



GENESIS Hands-on Part 2-1

Coarse-Grained Simulations in GENESIS

Cheng Tan

RIKEN Center for Computational Science

2024/07/01

IUPAB2024 Hands-on Training Program | CHARMM-GUI/GENESIS MD Tutorial

Schedule of GENESIS Tutorials (6/30-7/2)



GENESIS
Generalized-ensemble simulation system

06/30 Day 1	
13:30 – 15:00	GENESIS basics and GENESIS on Fugaku (Kobayashi)
	Lecture
	Hands-on tutorial on Fugaku
07/01 Day 2	
● 14:30 – 15:30	Coarse-grained simulations in GENESIS (Tan)
15:30 – 16:30	High-performance computation with GENESIS (Jung)
07/02 Day 3	
13:30 – 15:00	Generalized-ensemble simulations using GENESIS (Ito)



● **Part 2-1-I: Coarse-Grained (CG) Models in GENESIS**

- Introduction to CG
- Popular CG Models
- CG MD in GENESIS

Part 2-1-II: GENESIS CG Simulations Hands-on

- GENESIS-cg-tool installation
- Protein folding simulation
- Simple CG MD data analysis

CG: Fewer Details in Picturing Molecules

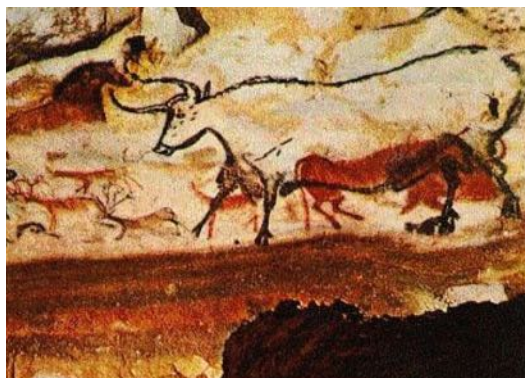
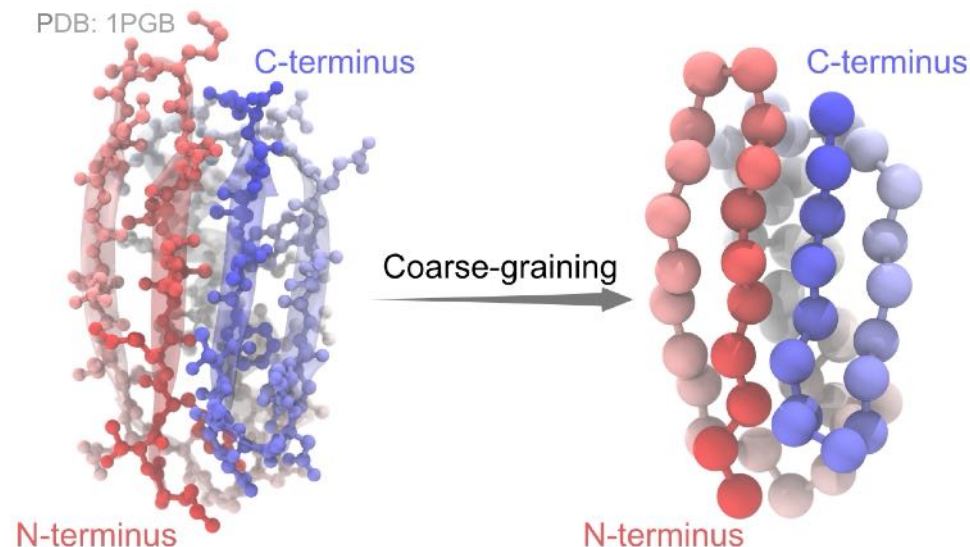


GENESIS
Generalized-ensemble simulation system

FROM THE PREHISTORIC DRAWINGS ON THE CAVE WALLS OF Lascaux to the paintings hanging in the Louvre in the late nineteenth century, the visual arts seemed to progress monotonically by adding ever more detail seemingly to give greater realism. A new movement then exploded near the turn of the twentieth century. Artists began to explore new ways of expressing truths about the visual world and the visual experience. These new ways involved generally more abstract and superficially less detailed techniques than those taught in the art schools. They required no less intellectual discipline, however. The new “modern art,” while not to everyone’s taste, nevertheless, in the opinion of many, brought into existence creations of great beauty and insight.

Peter Wolynes, 2018

「Coarse-Grained Modeling of Biomolecules」



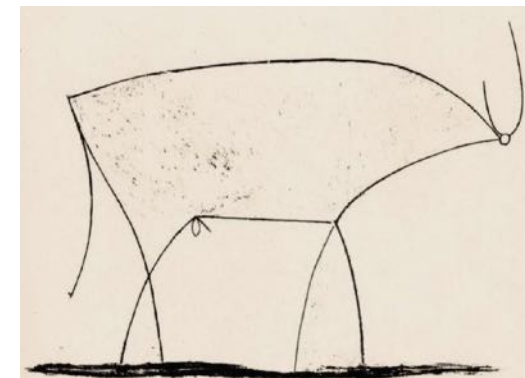
???

Anonymous, ~15000 BC



The Young Bull

Paulus Potter, 1647



The Bull

Pablo Picasso, 1945

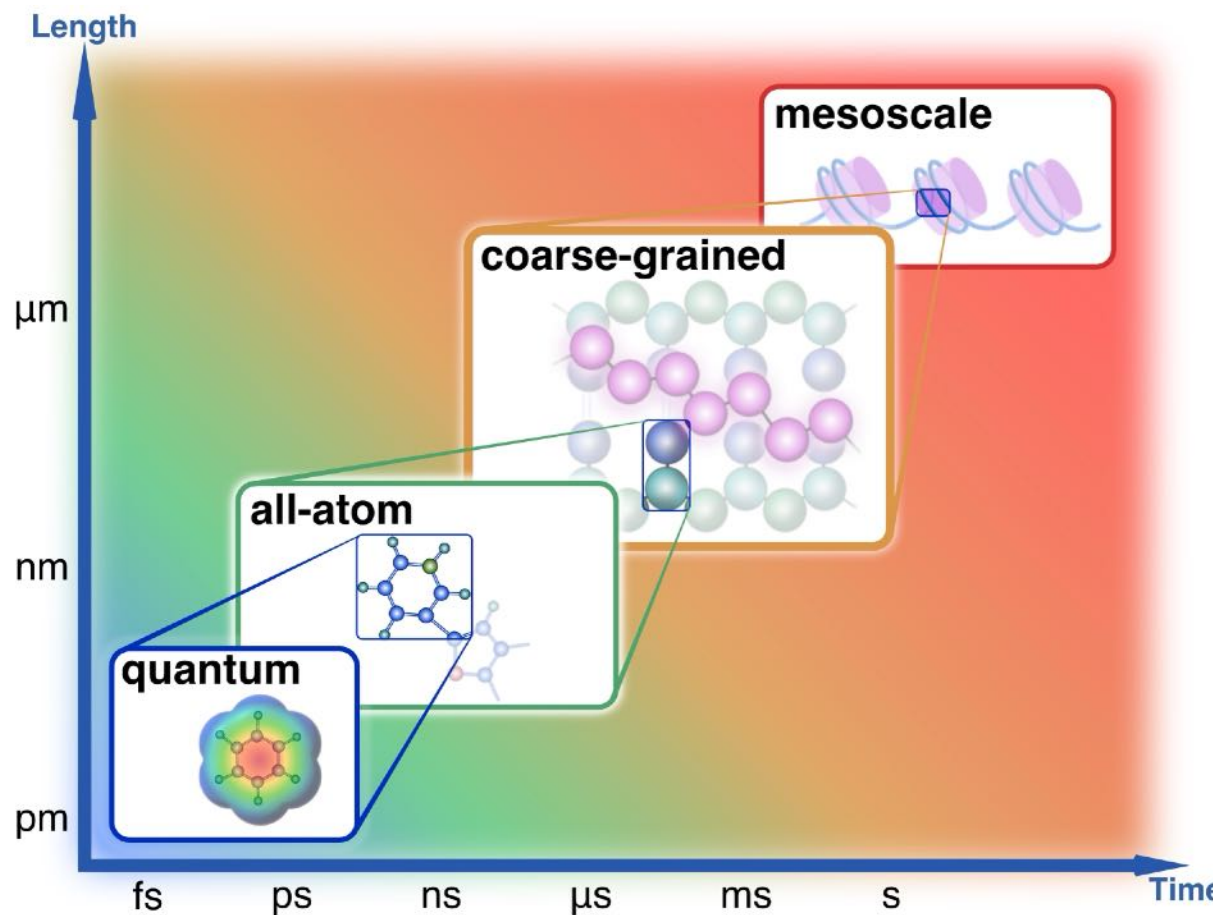
CG MD Simulations



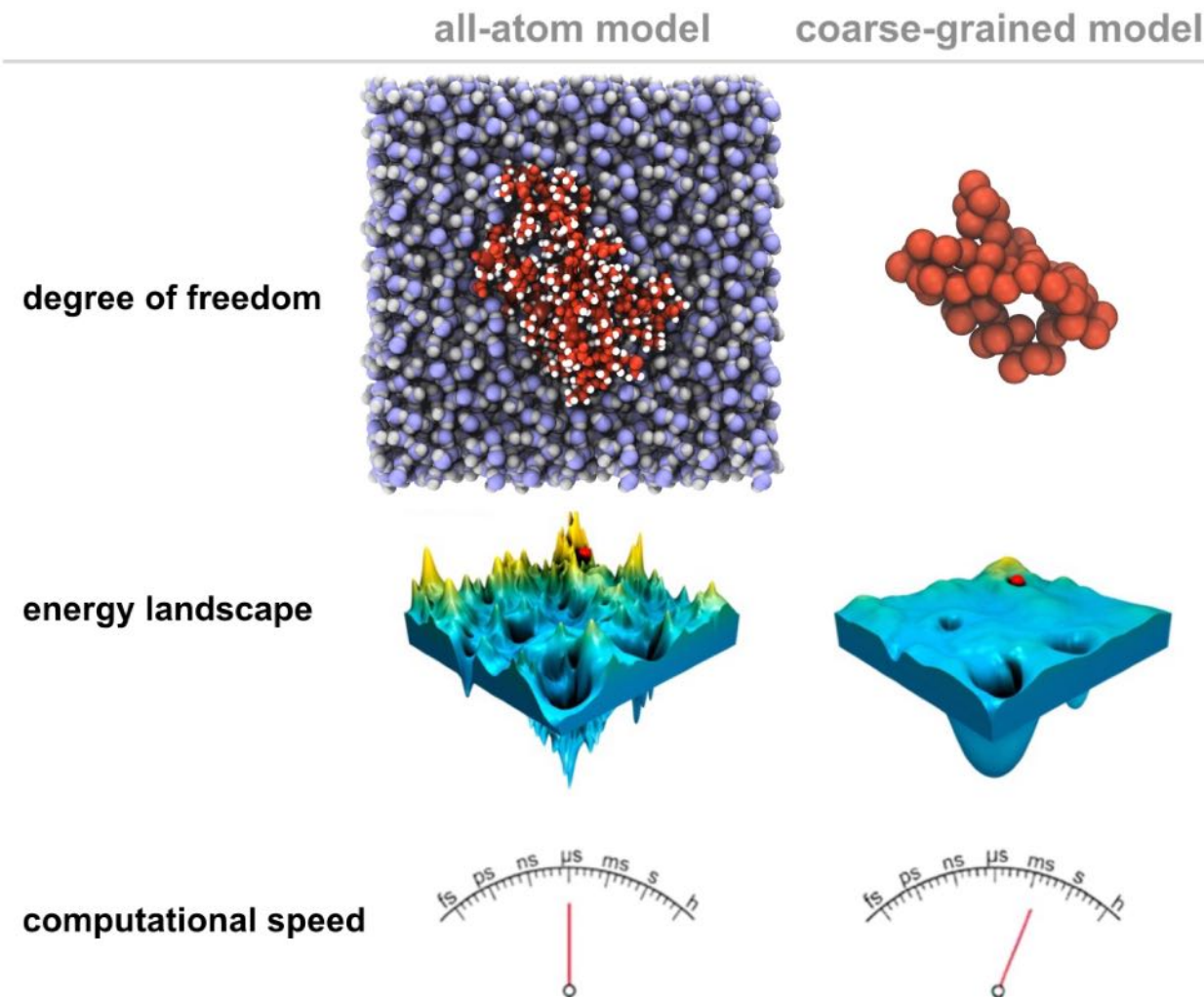
GENESIS
Generalized-ensemble simulation system

Task of MD: numerically solving Newton's equations of motion

$$M\ddot{X}(t) = -\nabla U(X) + f(X, \dot{X})$$



Adapted from S Kmiecik *et al.* 2016, *Chem. Rev.*



S Kmiecik *et al.* 2016, *Chem. Rev.*;

S Takada *et al.* 2015 *Accounts Chem. Res.*

Classification of CG Models



GENESIS
Generalized-ensemble simulation system

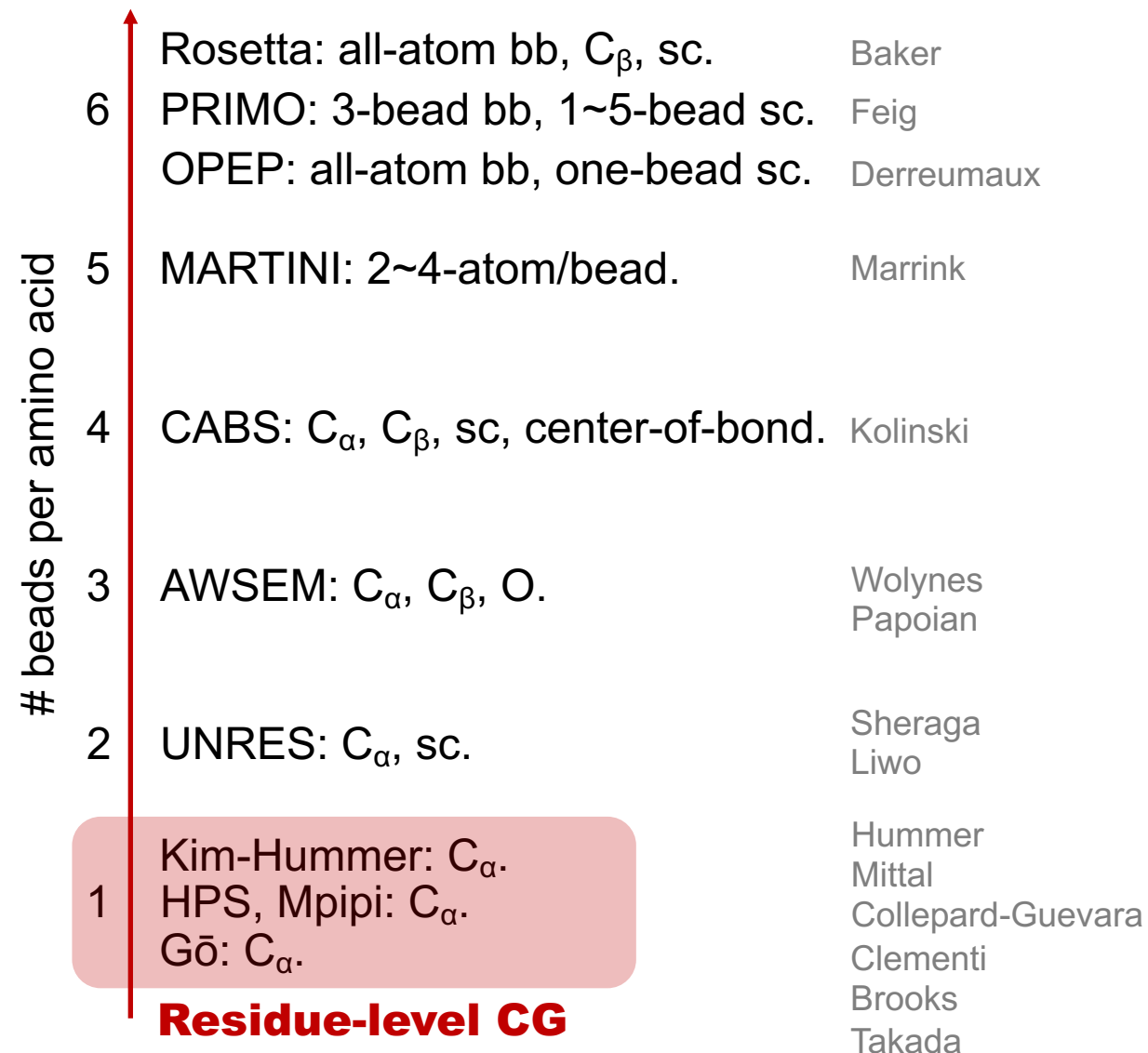
Coarse-graining approaches:

- Bottom-up
- Top-down

Modeling strategies:

- Dynamics-based (physics-based)
Systematic algorithms of AA \rightarrow CG mapping
 - Boltzmann Inversion
 - Inverse Monte Carlo
- Knowledge-based
Fitting experimental macroscopic properties
 - Miyazawa-Jernigan
 - Kim-Hummer
- Structure-based
Assumption of consistency principle
 - Network models
 - Gō models

Resolution



Residue-Level CG Models

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*.

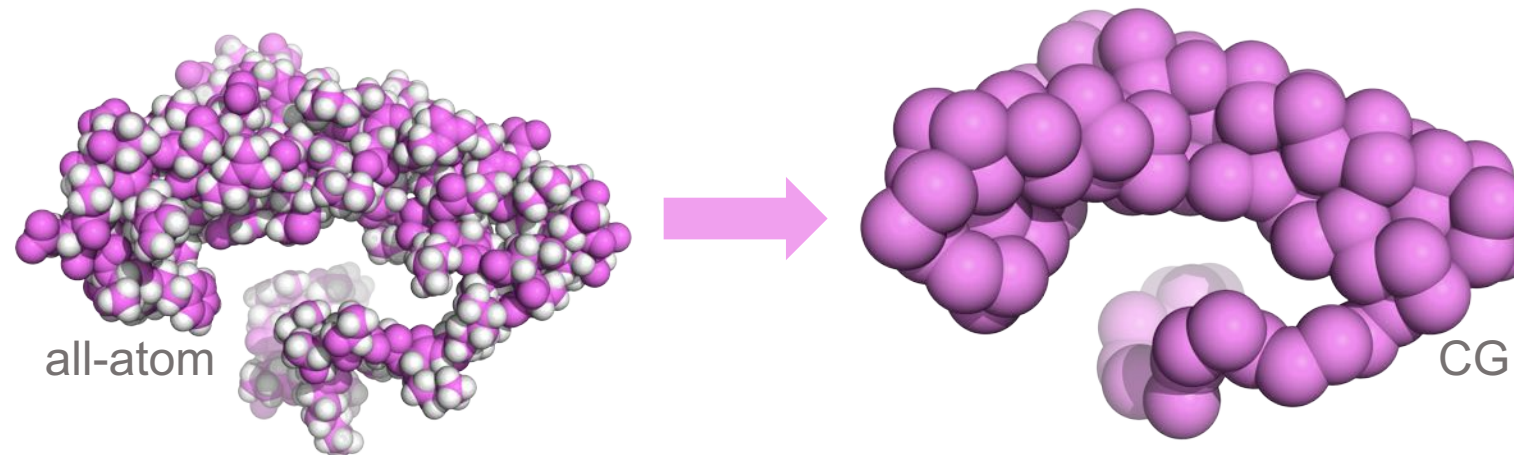
- RNA: Gō-like

N. Hori *et al.* 2012, *JCTC*.

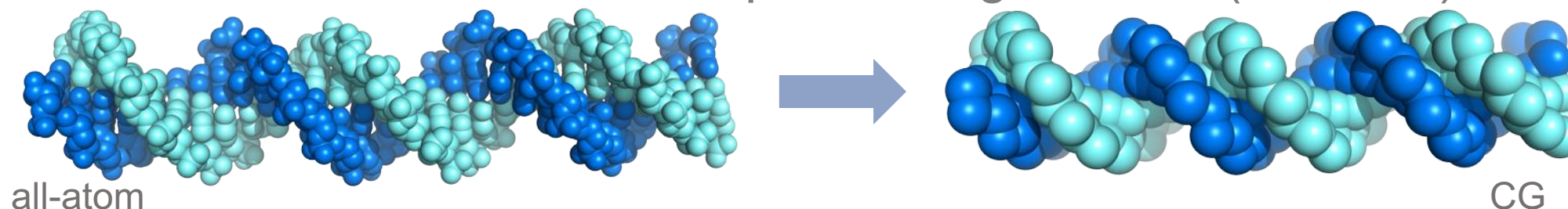
- IDP: HPS

Dignon *et al.* 2018 *PLoS Comput. Biol.*

1 bead / amino acid



3 beads / nucleotide: Phosphate, Sugar, Base (A,C,G,T)



- Protein-DNA: PWMcos C.Tan, S. Takada, 2018, *JCTC*.

Atomic Interaction-based Coarse-Grained (AICG) model

$$\begin{aligned}
 V_{AICG2+}(\mathbf{R}|\mathbf{R}_0) = & \sum_{ibd} K_{b,ibd} (b_{ibd} - b_{ibd,0})^2 + V_{loc}^{flp} \\
 & + \sum_{j=i+2} \varepsilon_{loc,ij} \exp \left(-\frac{(r_{ij} - r_{ij0})^2}{2W_{ij}^2} \right) + \sum_{j=i+3} \varepsilon_{loc,ij} \exp \left(-\frac{(\phi_{ij} - \phi_{ij0})^2}{2W_{\phi,ij}^2} \right) \\
 & + \sum_{i < j-3}^{nat \text{ contact}} \varepsilon_{go,ij} \left[5 \left(\frac{r_{ij0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij0}}{r_{ij}} \right)^{10} \right] \\
 & + \sum_{i < j-3}^{non-native} \varepsilon_{exv} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}.
 \end{aligned}$$

W Li *et al.*, *PNAS* 2011, 2012, 2014

Bonded interactions:

- Bond length (harmonic, structure based)
- 1-3 distance (Gaussian, physics + structure based)
- Dihedral angle (Gaussian, physics + structure based)
- Flexible local potentials (Boltzmann inv., statistical)
 - Angle
 - Dihedral

Nonbonded interactions:

- Native contact (12-10)
 - r_0 : native-structure based
 - ε_{go} : physics + structure based
- Nonnative contact (12-repulsion)
 - σ_{ij} : knowledge based

ε_{loc} for 1-3 distance and dihedral angle, and ε_{go} :

- Energy decomposition from atomistic simulations
- Count hydrogen-bond numbers

3SPN.2C DNA Model



GENESIS
Generalized-ensemble simulation system

3-Site-Per-Nucleotide (3SPN) model for double-stranded DNA (dsDNA)

$$U_b = U_{bond} + U_{ang} + U_{dih}. \quad \text{Bonded}$$

$$U_{bond} = \sum_i^{bonds} k_b(r_i - r_{i,0})^2 + 100k_b(r_i - r_{i,0})^4.$$

$$U_{ang} = \sum_i^{angles} k_\theta(\theta_i - \theta_{i,0})^2.$$

$$U_{dih,Gaussian} = \sum_i^{dihedrals} -k_{\phi,Gaussian} \exp\left(\frac{-(\phi_i - \phi_{i,0})^2}{2\sigma_\phi^2}\right).$$

$$U_{dih,periodic} = \sum_i^{dihedrals} k_{\phi,periodic} [1 + \cos(\phi_i - \phi_{i,0})].$$

$$U_{nb} = U_{bstk} + U_{bp} + U_{cstk} + U_{exv} + U_{ele}. \quad \text{Nonbonded}$$

$$U_{bstk} = \sum_{ij}^{n_{bstk}} U_m^{rep}(\epsilon_{BS}, \alpha_{BS}, r_{ij}) + f(K_{BS}, \Delta\theta_{BS}) U_m^{attr}(\epsilon_{BS}, \alpha_{BS}, r_{ij}).$$

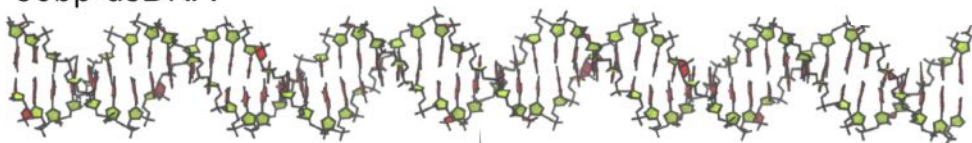
$$U_{bp} = \sum_{ij}^{n_{bp}} U_m^{rep}(\epsilon_{BP}, \alpha_{BP}, r_{ij}) + \frac{1}{2}(1 + \cos(\Delta\phi_1)) f(K_{BP}, \Delta\theta_1) f(K_{BP}, \Delta\theta_2) U_m^{attr}(\epsilon_{BP}, \alpha_{BP}, r_{ij}).$$

$$U_{cstk} = \sum_{kl}^{n_{cstk}} f(K_{BP}, \Delta\theta_3) f(K_{CS}, \Delta\theta_{CS}) U_m^{attr}(\epsilon_{CS}, \alpha_{CS}, r_{kl}).$$

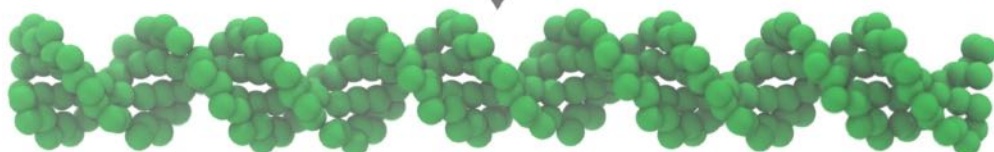
$$U_{exv} = \sum_{i<j} \begin{cases} \epsilon_r \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \epsilon_r & r_{ij} < \sigma_{ij}, \\ 0 & r_{ij} > \sigma_{ij}. \end{cases}$$

$$U_{ele} = \sum_{i<j}^{n_{ele}} \frac{q_i q_j e^{-r_{ij}/\lambda_D}}{4\pi\epsilon_0\epsilon(T, C)r_{ij}}.$$

50bp-dsDNA

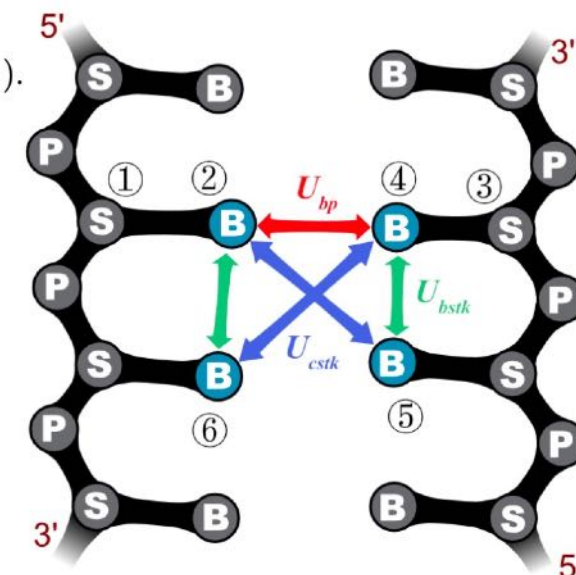


Coarse-graining



DM Hinckley *et al.*, JCP 2013

JS Freeman *et al.*, JCP 2014



Fitted experimental properties:

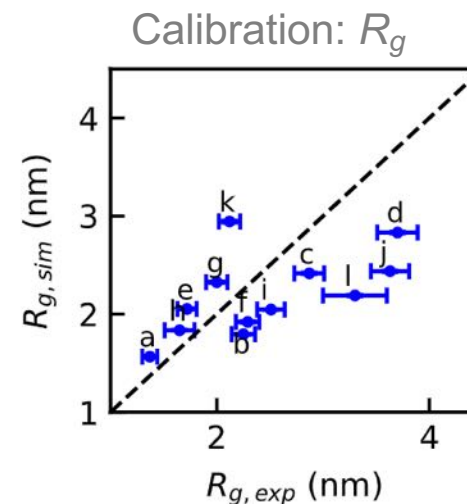
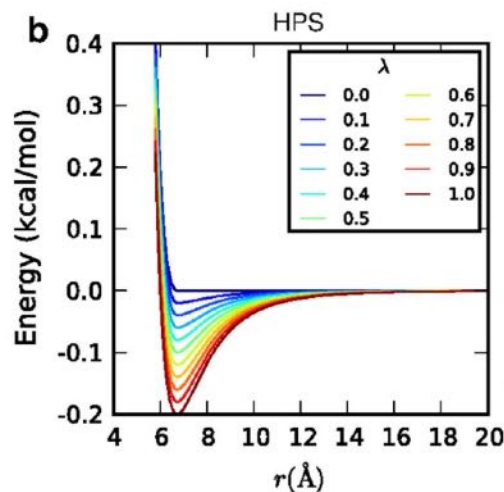
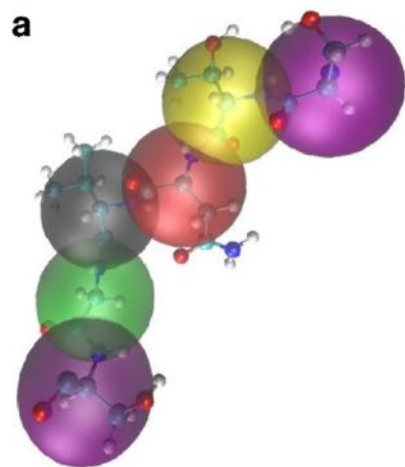
- Persistence length
- Melting temperature
- Hybridization rate constant
- **Sequence-dependent curvature**
- Local base-step flexibility

Weak points:

- High ionic concentration
- Single-stranded DNA

HPS/KH IDP Model

Hydrophobicity-scale (HPS) and **K**im-**H**ummer (KH) model for **I**ntrinsically **D**isordered **P**rotein (IDP)



Dignon *et al.* 2018 *PLoS Comput. Biol.*

Variations:

Regi *et al.* 2021 *Proteins*

Dannenhofer-Lafage *et al.* 2021 *JPCB*

Tesei *et al.* 2021 *PNAS*

Regi *et al.* 2020 *NAR* (protein and RNA)

Electrostatic interaction

$$E_{ij}(r) = \frac{q_i q_j}{4\pi D r} \exp(-r/\kappa),$$

HPS potential

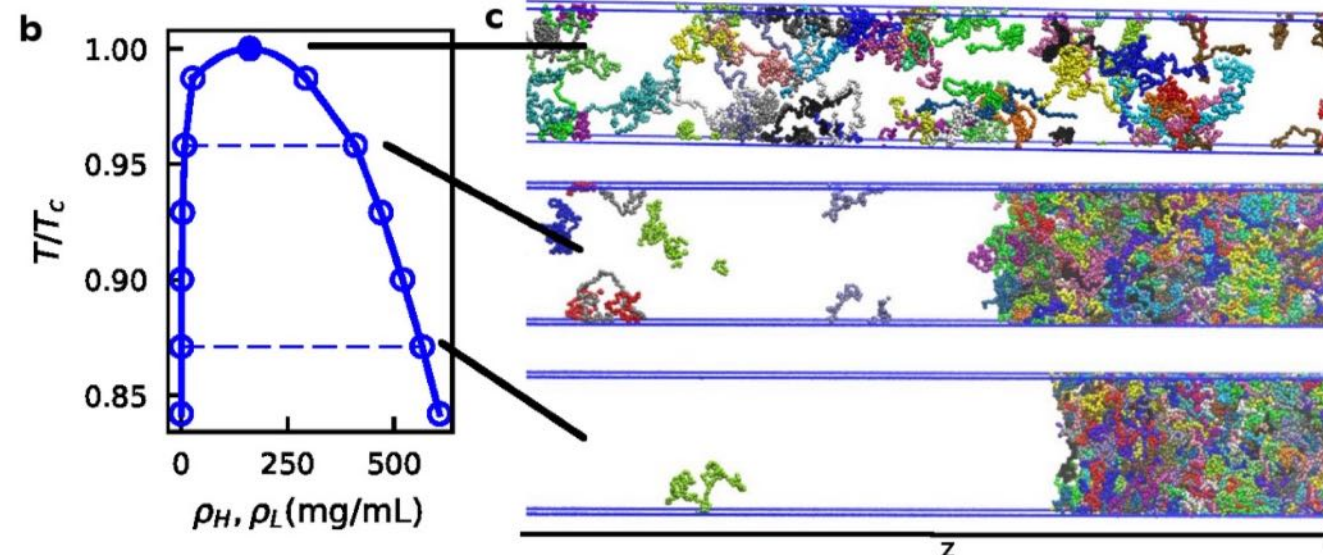
Ashbaugh-Hatch functional form

$$\Phi(r) = \begin{cases} \Phi_{LJ} + (1 - \lambda)\epsilon, & \text{if } r \leq 2^{1/6}\sigma \\ \lambda\Phi_{LJ}, & \text{otherwise} \end{cases}$$

Lennard-Jones potential

$$\Phi_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right].$$

Slab method: simulating multi-chain systems



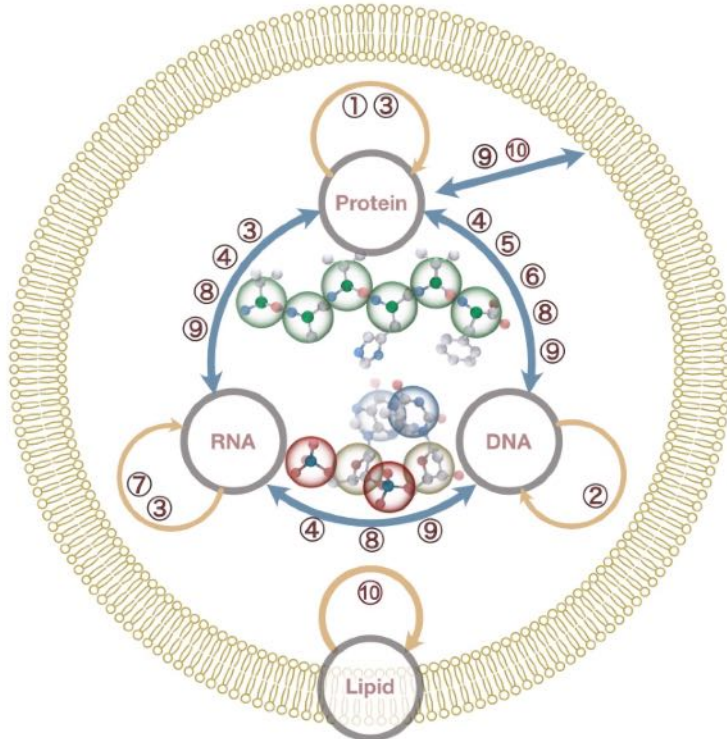
CG Models and MD in GENESIS



GENESIS
Generalized-ensemble simulation system

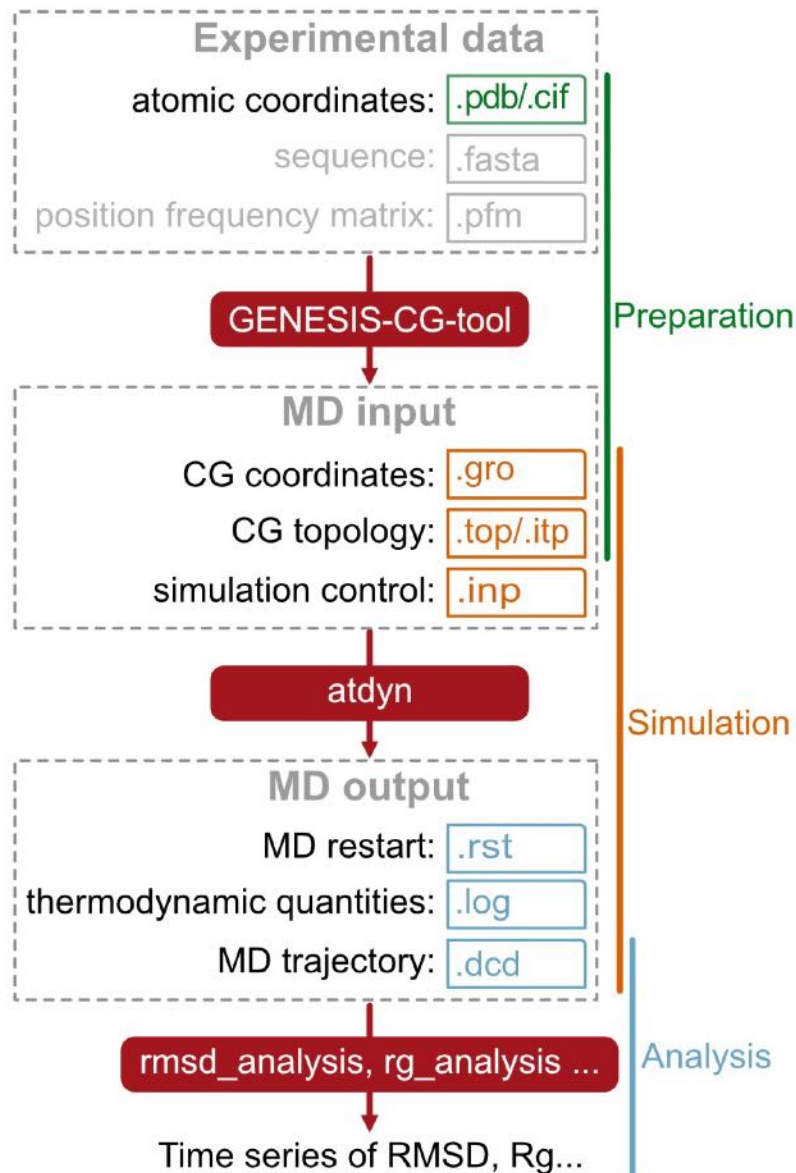
C.Tan *et al.* 2022, *PLoS Comput. Biol.*

Available Models

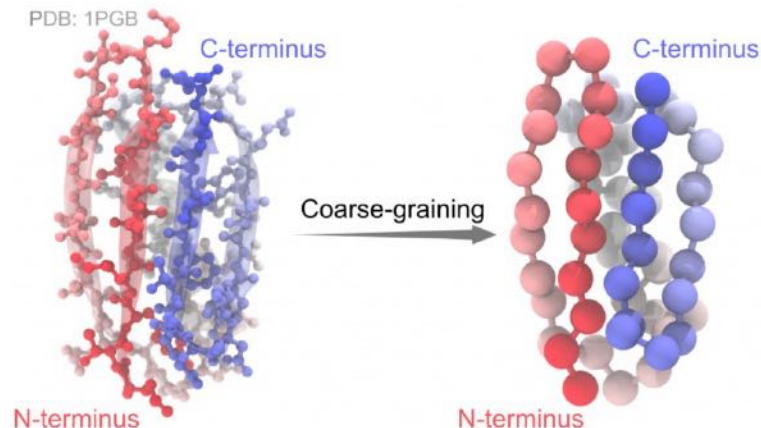


1. W. Li *et al.* (2014). *Proc. Natl. Acad. Sci.*
2. G.S. Freeman *et al.* (2014). *J. Chem. Phys.*
3. G.L. Dignon *et al.* (2018). *PLoS Comput. Biol.*
4. C. Clementi *et al.* (2000). *J. Mol. Biol.*
5. C. Tan & S. Takada (2018). *J. Chem. Theory Comput.*
6. G.B. Brandani *et al.* (2018) *Nucl. Acids 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.*

Information flow



One-line preparation example



```
$ ls
PR01.pdb

# run GENESIS_CG_TOOL
$ genesis_cg_tool/src/aa_2_cg.jl PR01.pdb

$ ls
PR01.pdb PR01_cg.gro PR01_cg.top PR01_cg.itp
```

GENESIS-cg-tool: (in Julia)

- https://github.com/genesis-release-r-ccs/genesis_cg_tool

Online wiki:

- https://github.com/genesis-release-r-ccs/genesis_cg_tool/wiki

GENESIS source code + manual:

- <https://www.r-ccs.riken.jp/labs/cbrt/genesis-version-2-1/>





Tutorials:

- <https://r-ccs.riken.jp/labs/cbrt/tutorials2022/>

<https://www.r-ccs.riken.jp/labs/cbrt/tutorials2022/>

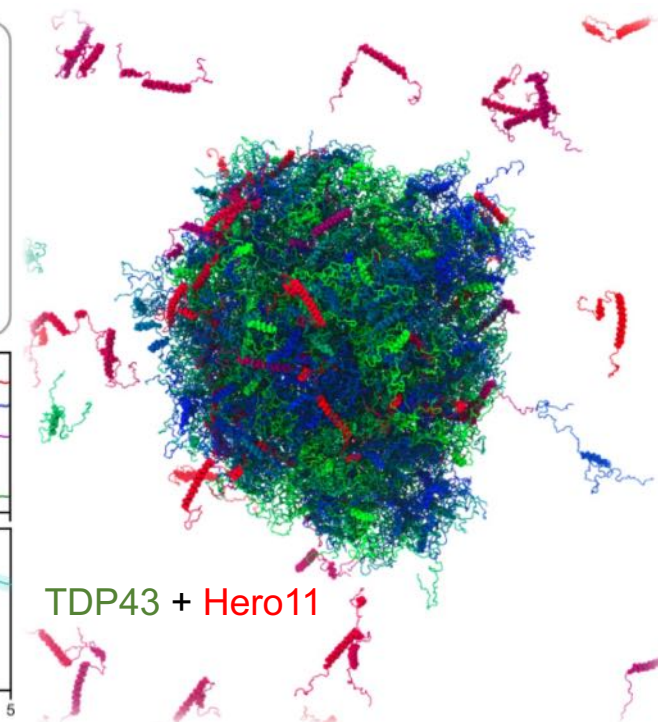
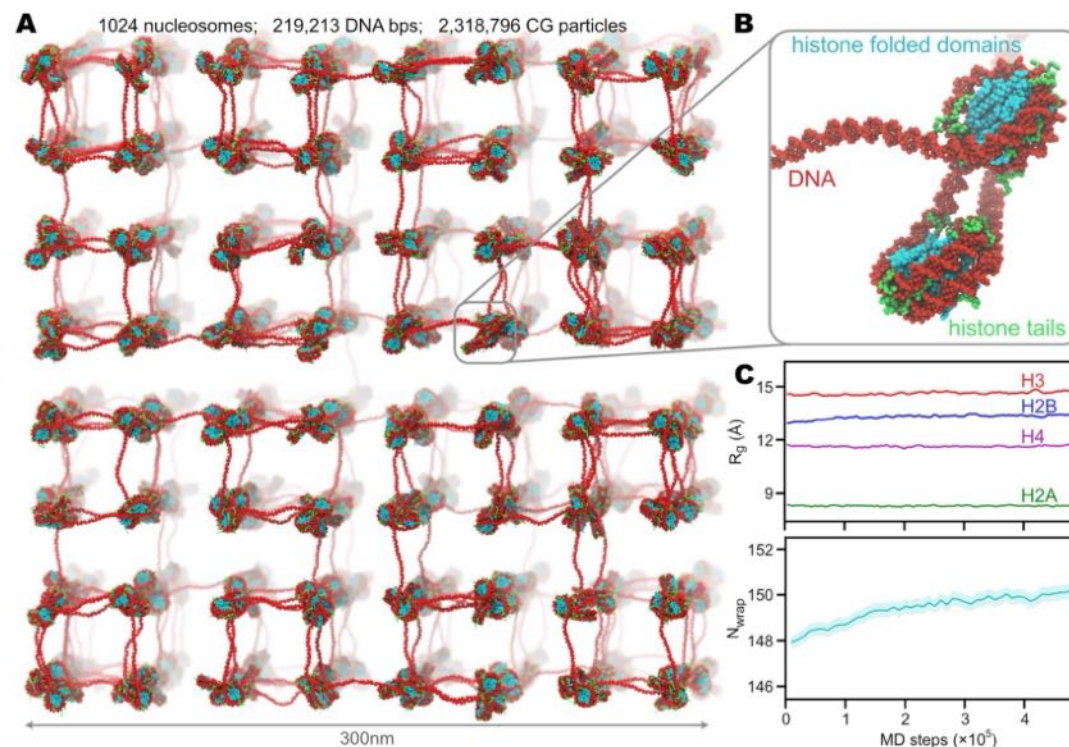
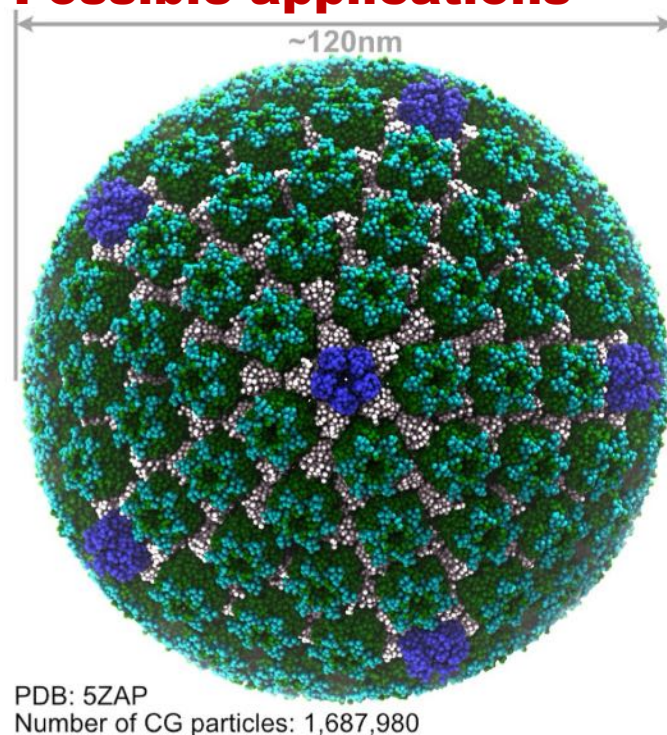
C.Tan *et al.* **2022**, *PLoS Comput. Biol.*

11. Advanced MD simulations with the coarse-grained model

- 11.1 Coarse-grained simulation of protein with AICG2+ model 
- 11.2 Coarse-grained simulation of double-stranded DNA with 3PSC.N model 
- 11.3 Coarse-grained simulation of protein-DNA interactions with PWMcos model 
- 11.4 Coarse-grained simulation of FUS condensation with HPS model 

- Protein folding
- DNA dynamics
- Protein-DNA recognition
- Phase behaviors of IDP

Possible applications



C.Tan, A. Niitsu, Y. Sugita, *JACS Au*, 2023.

Extra: High-Performance CG MD



GENESIS
Generalized-ensemble simulation system



GENESIS

Generalized-ensemble simulation system

This CG tutorial only
applies to ATDYN!

ATDYN

Atomic decomposition **d**ynamics
Memory limit

SPDYN

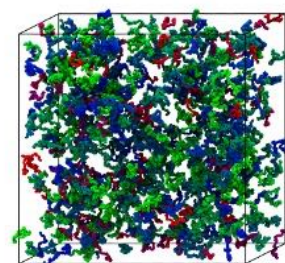
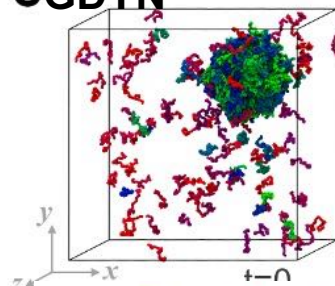
Spatial decomposition **d**ynamics
Load imbalance

CGDYN

CG molecular **d**ynamics

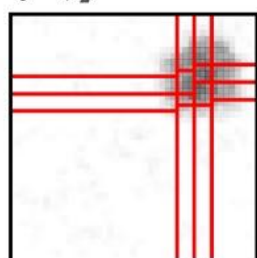
beta version

CGDYN

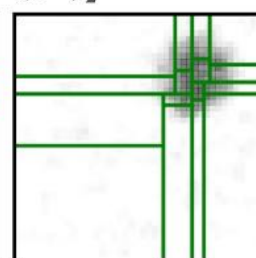


$t=9 \times 10^6$ steps

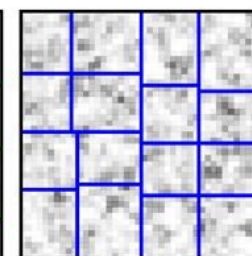
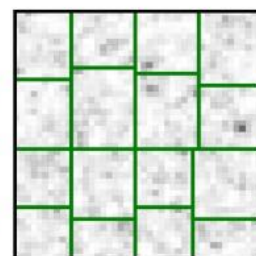
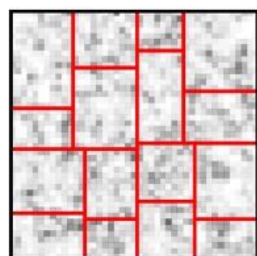
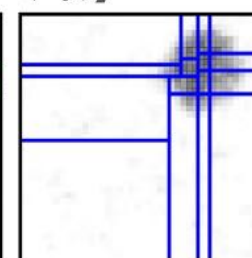
$(y-z)_2$



$(z-x)_2$



$(x-y)_2$



*Dynamic domain
decomposition scheme
to address the
nonuniform distribution*



Dr. Jaewoon Jung

Jung, Tan, Sugita, *Nat. Commun.* 2024