Chapter 10

Computational Biophysics Research Team

10.1 Members

Yuji Sugita (Team Leader (concurrent))*

Jaewoon Jung (Research Scientist (concurrent))**

Chigusa Kobayashi (Research Scientist)

Koichi Tamura (Special Postdoctoral Researcher)***

Ai Shinobu (Postdosctoral Researcher)

Cheng Tan (Postdosctoral Researcher)

Hiromi Kano (Assistant)

Mitsunori Ikeguchi (Guest Scientist)****

Michael Feig (Guest Scientist)*****

Takao Yoda (Guest Scientist)******

Naoyuki Miyashita (Guest Scientist)*******

* Chief Scientist in RIKEN CPR; Team leader in RIKEN BDR
** Senior Technical Scientist in RIKEN CPR
*** Special Postdoctoral Fellow in RIKEN CPR
**** The main affiliation is Yokohama City University
***** The main affiliation is Michigan State University
****** The main affiliation is Nagahama Bio Institute.
******* The main affiliation is Kindai University.

10.2 Overview of Research Activities

We have developed GENESIS (Generalized-ensemble simulation system) software for high-performance molecular dynamics (MD) simulations of various chemical and biological systems. Using GENESIS, we can carry out MD simulations based on atomistic and coarse-grained molecular models. The atomistic model allows us to include accurate molecular interactions between biomacromolecules, while the available system size and simulation time are limited due to the required computational resources. We therefore consider combining simulations based on the atomistic model with those with coarse-grained models for simulating cellular-scale phenomena on Fugaku supercomputer. Now, in GENESIS, various coarse-grained models, such as AICG2+ for protein, the 3-site-per-nucleotide (3SPN) model for DNA, which were previously developed in CafeMol, a coarse-grained MD simulator.

GENESIS also includes various enhanced conformational sampling methods for explore wider conformational spaces of biomacromolecules. In collaboration with Laboratory for biomolecular function simulation, RIKEN BDR, we added Gaussian accelerated replica-exchange umbrella sampling (GaREUS) method and free-energy perturbation (FEP) calculations in the latest version of GENESIS. These methods will be applicable to in-silico drug discovery.

The performance of GENESIS has been greatly improved in the new non-bonded interaction kernels of the next major release (GENESIS 2.0 or later). This kernel was optimized to the ARM CPU architecture in Fugaku, allowing us to simulate slow conformational dynamics of biomolecules in cellular environments. We also used GENESIS for biological applications, such as Calcium ATPase and Heme transporters in biological membranes.

10.3 Research Results and Achievements

10.3.1 Development of GENESIS on Fugaku supercomputer

In the last fiscal year, we developed GENESIS MD software to carry out cellular scale molecular dynamics (MD) on Fugaku supercomputer. It includes the optimization of program algorithm and modeling of the system.

10.3.1.1 Optimization of real-space non-bonded interaction

In MD with all atom model, there are bonded and non-boned interactions. Bonded interactions includes those of bond, angle, and dihedral angles. Non-bonded interactions consist of van der Waals and electrostatic interactions. Electrostatic interaction is separated into real- and reciprocal-space interactions. At moderate number of processors, the main bottleneck is the van der Waals and real-space electrostatic interactions. To accelerate the speed on Fugaku, we designed a new algorithm to minimize the operand waiting time. Array of Structure (AoS) and Structure of Array (SoA) types are appropriately assigned to maximize the overall performance. L1 cache prefetch is applied to minimize the memory access time. The new algorithm accelerates the speed on Fugaku more than twice compared to the algorithm used on K computer.

10.3.1.2 Optimization of reciprocal-space interaction

The reciprocal-space electrostatic interaction can be understood by applying fast Fourier transform (FFT). Because FFT requires global communications, it becomes the main bottleneck for very large number of processors. To accelerate the reciprocal-space interaction, we developed new algorithm of charge grid data generation. With the new algorithm, we can reduce the number of operations before/after forward/backward FFTs. It also allows us to reduce the FFT grid points. In addition, we made a tool to decide the best FFT scheme automatically from given number of processors. These developments accelerates the speed of the reciprocal-space interaction more than twice.

10.3.1.3 Optimization of multicopy enhanced sampling scheme on Fugaku

With multicopy enhanced sampling schemes, we prepare many replicas with different working conditions. To optimize the speed with these schemes, the communication among processors in the same replica should be minimized. For this, we defined three-dimensional indicies of MPI ranks and replicas, and mapped these indicies to multi-dimensional network topology of Fugaku.

10.3.1.4 parallel file input/output for large scale MD simulations

Parallel file input/output (I/O) are used in GENESIS for large scale MD to avoid memory problem in reading and writing files. The original parallel file I/O in GENESIS has the problem that we need to regenerate parallel files whenever working conditions are changed. For example, when MPI processor number is changed, we should rewrite the parallel file by spending at least a few hours. In the newly developed parallel file I/O, we do not need to regenerate parallel files in spite of changed working conditions.

10.3.1.5 modeling of cellular scale system

In living cells, a number of macromolecules such as proteins, nucleic acids, and lipids interact with each other. It is thought that their reactivity and stability are changed by recognizing each other. It is a fundamental issue to understand the mechanism of such processes. However, studying environments which include numerous

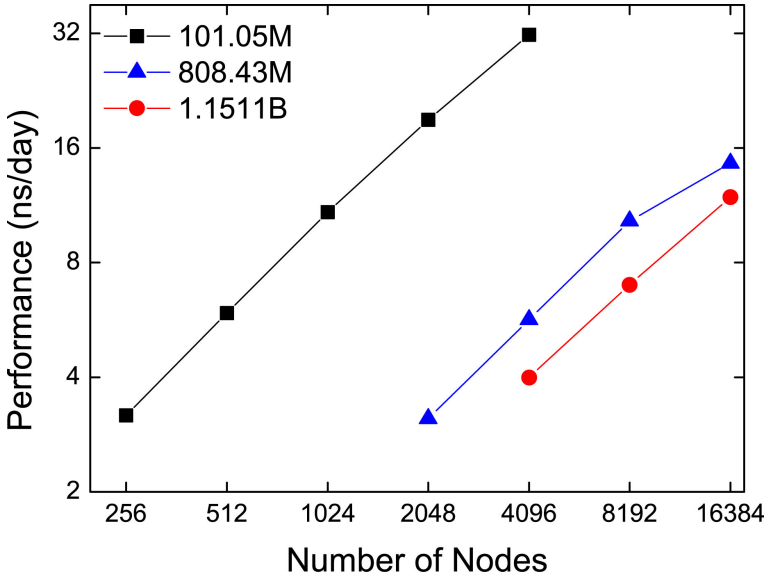

Figure 10.1: Strong scaling result of GENESIS on Fugaku.

macromolecules of different types is very difficult due to the enormous amount of calculation required. We previously developed a multi-scale modeling protocol that starts from coarse-grained (CG) model and increases the resolution step-by-step. This was applied successfully for assembling the first atomistic model of the cytoplasm of a small bacterium, Mycoplasma genitalium. We applied a very similar protocol as described previously for assembling proteins, nucleic acids, metabolites, ions, and water in a highly packed cytoplasmic model with much faster speed of equilibrium process. Finally, we could prepare the cytoplasmic systems with ∼100 million and ∼1.5 million atoms. This is used for a multi-copy enhanced sampling method on Fugaku.

10.3.1.6 benchmark result on Fugaku

Due to the developments described above, we obtained 31.77 ns/day, 14.59 ns/day, and 11.88 ns/day for 101.05 million atoms (101.05M), 808.43 million atoms (808.43M) and 1.1511 billion atoms (1.1511B) systems (Figure 10.1). When we apply a multi-copy enhanced sampling scheme named the generalized REST (gREST) with 54 replicas,we could obtaine 5.37 ns/day and 2.75 PFLOPS.

10.3.2 Investigatio of the Reaction Pathway on the E1/E2 transition of Ca^{2+}–ATPase

Sarco(endo)plasmic reticulum Ca^{2+}–ATPase is a representative protein of P-type ATPases, which transports Ca^{2+} across membrane against a large concentration gradient (104 times) by utilizing ATP hydrolysis. The protein consists of three cytoplasmic domains and ten transmembrane (TM) helices. According to the E1/E2 theory, it exhibits at least two different states, E1 and E2. Structural and biochemical studies have suggested functional roles of coupled motions between the cytoplasmic domains and the transmembrane helices in the reaction step between the two states. It is, however, difficult to observe the atomistic information of reaction pathway for experimental studies. Therefore, we calculated the reaction pathway on the step and the free-energy profile by using GENESIS (Figure 10.2) We analyzed the conformational changes on the reaction pathway. As a result, we suggest a series of structural changes, domain motions of cytoplasmic domains, and rearrangement of transmembrane helices, and then gating motion.

10.3.3 Simulating large-amplitude transitions in proteins with a coarse-grained Gō-model

Molecular dynamics (MD) simulations of biomolecules are widely used to investigate conformational dynamics and structure-function relationship. All-atom (AA) models provide the most accurate description of the underlying dynamics. However at present, even with the fastest computers, the time-scale obtainable with

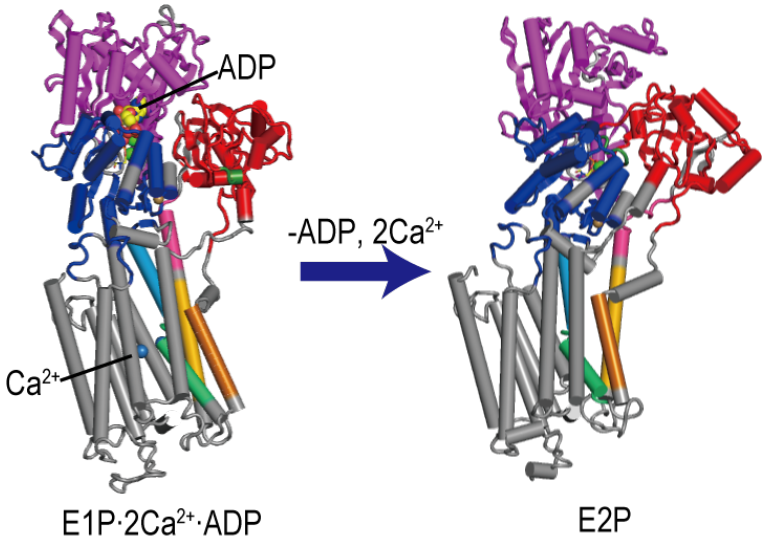

Figure 10.2: Reaction step between E1 $\sim$ P•2Ca^{2+}·ADP (left) and E2P (right) states of Ca^{2+}-ATPase.

an atomistic simulation is up to a millisecond, whereas biologically relevant motions occur at the time scale of milliseconds to seconds. To overcome this, coarse-grained (CG) modeling can be utilized. The use of CG models reduces the computational time by several orders of magnitude, allowing access to biologically relevant time-scales.

In the first work (Figure 10.3), we used the dual-basin (DB) Gō-model, which is a structural-based CG model, for simulating conformational transitions between two known structures of a protein. The DB potential is formed by mixing two single-basin potentials and includes system-dependent parameters. The determination of parameters however, is usually not straightforward and can be time consuming.

We developed an efficient scheme to determine the mixing parameters using the Multistate Bennett Acceptance Ratio (MBAR) analysis method after short simulations with a set of parameters. In the scheme, MBAR allows us to predict observables at various unsimulated conditions, which are useful to improve the mixing parameters in the next round of iterative simulations. The number of iterations that are necessary for obtaining the converged mixing parameters is significantly reduced in the scheme.

We applied the scheme to several proteins, for showing the effectiveness in parameter determination. After obtaining the converged parameters, the proteins show frequent conformational transitions between open and closed states, providing the theoretical basis to investigate structure dynamics-function relationships of the proteins.

Next, we extended the DB model to a general multi-basin (MB) model for realizing conformational transitions in multi-domain protein with more than two labile domains. For such proteins, the previously developed DB model could not describe the structure of intermediate states because experimental structures are usually not available for them.

In our second work (Figure 10.4), we described intermediate structures by assuming that they have partial inter-domain contacts with respect to both reference structures, i.e. their potential can be described as some combination of the end-states potentials, thus obviating the need for obtaining their experimental structures. In addition, we use the MBAR analysis method for an efficient determination of mixing parameters.

We demonstrated the use of the MB Gō (MBGo) potential on the enzyme Adenylate Kinase. We sampled multiple transitions between its open, closed and the intermediate states, and characterized transition pathways and structural ensembles at each basin. We demonstrated that with the MBGo potential, unconstrained time-dependent dynamics can be obtained in a reasonable simulation time. The obtained trajectories can be analyzed in any desired reaction coordinate, serving as a ground for endless possibilities to study structure function relationship in proteins.

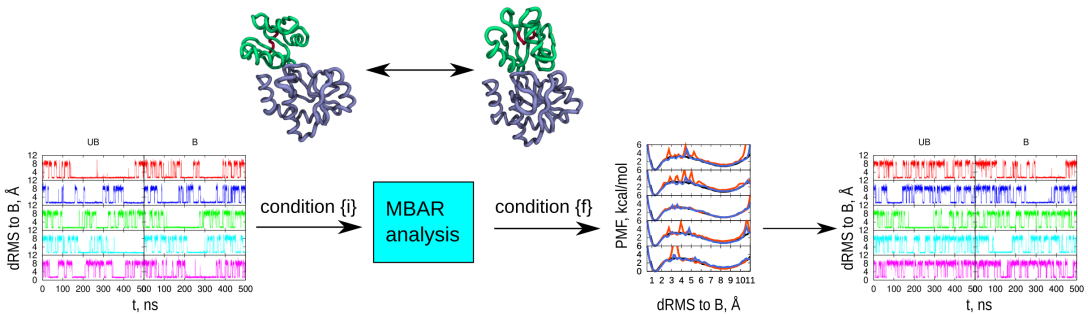

Figure 10.3: Top: Open and Closed conformations of the Glutamine-binding protein. Bottom: overall scheme describing mixing parameter optimization using MBAR analysis.

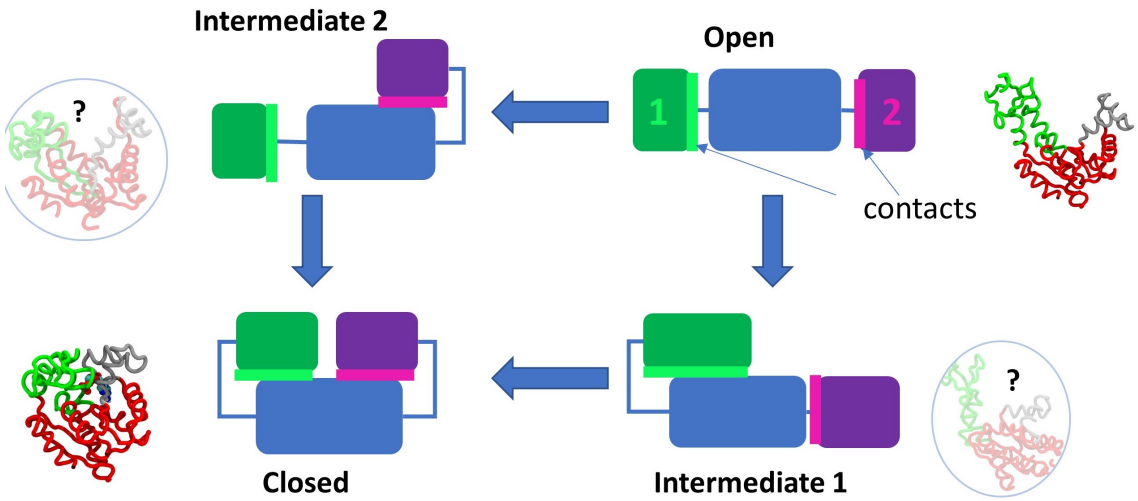

Figure 10.4: Schematic description of the conformational transition pathways of a protein with two labile domains.

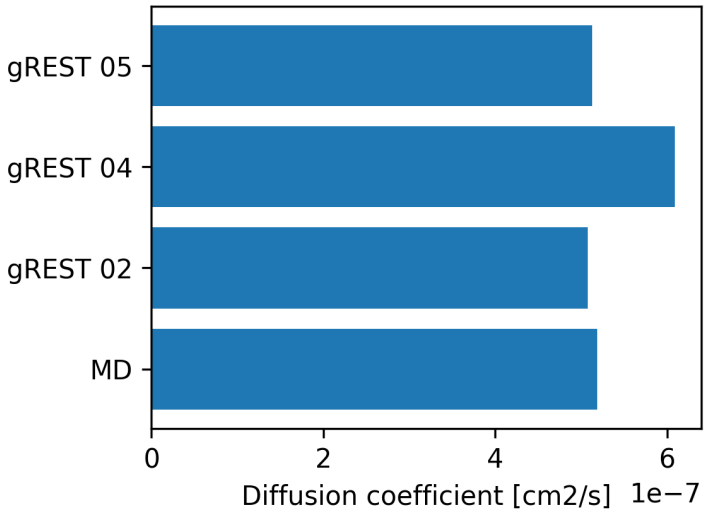

Figure 10.5: Comparison of diffusion coefficient calculated from the gREST and the conventional MD simulations.

10.3.4 Theoretical Study of Lateral Diffusion of Heme on Lipid Bilayer

In vertebrate, heme is abundant source of iron for infected bacterial pathogen. Bacteria have elaborated elegant system to acquire free heme from environment. Free heme that escape from the system tend to partition into the cell membrane and disrupt the structure of lipid bilayer. In this study, we aim to analyze the behavior of heme on the membrane. To this end, we adopted the gREST method to enhance lateral diffusion of heme. In the gREST simulation, the effective temperature of the specific part of the system ("solute region") is increased and therefore can effectively simulate the diffusion process of heme. We considered several conditions for gREST simulation and selected three efficient ones for further sampling ($\sim$ 50 ns/replica). For comparison, we performed five independent conventional MD simulations (300 ns each). Figure 10.5 summarize diffusion coefficients calculated from the trajectories. Although gREST slightly enhanced diffusion, further consideration is obviously necessary to drastically increase its efficiency.

10.3.5 Implementation of CG models in GENESIS

One of the long-standing challenges in computational biophysics is to achieve the balance between force-field accuracy and sampling adequacy. Among the multiscale approaches, coarse-grained (CG) models have achieved great success in studying the long time-scale dynamics of large biomolecules.During the last year, we implemented the most popular CG models into GENESIS, to provide the biophysical society a versatile platform for running general purpose MD simulations. Specifically, we focused on the recently developed CG models at the resolution of roughly 10 atoms per bead, which have been shown to work consistently and compatibly with each other. These models include the atomic interaction-based CG model version 2+ (AICG2+) for protein,the 3-site-per-nucleotide (3SPN) model for DNA, and a structure-based model for RNA. In addition, we employed the hydrophobicity scale (HPS) model and the Kim-Hummer (KH) model for the intrinsically disordered regions in protein.5 As for the inter-molecular interactions between protein and DNA, we considered the Debye-Huckel type electrostatics for the sequence-non-specific binding and the position-weight-matrix-complex-structure (PWMcos) integrated model for the sequence-specific recognition.6 We have implemented and tested all the above-mentioned models in GENESIS.

After the implementation of the interactions, we determined a unified set of parameters such as cut-off values for short-ranged and long-ranged interactions, respectively. We also developed cell-linked list method for utilizing the pair-list method for each interaction term. Via these developments, the CG models in GENESIS are now ready for use.

10.3.6 Development of a singularity-free dihedral angle potential in CG

The classical dihedral angle potential has a singularity problem when three adjoining particles are colinearly aligned. The CG MD simulations are more susceptible to this problem. By introducing a modulating function as a multiplier and found suitable parameters based on statistics, we solved the singularity problem and combined the new potential with existing CG models seamlessly. By apply our new potential in the simulations of designed and real biomolecules, we showed that our method was not only computationally efficient, but also robust in practical numerical calculations.

10.4 Schedule and Future Plan

In the next fiscal year, we plan to design a better parallelization scheme of coarse-grained MD for cell-scale systems with large number of particles. We also consider further optimization to gain better performance on the Fugaku. As applications, we will use the current models to study the phase behaviors of biomolecules in physiological processes such as genome organization and gene expression.

10.5 Publications

10.5.1 Articles/Journal

[1] Jung, J., Nishima, W., Daniels, M., Bascom, G., Kobayashi, C., Adedoyin, A., Wall, M., Lappala, A., Phillips, D., Fischer, W., Tung, C., Schlick, T., Sugita, Y. and Sanbonmatsu. K.Y,. Scaling Molecular Dynamics beyond 100,000 Processor Cores for Large–Scale Biophysical Simulations, Journal of Computational Chemistry, vol. 40, 1919–1930 (2019)
[2] Tamura, K., Sugitamoto, H., Shiro, Y. and Sugita, Y., Hemo-mechanical coupling in the transport cycle of a heme ABC transporter, The Journal of Physical Chemistry B, vol. 123, 7270–7281 (2019).
[3] Shinobu, A., Kobayashi, C., Matsunaga, Y. and Sugita, Y., Building a macro-mixing dual–basin Gō model using the multistate bennett acceptance ratio, Biophysics and Physicobiology, vol. 16, 310–321 (2019).
[4] Oshima, H., Re, S. and Sugita, Y., Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free–nergy Calculation of Biomolecules, Journal of Chemcal Theory and Computation, vol. 15, 5199–5208 (2019).
[5] Matsunaga, Y. and Sugita, Y., Use of single-molecule time-series data for refining conformational dynamics in molecular simulations, Current Opinon in Structural Biology, vol. 61, 153–159 (2020).
[6] Takada, S., Brandani, G. B. and Tan, C., Nucleosomes as allosteric scaffolds for genetic regulation. Current Opinion in Structural Biology, vol. 6, 93–101 (2020).
[7] Tamura, K. and Sugita, Y., Free Energy Analysis of a Conformational Change of Heme ABC Transporter BhuUV-T, Journal of Physical Chemistry Letter, vol. 11, 2824–2829 (2020).

10.5.2 Posters

[8] Shinobu, A., Simulating large-amplitude transitions in proteins with a coarse-grained model.Joint Annual Meeting of the 71st Japan Society for Cell Biology and 19th Protein Science Society of Japan, Kobe, Jun., 2019.
[9] Jung, J., Oshima, H., Kasahara, K,, Kobayashi, C., Mori. T. and Sugita, Y., A new MD integration enabling large time step from accurate temperature and pressure evaluations, The 57th Annual Meeting of The Biophysical Society of Japan, Miyazaki, Sep. 2019.
[10] Shinobu, A., Simulating large-amplitude transitions in proteins with a coarse-grained model, The 57th Annual Meeting of the Biophysical Society of Japan, Miyazaki, Sep. 2019.
[11] Tamura, K. and Sugita, Y., Free energy analysis of the conformational changes of a heme ABC importer BhuUV-T, The 57th Annual Meeting of The Biophysical Society of Japan, Miyazaki, Sep. 2019.
[12] Kobayashi, C., Elucidation of a mechanism of ion transport in membrane transport protein, The 6th project report meeting of the HPCI system including K computer,Tokyo, Nov. 2019.
[13] Tamura, K. and Sugita, Y., Free energy analysis of the conformational changes of a heme ABC transporter BhuUV-T, The 5th International Conference on Molecular Simulation (ICMS2019), Jeju, Korea, Nov. 2019.
[14] Sugita, Y., Mori, T., Yagi, K., Oshima, H., Jung, J., Kobayashi, C. and Matsunaga, Y., Development of Molecular Dynamics Software, GENESIS 1.4 for Large Biological Systems, The 64th Biophysical Society Annual Meeting, San Diego, USA. Feb. 2020

10.5.3 Invited Talks

[15] Sugita, Y., Machine Learning approach to link single–molecule FRET and MD simulations for understanding protein folding dynamics, Peking University Seminar, Beijing, China, May. 2019.
[16] Tamura, K. and Hayashi, S., Atomistically deciphering alternation access mechanism of the mitochondrial ADP/ATP carrier with molecular simulations, The 2nd Tokyo ATPase Workshop, Tokyo, Sep. 2019.
[17] Sugita, Y., Machine learning approach to link MD simulation with single-molecule experiment for understanding protein folding and dynamics, The Fifth Korean-Polish Conference on "Protein Folding: Theoretical and Experimental Approaches", Seoul, Korea, Sep. 2019.
[18] Sugita, Y., Enhanced Conformational Sampling Methods for Free Energy Calculation, CHARMM-GUI KIAS School, Seoul, Korea, Sep. 2019.
[19] Sugita, Y., Free-Energy Calculations of Protein-Ligand Binding using GENESIS software, The 5th International Conference on Molecular Simulation (ICMS2019), Jeju, Korea, Nov. 2019.
[20] Jung, J. and Sugita, Y., Optimal temperature and pressure evaluations for an MD integration with a large time step, The 5th International Conference on Molecular Simulation (ICMS2019), Jeju, Korea, Nov. 2019.
[21] Sugita, Y., Free-energy landscapes of protein-ligand bindings along the positional and orientational ligand coordinates, Warsaw University CeNT Seminar, Warsaw, Poland, Nov. 2019.
[22] Sugita, Y., Pushing forward the limit of molecular dynamics simulation by the development of GENESIS, ENS Seminar, Paris, France, Nov. 2019.
[23] Sugita. Y., Development of GENESIS software for molecular simulations in cellular environments, Sorbonne University Seminar, Paris, France, Nov. 2019.
[24] Tamura, K., Comprehensive analysis of a heme acquisition process by a heme transporter, The 6th project report meeting of the HPCI system including K computer, Tokyo, Nov. 2019.
[25] Kobayashi, C., Development of Molecular Dynamics Software, 11th Symposium on Automatic Tuning Technology and its Application, Tokyo, Dec. 2019.

10.5.4 Oral Talks

[26] Jung, J., Kobayashi, C., and Sugita, Y., The billion-atom simulation in biology with GENESIS on Intel Xeon Phi (KNL), The 33th Molecular Simulation conference, Nagoya, Dec. 2019
[27] Shinobu, A., Building a macro-mixing Dual-basin Gō model using the Multistate Bennett Acceptance Ratio, The 64th Biophysical Society Annual Meeting, San Diego, USA. Feb. 2020

10.5.5 Software

[28] Molecular dynamics and modeling software GENESIS, https://www.r-ccs.riken.jp/labs/cbrt (version 1.4)

10.5.6 Patents