A coarse-grained simulator: CafeMol

検崎博生[1],
古賀信康[1], 藤原慎司[1], 堀直人[1], 金田亮[1],
李 文飛[1,2], 岡崎圭一[1], 姚 新秋[1], 高田彰二[1,2]
[1]京都大学理学研究科生物物理学教室
[2] JST-CREST

CafeMol (www.cafemol.org)

Kenzaki

- CafeMol 2.1 (2013/7)
  source & manual released

- Features are;
  - Various CG models
    protein/DNA/RNA
    multiple basin model
    accurate CG model
  - Simulating protein-at-work
    “switching”

- Under development
  lipid

Overview of CafeMol

- General-purpose coarse-grained (CG) biomolecular modeling and simulation software
  - Protein: 1 bead / 1 amino acid
  - Nucleic acid:
    3 beads (sugar, base, phosphate) / nucleotide
  - Lipid: ~3 beads / lipid
- Written by FORTRAN90 with MPI and Open MP
- Large-scale simulation
  - ～”millisecond” event by K-computer
- Version 1.0 is released (only protein) (2010/12/27)
  - Version 2.0 (protein, DNA, RNA) (2012/5/31)

Menu

Models

Simulation methods

Implementation

Selected applications

In-progress models & methods

Menu

Models

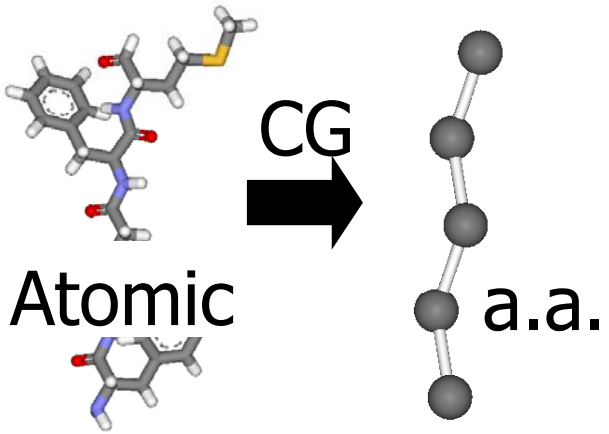
Simulation methods

Implementation

Selected applications

In-progress models & methods

Models and energy functions



1 beads / 1 amino acid

A. Off-lattice Go model
B. Atomic interaction based CG model
C. Multiple basin model
D. DNA/RNA model
E. Elastic network model
F. Electrostatic and hydrophobic interactions
G. Explicit and implicit ligands

Off-lattice Go model

C. Clementi, H. Nymeyer, and J.N. Onuchic, J. Mol. Biol. (2000)

Based on the energy landscape theory

Structure based

$$V_{protein} = V_{local} + V_{go} + V_{ex}$$

$$V_{local} = K_b \sum_i \left(r_{i,i+1} - r_{0i,i+1} \right)^2 + K_\theta \sum_i \left(\theta_i - \theta_{0i} \right)^2$$

$$+ K_\phi^1 \sum_i \left(1 - \cos\left(\phi_i - \phi_{0i} \right) \right) + K_\phi^3 \sum_i \left(1 - \cos 3\left(\phi_i - \phi_{0i} \right) \right)$$

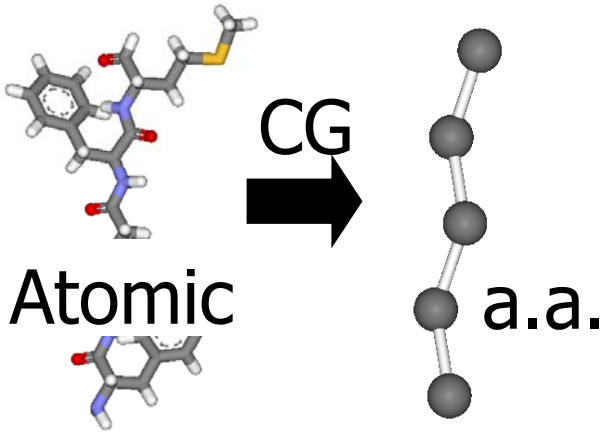
$$V_{go} = \varepsilon_{go} \sum_{i,j}^{native} \left[5 \left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0ij}}{r_{ij}} \right)^{10} \right]$$

$$V_{ex} = \varepsilon_{ex} \sum_{i,j}^{nonnative} \left(\frac{\sigma}{r_{ij}} \right)^{12}$$

θ: bond angle
ϕ: dihedral angle
(0 means native state)

$K_b = 100\varepsilon$
$K_\theta = 20\varepsilon$
$K_\phi^1 = \varepsilon$
$K_\phi^3 = 0.5\varepsilon$
$\varepsilon_{go} = 0.18\varepsilon$
$\varepsilon_{ex} = \varepsilon$
σ= 4A
ε=1.0kcal/mol

Atomic interaction based CG (AICG) model



W. Li, P. Wolynes, S. Takada, PNAS (2011)

$$V = \sum_i k_b^i (r^i - r_0^i)^2 + \sum_i k_a^i (\theta^i - \theta_0^i)^2$$
$$+ \sum_i \{\varepsilon_{\phi,1}^i [1 - \cos(\phi^i - \phi_0^i)] + \varepsilon_{\phi,3}^i [1 - \cos 3(\phi^i - \phi_0^i)]\}$$
$$+ \sum_{i>j-3}^{native} \varepsilon^{ij} [5(r_0^{ij}/r^{ij})^{12} - 6(r_0^{ij}/r^{ij})^{10}] + \sum_{i>j-3}^{non-native} \varepsilon (C/r^{ij})^{12}$$

Wenfei Li

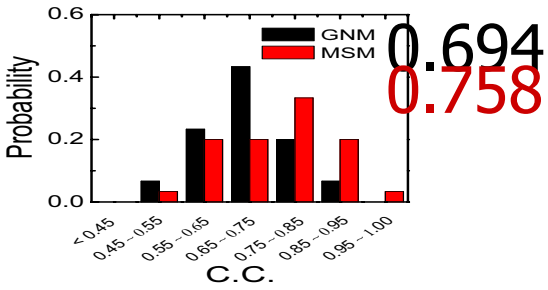
1) Contact energy ε_{ij} from pairwise all-atom (AA) energy

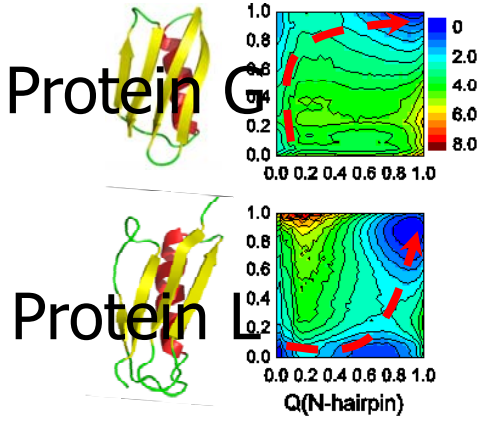
$$E^{IJ}(R_{IJ}) = \sum_{i \in I} \sum_{j \in J} u_{AA}(r_{ij}) \qquad u_{AA}(r) = V(r) + \Delta G_{pol}^{GB}(r) + \Delta G^{SA}(r)$$

2) Coefficients fitted by AA-derived fluctuation (23 proteins)

param	K_b	K_a^G	k_a^H	k_a^E	k_a^T	k_a^C	ε_ϕ^G	ε_ϕ^H	ε_ϕ^E	ε_ϕ^T	ε_ϕ^C	ε_{nloc}
Av.	109.94	13.40	40.03	17.32	19.35	11.70	0.29	1.76	1.32	0.82	0.81	0.37

Test for fluctuation, structural change, & folding





Multiple-basin model for proteins

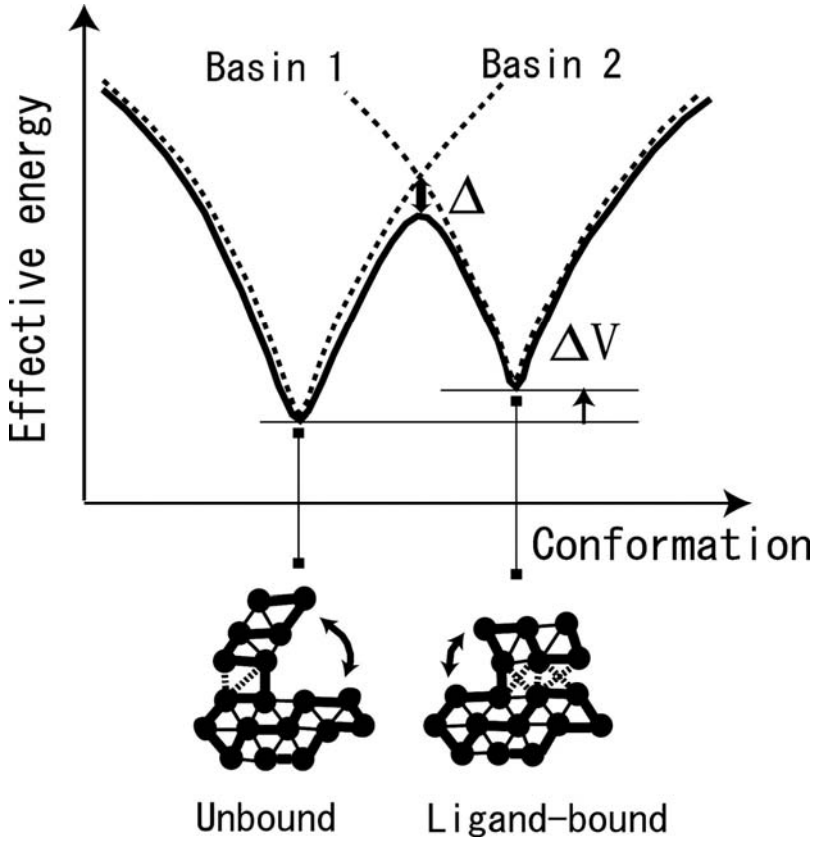
K. Okazaki, N. Koga, S. Takada, J.N. Onuchic, and P.G. Wolynes, PNAS (2006)

Use of 2 references

$$\begin{pmatrix} V(R|R_1) & \Delta \\ \Delta & V(R|R_2)+\Delta V \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = V_{MB}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

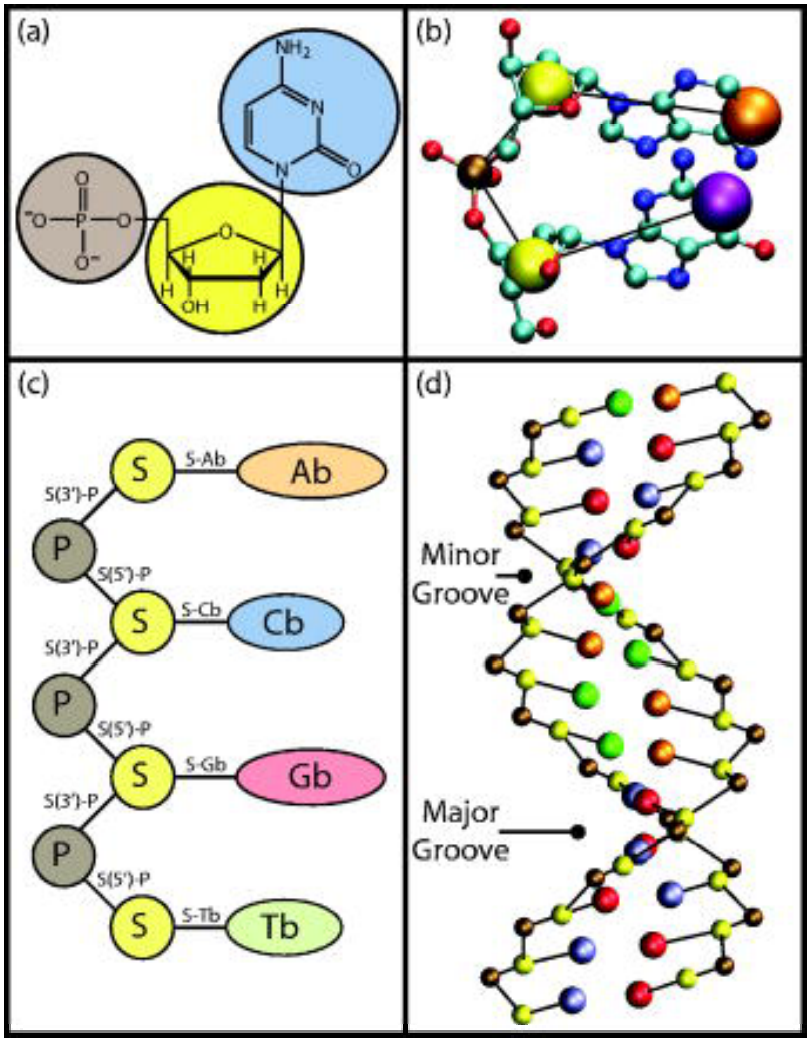
$$V_{MB} = \frac{V(R|R_1)+V(R|R_2)+\Delta V}{2} - \sqrt{\left(\frac{V(R|R_1)-V(R|R_2)-\Delta V}{2}+\Delta \right)}$$

$$\chi = \log\left(\frac{c_2}{c_1}\right)$$



CG DNA model

- Three interactions sites
  - Phosphate
  - Sugar
  - Base
- Reproduce various DNA behavior
  - Salt-dependent melting
  - Bubble formation
  - Mechanical properties



T.A. Knotts IV, N.Rathore, D.C. Shwartz, and J.J. Pablo, J. Chem. Phys. (2007)

3SPN.1 force field

E.J. Sambrisiki, D.C. Schwartz, and J.J. de Pablo,Knotts, Biophys J. (2009)

$$V_{dna} = V_{local} + V_{stack} + V_{bp} + V_{ex} + V_{qq} + V_{solv}$$

$$V_{local} = K_{b1}\sum_i \left(r_{i,i+1} - r_{0i,i+1}\right)^2 + K_{b2}\sum_i \left(r_{i,i+1} - r_{0i,i+1}\right)^4$$

$$+ K_\theta \sum_i \left(\theta_i - \theta_{0i}\right)^2 + K_\phi \sum_i \left(1 - \cos\left(\phi_i - \phi_{0i}\right)\right)$$

$$V_{stack} = 4\varepsilon_1 \sum_{i,j}^{N_{st}} \left[\left(\frac{\sigma_{0ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{0ij}}{r_{ij}}\right)^{6} \right]$$

$$V_{bp} = \sum_{i,j}^{N_{bp}} 4\varepsilon_{bpi} \left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right]$$

$$V_{ex} = 4\varepsilon_1 \sum_{i,j}^{N_{ex}} \left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6} \right] + \varepsilon_1 \ (if\ \ r_{ij} < d_{cut}),$$

$$= 0 \ (if\ \ r_{ij} > d_{cut})$$

θ: bond angle
ϕ: dihedral angle
(0 means B-type DNA)

Structre-based interaction

$K_{b1} = 1\varepsilon$
$K_{b2} = 100\varepsilon$
$K_\theta = 1400\varepsilon$
$K_\phi = 28\varepsilon$
$\varepsilon_{bpGC} = 2.532\varepsilon$
$\varepsilon_{bpAT} = 2.0\varepsilon$
ε =0.1839kcal/mol

3SPN.1 force field (electrostatic and solvation interaction)

$$V_{qq} = \sum_{i,j}^{N} \left(\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon(T,C) r_{ij}} \right) e^{-r_{ij}/\kappa D}$$

← Debye-Huckel theory

$$\varepsilon(T,C) = \varepsilon(T) a(C)$$

$$\varepsilon(T) = 249.4 - 0.788 T/K + 7.20 \times 10^{-4} (T/k)^2$$

$$a(C) = 1.000 - 0.2551 C/M$$

$$+ 5.151 \times 10^{-2} (C/M)^2 - 6.889 \times 10^{-3} (C/M)^3$$

$$V_{solv} = \sum_{i<j}^{N_{solv}} \varepsilon_s \left[1 - e^{-a(r_{ij} - r_s)} \right]^2 - \varepsilon_s$$

$$\varepsilon_s = \varepsilon_N A_I$$

$$e_N = e_0 (1 - [1.40418 - 0.268231 N_{nt}]^{-1})$$

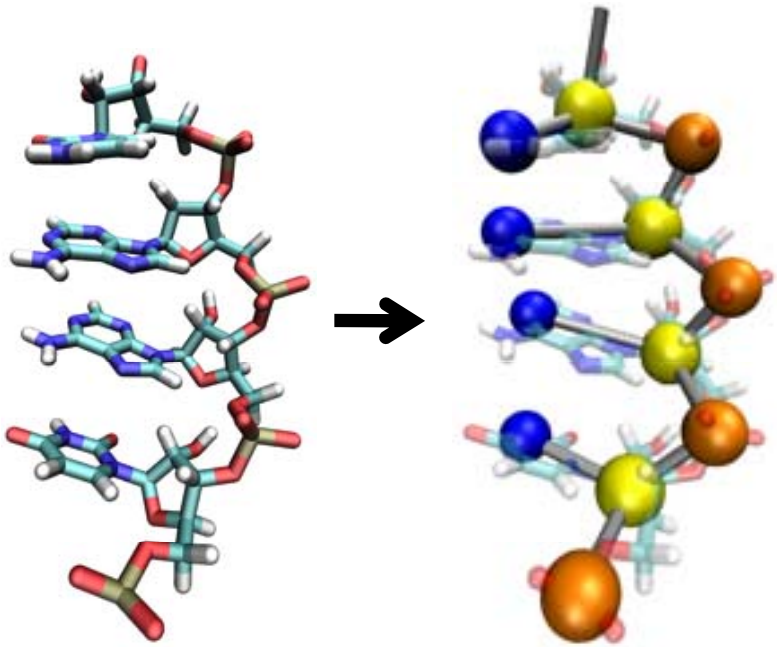
$$A_I = 0.474876 (1 + \{0.148378 + 10.9553 [Na^+]\}^{-1})$$

α^{-1} = 5.333A
r_s = 13.38A
$\varepsilon_0 = 0.504982\varepsilon$

RNA model

N. Hori and S. Takada (2012)

1 nucleotide = 3 beads



- Phosphate (phosphorus atom)
- Sugar (center of ribose ring)
- Base (pyrimidine: N1 atom)
  (purine: N3 atom)

RNA model (local)

$$V_{\text{total}} = V_{\text{local}} + V_{\text{stack}} + V_{\text{basepair}} + V_{\text{nonlocal}} + V_{\text{exclude}}$$

Local

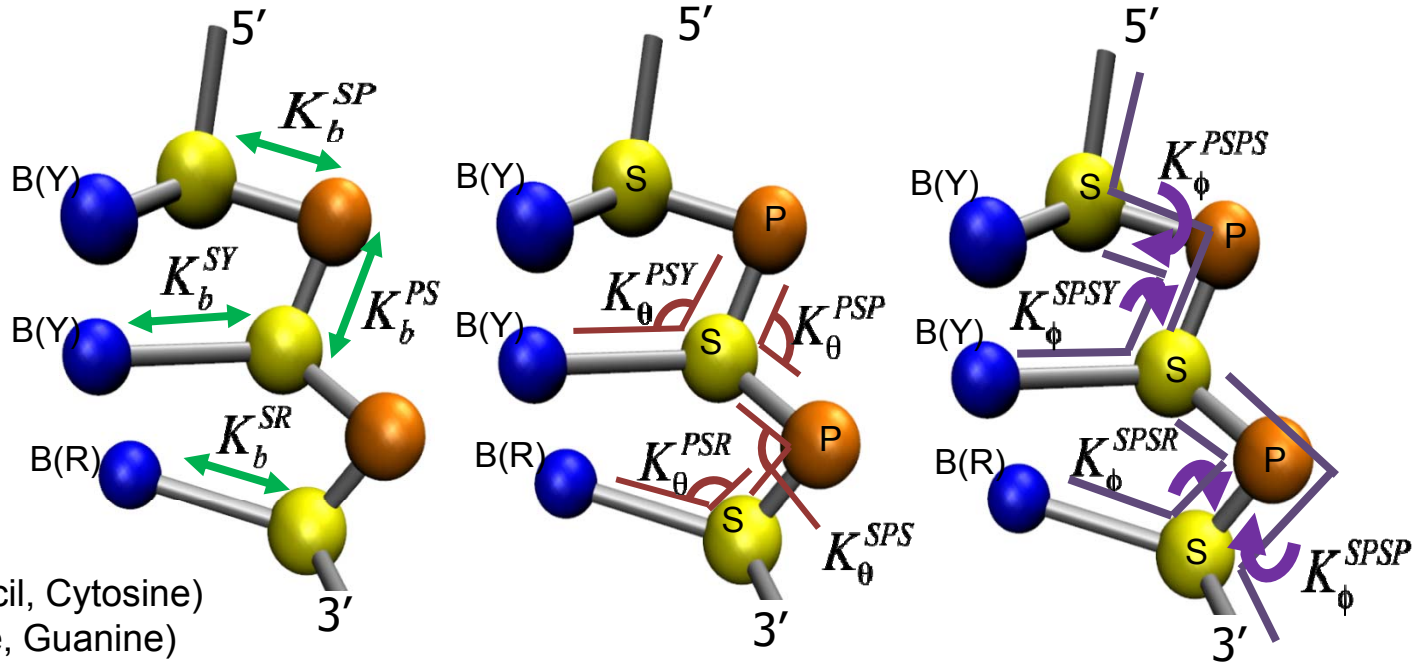
$$V_{local} = \sum_{ibd \in \substack{PS,\\ SP,\\ SB}} K_b^{\eta}(r_{ibd} - r_{ibd}^{(0)})^2 + \sum_{iba \in \substack{PSP,\\ SPS,\\ SPB}} K_{\theta}^{\eta}(\theta_{iba} - \theta_{iba}^{(0)})^2$$

$$+ \sum_{idih \in \substack{PSPS,\\ SPSP,\\ SPSB}} \{K_{\phi}^{\eta}(1 - cos(\phi_{idih} - \phi_{idih}^{(0)})) + 0.5K_{\phi}^{\eta}(1 - cos3(\phi_{idih} - \phi_{idih}^{(0)}))\}$$

Bond length

Bond angle

Dihedral angle



Base type
Y : Pyrimidine (Uracil, Cytosine)
R : Purine (Adenine, Guanine)

RNA model (nonlocal)

Nonlocal

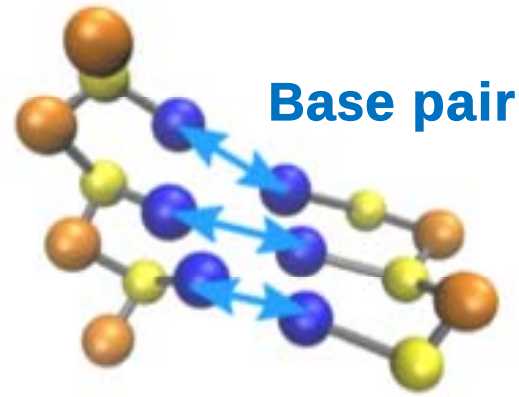
Reference (X-ray structure's) value

$$V_{\mathrm{contact}} = \sum_{i,j \in S,B}^{\text{native contact}} \varepsilon_\xi \left[5 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{10} \right]$$

$$\varepsilon_\xi = \varepsilon_{BP3}$$

$$\varepsilon_\xi = \varepsilon_{BP2}$$

Base pair



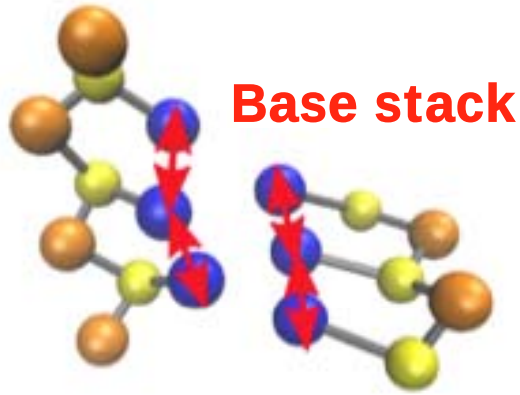
Base pair

$$\varepsilon_\xi = \varepsilon_{\mathrm{ST}}$$

Base stack

$$\varepsilon_\xi = \varepsilon_{con}^{\eta}$$

Contact (Other)



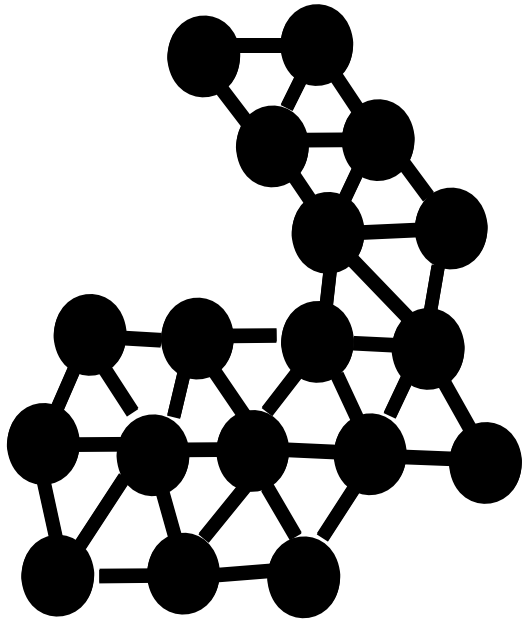
Base stack

$$V_{exclude} = \sum_{i,j}^{\text{non-native}} \varepsilon_{ex} \left(\frac{\sigma}{r_{i,j}} \right)^{12}$$

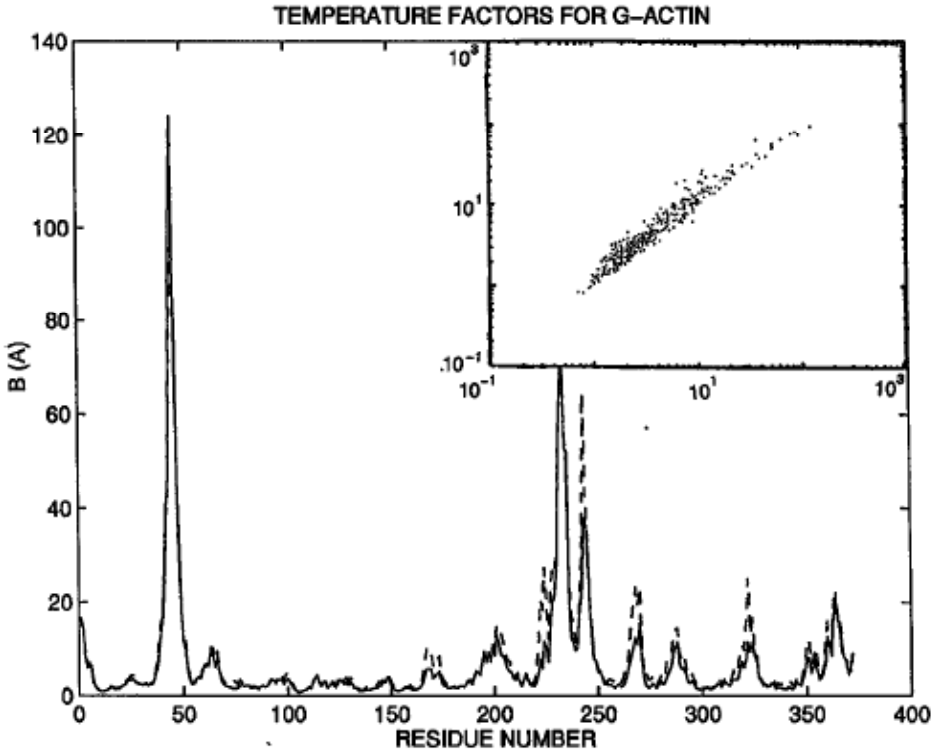
Excluded volume

Elastic network model

$$E = \sum_{ij, s.t. r_{ij}^0 < r_c} K(r_{ij} - r_{ij}^0)^2$$



Atomic fluctuation reproduced by ENM (Tirion96)



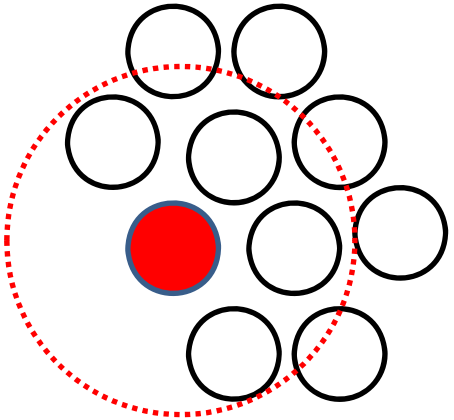
Electrostatic and hydrophobic interactions

Debye-Huckel form for electrostatics

$$V_{\text{ele}} = \sum_{i<j}^{N} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_k r_{ij}} e^{-r_{ij}/\kappa_D}$$

HP interactions analogous to ASA

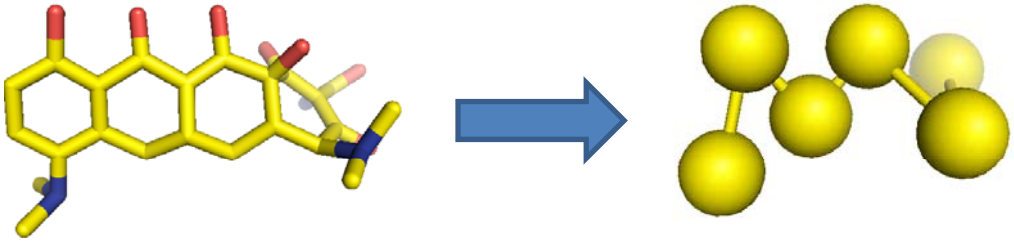
$$V_{\text{HP}} = -c_{\text{HP}} \sum_{i\in\text{HP}} \epsilon_{\text{HP},A(i)} S_{\text{HP}}(\rho_i)$$



Count coordination number
for each hydrophobic particle

Explicit and Implicit ligands

Explicit ligand; as a rigid molecule



Implicit ligand; MD-MC scheme with ligand-mediated contact

$$V_{protein} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} V_{protein} + V_{imp-lig}$$

$$V_{imp-lig} = \sum_{\substack{ligand-mediated \\ contact-pairs}} -c_{lig}\varepsilon_{go} \exp\left[-\frac{(r_{ij}/r_{0ij}-1)^2}{2(\sigma/r_{0ij})^2}\right]$$

Menu

Models

Simulation methods

Implementation

Selected applications

In-progress models & methods

Simulation method

- Dynamics
  - Newtonian dynamics with Berendsen thermostat
  - Langevin dynamics
  - Multi-Particle Collision dynamics (MPC)
- Time integration
  - velocity Verlet algorithm
- Run mode
  - Constant temperature simulation
  - Simulated annealing
  - Auto-search of Tf
  - Replica exchange method
  - Potential “switching”
- Useful option
  - anchor, bridge, pulling, fix, box

Menu

Models

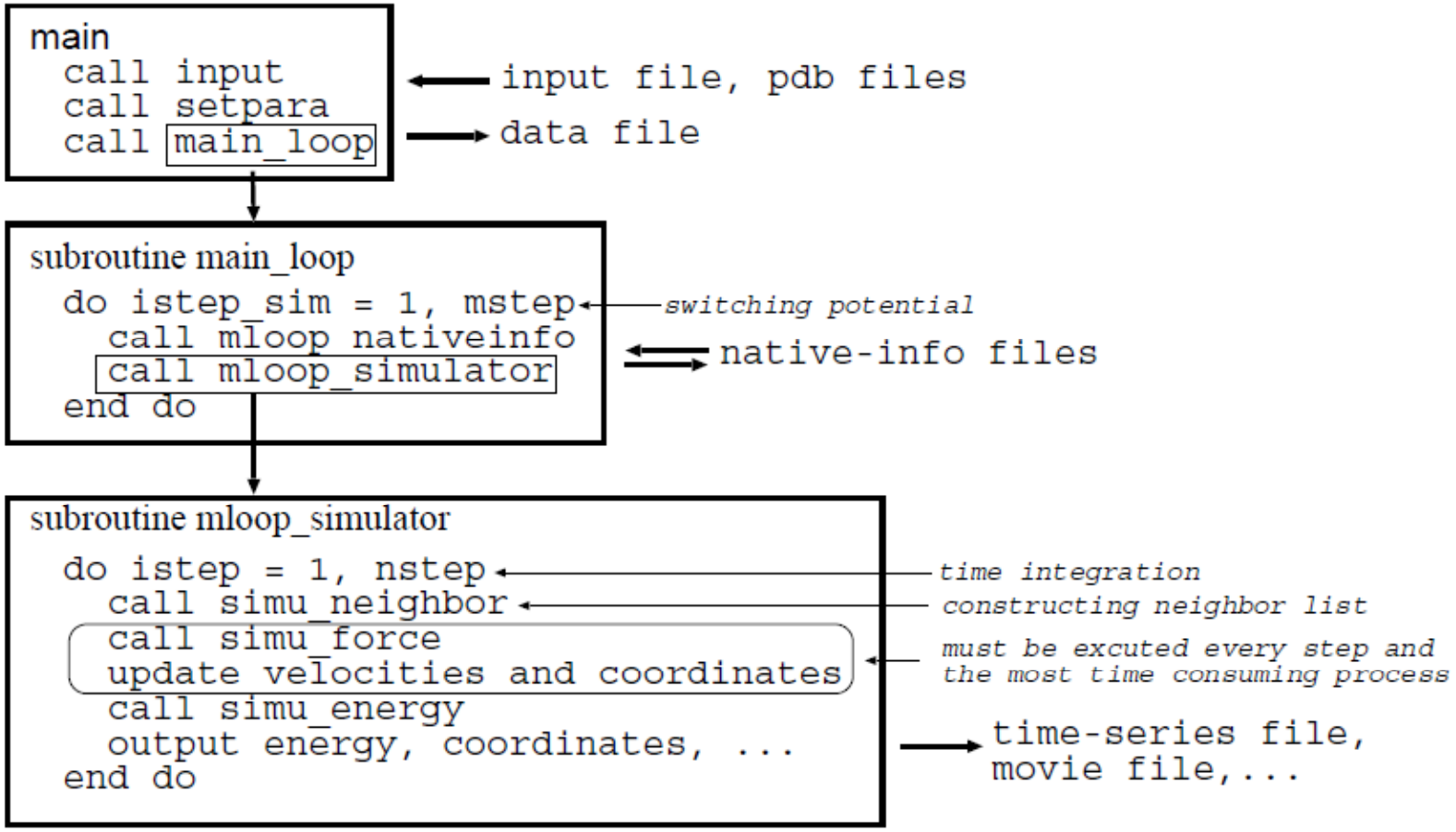
Simulation methods

Implementation

Selected applications

In-progress models & methods

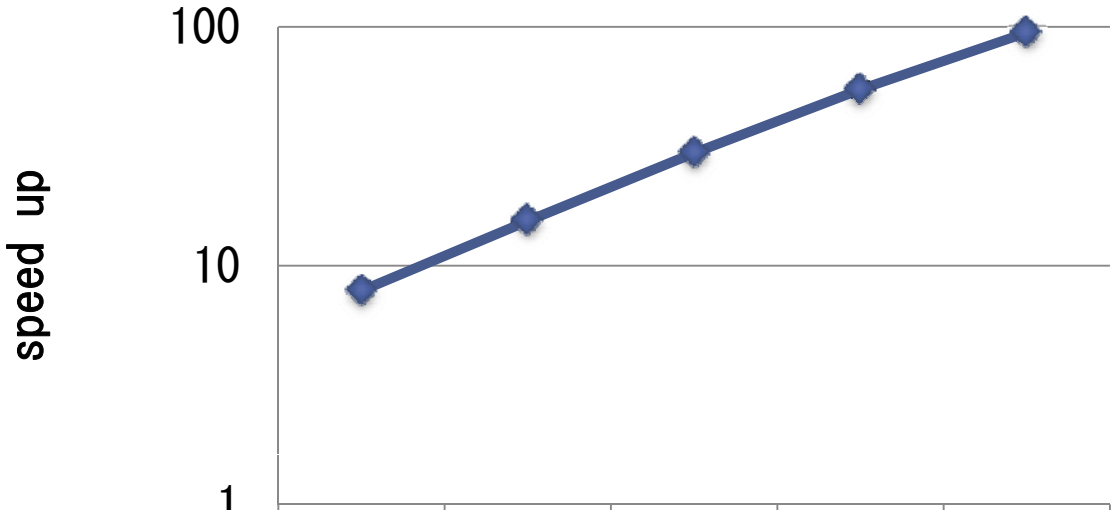
CafeMol code



- Parallelization
  - neighboring list, force, energy
    →hybrid(MPI+Open MP)
  - replica exchange
    →MPI(temperature/Hamiltonian REMD)

Performance of MPI parallelization

1300 base pairs DNA (7798 particles)
BG/L at Riken



	8	16	32	64	128
speed up	7.85	15.38	29.64	54.76	94.65

ope(s)	328.62	164.64	82.65	42.18	22.01
mpi(s)	5.44	5.90	5.84	5.71	5.70

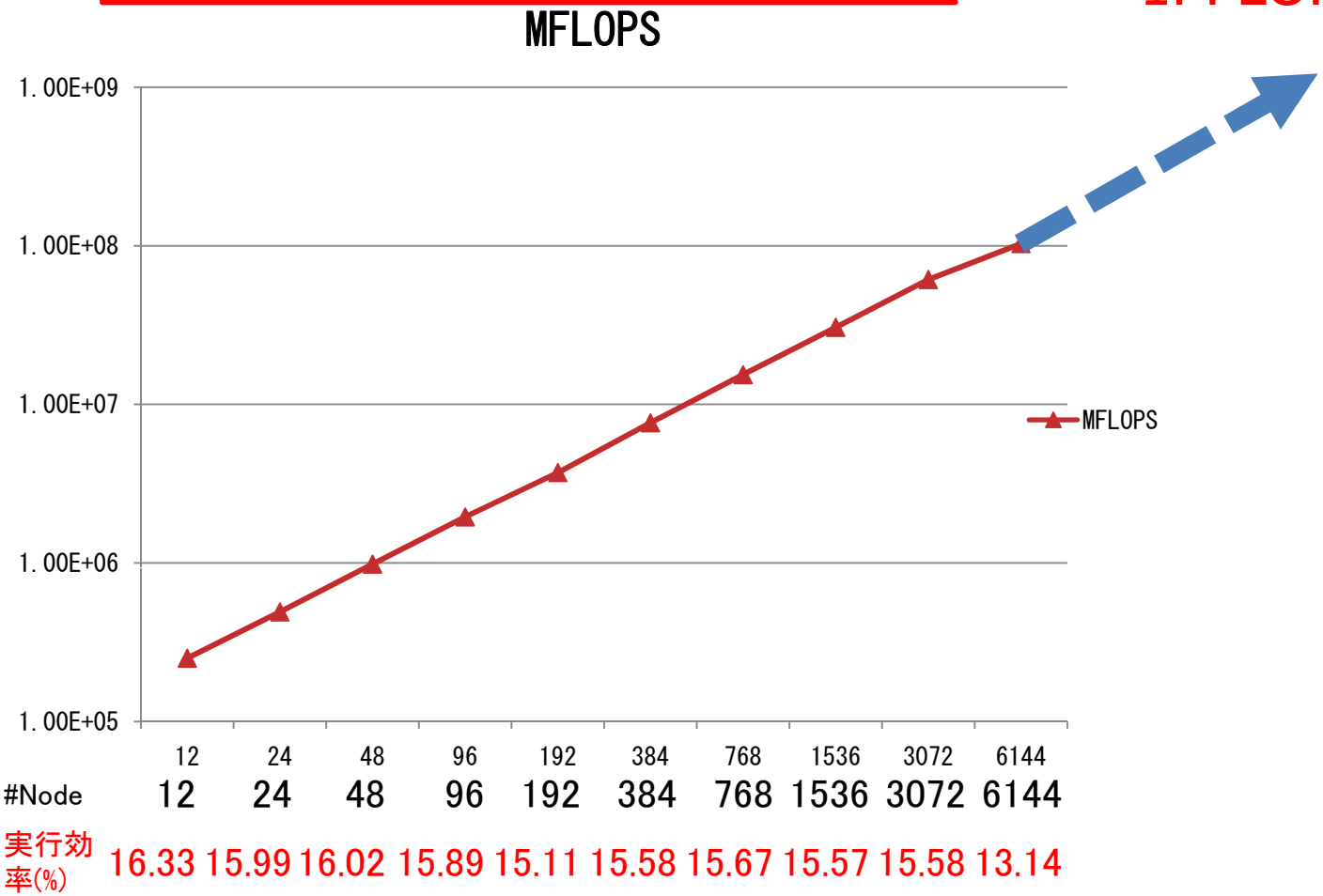
High parallelization efficiency

Performance of parallelization at K-Computer

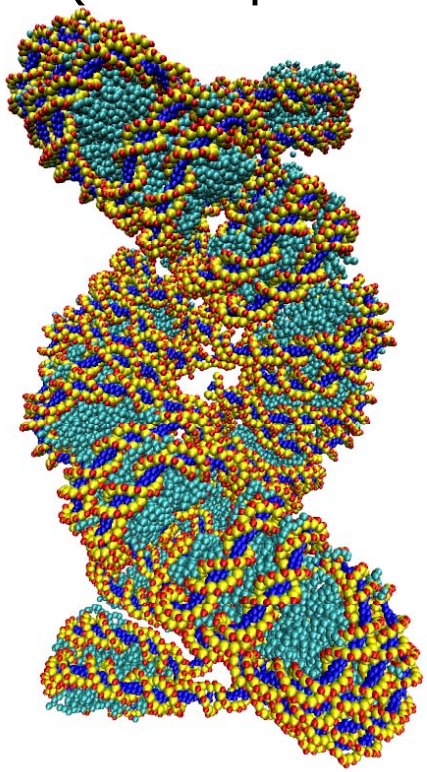
Before 2010
RICC at Riken (MPI+Open MP)
After 2011
at K-computer (SIMD)

Efficiency of parallelization:
99.996%

#node ~73,000 ⇒
~1PFLOPS !?



20 nucleosomes
(35918 particles)



Menu

Models

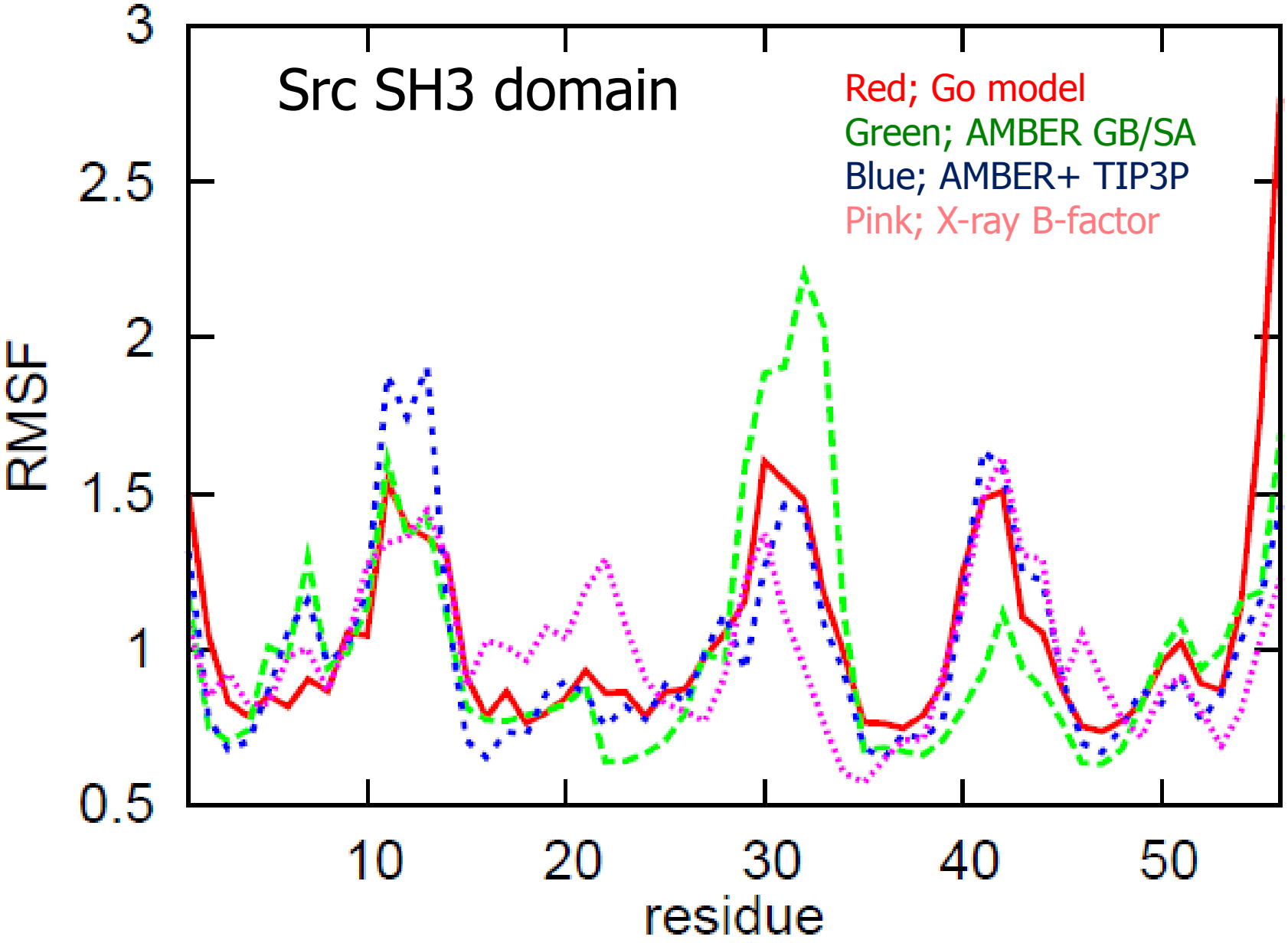
Simulation methods

Implementation

Selected applications

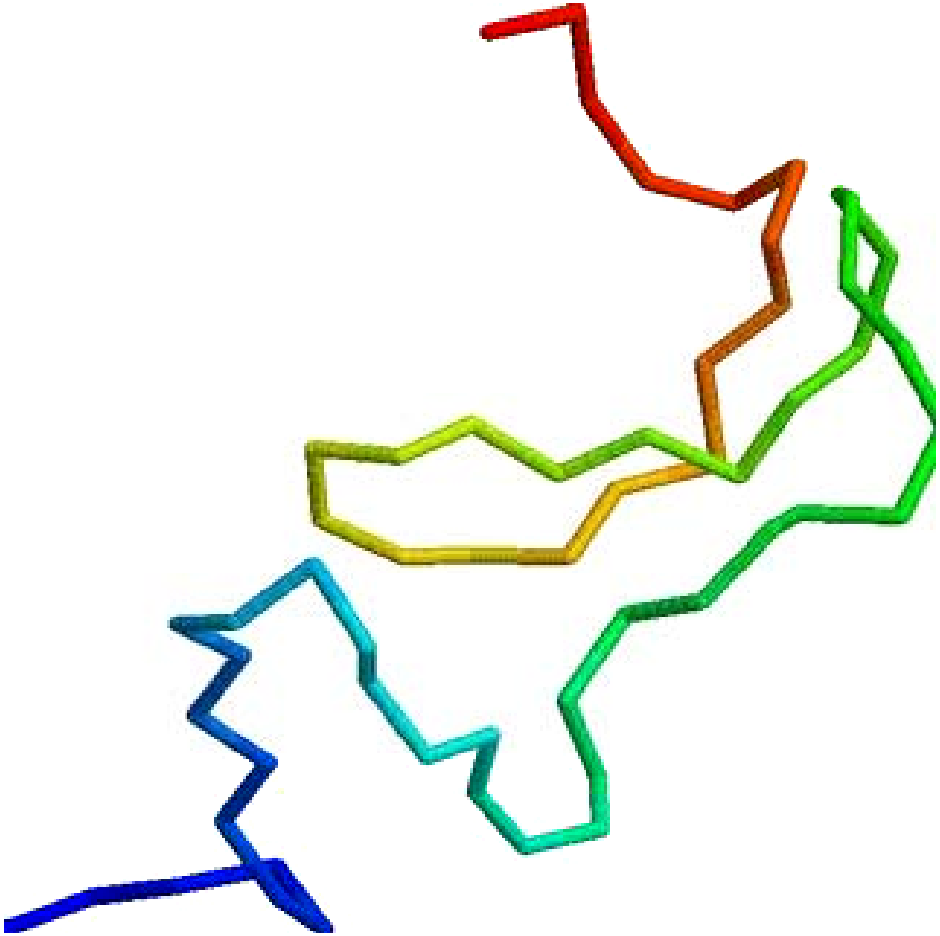
In-progress models & methods

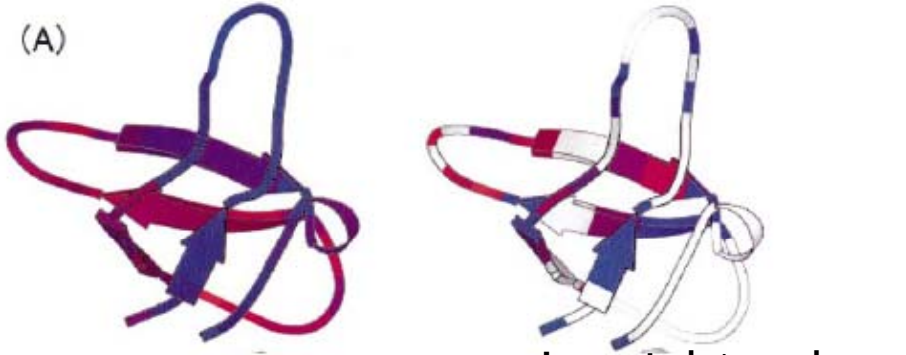
Native fluctuation by off-lattice Go model



Folding simulation of src SH3 domain

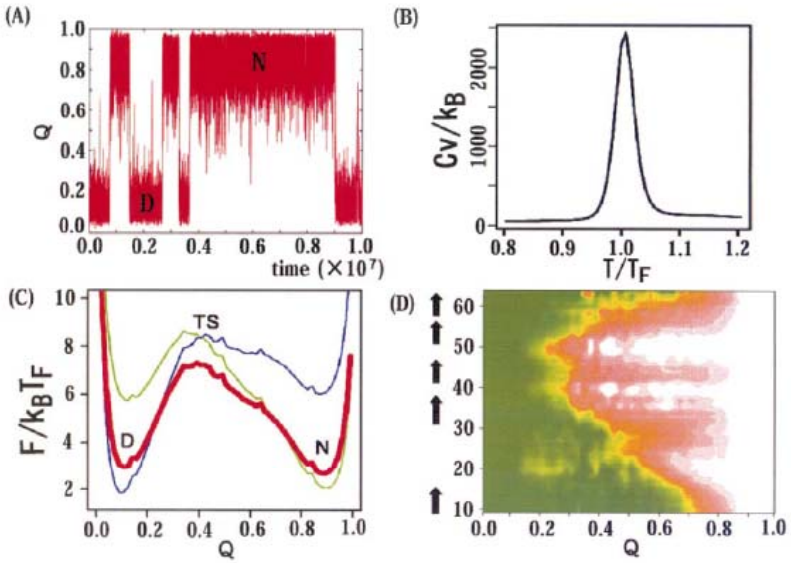
N. Koga, and S. Takada, J. Mol. Biol. (2001)



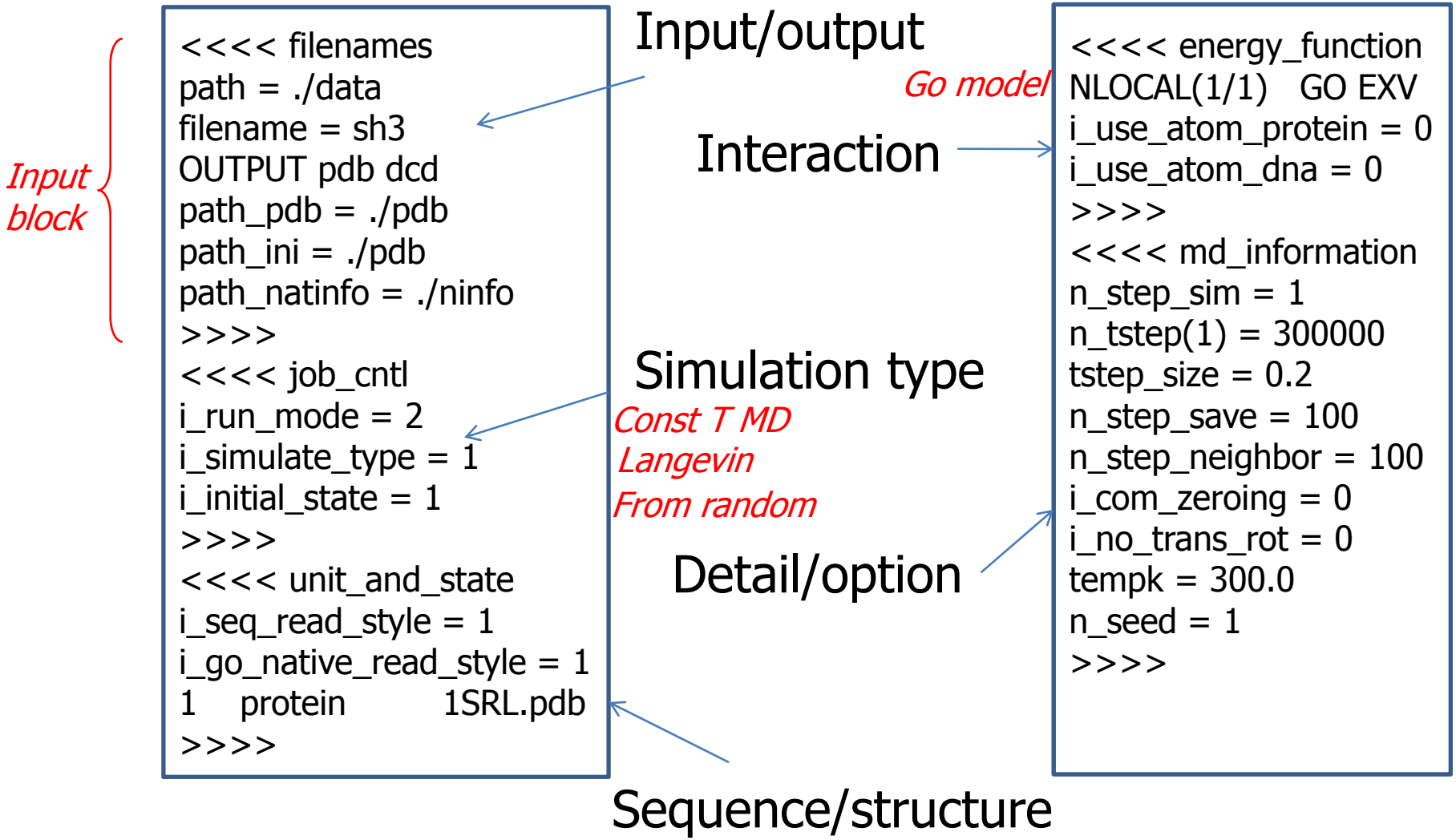


computed ϕ–values

experimental ϕ–values



Example of input file (folding simulation of src SH3)



Folding temperature of src SH3 (Auto-search of Tf)

Bi-section method

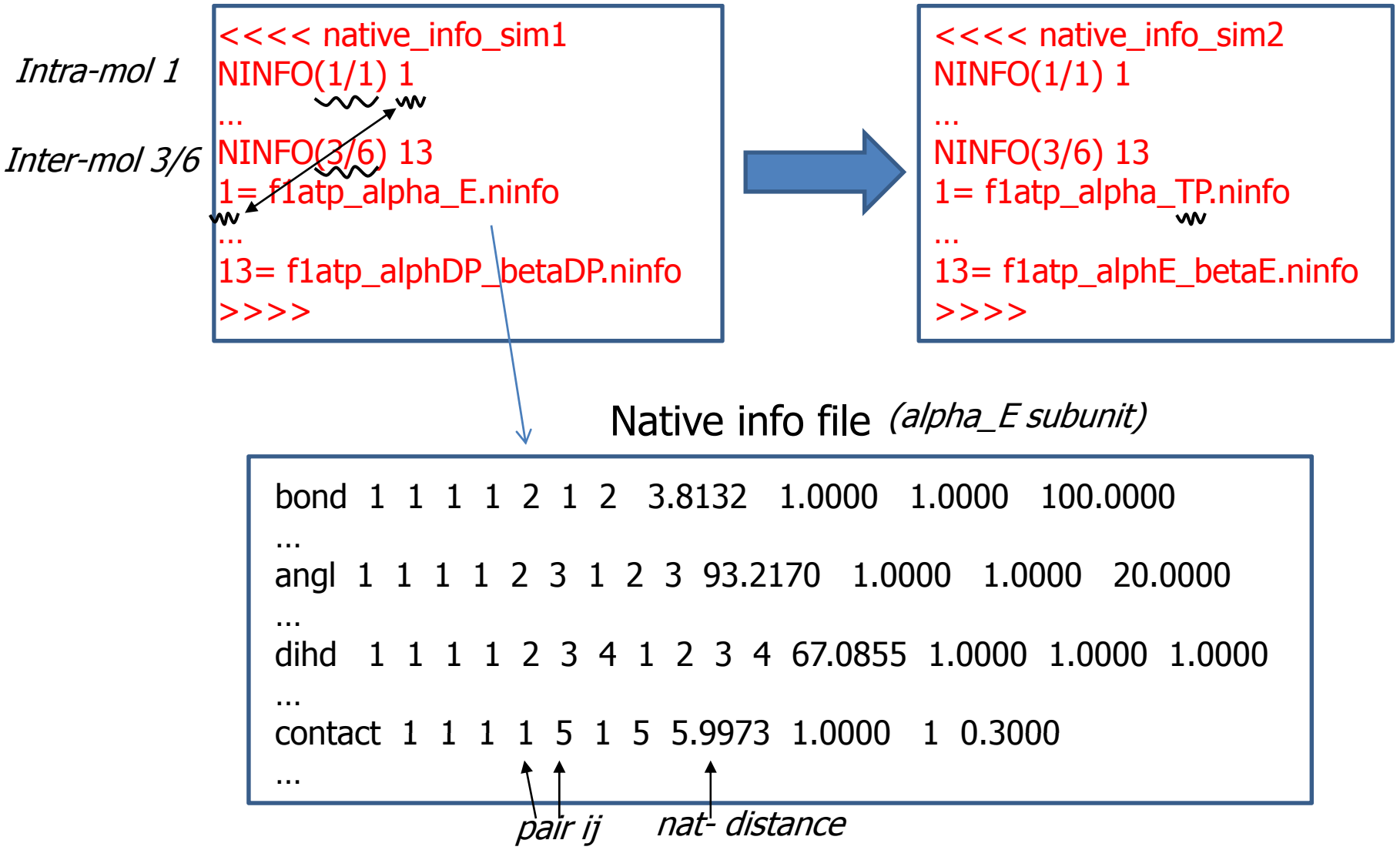
```
<<<< job_cntl
i_run_mode = 4
i_simulate_type = 1
i_initial_state = 1
>>>>
<<<< searching_tf
tempk_upper = 500.0
tempk_lower = 100.0
>>>>
```

```
**********************************
tf_out   tempk n_state d_state p_trans
tf_out 300.000     995       5       1
**********************************
tf_out   tempk n_state d_state p_trans
tf_out 400.000       1    1000       0
**********************************
tf_out   tempk n_state d_state p_trans
tf_out 350.000     166     835      78
**********************************
tf_out   tempk n_state d_state p_trans
tf_out 325.000     953      48      19
**********************************
...
...
**********************************
tf_out   tempk n_state d_state p_trans
tf_out 341.406     638     363      98
**********************************
```

Folding temperature of some proteins

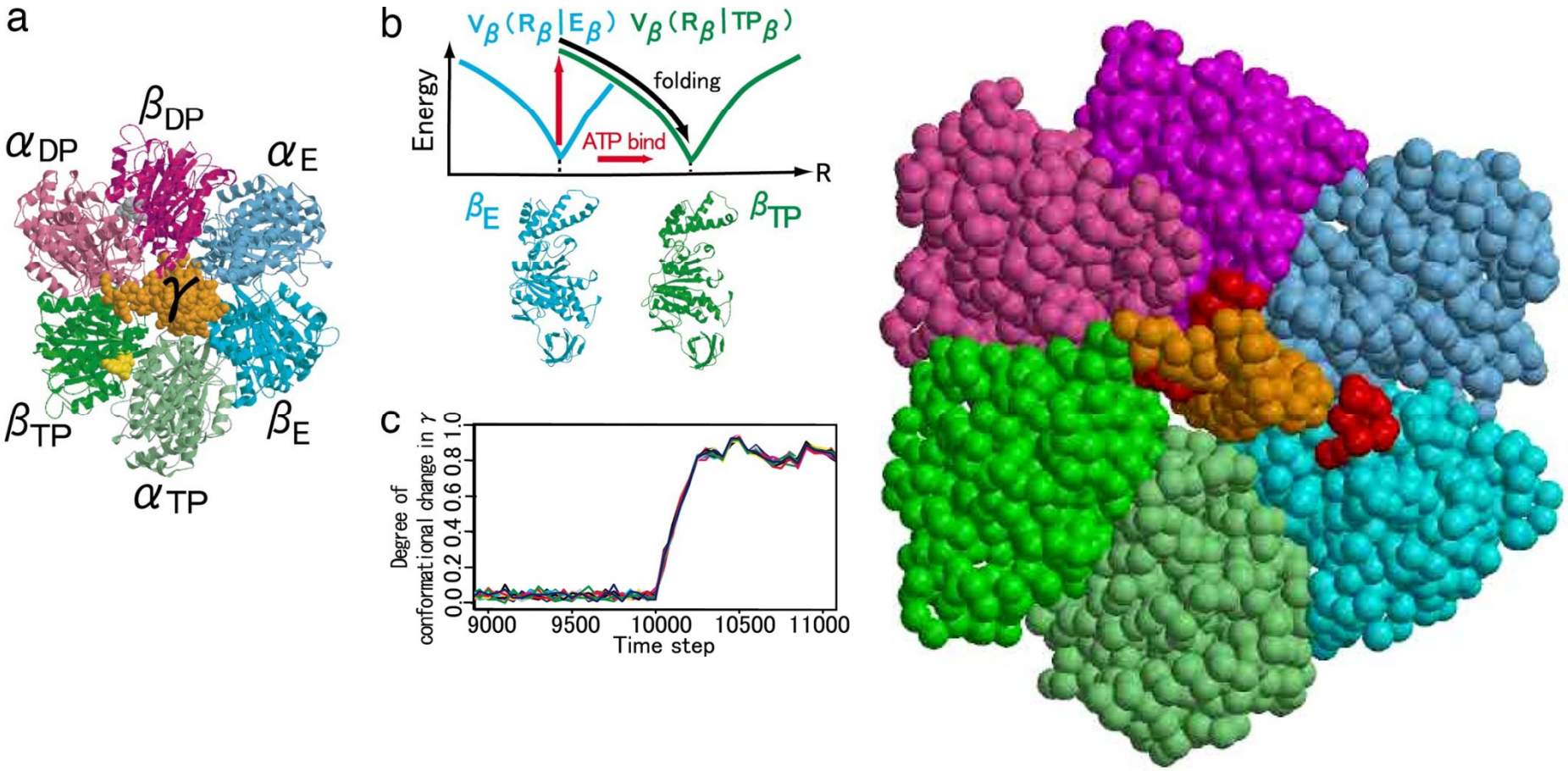
Protein	Number of amino acid	Folding temperature(K)
albumin binding domain	53	380.4
src SH3 domain	56	342.9
protein G	56	338.2
α-spectrin SH3 domain	57	360.1
Sso7d	64	332.0
protein L	78	374.2
Im9	86	382.0
cytochrom B562	106	352.2

“Switching” simulation



Rotation mechanism of F_1-ATPase by switching Go model

N. Koga, and S. Takada, PNAS (2006)



Conformational change by MBP

K. Okazaki, N. Koga, S. Takada, J.N. Onuchic, and P.G. Wolynes, PNAS (2006)

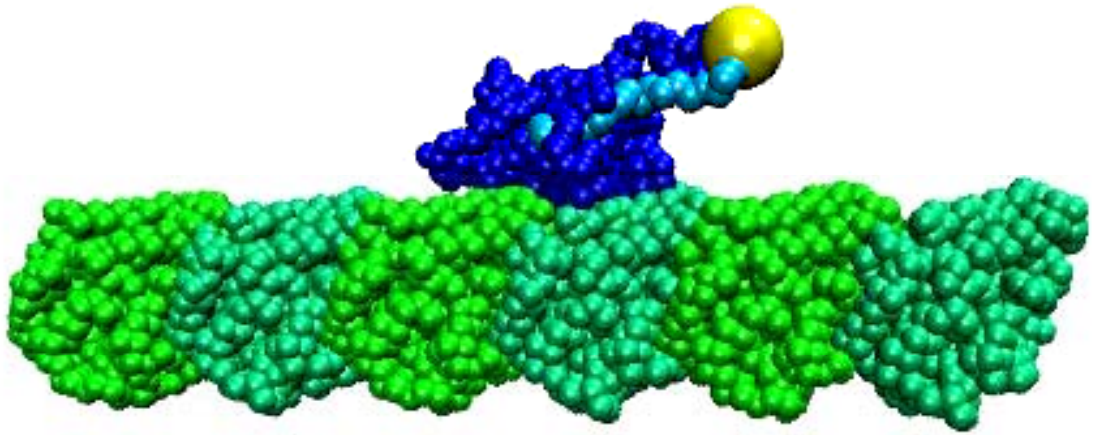
```
<<<< unit_and_state
i_seq_read_style = 1
i_go_native_read_style = 1
1a     protein        1GGG_2.pdb
1b     protein        1WDN_2.pdb
>>>>
<<<< energy_function
NLOCAL(1a/1a)    GO EXV
NLOCAL(1b/1b)    GO EXV
MULTIGO_SYSTEM(1a)   1a/1a
MULTIGO_SYSTEM(1b)   1b/1b
i_use_atom_protein = 0
i_use_atom_dna = 0
>>>>
<<<< multiple_go
bdemax_mgo = 100.0
baemax_mgo = 1.0
dihemax_mgo = 0.5
ENEGAP(1)(1)  0.0   -1.8
DELTA(1ab)    28.0
>>>>
```

Sliding movement of KIF1A

R. Kanada, et al PLOS Comp Biol (in press)

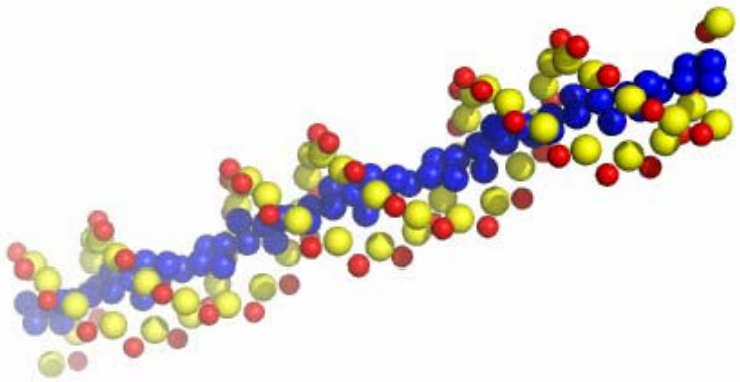
1 phase: multiple-basin (T, D)
2 phase: go(D)
3 phase: multiple-basin(D, phi)
4 phase: go(phi)
5 phase: go(T)

KIF1A:blue
tubulin:green
cargo:yellow



DNA duplex

- 30 bp oligomer of DNA
- Langevin dynamics (300K)
- [Na^+] = 69mM



```
<<<< unit_and_state
i_seq_read_style = 2
i_go_native_read_style = 3
1-2    dna          sequence
>>>>
<<<< energy_function
LOCAL(1-2)  L_BDNA
NLOCAL(1-2/1-2)    DNA ELE
i_use_atom_protein = 0
i_use_atom_dna = 0
>>>>
<<<< electrostatic
cutoff_ele = 20.0
ionic_strength = 0.069
diele_water = 78.0
>>>>
<<<< in_box
xbox = 120.0
ybox = 120.0
zbox = 120.0
boxsigma = 4.0
>>>>
```

DH

3SPN.1

Intra mol 1,2

Inter mol 1-2

Simulation of nucleosome

1KX5.pdb

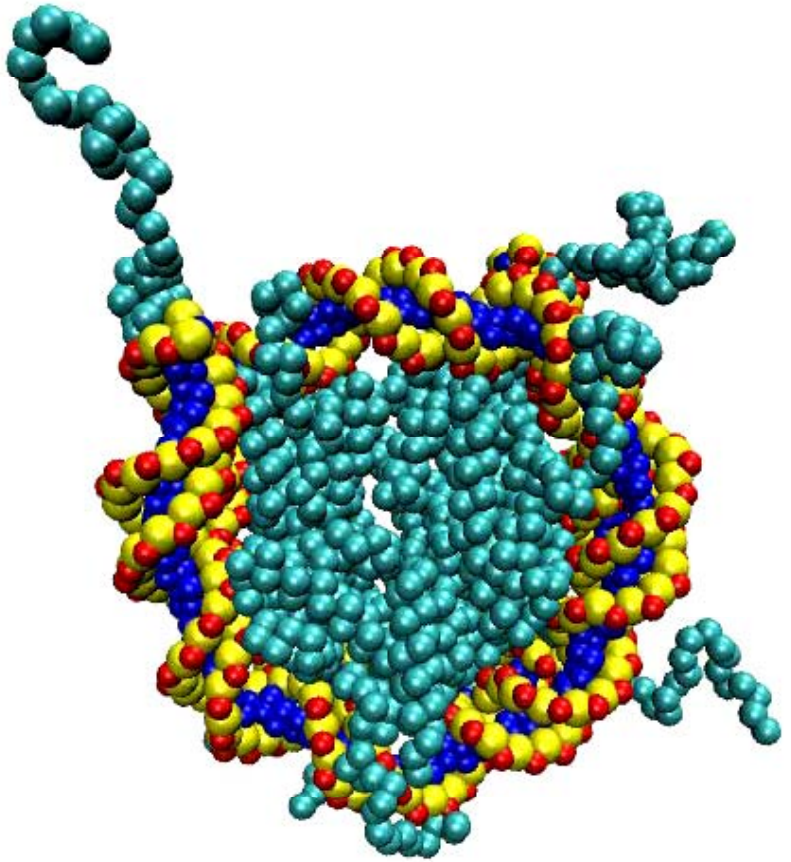
- Electrostatic interaction + Go potential

$\varepsilon_{go}^{\text{pro-dna}} = 0.5\varepsilon_{go}^{\text{pro}}$

$[Na^+]$ = 50mM

$\varepsilon_{go}^{\text{pro-dna}}$:coefficient of protein-DNA Go potential

```
<<<< energy_function
NLOCAL(1-2/1-2) ELE DNA
NLOCAL(1-2/3-10)  GO EXV ELE
NLOCAL(3-10/3-10)  GO EXV
i_use_atom_protein = 0
i_use_atom_dna = 0
>>>>
<<<< electrostatic
cutoff_ele = 20.0
ionic_strength = 0.05
diele_water = 78.0
>>>>
```



H. Kenzaki, et al unpublished data

Menu

Models

Simulation methods

Implementation

Selected applications

In-progress models & methods

Acknowledgement

CafeMol development has been supported by Research and Development of the Next-Generation Integrated Simulation of Living Matter, a part of the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology.