





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

BDR



Yuji Sugita







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

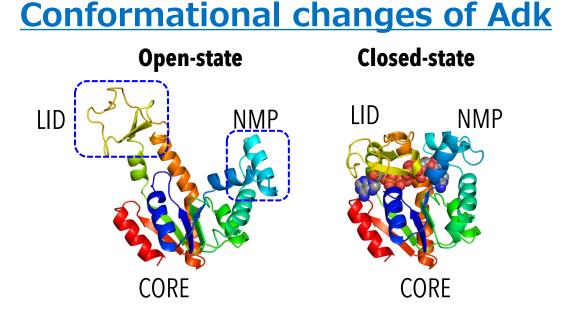


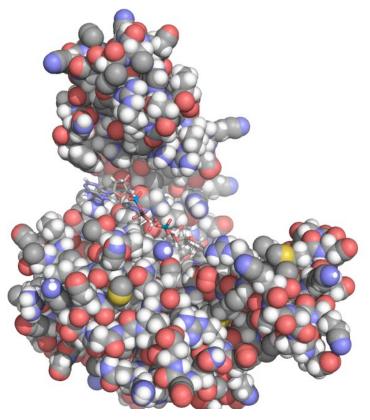
Protein Dynamics and Functions

- 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)

 $ATP + AMP \rightleftharpoons 2ADP$



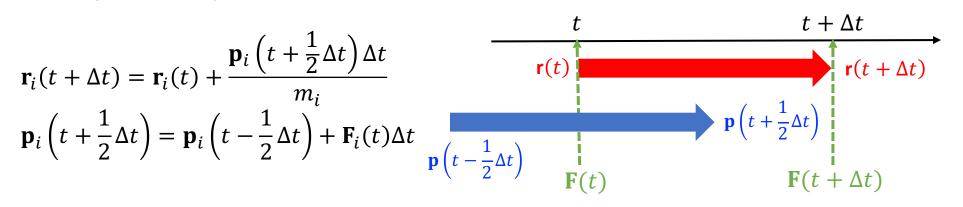


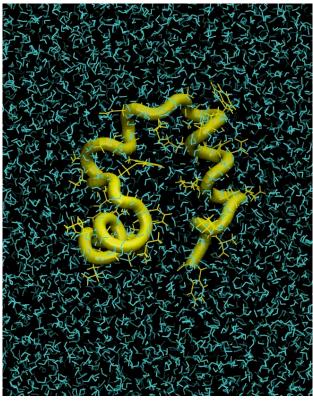
Molecular dynamics (MD) simulation

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

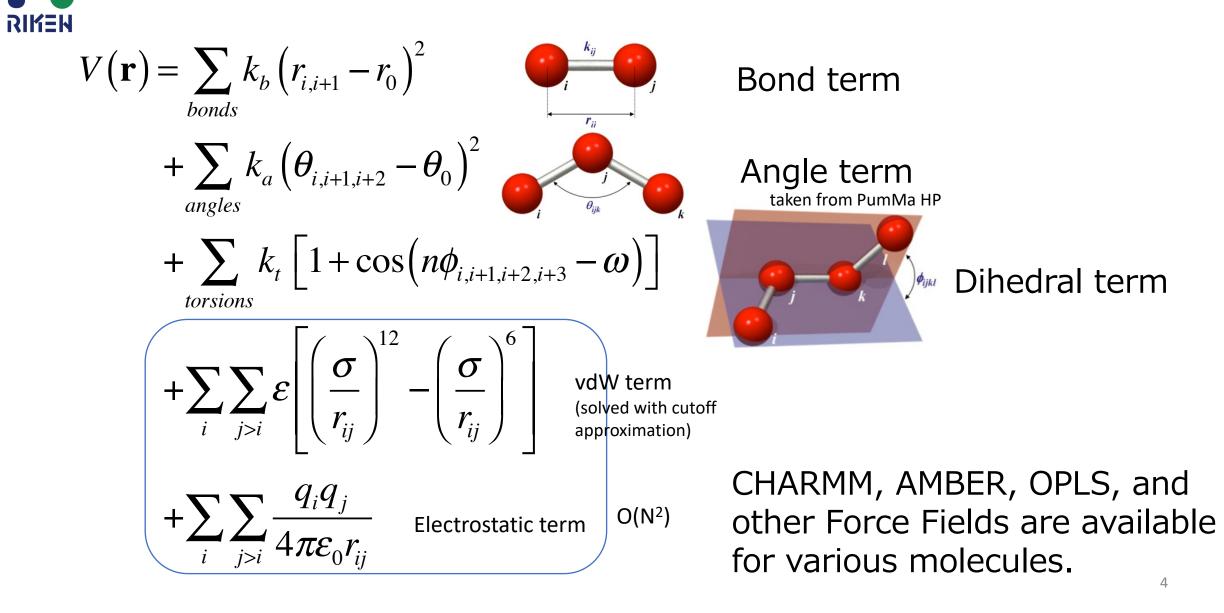
F = ma

 Time step △ t should be very short to reproduce fast vibrations in molecules. (10⁻¹⁵ sec). Many iterations are required to detect protein dynamics.





Molecular Force Field Function



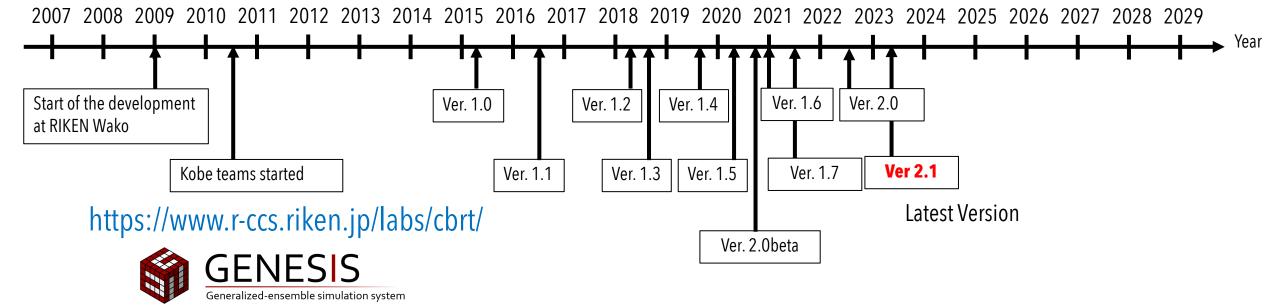


Development of GENESIS

- 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







GENESIS is freely available

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

Feb 20th, 2023skip the Level 1 tutorials, please complete at least Chapters 1.1, 2.1, and 2.2 before going to the next Level	denesis 2.1.0 released	Generalized-ensemble simulation systems Institution Usa (Tutorials & Samples retures Cearch eased eased	[GEI	Tutorials 2022 Here, we show basic, standard, and advanced MD tutorials using GENESIS version 2.0 . For version 1.7.1, see Tutorials 2019. Before starting the tutorials, please install VMD and gnuplot in your computer, which will be used for visualizing MD trajectories and plotting output data, respectively. All readers, especially students and young postdocs, are encouraged to first study all chapters in the "Level 1 Basic tutorials" without skipping any chapter,
	GENESIS 2.0.3 releasedother chapters. Note that Chapter 12 assumes that Chapter 3.2 has been completed. In the following chapter lists, you will find some symbols like , , , and , which means laptop, workstation, and super-computer, respectively.				and then go to the "Level 2 Standard MD tutorials" or "Level 3 Advanced MD tutorials". Even if you really want to skip the Level 1 tutorials, please complete at least Chapters 1.1, 2.1, and 2.2 before going to the next Level



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

RIKEN

Node	CPU	SPARC64™ VIIIfx 2GHz		
	Performance	128 GF (16 GF x 8 cores)		
	Memory	16 GB		
Number of I	Nodes per Rack	864		
Total Numb	er 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		

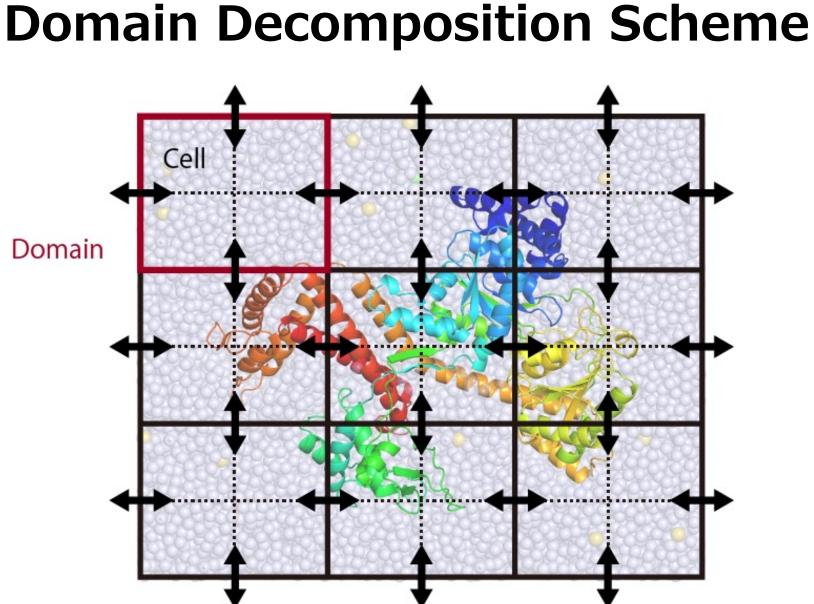


Top500	No.1 for 2 consecutive times since Jun. 2011
Graph500	No.1 in June 2014 & for 6 consecutive times since Jul. 2015
HPCG	No.2 for 4 consecutive times since Nov. 2014 No.1 for 3 consecutive times since Nov. 2016

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.



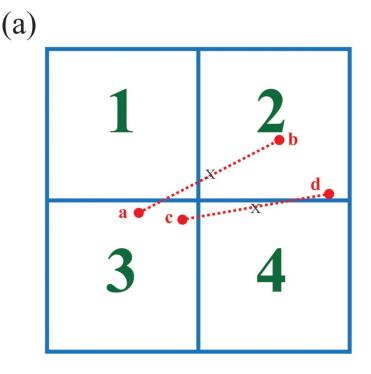
Domain



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



Midpoint Cell Method



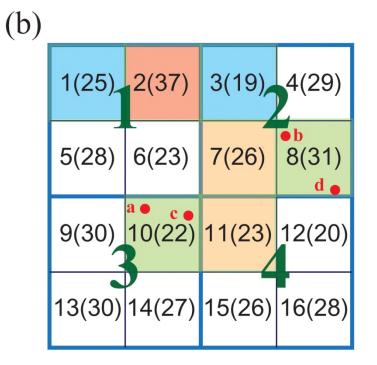
Midpoint Method:

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.

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



Midpoint Cell Method:

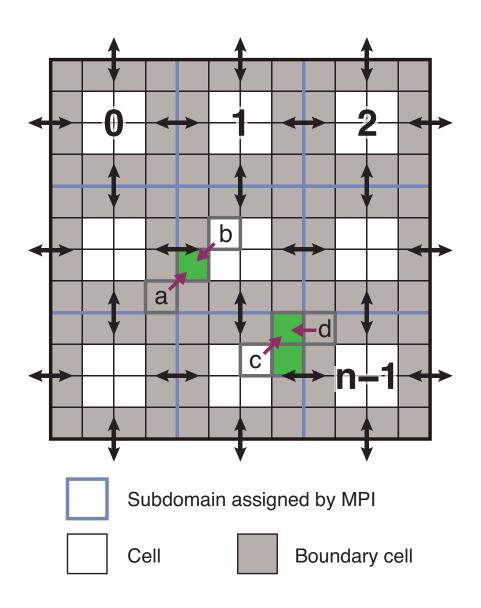
Interaction is computed at the midpoint cell for each cell pair.

- (**O**) Low communication cost
- (**O**) Midpoint cell is decided only once.

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



Subdomains and cells



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

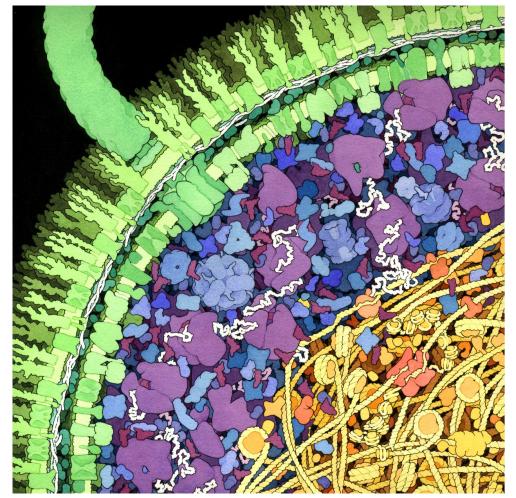


Illustration by Goodsell for E. coli cell

- 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/

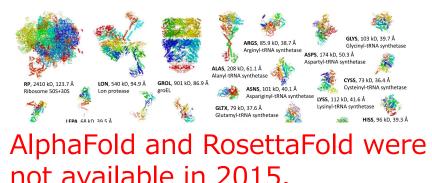
Feig, Harada, Mori, Yu, Takahashi, Sugita : J. Mol. Graph. Model 58, 1-9 (2015).

Atomistic Modeling of the Cytoplasm

Metabolites in M. Genitalium

Homology models of protein structures

RIKEN



Trigger factor FRR, 22 kD, 25.0 Å dnal chape Ribosome recycling factor MRNA, 33 kD, 33.4 Å PEPA, 295 kD, 56.4 Å Cytosol aminopeptidase		63.3 Å RNA TYRS, 136 kf Tyrosyl-tRNA synthetase	MT, 59 kD, 35.1 Å ATRN, 24 kD, 27.4
		32.2 Å thetase	Methionyl-tRNA tRNA formyl transferase
Thymidylate kinase DEOD 159 kD, 44.1 Å T Purine nucleoside phosphorylase	M. F (Michigan Stat	0), 36.8 Å -phosphate TPIA, 109 kD, 40.7 Å Triosephosphate 37.3 Å
UPP 111 kD, 39.6 Å Uracil phosphoribosyltransferase CDD 60 kD, 30.8 Å Cytidine deaminas	2. 2	ACKA, 89 kD, 38.2	isomerase Enolase
	55 kD, 36.1 Å ylate kinase Yate kinase		PYK, 229 kD, 54.6 Å NAMP, 104 kD, 38.8 Å Pyruvate kinase Nicotinamide phosphoribosyltransferase

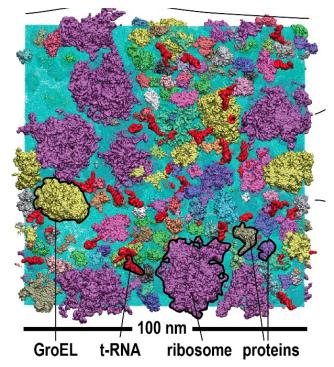
Disordered regions were not included in the models.

Nucleotidase	PRS 78 kD, 40 Å Phosphogluc Ribose phosphate pyrophosphokinase YID1 33 kC Sugar phos	1	STATES S	METK, 85 kD, 35.3 Å S-adenosylmethionine synthase PGM, 57 kD, 30.9 Ø Å Phosphoglucom	itase MTHF, 19 kD, 21.7 Å
снок	5 Å OHRB, 31 kD, 26.3 Å organic hydroperoxide resistance protein RPIB 34	Sugar phosphatase	TH Th FRED, 19 kD, 22.1 Å Flavin mononucleo	otide Methylenetetrahy dehydrogenase PTA, 71 kD, 35.5 Å	
GALU 130 kD, 43 UTP-glucose-1-pl uridylyltransferas	osphate Inorganic polyphospha	PPA, 65 kD, 36.1 Å Inorganic pyrophosphatase	COF, 33 kD, 25.3 Å Hydrolase	Phosphate acetyltransferase 10 nm	RIBF, 62 kD, 35.1 Å Riboflavin kinase

			Am	iino acids]	ß		2	
¥-	the state	+X	*	5	#	Ŧ	X	¥.	X.
PALA Alanine	PARG Arginine	PASN Asparagine	PASP Aspartate	PCYS Cysteine	PLEU Leucine	PLYS Lysine	PMET Methionine	PPHE Phenylalanine	PPRO Proline
T.	x	*	3	Mr.	X	5	±\$€	the	*
PGLN Glutamine	PGLU Glutamate	PGLY Glycine	PHSD/E Histidine	PILE	PSER Serine	PTHR Threonin	PTPF e Tryptophan	PTYR Tyrosine	PVAL Valine
ζφ-	2000	Dates	De la	^b xq	£*.	JX:	* **		3PG
ADE Adenine	DADO Deoxyadenosin	AMP e Adenosine monophosphate	DAMP Deoxyadenosin monophospha		me A A	ACOA cetyl-coa		vdroxyacetone (phosphate	Slycerate-3- phosphate
9 pt	you t	bay the	- Synty	אלי א	r 1	* Er	*	EMN	2PG
ADP Adenosine diphosphate		ATP e Adenosine E triphosphate	DATP Deoxyadenosin triphosphate	Fructose	e-6- Fla	FAD avin adenine	FBP Fructose-1,6-	Flavin G ononucleotide	lycerate-2-
4	-Er	the second	Jup"	phosph	nate di	inucleotide	biphosphate	Glucose6-	and the
CYN Cytosine	CTD Cytidine	DCTD Deoxycytidine	CMP Cytidine monophospha	FOF Forma		G1P lucose-1- phosphate	Glycerol-3-	phosphate	ADPG Adp-glucose
The second	Tot .	×	- T	14		×	* *	NAD Nicotinamide	4
DCMP Deoxycytidine		DCDP Deoxycytidine	CTP Cytidine	Glycerald phosph		GLY Glycerol L-h	LHC	adenine	ACP cetyl-phosphate
monophosphat		diphosphate	triphosphat	20%	~ ³ 4	Jack B	א <u>}</u> א א	33 Mgu	the
Deoxycytidine triphosphate		Guanosine monophosphate	DGMP Deoxyguanosii nonophospha	ne dipho	o-ribose-1 osphate Nic	 NADP cotinamide a ucleotide phi 	idenine	NADH Nicotinamide adenine dinucleotide	Spermidine
×.	Å.	A.	Ž.	4	t	۲ ۲	*	A	C
DCTP oxycytidine ohosphate	GMP Guanosine monophospha	DGMP Deoxyguanos ate monophospha	ine Guanosin ate diphospha	e pvr		RIP Ribose1- bhosphate	DR5P Deoxyribose-5- phosphate	Ribose-5 -phosphate	H2O2 Hydrogen peroxide
tter	Proto the	JAN A	*		$\frac{1}{2}$	4	R.	Q44×	4
DGDP xyguanosi diphosphate	GTP ine Guanosine triphosphate	DGTP Deoxyguanos triphosphate		Ribo		RU5P Ribulose5- phosphate	SAH S-adenosyl - homocysteine	SAM S-adenosyl- methionine	ETOH Ethanol
the second	the states	the state	- Fra	X	Sere !	the .	* Jose	202	\checkmark
Uracil triphosphate	Uridine diphosphate	Uridine Monophosphate	Thymidine monophospha		GL alactose U	UDPG dp-glucose	UDGF Udp - I galactofuranose	DR1P Deoxyribose-1- phosphate	ACE Acetate
₹¥	F	Nucleo		-	+	+	x	لل	Other metabolites
TDP Thymidine diphosphate	Thymidine triphosphate	Bases		Pho	9 12 osphate 9042-	PI1 Phosphate H2PO4 -	PPI Pyrophosphate	NH3 Ammonium	

Fully hydrated models

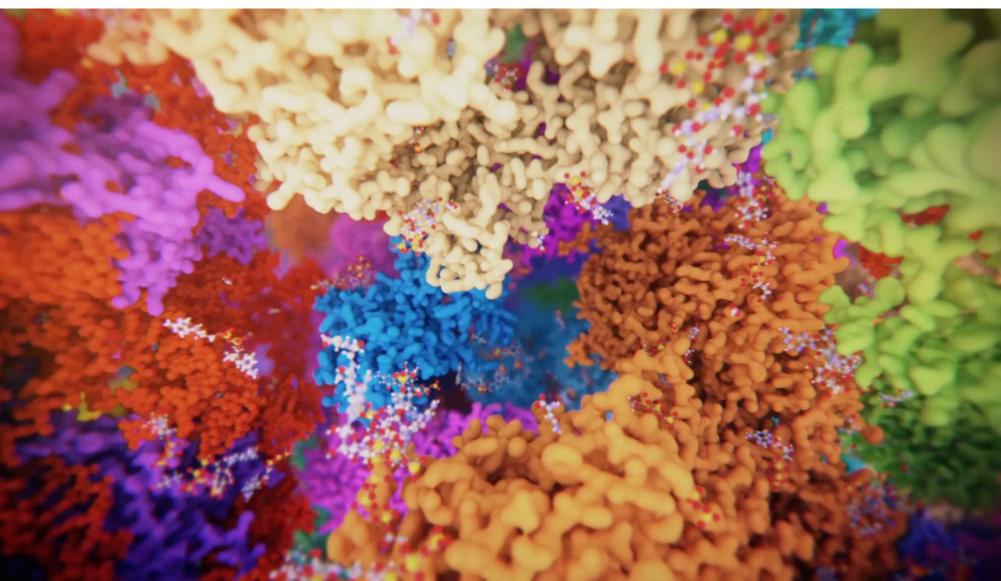
- •# of atoms: 103,708,785
- •# of proteins: 1238
- •# of RNAs: 284
- •# of metabolites: 41,006
- •# of ions: 214,000
- •# of water: 26,263,505





The cytoplasm in *M. genitalium*





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



Isseki Yu (RIKEN → Maebashi Inst. Tech.)

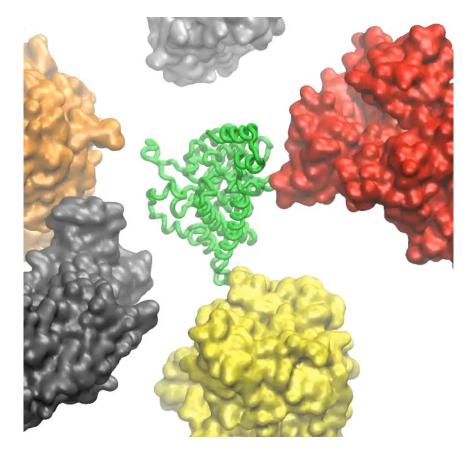


M. Feig (Michigan State University)



Destabilization in Crowded Environments

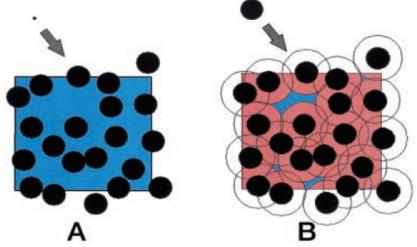
PDHA destabilization in Crowded Environments



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.



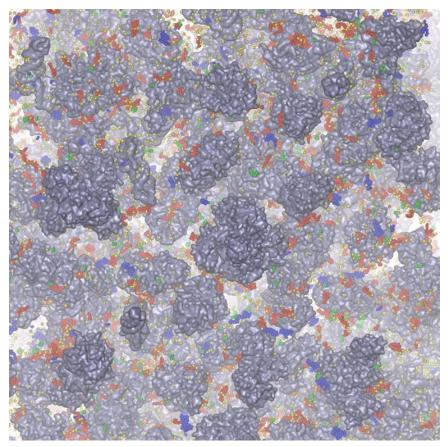
In the crowded condition, compact structures are preferred.

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



ATP controls solubility

ATP distributions on the Protein Surfaces

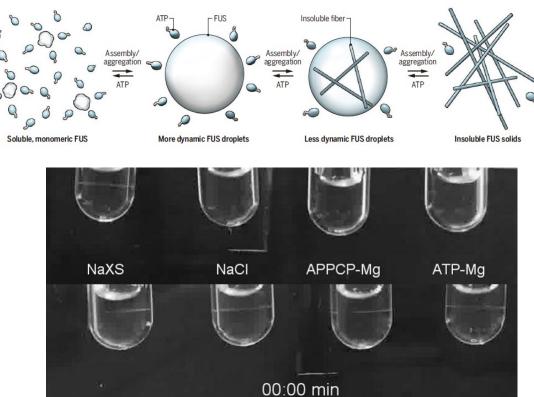


Interactions between proteins and ATPs are important. Yu *et al. eLife* **5**, e19274 (2016).

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

ATP controls solubility

K_m values for ATP-driven cell processes typically lie between 10 and 500 μM, whereas hydrotrope activity requires 2 to 8 mM of ATP. Cells maintain millimolar concentrations of ATP, perhaps to keep proteins (such as FUS) soluble by exploiting ATP as a biological hydrotrope.



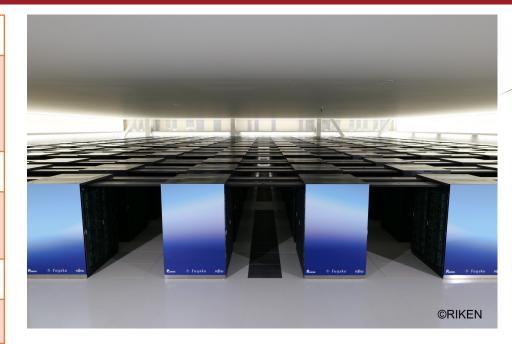
Patel et al. Science 356, 753-756 (2017)

RIKEN Center for Computational Science

The Supercomputer Fugaku

RIKEN

Node	CPU	Armv8.2-A SVE 512bit		
	Performance	3.072TF (DP) 6.144TF (SP) 48(+2)core		
	Memory	32 GB		
Number of I	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		



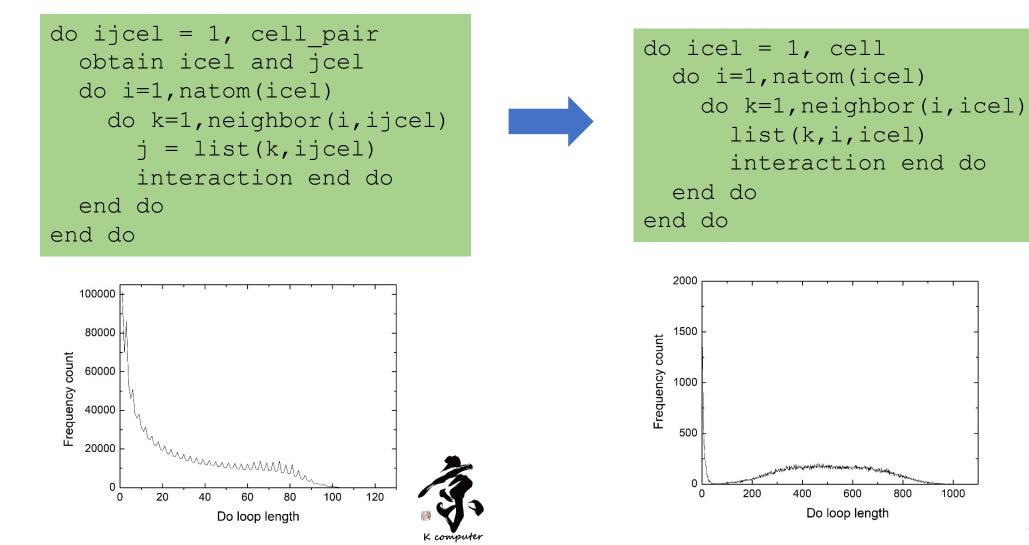
画后

Тор500	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



ApoA1 on one MPI processor

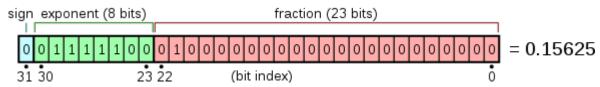
H

F

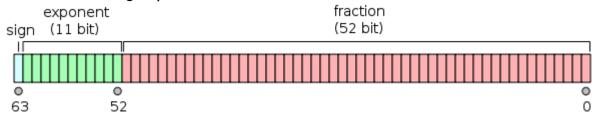


Single/Mixed/Double precision

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



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



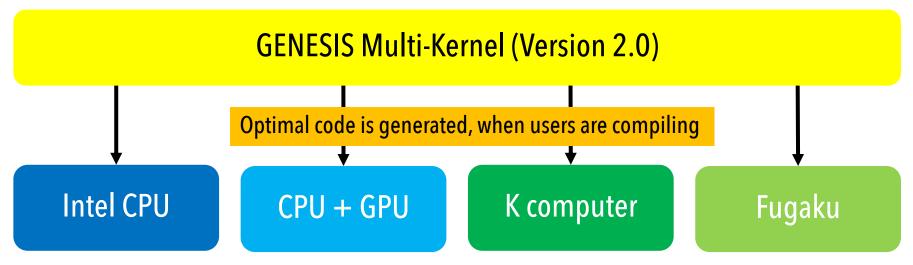
- We can define precision levels in GENESIS 2.0 according to our purposes
 - Single precision: All real numbers except energy/constraints are described by single precision
 - Mixed precision: Integration by double precision and force calculation by single precision
 - Double precision: All numbers are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

Picture is from wikipedia.



Multiple kernels for nonbonded energy calculations

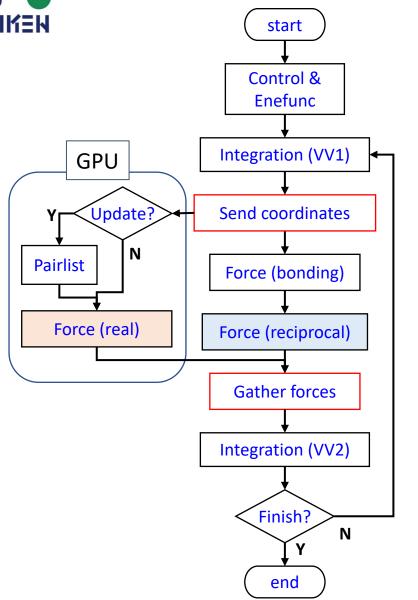
User Friendly Multiple Kernels for Nonbonded Energy Calculations



- 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.

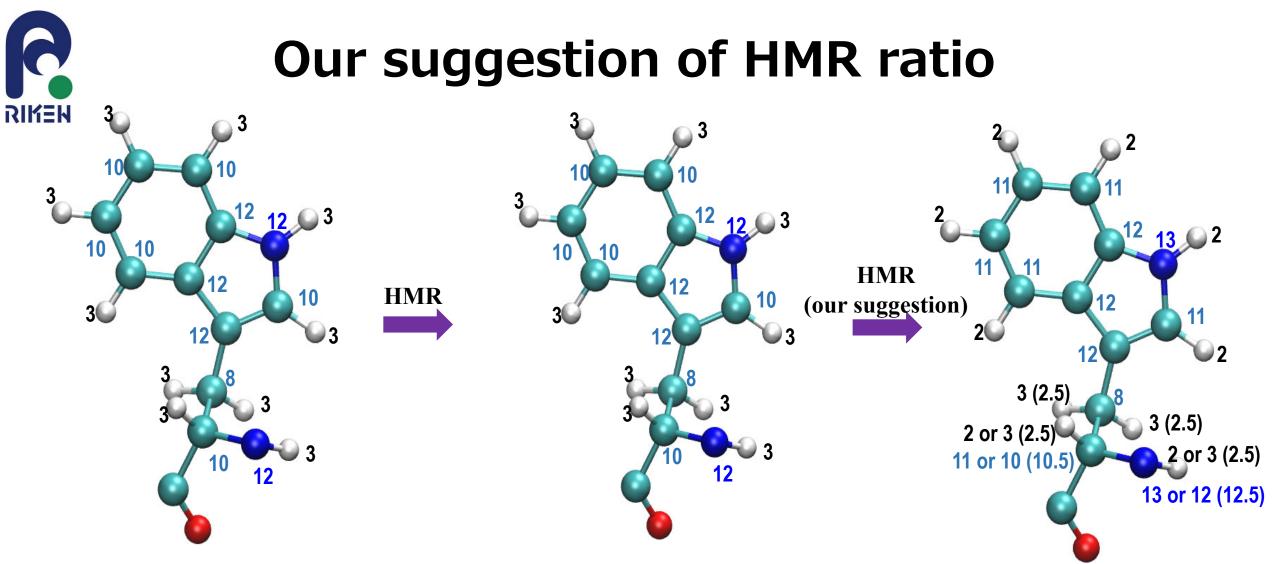
RIKEN

Hybrid GPU+CPU application in GENESIS



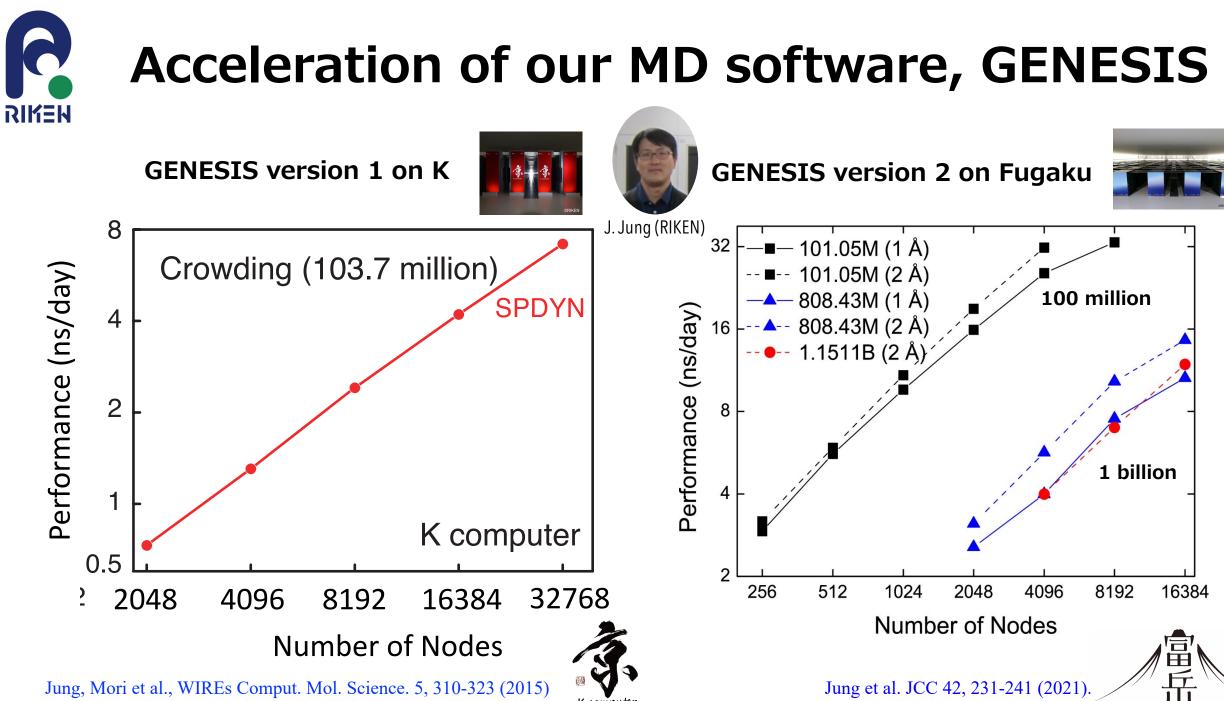
- 1. Computation intensive work : GPU
 - Pairlist
 - Real space non-bonded interaction
- 2. Communication intensive work : CPU
 - Reciprocal space non-bonded interaction with FFT
 - Bonded interaction
 - Exclusion list
- 3. Integration is performed on CPU due to file I/O.

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



- 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)

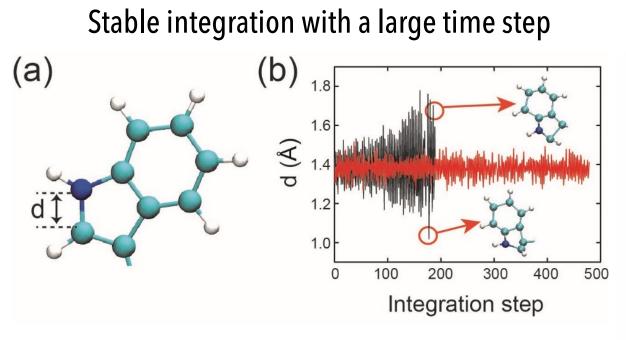


K computer

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



How accurate MD simulation with HMR?

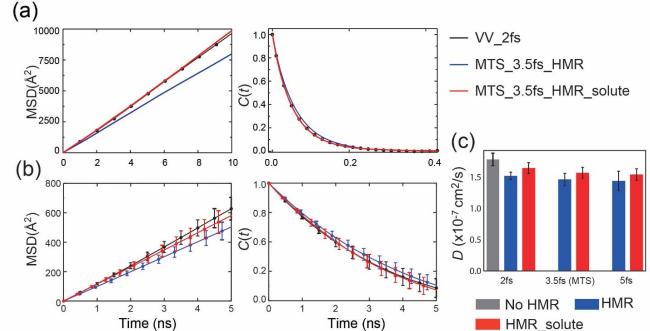


MD trajectories of a bond distance between nitrogen and carbon atoms in the pyrrole ring ("d" shown in Fig. 4(a)) of the ligand binding domain of AMPA receptor. The black line starts from 210 steps before the constraint error using HMR. In contrast, HMR_xh1 gives a stable trajectory of the distance,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

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. JCTC247, 5312-5321 (2021)

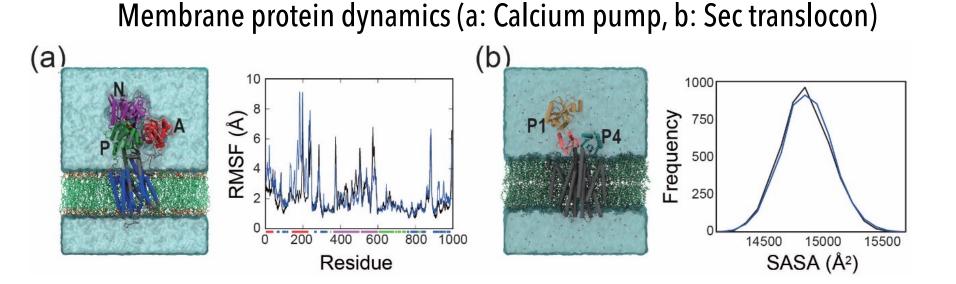
Diffusive motions of proteins and ligands



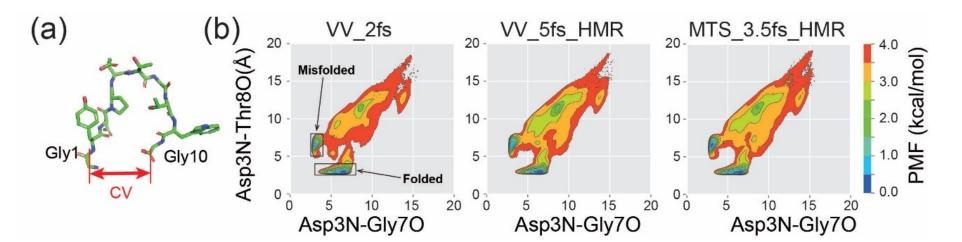
Jung et al. JCTC 17, 5312-5321 (2021)

How accurate MD simulation with HMR?

SIKEN

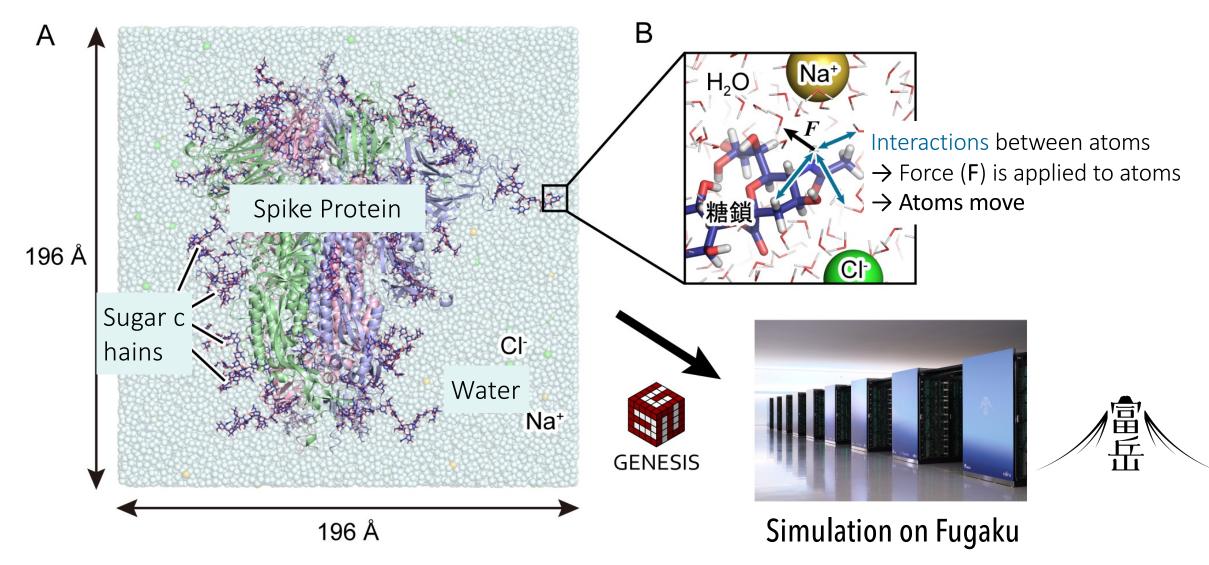


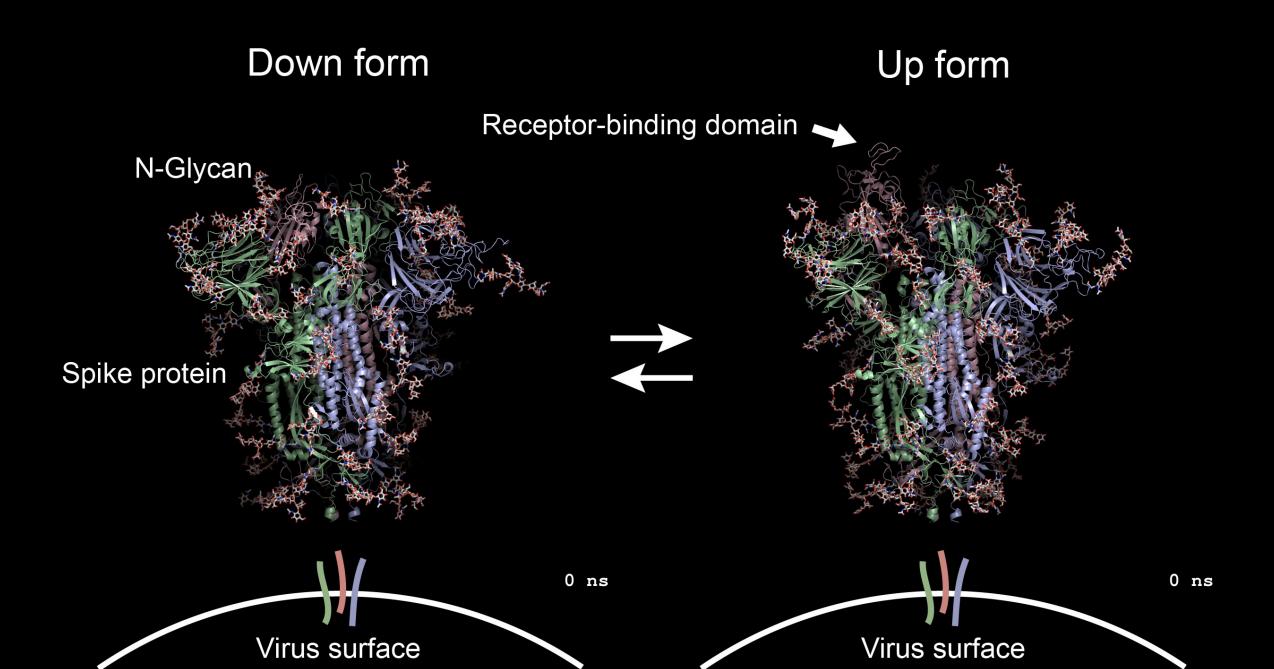
Free-energy landscape of Chignolin in water





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

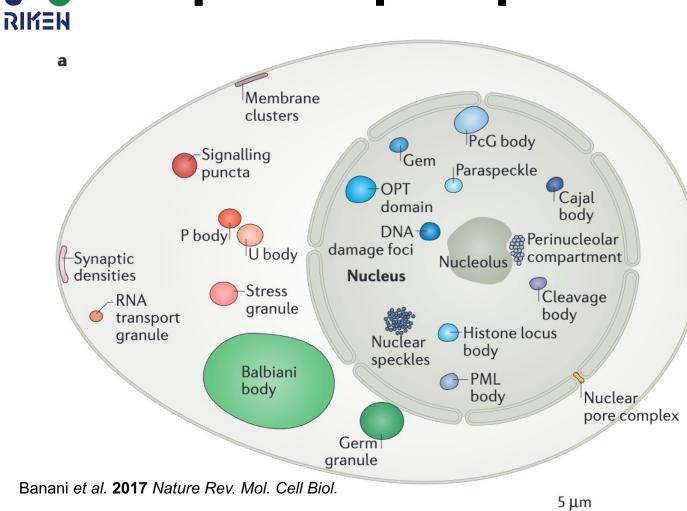






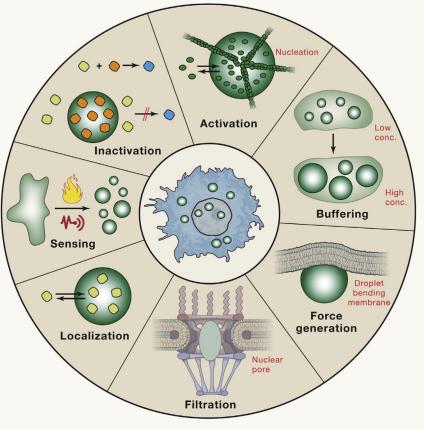
What can we study in biology (using a fast MD program on supercomputers)?

Liquid-liquid phase separation (LLPS)



¢





Alberti et al. 2019 Cell

Names of LLPS in biology:

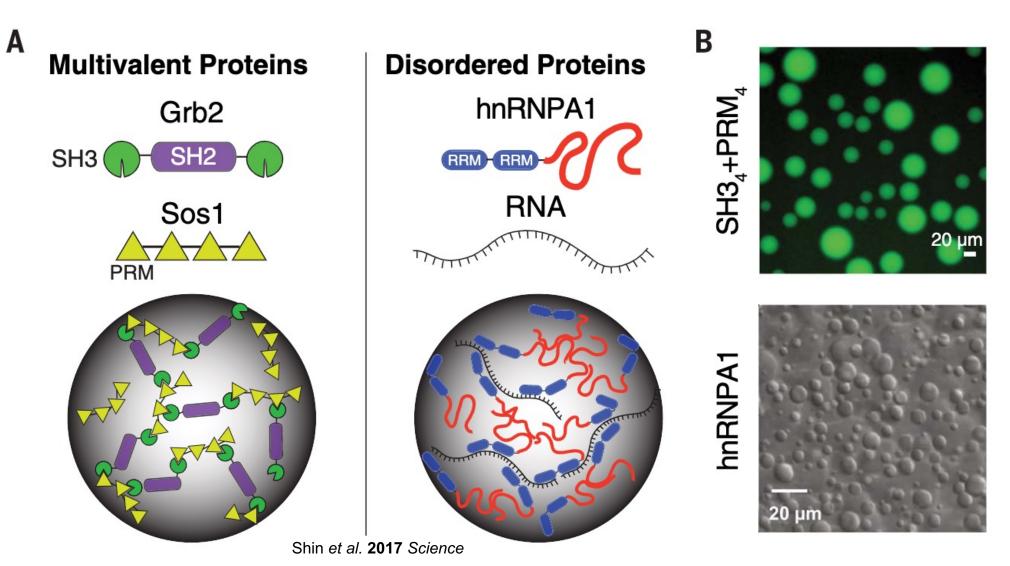
- Biomolecular Condensates
- Membraneless Organelles
- Liquid Droplets

Multivalency: the General Principle of LLPS

C

RIKEN

Weak transient interactions between multivalent proteins drive LLPS.

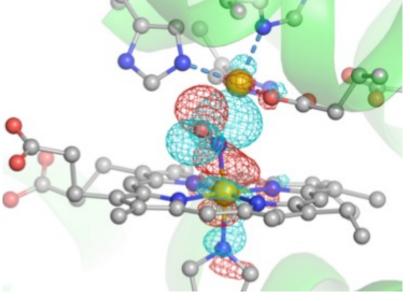


Multi-scale molecular models in GENESIS

Hybrid QM/MM MD

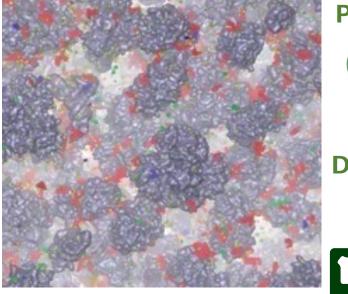
Atomistic MD

Coarse-Grained MD



Yagi et al., JCTC (2019)

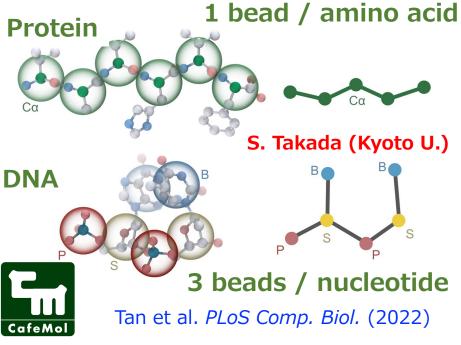
- \rightarrow Electronic structures
- \rightarrow Enzyme reactions
- \rightarrow Material designs
- \rightarrow Vibration and spectroscopy



Yu et al., eLife (2016)

- \rightarrow Membrane protein
- \rightarrow Protein/DNA complex
- \rightarrow Cellular environments

M. Feig (Michigan State U.)

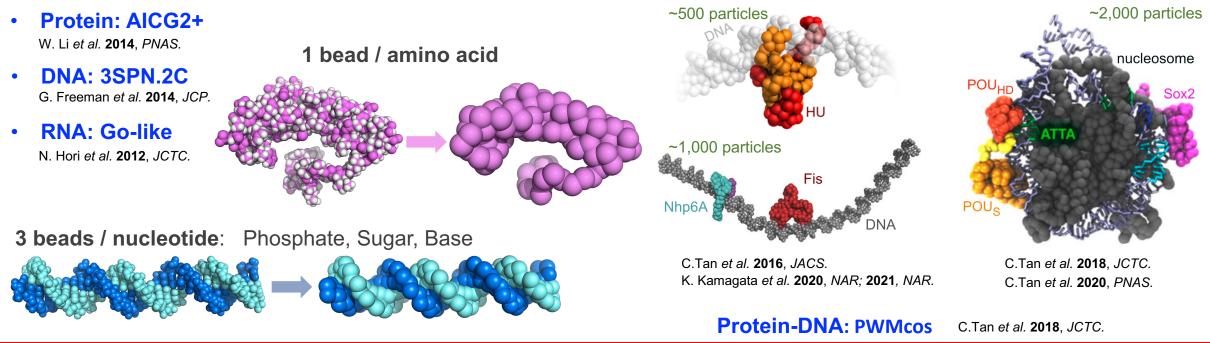


- \rightarrow Slow dynamics
- \rightarrow Macromolecular interaction
- \rightarrow Data-driven simulations
- \rightarrow Cellular-scale simulations



Structure-based Coarse-Grained Models for Biomolecules

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





https://www.cafemol.org/



Prof. S. Takada C. Tan (Kyoto U.) (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.**

Residue-level CG models: protein-DNA systems

We (RIKEN) are collaborating with the Takada group (Kyoto U) to develop CG MD simulations in GENESIS software.

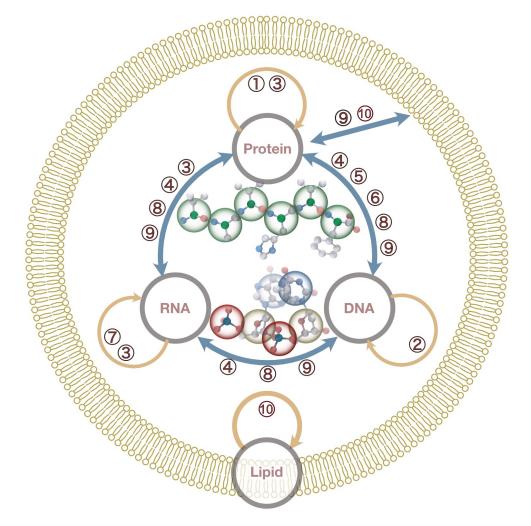


CG models in GENESIS MD software

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



C. Tan (RIKEN)







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!

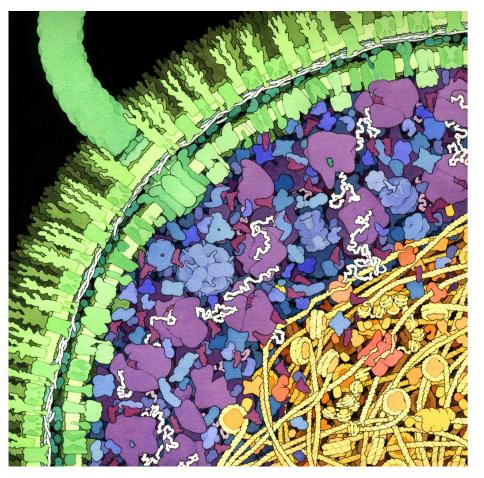


D. Ugarte La Torre (RIKEN)

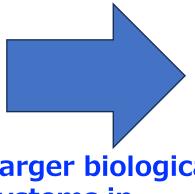


Summary and Perspectives

Macromolecular Crowding Effect

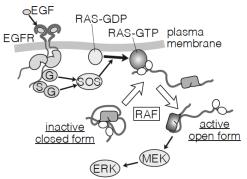


Multi-scale models (AA, CG, QM/MM) Enhanced Sampling (REMD, REUS, gREST, GaMD, etc.)



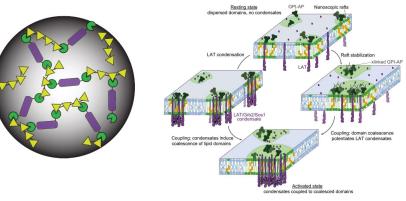
Larger biological systems in atomistic or residue-level MD simulations

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)

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.

SIKEN Collaborators

Acknowledgement

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)

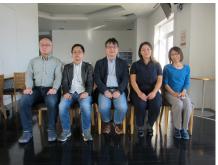
Research grants:

- 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

<u>Computer resources:</u>

- 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)

