Computational Biophysics Research Team

1. Team members

Yuji Sugita (Team Leader)

Osamu Miyashita (Senior Research Scientist)

Jaewoon Jung (Research Scientist)

Chigusa Kobayashi (Research Scientist)

Raimondas Galvelis (Postdoctoral Researcher)

Yasuhiro Matsunaga (RIKEN Special Postdoctoral Researcher)

Naoyuki Miyashita (Research Scientist (Concurrent))*

Tadashi Ando (Research Scientist (Concurrent))*

Yasuhito Karino (Postdoctoral Researcher (Concurrent))*

Yumi Kashihara (Postdoctoral Researcher (Concurrent))*

Takaharu Mori (Research Scientist (Concurrent))**

Takao Yoda (Visiting Scientist) ***

Mitsunori Ikeguchi (Visiting Scientist)****

Hiromi Kano (Assistant (Concurrent))*

* The main affiliation of these people is Laboratory for Biomolecular Function Simulation, Computational Biology Research Core, RIKEN Quantitative Biology Center.

** The main affiliation is RIKEN Theoretical Molecular Science Laboratory.

*** The main affiliation is Nagahama Bio Institute.

**** The main affiliation is Yokohama City University.

2. Research Activities

In molecular biology, atomic structures of proteins and other biomolecules provide essential information for understanding their biomolecular functions. Recently, MD simulations of biomolecules in solution or in biological membrane are often performed to elucidate the relationship between conformational dynamics and biomolecular functions. However, the conventional approaches have, at least, two major difficulties and cannot be compared directly to the experimental data. The first one is that simulation time of all-atom MD simulation is limited to about microsecond and this time scale is much shorter than that of slow conformational dynamics of proteins. The second difficulty is that the cellular environments are hardly involved in the MD simulations due to the size limitation of MD simulation. In this team, we have developed novel high-performance MD software, which we call GENESIS, to perform MD simulations of biomolecules efficiently on K computer. We aim to perform biomolecular simulations under realistic cellular environments as long as possible. The development of new

algorithms and the use of multi-scale and multi-resolution models are effective for large-scale MD simulations. In this team, we develop these methods and models in biomolecular simulations, also.

3. Research Results and Achievements

3.1. New Inverse Lookup Table for the evaluations of nonbonded interactions

We have developed a new lookup table for efficient short-range non-bonded interactions. Major bottleneck in MD is the calculation of non-bonded interactions of van der Waals and electrostatic. With spherical truncation (cutoff approximation) and particle mesh Ewald (PME), calculation order or van der Waals and real space electrostatic is reduced from O(n2) to O(n). However, these interactions are still the main bottleneck of MD, and they include very time-consuming inverse square roots and complementary error functions. To avoid such time-consuming operations while keeping accuracy, we proposed a new lookup table for short-range interaction in PME by defining energy and gradient as a linear function of inverse distance squared. In our lookup table approach, the table density is proportional to the inverse of squared distance. The new table increases accuracy by assigning large number of points at small pair distances where energy/gradients changes rapidly (Figure 1a). Despite of inverse operations in our approach, the new lookup table scheme allows fast evaluation due to small cache misses (Figure 1b). Overall, linear 1/R2 lookup table is highly promising for MD from the point of view of both accuracy and efficiency.

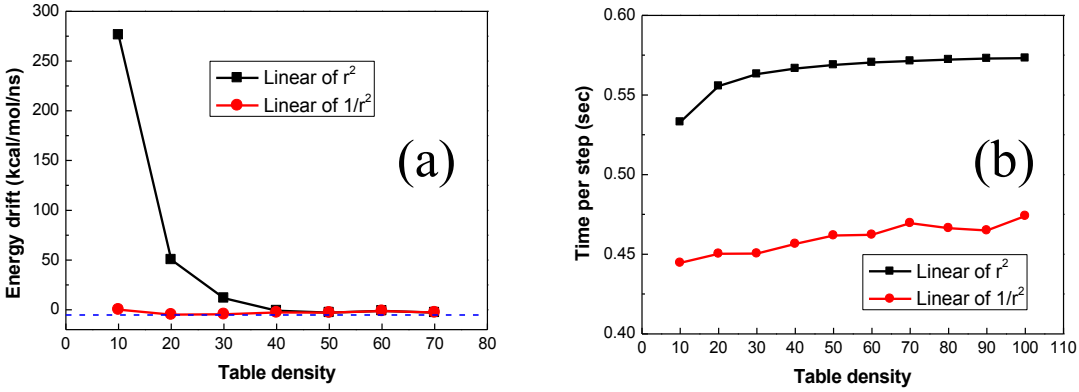

Figure 1. (a) Energy drift value according to the number of table points in unit section. (b) Computational time for one step calculation.

3.2 Midpoint Cell Method for hybrid parallelization

We have developed a new hybrid (MPI+OpenMP) parallelization scheme for molecular dynamics (MD) simulations by combining a cell-wise version of the midpoint method with pair-wise Verlet lists. In this scheme, which we call the midpoint cell method, simulation space is divided into subdomains, each of which is assigned to a MPI processor. Each subdomain is further divided

into small cells. The interaction between two particles existing in different cells is computed in the subdomain containing the midpoint cell of the two cells where the particles reside. In each MPI processor, cell pairs are distributed over OpenMP threads for shared memory parallelization. The midpoint cell method keeps the advantages of the original midpoint method, while filtering out unnecessary calculations of midpoint checking for all the particle pairs by single midpoint cell determination prior to MD simulations. Distributing cell pairs over OpenMP threads allows for more efficient shared memory parallelization compared with distributing atom indices over threads. Furthermore, cell grouping of particle data makes better memory access, reducing the number of cache misses. The parallel performance of the midpoint cell method on the K computer showed scalability up to 512 and 32,768 cores for systems of 20,000 and 1 million atoms, respectively. One MD time step for long-range interactions could be calculated within 4.5 ms even for a 1 million atoms system with PME electrostatics (Figure 2).

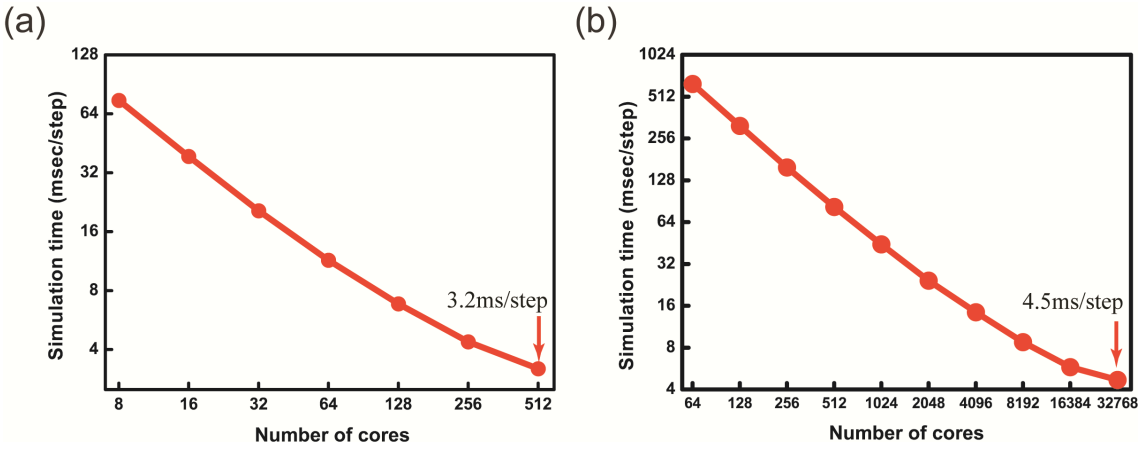

Figure 2. Simulation time (ms/step) for (a) 22,000 and (b) 1 million atoms systems.

3.3 Development of high-performance software GENESIS

GENESIS (Generalized Ensemble Simulation System) is a suite of computer program for carrying out MD for biomolecular systems. While most of MD programs have been parallelized for distributed memory parallelization for small and intermediate size systems (< 1 million atoms), GENESIS is optimized with hybrid parallelization by combining MPI with OpenMP for large-scale simulations. For fast evaluation of MD, we introduced lookup table approach and domain decomposition named midpoint cell method, which are already written in the above sections. In GENESIS, we have two simulators: ATDYN (atomic decomposition dynamics) and SPDYN (spatial decomposition dynamics). The former is easy to be modified for developing new methods due to simple parallelization using atomic decomposition, and enhanced sampling algorithms like replica-exchange molecular dynamics (REMD) is available. In ATDYN, in particular, there are special generalized ensemble algorithms named "surface-tension replica-exchange" developed

in our group. SPDYN was written mainly for efficient parallelization and fast evaluation for large systems. For efficient parallelization, Fast Fourier Transform (FFT) that shows the best parallel performance out of all MD programs is also optimized. SPDYN is optimized for K supercomputer, leading 6 ns/day for 100 million atoms system. This is an impressing result because the performance is almost twice faster than that of NAMD on Blue Gene/Q.

GENESIS has the following features:

1) Coarse-grained as well as explicit all-atom MD is available in ATDYN.
2) Parallel input/output (I/O) is available for very large system for efficient memory usage and fast setup.
3) GENESIS is optimized for K supercomputer, but it is also available on PC-clusters.
4) Everything is written in Fortran 90/95/2003 with dynamics memory allocation.
5) GENESIS is free software licensed under GPL version 2.

3.4 Data assimilation algorithm for analyzing conformational dynamics of biomolecules

We have been developing an algorithm for data-assimilation simulations incorporating single-molecule Förster resonance energy transfer (smFRET) measurements. SmFRET measurement is a powerful technique to investigate dynamic behavior of biomolecules as a function of time. However, the interpretation of smFRET data is sometimes difficult since the information is limited only to the distance-like information between two fluorescence dyes. We have been developing a data-assimilation technique, based on the particle filter, to interpret the smFRET data in terms of coarse-grained protein models. This year, we have formulated a likelihood function for smFRET photon counting data, by modeling the numbers of observed photons from the two dyes as inhomogeneous Poisson processes. We have implemented the likelihood function in GENESIS and tested the performance of the algorithm by using a simulated FRET-like photon counting data on K computer. Using polyproline as a test case, we have confirmed the performance using 131,072 replicas (particles) and 8,192 nodes of K computer.

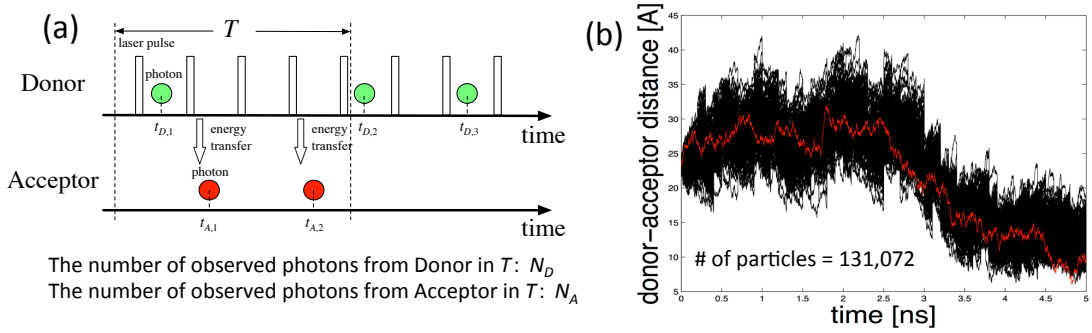

Figure 3. (a) Schematic picture of smFRET photon counting data. (b) Inference for the distance between two fluorescence dyes attached to polyproline from simulated smFRET photon

counting data. The red line indicates the true answer, and black lines are replicas (particles).

3.5 Motion tree algorithm for analysis of large domain motions of proteins

Proteins are known to take their own three-dimensional structures in physiological conditions. The structures are experimentally determined in crystal conditions (with X-ray crystallography) or in solution (with NMR or others). However, in physiological conditions or in cellular environments, proteins don't behave as rigid bodies but show significant flexibility due to thermal noises. Furthermore, some proteins undergo large domain motions in their reaction cycle, utilizing ATP hydrolysis or proton motive forces. Sarco(endo)plasmic reticulum Ca^{2+}-ATPase (SERCA), which transports Ca^{2+} across biological membranes against a large concentration gradient, is one of the best-studied membrane proteins. In classical E1/E2 theory, SERCA takes at least two different physiological states, E1 and E2: in the E1 state, the transmembrane binding sites have high affinities for Ca^{2+}, whereas the affinities are greatly reduced in the E2 state. In addition to the binding and release of Ca^{2+}, ATP hydrolysis and dephosphorylation at the phosphorylation residue, Asp351, introduces more physiological states.

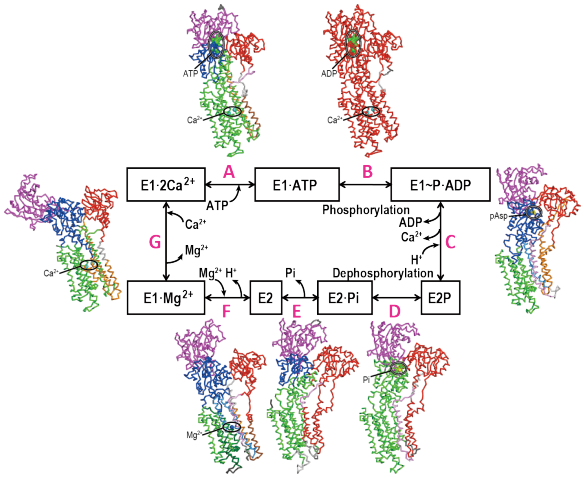

Figure 4. Reaction cycle of SERCA and rigid domains in MTs on reaction steps

To characterize its conformational motions, we illustrate 'Motion Tree (MT)' based on seven crystal structures of SERCA. MT is a tree diagram that represents hierarchical domain-motion. (Koike *et al.*, *J. Mol. Biol.*, 2014) We investigate the relationship between local conformational changes and function of SERCA based on MTs. In addition, we determine 'common rigid domains (CRD)' that keep their structural rigidity during the whole reaction cycle. The analysis allows discussion of how the protein utilizes both structural rigidity and flexibility for pumping Ca^{2+} across the membrane. We also investigate local conformational changes upon a dissociation of Pi and Mg^{2+} from the nucleotide-binding site using atomistic molecular dynamics (MD) simulations. The simulations reinforce the notion of a conformational change upon binding/dissociation of the ligands. We emphasize that MT detects such motions automatically without extensive biological knowledge, suggesting general applicability to domain movements in other membrane proteins to deepen the understanding of protein structure and function.

3.6 Development of new meta-dynamics algorithms

The understanding of biological systems by atomistic-level simulations requires free energy calculations, which inherently is a problem of conformation space sampling. Metadynamics, an adaptive-biasing technique, has proven its efficiency to accelerate sampling. The method estimates the free energy by iteratively updating a biasing potential in a predefined collective variable space. In particular, we were focused on the multi-replica algorithms of metadynamics, which could be efficiently implemented on the massively parallel computers (such as K computer). Currently ATDYN supports several mutil-replica algorithms: multiply-walker, parallel-tempered, and bias-exchange. We have demonstrated the advantages of metadynamics to enhance and parallelize sampling effort with several systems, including alanine pentapeptide (Figure 5). Finally, novel replica-exchange schemes are being investigated to increase efficiency of the multi-replica metadynamics by optimizing exchange rates and patterns. Additionally, this allows a larger number of collective variables to be used, enabling the efficient simulations of more complex systems and phenomena.

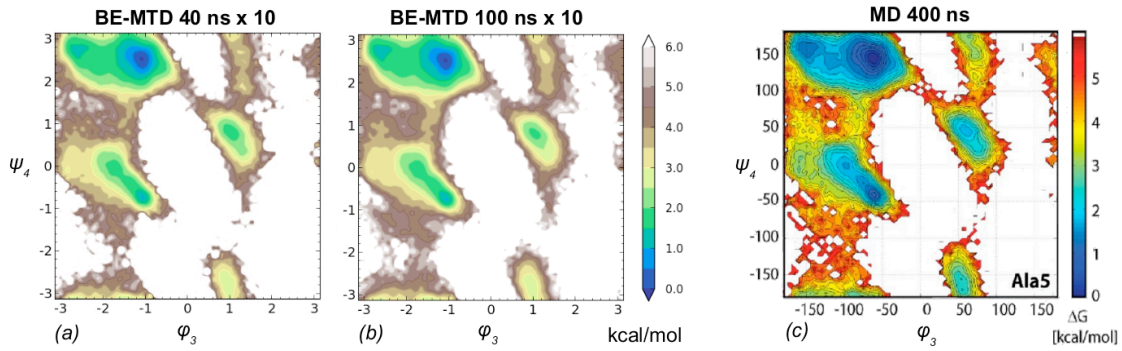

Figure 5. Free energy surface along two backbone dihedral angles (φ_3, ψ_4) of alanine pentapeptide (Ala5) obtained from (a) 40 ns and (b) 100 ns of 10-replica bias-exchange metadynamics (BE-MTD) simulations, and, for comparison, from (c) 400 ns of an MD simulation (adapted from R.B. Best, et al., *J. Chem. Theory Comput.*, 2012, 8(9): 3257–3273).

3.7 Computational analysis of low-resolution structural data from XFEL and EM

We have been developing algorithms to construct atomistic models from low-resolution structural data. Cryo-EM and newly emerging XFEL experiments provides new structural information that are not available in traditional X-ray crystallography, since these experiments can be performed without the crystallization of target systems. However, on the other hand, the data from Cryo-EM and XFEL are at low-resolution without atomic details, and thus need to be complimented by other information to construct atomic models. We have been implementing the algorithms in GENESIS to perform flexible fitting of atomic structures into such low-resolution data. Using generalized ensemble algorithms embedded in GENESIS, the accuracy and efficiency of fittings can be enhanced.

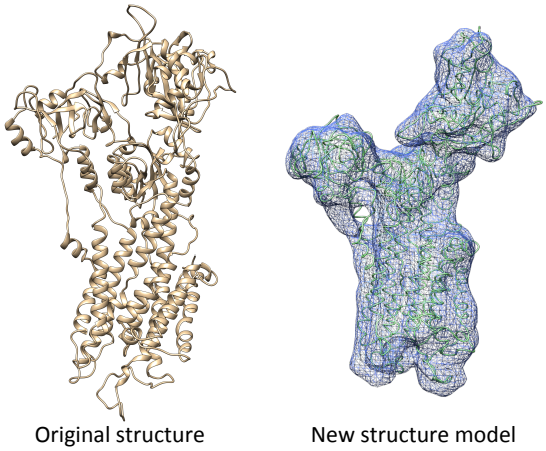

Figure 6. A result of flexible fitting using GENESIS. The original structure (left) is deformed using molecular dynamics simulation to fit the low-resolution data.

4. Schedule and Future Plan

We have released the first version of GENESIS program as free software under GPL license v2. We continue to develop the software for improving its performance in MD simulations and adding new functions and molecular models. To show the performance and reliability of GENESIS on K computer, we will perform simulations of large biomolecules like Ribosomes, membrane proteins, and so on.

We have already implemented the efficient evaluations of non-bonded interactions and their forces. Another time consuming part is non-bonded reciprocal interactions in particle mesh ewald (PME) approximation. In PME, the reciprocal interaction is evaluated using FFT (Fast Fourier Transform) computation, which usually show poor parallel scalability. We plan to improve the performance of FFT on parallel computers.

Multi-scale and multi-resolution models are important to simulate large-scale conformational changes of membrane proteins or protein complexes under cellular environment. In GENESIS, we will introduce these models for efficient conformational sampling of large biomolecular systems. We also plan to implement QM/MM hybrid simulations on GENESIS for simulating enzyme reactions.

5. Publication, Presentation and Deliverables

(1) Journal Papers

[1] Jaewoon Jung, Takaharu Mori, and Yuji Sugita: "Midpoint Cell Method for Hybrid (MPI+OpenMP)

Parallelization of Molecular Dynamics Simulations", J. Comput. Chem., in press.

[2] L.S. Ahlstrom, and Osamu Miyashita: "Packing interface energetics in different crystal forms of the λ Cro dimer", Proteins, in press.

[3] Takaharu Mori, Jaewoon Jung, and Yuji Sugita: "Surface-tension replica-exchange molecular dynamics method for enhanced sampling of biological membrane systems", J. Chem. Theory Comput. **9** (2013) 5629-5640.

[4] Jaewoon Jung, Takaharu Mori, and Yuji Sugita: "Efficient lookup table using a linear function of inverse distance squared", J. Comput. Chem. **34** (2013) 2412-2420.

[5] Logan S. Ahlstrom, Joseph Lee Baker, Kent Ehrlich, Zachary T. Campbell, Sunita Patel, Ivan I. Vorontsov, Florence Tama, Osamu Miyashita: "Network visualization of conformational sampling during molecular dynamics simulation", J. Mol. Graph Model. **46** (2013) 140-149.

[6] Yasuhiro Matsunaga, Akinori Baba, Chun-Biu Li, John E. Straub, Mikito Toda, Tamiki Komatsuzaki, and R. Stephen Berry: "Spatio-temporal hierarchy in the dynamics of a minimalist protein model", J. Chem. Phys. **139** (2013) 215101.

(2) Invited Talks

[7] Yasuhiro Matsunaga: "Finding Conformational Transition Pathways in Biomolecules with the String Method and Sequential Data Assimilation", Rare Event Sampling and Related Topics I, ISM, Tokyo, Japan, March 4-5, 2014.

[8] Jaewoon Jung: "Development of GENESIS for large scale molecular dynamics simulation", Bio-Supercomputing Winter school, Atagawa, January 23-24, 2014.

[9] Osamu Miyashita, Atsushi Tokuhisa, Florence Tama: "An Overview of Single Biomolecular Imaging by X-ray Free Electron Laser", Institut de Minéralogie et de Physique des Milieux Condensés, France, December 16, 2013.

[10] Jaewoon Jung: "Development of GENESIS for large scale molecular dynamics simulation", Workshop on Molecular Simulations of Biophysics and Biochemistry, Kobe, November 21, 2013.

[11] Yasuhiro Matsunaga: "Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations", Workshop on Molecular Simulations of Biophysics and Biochemistry, Kobe, November 21, 2013.

[12] Yuji Sugita, Ryuhei Harada, Isseki Yu, Takaharu Mori, Jaewoon Jung, and Michael Feig: "Biomolecular Simulations under Cellular Crowding Environment", ICMS 2013, Kobe, November 18-20, 2013.

[13] Jaewoon Jung, "Development of GENESIS for large scale molecular dynamics simulation", CMSI International Satellite Meeting 2013 in Nagoya, Nagoya, October 17-19, 2013.

[14] Yuji Sugita: "Replica-Exchange Molecular Dynamics Simulations of Membrane Protein Systems", The Snowmass Biophysics Workshop on Free-Energy Calculations, Snowmass,

Colorado, USA, July 15-19, 2013.

[15] Yuji Sugita: "Molecular Dynamics Simulations of MATE multi-drug transporter", The Snowmass Biophysics Workshop on Membrane and Membrane Proteins, Snowmass, Colorado, USA, July 22-26, 2013.

[16] Osamu Miyashita: "Effect of Crystal Packing on Protein Conformation and Dynamics", Nagoya University, Nagoya, May 27, 2013.

[17] Yuji Sugita and Takaharu Mori: "Surface Area in Protein-Membrane Simulation Systems, Biophysical Society Meeting on Membrane Protein Folding", Seoul, South Korea, May 19-22, 2013.

(3) Posters and Presentations

[18] Chigusa Kobayashi and Yuji Sugita: "Conformational change of SERCA upon alternating protonation states in Ca^{2+}-binding site", The 4th AICS International Symposium, Kobe, December 2-3, 2013.

[19] Jaewoon Jung, Takaharu Mori, and Yuji Sugita, "Midpoint cell method for hybrid (MPI+OPENMP) parallelization of Molecular Dynamics", The 4th AICS International Symposium, Kobe, December 2-3, 2013.

[20] Jaewoon Jung, Takaharu Mori, and Yuji Sugita: "Midpoint cell method for hybrid (MPI+OPENMP) parallelization of Molecular Dynamics", ICMS 2013, Kobe, November 18-20, 2013.

[21] Isseki Yu, Takaharu Mori, Jaewoon Jung, Ryuhei Harada, Yuji Sugita, and Michael Feig: "All-atom Modelling and Molecular Dynamics Simulation of the Cytoplasm of Mycoplasma Genetalium", ICMS 2013, Kobe, November 18-20, 2013.

[22] Chigusa Kobayashi and Yuji Sugita: "Conformational change of SERCA upon alternating protonation states in Ca^{2+}-binding site", ICMS 2013, Kobe, November 18-20, 2013.

[23] Chigusa Kobayashi and Yuji Sugita: "Conformational change of SERCA upon alternating protonation states in Ca^{2+}-binding site", ICMS 2013, Kobe, November 18-20, 2013.

[24] Yasuhiro Matsunaga, Takaharu Mori, Jaewoon Jung, and Yuji Sugita: "Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations", Workshop on Modeling Biomolecular Systems in Cellular Environments, Kyoto, October 31 - November 1, 2013.

[25] Raimondas Galvelis and Yuji Sugita: "Metadynamics: Implementation in GENESIS and Demonstration of Efficient Simulations", Workshop on Modeling Biomolecular Systems in Cellular Environments, Kyoto, October 31 - November 1, 2013.

[26] Yasuhiro Matsunaga, Takaharu Mori, Jaewoon Jung, and Yuji Sugita: "Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations", Workshop on Modeling Biomolecular Systems in Cellular Environments, Kyoto,

October 31 - November 1, 2013.

[27] Jaewoon Jung, Takaharu Mori, and Yuji Sugita: "Efficient Lookup Table using a Linear Function of Inverse Distance Squared", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[28] Isseki Yu, Takaharu Mori, Jaewoon Jung, Ryuhei Harada, Yuji Sugita, and Michael Feig: "All-Atom Molecular Dynamics Simulation of Bacterial Cytoplasm", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[29] Takaharu Mori, Jaewoon Jung, and Yuji Sugita: "Acceleration of lipid lateral diffusion by generalized-ensemble molecular dynamics simulation", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[30] Chigusa Kobayashi and Yuji Sugita: "Conformational change of SERCA upon alternating protonation states in Ca^{2+}-binding site", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[31] Raimondas Galvelis and Yuji Sugita: "Metadynamics: Implementation in GENESIS and Demonstration of Efficient Simulations", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[32] Yasuhiro Matsunaga, Takaharu Mori, Jaewoon Jung, and Yuji Sugita: "Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations", The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.

[33] 小林千草、小池亮太郎、太田元規、杉田有治: "Motion Tree 法を用いた SERCA のリガンド解離における構造変化の解析", 第 13 回日本蛋白質科学会年会, 鳥取、2013 年 6 月 12 日.

[34] Chigusa Kobayashi, Ryotaro Koike, Motonori Ota, and Yuji Sugita: "Conformational changes of SERCA upon dissociation of ligand analyzed with Motion Tree method", Membrane protein folding meeting, Korea, May 20, 2013.

(4) Patents and Deliverables

[35] Generalized-Ensemble Simulation System (GENESIS) is released. 2014/03.
https://aics.riken.jp/labs/cbrt/
http://www.riken.jp/TMS2012/cbp/en/research/software/genesis/index.html