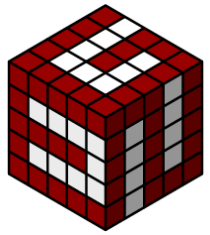


GENESIS

Generalized-ensemble simulation system

Development of GENESIS on Fugaku supercomputer and its application of Spike protein on the surface of SARS-CoV-2 in solution

2020年度第1回計算科学フォーラム
2020/11/24

Jaewoon Jung

Computational Biophysics Research Team,

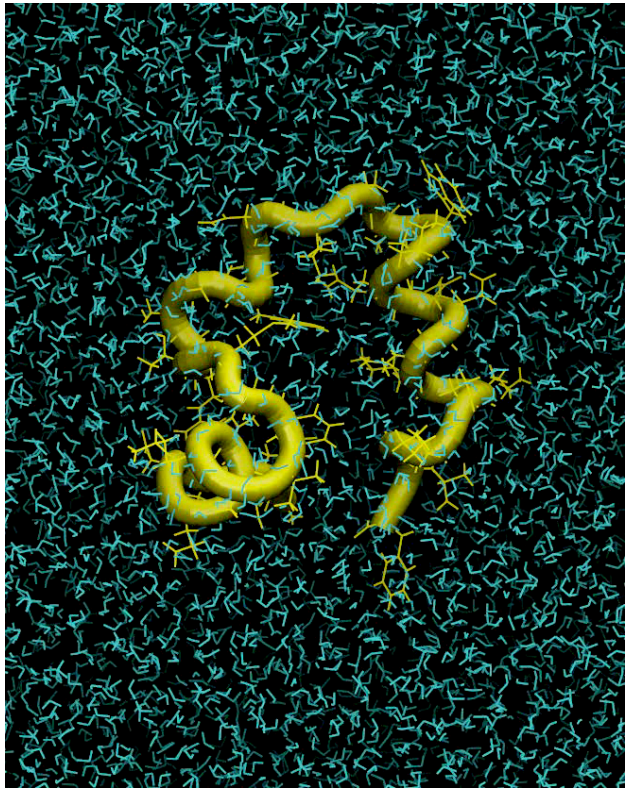
RIKEN Center for Computational Science

Molecular dynamics (MD) simulation

- MD is useful to predict molecular motions at the atomic resolutions.
- In classical MD, newton's equation of motion is solved iteratively.

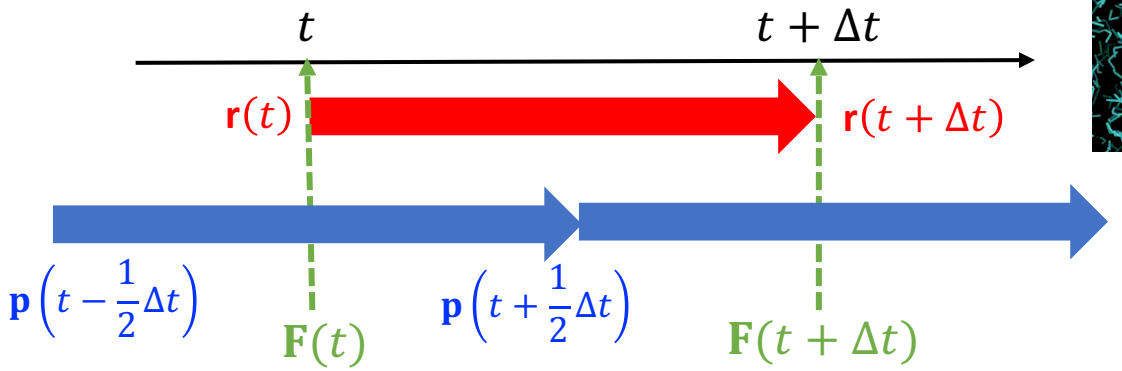
$$F = ma$$

- Time step Δt should be very short to reproduce fast vibrations in molecules. (10^{-15} sec). Many iterations are required to detect protein dynamics.



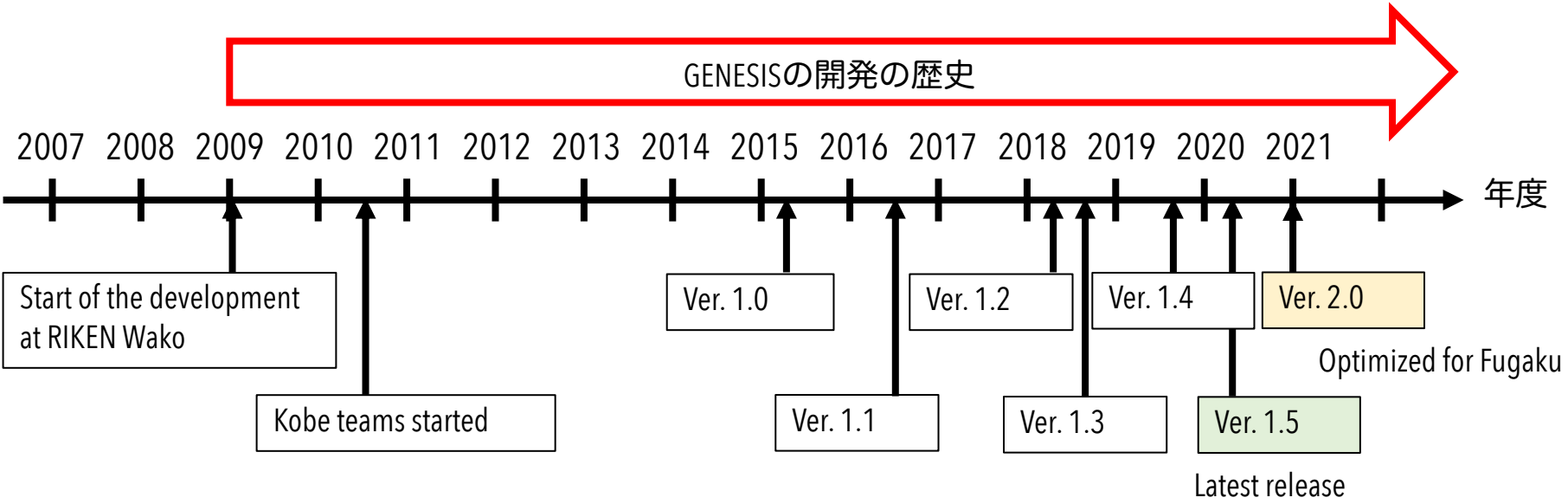
$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{p}_i\left(t+\frac{1}{2}\Delta t\right)\Delta t}{m_i}$$

$$\mathbf{p}_i\left(t+\frac{1}{2}\Delta t\right) = \mathbf{p}_i\left(t-\frac{1}{2}\Delta t\right) + \mathbf{F}_i(t)\Delta t$$

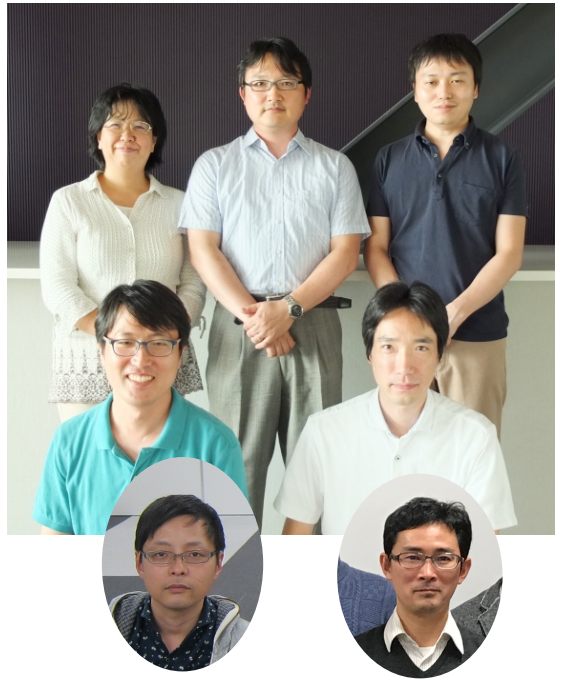


MD software GENESIS

- GENESIS has been developed in RIKEN.
- It allows high-performance MD simulations on parallel supercomputers like K, Fugaku, Tsubame, etc.
- It is free software under LGPL license.



GENESIS developers



GENESIS
Generalized-ensemble simulation system

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

GENESIS 2.0beta: Fugaku Codesign

Continuous effort on the acceleration of GENESIS on Fugaku

FUGAKU Co-design

System

Basic Arch.
Design
Manufacturing

Collaboration

Application

Start-up
Investigation
Development

From K to Fugaku

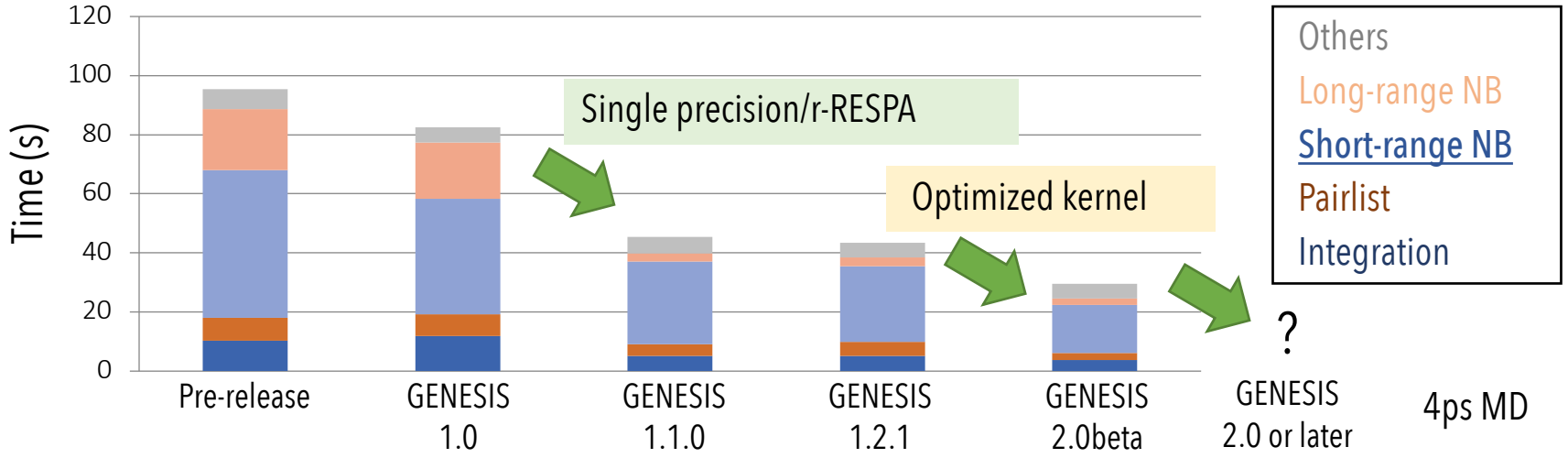
11.3 peta-flops X 35 400 peta-flops

GENESIS (2014) → GENESIS (2020)

x 125 (δt=2.5fs)

The acceleration was very important for flagship 2020 project.

Speed-up of GENESIS on Intel Xeon Gold 6148 (1 node, 40 cores)



How to optimize GENESIS 2.0beta on Fugaku?

Algorithm1 (X)

```
do i = 1, 16
  do j = 1, 16
     f(i,j)
  end do
end do
```

Algorithm2 (O)

```
do ij = 1, 256
  index i and j from ij
    f(i,j)
  end do
end do
```

1. In principle, there is no difference between Algorithm1 and Algorithm2 because of the same operation amount.
2. On Fugaku, Algorithm2 has better performance than Algorithm1.
3. The main reason is the operand waiting time for each do loop. On Fugaku, there are non-negligible waiting time before executing calculations in the do loop.
4. To minimize the waiting time, the most inner do loop length should be long.
5. If the most inner do loop length is small, it would be better not to vectorize when compiling.

```
do i = 1, 3
  f(i)
end do
```

→

```
!$ocl nosimd
do i = 1, 3
  f(i)
end do
```

```
f(1:3)
```

→

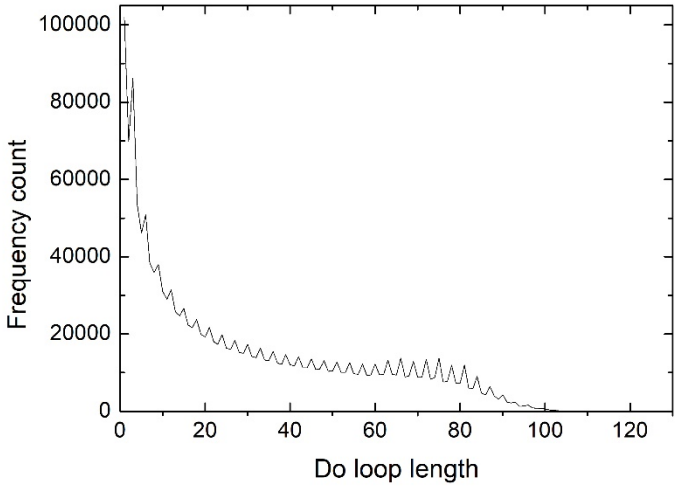
```
!$ocl nosimd
f(1:3)
```

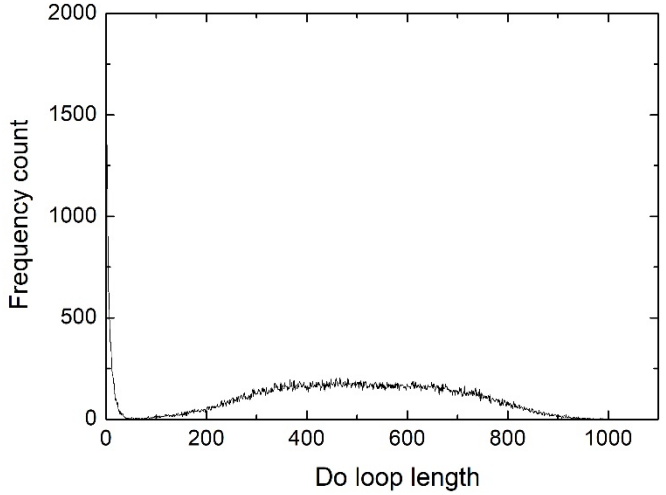
```
f(1), f(2), f(3)
```

New non-bonded interaction algorithm in GENESIS 2.0beta

```
do ijcel = 1, cell_pair
  obtain icel and jcel
  do i=1,natom(icel)
    do k=1,neighbor(i,ijcel)
      j = list(k,ijcel)
      interaction end do
  end do
end do
```

```
do icel = 1, cell
  do i=1,natom(icel)
    do k=1,neighbor(i,icel)
      list(k,i,icel)
      interaction end do
  end do
end do
```





ApoA1 on one MPI processor

Two Kinetic energy evaluations in MD

Integration

$$\tilde{\mathbf{p}}_i\left(t+\frac{1}{2}\Delta t\right)=\mathbf{p}_i(t)+\frac{1}{2}\mathbf{F}_i(t)\Delta t$$

$$\mathbf{r}_i(t+\Delta t)=\mathbf{r}_i(t)+\frac{\tilde{\mathbf{p}}_i\left(t+\frac{1}{2}\Delta t\right)\Delta t}{m_i}$$

$$\mathbf{p}_i(t+\Delta t)=\tilde{\mathbf{p}}_i\left(t+\frac{1}{2}\Delta t\right)+\frac{1}{2}\mathbf{F}_i(t+\Delta t)\Delta t)$$

Half-time step kinetic energy

$$K_{\text{full}}(t)=\sum_{i=1}^{N}\frac{\mathbf{p}_i(t)^2}{2m_i}$$

$$N_f k_B T=2\langle K_{\text{full}}(t)\rangle-\frac{\Delta t^2}{6}\sum_{i=1}^{N}\langle\mathbf{p}_i(t)\cdot\mathbf{r}_i{}^{(3)}(t)\rangle+O(\Delta t^4)$$

Full-time step kinetic energy

$$K_{\text{half}}(t)=\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\tilde{\mathbf{p}}_i\left(t-\frac{1}{2}\Delta t\right)^2}{2m_i}+\frac{\tilde{\mathbf{p}}_i\left(t+\frac{1}{2}\Delta t\right)^2}{2m_i}\right)$$

$$N_f k_B T=2\langle K_{\text{half}}(t)\rangle+\frac{\Delta t^2}{12}\sum_{i=1}^{N}\langle\mathbf{p}_i(t)\cdot\mathbf{r}_i{}^{(3)}(t)\rangle+O(\Delta t^4)$$

1. Full- and half-time step kinetic energies have the Δt^2 perturbation terms with opposite signs.
2. The perturbation terms are highly related to the high frequency vibrational motions.

Our suggestion of accurate temperature

$$N_f k_B T = \frac{4}{3}\langle K_{\text{half}}(t)\rangle + \frac{2}{3}\langle K_{\text{full}}(t)\rangle + O(\Delta t^4)$$

1. By combining full- and half-time step kinetic energies, we can obtain new temperature estimation which is accurate up to the third power of Δt.
2. The new temperature estimation does not require further computational cost.
3. With the new temperature estimation, we can enlarge the time step (Δt) up to **3.5 fs** for short-range force and **7.0 fs** for long-range force. (In conventional methods, Δt is limited up to 2.5 fs)

Similarly, we can estimate pressure in a more accurate way:

X $P(t) = \frac{1}{3V}(2K_{\text{full}}(t) + Vir(t))$

However, the error could be negligible with group pressure up to 2.5 fs

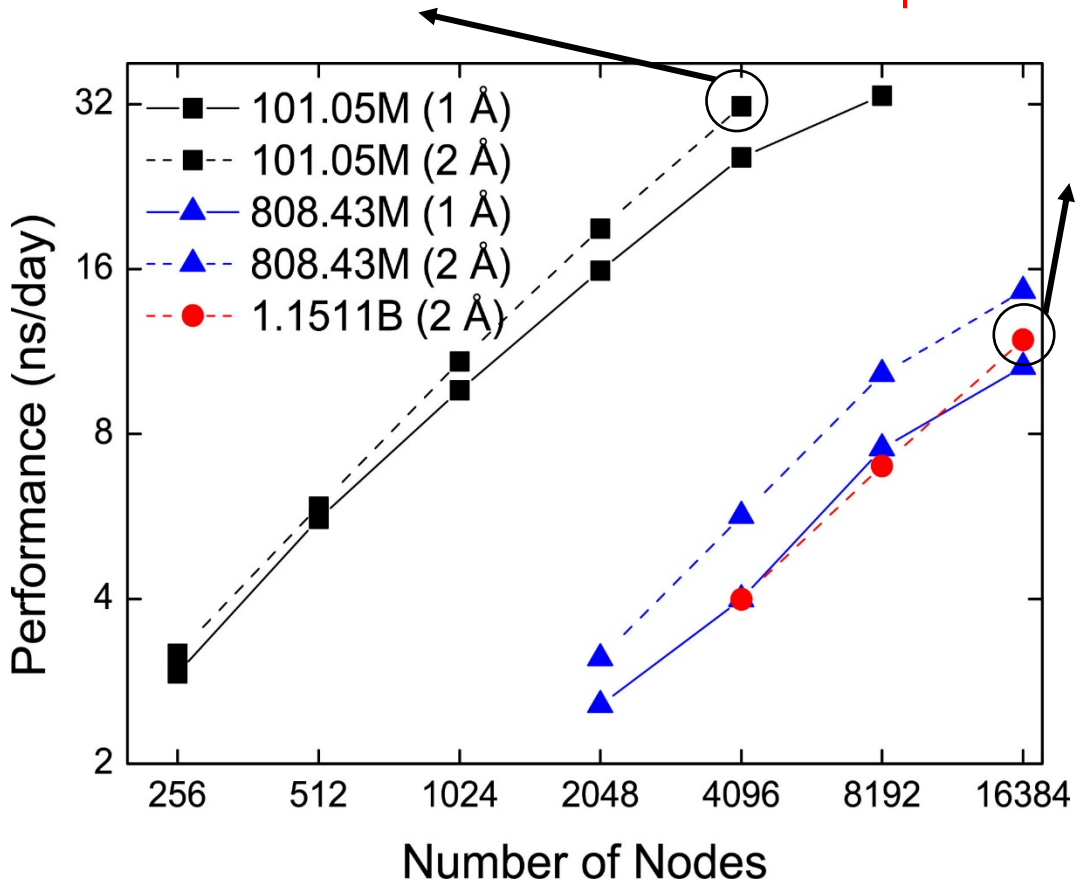
O $P(t) = \frac{1}{3V}(2K_{\text{half}}(t) + Vir(t))$

Ref: JCP, 148, 164109 (2018)
Ref: JCTC, 15, 84 (2019)

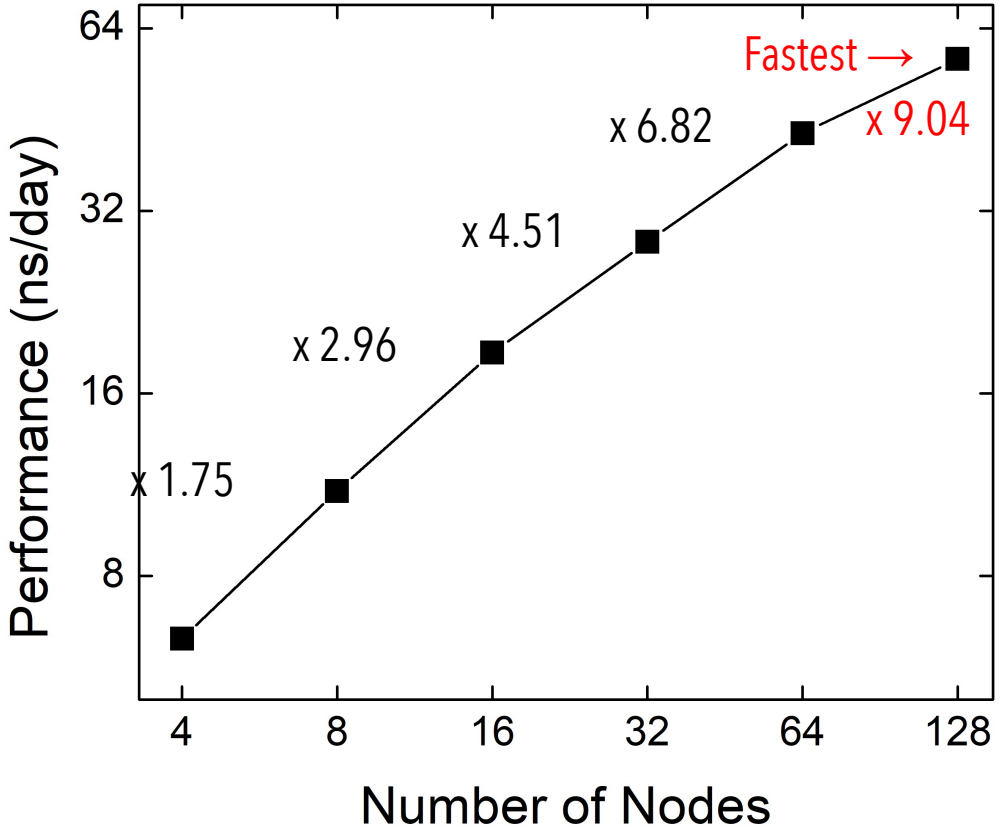
Strong Scaling of GENESIS 2.0beta on Fugaku

11.9 ns/day
(more than twice better performance than NAMD)

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



Spike-protein in solution on Fugaku



Performance between Fugaku and Oakforest-PACS

Spike-protein in solution on Oakforest-PACS

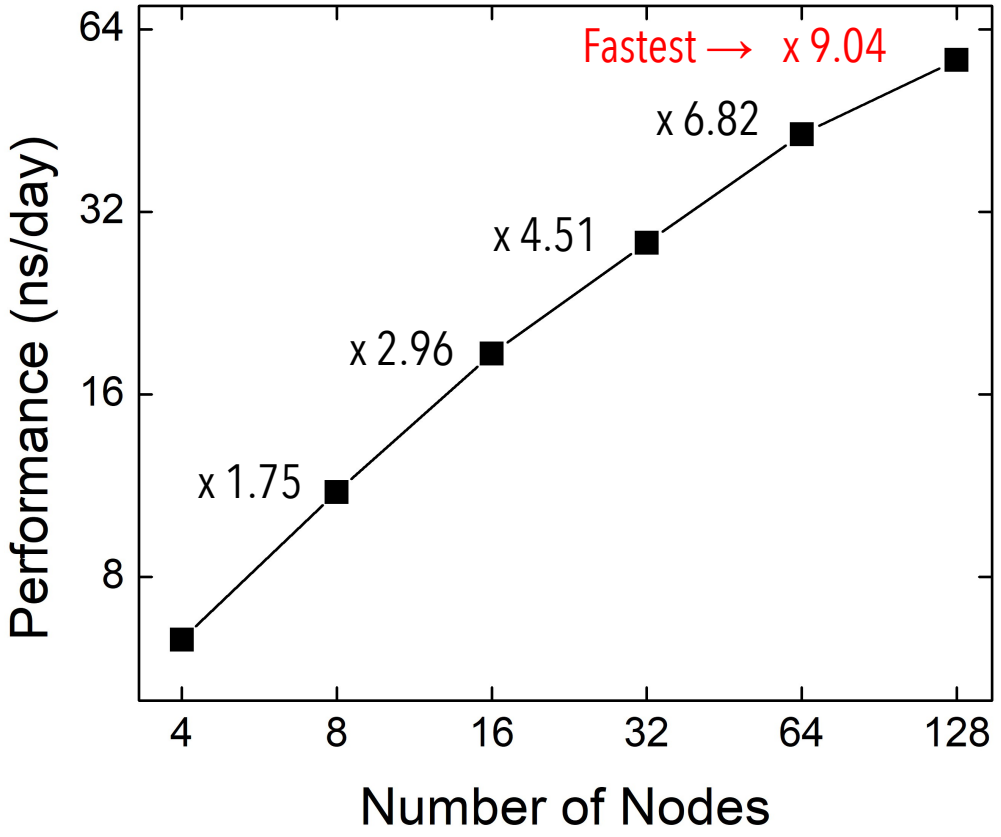
Nodes	Processes	Threads	Spline order	# of grids	Performance (ns/day)	PME scheme
8	512	4	6	128	4.59	OPT_1D
16	512	8	6	128	7.76	OPT_1D
32	512	16	6	128	11.08	OPT_1D

Nodes	Processes	Threads	Spline order	# of grids	Performance (ns/day)	PME scheme
8	512	4	4	192	4.52	NOOPT_1D
16	512	8	4	192	7.62	NOOPT_1D
32	512	16	4	192	11.07	NOOPT_1D

Domains_xyz = (8,8,8)
Ncell_xyz = (24,24,24)

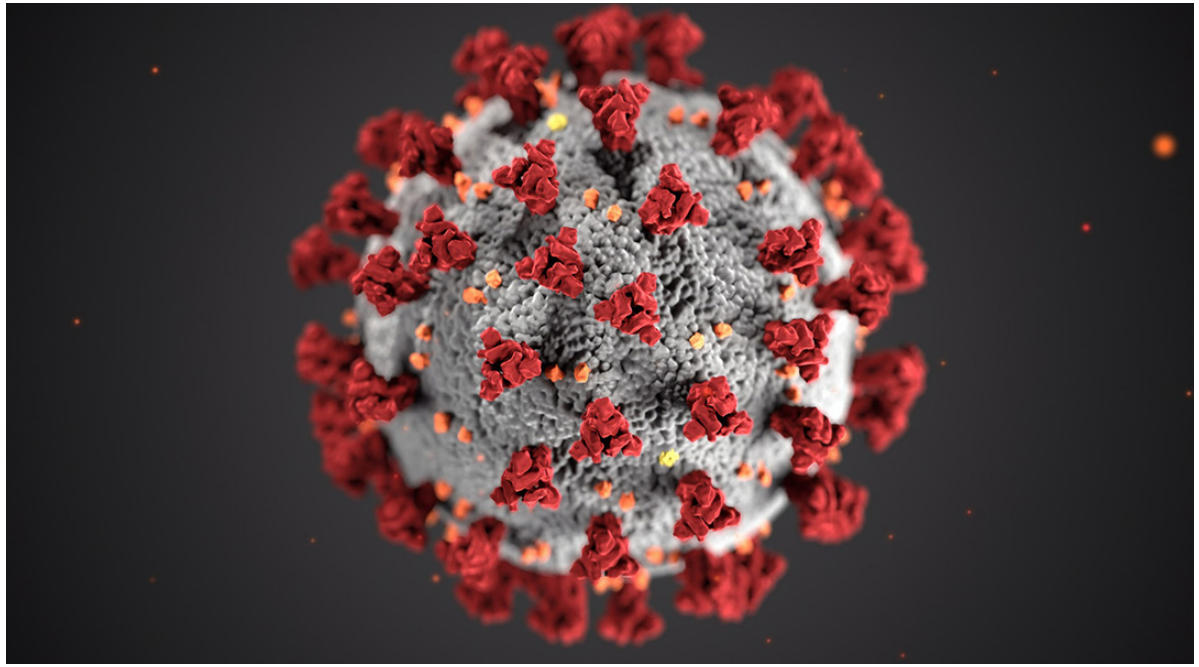
64 core is used for each node.

Spike-protein in solution on Fugaku



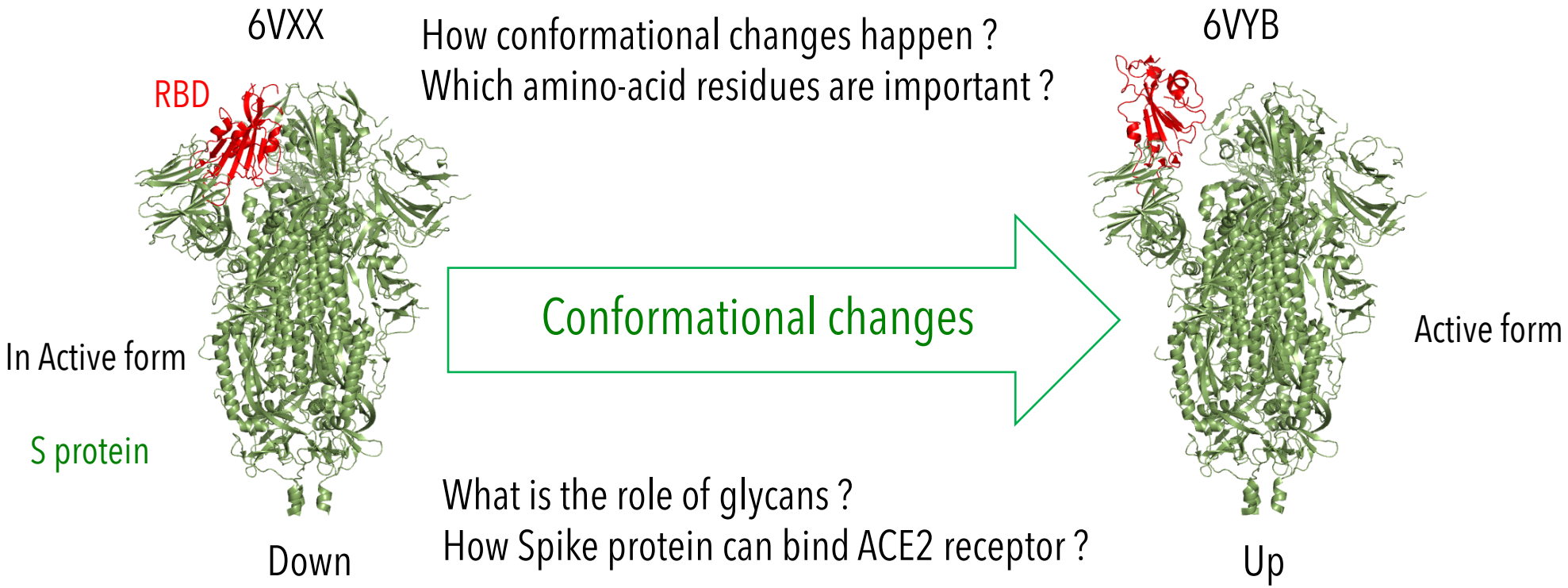
Entry of SARS-CoV-2 into Human Cell

A structural model of SARS-CoV-2



- A key structural features of SARS-CoV-2 ⇨ Many spikes on the surface
- Binding between Spike protein and a receptor in human cell

Purpose of our simulations



Two purposes:

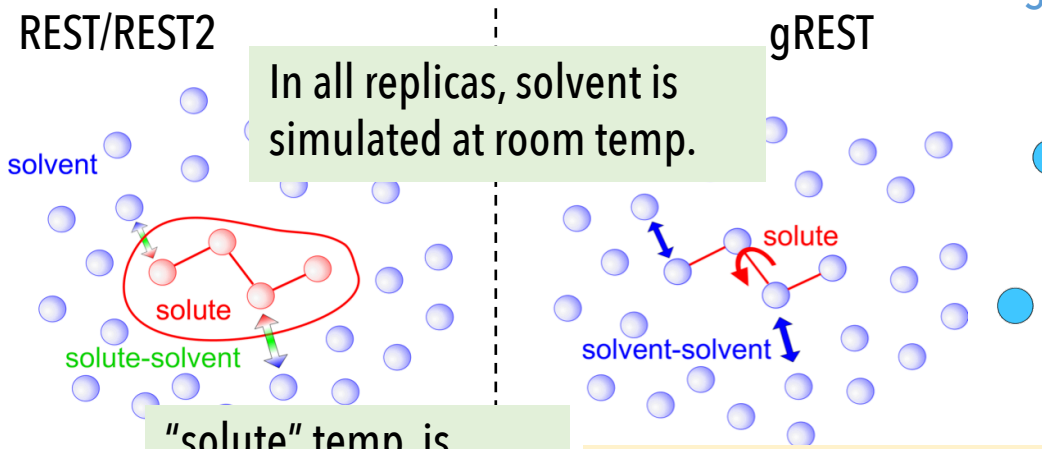
- Prediction of conformational changes between the inactive and active forms.
- Elucidation of functional role of glycans in spike protein

generalized REST (gREST) method (theory)

Replica-Exchange with Solute Tempering (REST2):

T. Terakawa et al., *J. Comput. Chem.* **32**, 1228-1234 (2011),
S. L. C. Moors et al., *J. Chem. Theor. Comput.* **7**, 231-237 (2011),
L. Wang et al., *J. Phys. Chem. B* **115**, 9431-9438 (2011).

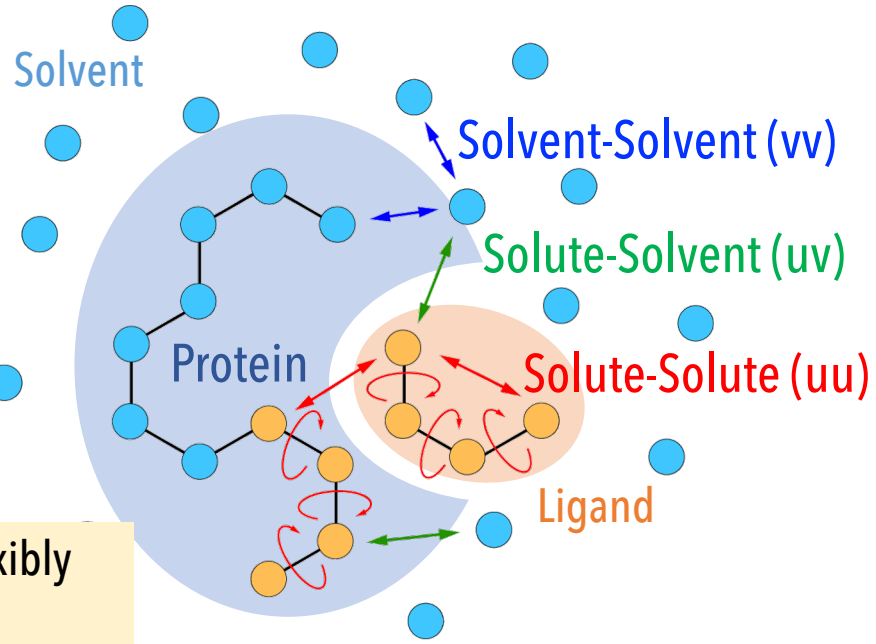
$$E_{\mathrm{REST2}} = \frac{\beta_m}{\beta_0} E_{uu} + \left(\frac{\beta_m}{\beta_0}\right)^{1/2} E_{uv} + E_{vv}$$



generalized REST (gREST):

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

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



Solute = Ligand + Sidechains at the Binding Sites

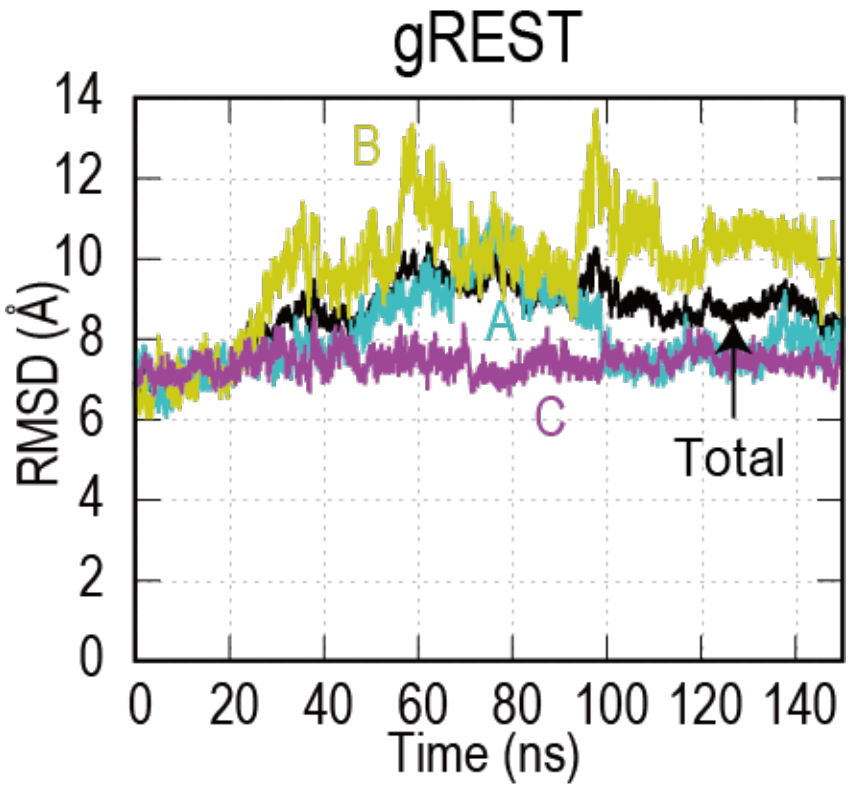
Conformational Changes of S-protein in solution

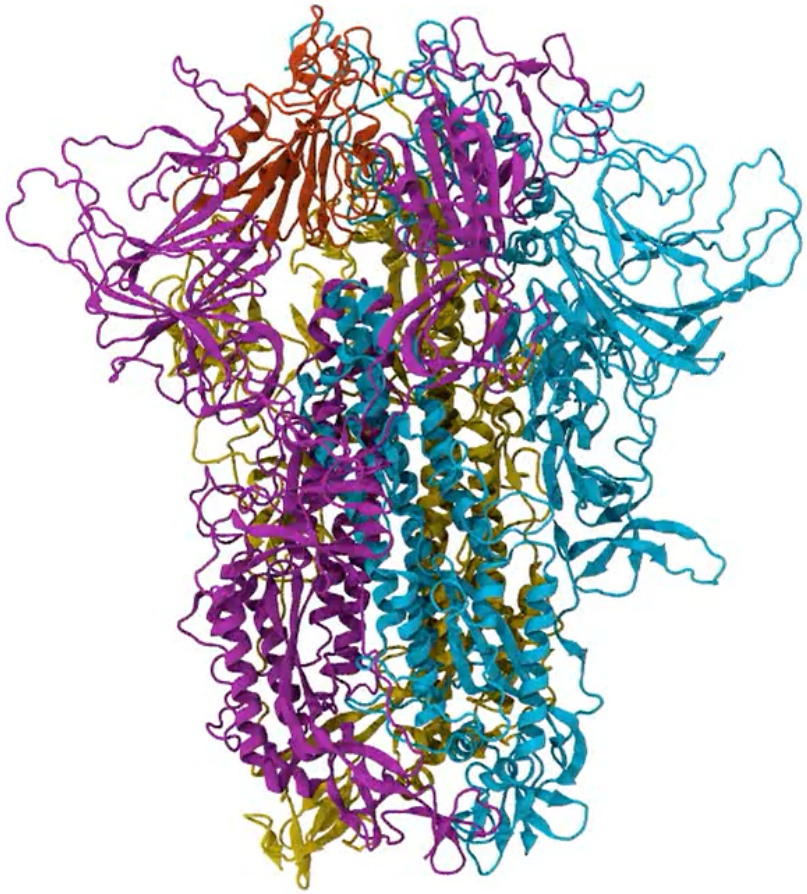
gREST simulations on S-protein

- We used only Down form (6VXX) and could predict similar structures of Up form.

⇨ Only chain B underwent the conformational change.

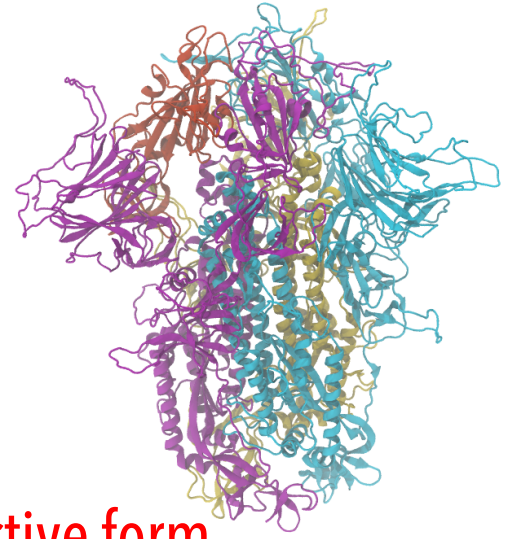
(N-glycans were not added in the gREST simulations.)





Summary of gREST simulations on Fugaku

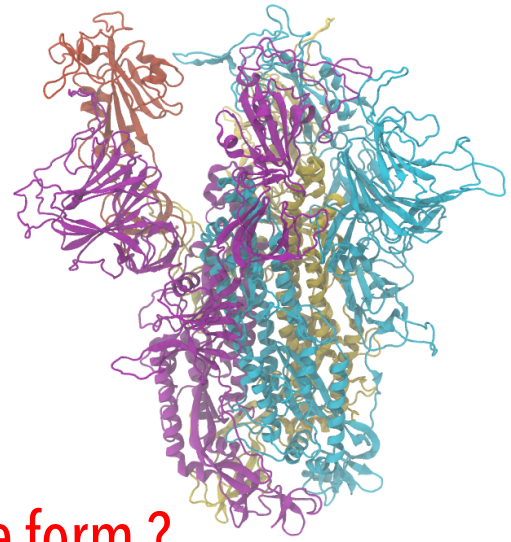
Down (Initial Structure)



Inactive form

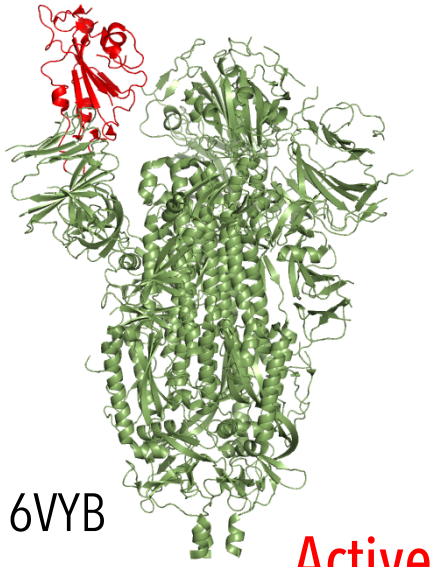
gREST

Open (MD Snapshot)



Active form ?

Up (cryo-EM)



6VYB

Active form

- We predicted large conformational changes from Down to Open forms using Fugaku just for 3 weeks.
- The open form predicted using gREST is similar to Up form by cryo-EM.
- We would like to examine molecular mechanisms for the conformational changes.

Summary

- We have developed MD software GENESIS and optimized GENESIS into K, Fugaku, Oakforest, and other supercomputers.
- Conformational changes from Down to Up were predicted using gREST simulations on Fugaku.
- N-glycans on the surface of S-protein stabilize either Up or Down conformation, forming different contact and hydrogen-bonded pairs.
- We hope to contribute the development of vaccines or anti-virus drug.

Acknowledgements

- RIKEN CPR（理研開拓研究本部）

Dr. Hisham Dokainish

Dr. Takaharu Mori

- RIKEN R-CCS（理研計算科学研究センター）

Dr. Chigusa Kobayashi

- Computational Resources

Fugaku, Oakforest-PACS

- NIBIOHN （医薬基盤研究所）

Dr. Suyong Re

- Computational Resources

Fugaku, Oakforest-PACS

- **All works are organized by Dr. Sugita (RIKEN CPR, BDR, R-CCS)**