Usage Report for Fiscal Year 2023

Project Title:

Development of multi-scale simulation methods for large biomolecular systems

Name:

○Cheng Tan (1), Diego Ugarte La Torre (1), Jaewoon Jung (1,2), and Chigusa Kobayashi (1)

Laboratory at RIKEN:

(1) RIKEN Center for Computational Science, Computational Biophysics Research Team

(2) RIKEN Cluster for Pioneering Research, Theoretical Molecular Science Laboratory

1. Background and purpose of the project, relationship of the project with other projects

Molecular dynamics (MD) simulation is a useful tool for delving into biological processes, utilizing meticulously designed computational models and algorithms. In our current project (Q23536), we have advanced the MD methods within the GENESIS software, enabling the detailed modeling of large-scale biomolecular systems. This includes the integration of state-of-the-art models across various resolutions and optimizing the software for peak performance across diverse computing platforms.

The advent of cryo-electron microscopy (cryo-EM) experimental studies has uncovered a wealth of structures of large proteins in myriad physiological conditions. Coupled with high-speed algorithms on cutting-edge hardware, the scope of simulations has broadened dramatically, both in terms of time and spatial scales. However, the analysis of such extensive datasets remains a formidable challenge. To address this, we have developed a method that leverages the geometric mean of protein domains, allowing for the extraction of significant structural changes within these complex datasets.

Our efforts in refining simulation and analysis methodologies have led us to explore several key areas: such as the liquid-liquid phase separation (LLPS) of biomolecules and the conformational dynamics of spike proteins in coronaviruses. These applications underscore the versatility and depth of our developments in the field of molecular dynamics.

Specifically, we have conducted the following studies:

1) Improvement of the coarse-grained lipid model, iSoLF, by increasing the number of available lipids and including electrostatic interactions (by Ugarte)

2) Development of efficient parallelization for coarse-grained MD simulation (by Jung)

3) Development of coarse-grained method to express domain motions of multi-domain/multi-chain proteins (by Kobayashi)

4) Application of multi-scale MD simulations to study the phase behavior of biomolecules (by Tan)

2. Specific usage status of the system and calculation method

1) To improve the implicit solvent lipid force field, iSoLF, we used a previously generated database of different lipid properties as target values for fitting parameters. We used the implementation of iSoLF in GENESIS and performed coarse-grained simulations of 30 different lipids at different temperatures. Then, from each simulation, we calculated a score value that was used to assess the quality of the parameters. Finally, we

repeated this procedure by gradually adjusting the parameters in the force field until it was possible to reproduce the area per lipid and the membrane thickness.

2) To improve the performance of residue-level coarse-grained (CG) MD simulations of large biomolecules, we have developed a domain decomposition scheme with dynamic load-balancing. The simulation system is divided into subdomains, called cells. Unlike the conventional all-atom MD, different number of cells are assigned to different processes such that all process has the similar number of particles. During MD simulations, the particle densities can change, and we adjust the domain size accordingly. The developed program has been implemented into the CGDYN simulator of GENESIS MD software.

3) Henderson *et al.* (*Nat. Str. Mol. Biol.* 2020, *Science* 2021) proposed a method to represent a multi-domain/multi-chain protein S-protein in CG particles. These particles are calculated from the Cα centroids of given residues. We have modified it so that it can be applied in the trimer form of S-protein. We show that it describes the domain motions of S-protein well. In this year, we develop a method to set up CG particles of arbitrary proteins by using MD trajectories.

4) In our investigation of large-scale droplet fusion in TDP-43, we employed the HPS model within GENESIS. This study made use of our recent innovation, GENESIS CGDYN, an MD engine characterized by dynamic load balancing and a non-uniform domain decomposition scheme. We constructed droplet structures comprising TDP-43 and assembled expansive systems that included hundreds of droplets of varying sizes. Subsequently, we simulated the fusion behaviors of these droplet systems across different densities

3. Result

1) We improved the iSoLF force field by increasing the number of available lipids from 2 to 30 (table 1).

TARGET LIPIDS		LIPID HEAD				
		PA	PC	PE	PG	PS
LIPID TAIL	DL	DLPA	DLPC	DLPE	DLPG	DLPS
	DM	DMPA	DMPC	DMPE	DMPG	DMPS
	DO	DOPA	DOPC	DOPE	DOPG	DOPS
	DP	DPPA	DPPC	DPPE	DPPG	DPPS
	PO	POPA	POPC	POPE	POPG	POPS
	SO	SOPA	SOPC	SOPE	SOPG	SOPS

Table 1. Target lipids. By combining both head and tail beads, 30 different lipids become available in the improved model iSoLFv2.

Additionally, the new parameters in the iSoLF coarse-grained (CG) model reproduce different properties calculated from all-atom (AA) simulations at different temperatures (Figure 1).

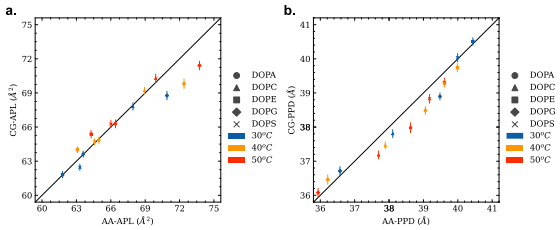

Figure 1. Property comparison between AA and CG lipid models. (a) and (b) show the area per lipid and membrane thickness for DO lipids at three different temperatures.

2) We performed benchmark tests on (a) 120 DPS proteins (222,360 particles), (b) 5000 chains of 100 amino acid IDPs (500,000 particles), (c) 512 nucleosomes (1,044,480 particles), and (d) multiple DNA systems. CGDYN, which we developed in this fiscal year, shows better performance than ATDYN, our previously implemented program in GENESIS. CGDYN also shows much improved performance than Open3SPN2, which is a CG program implemented on OpenMM for GPU use (Figure

2).

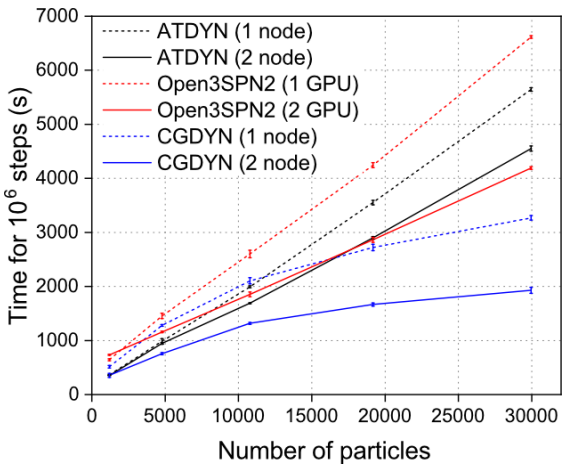

Figure 2. Comparison of the performance of multiple DNA chains on Hokusai.

3) We categorized each part of the protein into three types: domains, hinges, and surfaces. As like previous studies, domains can be calculated by Ca centroid of given residues. A hinge is placed between adjacent domains. A surface indicates the direction of the domain motion. We propose a method to find optimal surfaces using MD trajectories. We apply the method to three proteins with multi-domain and show that the domain motion of the protein can be well represented.

4) We have implemented many residue-level coarse-grained (CG) models within the molecular dynamics (MD) software GENESIS, including the popular models for intrinsically disordered proteins (IDPs). Leveraging our implementation, we explore the anti-aggregation properties of Hero11, a heat-resistant obscure protein, and its interactions with the client protein TDP-43. Our research uncovers how Hero11 regulates the phase separation of TDP-43 in both dense and dilute environments, highlighting the potential influence of Hero11's surface distribution within condensates. Additionally, by employing a hierarchical domain decomposition approach and dynamic load balancing, we can investigate the fusion behaviors of near-realistic sized droplets of these proteins (Figure 3). Our advancements in MD simulation provide a new framework for examining various biomolecular phase separation phenomena.

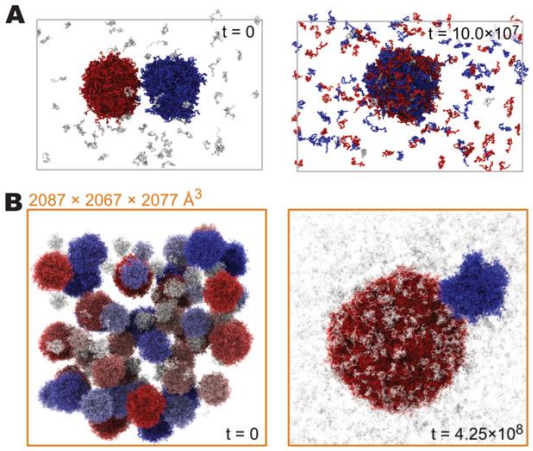

Figure 3. Large-scale CG MD simulations of the fusion behaviors of two droplets (A) and multiple droplets (B). All the systems are consisting of TDP-43 proteins.

4. Conclusion

In our work with Hokusai, we've made significant advancements in developing multi-scale MD simulation methods within the GENESIS software. Our progress includes:

1) Improving the iSoLF CG lipid model, making it a better tool for the multiscale simulation of biological membrane systems.
2) Efficient parallelization of residue-level MD simulation for understanding mesoscopic biological phenomena such as droplet fusions.
3) We develop a method for selecting CG particles that can represent domain motion. This method not only reduces the amount of data, but also enables the extraction of information of domain motion.
4) Conducting simulations of biomolecular condensates using the residue-level CG models for proteins. These simulations have unveiled potential mechanisms determining the fusion of biomolecular condensations

formed via LLPS.

Our developments and application studies represent a significant contribution to the computational biophysical community, offering a robust framework for MD simulation and analysis.

5. Schedule and prospect for the future

Looking ahead, we aim to enhance our multi-scale implementation, focusing on achieving even greater performance efficiency. Additionally, we plan to extend our coarse-grained analysis method to encompass a broader range of multi-domain proteins and their interactions with lipid membranes. Following these developments, we will also conduct a comprehensive examination of the methods applied to the proteins analyzed thus far.

6. If no job was executed, specify the reason.

Usage Report for Fiscal Year 2023

Fiscal Year 2023 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1. Cheng Tan, Ai Niitsu, and Yuji Sugita. "Highly charged proteins and their repulsive interactions antagonize biomolecular condensation." *JACS Au* 3, 3 (2023): 834-848.
2. Azuki Mizutani, Cheng Tan, Yuji Sugita, and Shoji Takada. "Micelle-like clusters in phase-separated Nanog condensates: A molecular simulation study." *PLoS Computational Biology* 19, 7 (2023): e1011321.
3. Jaewoon Jung, Chigusa Kobayashi, and Yuji Sugita. "Acceleration of generalized replica exchange with solute tempering simulations of large biological systems on massively parallel supercomputer." *J. Compute. Chem.* 44, 1740-1749 (2023).
4. Diego Ugarte La Torre, Shoji Takada and Yuji Sugita. "Extension of the iSoLF implicit-solvent coarse-grained model for multicomponent lipid bilayers." *J. Chem. Phys.* 159, 075101 (2023).

[Oral presentation]

1. C. Tan, A. Niitsu, J. Jung, Y. Sugita, “Regulation of Biomolecular Condensation Studied with Large-Scale Coarse-Grained Molecular Dynamics Using GENESIS”, The 61st Annual Meeting of Biophysical Society of Japan, Nagoya, 2023/11/15.
2. C. Kobayashi, K. Inaba, Y. Sugita, “Molecular dynamics (MD) simulations of structural changes in the E1P-E2P transition of SR-Ca^{2+}-ATPase.” The 61st Annual Meeting of Biophysical Society of Japan, Nagoya, 2023/11/15.

[Poster presentation]

1. C. Tan, Y. Sugita, “Development of Coarse-Grained Models in Molecular Dynamics Software GENESIS and Applications in Biomolecular Condensation Simulations”, BioNano8, Kobe, 2023/06/14.
2. C. Tan, Y. Sugita, “Highly Charged Proteins and Their Repulsive Interactions Antagonize Biomolecular Condensation”, The 23rd Annual Meeting of Protein Science Society of Japan, Nagoya, 2023/07/06.
3. C. Tan, Y. Sugita, “Development of Coarse-Grained Models in Molecular Dynamics Software GENESIS and Applications in Biomolecular Condensation Simulations”, CCP2023, Kobe, 2023/08/07.
4. C. Kobayashi, H. M. Dokainish, S. Re, T. Mori, J. Jung, Y. Sugita, “Analysis of structural changes of multi-chain/multi-domain proteins by coarse-grained description.”, CCP2023, Kobe, 2023/08/07.