7. Computational Biophysics Research Team

7.1. Team members

Yuji Sugita (Team Leader)
Jaewoon Jung (Research Scientist)
Ryuhei Harada (Postdoctoral Researcher)
Yasuhiro Matsunaga (RIKEN Special Postdoctoral Researcher)
Naoyuki Miyashita (Research Scientist (Concurrent))*
Takaharu Mori (RIKEN Special Postdoctoral Researcher (Concurrent)*
Hiromi Kano (Assistant (Concurrent))*
* The main affiliation of these people is Laboratory for Biomolecular Function Simulation, Computational Biology Research Core, RIKEN Quantitative Biology Center.

7.2. Research Activities

Recent advances in structural biology, atomic structures of macromolecules can be determined by X-ray crystallography and nuclear magnetic resonance (NMR). These coordinates are stored in protein data bank (PDB) and are utilized for academic researches or industrial usages like drug design. This information is usually quite useful to understand molecular mechanisms underlying protein stability, large-scale conformational changes, ligand-receptor binding, and enzymatic reactions. However, due to the complexity of the structures, the static structural information is, in some cases, not enough to understand biological phenomena. Some of the proteins show large conformational changes during their functional cycles.

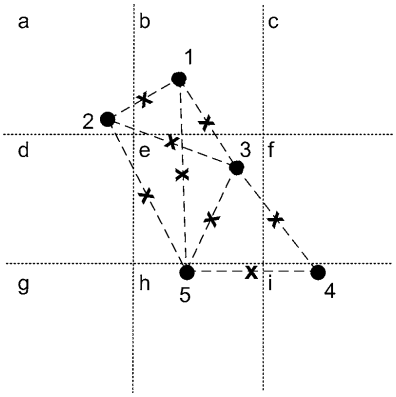
Computer simulations based on molecular dynamics (MD) method using the macromolecular structures now become important research tools in biological sciences. By performing the MD simulations in supercomputer, we can simulate molecular motions of proteins, which occur on the time scale between nsec and μsec. However, much longer simulations (from msec to sec) are highly desired to simulate whole processes of most of biological phenomena.

In our team, we aim to develop software for MD simulations to use K computer most efficiently. We also try to introduce novel algorithms or models into the software. Our software allows us not only to simulate dynamical properties of macromolecules, but also to provide thermodynamic quantities at physiological conditions. Currently, we focus on the parallelization of MD simulation code, the development of replica-exchange molecular dynamics (REMD) method or reaction-path search method and data-assimilation algorithm, and large-scale MD simulations of biological systems in cellular environment.

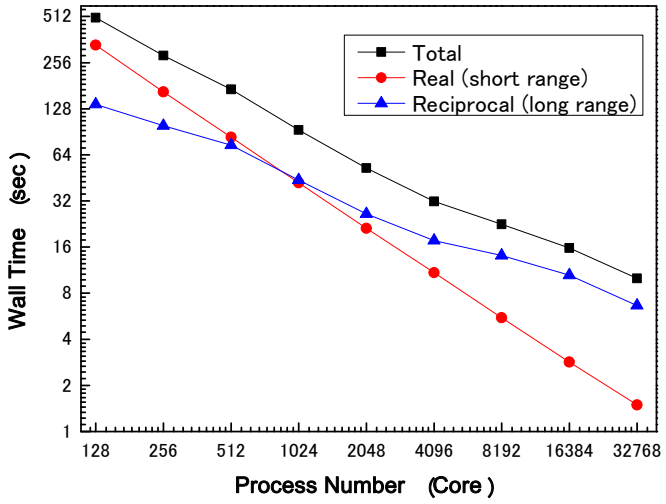
7.3. Research Results and Achievements

7.3.1. Parallelization of Molecular dynamics simulation codes

In SPDYN, we use the midpoint method as a basic scheme. In the midpoint method, two particles interact on a particular box if and only if the midpoint of the segment connecting them falls within the region of space associated with that box.

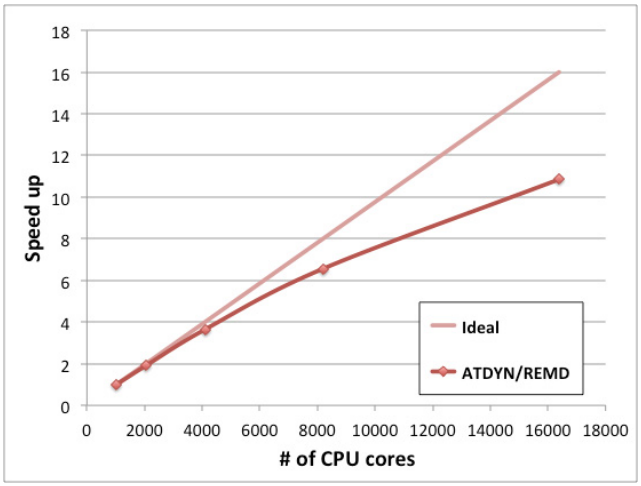


In this figure, each pair of particles separated by a distance less than R (cutoff distance) is connected by a dashed line segment, with "x" at its center lying in the box which will compute the interaction of that pair. We apply this scheme not only non-bonded but also bonded interactions. With this scheme, we could get a good scalability for short-range interactions. As for the long range electrostatic intreactions, smooth particle mesh Ewald method is assigined, in which real part is calculated with cutoff (short-range) and reciprocal part uses 3D FFT. For the efficient parallelization of reciprocal part calculation, volume decomposition scheme with MPI_Allgather is used for FFT.

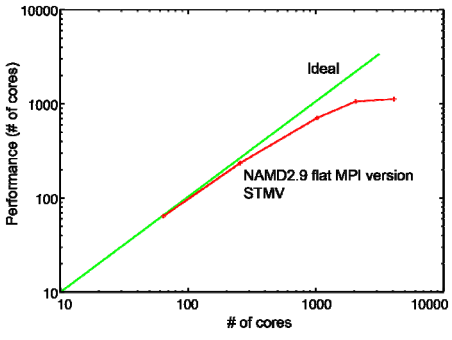


Parallel performance of virus (1,000,000 atoms) with SPDYN using K computers. Wall time is checked for 1000 steps MD calculations.

In the ATDYN program, computation of the energy and forces acting on atoms are parallelized for the number of interactions in each term of the potential energy function. We introduced not only an all-atom model but also coarse-grained model (Go model) in ATDYN. We also implemented a standard molecular dynamics method and the replica-exchange molecular dynamics method (REMD). REMD is one of the efficient sampling methods for biomolecular systems. In this method, some replicas of the system are prepared, where each replica has different parameters. During the REMD simulation, such parameters are exchanged between the neighboring replicas at certain intervals. REMD can sample structures of biomolecules in a wide range of the configurational space without getting trapped in local-minimum free-energy states. The following figure shows a parallel performance for the system containing 92,224 atoms using REMD method (128 replicas) with ATDYN/REMD on K computers. We found that the computational speed is scalable even if we use more than 10,000 CPU cores.



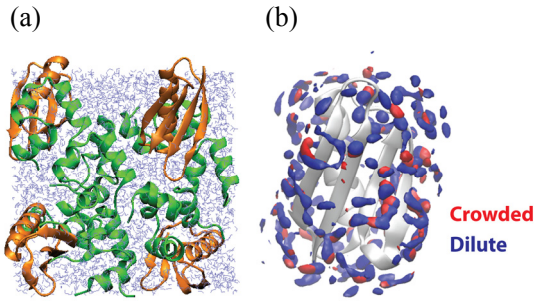
NAMD2, which have been developed by Theoretical and Computational Biophysics Group in University of Illinoi, is a highly parallelized molecular dynamics code designed for high-performance simulation of large biomolecular system, and it is widely used in biophysics field. Comparison of performance between GENESIS and NAMD2 is useful not only for the developers of GENESIS but also for many NAMD2 users in K computer. We have installed the flat MPI version of NAMD2.9 in K computer. Though it is based on Charm++ parallel objects, it has been compiled using MPI. The figure shows the parallel performance of virus, which has 1M atoms, with NAMD2.9 flat MPI version using K computer. We are now trying to compile the program using hybrid parallel scheme based on pthread/MPI library in collaboration with Dr. Kamada at



Kobe University.

7.3.2. Large-scale molecular dynamics simulations for biological systems

As future applications in terms of GENESIS with Replica-Exchange Molecular Dynamics (REMD) on K computer, we are planning to perform large-scale molecular dynamics simulations for biological systems such as crowded cellular environments. As a first step to simulate actual cellular environments, we investigated simple crowding systems consisting of protein G and villin headpiece sub-domain (Figure (a)). The results of crowding simulations showed significantly differences between crowded and dilute conditions. For instance, the distributions of water were perturbed due to crowding (Figure (b)), indicating that crowding effects hydration under crowded cellular environments.



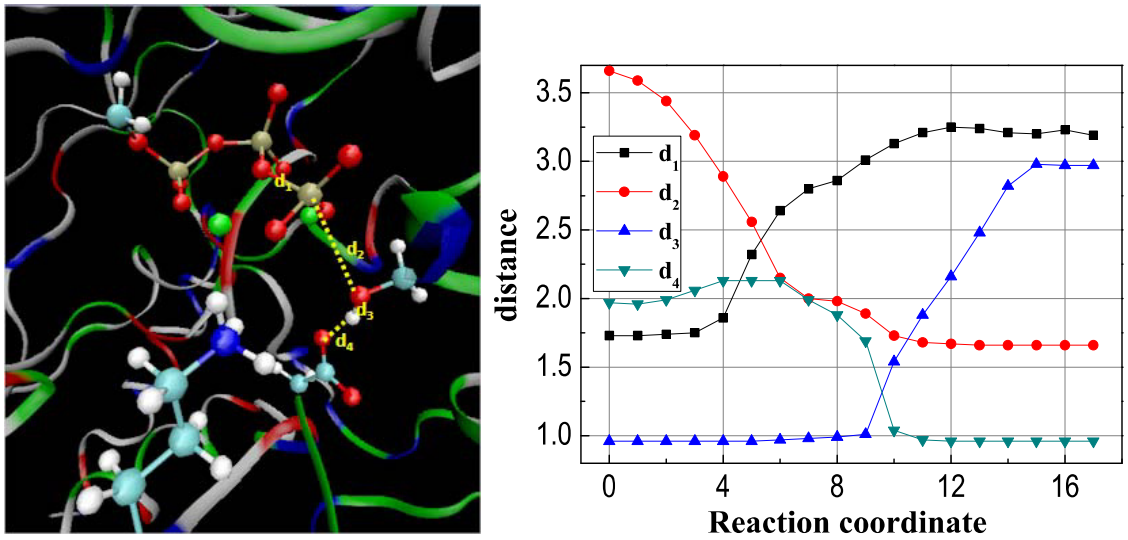
(a) A protein crowding system consisting of two proteins, protein G (brown) and villin headpiece sub-domain (green).

(b) Perturbations of three dimensional distributions of water due to crowding around protein G (white). Each color corresponds to the distributions of water under crowded (red) and dilute (blue) conditions.

7.3.3. Development of the replica-path optimization

The nudged elastic band (NEB) and string methods are widely used to obtain the minimum-energy path of chemical reactions and phase transitions. In these methods, however, it is difficult to define an accurate Lagrangian to generate the conservative forces, resulting in slow convergence. On the other hand, the constrained optimization with locally updated planes scheme (CO-LUP) defines the target function properly, although the method does have problems of inaccurate estimation of reactions and inappropriate accumulation of images around the energy minimum. We introduce three modifications into CO-LUP to overcome these problems: (1) An improved tangent estimation of the reaction path, which is used in the NEB method, (2) Redistribution of images using an energy-weighted interpolation before updating local tangents, and (3) Reduction of the number of constraints, in particular translation/rotation constraints, for improved convergence. The present method benefits from a micro-iteration scheme for protein environments in QM/MM optimization.

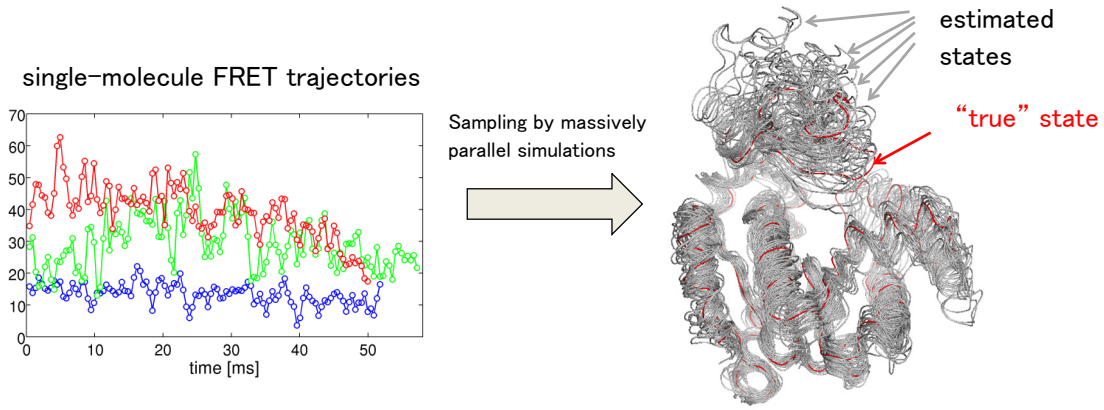
We test the method on the isomerization of alanine dipeptide. We also apply the method for defining the reaction paths of the rearrangement reaction catalyzed by chorismate mutase (CM), and of the phosphoryl transfer reaction catalyzed by cAMP-dependent protein kinase. In both cases, the results are consistent with previous QM/MM calculations.



Changes of key structural parameters of PKA during the reaction process.

7.3.4. Data assimilation algorithm

Förster resonance energy transfer (FRET) is a powerful method for determining various mechanistic properties of biomolecules. However, except for the simple cases of well-separated, low-noise, and limited to only a few states, it is usually difficult to determine the conformational states and the transition rates between them from FRET trajectories. The general state space hidden Markov models have opened a possibility to analyze such a complex time series. With the help of the recent advances in computer power and sampling techniques, it has become possible to determine the hidden states of the high-dimensional models even from low-dimensional information (Figure (a)). As future applications with GENESIS and K computer for FRET trajectories, we are developing a sampling scheme utilizing the sequential Monte Carlo filtering (particle filtering (Figure (b))). As a first demonstration of the algorithm, we have applied the method to the high-resolution single-molecule FRET trajectories (J.A. Hanson *et al.*, *PNAS* 2007) of the domain movements in adenylate kinase.



A schematic picture of the sequential Monte Carlo estimation of the high-dimensional states (i.e., protein conformations in this case) from one-dimensional FRET trajectories. The protein conformations represented by gray lines indicate the estimated conformations, and the red line corresponds to the “true” conformation.

7.4. Schedule and Future Plan

The major goal of research in next financial year (2012) is to finish the parallelization of our MD code and perform large-scale MD simulations of biomolecules in cellular environment using K computer. Before the large-scale applications, we need to perform test calculations on several different molecules under various conditions. By the end of FY2013, we are planning to open the source code of our MD program for academic researchers as well as industrial users under the license of GPL. We also continue to develop the MD code by introducing enhanced conformational sampling techniques like the generalized-ensemble method and path-optimization method. These methods are in particular useful for the free-energy calculations of biomolecules.

We are also planning to develop QM/MM molecular dynamics module in our code in collaboration with Dr. Nakajima’s team at RIKEN AICS. By combining with parallelized QM code developed by Dr. Nakajima’s group, we can perform highly parallelized QM/MM molecular dynamics simulations or QM/MM free-energy calculations of biomolecules or other molecular systems.

7.5. Publication, Presentation and Deliverables

(1) Journal Papers

1. R. Harada, Y. Sugita, and M. Feig, “Protein crowding affects hydration structure and dynamics”, Journal of American Chemical Society, 134, 482-4849 (2012).
2. Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, and A. Kidera, “Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase”, PLoS Computational Biology 8, e1002555 (2012).

(2) Conference Papers

- None

(3) Invited Talks (From April 2011 to March 2012)

1. R. Harada, Y. Sugita and M. Feig, “Protein crowding affects hydration structure and dynamics”, The 4th Applied Systems Biology Workshop at RIKEN AICS, November 7, 2011.
2. Y. Matsunaga, “Basics and advances of molecular dynamics simulation of biomolecules”, Bio super-computing summer school 2011 in Awaji, September 26-27, 2011.
3. Y. Matsunaga, “Sampling of conformational transitions in proteins: string method and data assimilation”, IMS workshop at The Institute of Statistical Mathematics, March 7, 2012.

(4) Posters and presentations

1. J. Jung, S. Re, Y. Sugita, and S. Ten-no, “New implementation of the reaction path determination in the QM/MM method”, The 49th Annual meeting of the Biophysical Society of Japan in Himeji, September 16-18 (2011).
2. R. Harada, Y. Sugita and M. Feig, “Protein crowding affects hydration structure and dynamics”, The 25th annual meeting of the molecular simulation society of Japan at Tokyo Institute of Technology, December 6, 2011.
3. Y. Matsunaga, H. Fujisaki, T. Terada, and A. Kidera, “Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase”, The 25th annual meeting of the molecular simulation society of Japan at Tokyo Institute of Technology, December 6, 2011
4. Y. Matsunaga, H. Fujisaki, T. Terada, and A. Kidera, “Conformational Transition Pathways of Adenylate Kinase Explored by the String Method”, Biophysical Society 56th Annual Meeting in San Diego, USA, February 24-29, 2012.

(5) Patents and Deliverables

- None