Usage Report for Fiscal Year 2022

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 powerful tool for exploring biological processes with carefully conceived computational models and algorithms. In the current project, we develop MD methods in the GENESIS software for modeling large-scale biomolecular systems at the molecular level. We implement the latest models at different resolutions and optimize the code for high performance on various platforms.

Recently, cryo-EM experimental studies have revealed many structures of large proteins under various physiological conditions. In addition, high-speed algorithms on new machines expand the range of simulation in both time and space. However, analyzing such extensive data is still challenging. Therefore, we introduce an analysis method that uses the geometric average of protein domains to extract critical structural changes.

As applications of our simulation and analysis method developments, we investigate the liquid-liquid phase separation (LLPS) of biomolecules and conformational dynamics of the spike proteins of coronaviruses.

Specifically, we have conducted the following studies:

1) Development of an implicit solvent coarse-grained (CG) lipid model in GENESIS (by Ugarte)
2) Development of an efficient scheme of multi-copy enhanced sampling simulations (by Jung)
3) Development of a coarse-graining method for protein structural dynamics analysis (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) The implicit solvent lipid force field, iSoLF, is a model for performing CG MD simulations of lipidic systems. We have further improved and extended this model by parameterizing 35 different lipids using a new CG mapping by performing all-atom (AA) MD simulations using GENESIS with the CHARMM36 force field and applying the multi-state Boltzmann Inversion method and fitting physical properties with the resulting AA trajectories.

2) We developed an efficient scheme of multi-copy enhanced sampling with improved performance. Among various multi-copy sampling schemes, we improved generalized replica exchange with solute tempering (gREST) by on-the-fly energy analysis. In gREST, free energy can be evaluated by the multi-state bennet acceptance ratio (MBAR). We must evaluate energy values of all temperature

parameters from trajectories generated by gREST MD simulations. The overall procedure of gREST simulations can be described by the following five steps:

(1) gREST MD simulations
(2) gREST energy calculations with all temperature parameters
(3) gREST energy rearrangements according to temperature parameters
(4) free energy estimation
(5) free energy calculation according to collective variable

Usually, we consider multiple processes for step (1) and a single process for steps (2)-(5). Unfortunately, step (2) requires a large amount of calculation, and calculation is not even possible for a very large system. To overcome this, we have implemented an on-the-fly energy analysis scheme where steps (1) and (2) are combined into one (Figure 1). Using this way, we can significantly increase performance.

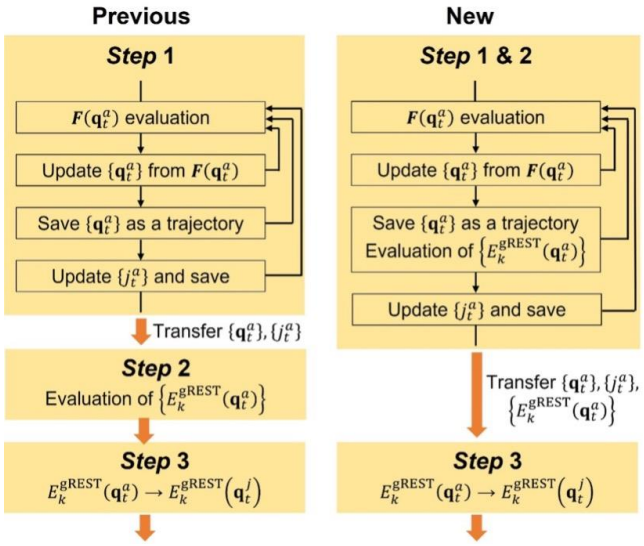

Figure 1. gREST free energy calculation flowchart using conventional (left) and our new scheme (right).

3) Henderson *et al.* (*Nat. Str. Mol. Biol.* 2020, *Science* 2021) compared the structures of spike proteins (S-protein) in different species of coronaviruses (COV). They represented the monomers as nine CG particles from the Cα centroids of given residues. We have modified the method to apply to the trimers based on the selection. We have also developed and released the program as an analysis tool in GENESIS.

4) We have implemented the residue-level CG models for protein and nucleic acids in GENESIS. We also introduced a CG model (named HPS) for the intrinsically disordered protein (IDP) and tested several different parameter-sets. We then utilized our CG development in GENESIS to study the regulation of IDPs' LLPS by a class of heat-resistant obscure proteins (Hero proteins, Tsuboyama *et al. PLoS Biol.* 2020).

3. Result

1) We extended the iSoLF model to 35 lipids by the parameterization of tail and head groups that act as building blocks for representing lipidic molecules. The resulting new model qualitatively matches the area per lipid and membrane thickness obtained from AA simulations (Figure 2). We found that lipids that naturally occurred in the liquid phase are better represented by our CG model. Furthermore, when performing simulations of multi-component lipid bilayers with different compositions of DOPC and DPPC, we could observe the phase coexistence of gel and liquid-disordered phases, represented by the location of the peaks of the distributions for the order parameter (Figure 3).

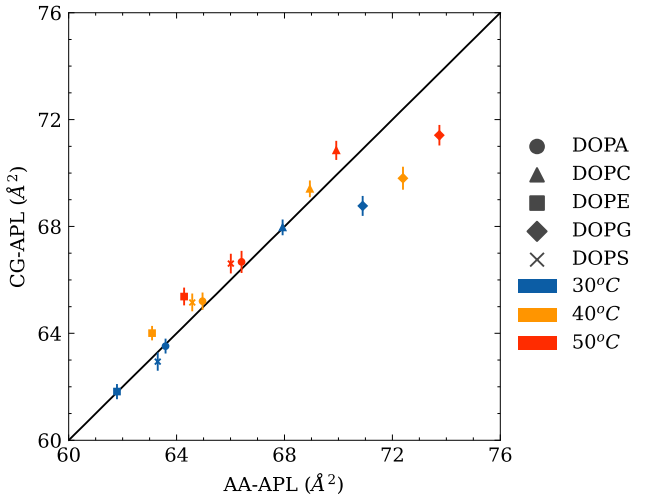

Figure 2. Area per lipid. Representative plot for the comparison of the area per lipids obtained at different temperatures with our new iSoLF model and AA simulations using the CHARMM36 force field for different DO lipids.

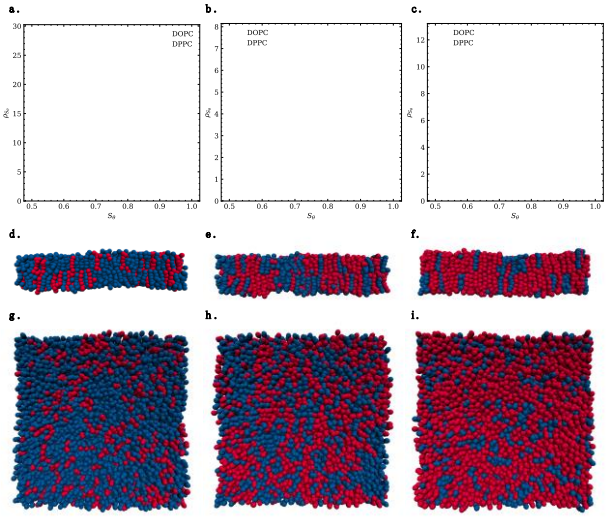

Figure 3. Phase coexistence. Distribution of the order parameter for multi-component lipid bilayers composed of DOPC (blue) and DPPC (red) lipids. By gradually increasing the relative concentration of DPPC, a gel phase is observed, presented by the appearance of a peak near 0.95. (a-c) Distributions with 25%:75%, 50%:50%, and 75%:25% of DOPC and DPPC, respectively. (d-f) Side view of the bilayers. (g-i) Top view of the bilayers.

2) We compared our new scheme of multi-copy enhanced sampling simulation with the previous one and found two-order performance improvements on Hokusai and Fugaku. We also found that our new scheme generates the same free energy results as the conventional scheme (Figure 4).

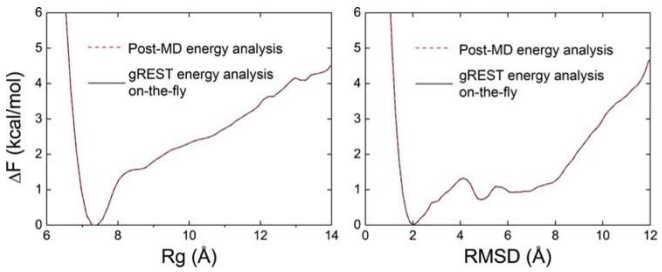

Figure 4. Free energy landscape of trp-cage in terms of the radius of gyration (left) and root mean square deviation (right) from gREST MD simulations.

3) We applied our CG analysis method using the original selection by Henderson to Cryo-EM structures of spike proteins on the surface of COVID-19. By using this method, we can clearly show the up/down of the three RBDs from experimental structures with missing residues. To find a good choice of domains, we also increased the CG particles and analyzed the Cryo-EM structures and simulation trajectories.

4) We implemented the latest CG models for the study of LLPS in GENESIS. We then applied our MD simulations to study the phase behavior of Hero11, a Hero protein, and the C-terminal low-complexity domain (LCD) of the transactive response DNA-binding protein 43 (TDP-43), a client protein of Hero11, under various conditions to examine their interactions with each other. Based on our simulation results, we have proposed three possible mechanisms for Hero11's regulatory functions, based on the changes of TDP-43-LCD induced by Hero11, in the dense and dilute phases, and on the surface of condensates (Figure 5).

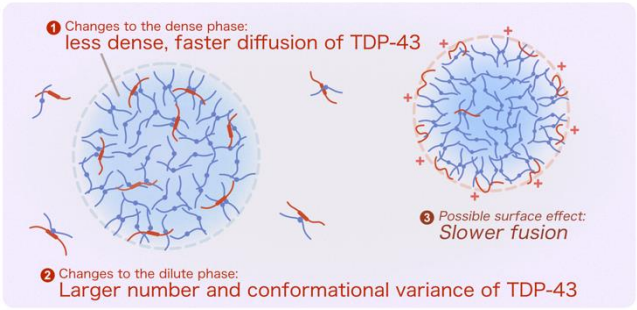

Figure 5. Three possible mechanisms of Hero11's anti-aggregation function. 1) Hero11 joins the condensate of TDP-43-LCD as a ligand and decreases the intermolecular contacts. Consequently, the diffusion of TDP-43-LCD is promoted. 2) Hero11 raises the probability for TDP-43-LCD to be in the dilute phase and induces larger variance in conformation of TDP-43-LCD. 3) In the case of less helical structures, Hero11 tends to stay at the surface of the droplets, which may have some surface effects such as preventing droplet fusion.

4. Conclusion

We have utilized Hokusai in developing multi-scale MD simulation methods in the MD software GENESIS. 1) We have designed a new scheme of multi-copy enhanced sampling and improved the performance. 2) We have developed a new CG model of lipids in GENESIS and improved the existing parameter set. 3) We have introduced an analysis method that uses the geometric average of protein domains to extract critical structural changes. We applied the technique to Cryo-EM structures of spike proteins on the surface of COVID-19. The method shows the up/down of the three RBDs from experimental structures. 4) We have simulated the biomolecular condensates by developing the residue-level CG models for protein and nucleic acids. Our simulation results revealed possible mechanisms for the regulation of LLPS. Our developments and application studies will contribute to the computational biophysical society with a robust MD simulation and analysis framework.

5. Schedule and prospect for the future

In the future, we will further optimize our multi-scale implementation to gain better performance. We will also continue developing the CG analysis method for other multi-domain/multi-chain proteins. Then, we will examine the methods for proteins calculated so far.

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

Usage Report for Fiscal Year 2022

Fiscal Year 2022 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1) H. M. Dokainish, S. Re, T. Mori, C. Kobayashi, J. Jung, Y. Sugita, "The inherent flexibility of receptor binding domains in SARS-Cov-2 spike protein", *Elife*, 11, e75720 (2022)
2) C. Tan, J. Jung, C. Kobayashi, D. Ugarte, S. Takada, Y. Sugita, "Implementation of residue-level coarse-grained models in GENESIS for large-scale molecular dynamics simulations", *PLoS Comput. Biol.*, 18, e1009578 (2022)
3) Y. Matsunaga, M. Kamiya, H. Oshima, J. Jung, S. Ito, Y. Sugita, "Use of multistate Bennett acceptance ratio method for free-energy calculations from enhanced sampling and free-energy perturbation", *Biophys. Rev.*, 14, 1503-1512 (2022)
4) Y. Zhang, C. Kobayashi, X. Cai, S. Watanabe, A. Tsutsumi, M. Kikkawa, Y. Sugita, and K. Inaba, "Multiple sub-state structures of SERCA2b reveal conformational overlap at transition steps during the catalytic cycle", *Cell Reports*, 41, 111760 (2022)
5) C. Tan, A. Niitsu, Y. Sugita, "Repulsive interaction and secondary structure of highly charged proteins in regulating biomolecular condensation", *JACS Au*, (2023) (in press).

[Oral presentation]

1) C. Tan, A. Niitsu, Y. Sugita, "Multi-scale molecular dynamics study of repulsive interactions in regulating biomolecular condensation", The 32nd Tokyo RNA Club, RIKEN Symposium, Tokyo, 2022/11/12.
2) C. Tan, J. Jung, C. Kobayashi, D. Ugarte, S. Takada, Y. Sugita, "Implementation of residue level coarse-grained models in GENESIS for large-scale molecular dynamics simulations", 第 36 回分子シミュレーション討論会, Tokyo, 2022/12/07.

[Poster presentation]

1) Diego Ugarte La Torre, Shoji Takada, Yuji Sugita, "Extension of the Implicit Solvent Lipid Force Field, iSoLF, for the simulation of large multi-component lipidic systems using GENESIS", The 60th Annual Meeting of the Biophysical Society of Japan, Hokodate, Sep. 28-30 (2022)
2) Diego Ugarte La Torre, Shoji Takada, Yuji Sugita, "Molecular dynamics simulation of large multicomponent lipid bilayers in GENESIS using the iSoLF coarse-grained model", The 5th R-CCS international symposium, Kobe, Feb. 6. (2023)
3) J. Jung, C. Tan, C. Kobayashi, D. Ugarte, Y. Sugita, "Acceleration of residue-level coarse-grained molecular dynamics by efficient parallelization", 67th Biophysical Society Annual Meeting, San Diago, Feb. 18-22 (2023)
4) J. Jung, C. Kobayashi, S. Sugita, "Acceleration of gREST simulations of large biological systems on massively parallel computers", 第 36 回分子シミュレーション討論会, Tokyo, Dec. 5-7 (2022)

5) J. Jung, C. Tan, C. Kobayashi, D. Ugarte, Y. Sugita, "Acceleration of residue-level coarse-grained molecular dynamics by new development of parallelization", The 60th Annual Meeting of the Biophysical Society of Japan, Hokodate, Sep. 28-30 (2022)
6) C. Kobayashi, H. Dokainish, S. Re, T. Mori, J. Jung, Y. Sugita, "Analysis of structural changes of multi-chain/multi-domain proteins.", The 5th R-CCS international symposium, Kobe, Feb. 6. (2023)
7) C. Kobayashi, H. Dokainish, S. Re, T. Mori, J. Jung, Y. Sugita, "A method for analyzing structural changes of protein with multi-chains/multi-domains", The 60th Annual Meeting of the Biophysical Society of Japan, Hakodate, Sep. 28. (2022)
8) C. Tan, J. Jung, C. Kobayashi, D. Ugarte, S. Takada, Y. Sugita, "Implementation of residue level coarse-grained models in GENESIS for large-scale molecular dynamics simulations", The 5th R-CCS international symposium, Kobe, Feb. 6. (2023)
9) C. Tan, A. Niitsu, Y. Sugita, "Multi-scale molecular dynamics study of repulsive interactions in regulating biomolecular condensation", The 60th Annual Meeting of the Biophysical Society of Japan, Hakodate, Sep. 28. (2022)
10) C. Tan, A. Niitsu, Y. Sugita, "Multi-scale molecular dynamics study of repulsive interactions in regulating biomolecular condensation", 67th Biophysical Society Annual Meeting, San Diago, Feb. 18-22 (2023)

[Others (Book, Press release, etc.)]

1) C. Tan, A. Niitsu, T. Yoda, Y. Sugita, 分子シミュレーションで理解する相分離のメカニズム. In: "相分離プロトコール, 発展編 5", 235-241. 羊土社 (2022).