# Large-Scale Molecular Dynamics Simulation of Biomolecules in Cellular Environments

Accelerated Data Analytics and Computing Workshop 8
University of Tokyo, Kashiwa Campus, Kashiwa, Japan, October 30-31, 2019

**Yuji Sugita**

Laboratory for Biomolecular Function Simulation, RIKEN BDR
Computational Biophysics Research Team, RIKEN R-CCS
Theoretical Molecular Science Laboratory, RIKEN CPR

GENESIS
Generalized-ensemble simulation system

# To Accelerate Biological MD Simulations

- Accelerate MD simulation
  - Why MD is so slow ?
  - How to accelerate MD ?

J. Jung
(CPR / R-CCS)

C. Kobayashi
(R-CCS)

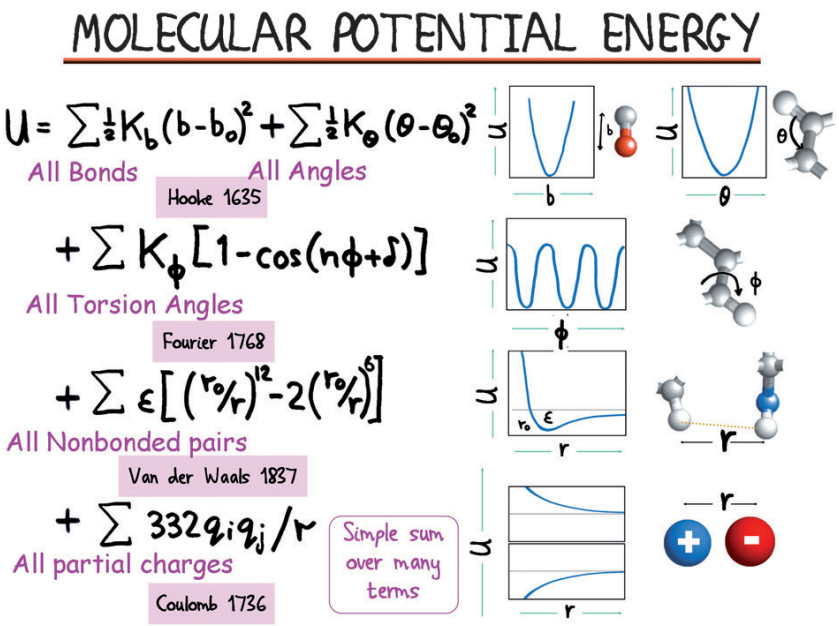
T. Mori
(CPR)

- Efficient MD algorithms
  - Enhanced Conformational Sampling
  - Application to Protein-Ligand Binding

# Why MD is so slow ?

## [I] Atomic Force Evaluation

Atomic force is evaluated as a derivative of Molecular Potential Energy Function.




## [II] Time Integration

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

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

## Equation of Motions

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

## 1) Large Number of Atoms

- At least $10^4$ atoms
- Sometimes more than 1M atoms

## 2) Nonbonded Interactions

- Nonbonded Interaction is the slowest calculations in Energy/Force Evaluations.

## 3) $\Delta t$ must be small

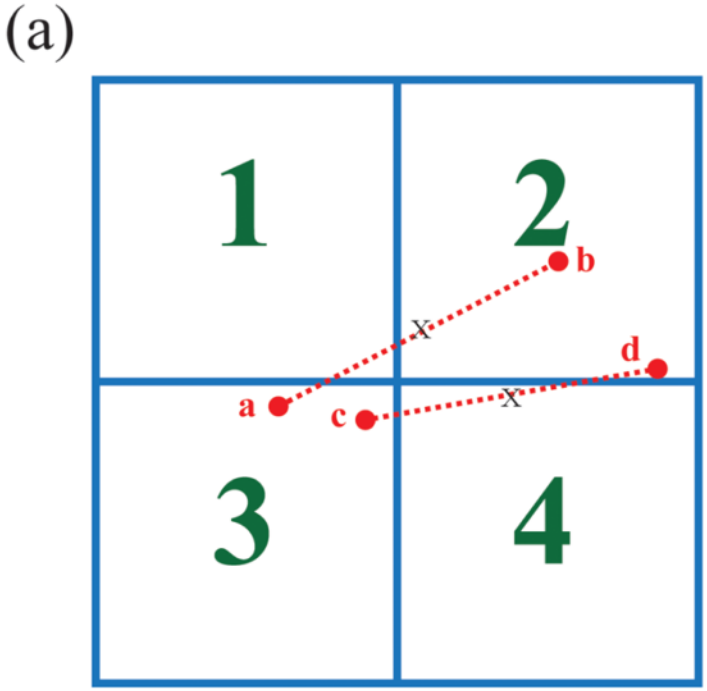
- Fast vibrational motions
- $\Delta t$ is 1~2 fsec = 1~2 x $10^{-15}$ sec

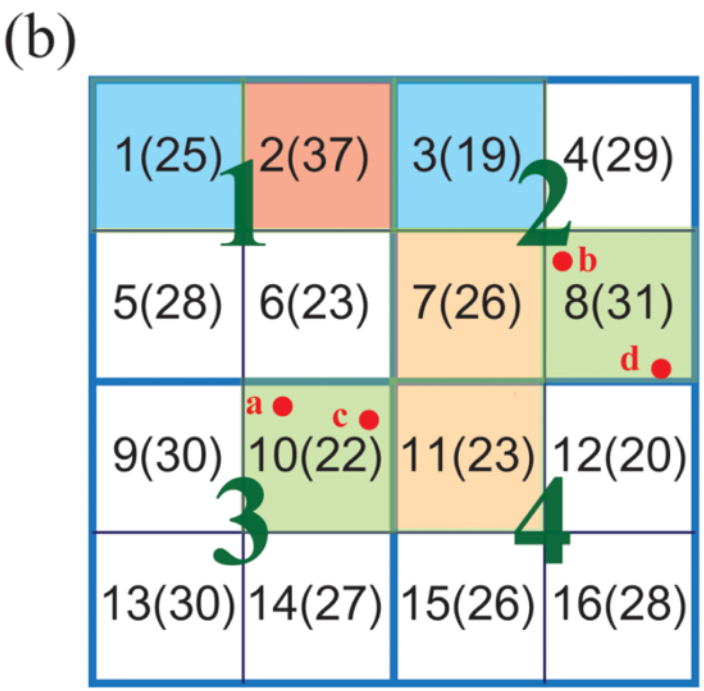
# How to accelerate MD ?

1) Large Number of Atoms
- At least $10^4$ atoms
- > 1M atoms, sometimes

## 1) Hybrid Parallelization

Hierarchical Space Decomposition ⇨ Each CPU computes Local Interaction




J. Jung
(CPR / R-CCS)

Nonbonded Interactions are evaluated in the midpoint cells.
→ Optimize the balance between computation and communication.

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

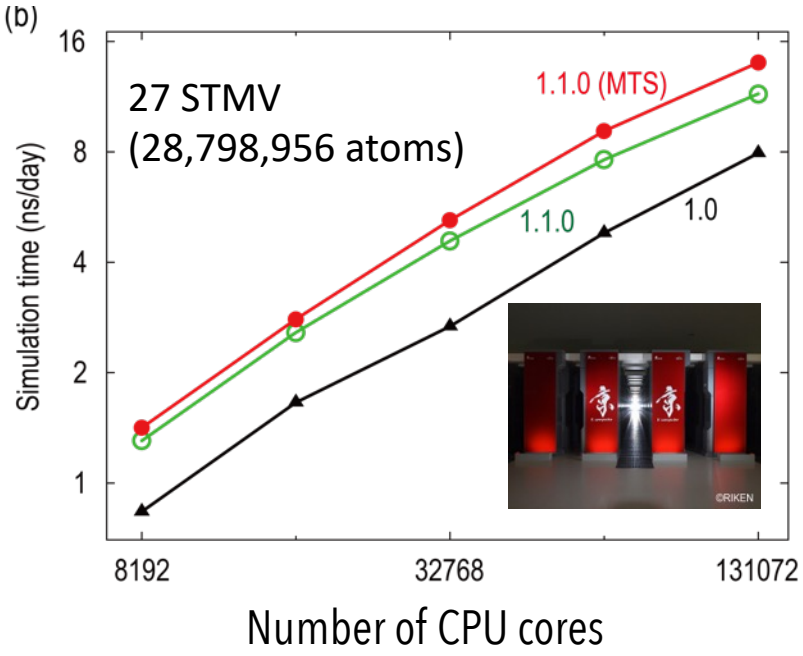
# How to accelerate MD ?

**1) Large Number of Atoms**
- At least $10^4$ atoms
- > 1M atoms, sometimes

## 1) Hybrid Parallelization

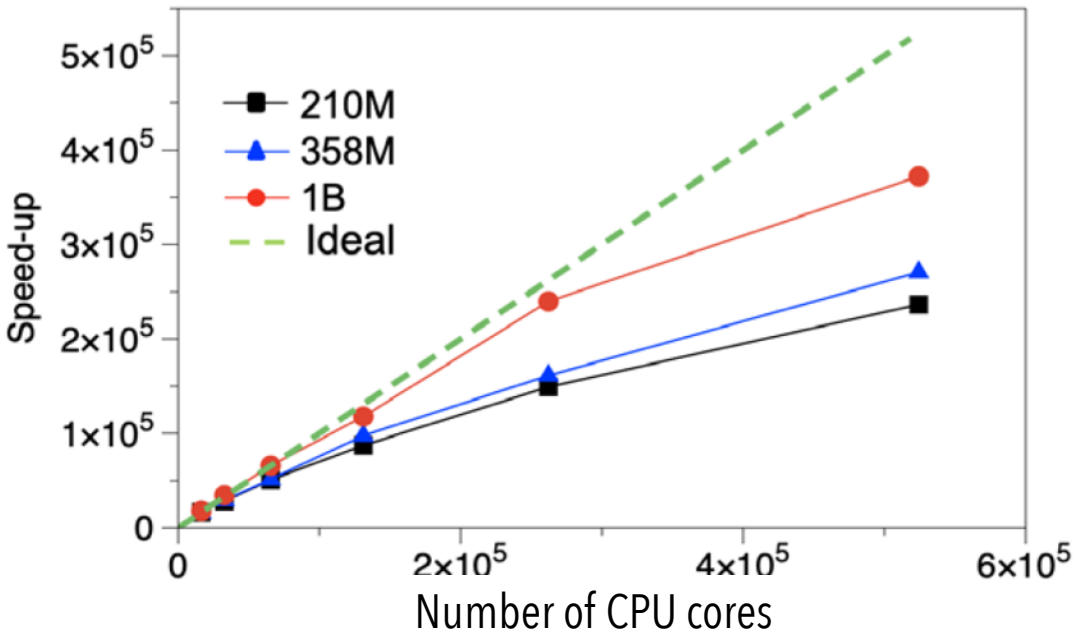
Hierarchical Space Decomposition ⇨ Each CPU computes Local Interaction

K computer @ RIKEN




C. Kobayashi, J. Jung et al. J. Comp. Chem. (2015)

Trinity @ LANL




J. Jung et al. J. Comp. Chem. (2019)

On K computer at RIKEN and Trinity at LANL, we could show good weak scaling of GENESIS.

# How to accelerate MD ?

## 2) Optimization of the program

### Optimized Nonbonded Kernel for each CPU

**Algorithm. Real-space non-bonded interaction kernel used for GENESIS 1.0-1.3**

```
do ij = 1, M
  icel = cell_index(1,ij)
  jcel = cell_index(2,ij)
  do i = 1, N(icel)
    force_temp(1:3) = 0.0
      do k = 1, Neighbor(i,ij)    Number of neighbors of i-th atm in icel-th cell
        j = Neighbor_list(k,i,ij) Neighbor of i-th atom in icel-th cell
        rij(1) = coord(1,i,icel)-coord(1,j,jcel)
        rij(2) = coord(2,i,icel)-coord(2,j,jcel)
        rij(3) = coord(3,i,icel)-coord(3,j,jcel)
        dij = sqrt(rij(1)^2+rij(2)^2+rij(3)^3)
        calculate f (1:3):force component from given distance
        force_temp(1) = force_temp(1) – f (1)
        force_temp(2) = force_temp(2) – f (2)
        force_temp(3) = force_temp(3) – f (3)
        force(1,j,jcel) = force(1,j,jcel) + f (1)
        force(2,j,jcel) = force(2,j,jcel) + f (2)
        force(3,j,jcel) = force(3,j,jcel) + f (3)
      end do
      force(1,i,icel) = force(1,i,icel) + force_temp(1)
      force(2,i,icel) = force(2,i,icel) + force_temp(2)
      force(3,i,icel) = force(3,i,icel) + force_temp(3)
   end do
end do
```

Nonbonded Energy Kernel in GENESIS 1.3

## 2) Nonbonded Interaction

- Nonbonded Interaction is the slowest calculations in Energy/Force Evaluations.

# How to accelerate MD ?

## 2) Optimization of the program

Optimized Nonbonded Kernel for each CPU

**Algorithm. Real-space non-bonded interaction kernel used for KNL**

```
do ij = 1, M
  icel = cell_index(1,ij)
  jcel = cell_index(2,ij)
  do i = 1, N(icel)
    if (Neighbor(i,ij) == 0) cycle
    force_temp(1:3) = 0.0
    do j = 1, N (jcel)
      rij(1) = coord(i,1,icel)-coord(j,1,jcel)
      rij(2) = coord(i,2,icel)-coord(j,2,jcel)
      rij(3) = coord(i,3,icel)-coord(j,3,jcel)
      dij = sqrt(rij(1)^2+rij(2)^2+rij(3)^3)
      calculate f (1:3):force component from given distance
      force_temp(1) = force_temp(1) – f (1)
      force_temp(2) = force_temp(2) – f (2)
      force_temp(3) = force_temp(3) – f (3)
      force(j,1,jcel) = force(j,1,jcel) + f (1)
      force(j,2,jcel) = force(j,2,jcel) + f (2)
      force(j,3,jcel) = force(j,3,jcel) + f (3)
    end do
  force(i,1,icel) = force(i,1,icel) + force_temp(1)
  force(i,2,icel) = force(i,2,icel) + force_temp(2)
  force(i,3,icel) = force(i,3,icel) + force_temp(3)
  end do
end do
```

Nonbonded Energy Kernel for Intel PHI

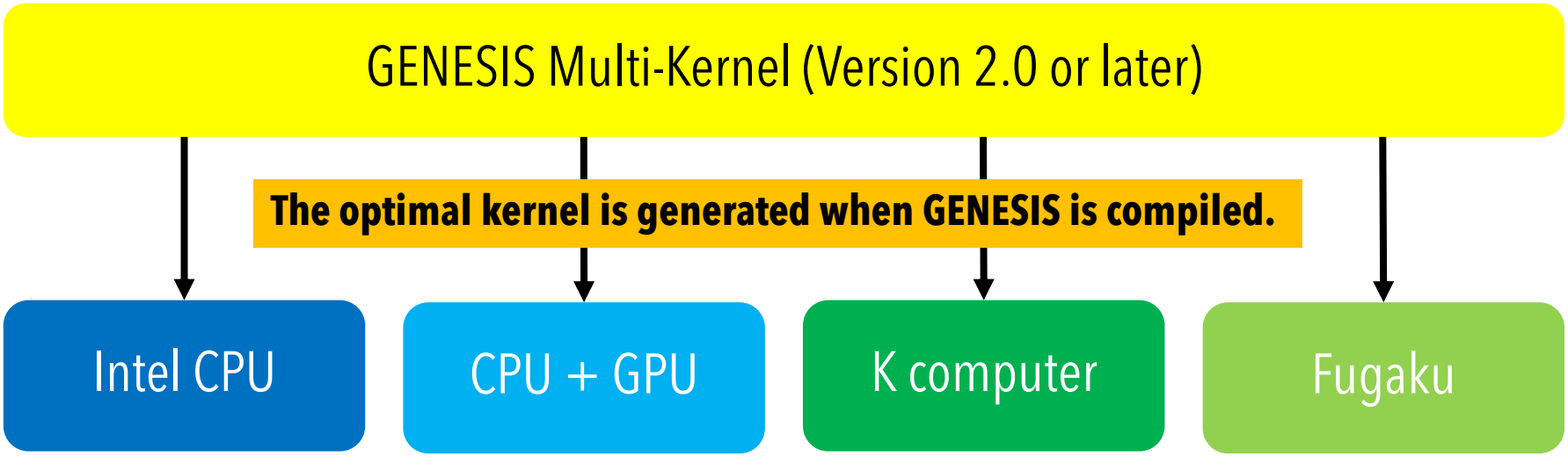
## 2) Nonbonded Interaction

- Nonbonded Interaction is the slowest calculations in Energy/Force Evaluations.

# How to accelerate MD ?

## 2) Optimization of the program

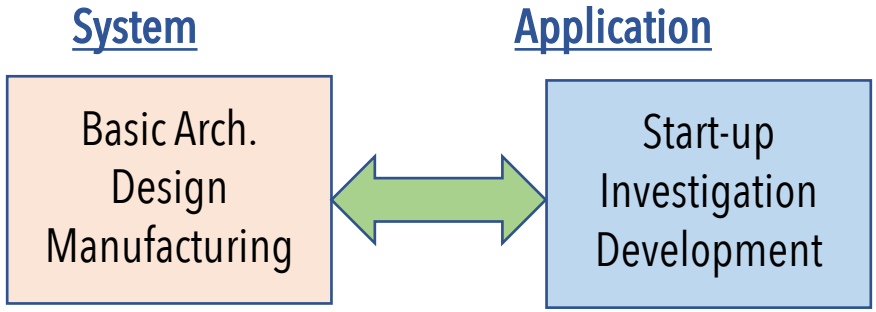
Optimized Nonbonded Kernel for each CPU




## 2) Nonbonded Interaction

- Nonbonded Interaction is the slowest calculations in Energy/Force Evaluations.

## FUGAKU Co-design




## From K to Fugaku

11.3 peta-flops X 35 400 peta-flops

GENESIS 1.3.0 → GENESIS 2.0

**x 125 ($\delta t$=2.5fs)**

# New Methods Implemented in GENESIS

- Solve the system-size problem
  - Inverse Lookup Table: Jung et al. JCC 34, 2414-2420 (2013).
  - Midpoint Cell Method: Jung et al. JCC 35, 1064-1072 (2014).
  - Volumetric 3D FFT: Jung et al. CPC 200, 57-65 (2016).
  - GPU parallelization: Jung et al. JCTC 12, 4947-4958 (2016).
  - Multiple program/multiple data: Jung et al. JCC 38, 1410-1417 (2017).
  - Kinetic energy definition: Jung et al. JCP 148, 164109 (2018).
  - Optimal temperature: Jung et al. JCTC 15, 84-94 (2019).
  - KNL parallelization: Jung et al. JCC 40, 1919-1930 (2019).

J. Jung
(RIKEN)


- Solve the time-scale problem
  - Reaction Path Method: Matsunaga et al. JPCLett 7, 1446-1451 (2016).
  - Domain Motion Enhanced model: Kobayashi et al. JPCB 119, 14584-14593 (2016).
  - RSE-MTD: Galvelis et al. JCC 36, 1446-1455 (2015) ; Galvelis et al. JCTC 13, 1934-1942 (2017).
  - gREST, gREST/REUS: Kamiya et al. JCP 149, 072304 (2018).
  - GaREUS: Oshima et al. JCTC in revision
- Apply biological problem
  - QM/MM: Yagi et al. JCTC 15, 1924-1938 (2019).
  - Cryo-EM flexible fitting: Mori et al. Structure 27, 161-174 (2019).

Fugaku is a nick name of Mt. Fuji

# GENESIS for High-Performance MD

## Efficient Weak Scaling for Supercomputers w/o GPU

GENESIS
Generalized-ensemble simulation system

**Leader:** Y. Sugita

**Current main developers:**
J. Jung, T. Mori,
C. Kobayashi,, Y. Matsunaga,
H. Oshima, K. Kasahara,
K. Yagi

This is free software under GPLv2 License.
https://www.r-ccs.riken.jp/labs/cbrt/

**K computer (2011 – 2019)**

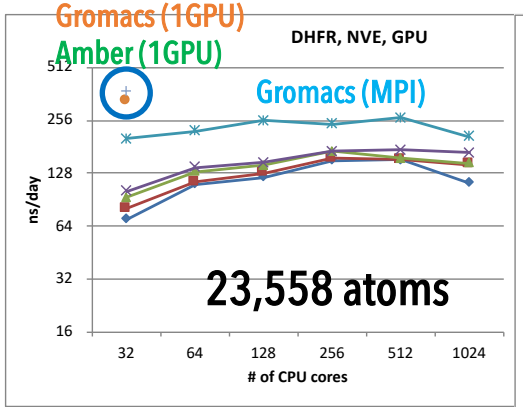

**Post K computer (Fugaku) (2021 – )**

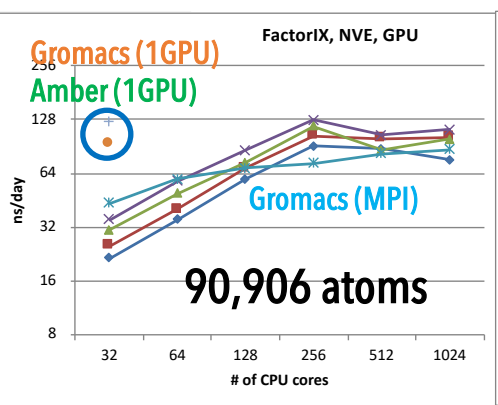

11.3 peta-flops
GENESIS 1.3.0

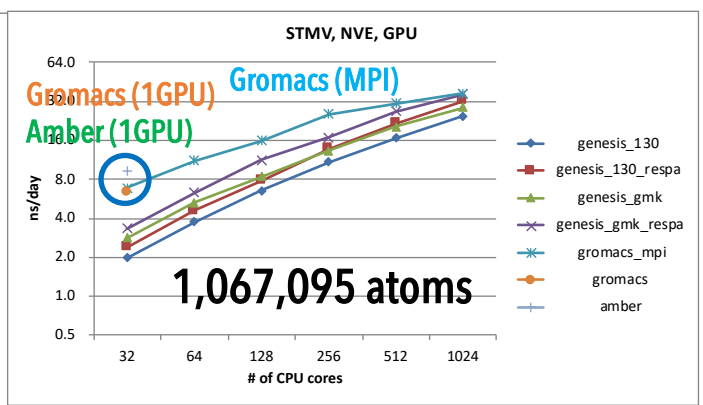
X 35
x 125

400 peta-flops
GENESIS 2.0

**GENESIS Benchmark on Infini-band PC-cluster (Intel Gold 6142 2.6 GHz 32 core, GeForce GTX-1080 Ti)**



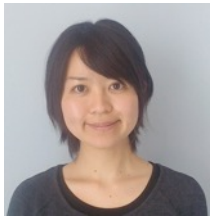
# To Accelerate Biological MD Simulations

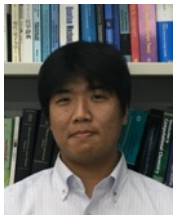
- Accelerate MD simulation
  - Why MD is so slow ?
  - How to accelerate MD ?

- Efficient MD algorithms
  - Enhanced Sampling Methods
  - Protein-Ligand Binding Simulation


S. Re
(BDR)

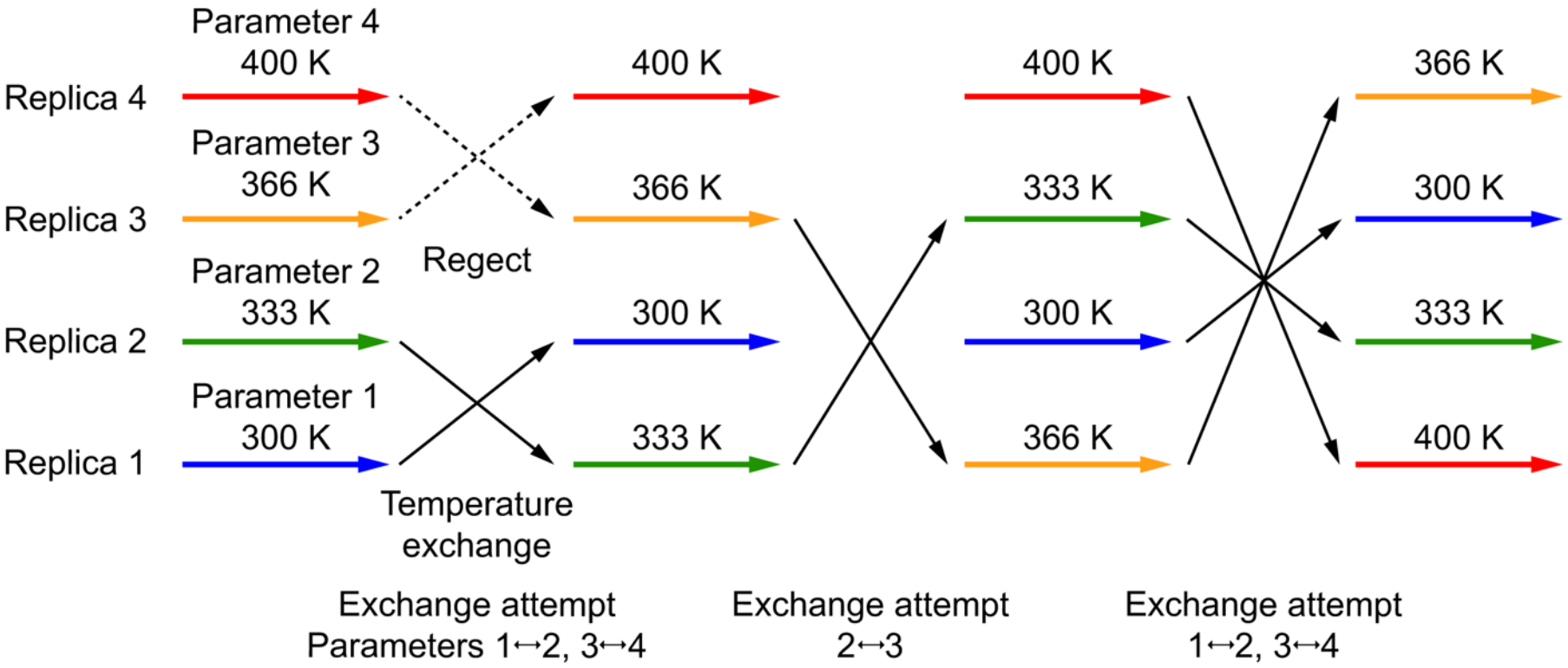

H. Oshima
(BDR)


A. Niitsu
(CPR)


K. Kasahara
(BDR)

# Replica-Exchange MD (REMD)

Upon the exchange of temperatures 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)

# generalized REST (gREST)

## for protein-folding dynamics

M. Kamiya (RIKEN → IMS)

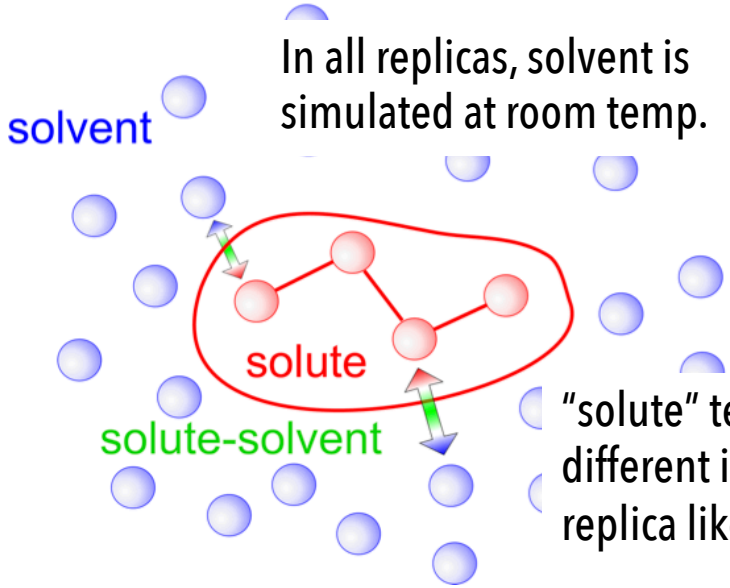
**Replica-Exchange with Solute Tempering (REST/REST2):**

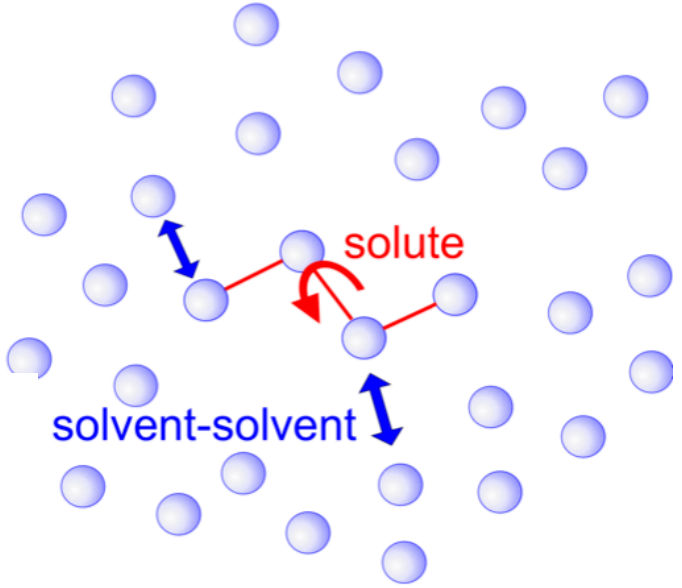
T. Terakawa et al. J. Comput. Chem. 32:1228-1234 (2011),
S. L. C. Moors et al. J. Chem. Theory Comput. 7:231-237 (2011),
L. Wang et al. J. Phys. Chem. B 115:9431-9438 (2011).

**generalized REST (gREST)**

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

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




gREST can define the solute region in a more flexible manner.
The solute region is selected as a part of a molecule or a part of potential energy function.

# generalized REST (gREST)
## for Protein-Ligand Binding

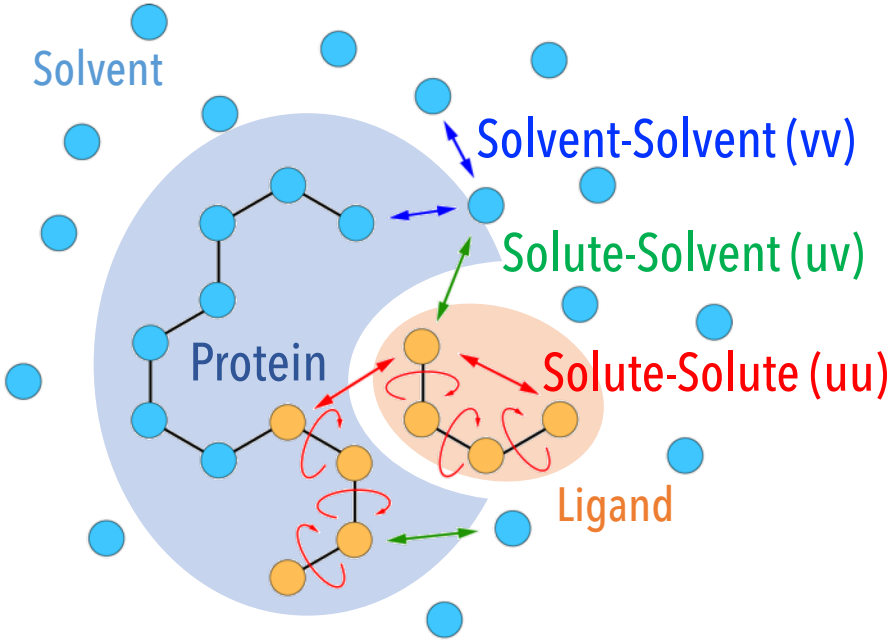
S. Re (RIKEN)

### generalized REST (gREST)

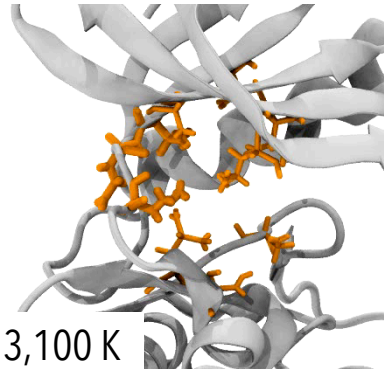
M. Kamiya and Y. Sugita, J. Chem. Phys. 149, 072304 (2018)

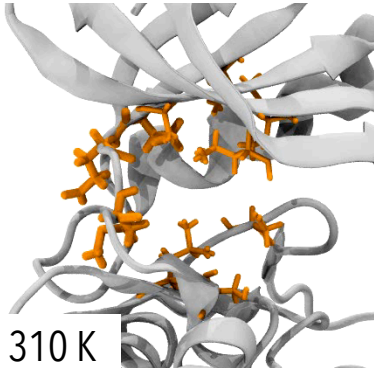
$$E_{\mathrm{gREST}} = \frac{\beta_m}{\beta_0} E_{uu} + \left(\frac{\beta_m}{\beta_0}\right)^{1/n} E_{uv} + E_{vv}$$

### Solute = Ligand + Sidechains at the Binding Sites

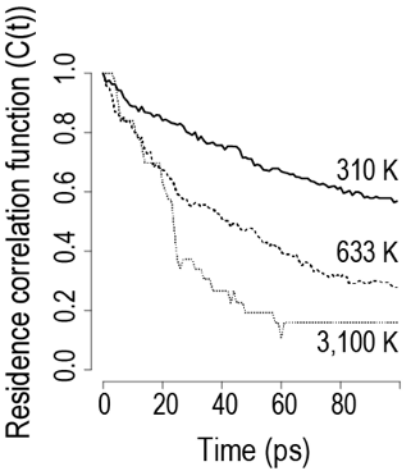


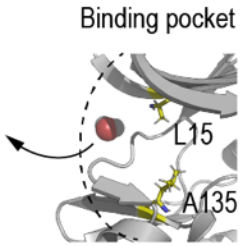

### Sidechain motion is enhanced in gREST



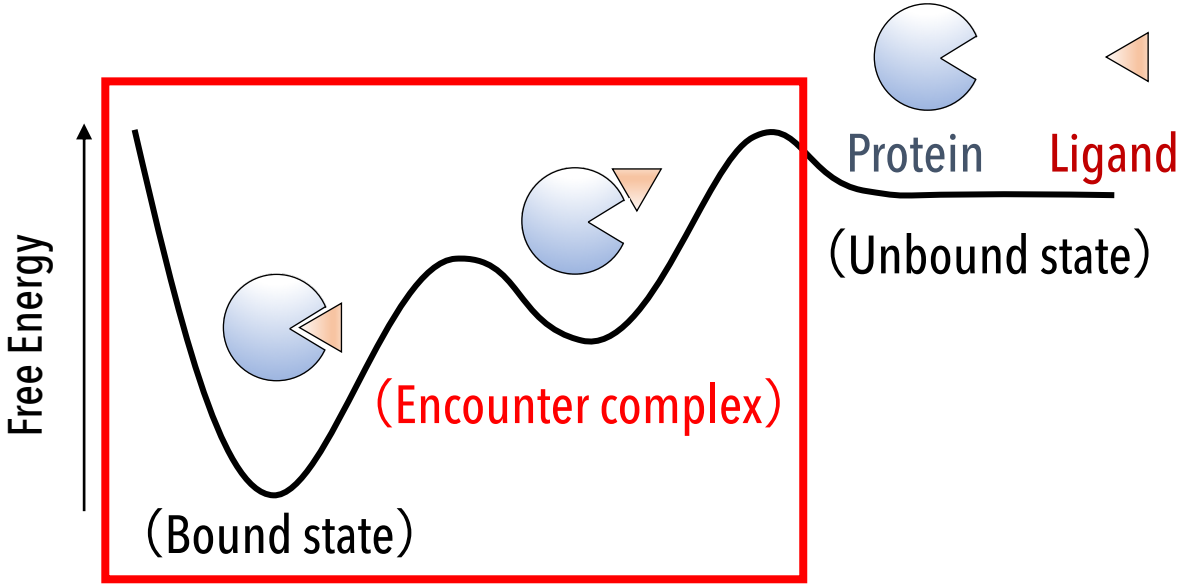

### Water dynamics is accelerated in gREST



# Mechanisms for Protein-Ligand Binding

Binding of a ligand to a protein can be understood via three different states, namely, Bound, Unbound, and Encounter complex.

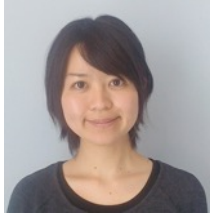



Many questions to be addressed by theoretical chemistry

- Prediction of the Bound and Encounter complex states
- Binding pathways or free-energy landscapes
- Effect of protein structural flexibility
- Kinetics of binding processes

# Binding Prediction with gREST

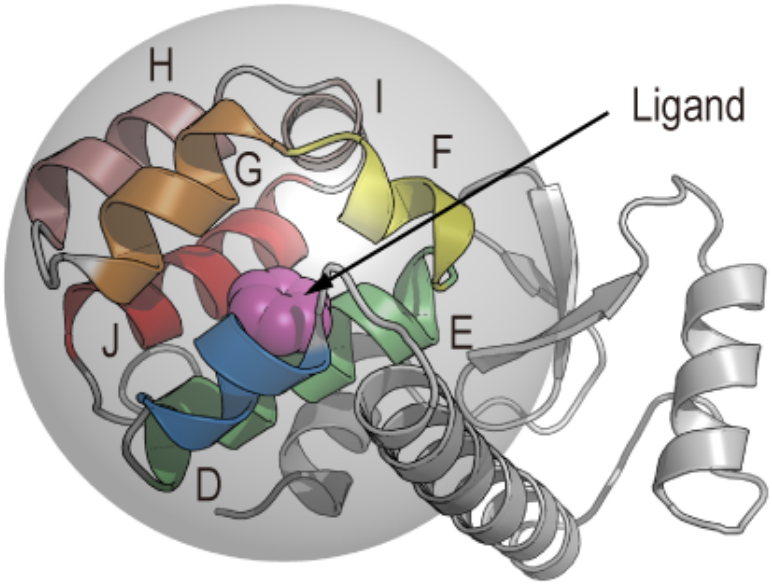
## Less expensive, straight forward approach



A. Niitsu (RIKEN)

S. Re (RIKEN)

L99A T4 Lysozyme (T4L L99A)




**Solute region:**
Ligand+Helix D, E, F and G
Dihedral, LJ and CMAP terms
8 replicas (300 K~520 K)
300 ns / replica
**2.4 μs in total (300 ns x 8 replicas)**

**Flat bottom restraint potential:**
Center of pocket – Ligand COM
Reference distance: 15 Å
Force constant: 1 kcal/mol/Å$^2$

L99A T4 Lysozyme

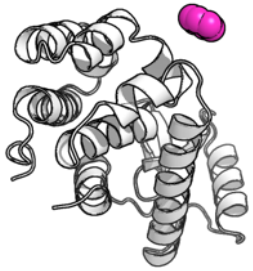
A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

- A well studied benchmark/test system.
- Recently revisited for understanding the conformation-binding coupling.

e.g. G. Bouvignies et al (2011) Nature 477:111–117, M. Merski et al (2015) PNAS 112:5039–5044. A. Nunes-Alves et al (2018) Biophys J 114:1058–1066. Y. M. Huang et al (2018) J Chem Theory Comput 14:1853–1864. Y. Wang et al (2016) Elife 5:1–35. P. Vallurupalli et al (2016) Chem Sci 7:3602–3613. J. M. Schiffer et al (2016) Biophys J 111:1631–1640.

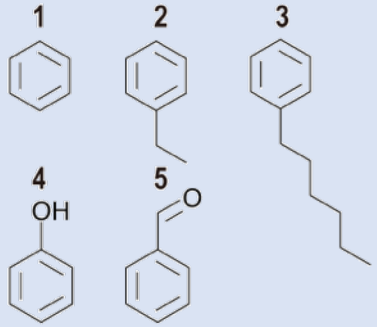
# Blind Prediction of Five Ligands

## A distinction between binders and non-binders possible?

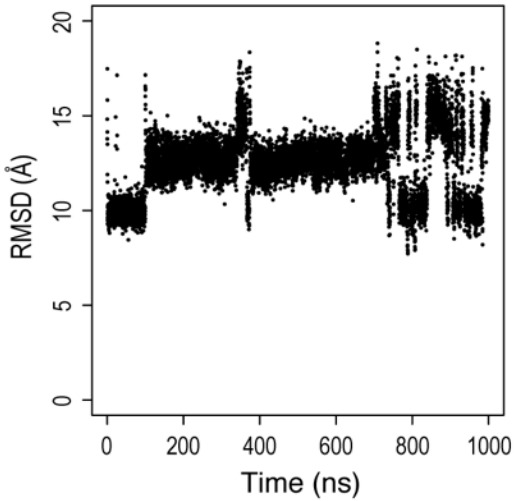


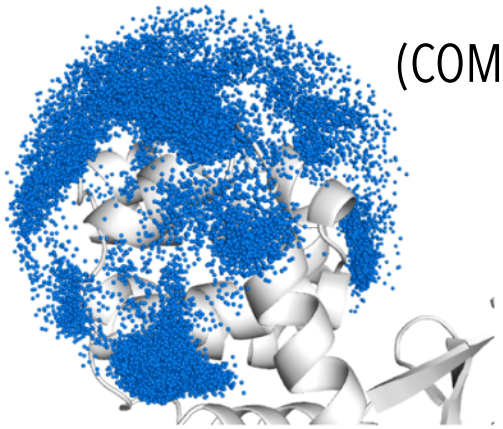
(initial)

| | | |
|---|---|---|
| Binder | Benzene (**1**) | |
| | Ethyl benzene (**2**) | |
| | *n*-hexylbenzene (**3**) | |
| Non-binder | Phenol (**4**) | |
| | Benzaldehyde (**5**) | |




No binding event is observed for 1 μs conventional MD of benzene



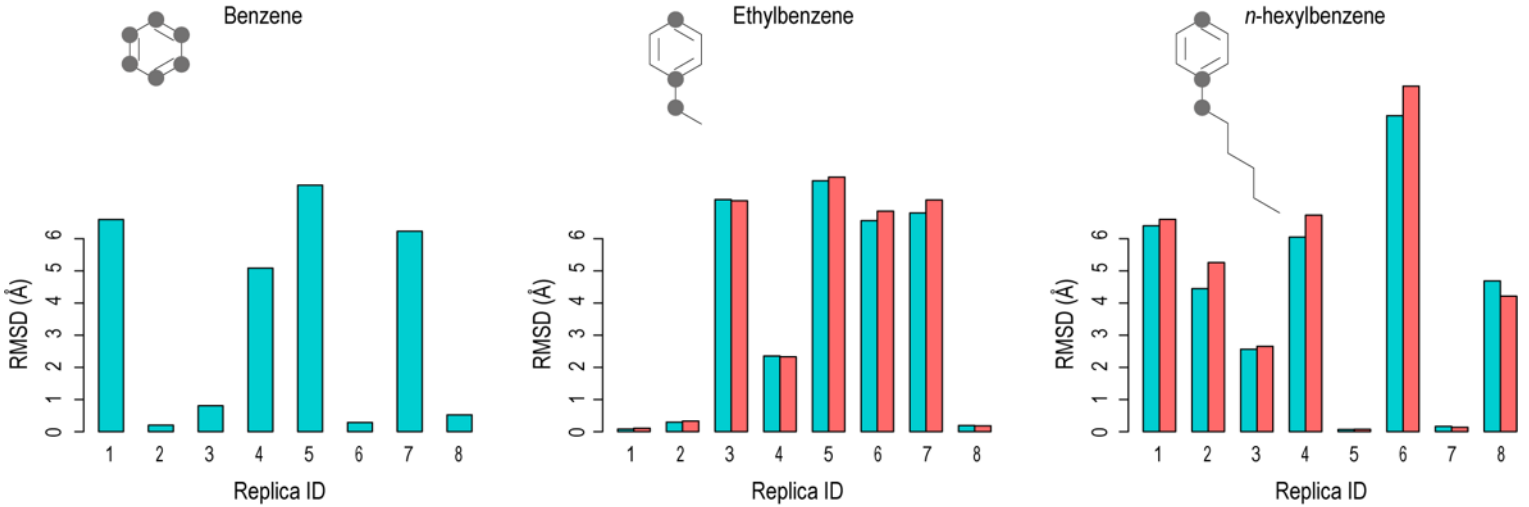



(COMs of ligand)

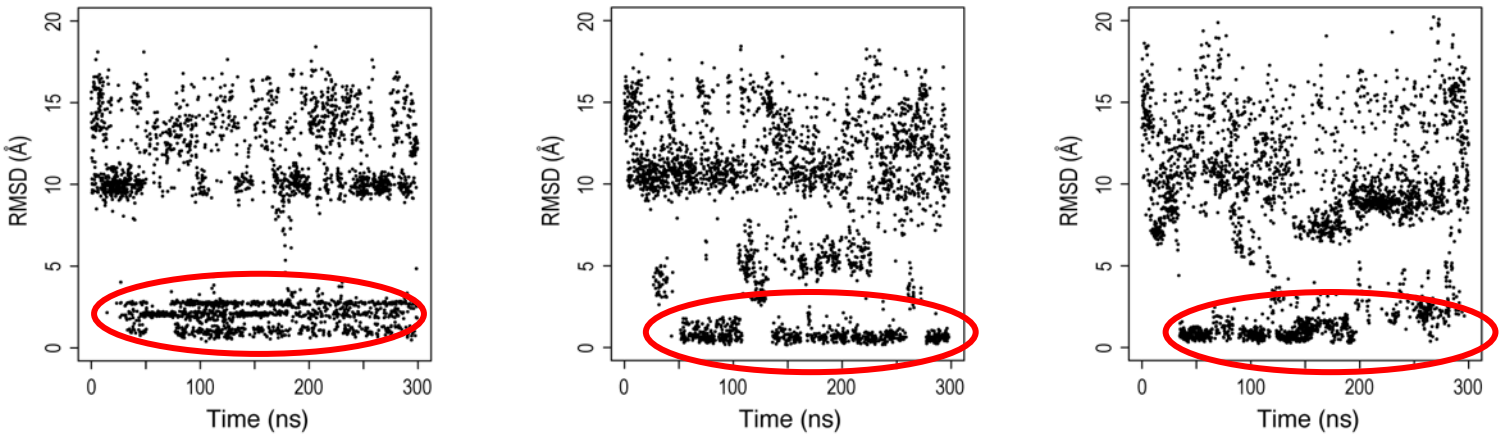
A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

# Prediction for Binders

## Multiple replicas find "native" pose



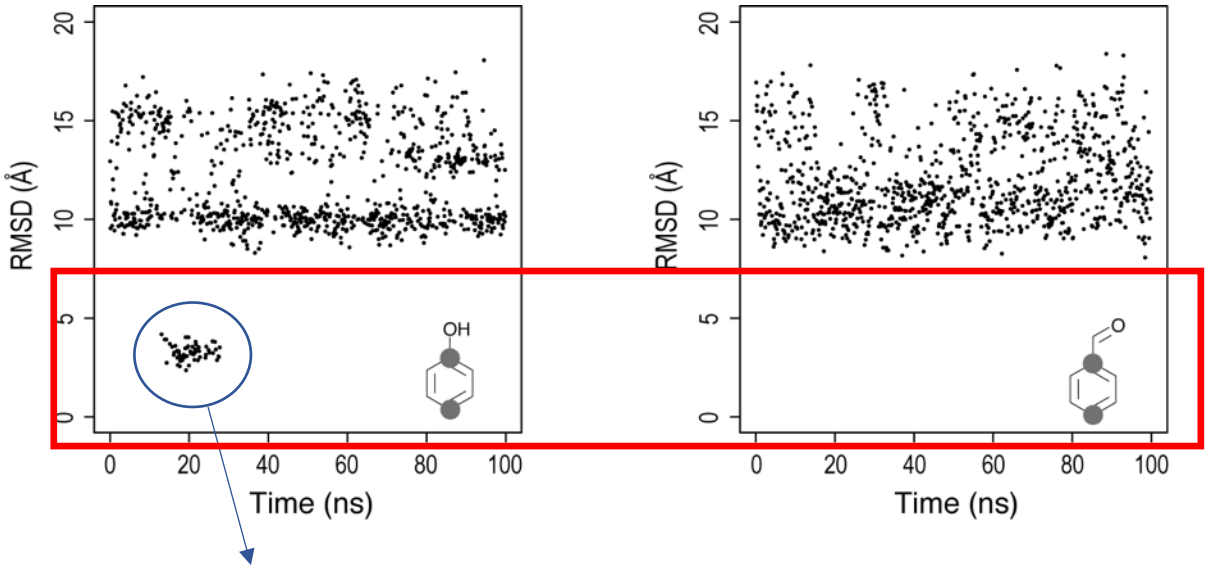

(300 K)

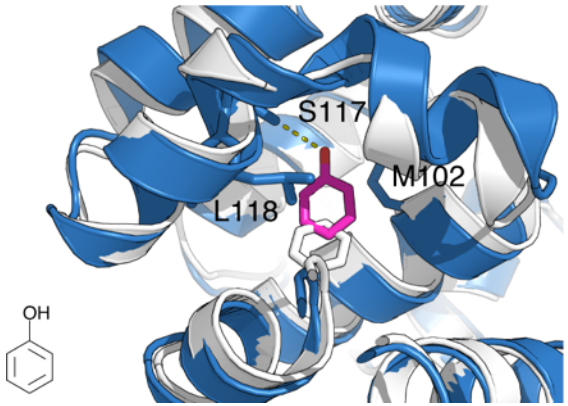



A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

# Prediction for Non-Binders

## Ligands rarely enter the cavity



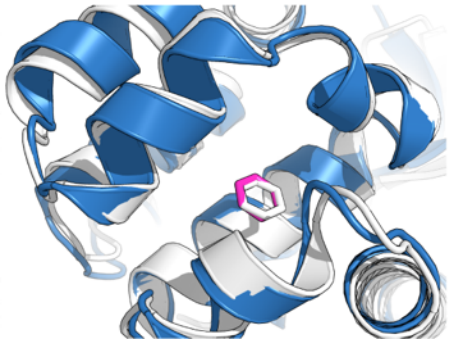

(Comparison with the X-ray structure for the benzene)

A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

# Predicted Binding Poses

## X-ray structures are accurately reproduced

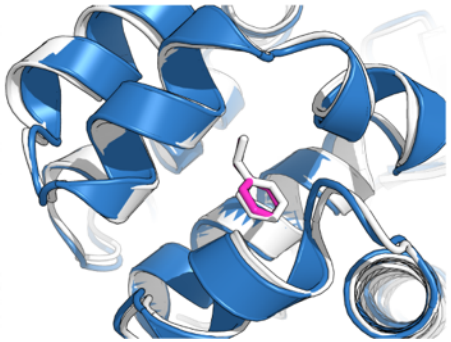
Benzene



X-ray structure
(PDB ID: 181L)

RMSD = 0.25 Å

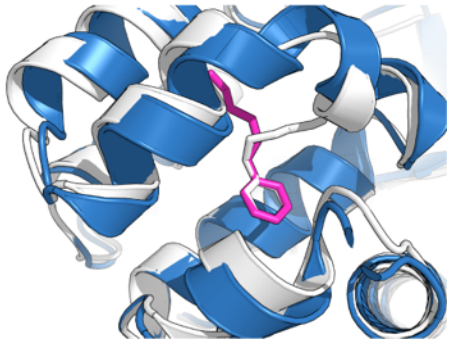
Ethylbenzene



X-ray structure
(PDB ID: 4W54)

RMSD = 0.11 Å

*n*-hexylbenzene



X-ray structure
(PDB ID: 4W59)

RMSD = 0.07 Å

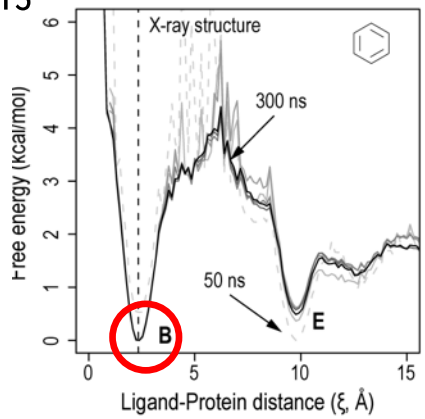
Blue and magenta represent structures with the smallest RMSDs

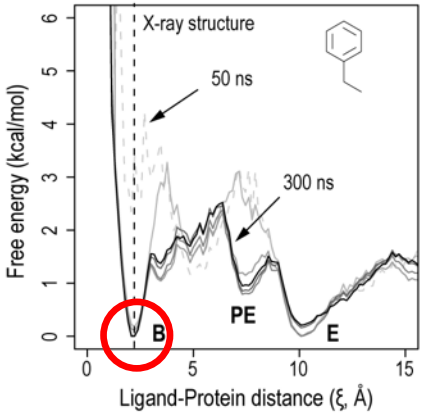
A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

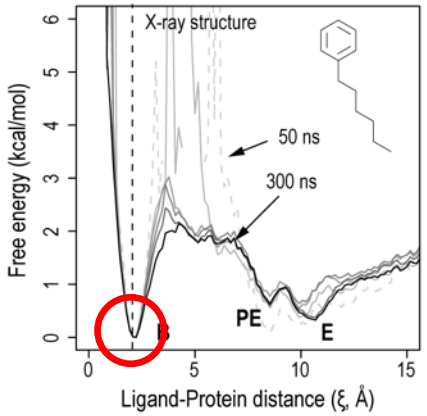
# Distinction of Binders and Non-Binders Based on Free-Energy Profiles

Free-energy profiles at 300 K obtained from MBAR analysis

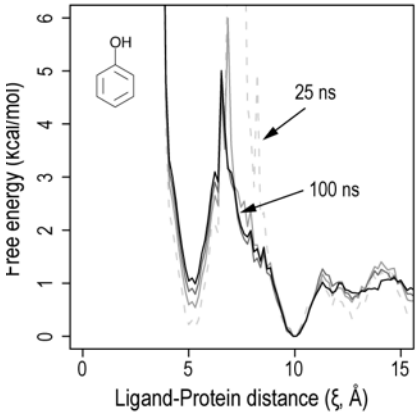
Binders

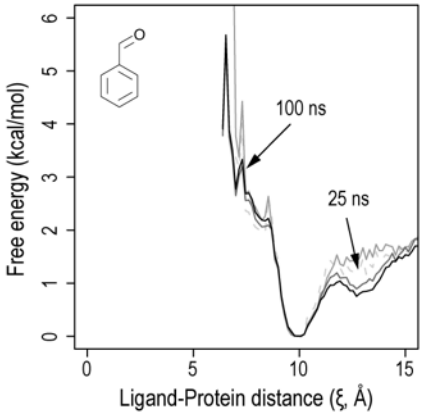


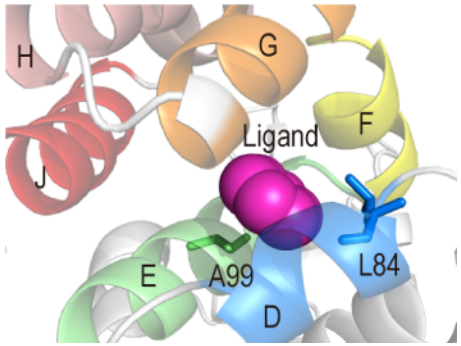




Non-binders



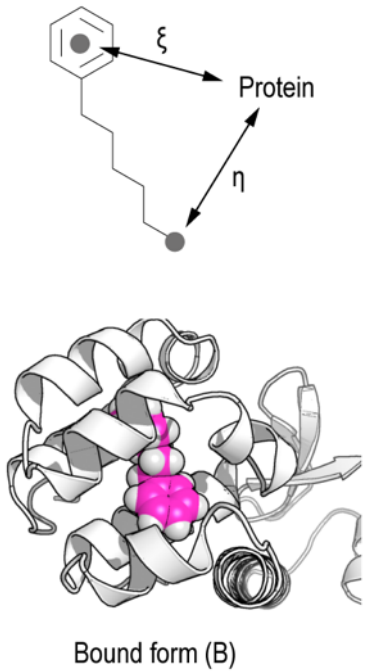

Ligand-protein distance



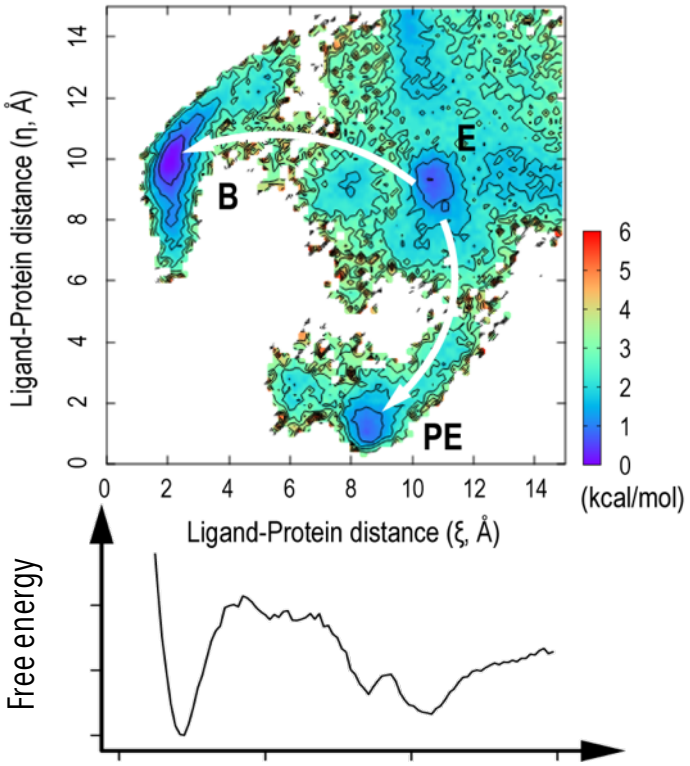
# Binding Pathways and Intermediates

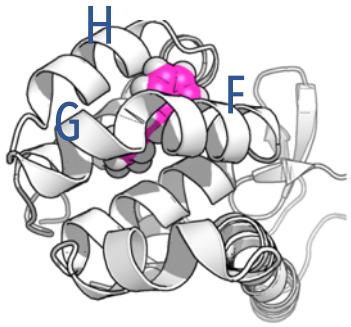
## Favors FGH path, Encounter complex is important

Free-energy landscapes at 300 K obtained from MBAR analysis

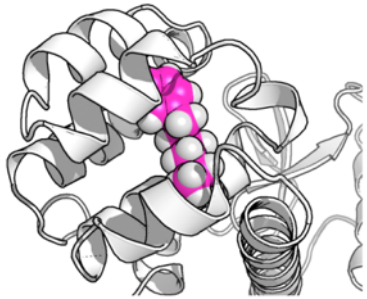



Bound form (B)




Encounter complex (E)



Post Encounter complex (PE)

A. Niitsu, S. Re , Oshima, Kamiya, Sugita J. Chem. Info. Model. (2019) in press.

# Summary

- In GENESIS, we implemented efficient hybrid parallelization and optimization for each CPU architecture.
- The optimal temperature definition allows us to extend the time-step in the time integration.
- Enhanced conformational sampling method fills the gap between simulation and experiment.
- Using GENESIS on FUGAKU, we study cellular-scale biology and drug discovery in collaboration with experimental scientists.

# Acknowledgement


Ligand-Binding Simulations
Dr. Suyong Re (RIKEN BDR)
Dr. Hiraku Oshima (RIKEN BDR)
Dr. Kento Kasahara (RIKEN BDR)
Dr. Motoshi Kamiya (RIKEN → IMS)
Dr. Ai Niitsu (RIKEN CPR)

GENESIS developments
Dr. Jung Jaewoon (RIKEN CPR/R-CCS)
Dr. Chigusa Kobayashi (RIKEN R-CCS)
Dr. Takaharu Mori (RIKEN CPR)
Dr. Kiyoshi Yagi (RIKEN CPR)
Prof. Yasuhiro Matsunaga (RIKEN → Saitama Univ.)

Simulations in Crowded Cellular Environments
Prof. Isseki Yu (RIKEN → Maebashi Inst. of Tech.)
Prof. Tadashi Ando (RIKEN → Tokyo University of Science)
Prof. Ryuhei Harada (RIKEN → Tsukuba Univ.)
Prof. Michael Feig (MSU, USA)
Dr. Greg. Nawrocki (MSU, USA)
Prof. Wonpil Im (Lehigh Univ., USA)
Dr. Karissa Sanbonmatsu (LANL, USA)

Fundings (to YS)
RIKEN R-CCS, BDR, iTHES
RIKEN Pioneering Projects
MEXT KAKENHI 26119006
HPCI strategic research project 1
post-K priority issue 1
JST CREST

Computational resources: K computer, FX10 in University of Tokyo by HPCI system, and HOKUSAI by RIKEN Advanced Center for Computing and Communication.