GENESIS
Generalized-ensemble simulation system

マルチスケール分子動力学ソフトウェアGENESISの並列化

Supercomputing Japan 2024
2024/03/12 タワーホール船堀

Yuji Sugita

RIKEN Center for Computational Science
RIKEN Center for Biosystems Dynamics Research
RIKEN Cluster for Pioneering Research
RIKEN iTHEMS

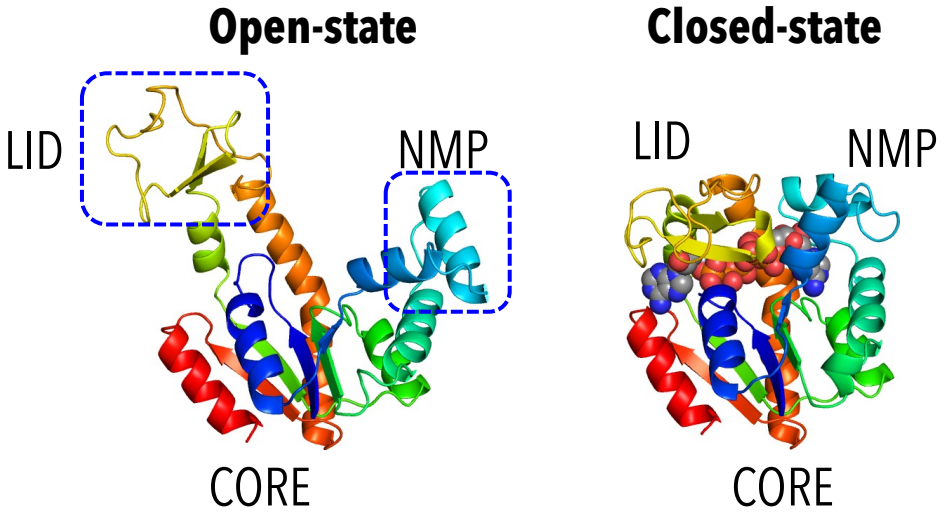
Protein Dynamics and Functions

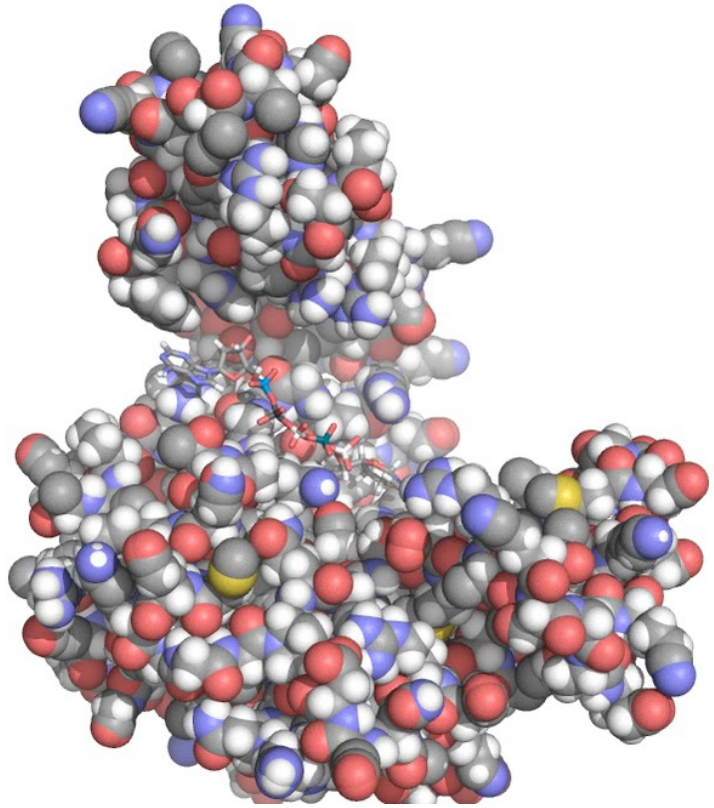
- Protein structures are determined by X-ray crystallography, NMR , and cryo-EM at atomic resolutions.
- Proteins are flexible for their functions in a living cell .

Enzyme reaction by Adenylate Kinase (Adk)

$$ATP + AMP \rightleftharpoons 2ADP$$

Conformational changes of Adk





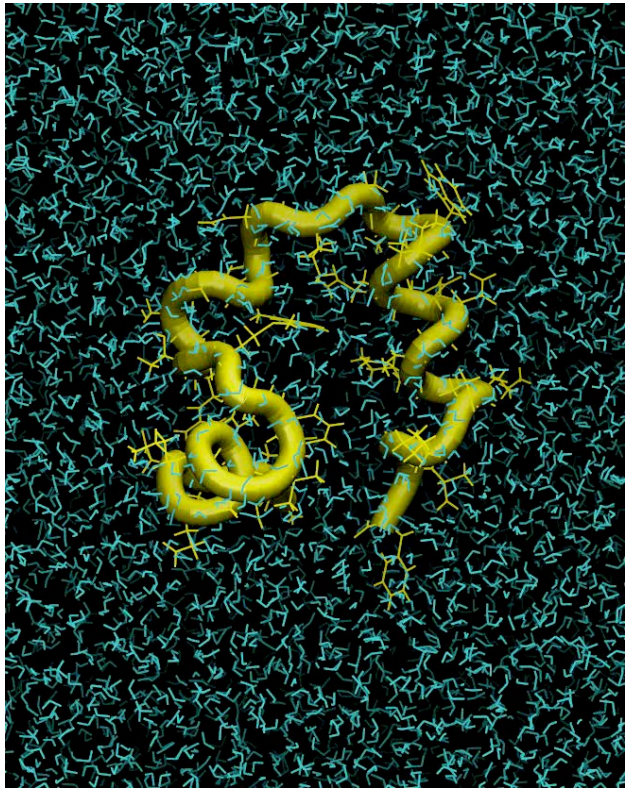
Y. Matsunaga, et al., *PLoS Comput. Biol.* **8**, e1002555 (2012).

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.

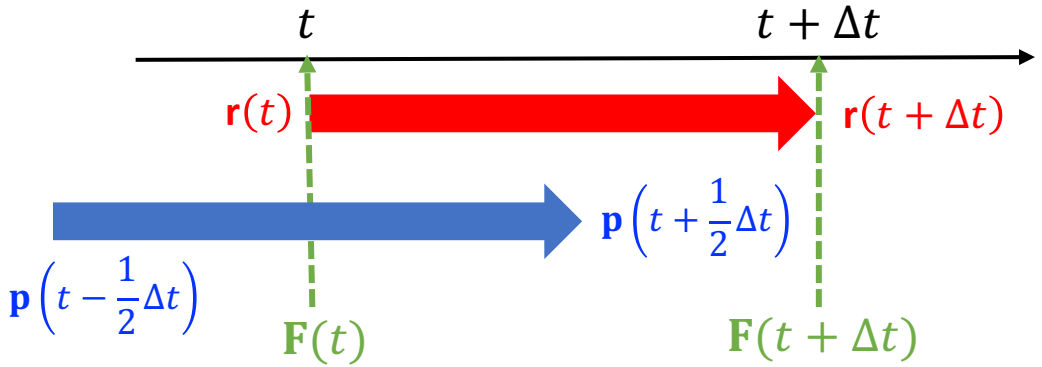
$$F = ma$$

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



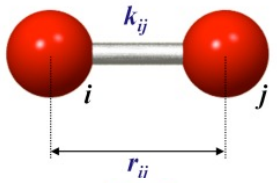
$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{p}_i\left(t+\frac{1}{2}\Delta t\right)\Delta t}{m_i}$$

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



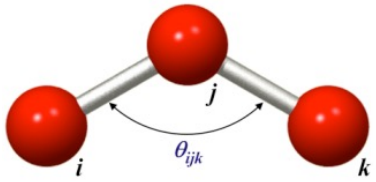
Molecular Force Field Function

$$V(\mathbf{r}) = \sum_{bonds} k_b \left(r_{i,i+1} - r_0 \right)^2$$



Bond term

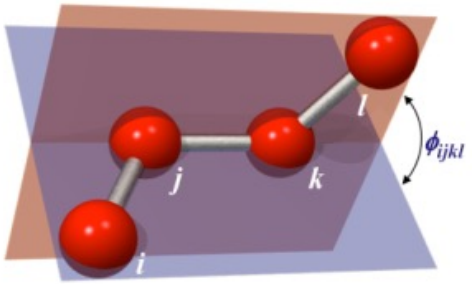
$$+ \sum_{angles} k_a \left(\theta_{i,i+1,i+2} - \theta_0 \right)^2$$



Angle term

$$+ \sum_{torsions} k_t \left[1 + \cos\left(n\phi_{i,i+1,i+2,i+3} - \omega \right) \right]$$

taken from PumMa HP



Dihedral term

$$+ \sum_{i} \sum_{j>i} \varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]$$

vdW term
(solved with cutoff approximation)

$$+ \sum_{i} \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Electrostatic term

$O(N^2)$

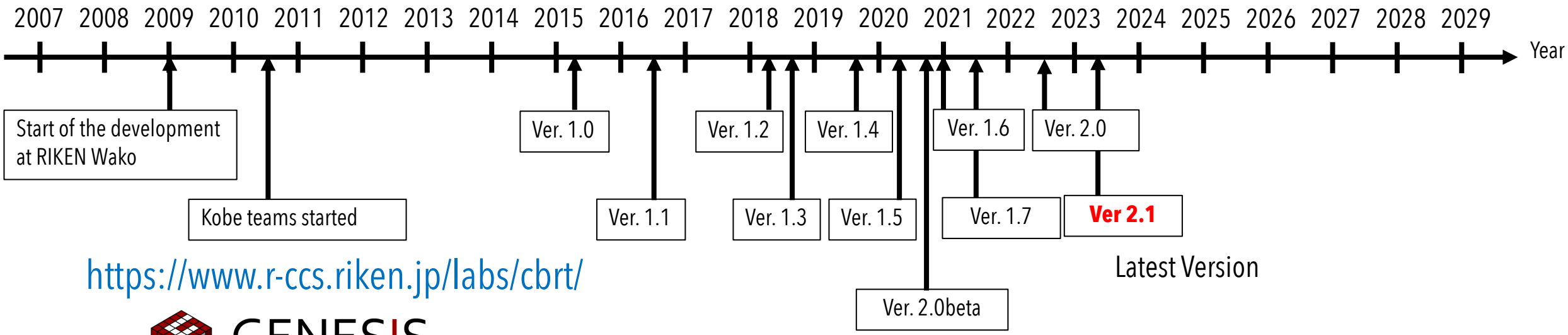
CHARMM, AMBER, OPLS, and other Force Fields are available for various molecules.

Jung, Mori et al., WIREs Comput. Mol. Science. (2015); Kobayashi, Jung et al., JCC. (2017)

Development of GENESIS

- We have developed GENESIS software since 2019.
- GENESIS can show good performances on K and Fugaku.
- GENESIS is freely available for everyone under LGPLv3 license.

Initial Developers

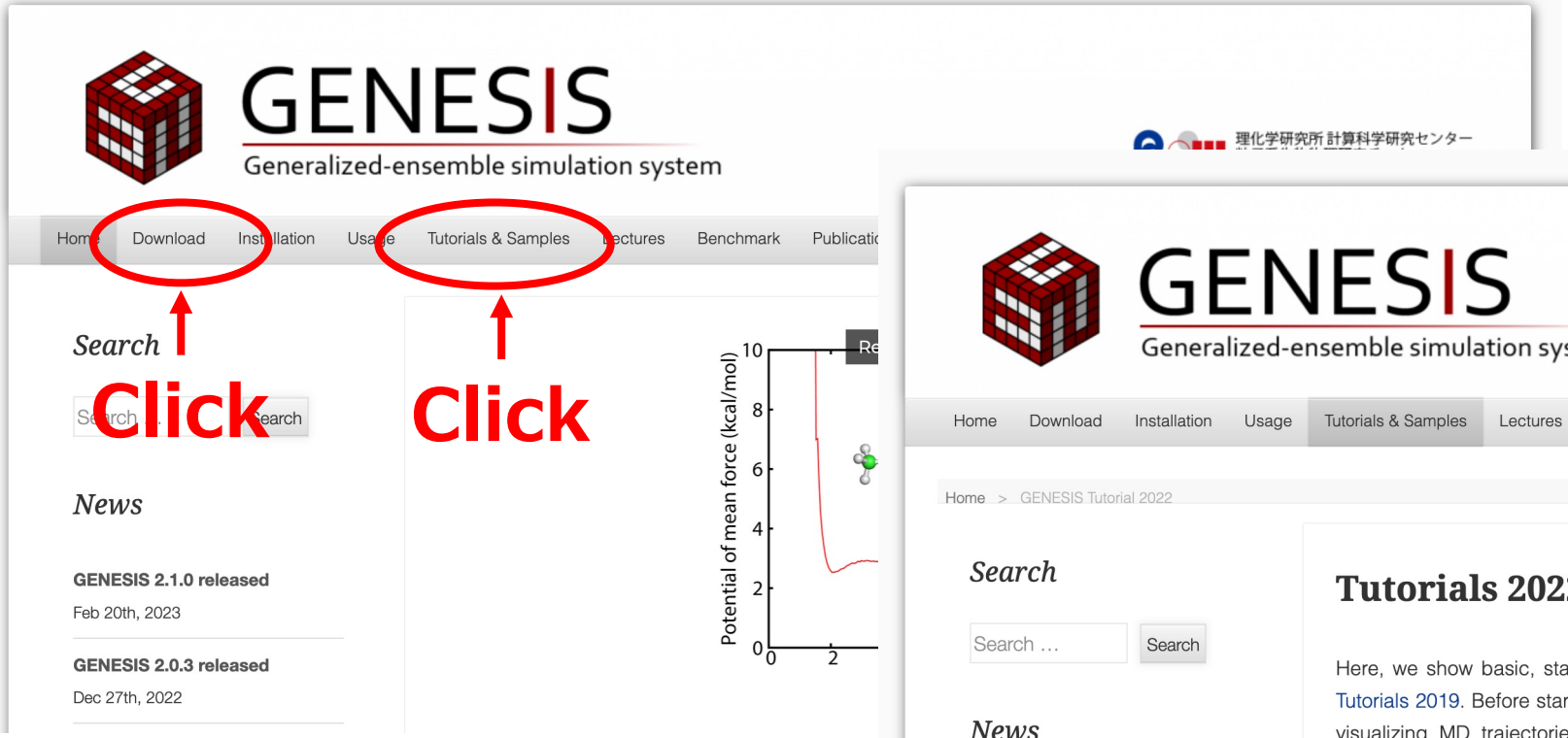


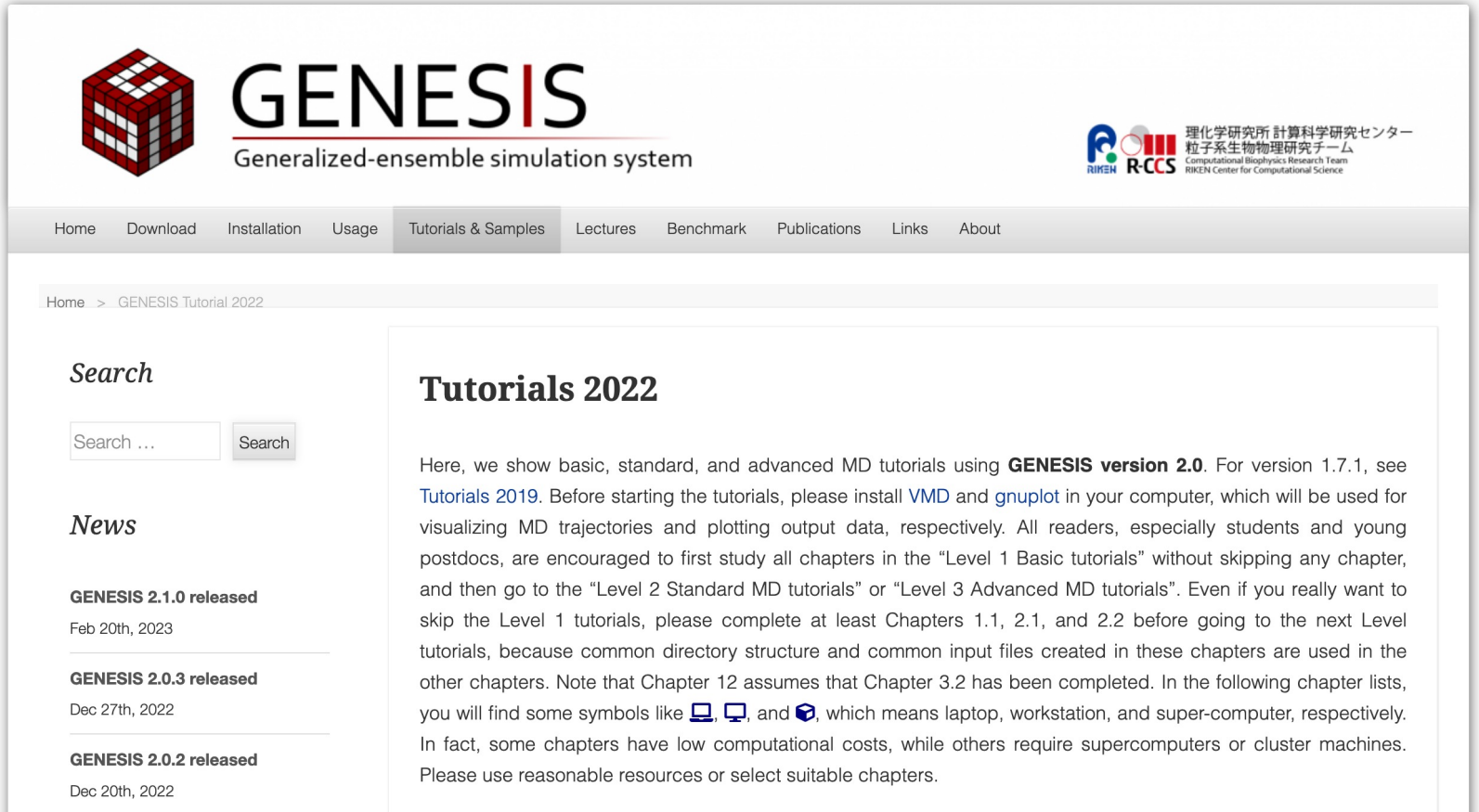
https://www.r-ccs.riken.jp/labs/cbrt/

GENESIS
Generalized-ensemble simulation system

GENESIS is freely available

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





How can we accelerate MD simulations to understand the slow conformational changes of proteins (if we can use supercomputers) ?

RIKEN Center for Computational Science

The Supercomputer K

Node	CPU	SPARC64™ VIIIfx 2GHz
	Performance	128 GF (16 GF x 8 cores)
	Memory	16 GB
Number of Nodes per Rack		864
Total Number of Nodes		82,944
Network		Tofu Interconnect (6D Mesh/Torus)
Peak Performance		10.62 PF
Total Memory		1.26 PB
File IO		Fujitsu Exabyte File System (FEFS)
Storage		30 PB

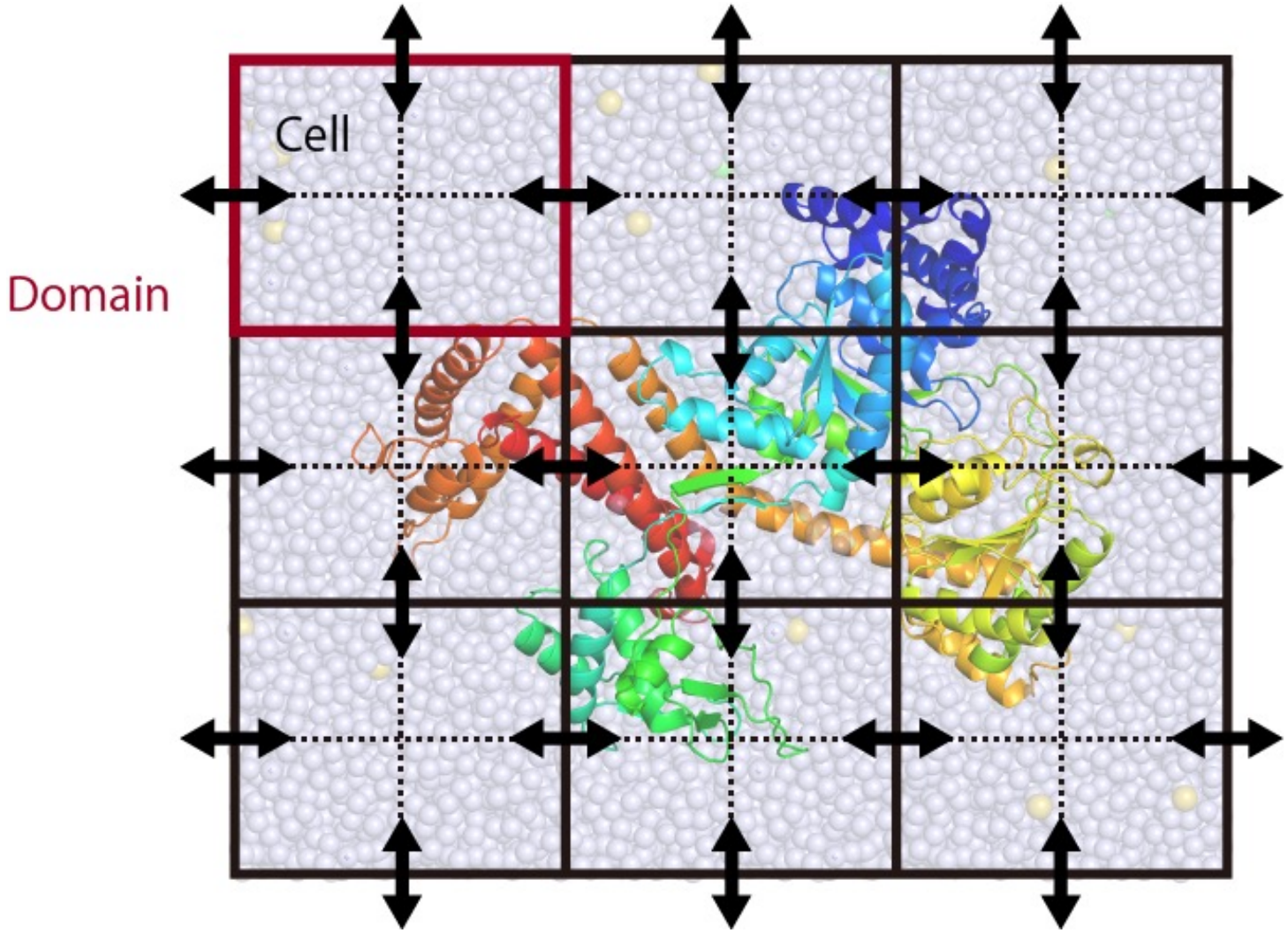
K computer

Top500	No.1 for 2 consecutive times since Jun. 2011
Graph500	No.1 in June 2014 & for 6 consecutive times since Jul. 2015
HPCG	No.2 for 4 consecutive times since Nov. 2014 No.1 for 3 consecutive times since Nov. 2016

The K computer strikes a balance between performance in calculation, memory, and communication.

Shared use of the K computer has ended on August 16, 2019.

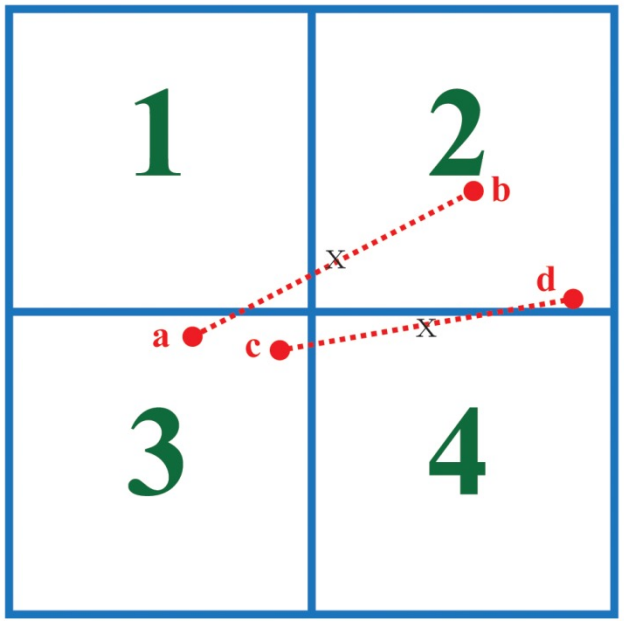
Domain Decomposition Scheme



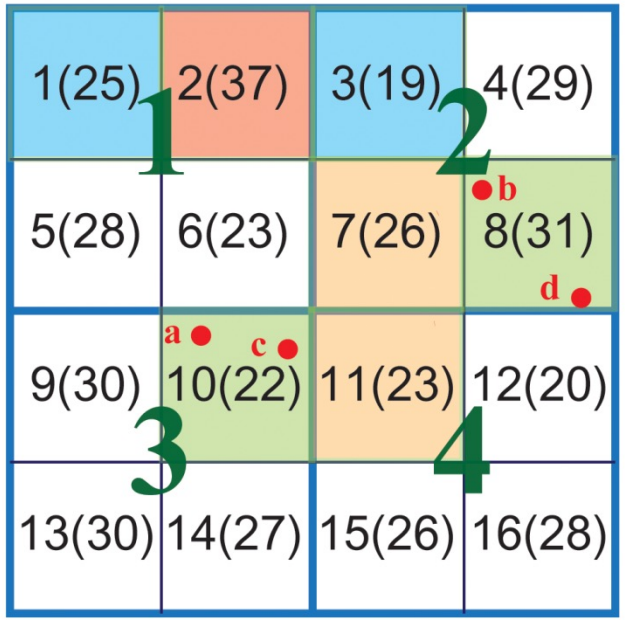
Divide a whole space into subdomains and cells → Greatly reduce the memory

Midpoint Cell Method

(a)



(b)



Midpoint Method:

Interaction is computed at the cell where the midpoint of two interacting particles is located.

(O) Low communication cost

(X) Need to decide the midpoint every time step.

Midpoint Cell Method:

Interaction is computed at the midpoint cell for each cell pair.

(O) Low communication cost

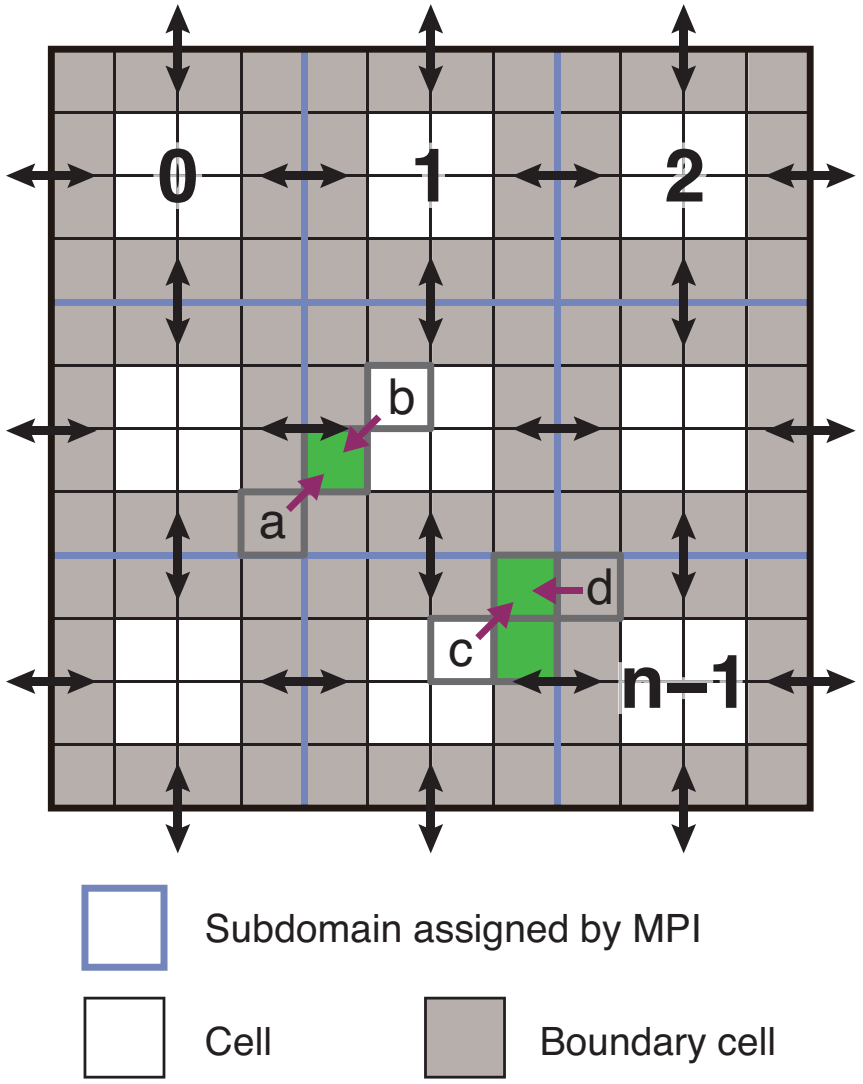
(O) Midpoint cell is decided only once.

Kevin J. Bowers et al. , ***J. Chem. Phys.*** 124, 184109 (2006).

J. Jung, T. Mori and Y. Sugita, ***J. Comp. Chem.*** 35, 1064 (2014)

Subdomains and cells

(a)



1. The basic domain decomposition scheme is the midpoint cell method.

2. A whole system is divided into subdomains and cells.

3. Information of the particles residing at each subdomain (plus surrounding cells) is stored at each MPI process.

4. Each subdomain is further divided into cells for OpenMP thread parallelization.

This method is suitable to Hybrid parallelization.

Intracellular Environments

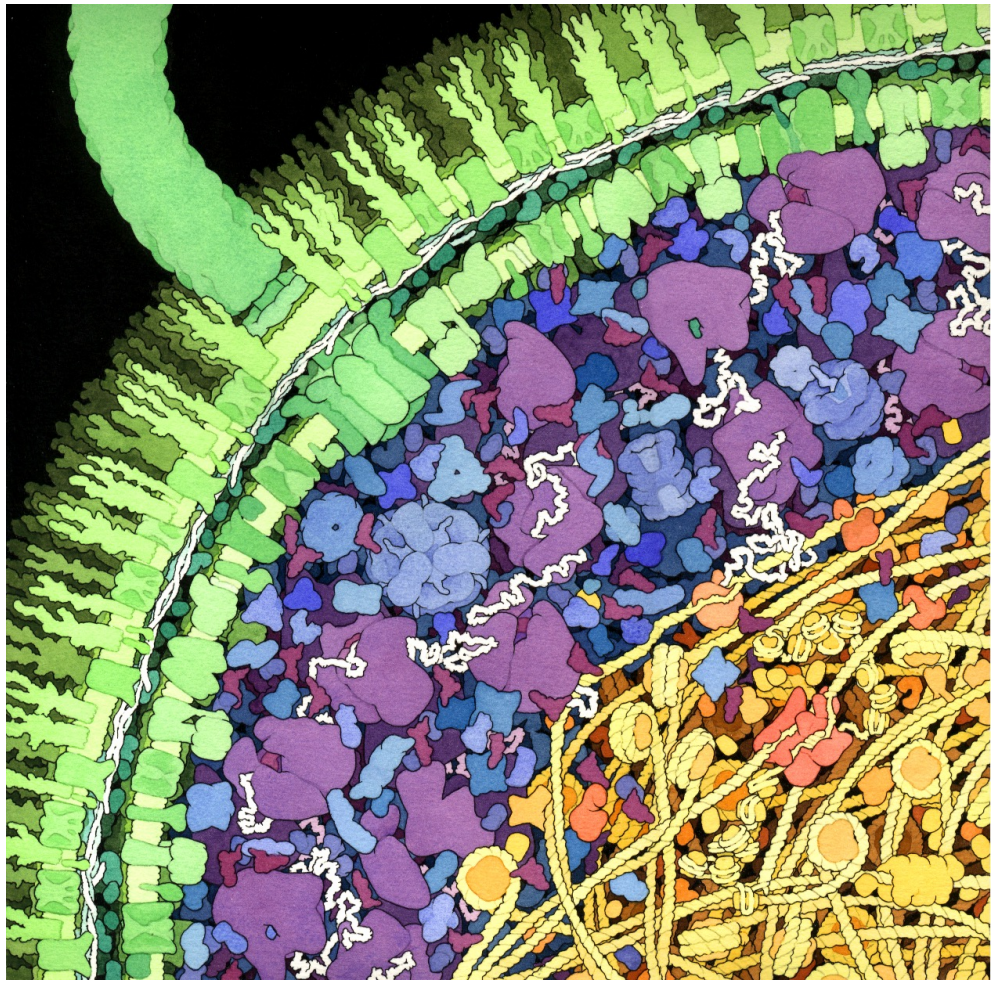


Illustration by Goodsell for E. coli cell

- Concentration of proteins and RNA in the E. coli cell is 300-400mg/ml.
- 20-40% of the volume inside of the cell is occupied by macromolecules.
- This condition has been called as **macromolecular crowding**, which can affect structure, dynamics, and functions of macromolecules in the cell.

Zimmerman and Minton, *Ann. Rev. Biophys. Biomol. Struct.* (1993) 22: 27-65.

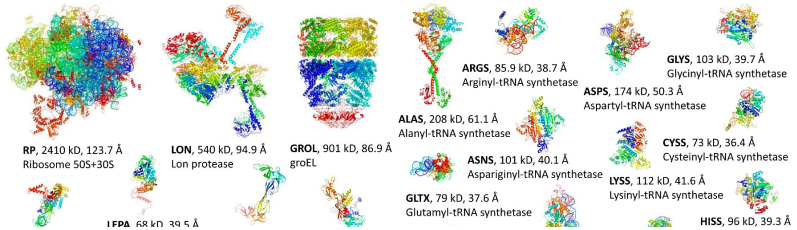
Zhou, H.X., Rivas, G., Minton, A.G., *Ann. Rev. Biophys.* (2008) 37, 375-397.

Go to see more arts in his homepage: https://ccsb.scripps.edu/goodsell/

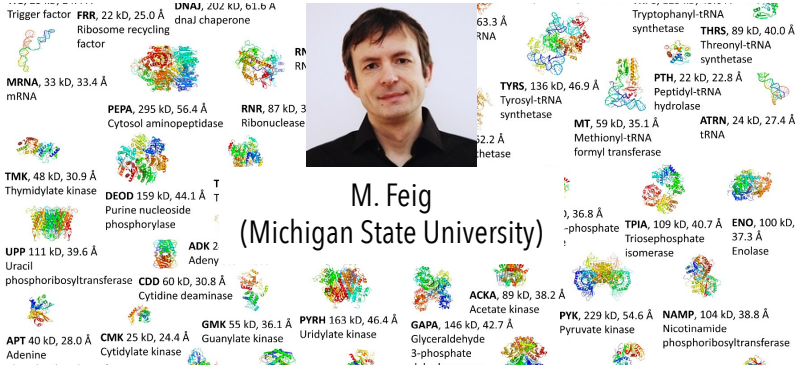
Feig, Harada, Mori, Yu, Takahashi, Sugita : *J. Mol. Graph. Model* **58**, 1-9 (2015).

Atomistic Modeling of the Cytoplasm

Homology models of protein structures

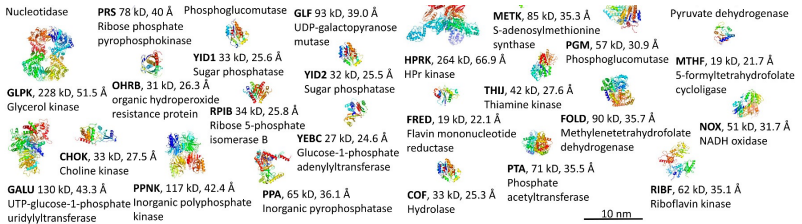


AlphaFold and RosettaFold were not available in 2015.

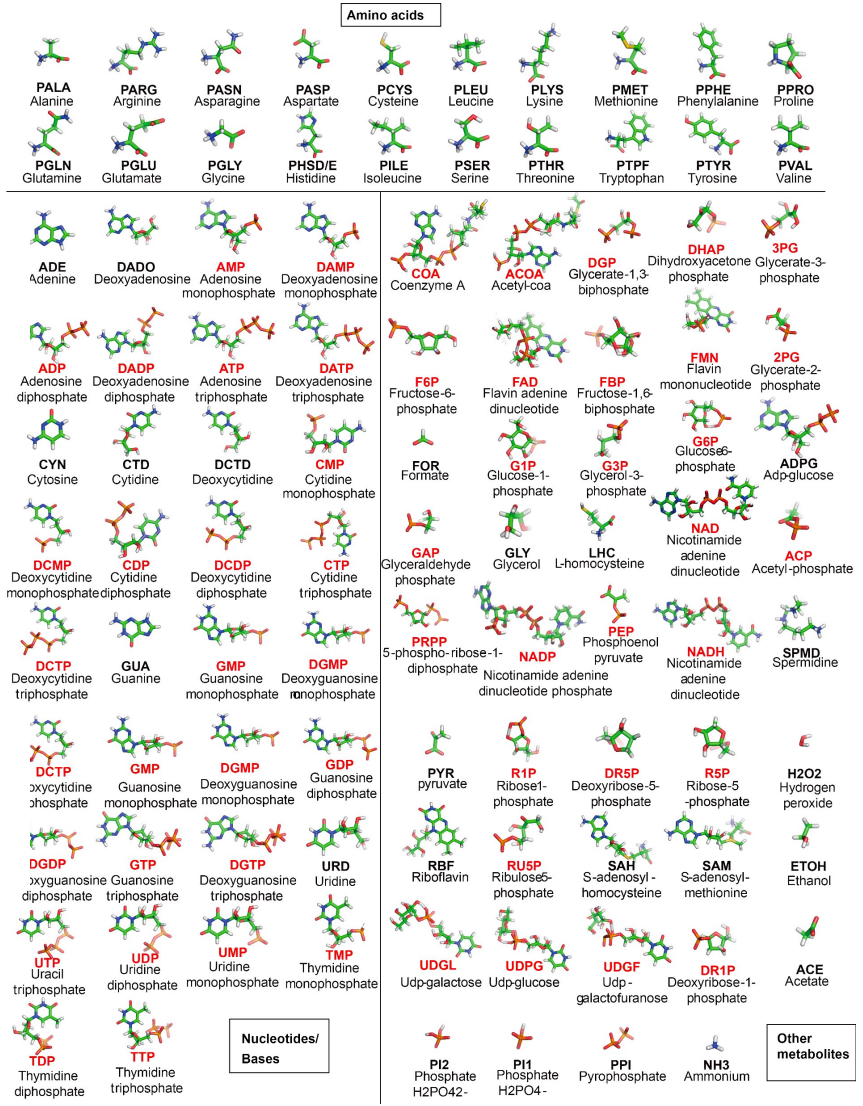


M. Feig
(Michigan State University)

Disordered regions were not included in the models.

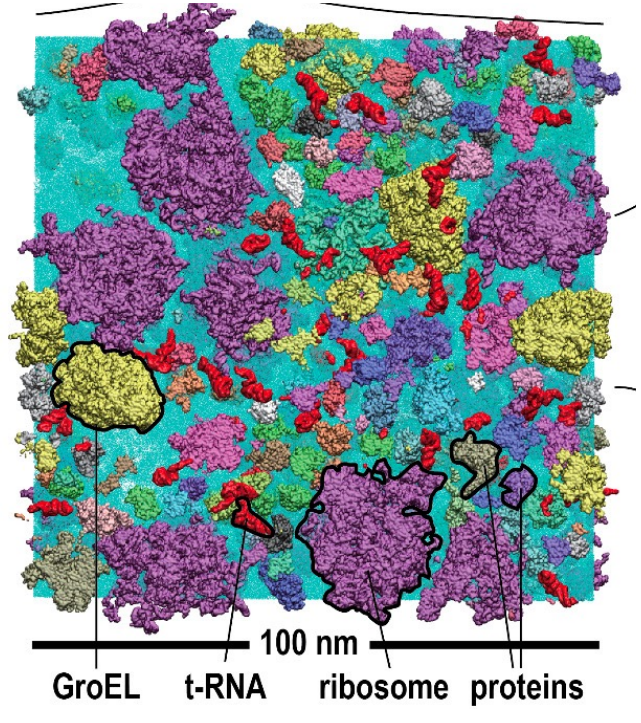


Metabolites in *M. Genitalium*



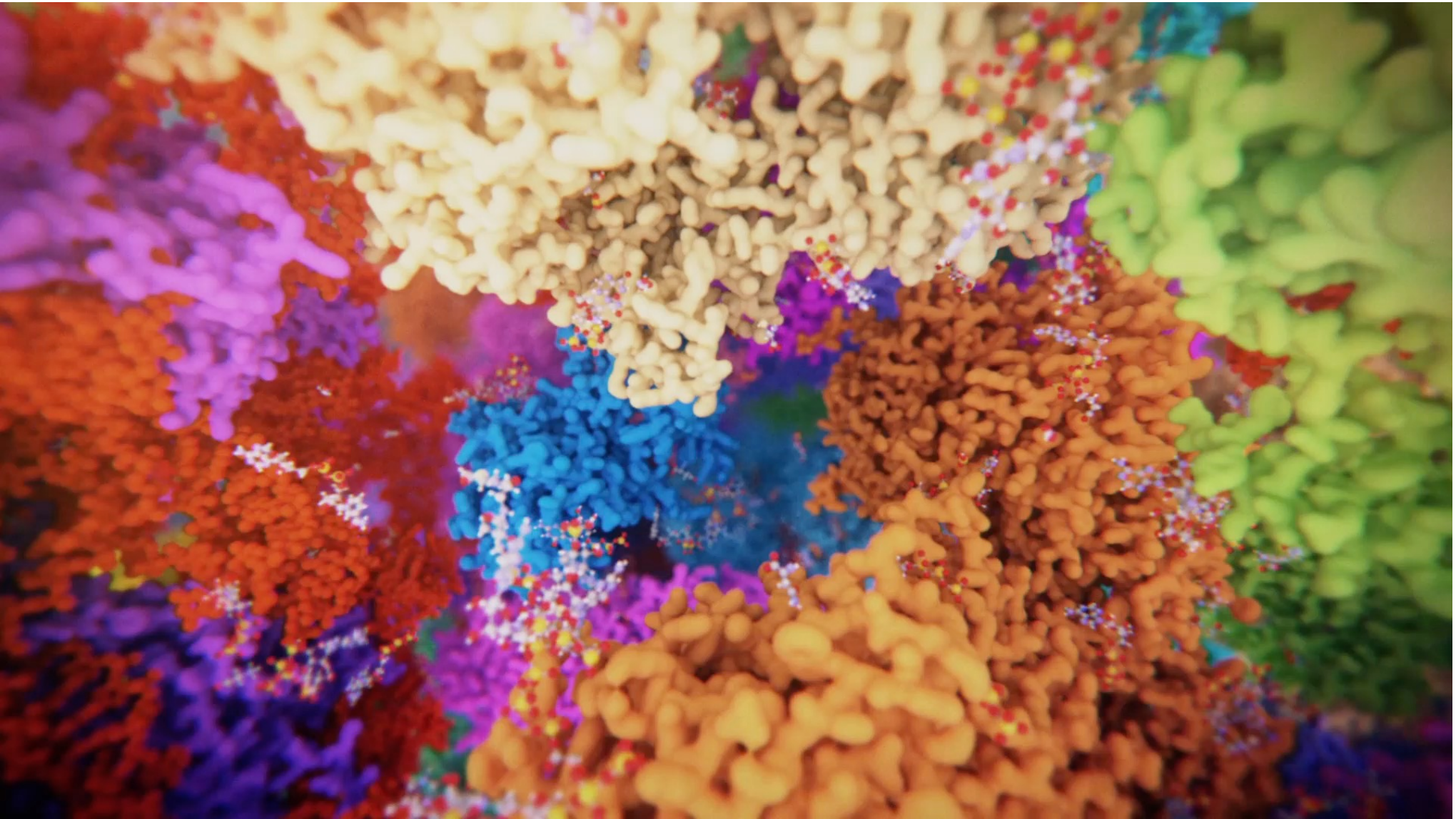
Fully hydrated models

- # of atoms: 103,708,785
- # of proteins: 1238
- # of RNAs: 284
- # of metabolites: 41,006
- # of ions: 214,000
- # of water: 26,263,505



Yu *et al. eLife* **5**, e19274 (2016).

The cytoplasm in *M. genitalium*



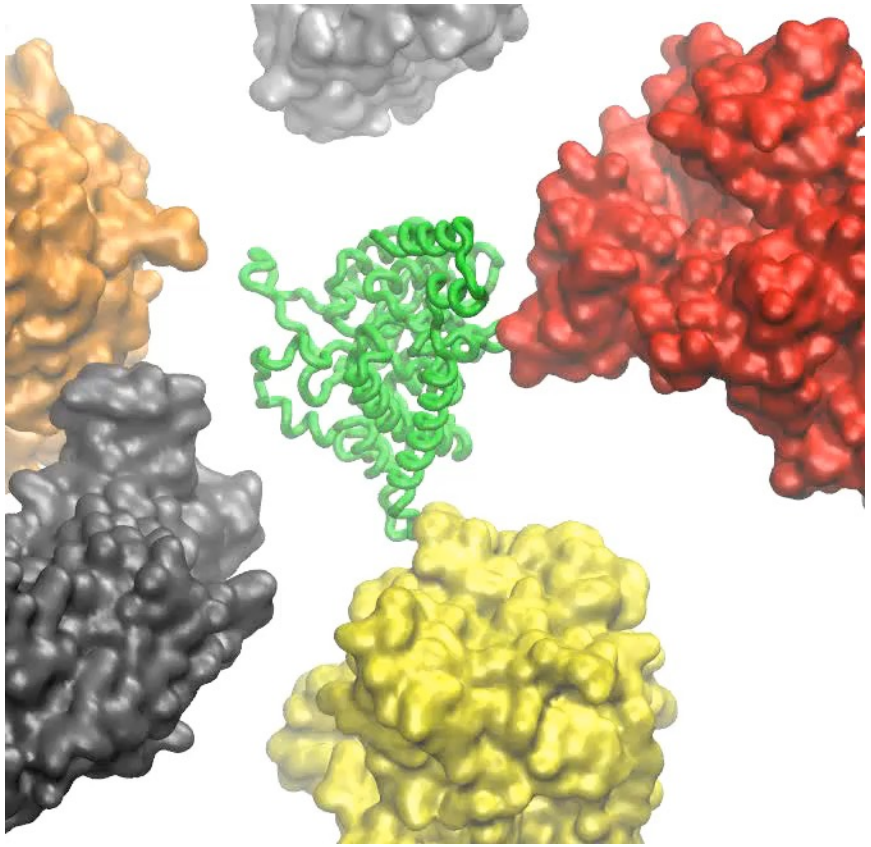
The first all-atom MD calculation of the bacterial cytoplasm (> 100 M atoms)

Isseki Yu (RIKEN
→ Maebashi Inst. Tech.)

M. Feig
(Michigan State University)

Destabilization in Crowded Environments

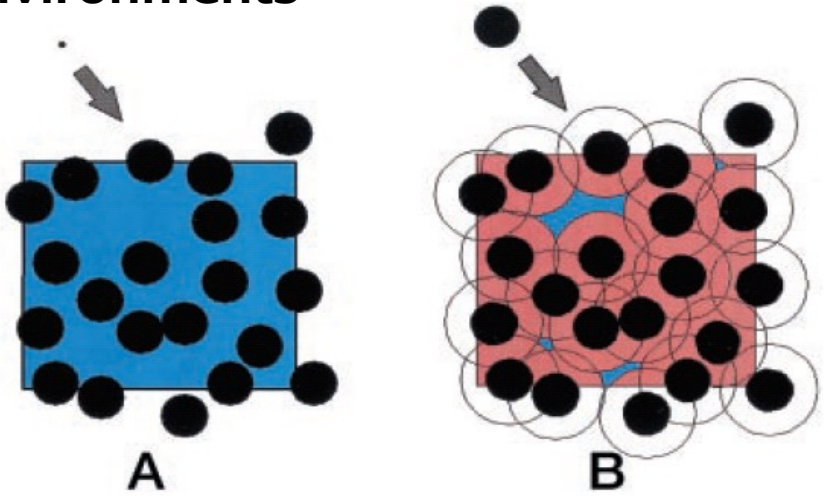
PDHA destabilization in Crowded Environments



This is inconsistent with the excluded volume effect.

Yu *et al. eLife* **5**, e19274 (2016).

Volume Exclusion Effect in crowded environments



Blue: available, Red: excluded (Minton, 2001)

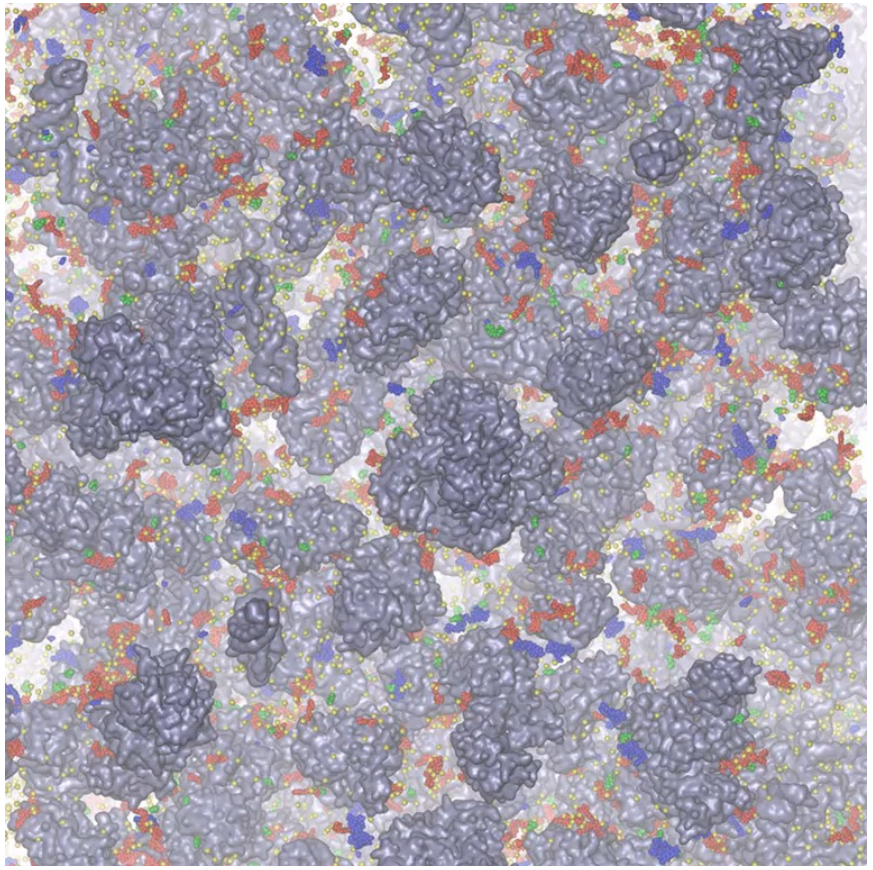
Free space for large molecules is very limited.

In the crowded condition, compact structures are preferred.

Zimmerman and Minton, *Ann. Rev. Biophys. Biomol. Struct.* (1993) 22: 27-65.

ATP controls solubility

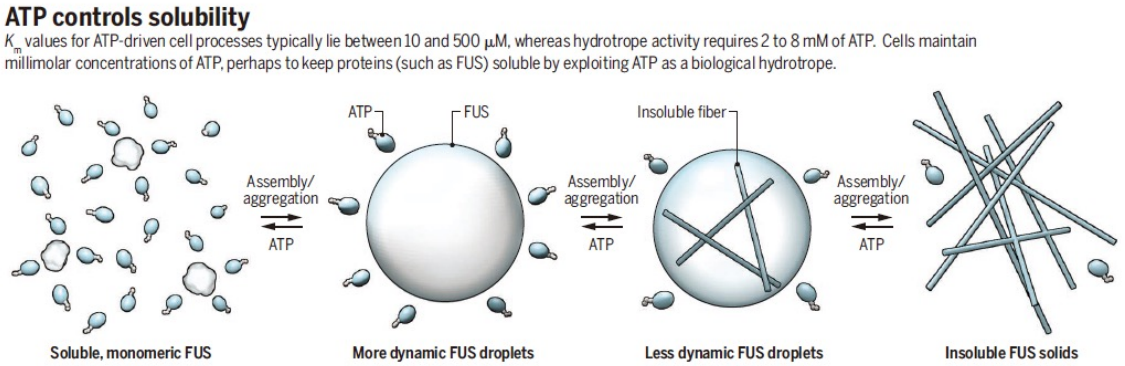
ATP distributions on the Protein Surfaces

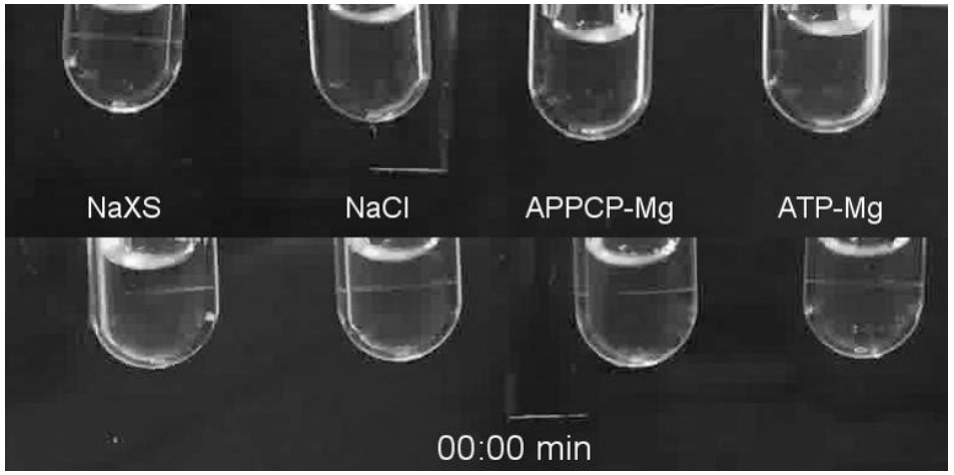


Interactions between proteins and ATPs are important.

Yu *et al. eLife* **5**, e19274 (2016).

A hidden function of ATP in the cells (hydrotrope in crowded conditions)



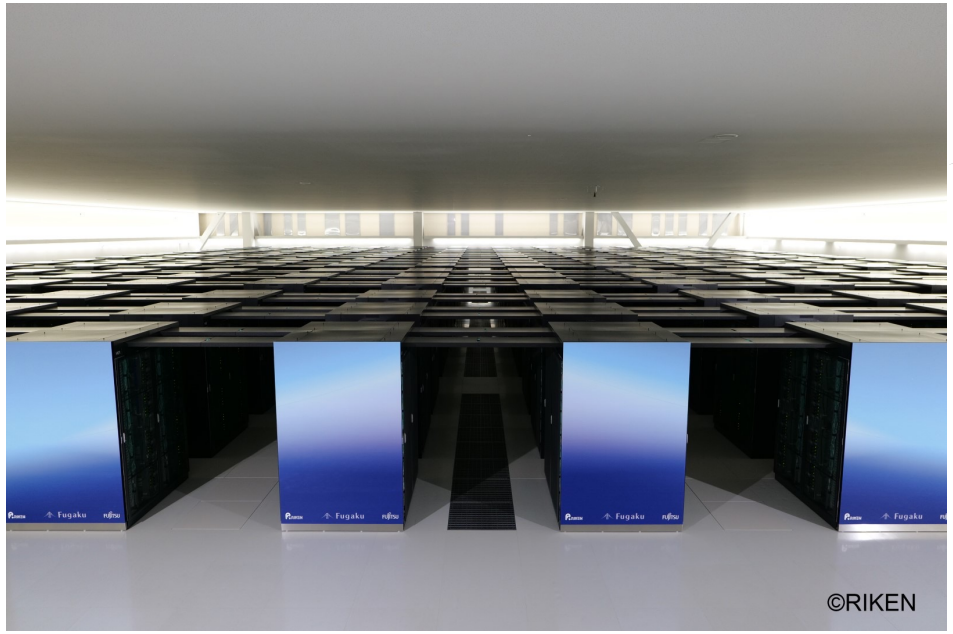


Patel et al. *Science* **356**, 753-756 (2017)

RIKEN Center for Computational Science

The Supercomputer Fugaku

Node	CPU	Armv8.2-A SVE 512bit
	Performance	3.072TF （DP） 6.144TF （SP） 48 (+2) core
	Memory	32 GB
Number of Nodes per Rack		384 node x 396 rack 192 node x 36 rack
Total Number of Nodes		158,976
Network		Tofu Interconnect D
Peak Performance		488 PF （DP） 977 PF （SP）
Total Memory		4.85 PB
File IO		LLIO
OS		Red Hat Enterprise Linux 8



富岳

Top500	No. 1 in June and Nov. 2020
Graph500	No. 1 in June and Nov. 2020
HPL-AI	No. 1 in June and Nov. 2020
HPCG	No. 1 in June and Nov. 2020

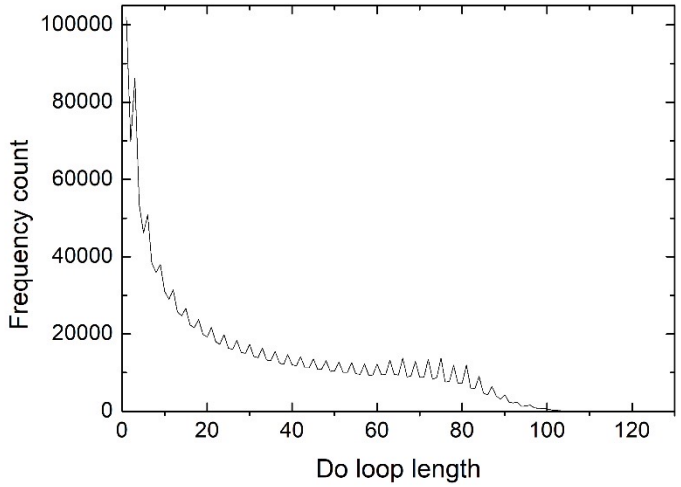
Fugaku is now the fourth fastest supercomputer in the world.

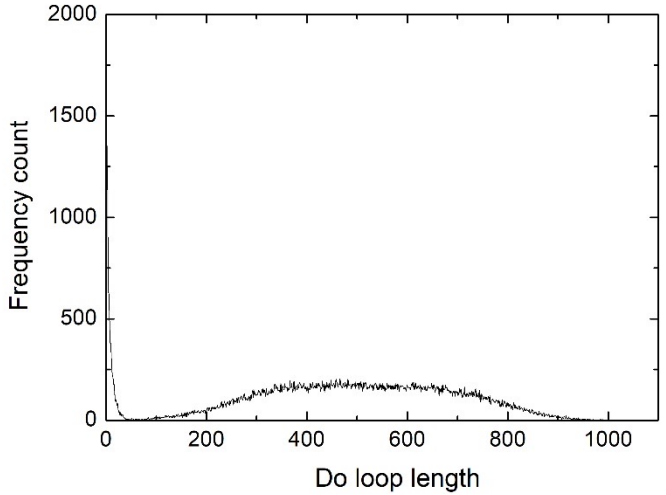
Optimization of GENESIS on Fugaku

New non-bonded interaction algorithm in GENESIS

```
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
```





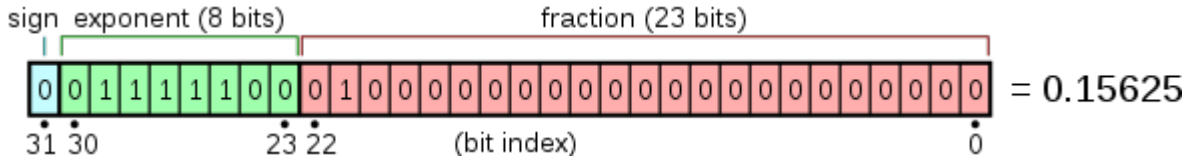
K computer

富岳

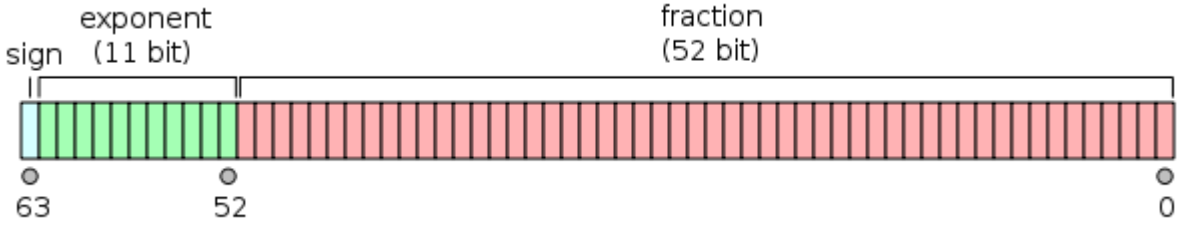
ApoA1 on one MPI processor

Single/Mixed/Double precision

- Single precision
  - 32 bit
  - Fast evaluation is possible by SIMD.
  - Less accurate than double precision



- Double precision
  - 64 bit
  - More accurate than single precision

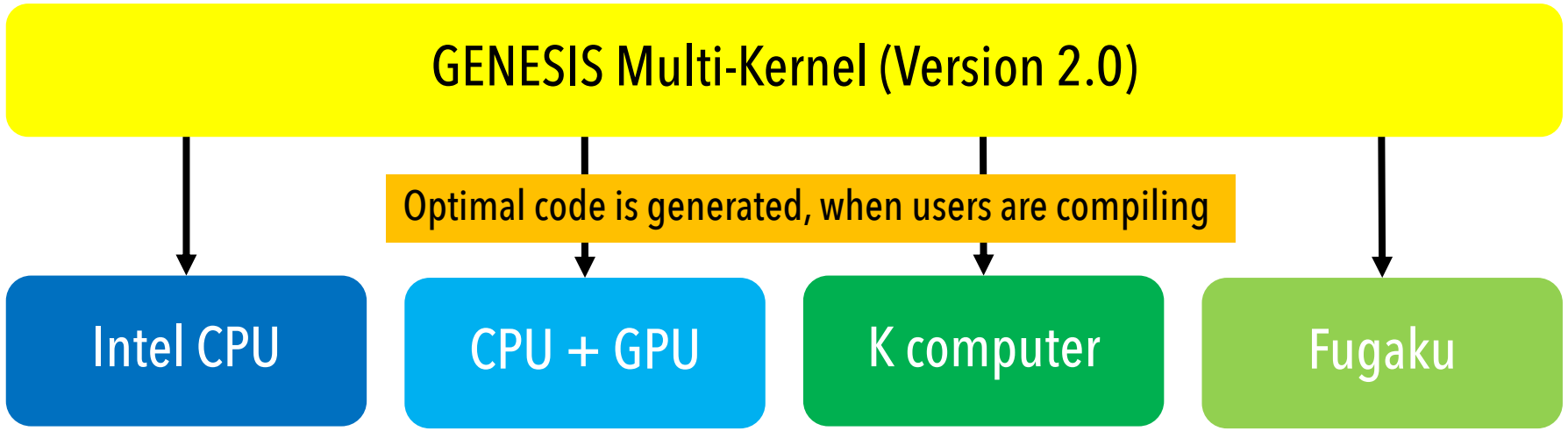


- We can define precision levels in GENESIS 2.0 according to our purposes
  - Single precision: All real numbers except energy/constraints are described by single precision
  - Mixed precision: Integration by double precision and force calculation by single precision
  - Double precision: All numbers are described by double precision

Picture is from wikipedia.

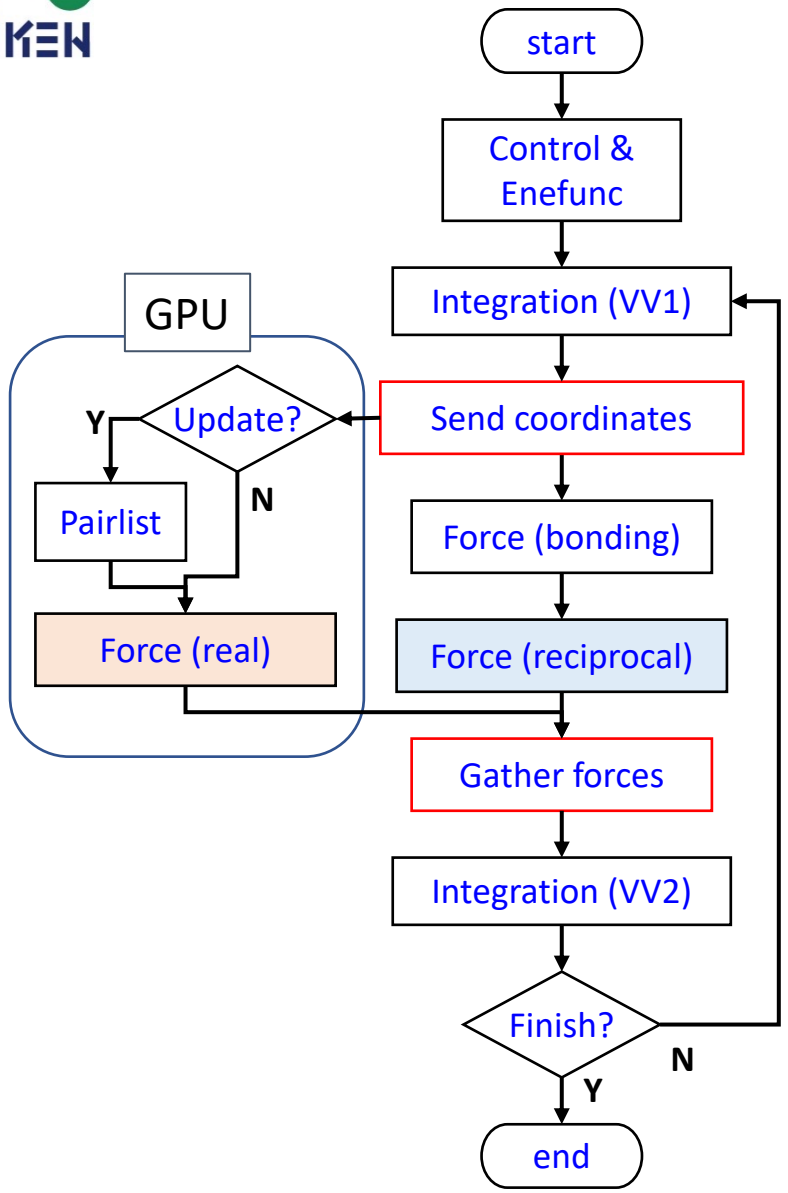
Multiple kernels for nonbonded energy calculations

User Friendly Multiple Kernels for Nonbonded Energy Calculations



- GENESIS 2.0 (the latest version) includes multiple kernels for nonbonded energy calculations.
- Users need to select a proper option when they compile GENESIS on each computational platform.
- Then, optimal code is generated automatically.
- Precision is also needed to select when compile GENESIS on each platform. We recommend single_precision for intel compiler and mixed_precision for Fugaku, considering the performance as well as accuracy.

Hybrid GPU+CPU application in GENESIS



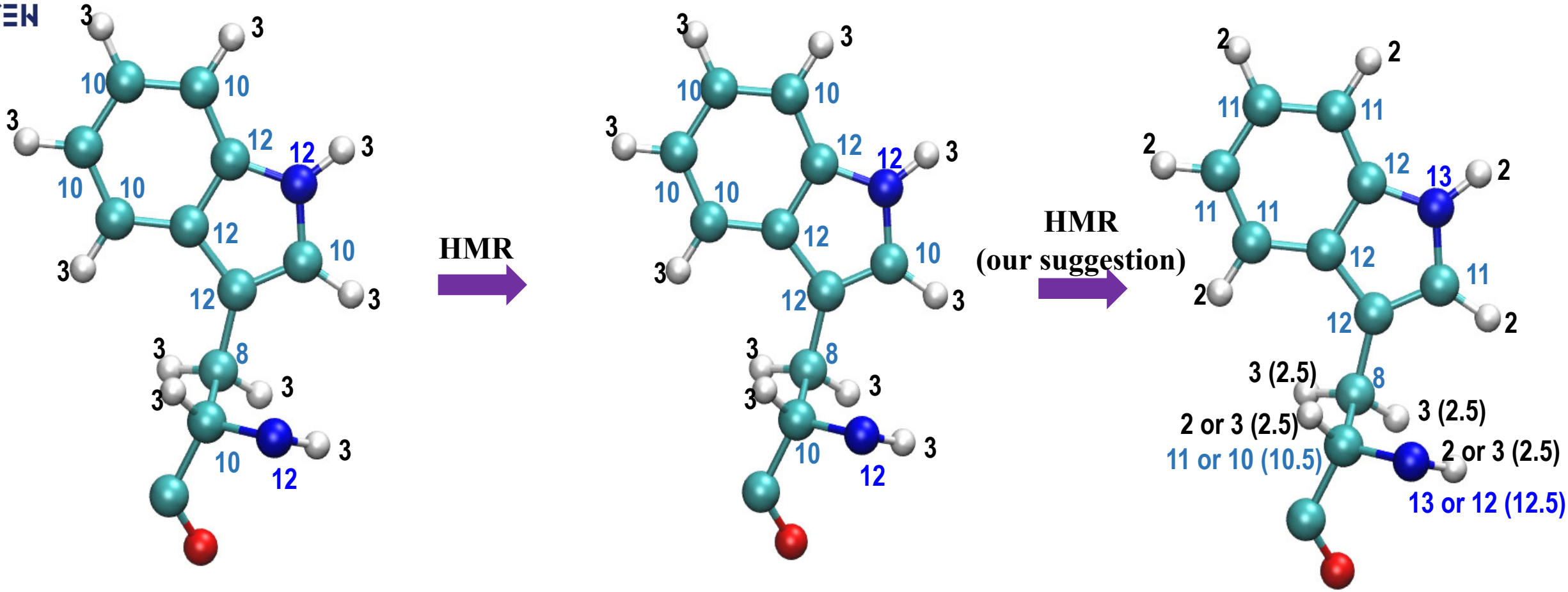
1. Computation intensive work : GPU
   - Pairlist
   - Real space non-bonded interaction

2. Communication intensive work : CPU
   - Reciprocal space non-bonded interaction with FFT
   - Bonded interaction
   - Exclusion list

3. Integration is performed on CPU due to file I/O.

Jung et al. JCTC 12, 4947-4958 (2016).

Our suggestion of HMR ratio



1. HMR ratio is 2 for XH_1 type or ring case
2. For XH_2 or XH_3, we assign HMR ratio 3 (CHARMM) or 2.5 (AMBER)
3. It should be noted that this does not confirm the reliability of 5 fs all the time. To avoid shake error all the time, 3.5 fs with RESPA MTS could be better solution.

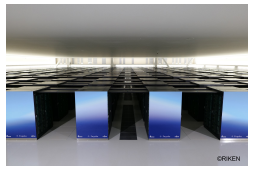
Jung et al. JCTC 17, 5312-5321 (2021)

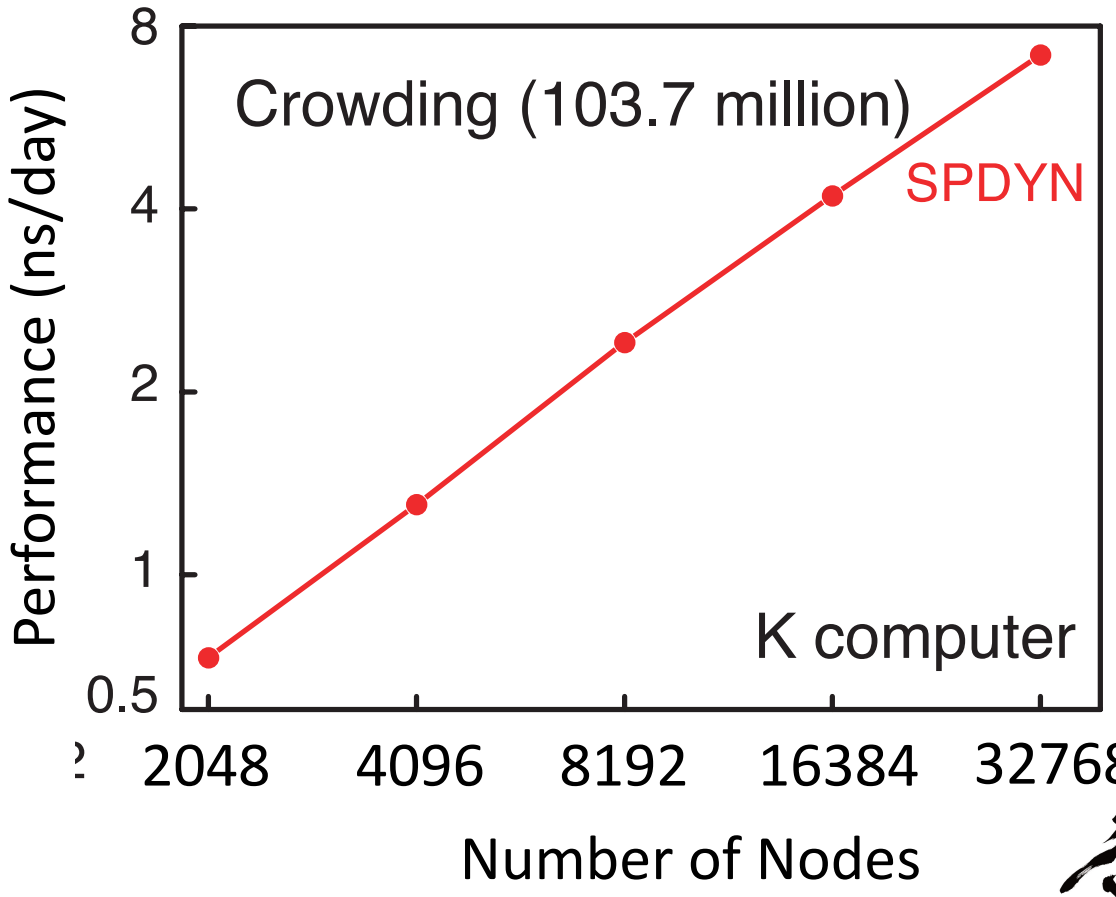
Acceleration of our MD software, GENESIS

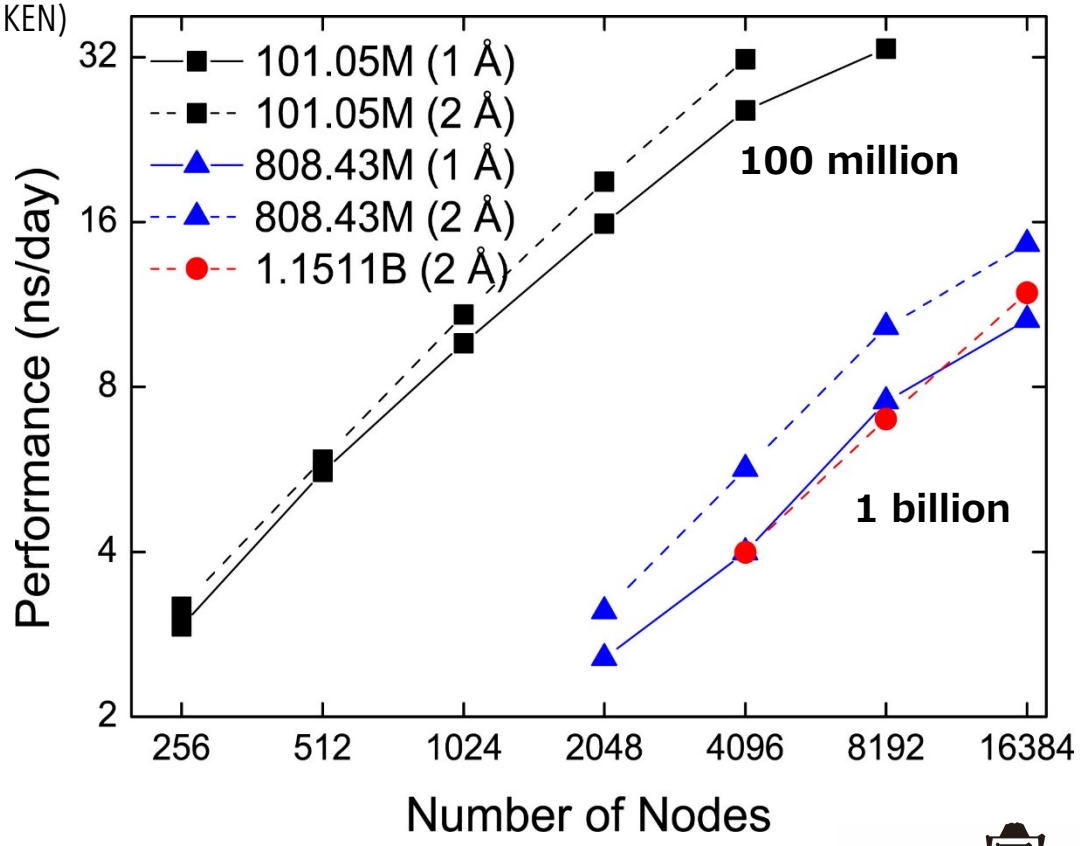
GENESIS version 1 on K

J. Jung (RIKEN)

GENESIS version 2 on Fugaku





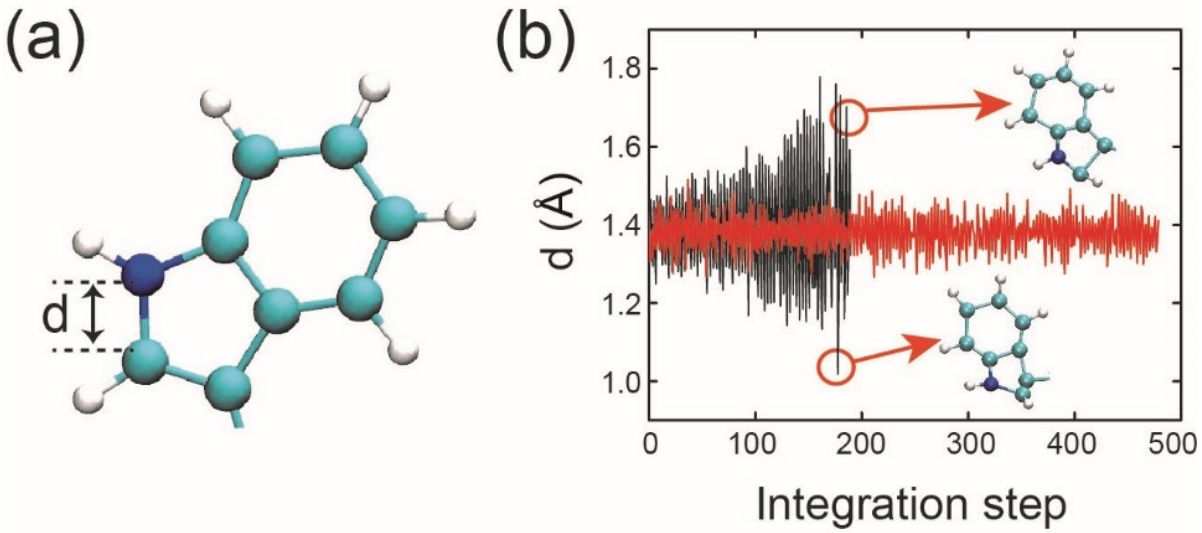


Jung, Mori et al., WIREs Comput. Mol. Science. 5, 310-323 (2015)

Jung et al. JCC 42, 231-241 (2021).

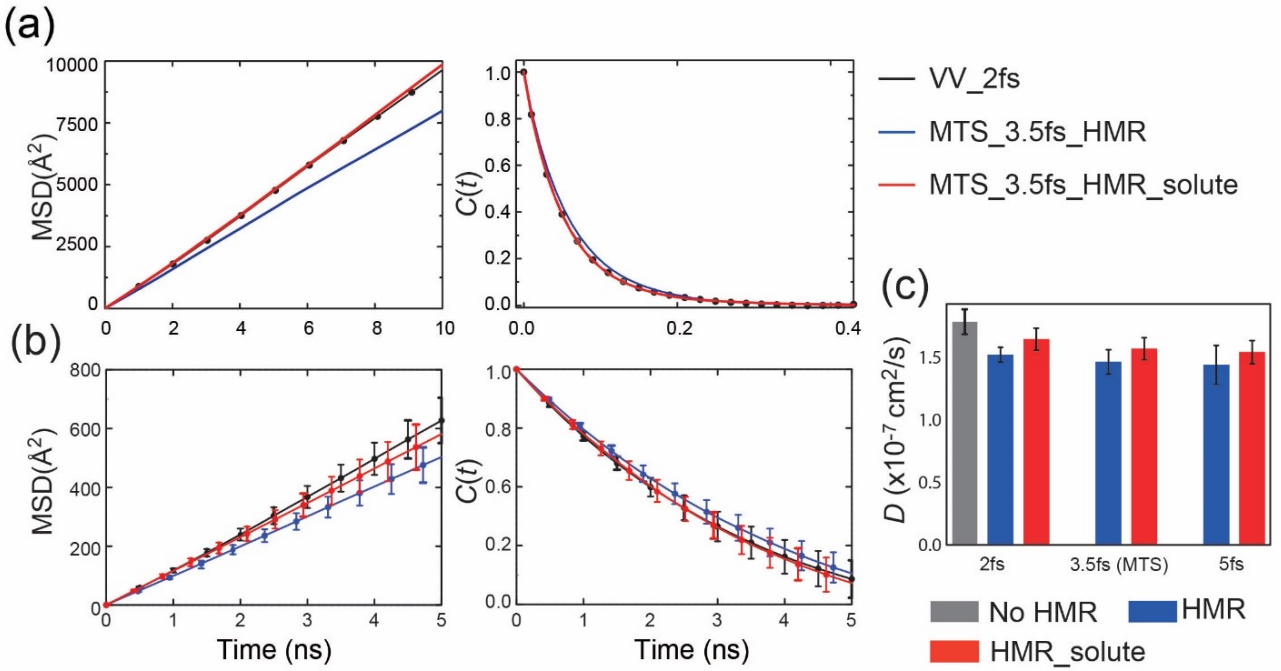
How accurate MD simulation with HMR?

Stable integration with a large time step



MD trajectories of a bond distance between nitrogen and carbon atoms in the pyrrole ring ("d" shown in Fig. 4(a)) of the ligand binding domain of AMPA receptor. The black line starts from 210 steps before the constraint error using HMR. In contrast, HMR_xh1 gives a stable trajectory of the distance, starting from the same structure. (Inset) Two structures of distorted indole rings.

Diffusive motions of proteins and ligands



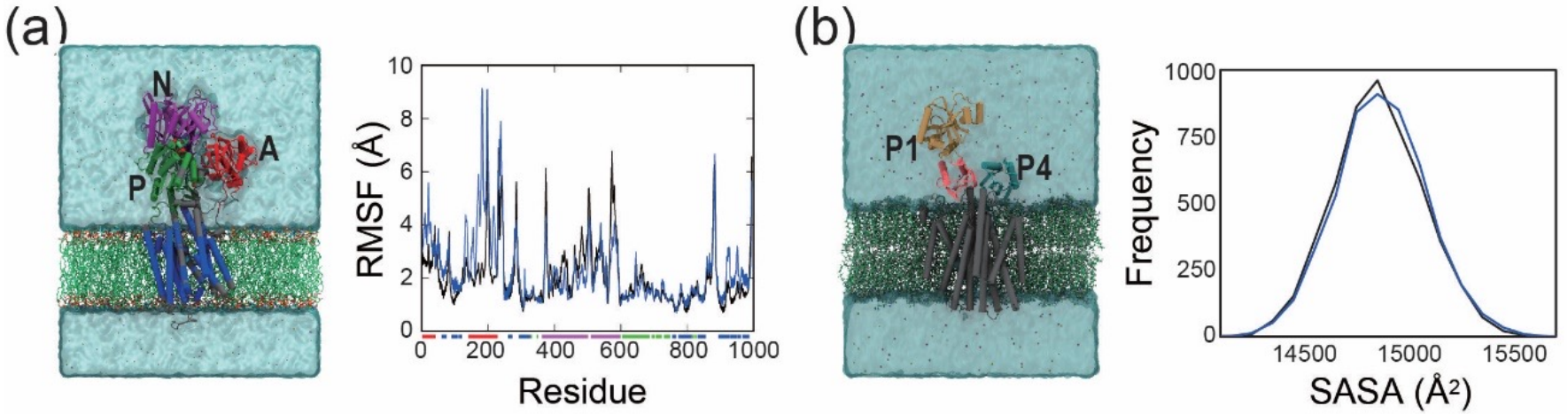
Translational mean squared displacements (left) and rotational correlational function (right) of (a) PP1 and (b) CI2. (c) Lateral diffusion coefficients of DPPC in 160DPPC systems in seven simulations.

Jung et al. JCTC 17, 5312-5321 (2021)

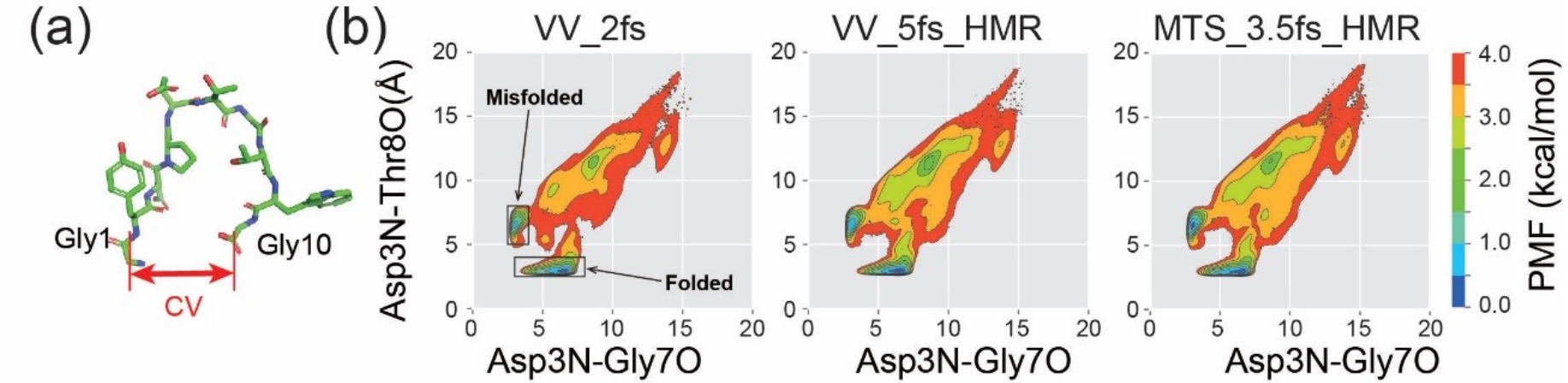
Jung et al. JCTC 17, 5312-5321 (2021)

How accurate MD simulation with HMR?

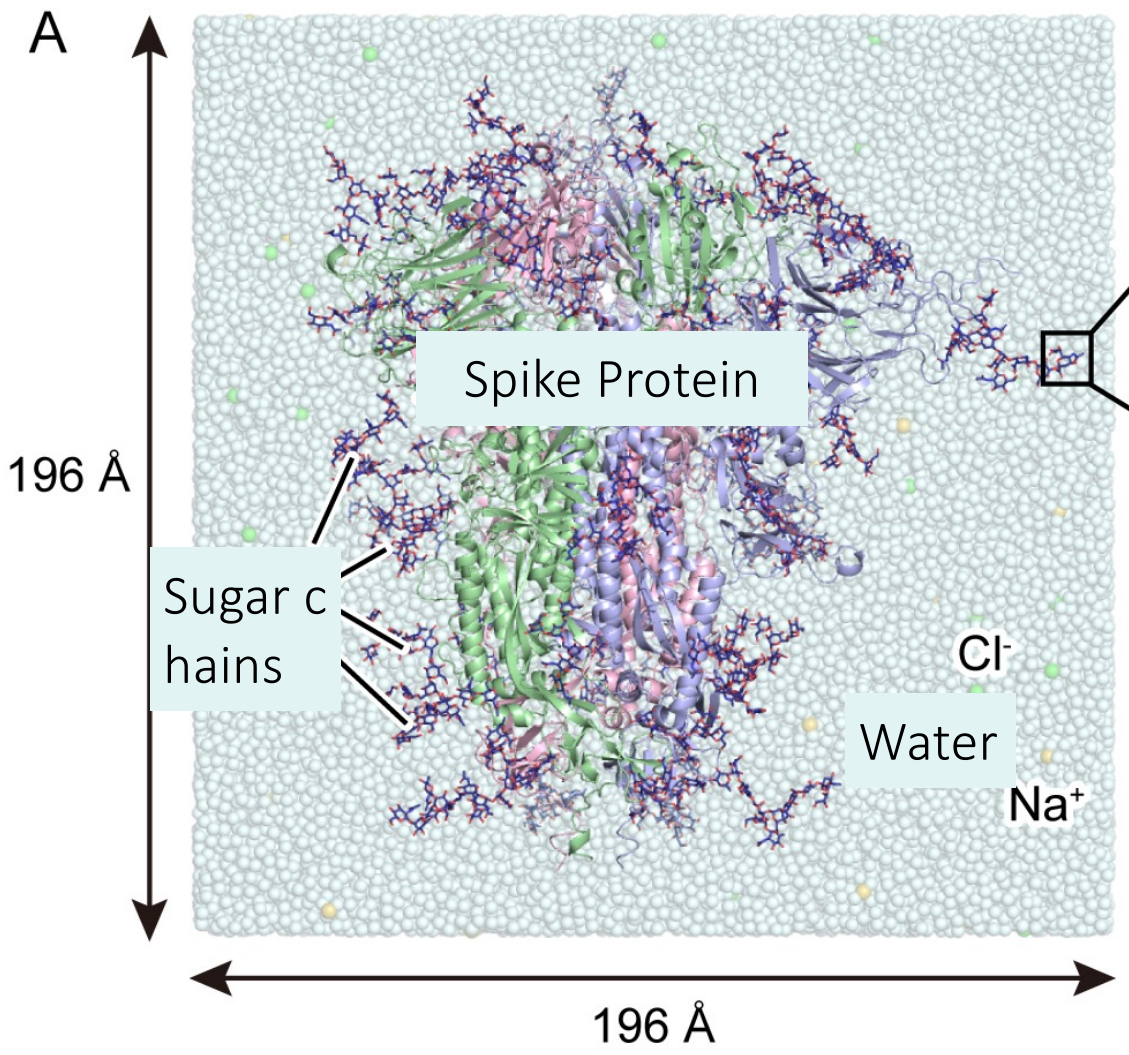
Membrane protein dynamics (a: Calcium pump, b: Sec translocon)

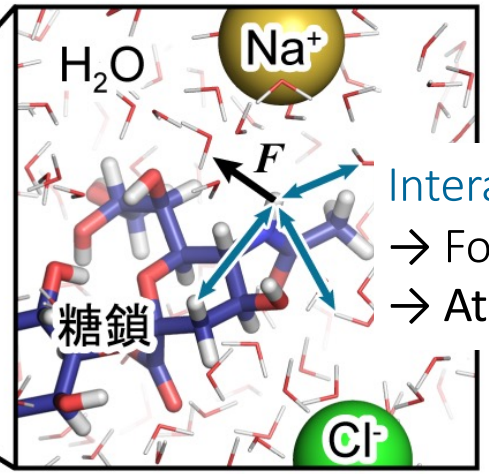


Free-energy landscape of Chignolin in water



MD simulations of SARS-CoV-2 S-protein on Fugaku





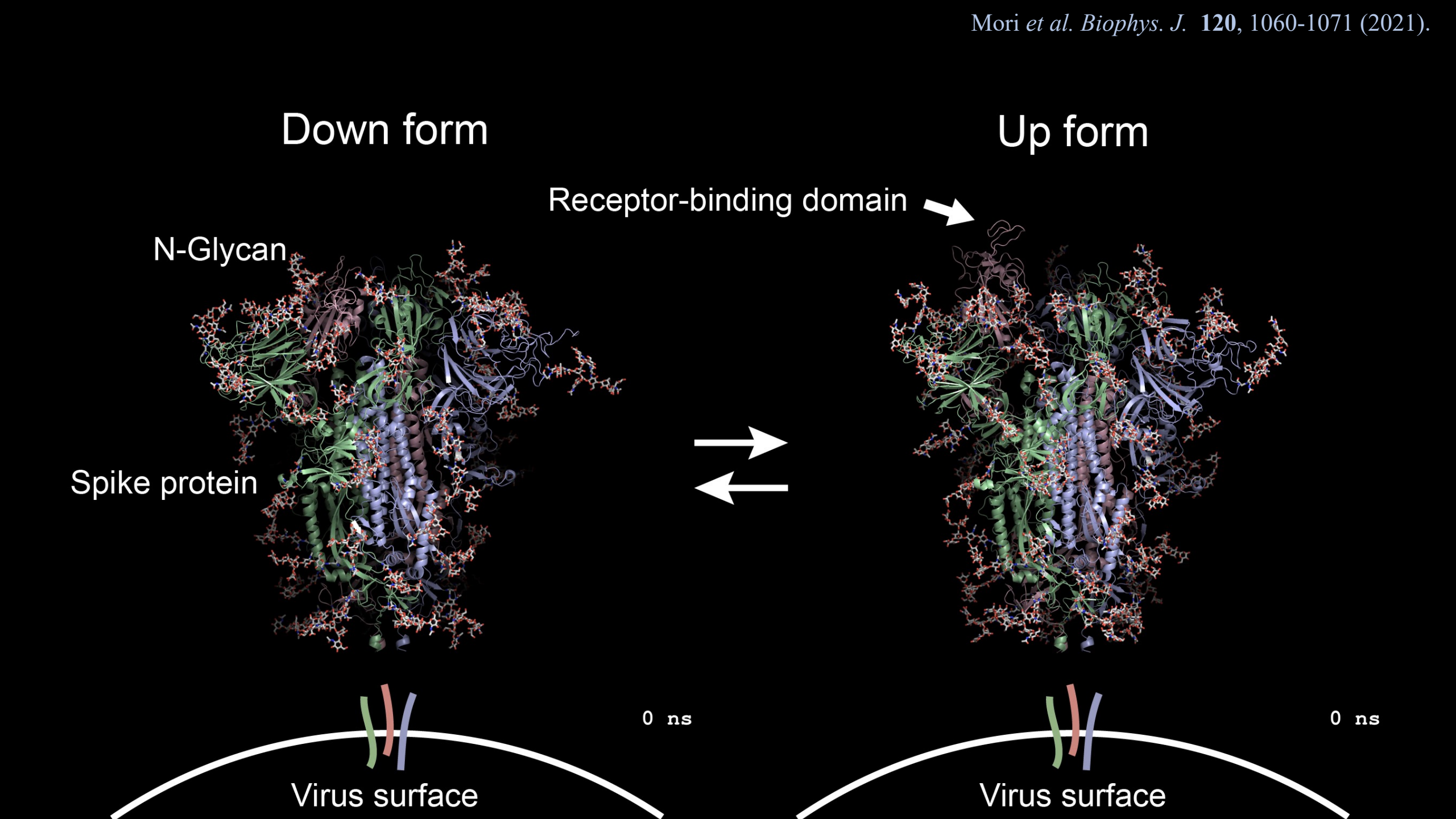
Interactions between atoms
→ Force (F) is applied to atoms
→ Atoms move

GENESIS

Simulation on Fugaku

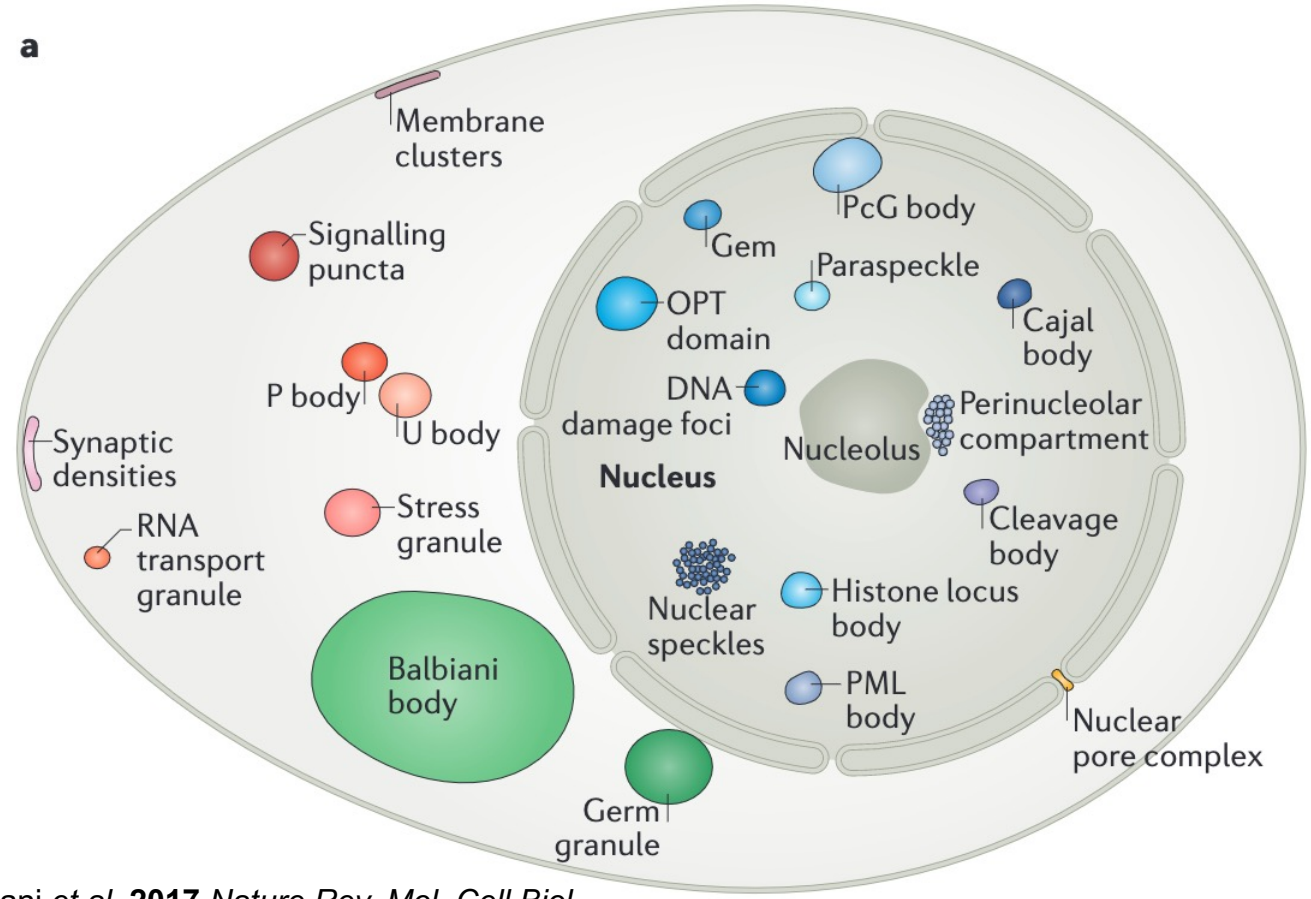
富岳

Mori *et al. Biophys. J.* **120**, 1060-1071 (2021).

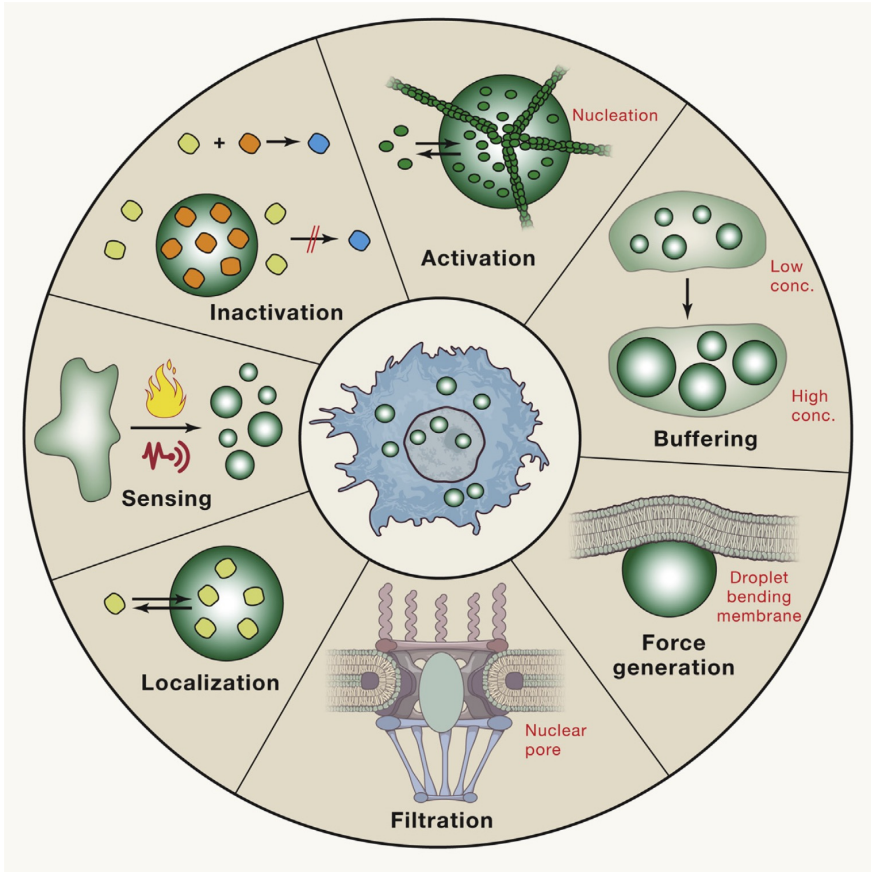


What can we study in biology (using a fast MD program on supercomputers)?

Liquid-liquid phase separation (LLPS)



Banani *et al.* **2017** *Nature Rev. Mol. Cell Biol.*



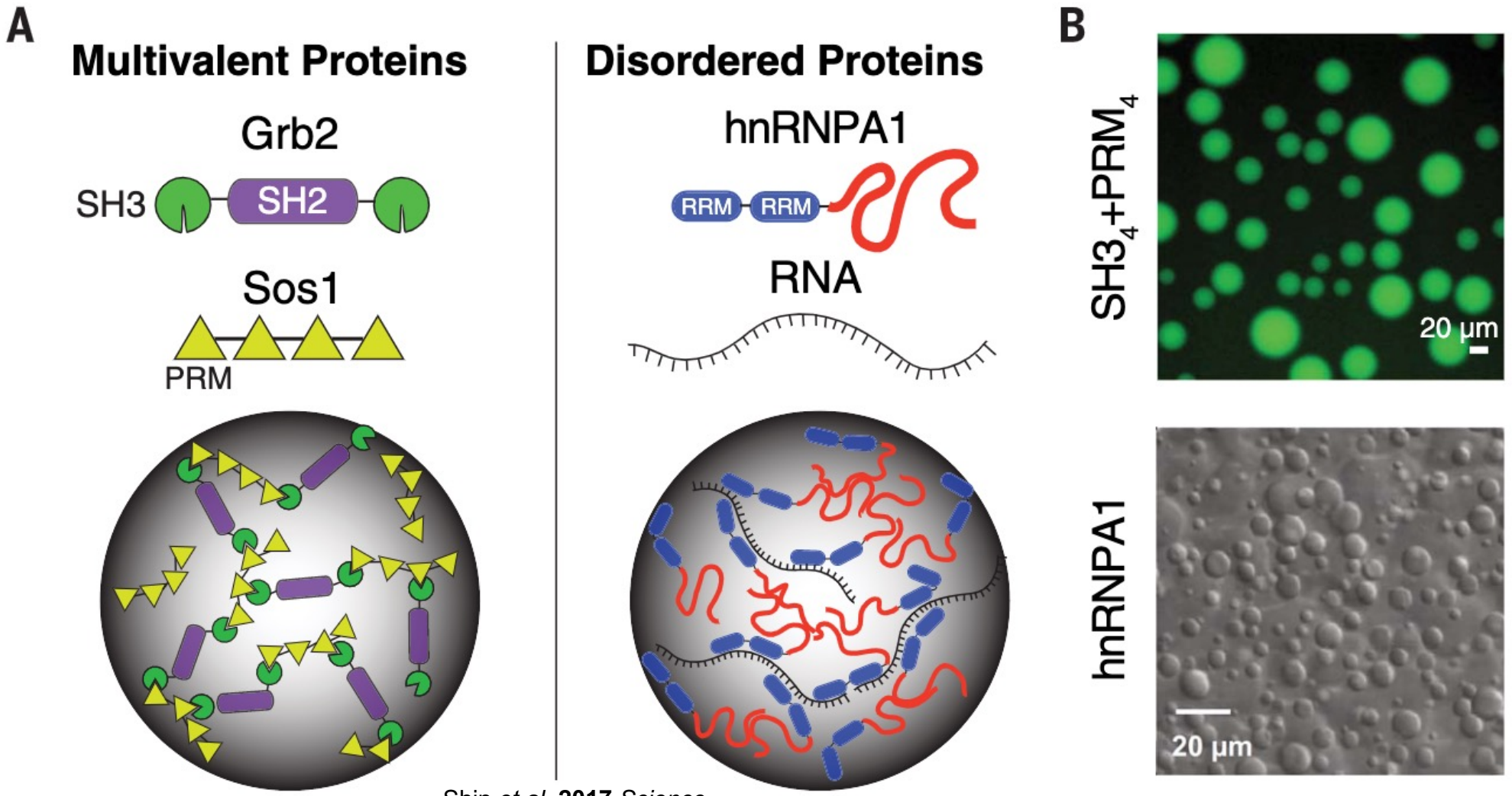
Alberti *et al.* **2019** *Cell*

Names of LLPS in biology:

- Biomolecular Condensates
- Membraneless Organelles
- Liquid Droplets

Multivalency: the General Principle of LLPS

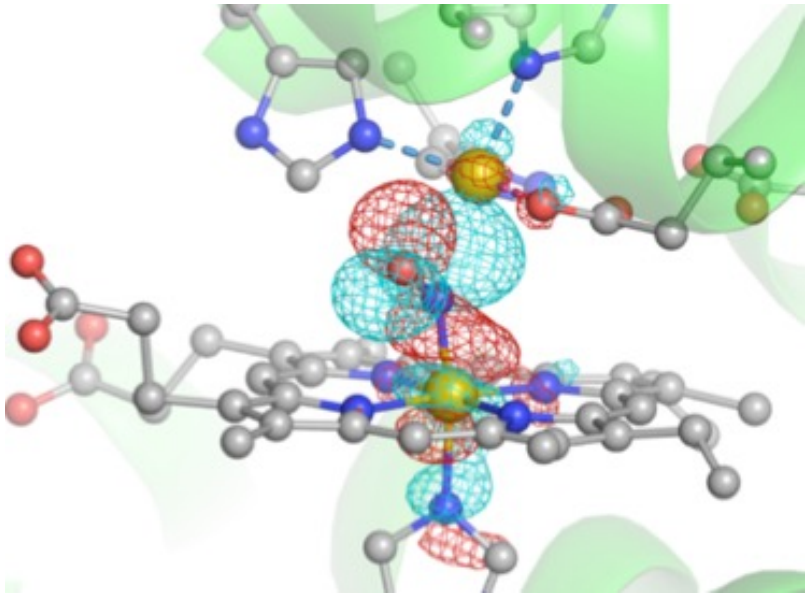
Weak transient interactions between multivalent proteins drive LLPS.



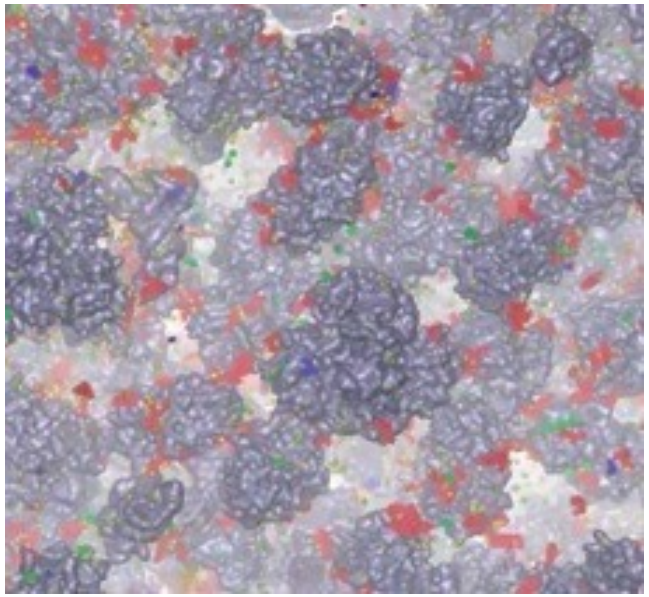
Shin *et al.* **2017** *Science*

Multi-scale molecular models in GENESIS

Hybrid QM/MM MD ⟷ **Atomistic MD** ⟷ **Coarse-Grained MD**



Yagi et al., *JCTC* (2019)

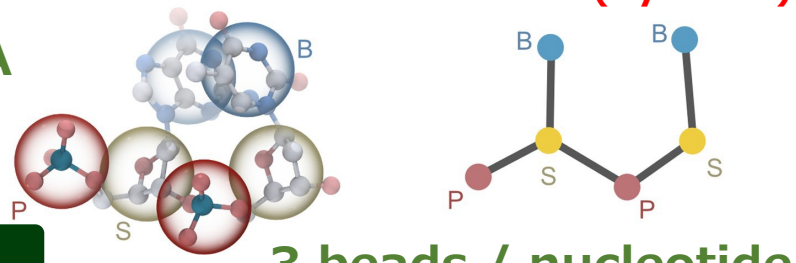


Yu et al., *eLife* (2016)

Protein — **1 bead / amino acid**

S. Takada (Kyoto U.)

DNA



3 beads / nucleotide

Tan et al. *PLoS Comp. Biol.* (2022)

CafeMol

→ Electronic structures
→ Enzyme reactions
→ Material designs
→ Vibration and spectroscopy

→ Membrane protein
→ Protein/DNA complex
→ Cellular environments

M. Feig (Michigan State U.)

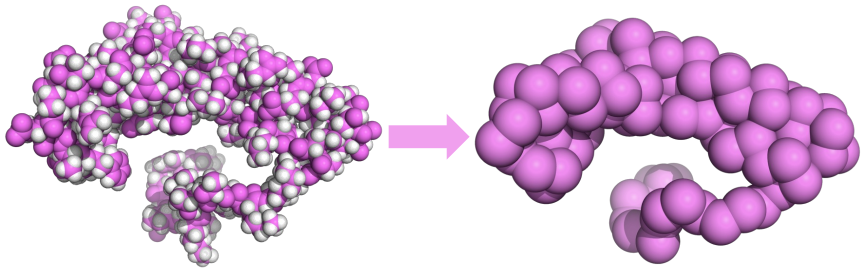
→ Slow dynamics
→ Macromolecular interaction
→ Data-driven simulations
→ Cellular-scale simulations

Structure-based Coarse-Grained Models for Biomolecules

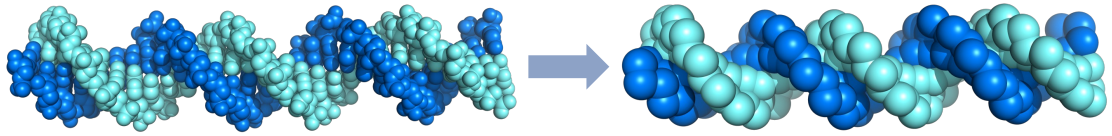
Residue-level coarse-graining: ~10 atoms / CG particle

- **Protein: AICG2+**
  W. Li *et al.* **2014**, *PNAS.*
- **DNA: 3SPN.2C**
  G. Freeman *et al.* **2014**, *JCP.*
- **RNA: Go-like**
  N. Hori *et al.* **2012**, *JCTC.*

1 bead / amino acid

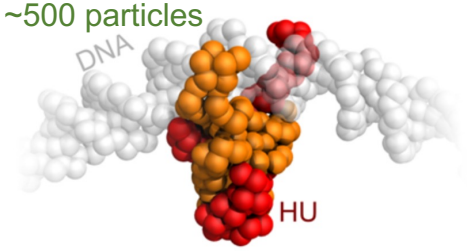


3 beads / nucleotide: Phosphate, Sugar, Base

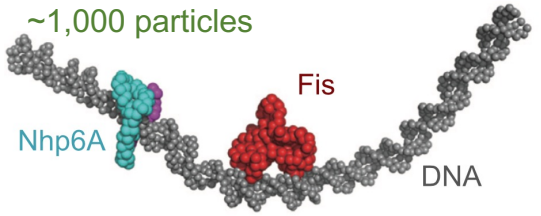


Residue-level CG models: protein-DNA systems

~500 particles

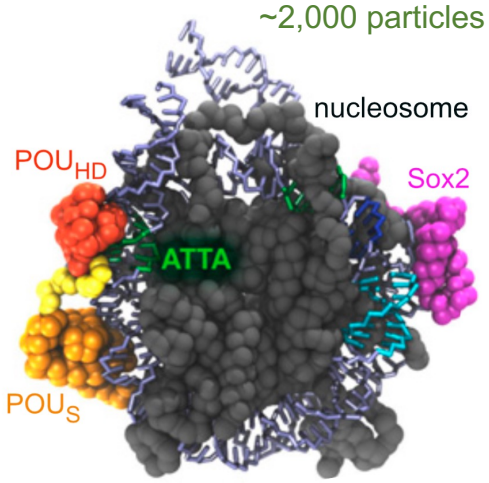


~1,000 particles



C.Tan *et al.* **2016**, *JACS.*
K. Kamagata *et al.* **2020**, *NAR;* **2021**, *NAR.*

~2,000 particles



C.Tan *et al.* **2018**, *JCTC.*
C.Tan *et al.* **2020**, *PNAS.*

Protein-DNA: PWMcos C.Tan *et al.* **2018**, *JCTC.*

CafeMol
https://www.cafemol.org/

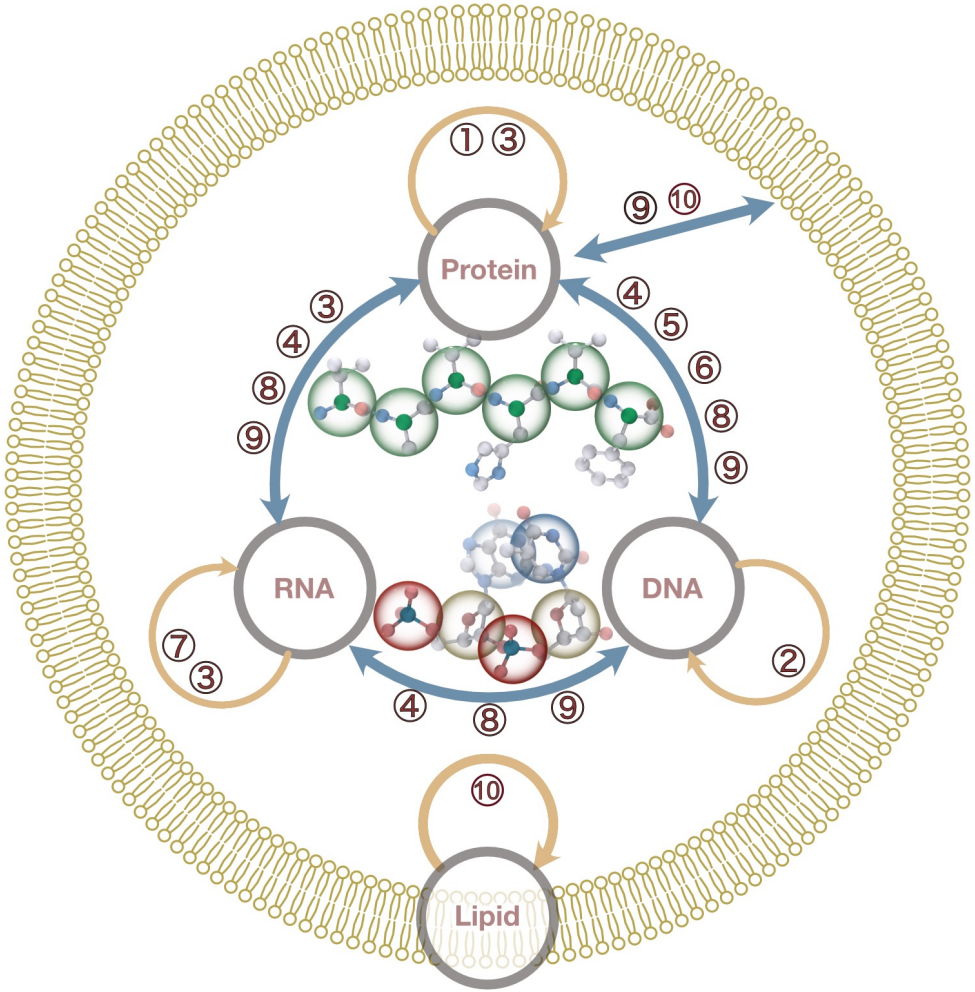

Prof. S. Takada
(Kyoto U.)


C. Tan
(Kyoto U. → RIKEN)

These CG models, which were originated from the structure-based Go-model, have been developed and used in **the CafeMol program by Prof. Shoji Takada's group.**
We (RIKEN) are collaborating with the Takada group (Kyoto U) to develop CG MD simulations in GENESIS software.

CG models in GENESIS MD software

GENESIS supports CG models of protein, RNA, DNA and Lipid



GENESIS

Generalized-ensemble simulation system

Residue-level coarge-grained models

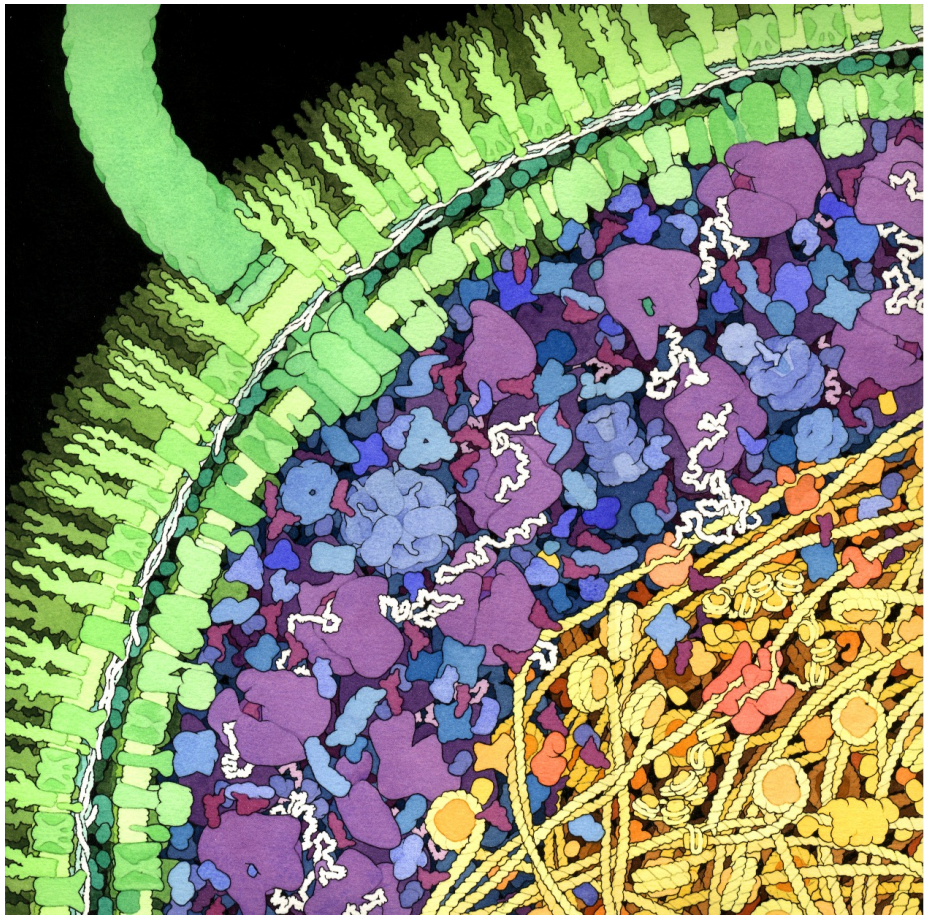
1. W. Li *et al.* (2014). *Proc. Natl. Acad. Sci.*
2. G.S. Freeman *et al.* (2014). *J. Chem. Phys.*
3. G.L. Dignon *et al.* (2018). *PLoS Comput. Biol.*
4. C. Clementi *et al.* (2000). *J. Mol. Biol.*
5. C. Tan & S. Takada (2018). *J. Chem. Theory Comput.*
6. G.B. Brandani *et al.* (2018) *Nucl. Acids Res.*
7. N. Hori. & S. Takada (2012). *J. Chem. Theory Comput.*
8. C. Tan & S. Takada (2016). *J. Am. Chem. Soc.*
9. P. Debye & E. Hückel (1923). *Physikalische Zeitschrift*
10. D Ugarte La Torre *et al.* (2023). *J. Chem. Phys.* ← **New, for lipids!**

C. Tan (RIKEN)

D. Ugarte La Torre (RIKEN)

Summary and Perspectives

Macromolecular Crowding Effect



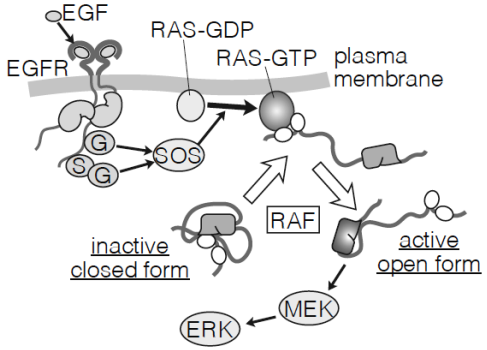
Zimmerman and Minton, *Ann. Rev. Biophys. Biomol. Struct.* (1993) 22: 27-65.

Zhou, H.X., Rivas, G., Minton, A.G., *Ann. Rev. Biophys.* (2008) 37, 375-397.

Multi-scale models (AA, CG, QM/MM) Enhanced Sampling (REMD, REUS, gREST, GaMD, etc.)

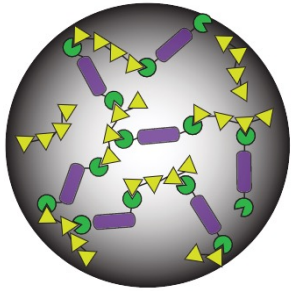
Larger biological systems in atomistic or residue-level MD simulations

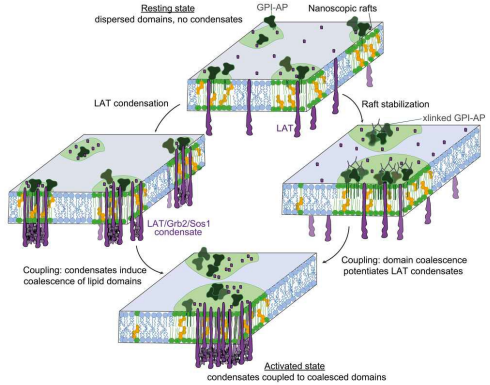
Signal Transduction Pathway



Okamoto, Hibino, Sako. BBA – General Subjects 1864 (2020) 129358.

Protein Condensates





Shin et al., Science 357, 1253 (2017)

Wang et al., Sci. Adv. 9, eadf6205 (2023)

Acknowledgement

Collaborators

RIKEN:

- Dr. Jaewoon Jung (RIKEN CPR/R-CCS)
- Dr. Chigusa Kobayashi (RIKEN/R-CCS)
- Dr. Cheng Tan (RIKEN/R-CCS)
- Dr. Diego Ugarte La Torre (RIKEN/R-CCS)
- Dr. Ai Niitsu (RIKEN/CPR)
- Dr. Isseki Yu (RIKEN/CPR→Maebashi Inst. Tech)
- Dr. Kiyoshi Yagi (RIKEN/CPR)
- Dr. Takaharu Mori (RIKEN/CPR→Tokyo U. Sci)

For Bacterial cytoplasm

- Prof. Michael Feig (Michigan State University)

For CafeMol

- Prof. Shoji Takada (Kyoto University)

Research grants:

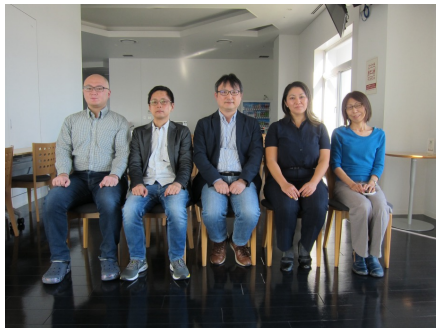
- MEXT "Priority Issue 1 on post-K computer"
- MEXT "Program for Promoting Researches on the Supercomputer Fugaku"
- MEXT Grant-in-Aid for Scientific Research (S) (19H05645)
- MEXT Grant-in-Aid for Transformative Research Area (A), Cross-Scale Biology (21H05249)
- RIKEN Pioneering Projects (Dynamic Structural Biology/Glycolipidologue Initiative/Biology of Intracellular Environments)
- Research grants from RIKEN R-CCS and BDR

Computer resources:

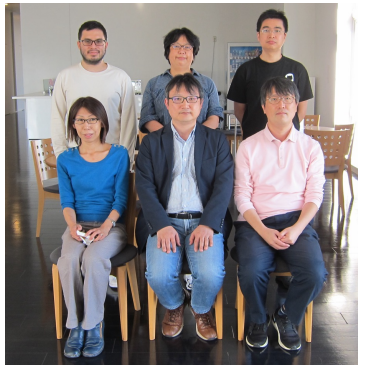
- RIKEN HOKUSAI GreatWave, HOKUSAI BigWaterfall, K computer, and Fugaku supercomputer
- HPCI supercomputer resources (hp170254, hp180201, hp180274, hp190097, hp190181,hp200129, hp200135, hp210107, hp210177, hp220087).

RIKEN Sugita Lab

BDR @ Kobe (2022)



R-CCS @ Kobe (2022)



CPR @ Wako (2023)