PKA activation, and which thus would facilitate long-term potentiation in the cortico-striatal synapses. Our simulation experiments furthermore help to identify during what circumstances the DA transients are expected to promote synaptic plasticity during reward dependent learning paradigms.

Breast Cancer Phenotype Characterization by Network-Based Omics Data Analysis

Sun Kim
Professor, Computer Science and Engineering;
Director, Bioinformatics Institute
Seoul National University
KOREA

Sequencing technology is routinely used to produce a genome-wide measurement of biological events, known as omics data. Omics data provides a complete picture of genetic and epigenetic events in the whole cell, so it is invaluable for research in medical and biological sciences to characterize biological mechanisms underlying phenotypes, e.g., cancer subtypes. Unfortunately, integrating omics data is a challenging data mining task and we need to employ novel approaches. In particular, association of different genetic and epigenetic events increases the complexity of the computational problem dramatically and we need powerful computational techniques. Network based approaches are most promising since associations are naturally modeled as edges in the network. In this talk, we show how network mining can be useful to characterize phenotypes and biological mechanisms underlying breast cancer phenotypes. Three phenotypes in breast cancer are characterized by using protein-protein interaction (PPI) networks, TF-target gene networks, and miRNA-target gene networks, respectively. The first study is to show that decomposition of biological pathways using PPI networks can effectively enhance patient survival prediction in breast cancer. Recent studies show that pathway based cancer analysis is effective. However, some pathways consist of several hundred genes, thus these pathways involve multiple biological functions. We used PPI information to decompose pathways into functionally coherent components and used them for breast cancer patient survival prediction. The second study is to show that sub-network mining of TF-target gene networks can effectively classify breast cancer grades. This is to investigate on biological functions underlying breast cancer tumor grades in terms of transcriptional mechanisms and distinct biological functions of TF target genes. The third study is to show that use of miRNA-target gene networks can be useful for characterizing breast cancer subtypes and roles of miRNAs in breast cancer subtypes. In this study, to leverage characteristics of sequencing data that can measure transcription quantities, we used a novel algorithm that considers miRNA-target gene split ratios to distribute transcription quantities of miRNAs to their target genes in an optimal way. The analysis results not only suggest miRNAs that are related to tumor subtypes but also regulatory mechanisms by miRNAs in different breast cancer subtypes.

Supercomputers for Cancer Research and Cancer Big Data

Satoru Miyano

Professor, Human Genome Center, The Institute of Medical Science, The University of Tokyo
JAPAN

In our cancer research, we have been using K computer of RIKEN and the supercomputer system Shirokane 1&2 (225 TFLOPS, 3PB Lustre file system, 2009-2015) and Shirokane 3 (440 TFLOPS from April 2015) with 12PB Lustre file system together with a nearline storage (1PB) connected to the IBM tape archive system extensible to 100PB of our Human Genome Center. In this talk, we will first present highly parallel software applications developed in the project “HPCI Strategic Programs for Innovative Research Field 1 “Supercomputational Life Science” (2011- 2015) and the Grand Challenge Project for Life Science “Next-Generation Integrated Simulation of Living Matters (2006-2012). The first series of applications include various gene network estimation software applications on both supercomputers. The second is a series of software applications and pipelines for cancer genome analysis on these supercomputers. The power of K computer led us to a challenge and the Shirokane seires efficiently accelerated our cancer genome sequence data analysis. We present some innovative contributions in cancer systems biology which could not be achieved without K computer and important discoveries on cancer genomes which were made by Professor Seishi Ogawa with the Shirokane series. The talk also includes our story of struggles behind.