GENESIS
Generalized-ensemble simulation system

# GENESIS Hands-on
# Part 1: GENESIS basics and GENESIS on Fugaku
# - Lecture -

**Chigusa Kobayashi**

RIKEN Center for Computational Science

2024/06/30

**IUPAB2024 Hands-on Training Program CHARMM-GUI/GENESIS MD Tutorial**

# Schedule of GENESIS parts (6/30-7/2)

| | |
|---|---|
| 06/30 Part1 | |
| 13:30 – 15:00 | **GENESIS basics and GENESIS on Fugaku (Kobayashi)** |
| | **Lecture** |
| | Hands-on tutorial on Fugaku |
| | |
| 07/01 Part 2 | |
| 14:30 – 15:30 | Coarse-grained simulations in GENESIS (Tan) |
| 15:30 – 16:30 | High-performance computation with GENESIS (Jung) |
| | |
| 07/02 Part 3 | |
| 13:30 – 15:00 | Generalized-ensemble simulations using GENESIS (Ito) |

# Contents

- Molecular Dynamics (MD)
- MD software, GENESIS
- What we can do by using GENESIS

# Molecular Dynamics (MD)

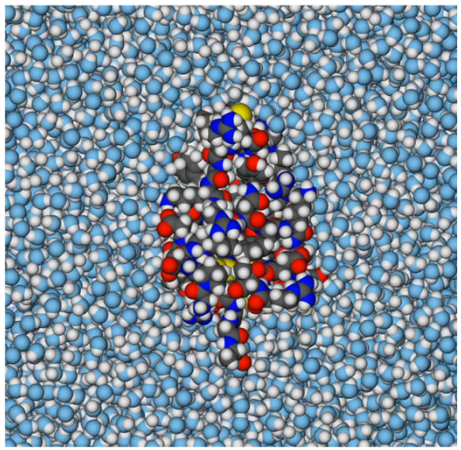
- calculates motion of particles based on the Newton's equation of motion.

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i}$$
$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i$$

→

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{p}_i}{m_i}\Delta t$$
$$\mathbf{p}_i(t+\Delta t) = \mathbf{p}_i(t) + \mathbf{F}_i\Delta t$$

→

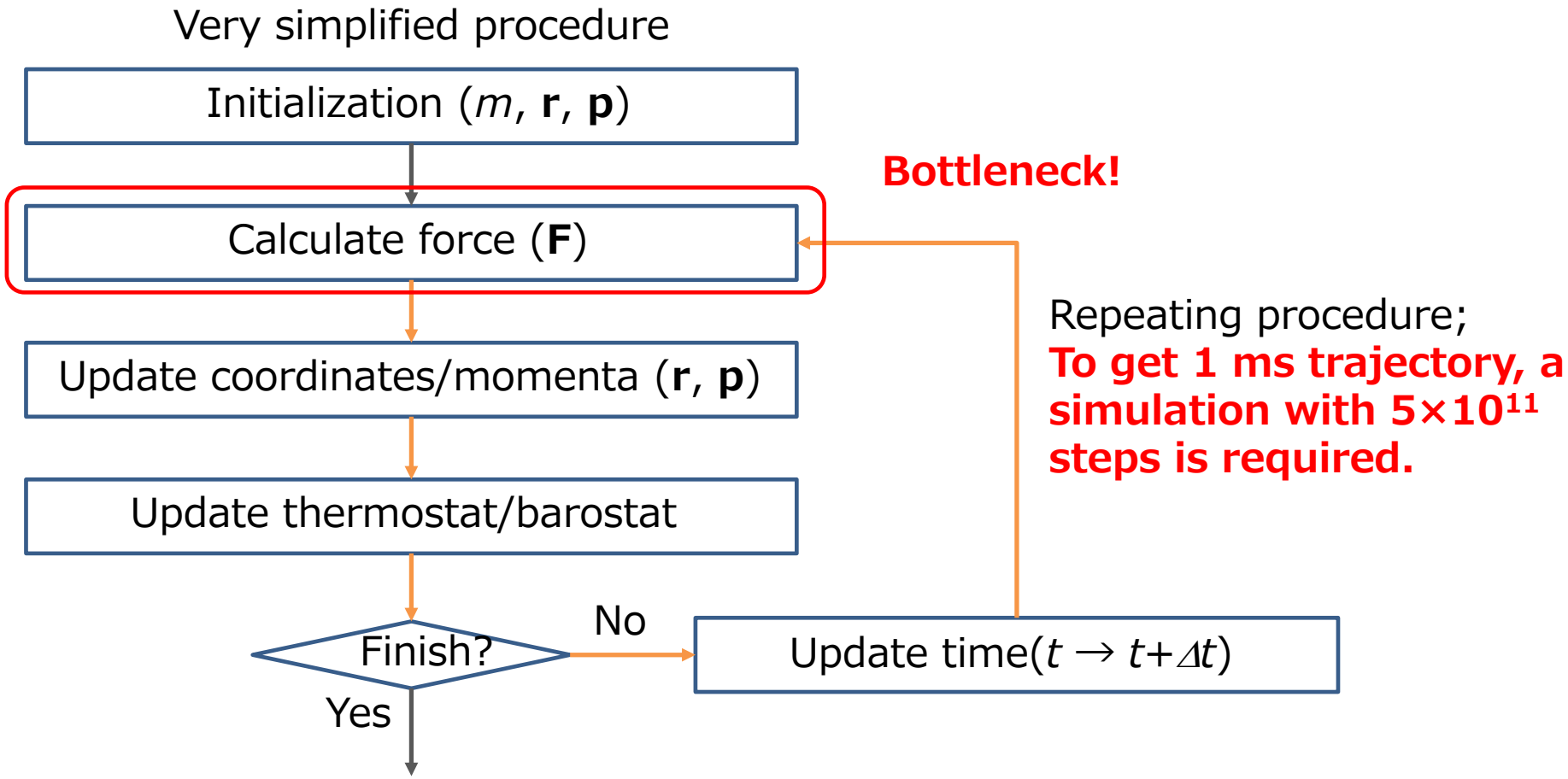


**Equation of motion** **Integration** **Motions of particles**

$\mathbf{F}_i$ : force of *i*-th particle
$\mathbf{r}_i$ : coordinate of *i*-th particle
$\mathbf{p}_i$ : momentum of *i*-th particle
$m_i$ : atomic mass of *i*-th particle
$t$ : time
$Dt$ : timestep size

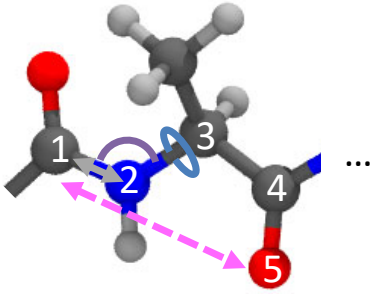
# Procedure of MD

In MD, numerical integration is repeated with a small timestep.



# Force calculation for biological systems

Force can be divided into bonding and nonbonding interactions.




...

Hydrogen: Light Gray
Carbon: Dark Gray
Oxygen: Red
Nitrogen: Blue

$$E_{\text{total}} = \sum_{\text{bonds}} k_b (b - b_0)^2$$

**Bond** (ex. 1-2, 2-3, ...)

$$+ \sum_{\text{angles}} k_a (\theta - \theta_0)^2$$

**Angle** (ex. 1-2-3, 2-3-4,..)

$$+ \sum_{\text{dihedrals}} V_n [1 + \cos(n\omega - \gamma)]$$

**Dihedral** (ex. 1-2-3-4, ...)

Bonding
O(***N***)

$$+ \sum_{i,j \notin \text{bonding}} \left\{ \varepsilon_{ij} \left[ \left( \frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\}$$

**van der Waals** (ex. 1-5, ⋯) :

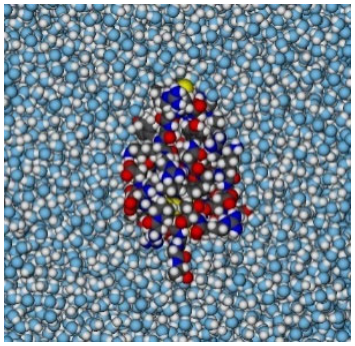
**Coulomb** (ex. 1-5, ⋯) :

Nonbonding
O(***N²***)

# Nonbonded interactions

- Nonbonded interactions (Coulomb + vdW) are the main **bottleneck** of simulations.
- Nonbonded interactions are divided into those of real space with cutoff-distance and reciprocal lattice space.

All pairs
O(***N***$^2$)



Pairs within cutoff-distance +
O(C***N***)

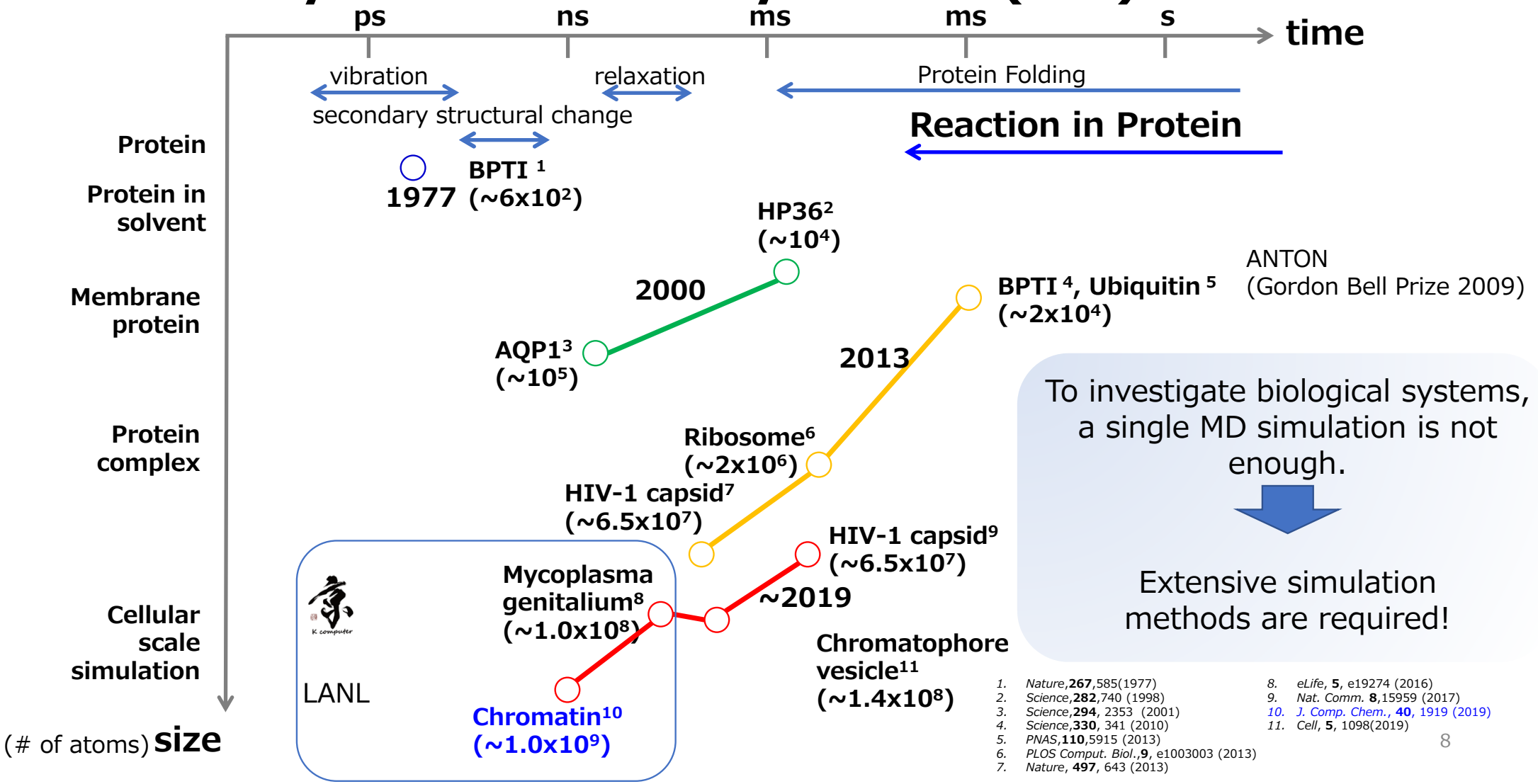
Long-range interactions

$$\sum_{|r_{ij}|<R\&i,j\notin bonding}\left\{\varepsilon_{ij}\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12}-2\left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right]f(r_{ij})+\frac{q_iq_j\mathrm{erfc}(\alpha r_{ij})}{r_{ij}}\right\}+\sum_{\mathbf{k}\neq 0}\frac{\exp(-\mathbf{k}^2/4\alpha^2)}{\mathbf{k}^2}\mathrm{FFT}(Q(\mathbf{k}))$$

**Real space, O(C*N*)** | **Reciprocal space, O(*N*log*N*)**

Darden *et al.*, *J. Chem. Phys.*, **98**, 10089-10092 (1993).
Essmann *et al.*, *J. Chem. Phys.*, **103**, 8577-8593 (1995).

# History of molecular dynamics (MD) simulation




To investigate biological systems, a single MD simulation is not enough.

Extensive simulation methods are required!

1. *Nature*,**267**,585(1977)
2. *Science*,**282**,740 (1998)
3. *Science*,**294**, 2353 (2001)
4. *Science*,**330**, 341 (2010)
5. *PNAS*,**110**,5915 (2013)
6. *PLOS Comput. Biol.*,**9**, e1003003 (2013)
7. *Nature*, **497**, 643 (2013)
8. *eLife*, **5**, e19274 (2016)
9. *Nat. Comm.* **8**,15959 (2017)
10. *J. Comp. Chem.*, **40**, 1919 (2019)
11. *Cell*, **5**, 1098(2019)

# Contents

- Molecular Dynamics (MD)
- MD software, GENESIS
- What we can do by using GENESIS

# **Gen**eralized **E**nsemble **Si**mulation **S**ystems (GENESIS)

- MD software developed in RIKEN from 2009
  **https://www.r-ccs.riken.jp/labs/cbrt/**

Version 1.0: Jung, Mori, *et al*. *WIREs Comput. Mol. Sci.* **5**:310-323, (2015)
Version 1.1.x: Kobayashi, Jung, *et al. J. Comput. Chem.* **38**, 2193-2206 (2017)
Version 2.x : Jung, Yagi, Tan, *et al. J. Phys. Chem. B* **128**, 6028-6048 (2024)

**Please find the website by search engine, "GENESIS RIKEN".**

- Free software under LGPLv3
- Two key features:
- High performance in large-scale biomolecular simulations.
- Efficient conformational sampling based on various multi-copy methods.
- → Take the advantage of high performances of supercomputers

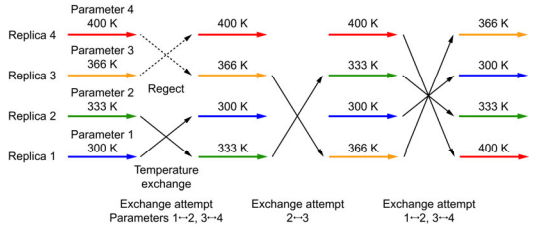
# Key features

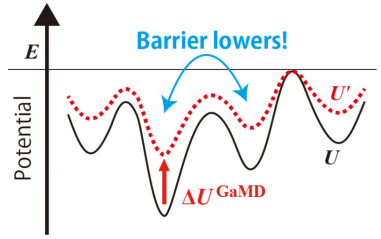
## High performance

富岳 GPU K computer

→ 2nd day (*Jung*)

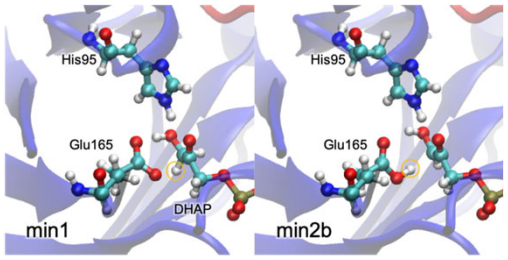
## Efficient conformational sampling methods



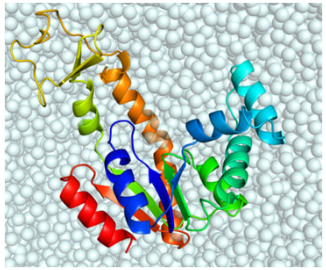




→ 3rd day (*Ito*)

## Enable QM/MM, all-atom and coarse-grained models

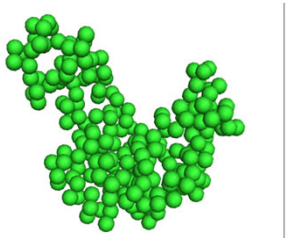



QM/MM



All-atom (AA)

→ Today

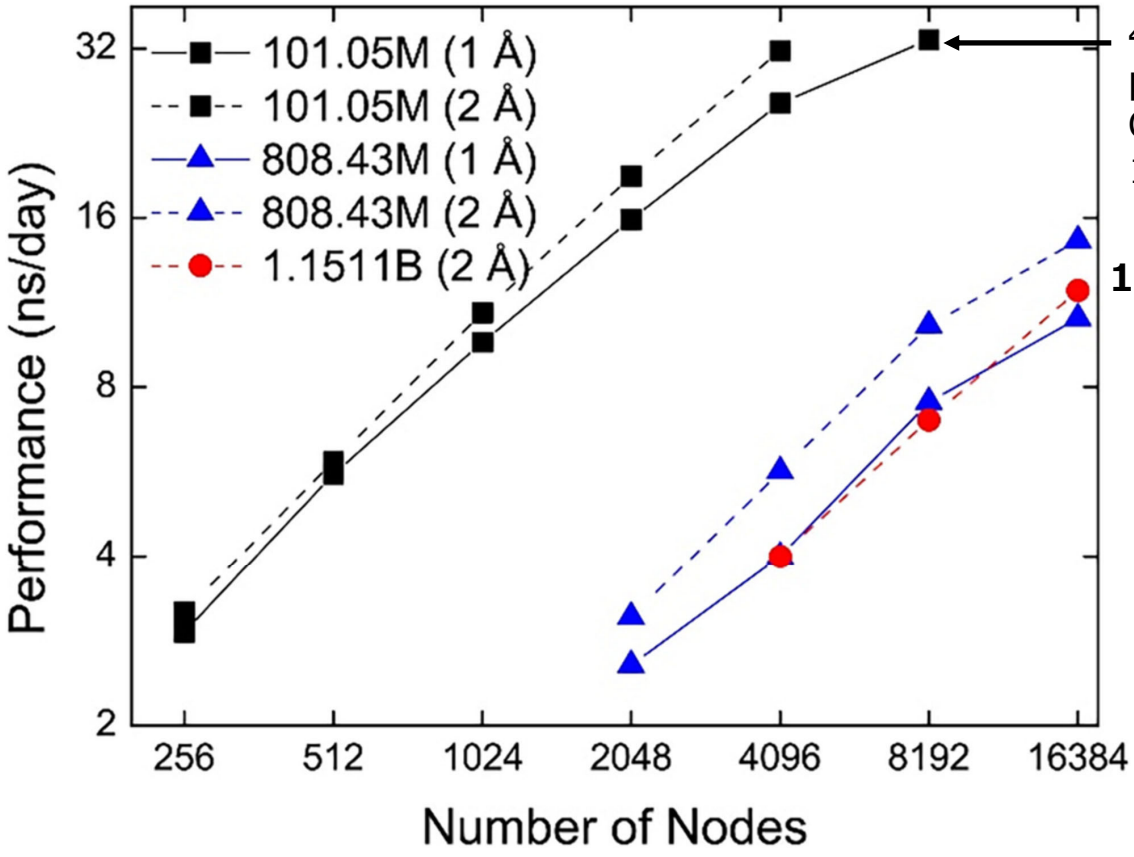


Coarse grained (CG)

→ 2nd day (*Tan*)

# Performance of AA model on Fugaku

Jung *et al.* *J. Comput. Chem.* ***42***, *231-240 (2021)* https://doi.org/10.1002/jcc.26450

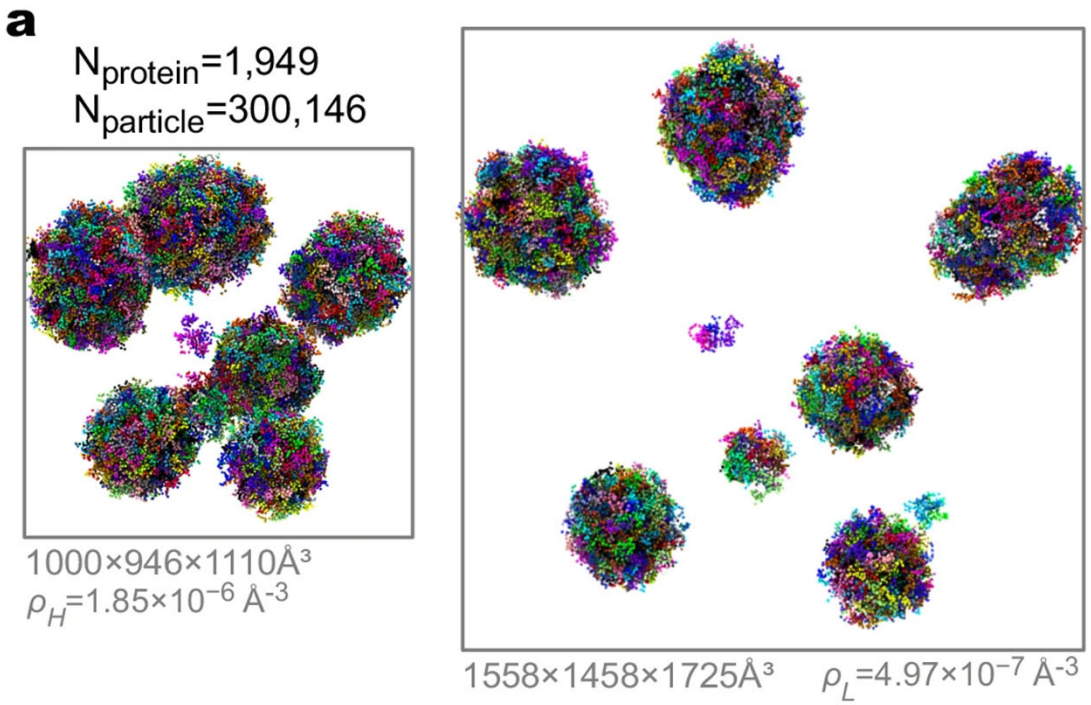


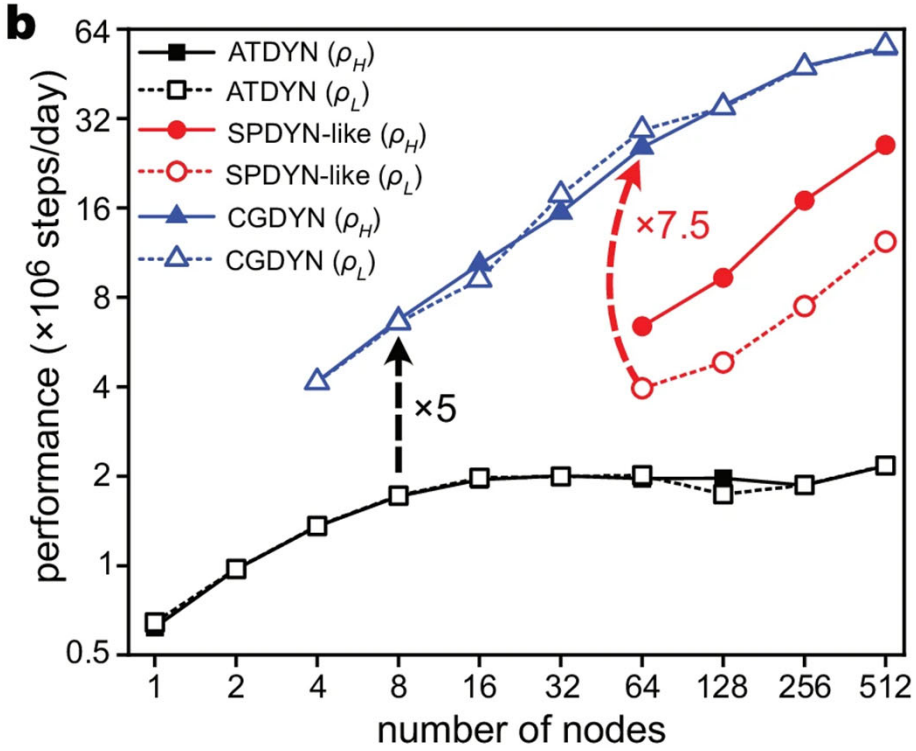

4 times better performance than GENESIS1.0 on K using 1/8 nodes

**11.9 ns/day**

# Performance of CG model on Fugaku

Jung *et al. Nat. Commun.* **15**, 3370 (2024).



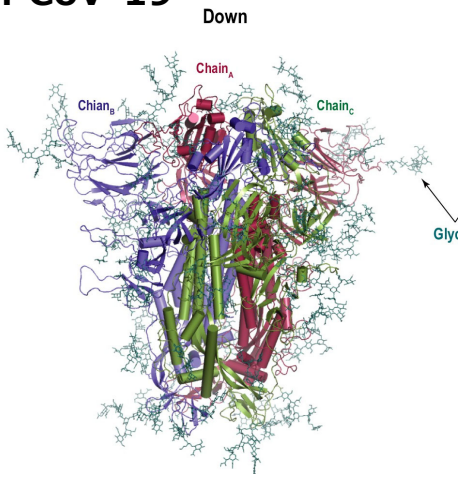
# Efficient conformational sampling methods

Sampling methods in GENESIS have been applied to various biological systems.

## gREST

Kamiya and Sugita, *J. Chem. Phys.* **149**, 072304 (2018)

### S-protein on surface of CoV-19

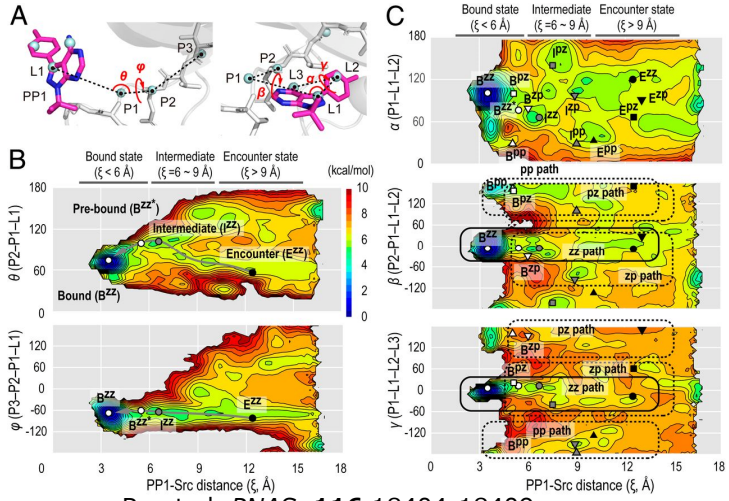



Dokainish *et al.*, *eLife,* **11**, e75720 (2022)

## gREST/REUS

REMD : Sugita and Okamoto, *Chem. Phys. Lett.* **314**, 141-151 (1999)
REUS : Sugita *et al.*, *J. Chem. Phys.* **113**, 6042-6051 (2000)

### Kinase-inhibitor binding

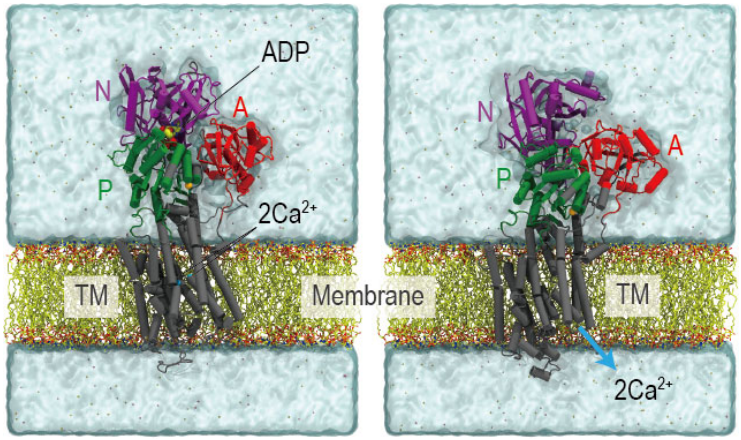



Re et al. *PNAS*, **116** 18404-18409 (2019)

## String method

Maragliano et al., *J. Chem. Phys.* **125**, 024106 (2006)

### Conformational changes of $Ca^{2+}$-ATPase




Kobayashi *et al.*, *PNAS, **118**,* e2105507118 (2021)

**⋯ and there's more!**

# Enable multiple models

Various models with different resolutions are available in GENESIS.

| Force field | Input files | Setup tool |
|---|---|---|
| CHARMM | top, par, psf, pdb (or crd), str | VMD, PSFGEN, CHARMM-GUI, CHARMM |
| AMBER | prmtop, pdb, (or ambcrd) | LEaP |
| KB Go-model | top, par, psf, pdb | MMTSB server |
| All-atom Go-model | grotop, grocrd (or pdb) | SMOG server, SMOG2 |

From GENESIS manual

AA FFs CG models

Force fields: CHARMM/CHARMM19/AMBER/MARTINI/Ca GO/All-atom GO/RESIDCG

For QM/MM : the following QM software can be used.

- Gaussian09/Gaussian16 (http://gaussian.com)
- Q-Chem (http://www.q-chem.com)
- TeraChem (http://www.petachem.com)
- DFTB+ (https://www.dftbplus.org)
- Qsimulate (https://qsimulate.com/academic)

# Contents

- Molecular Dynamics (MD)
- MD software, GENESIS
- What we can do by using GENESIS

# Which machines we can execute GENESIS?

GENESIS can work on various machines!

## Operating systems

- Linux
- Mac OSX
- Windows 10, 11 (spdyn, ver > 1.7.1)

## Fortran and C compilers

- GCC compiler: gfortran, *gcc* (version > 7.0)
- Intel compiler: *ifort, icc*
- Fujitsu compiler: *frtpx, fccpx*
- Cygwin/mingw (spdyn, ver > 1.7.1)

## Supercomputers

- Fugaku (RIKEN)
- Flow (Nagoya)
- TSUBAME3.0 (Tokyo Tech.)
- Cygnus (Tsukuba)
- Oakbridge-CX (Tokyo)

etc

## Hardware

- Intel® Xeon ® (> SSE4.2), Xeon Phi ®
- AMD EPYC™
- ARM (Armv8.2-A SVE)
- NVIDIA GPU (CC >= 3.5)

# What we can do by using GENESIS? (1)

We have two MD applications; **atdyn** and **spdyn**.

- ATDYN : Many models (QM/MM, AA, CG) & readable codes **CG, QM/MM, Cryo-EM**
- SPDYN : High performance **Basic, enhanced-sampling**

Table 3.1: Available functions in **atdyn** and **spdyn**

| Function | atdyn | spdyn |
|---|---|---|
| Energy minimization | ○ (SD and LBFGS) | ○ (SD) |
| All-atom molecular dynamics | ○ | ○ |
| Coarse-grained molecular dynamics | ○ | ○ (Only Martini) |
| Implicit solvent model | ○ | – |
| Replica-exchange method | ○ | ○ |
| Gaussian accelerated MD | ○ | ○ |
| Reaction path search | ○ (MEP and MFEP) | ○ (MFEP) |
| QM/MM calculation | ○ | – |
| Vibrational analysis | ○ | – |
| Cryo-EM flexible fitting | ○ | ○ |
| Precision | double | double/mixed/single |
| GPGPU calculation | – | ○ (All-atom MD) |
| Parallel I/O | – | ○ |

# What we can do by using GENESIS? (2)

We can analyze trajectories in GENESIS.

**Today (*Kobayashi*)**
**3rd day (*Ito*)**

Most tools can execute on desktop and laptop;

- Distance/angle/dihedral angles : **trj_analysis**
- RMSD: **rmsd_analysis**
- Conversion of trajectory files : **crd_convert**, **remd_convert**
- PCA: **avecrd_analysis**, **flccrd_analysis**, **eigmat_analysis**, **prjcrd_analysis** (**avecrd_analysis**, **flccrd_analysis** can be used to RMSF)
- Free energy calculation: **mbar_analysis**, **wham_analysis**, **pmf_analysis** …
- Diffusion: **msd_analysis**, **diffusion_analysis**
- Lipid properties: **lipidthink_analysis**, **tilt_analysis**

Highlighted tools will be shown in hands-on

# What we can do by using GENESIS? (3)

Modeling with coarse grained model can be done in GENESIS.

## GENESIS CG tools

**2nd day (*Tan*)**

Tan *et al.*, *PLoS Comp. Biol.* **18**, e1009578 (2022)

https://github.com/genesis-release-r-ccs/genesis_cg_tool

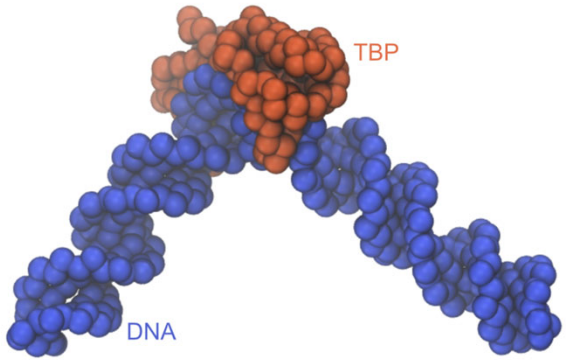
**Experimental data**

- .pdb :atomic coordinates
- .fasta :sequence
- .pfm :position frequency matrix

GENESIS_CG_tools

**MD input**

- .gro :CG coordinates
- .top :CG topology main file
- .itp :CG topology include file

# Tutorial of GENESIS (1)

Please find “Tutorial and Samples” in our website.


**Computer resources**

- : Suitable for laptop or small desktop machine (less than 4 CPU cores)
- : Suitable for typical Linux workstation (~16 CPU cores)
- : Suitable for cluster machine or super-computer (more than 64 CPU cores)

**Level 1: Basic tutorials** (Level1-3)

1. Getting started
   - 1.1 Installation of GENESIS for Tutorials
   - 1.2 Let’s take a quick look at the source code of GENESIS
2. Preparation of the input files for GENESIS
   - 2.1 3D structure of biological molecules
   - 2.2 Force field parameters of biological molecules
   - 2.3 Building the initial structure for MD simulation
3. MD simulations of peptides and proteins with the all-atom CHARMM force field
   - 3.1 Ala-dipeptide in the gas-phase

https://www.r-ccs.riken.jp/labs/cbrt/tutorials2022/

# Tutorial of GENESIS (2)

We can find how to setup MD inputs in the tutorial site.

2. Preparation of the input files for GENESIS
   - 2.1 3D structure of biological molecules
   - 2.2 Force field parameters of biological molecules
   - 2.3 Building the initial structure for MD simulation

(skip)

5. Preparation of the input files for various systems
   - 5.1 Creating input files of MD simulations with the CHARMM force field
   - 5.2 Creating input files of MD simulations with the AMBER force field
   - 5.3 Creating input files of MD simulations using the Gromacs input files
6. MD simulations of various biomolecules with all-atom models
   - 6.1 POPC lipid bilayer
   - 6.2 GPCR in a lipid bilayer
   - 6.3 N-glycan in water
   - 6.4 RNA in water

Setup by vmd[1]
(CHARMM FF: Soluble protein, DNA-protein)

Setup by AmberTools[2]
(AMBER FF : Soluble protein, DNA-protein)

Setup by Gromacs tool[3]
(AMBER FF : Soluble protein, DNA-protein)

Setup by CHARMM-GUI[4]
(CHARMM FF : Lipid bilayers, membrane protein, N-glycan)

1. https://www.ks.uiuc.edu/Research/vmd/
2. http://ambermd.org/AmberTools.php
3. https://www.gromacs.org/
4. https://www.charmm-gui.org/

Please check the license issue by yourselves.

# Summary

- GENESIS -
  - is MD software for biomolecule system.
  - has three key features ;
    - High performance.
    - Efficient sampling methods.
    - Available QM/MM, all-atom, and coarse-grained models.
  - includes two MD applications and more than 30 analysis tools.
- You can find how to use GENESIS in the GENESIS website.

**Please enjoy this tutorial!**