Theoretical Molecular Science Laboratory
Chief Scientist: Yuji Sugita (D.Sci.)

(0) Research field

CPR Subcommittee**:** Chemistry

Keywords: Molecular dynamics simulation, *ab initio* quantum chemistry, multi-scale simulation, crowded cellular environment, integrative dynamic structural biology

(1) Long-term goal of laboratory and research background

Our group has developed multi-scale simulation methods and software, GENESIS, for chemical and biological problems. The multi-scale simulation methods include hybrid quantum mechanics/molecular mechanics (QM/MM), all-atom molecular dynamics (AAMD), and coarse-grained MD (CGMD) methods. Our software GENESIS also allows extensive conformational sampling of biomacromolecules using advanced methods, which is especially useful for investigating slow conformational changes of biomacromolecules using Fugaku supercomputer. We are also developing data-driven simulations and machine learning methods for understanding structure-dynamics-function relationships of molecular systems in various environments, such as intracellular environments of the cells.

(2) Current research activities (FY2021) and plan (until Mar. 2025)

(A) Development of multi-scale MD simulation software GENESIS

In FY2021, we mainly developed (1) a new integration scheme enabling a large time step, and (2) efficient parallelization of CG MD simulations. For a stable MD with a large time step, we developed a hydrogen mass repartitioning (HMR) scheme, where the mass of hydrogen atoms is increased from twice to three times to reduce high-frequency motion while the mass of heavy atom boned to hydrogens is reduced to conserve the total mass. Existing HMR schemes have been stable up to 4 fs time step, whereas we could extend it up to 5 fs time step with a modified ratio of hydrogen mass scaling. By combining this scheme with accurate temperature and pressure evaluations, we could obtain the same thermodynamic and kinetic properties as 2 fs time step case. For CGMD simulations, we developed a new domain decomposition scheme, which we call cell-based *kd*-tree scheme. The system is divided into two subsystems with similar number of particles. This division is iterated until the number of subsystems is equal to the number of MPI processes. Our preliminary benchmark results of a chromatin consisting of 760,000 CG particles show that this scheme can increase the performance about three times better than the existing one when proper load balance is kept.

(B) Hybrid QM/MM calculations

The QM/MM method is a computational method for determining the interaction between atoms and molecules by treating the spatial region of interest with highly accurate quantum chemical (QM) calculations and the surrounding environment with a molecular mechanics (MM) force field. The method compliments the drawback of the MM force field, and has made various calculations of biomolecules feasible, such as chemical reactions, spectroscopic properties, etc. However, despite restricting the QM region to a small space, the computational cost of QM calculations is extremely large, and most conventional QM/MM calculations have been limited to static analyses. We have implemented the QM/MM method into GENESIS and interfaced it with QSimulate, a recently developed highly parallelized QM program. The combination of GENESIS and QSimulate together with a massively parallel computer has achieved a speedup of more than 10 times faster than the conventional method. GENESIS/QSimulate enables QM/MM calculations that consider dynamical effects and has been applied to the calculation of reaction free energies for enzyme reaction and vibrational spectra of membrane proteins.

(C) Data-driven approaches

Recently, single particle analysis using cryo-EM has been widely used in structural analysis of protein. The method reconstructs a three-dimensional (3D) density map from many two-dimensional (2D) images of the target protein taken by electron microscopy. To determine the molecular structure from the density map, careful and accurate structural modeling based on computational science is necessary. Especially, *de novo* modeling is necessary when there is no known structure to refer to, and the 3D structure should be predicted directly from the density map based on the amino acid sequence. In addition, flexible fitting based on an MD algorithm is widely used as a method to optimize the

predicted structure, where MD simulation is performed with a biasing potential for fitting the atoms to the density map. In collaboration with Prof. Daisuke Kihara group at Purdue University, we proposed the SAUA-FFR method as a protocol for efficiently optimizing candidate structures of Cα models obtained by *de novo* modeling using flexible fitting. In this method, the structure generated from the Cα model is optimized by repeating the simulated annealing method several times using the united-atom model. Furthermore, we proposed a protocol for selecting the best model from many candidate structures, which enables the efficient selection of a model with less overfitting. When the SAUA-FFR method was used to optimize the candidate structures obtained by MAINMAST, it was found that the SAUA-FFR method promotes the formation of secondary structures and gets closer to the native structure than the method using the all-atom model (SAAA-FFR), indicating that our method enables more reliable structural modeling than conventional methods.

(D) Enhanced conformational sampling of spike protein dynamics

Spike protein on the surface of SARS-CoV-2 is an essential protein for the virus entry. We carried out atomistic MD simulations of the spike protein using Fugaku to investigate the role of N-glycans for protein conformational stability and the mechanisms for conformational changes from the inactive Down to the active Up forms. First, we carried out conventional MD simulations of spike protein starting from the Down and Up forms for 1μs each and found that three glycans have essential roles for the stabilization of spike protein conformations. Importantly, the Down and Up structures are stabilized with different glycans. The transition from Down to Up forms takes longer time than that is accessible with the conventional MD simulations. Therefore, we applied the gREST method, which is one of the generalized-ensemble algorithms for enhancing conformational sampling of biomacromolecules. In the gREST simulations of spike proteins, we can cover not only the Down and Up structures but also the intermediate structures between them for elucidating molecular mechanisms for the conformational transitions. The intermediate structures show different intra-molecular interactions and accessible surface areas, which might be useful for the design of new antibodies or anti-virus drugs.

Future Plan

We will continue the developments of GENESIS multi-scale MD simulation program, which includes the simulations using QM/MM, atomistic, and coarse-grained models of various molecules. In addition, we need to connect MD simulations with two different resolutions, for instance, QM/MM and atomistic simulations, or atomistic and coarse-grained ones. For this purpose, machine learning or other computational approaches would be useful.

We have great interest on the formations and collapses of condensates (or droplets), which are formed with multi-domain proteins, intrinsically disordered proteins/regions, and RNAs in the cell. The droplets formed by the phase separation are known to play important roles in various biological functions. In the simulations of condensates (or droplets) in the cell, integrated multi-scale MD simulations are, in particular, useful for understanding the detailed molecular mechanisms.

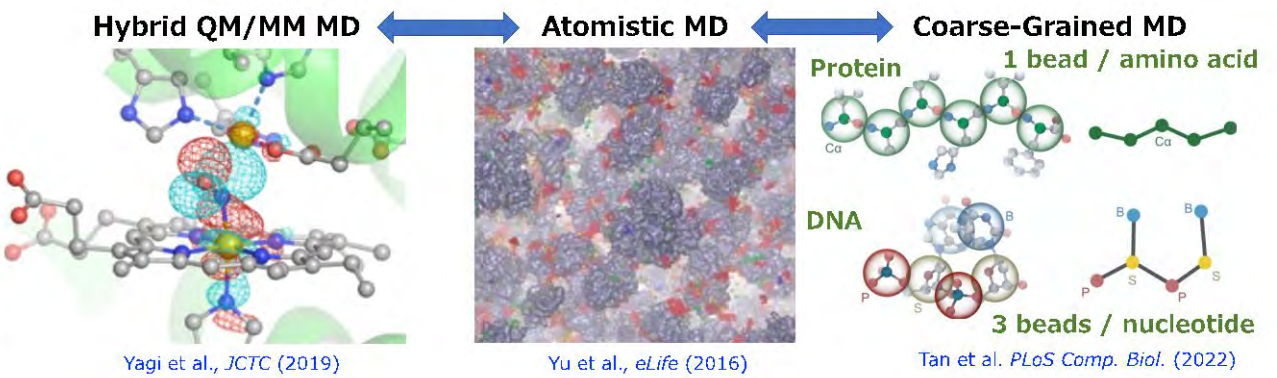

Figure 1. Multi-scale models available in GENESIS simulation software.

(3) Members **(FY2021)**

(Chief Scientist)
Yuji Sugita
(Senior Research Scientist)
Kiyoshi Yagi, Takaharu Mori
(Senior Technical Scientist)
Jaewoon Jung
(Special Postdoctoral Researcher)
Ai Niitsu, Mao Oide, Yosuke Sumiya
(Postdoctral Researcher)
Hisham M. Dokainish, Weitong Ren, Shingo Ito, Yaokun Lei, Haeri Im
(Trainee)
Masahiro Motohashi
(Assistant)
Machiko Ishigaki, Hiromi Kano

(4) Representative research achievements

1. "The inherent flexibility of receptor binding domains in SARS-CoV-2 spike proteins", H.M. Dokainish, S. Re, T. Mori, C. Kobayashi, J. Jung, Y. Sugita, *eLife* 11 (2022), e75720.
2. "Towards complete assignment of the infrared spectrum of the protonated water cluster $H^+(H_2O)_{21}$", J. Liu, J. Yang, X.C. Zeng, S.S. Xantheas, K. Yagi, X. He, *Nat. Commun.* 12 (2021) 6141.
3. "Optimized Hydrogen Mass Repartitioning Scheme Combined with Accurate Temperature/Pressure Evaluations for Thermodynamic and Kinetic Properties of Biological Systems", J. Jung, K. Kasahara, C. Kobayashi, H. Oshima, T. Mori, Y. Sugita, *J. Chem. Theory Comput.* 17 (2021) 5312-5321.
4. "Efficient Flexible Fitting Refinement with Automatic Error Fixing for De Novo Structure Modeling from Cryo-EM Density Maps", T. Mori, G. Terashi, D. Matsuoka, D. Kihara, and Y. Sugita, *J. Chem. Inf. Model.* 61 (2021) 3516-3528.
5. "Exploring the Minimum-Energy Pathways and Free-Energy Profiles of Enzymatic Reactions with QM/MM Calculations.", K. Yagi, S. Ito, Y. Sugita, *J. Phys. Chem. B* 125 (2021) 4701-4713.

Supplementary
Group Photo (taken in 2019)

Laboratory Homepage
https://www.riken.jp/en/research/labs/chief/theor_mol_sci/index.html
https://tms.riken.jp/en/