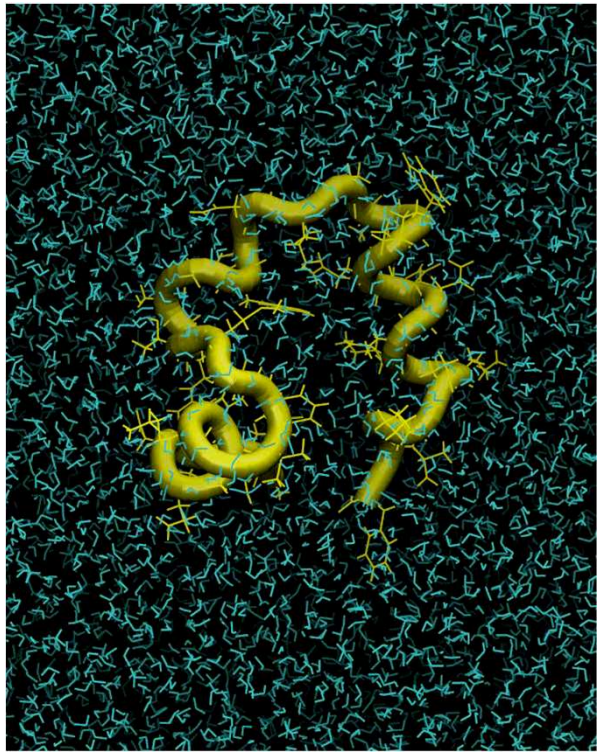
GENESIS
Generalized-ensemble simulation system

Acceleration of large-scale MD simulation for biological functions

Jaewoon Jung and Yuji Sugita
Computational Biophysics Research Team
RIKEN Center for Computational Science

MD simulations for biomolecules



In MD, we solve classical equations of motions to examine protein structure, dynamics, and function relationship.

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

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

Equation of motion

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

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

Integration



Nobel Prize in Chemistry 2013 was awarded to M. Karplus, M. Levitt, and A. Warshel “for the development of multiscale models for complex chemical systems”.

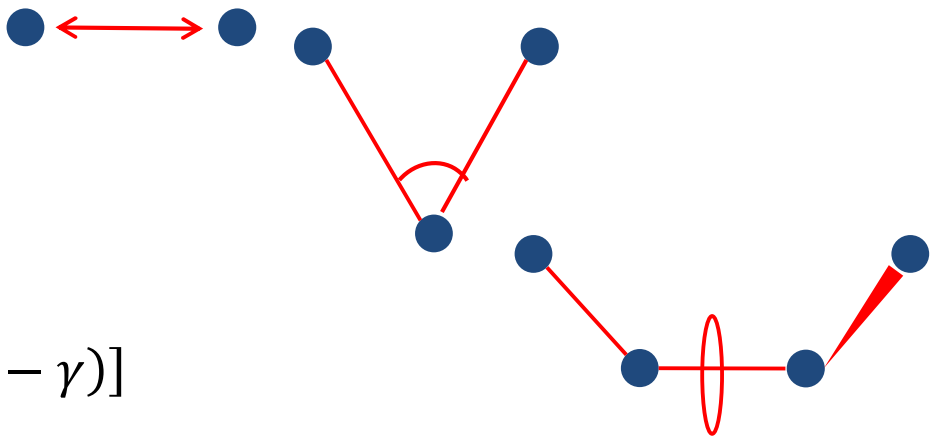
Potential energy in MD

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

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

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

$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left[\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_i q_j}{r_{ij}} \right]$$



O(*N*)

O(*N*)

O(*N*)

O(*N*²)

Main bottleneck in MD

$$\sum_{\substack{ij \\ |r_i - r_j| < R}}^{N} \left[\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_i q_j \text{erfc}(\alpha r_{ij})}{r_{ij}} \right]$$

Real space, O(C*N*)

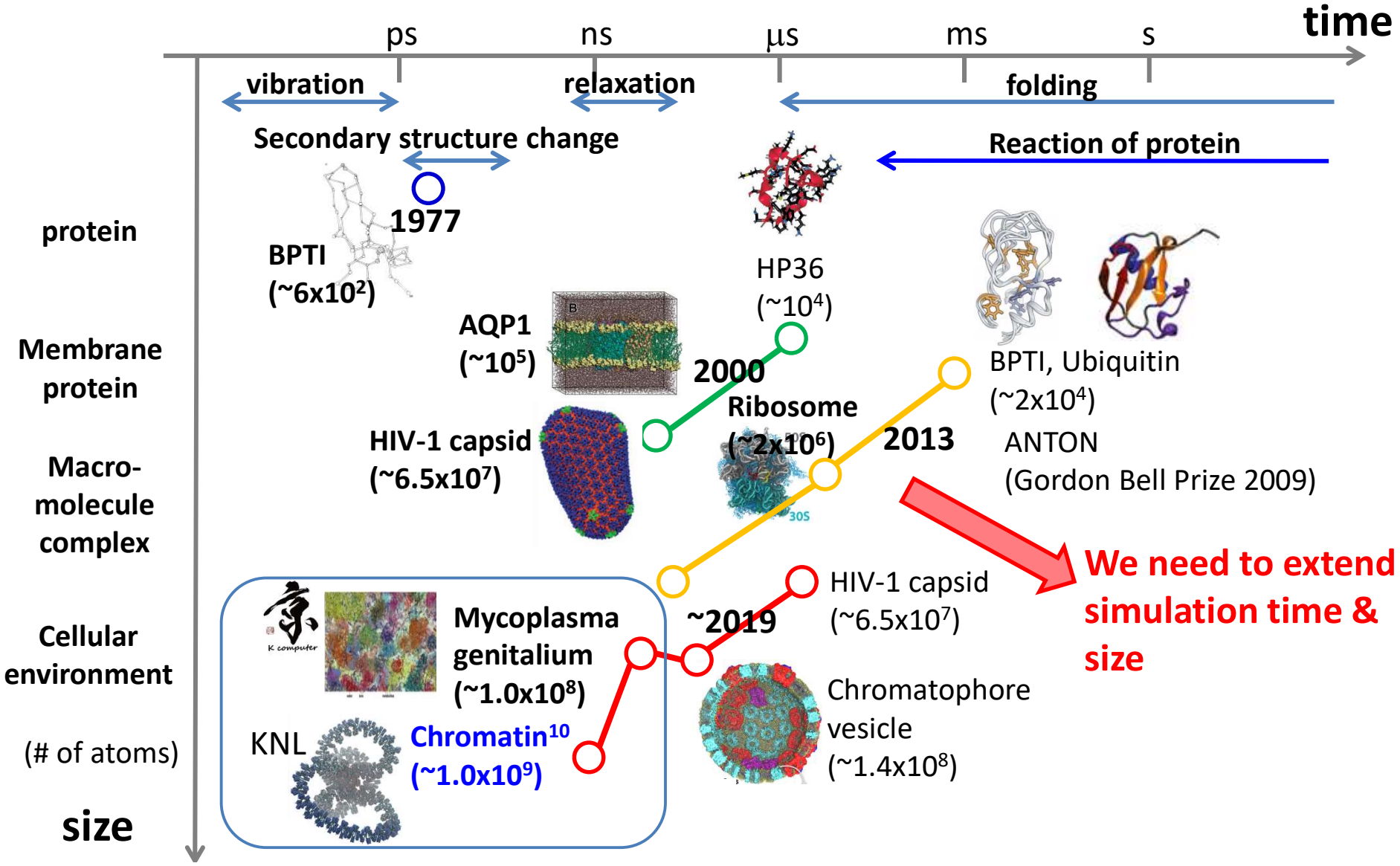
$$+ \sum_{\mathbf{k} \neq 0} \frac{exp(-k^2/4\alpha^2)}{k^2} FFT(Q(\mathbf{k}))$$

Reciprocal space, O(*N*log*N*)

GENESIS
Generalized-ensemble simulation system

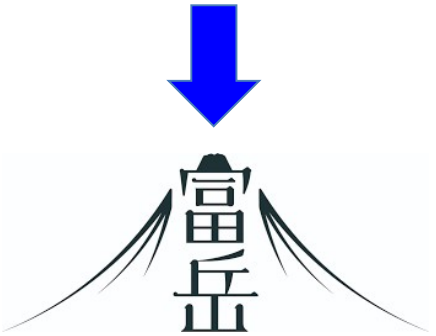
History of Biological MD simulations

(1977 to current)

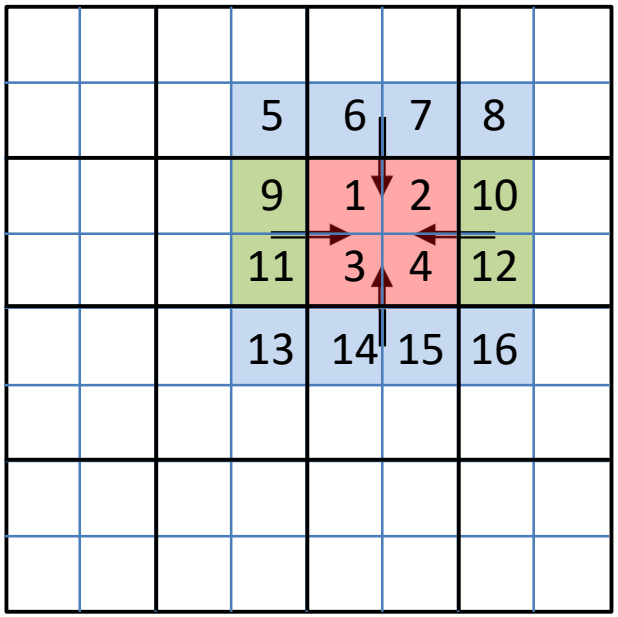


GENESIS is highly parallelized MD software for biomolecular simulations

- We have recently developed new MD software, GENESIS (Generalized-Ensemble Simulation System).
  - Jaewoon Jung, Takaharu Mori, et al. WIREs CMS, 5, 310-323 (2015)
- The development of GENESIS was motivated to for large scale MD simulation on K computer
- We intend to solve the two problems in MD.
  - Time-Scale Problem
    - The enhanced conformational sampling algorithms
    - Parallelization with a number of “Replicated” MD simulations.
  - Size Problem
    - The domain decomposition methods for nonbonded interactions
    - Parallelization for a single MD simulation.
- Using Fugaku, we can more extend the time-scale and target system size.



Parallelization of GENESIS (real-space)



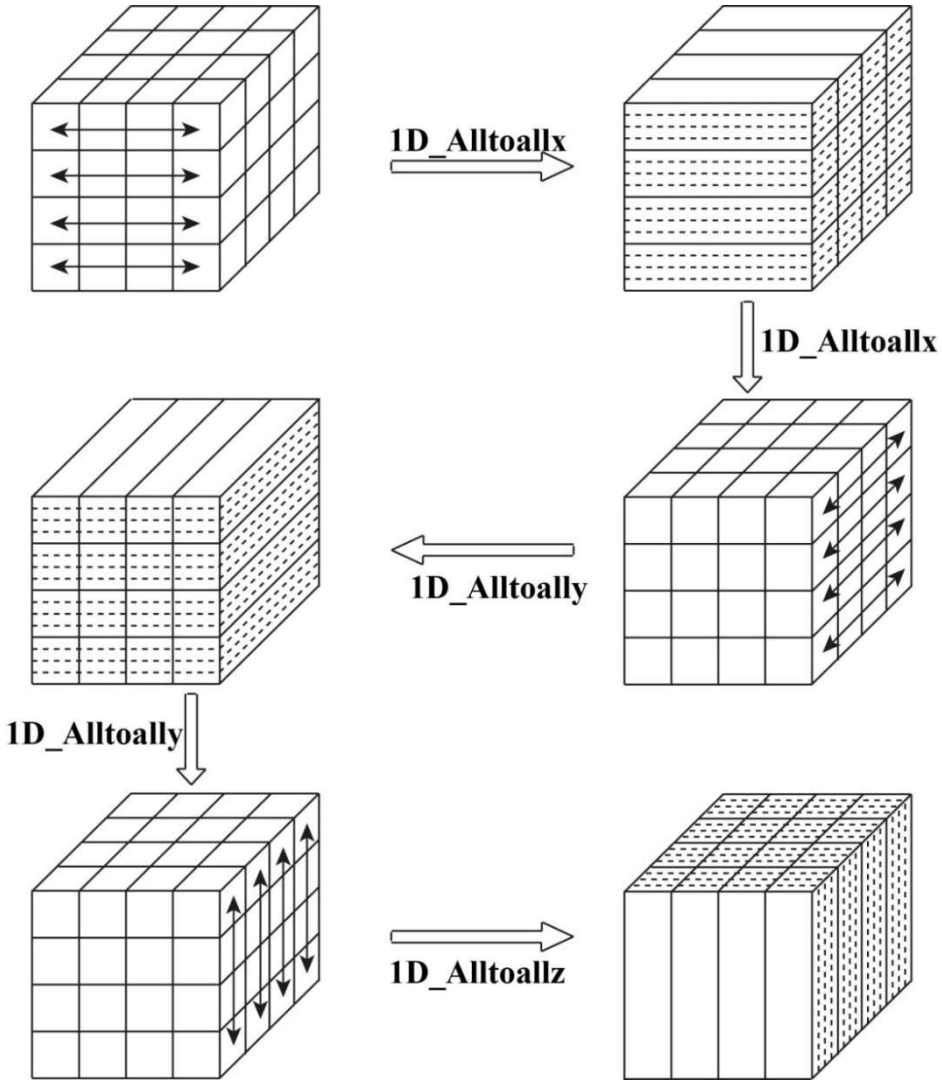
Unit cell

Subdomain of each MPI

Cell pair	Midpoint cell?	OpenMP parallelization
(1,1)	Yes	← thread1
(2,2)	Yes	← thread2
…	…	
(1,2)	Yes	← thread4
(1,3)	Yes	← thread1
…	…	
(1,5)	Yes	← thread2
(1,6)	No	
(1,7)	No	
(1,8)	Yes	← thread3
(1,9)	No	
(1,10)	Yes	← thread4
…	…	

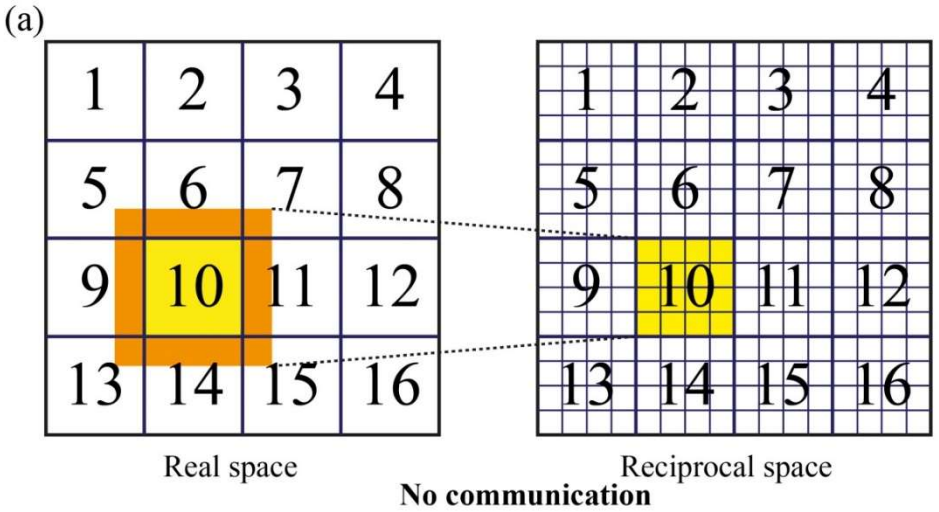
J. Jung, T. Mori, and Y. Sugita, *J. Comput. Chem.* **35**, 1064-1072 (2014).
J. Jung et al., *Wiley Interdiscip. Rev. Mol. Sci.* **5**, 310-323 (2015).
C. Kobayashi et al., *J. Comput. Chem.* **38**, 2193-2206 (2017).

Parallelization of GENESIS (reciprocal-space)



1. More frequent communications than existing MD programs
2. MPI Alltoall communications only in one-dimensional space (existing : communications in two/three-dimensional space)
3. Reduce communicational cost for large number of processors by reducing the number of processes involved in communications
4. For Fugaku, GENESIS choose the best FFT scheme automatically according to the process numbers.

Connection between real- and reciprocal-space parallelization in GENESIS



GENESIS

Real-space :
Volumetric decomposition
Reciprocal-space:
Volumetric decomposition

No communication is needed.

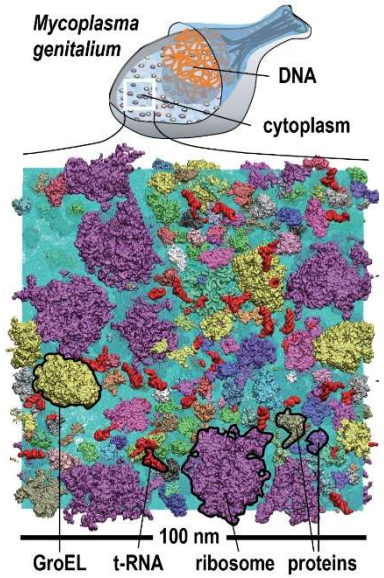
Other MD programs

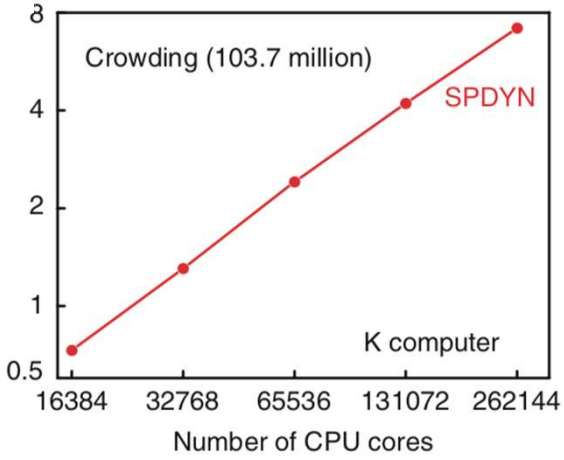
Real-space :
Volumetric decomposition
Reciprocal-space:
Pencil or slab decomposition

Communication is necessary.

J. Jung, C. Kobayashi, T. Imamura, and Y. Sugita, *Comp. Phys. Comm.200*, 57-65 (2016).

Performance of GENESIS on K computer



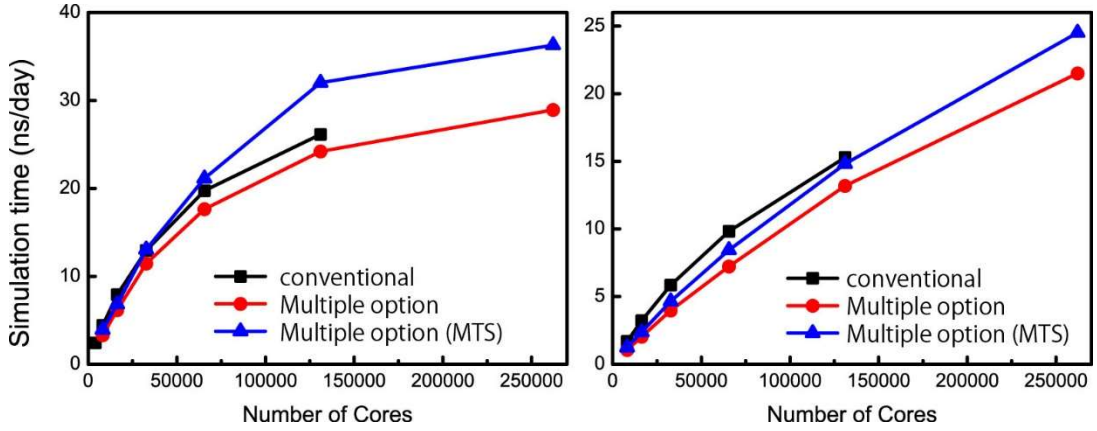


The first 100 million atom simulation of biomolecules

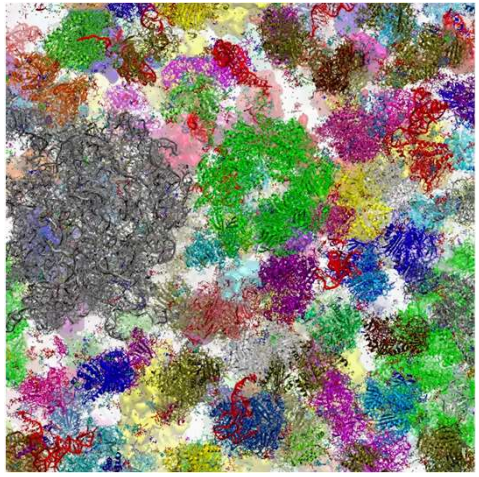
GENESIS shows good parallel efficiency on K computer, in particular, for very large biological systems.

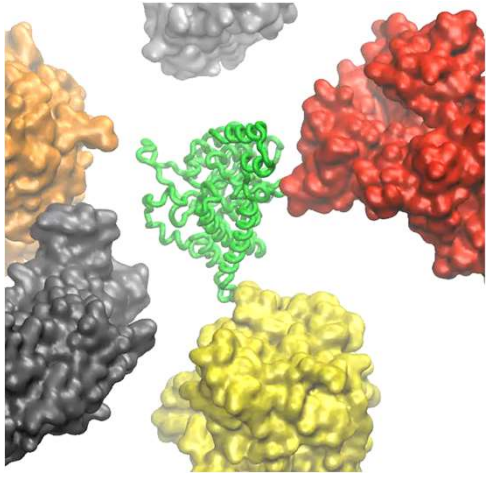
8.5 million atoms

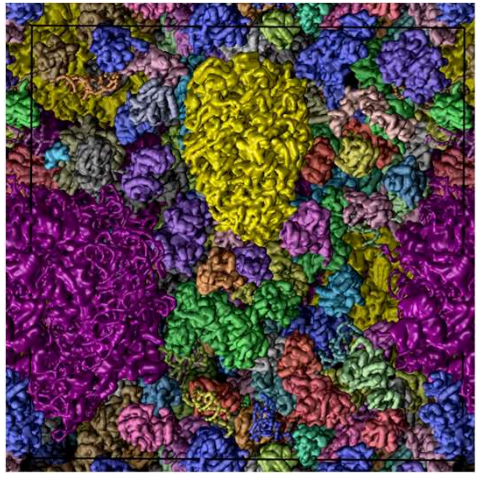
28 million atoms

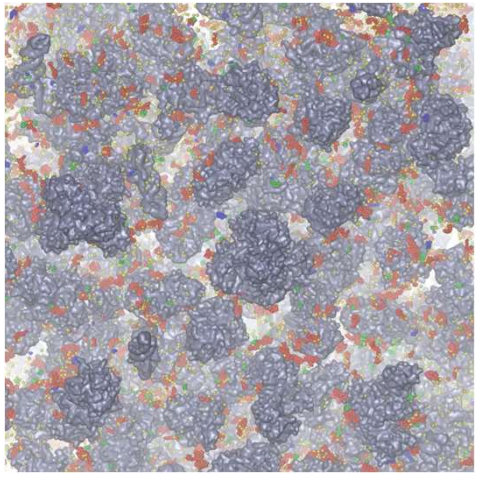


MD simulation example on K (cytoplasm)



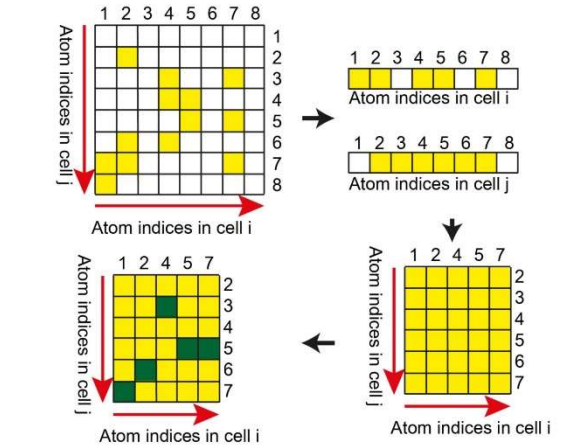


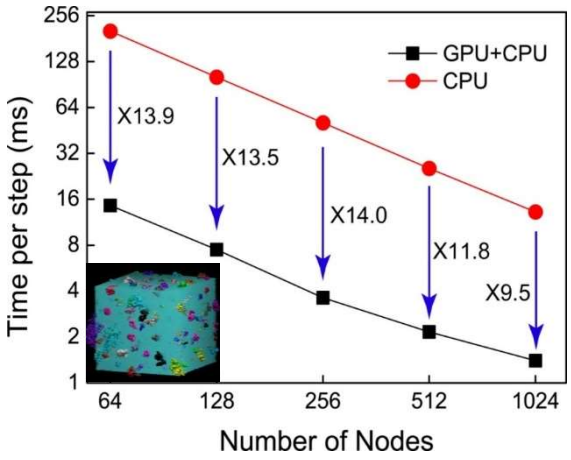




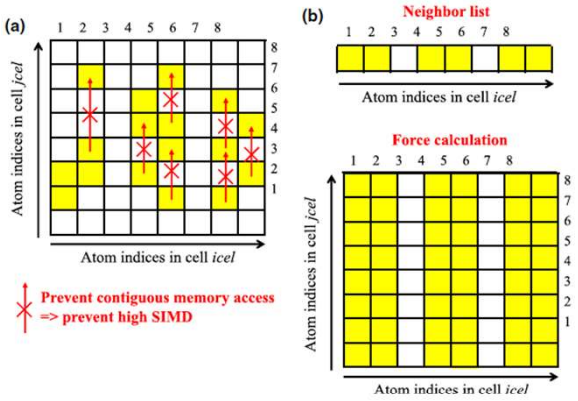
GENESIS performance developments for various hardware architectures

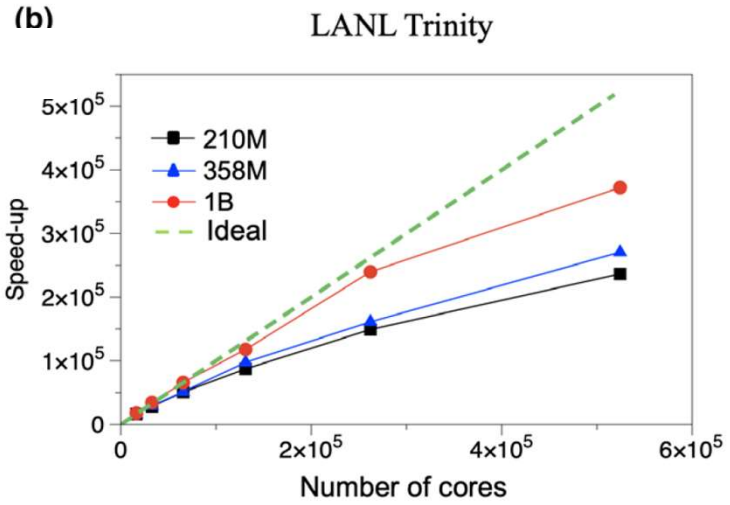
GPU non-bonded scheme





KNL non-bonded scheme

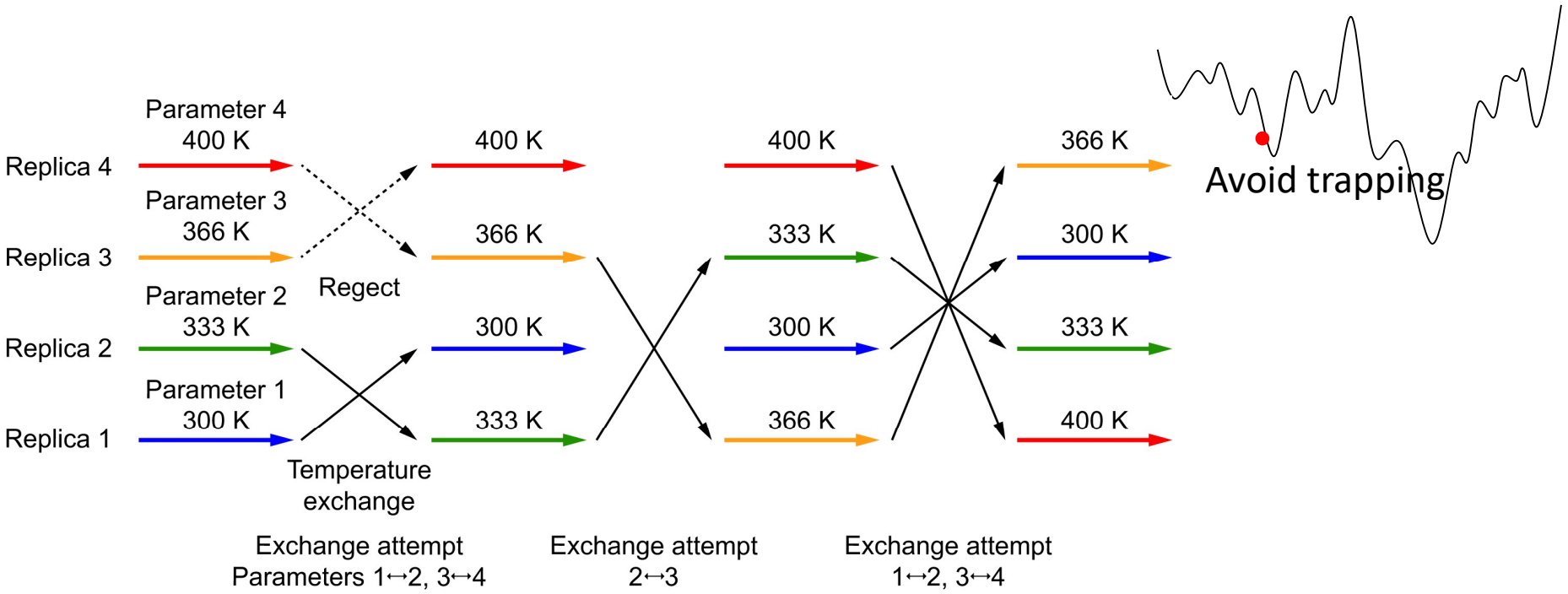




J. jung, A. Naruse, C. Kobayashi and Y. Sugita, *J. Chem. Theory Comput.* **12**, 4947-4958 (2016).
J. Jung et al. *J. Comput. Chem.* **40**, 1919-1930 (2019)

Enhanced sampling scheme with replicas

1. From conventional MD without very long time-scale, simulation can be trapped at one of the local energy minimum.
2. Using the exchange of temperatures or other parameters between replicas, we can sample a wider conformational space than the conventional MD.

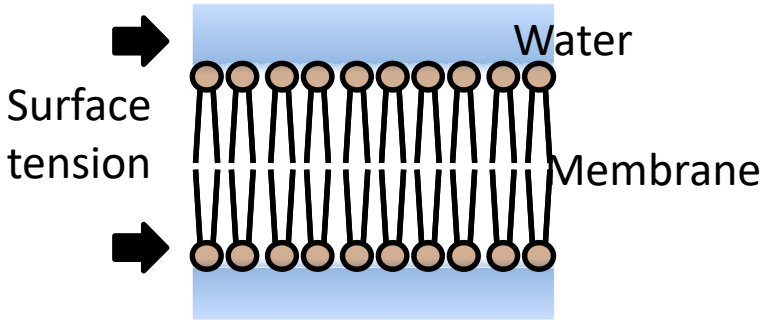


Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141-151 (1999).

Enhanced sampling scheme developments in GENESIS

1. Temperature REMD
2. Surface-tension REMD
3. Generalized REST (replica exchange with solute tempering)
4. Multidimensional replica exchanges (REST/REUS, gREST/REUS and so on)
5. Reaction method with string method

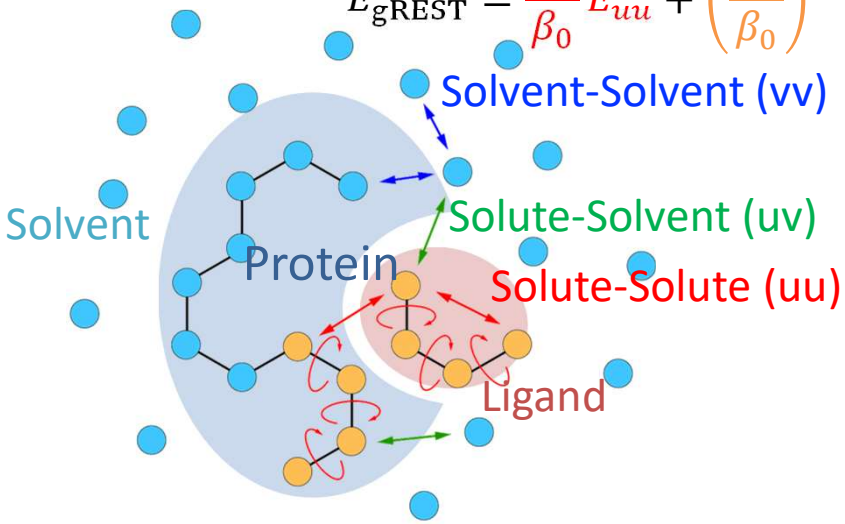
Surface-tension REMD



Exchange surface tensions in NPγT (surface tension constant) ensemble

gREST

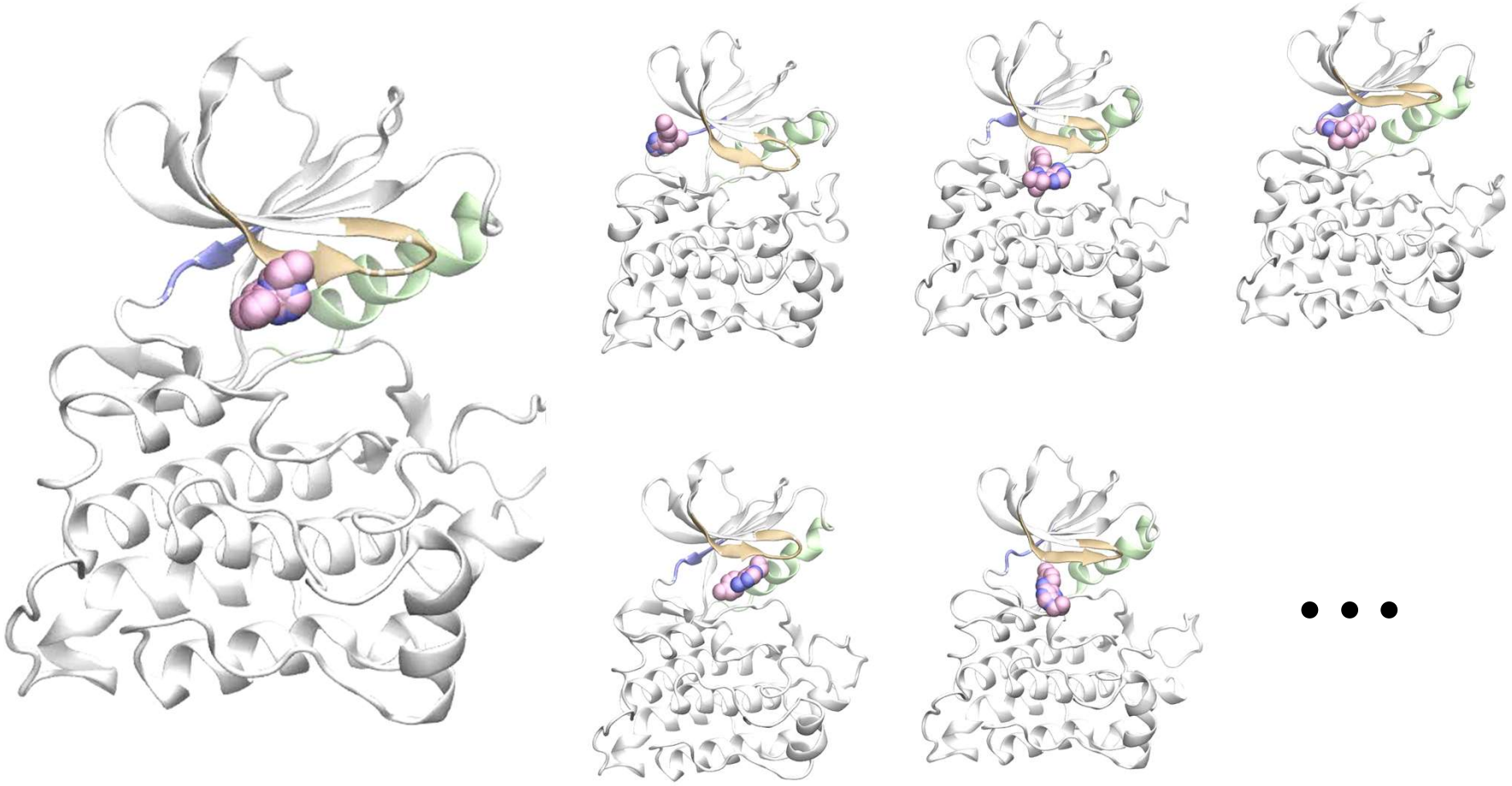
$$E_{\mathrm{gREST}} = \frac{\beta_m}{\beta_0} E_{uu} + \left(\frac{\beta_m}{\beta_0}\right)^{l/n} E_{uv} + E_{vv}$$



T. Mori, J. Jung, and Y. Sugita, *J. Chem. Theory Comput.* **9**, 5620-5640 (2013).
Y. Matsunaga et al. *J. Phys. Chem. Lett.* **7**, 1446-1451 (2016)
M. Kamiya and Y. Sugita, *J. Chem. Phys.* **149**, 072304 (2018)

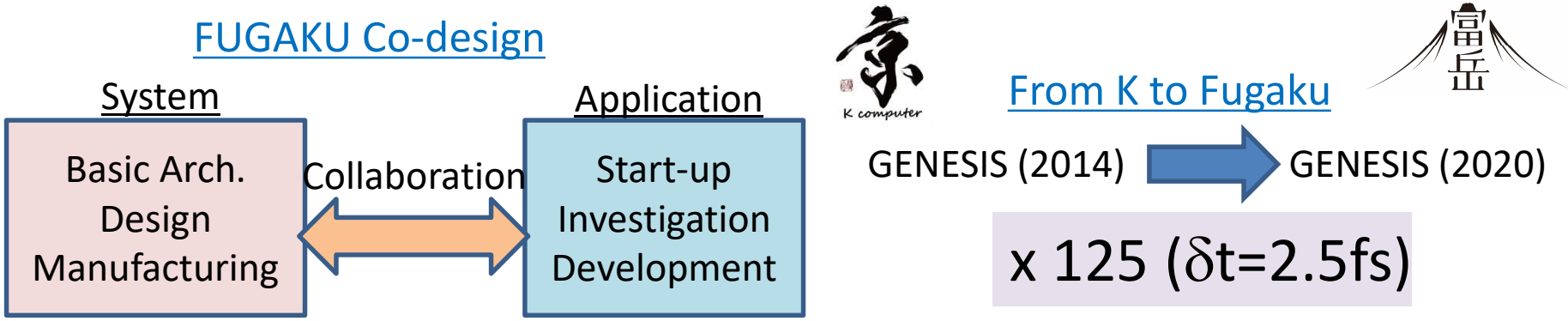
Sampling Hundred of Binding Events

Movies from arbitrary selected replicas



GENESIS development on Fugaku

1. Optimization of GENESIS for Fugaku supercomputer: Continuous effort from Fugaku co-design
   1) New non-bonded interaction kernel with increased SIMD width.
   2) Optimization of parallelization of Fugaku



2. New functions/libraries
   1) New parallel I/O to deal with inputs/outputs for very huge system.
   2) Stable integration with increased time step.
   3) Enhanced sampling schemes using replicas.

Summary

1. GENESIS is developed for high performance large-scale MD on K computer.

2. With efficient parallelization, GENESIS can perform MD simulations for 100 million and 1 billion atom systems.

3. Various enhanced sampling schemes are also developed in GENESIS to overcome current MD time-scale limitation.

4. With the development of GENESIS on Fugaku, we expect 125 times performance improvement and more realistic MD simulations for huge systems.