Computational Biophysics Research Team

1. Team members

Yuji Sugita (Team Leader (Concurrent))*, **
Osamu Miyashita (Senior Research Scientist)
Jaewoon Jung (Research Scientist)
Chigusa Kobayashi (Research Scientist)
Yasuhiro Matsunaga (Research Scientist)
Naoyuki Miyashita (Research Scientist (Concurrent))*
Tadashi Ando (Research Scientist (Concurrent))*
Yasuhito Karino (Postdoctoral Researcher (Concurrent))*
Yumi Kashihara (Postdoctoral Researcher (Concurrent))*
Hiromi Kano (Assistant (Concurrent))*
Takaharu Mori (Research Scientist (Concurrent))**
Isseki Yu (Research Scientist (Concurrent))**
Yasuaki Komuro (Student Trainee (Concurrent))**
Raimondas Galvelis (Postdoctoral Researcher (Concurrent))**
Takao Yoda (Visiting Scientist) ***
Mitsunori Ikeguchi (Visiting Scientist)****

* 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

Molecular dynamics (MD) simulation of biomolecules has become one of the essential research tools in biophysics and biochemistry. The simulation is, in particular, useful for investigating relationships between biomolecular structure and function as well as those between the conformational dynamics and function. One of the major advantages in the simulation is its atomically detailed description of biomolecular structures and dynamics, although the time-scale (< microseconds) is relatively shorter than the typical experimental ones (> milliseconds). Many researchers, therefore, have tried to extend the simulation sizes and lengths by optimization of MD software and/or its algorithms. In our research team, we have developed new high-performance MD software, which we call GENESIS (GENeralized-Ensemble SImulation System) to perform MD simulations of biomolecules efficiently on K computers. At the end of last fiscal year (March, 2014),

we provided the pre-released version of GENESIS as free software under the license of GPLv2 on the team website. In this fiscal year, we developed new algorithms for further optimizations on K computer, sequential data assimilation for combing experimental measurements with simulations, and coarse-grained MD simulations.

3. Research Results and Achievements

3.1 Volumetric decomposition FFT in GENESIS

Many biological applications such as protein folding, aggregation, protein-protein interaction, and so on demand us longr time-scale molecular dynamics (MD) simulations, because these biological phenomena typically occur in the time scale of μsec or msec. For the long time MD simulation, efficient parallelization of MD is essential. However, in the particle-mesh ewald (PME) calculation for long-range electrostatic interaction, the conventional parallelization schemes of fast Fourier transform (FFT) have limited the total performance of MD simulations in massively parallel supercomputers. To overcome the problem, we have developed a new parallelization scheme of FFT. Our method basically makes use of the volumetric decomposition scheme with hybrid parallelization (MPI+OpenMP), which is particularly useful when used in conjunction with the midpoint cell scheme developed in GENESIS. The newly developed parallelization of FFT includes two schemes: 1d_Alltoall with five all-to-all communications in one dimension, and 2d_Alltoall with one two-dimensional all-to-all communications combined with two all-to-all communications in one dimension. Both schemes show comparable or superior performance than existing FFT libraries. The parallel performance on K computer shows the world fastest records of FFT calculations for 512^3 grids and 1024^3 grids, to our best knowledge, improving the performance on large-scale MD simulations greatly. Due to efficient parallelization of FFT, now GENESIS can finish one MD cycle for a virus system containing more than 1 million atoms within 5 ms using 32,768 cores on K computer.

3.2. Development of GENESIS on hybrid CPU+GPUs

The main bottleneck in MD simulations is the calculation of pair-wise non-bonded interactions. Recently, graphics processing units (GPU) become very powerful tools to accelerate the time-consuming real space non-bonded interactions. Nowadays, a lot of MD programs including CHARMM, NAMD, Gromacs, Amber, AceMD, OpenMM, and so on support the use of GPUs. Some of them focus on optimizing on single node without using CPU to remove the data transfer between CPU and GPU. Unlike these programs optimized for single node, we have developed hybrid CPU+GPU calculations on multiple nodes for large-scale MD simulations. First, we introduced new non-bonded interaction scheme suitable for GPU. CPUs have large amounts of cache memory to optimize random memory access. On the other hand, GPUs include very small amount of

cache memory, and instead are optimized for in-order memory access. The global memory size of GPU is usually less than that of CPU. Considering these differences between CPU and GPU, we introduced a new pair-list scheme, which depends on not the pair of atoms but just atom list itself. Although this scheme increases the amount of calculations, computational time could be reduced due to efficient thread calculations on GPU. Moreover, we properly assigned mixed precisions where calculation on GPU is performed using single precision floating points while accumulation is dealt with double precision floating points. Our program is tested on TSUBAME for 1 million atom systems, showing the expected MD simulation time is around 30ns/day using 128 nodes, which is equivalent to the production using 2048 nodes of K computer.

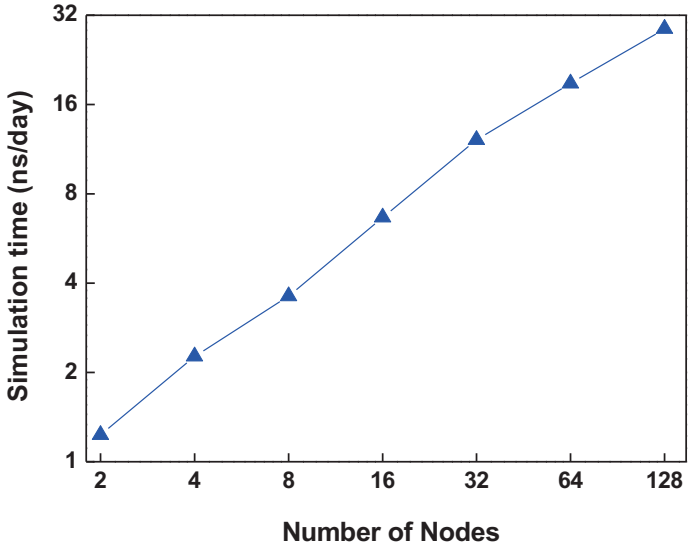

Figure 1. Simulation time versus number of nodes for hybrid CPU+GPU usage of MD simulation.

3.3 Sequential data assimilation for single-molecule FRET data

We have been developing a framework for data assimilation of single-molecule Forster resonance energy transfer (smFRET) data. Data assimilation is a statistical method designed to improve the quality of numerical simulations in combination with real observations. We develop a sequential data assimilation method that incorporates one-dimensional time-series data of smFRET photon-counting into conformational ensembles of biomolecules derived from “replicated" molecular dynamics (MD) simulations. A particle filter using a large number of “replicated" MD simulations with a likelihood function for smFRET photon-counting data is employed to screen the conformational ensembles that match the experimental data. We examined the performance of the method using emulated smFRET data and coarse-grained (CG) MD simulations of a dye-labeled

polyproline-20. The method estimated the dynamics of the end-to-end distance from smFRET data as well as revealing that of latent conformational variables. We also confirmed that the particle filter is also able to correct model parameter dependence in CG MD simulations.

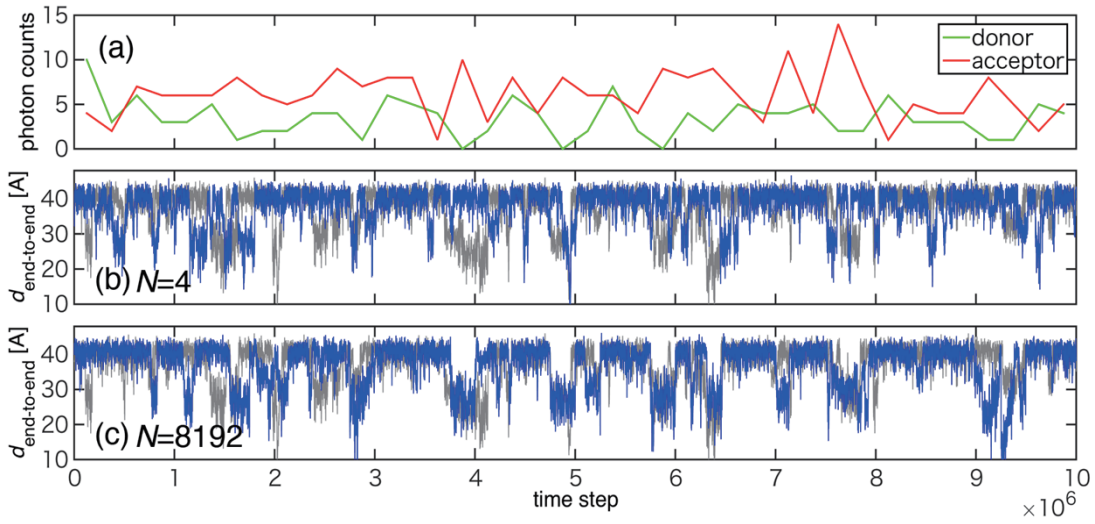

Figure 2. MAP estimates for the end-to-end distance obtained by the particle filter using N=4, and 8192 particles. (a) Emulated smFRET photon-counting data. The photon-counts from the donor and acceptor dyes are indicated by the green and red lines, respectively. (b) The true end-to-end distance obtained in the emulation simulation for smFRET data. (b) MAP estimates by the particle filter using *N*=4 particles is shown with the blue line. The target end-to-end distance (the emulation trajectory data) is indicated with the gray line. (c) MAP estimates by using N=8192 particles.

3.4. Development of coarse-grained MD simulations

Large conformational changes of multi-domain proteins are difficult to be simulated using all-atom models due to the slow time scale of motions. We propose a simple modification in the structure-based coarse-grained (CG) model for stable but efficient MD simulations of those proteins. We first obtain information on the definition of "dynamic domains" and the magnitude of local domain motions in a protein through "Motion Tree", a tree diagram that describes conformational changes in a hierarchical manner from two structures. (Koike *et al.*, *J. Mol. Biol.*, 2014) We recently investigated conformational changes of Ca^{2+}-pump, a multi-domain membrane protein, based on the method and showed that the analysis detects functional motions without extensive biological knowledge of experts. (Journal #2)

In our new CG model, which we call DoME (Domain-Motion Enhanced) model, only the native-contact interaction parameters between two domains are modified as being inversely proportional to the magnitude in "Motion Tree". We applied simulations based on DoME to a small

soluble protein, Adenylate Kinase (AdK), examining the conformational fluctuation in the open state and structural transition between its open and closed states. The simulation results based on DoME are consistent with a 10μsec all-atom MD, previous CGMD simulations, and experimental data. In addition, the simulations based on the model are shown to provide stable trajectories regardless of some temperature differences. We also investigate the conformational transition from the open to closed states in AdK by using a dual-basin Go-model via a perturbation approach proposed by Whitford et al. The comparison suggests that DoME allows us to apply CG MD simulations more easily for multi-domain proteins and provides a stable trajectory that is comparable to experimental studies and long-time all-atom MD simulations.

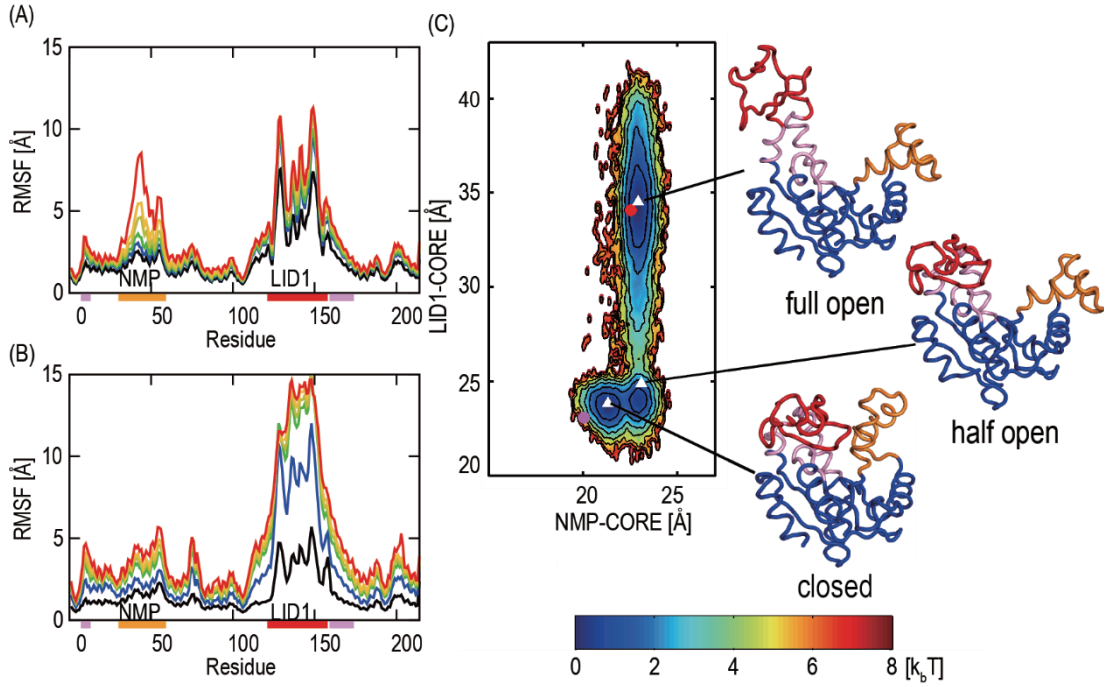

Figure 3. (A-B) RMSFs with different models; (A) the DoME model, (B) the KB (previous) model. Black, blue, green, yellow-green, orange, and red lines are corresponds at different temperatures (C) Free energy surfaces of domain motions with the perturbation DoME model. Vertical and horizontal axes indicate the distance between centers of mass (COMs) of CORE and LID1, and that between COMs of CORE and NMP, respectively. Right side shows structures of three centroids on the free energy surface. (White triangles) Full open, half open, and closed structures are displayed.

3.5. Implementation of new functions in GENESIS

After opening the first version of GENESIS (journal #1), we have implemented the following new functions; 1. GPGPU implementation, (explained in 3.2), 2. Multiple time step integrator (Reversible System Propagator Algorism, RESPA). 3. Enhance available force fields with all atom and CG

models. For efficient sampling with multiple time step integrator, we basically implemented RESPA (Tuckerman *et al, J. Chem. Phys.* 1992) in GENESIS. Long-range interaction described by reciprocal-space interaction is defined as slow motion force and skipped periodically. Based on RESPA, we assigned Langevin thermostat to generate statistically meaningful ensemble spaces. It is revealed that our Langevin RESPA shows better thermodynamic and dynamic properties than the multiple time step integrators using Langevin thermostat and extrapolation. It is also shown that the time step of the slow motion could be increased up to 8 fs with Langevin thermostat while 4-6 fs is the maximum time step with the conventional scheme. In the first version of GENESIS, the standard CHARMM force field files (topology and parameter) are available as input parameters. For coarse-grained model simulations, a Go-model potential developed by Karanicolas and Brooks (Karanicolas and Brooks, *J. Mol. Biol.* 2003) is available. In order to simulate more realistic system, e.g. complex with nucleicacids and glycol-proteins, modules for AMBER and Gromacs force field files have been newly installed. As for CG models, in addition DoME model (explained in 3.4), SMOG (Whitford *et al. Proteins*, 2008) and MARTINI (Marrink *et al . J. Chem. Theo. Comp.*, 2008) models are installed. SMOG and MARTINI models have higher resolution than Karanicolas and Brooks and DoME models. By implementation of different types of CG models, various simulations with multi-scale are available in GENESIS.

4. Schedule and Future Plan

In the next fiscal year, we plan to implement new functions in GENESIS, namely, the reaction path sampling, the flexible collective variables, hybrid QM/MM calculations, and so on. The 2nd version of GENESIS will be released at the end of this fiscal year (March, 2015) also as free software under GPL license v2. For this, we need to test many biological simulations for confirming stable MD simulations using GENESIS.

We will try to simulate slow conformational dynamics of membrane proteins and supramolecules like Ribosome using multi-resolution simulations. So far, we have developed high-performance MD simulations and coarse-grained MD simulations separately. In this fiscal year, we integrate these two simulation modules as well as the reaction path sampling module for this purpose. We will simulate large-scale free-energy calculations of membrane proteins and Ribosomes, using large number of CPUs on K computer.

Hybrid QM/MM calculation is also very important for investigating biological functions of enzymes and proteins. In this fiscal year, new functions including ab initio vibrational analysis and reaction path analysis will be installed in GENESIS in combined with the existing QM programs.

5. Publication, Presentation and Deliverables

(1) Journal Papers

1. Jaewoon Jung, Takaharu Mori, Chigusa Kobayashi, Yasuhiro Matsunaga, Takao Yoda, Michael Feig, and Yuji Sugita, “GENESIS: A hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations”, WIREs Computational Molecular Science 2015.
2. Chigusa Kobayashi, Ryotaro Koike, Motonori Ota and Yuji Sugita, “Hierarchical domain-motion analysis of conformational changes in sarcoplasmic reticulum Ca^{2+}-ATPase”, Proteins: Structure, Function, and Bioinformatics, 83, pp 746–756, (2015).

(2) Conference Papers

None.

(3) Invited Talks

1. “Molecular mechanism for the drug extrusion by MATE multi-drug transporter”, Yuji Sugita, Functional Computational Changes in Complex Biological Systems, Hong Kong, China, Apr. 2014.
2. “Replica-exchange molecular dynamics simulations for enhanced sampling of biological membrane systems”, Yuji Sugita, Computational Biophysics to Systems Biology (CBSB14), Gdansk, Poland, May. 2014.
3. “Atomistic motions of proteins and water in crowded environments”, Yuji Sugita, Telluride Workshop on Protein and Peptide Interactions in Cellular Environments, Telluride, USA, Jun. 2014.
4. “Domain motion analysis of proteins based on multiple atomic structures and coarse-grained molecular dynamics simulations”, Yuji Sugita, Telluride Workshop on Coarse-grained modeling of structure and dynamics of biomolecules, Telluride, USA, Aug. 2014.
5. “理論・計算化学に基づく脂質ダイナミクスの解析”, Yuji Sugita, 第 87 回日本生化学会大会, Kyoto, Japan, Oct. 2014.
6. “分子動力学計算の高速化と大規模生体分子シミュレーション”, Yuji Sugita, バイオスーパーコンピューティング研究会 2014(第 6 回)総会・講演会, Saitama, Japan, Oct. 2014.
7. 松永康佑 "最小自由エネルギー経路探索法による多剤排出トランスポーターの薬剤排出機構の解明", 第 1 回「京」を中核とする HPCI システム利用研究課題 成果報告会 東京, Oct. 2014.
8. Yasuhiro Matsunaga, “Molecular dynamics simulation studies using multi-copy based methods: string method and sequential data assimilation”, 5th AICS International Symposium, Kobe, Japan, Dec. 2014.

9. “Domain motion analysis of membrane proteins by bioinformatics and molecular dynamics simulations”, Yuji Sugita, The international workshop on computational science and engineering, Hong Kong, China, Dec. 2014.
10. “スーパーコンピュータを用いた細胞内分子ダイナミクスの解析”, Yuji Sugita, ISSP ワークショップ　機能物性融合科学研究会シリーズ(１)「光機能」, Kobe, Japan, Dec. 2014.
11. “Optimization of Molecular Dynamics Program “GENESIS” and its Application to Biomolecular System”, Yuji Sugita, WINTech2015, Kobe, Japan, Mar. 2015.
12. “細胞内分子ダイナミクスのシミュレーションの現状と今後の課題”, Yuji Sugita, よこはま NMR 研究会　第 52 回ワークショップ, Yokohama, Japan, Mar. March 2015.
13. Multiscale molecular dynamics simulations of transmembrane structures of amyloid precursor protein in biological membrane”, Yuji Sugita, 249th ACS National Meeting & Exposition, Denver, USA, Mar. 2015.
14. Jaewoon Jung and Yuji Sugita, “Multiple time step integrator in molecular dynamics (MD) program GENESIS”, IMS2015, Kyoto, March 2015.
15. Yasuhiro Matsunaga, “Drug extrusion mechanism of multidrug transporter AcrB studied by molecular dynamics simulations”, Rare Event Sampling and Related Topics II, The Institute of Statistical Mathematics, Tokyo, Japan, March 2015.

(4) Posters and presentations

1. 小林千草、小池亮太郎、太田元規、杉田有治、 “Conformational changes of SERCA in response to reactions described by hierarchical domain-motion analysis.”, 第 51 回日本生物物理学会年会, 北海道、2014/9/26.
2. 小林千草、小池亮太郎、太田元規、杉田有治、 “Conformational changes of SERCA in response to reactions described by hierarchical domain-motion analysis.”, AICS symposium, Kobe、2014/12/8-12/9.
3. Y. Matsunaga, and Y. Sugita "Sequential data assimilation of single-molecule FRET photon-counting data by using molecular dynamics simulations", Telluride Workshop on Coarse-Grained Modeling of Structure and Dynamics of Biomacromolecules, (口頭発表), 2014/8.
4. 松永康佑 "データ同化技術を用いた 1 分子 FRET 計測融合シミュレーションによるタンパク質動態の解明" 第 1 回「京」を中核とする HPCI システム利用研究課題 成果報告会 東京, 2014/10/31.
5. 松永康佑 "1 分子 FRET 光子計数データのモデリング" 新学術領域研究 スパースモデリングの深化と高次元データ駆動科学の創成 2014 年度 公開シンポジウム 東京工業大学, 2014/12/15-17.

6. 松永康佑 "1分子FRET光子計数データのモデリング" 新学術領域研究　スパースモデリングの深化と高次元データ駆動科学の創成 第 2 回領域会議 東京大学, 2014/6/19-21.2

(5) Patents and Deliverables

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