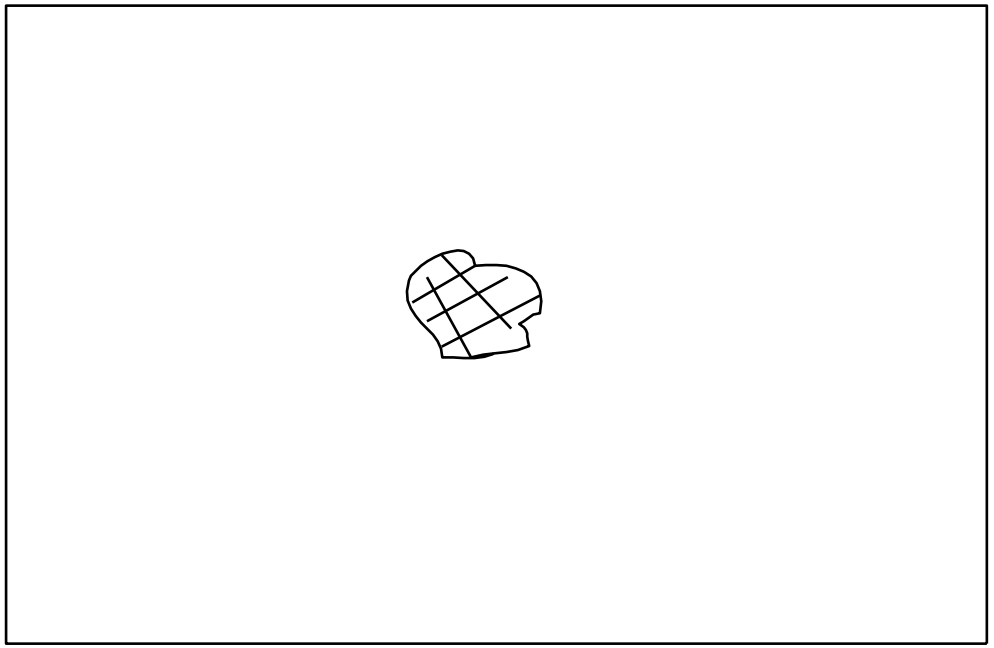
バイオスーパーコンピューティング名古屋2015
名古屋工業大学、January 22, 2015

拡張アンサンブル法による生体分子シミュレーション

岡本 祐幸（ Yuko OKAMOTO ）

名古屋大学 大学院理学研究科 物質理学専攻(物理系)
同 大学院理学研究科 構造生物学研究センター
同 大学院工学研究科 計算科学連携教育研究センター
同 情報基盤センター
e-mail: okamoto{a}phys.nagoya-u.ac.jp
URL: http://www.tb.phys.nagoya-u.ac.jp/

Canonical Ensemble
カノニカルアンサンブル

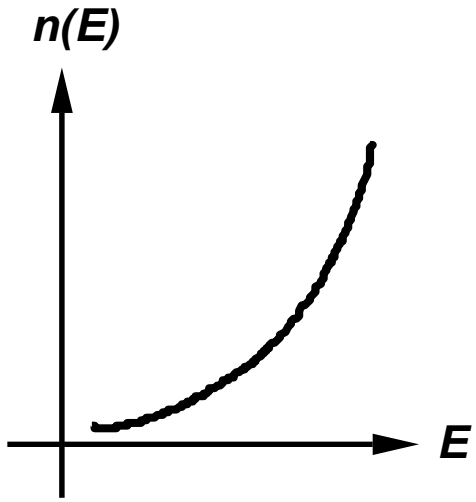


System in Heat Bath (Exchange Energy w/ Heat Bath)

熱浴中の系（熱浴とエネルギーをやりとり）:

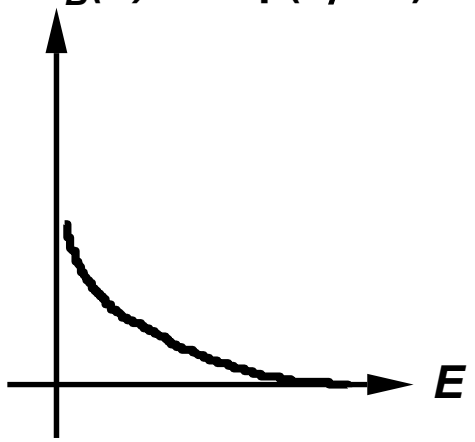
T = const = 一定

Canonical Ensemble at Temperature T



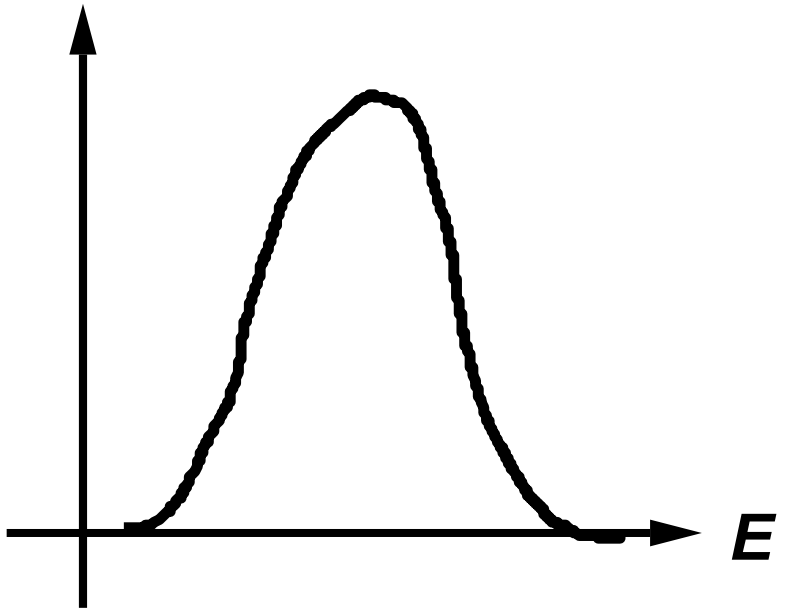
Density of States

$W_B(E) = \exp(-\beta E)$



Boltzmann Factor

$P_B(E) = n(E)W_B(E)$

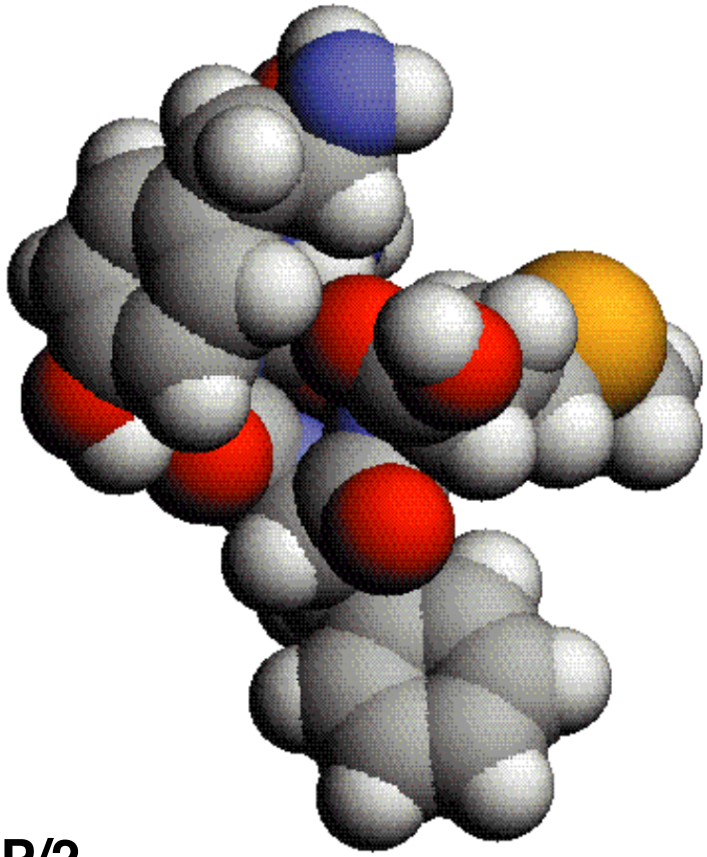


Canonical Probability Distribution

Met-Enkephalin:
Global Minimum Structure in Gas Phase

Tyr-Gly-Gly-Phe-Met

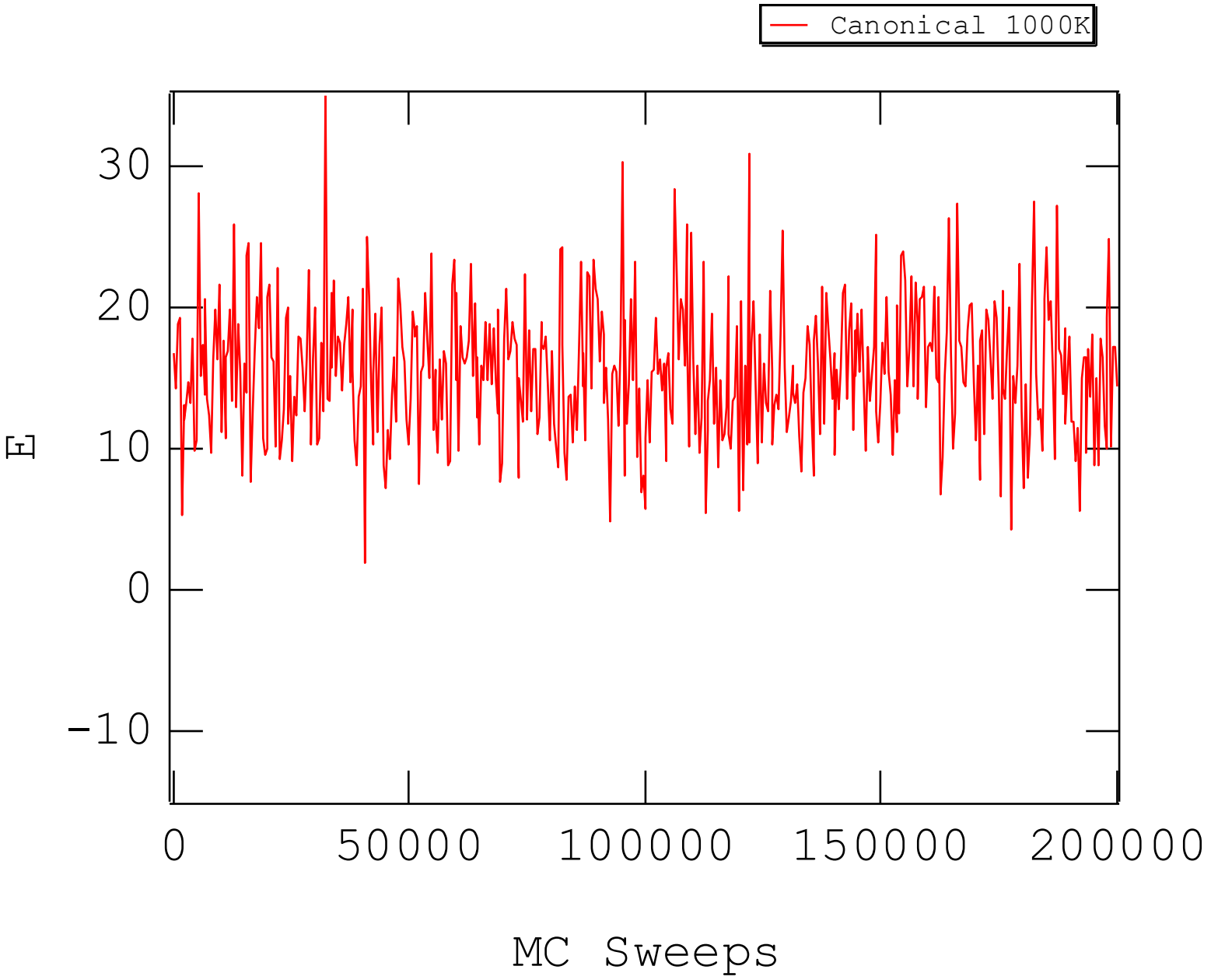
(N = 5)

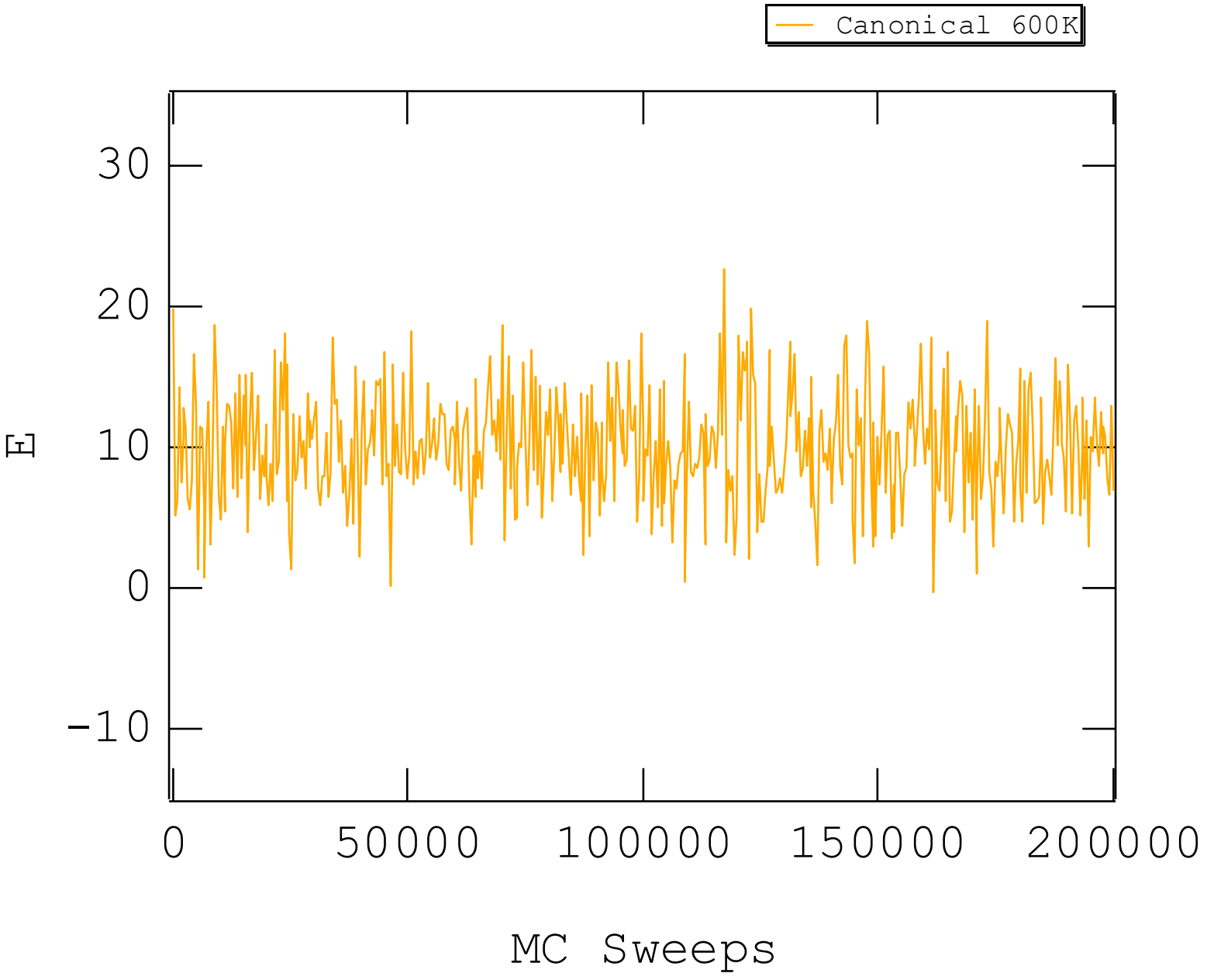


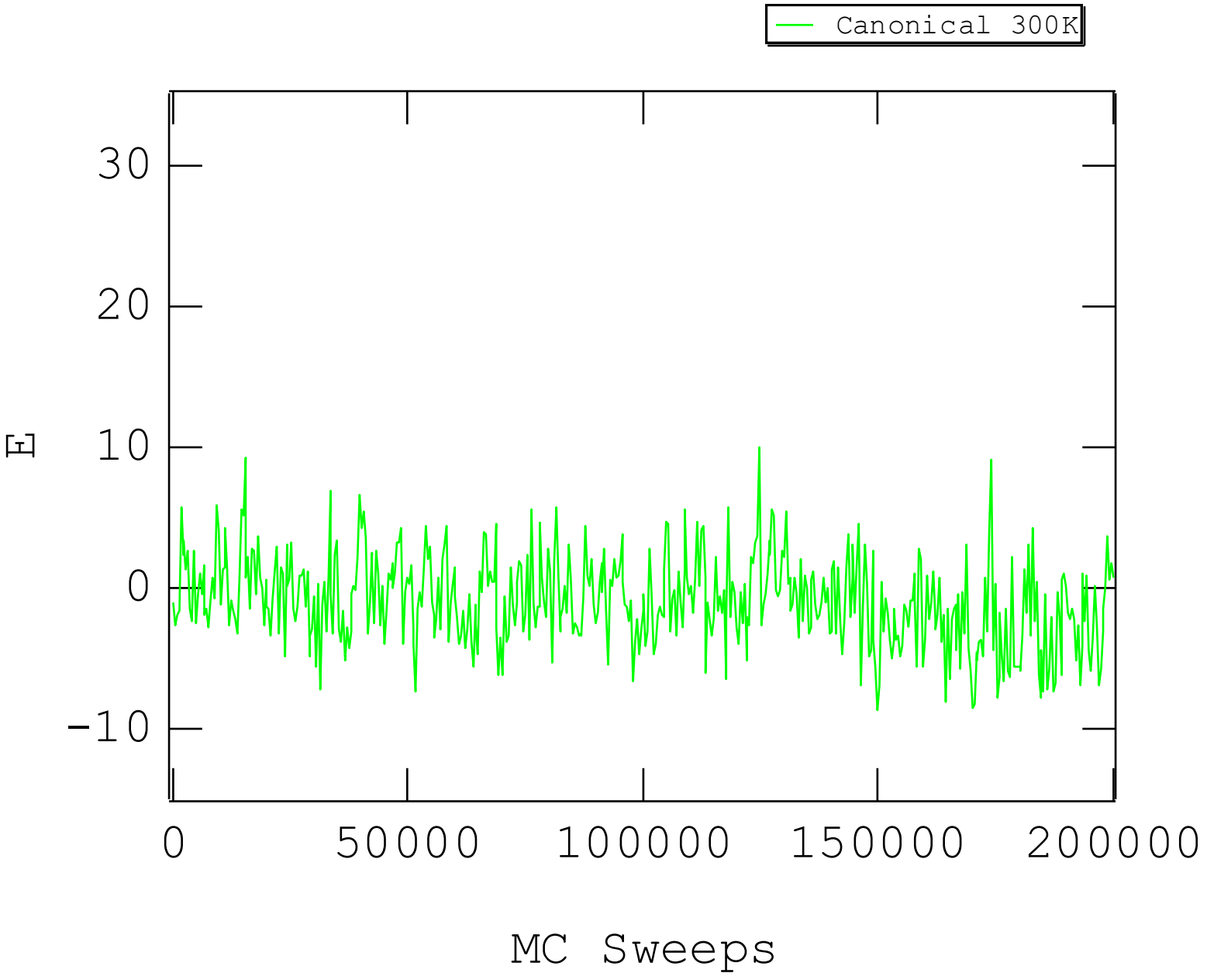
E = -12 kcal/mol

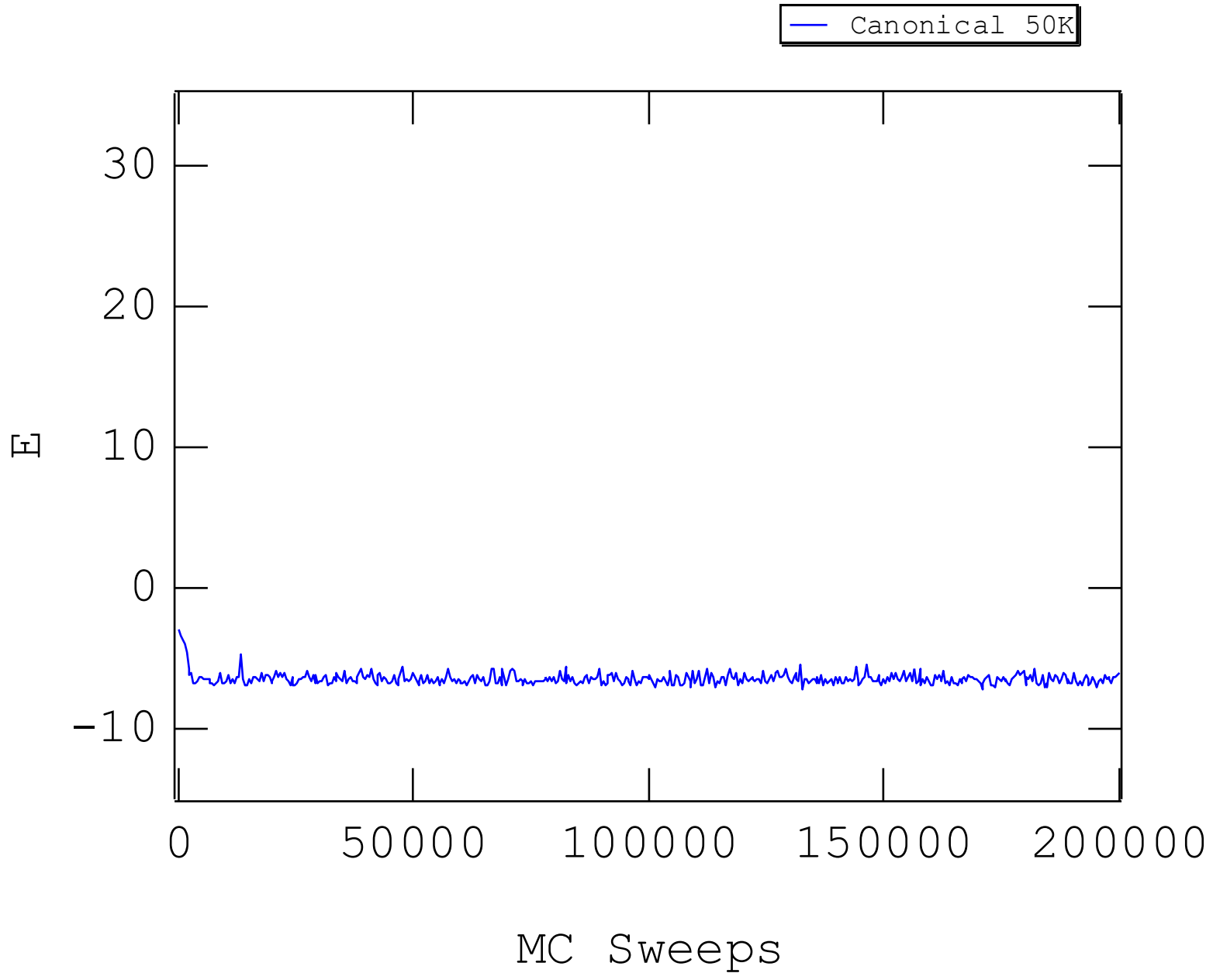
potential energy: ECEPP/2

degrees of freedom: dihedral angles


Canonical 1000K
30
20
10
0
-10
E
0
50000
100000
150000
200000
MC Sweeps


Canonical 600K
30
20
10
0
-10
E
0
50000
100000
150000
200000
MC Sweeps


Canonical 300K
30
20
10
0
-10
E
0
50000
100000
150000
200000
MC Sweeps


Canonical 50K
30
20
10
0
-10
E
0
50000
100000
150000
200000
MC Sweeps

Simulated Annealing

徐冷法（Simulated Annealing)

S. Kirkpatrick, C. Gelatt, Jr. & M. Vecchi, *Science* **220**, 671 (1983).

Reproduce a Crystal-Making Process on a Computer

Application of Simulated Annealing to Systems of Biopolymers

- H. Kawai, T. Kikuchi & Y.O., *Protein Eng*. **3**, 85 (1989).

See also:

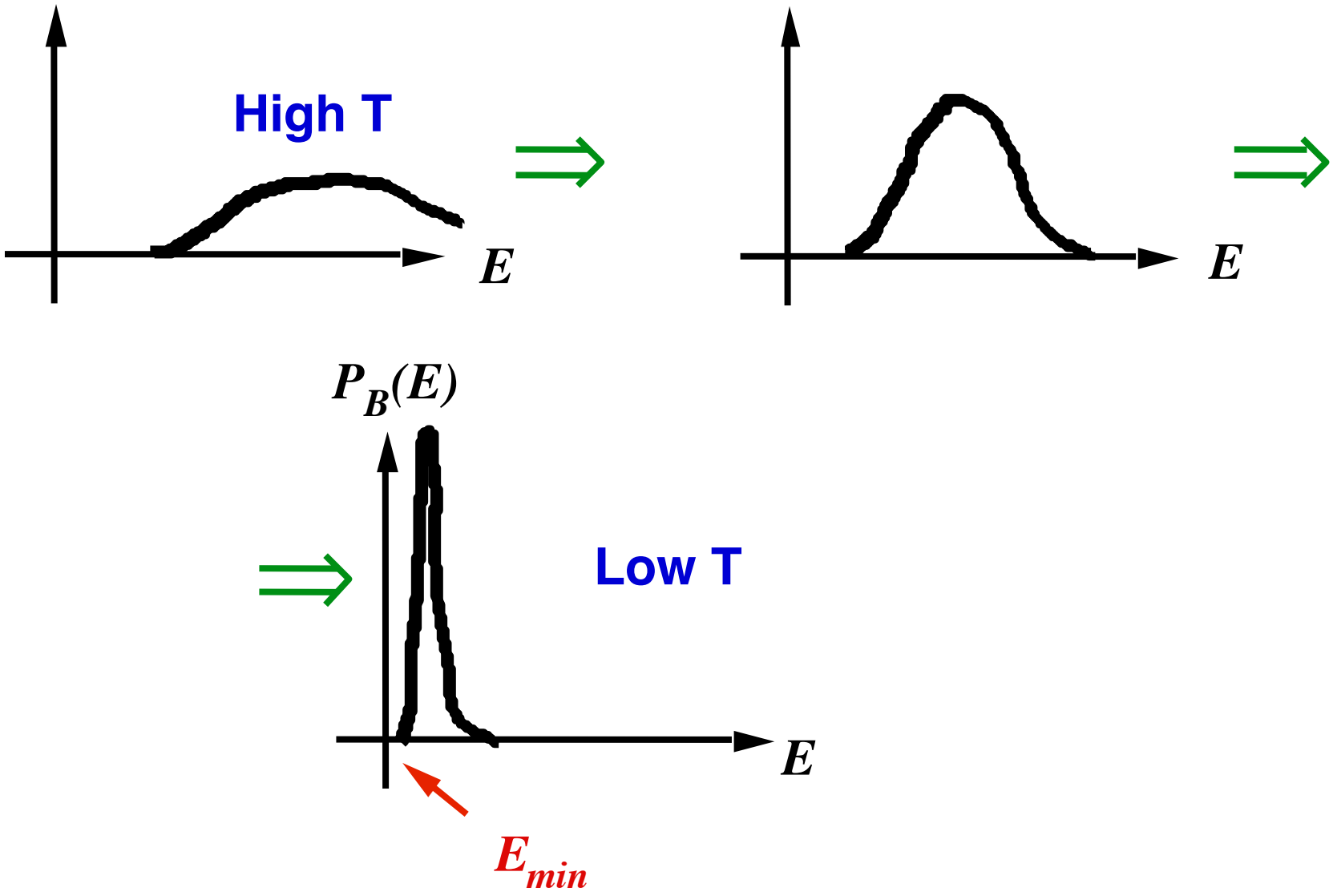
- S. Wilson, W. Cui, J. Moskovitz & K. Schmidt, *Tetrahedron Lett.* **29**, 4373 (1988).
- C. Wilson & S. Doniach, *Proteins* **6**, 193 (1989).
- A. Brunger, *J. Mol. Biol.* **203**, 803 (1988).
- M. Nilges, G. Clore & A. Gronenborn, *FEBS Lett.* **229**, 317 (1988).

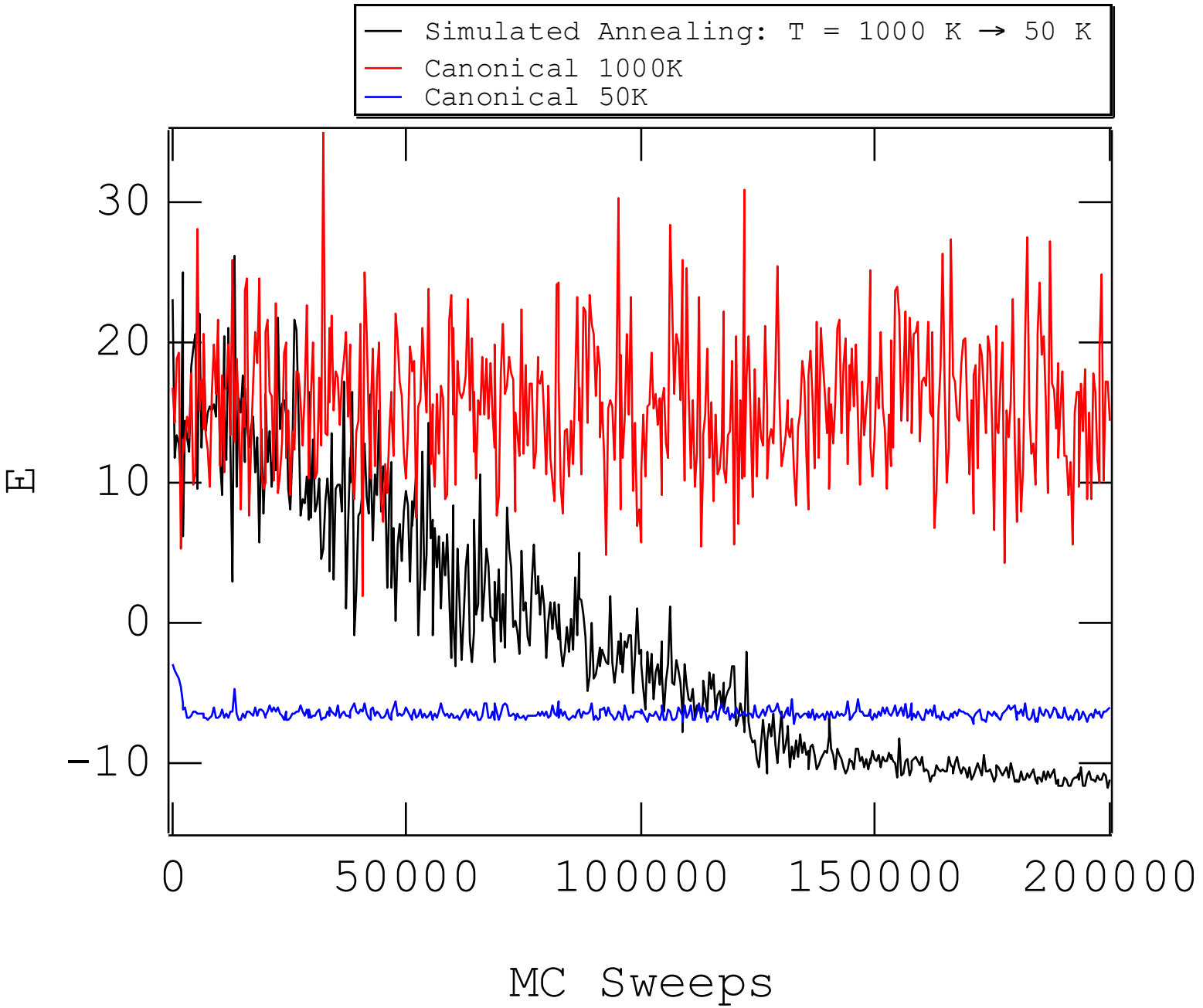
For a review see:

- Y.O., *Recent Res. Devel. In Pure & Applied Chem.* **2**, 1 (1998).

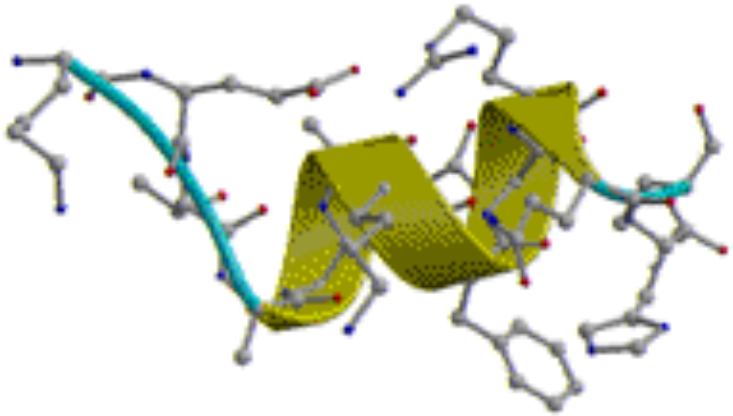
Simulated Annealing（徐冷法）

$P_B(E) = n(E)W_B(E)$




Simulated Annealing: T = 1000 K → 50 K
Canonical 1000K
Canonical 50K
E
30
20
10
0
-10
0
50000
100000
150000
200000
MC Sweeps

C-Peptide (N = 13): X-Ray



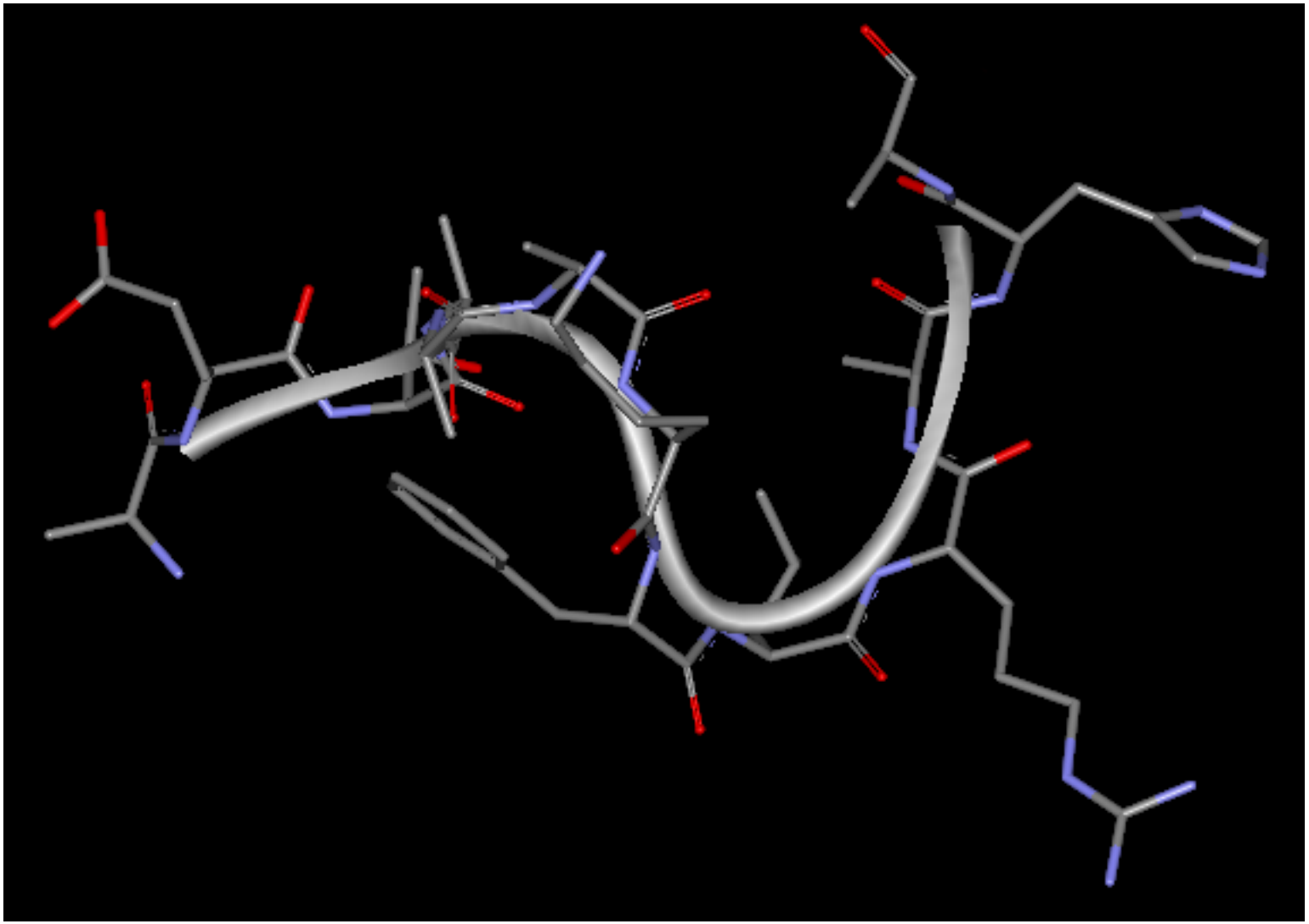
C-Peptide of Ribonuclease A

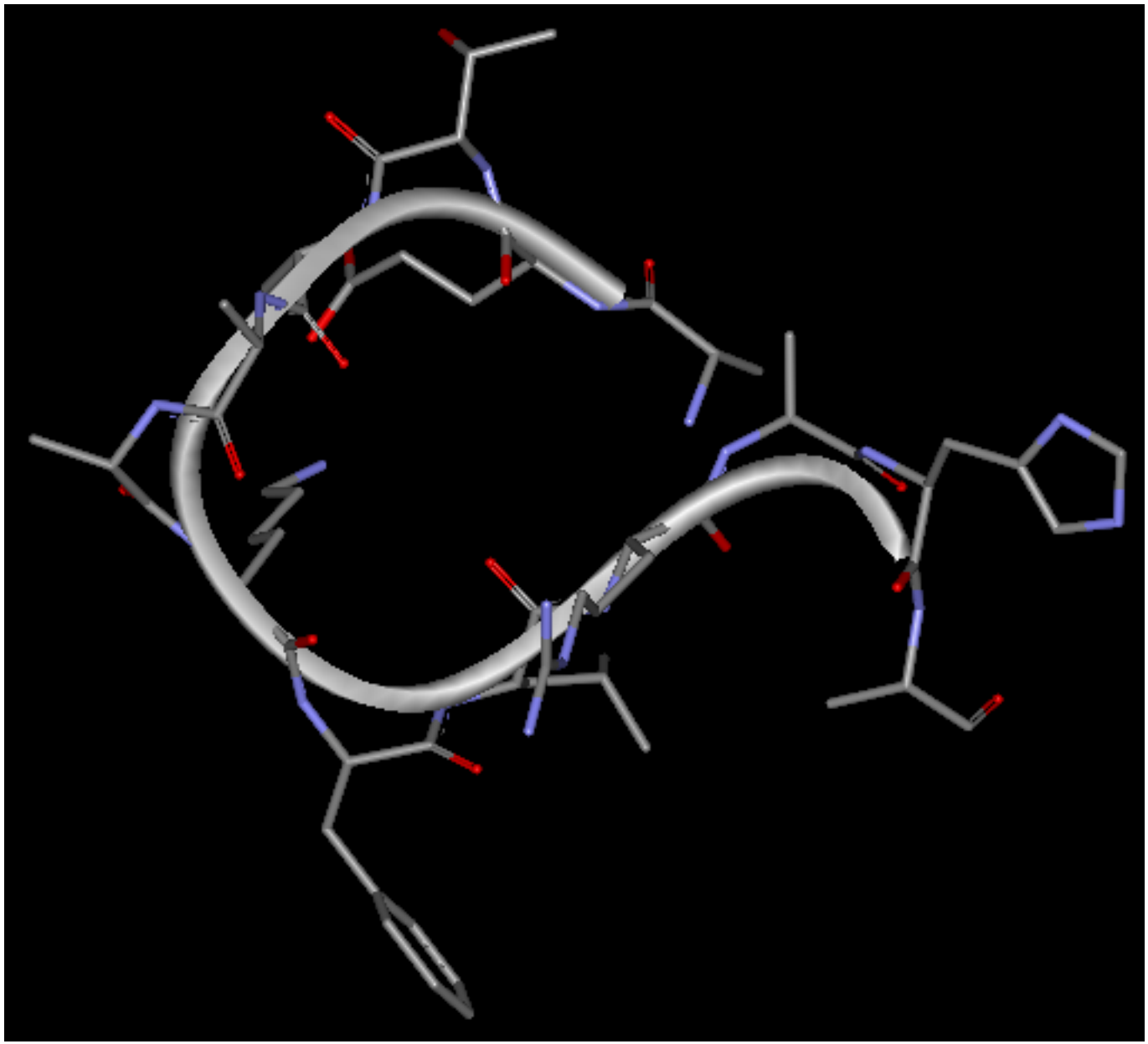
Amino-Acid Sequence:

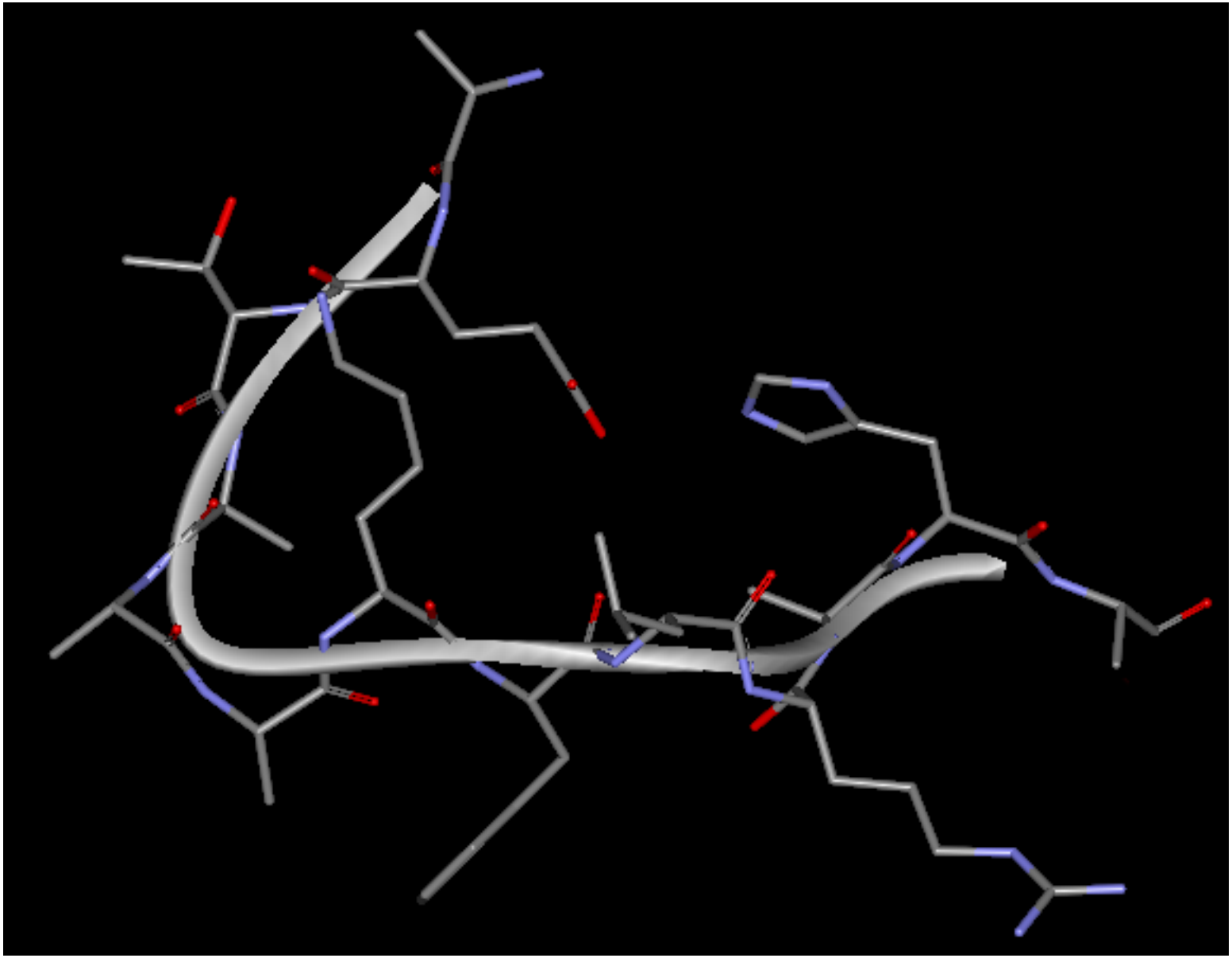
Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-

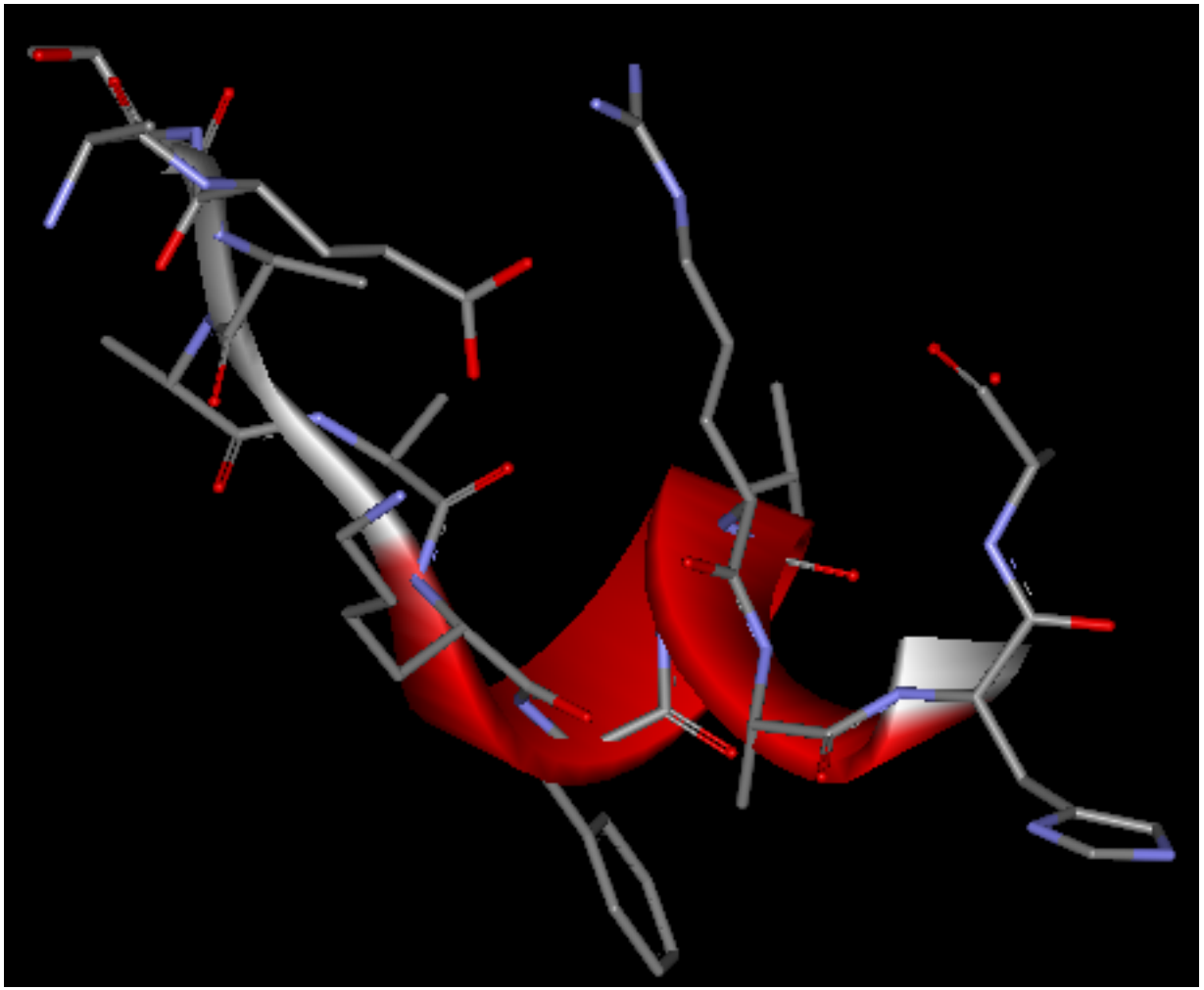
Gln-His-Met

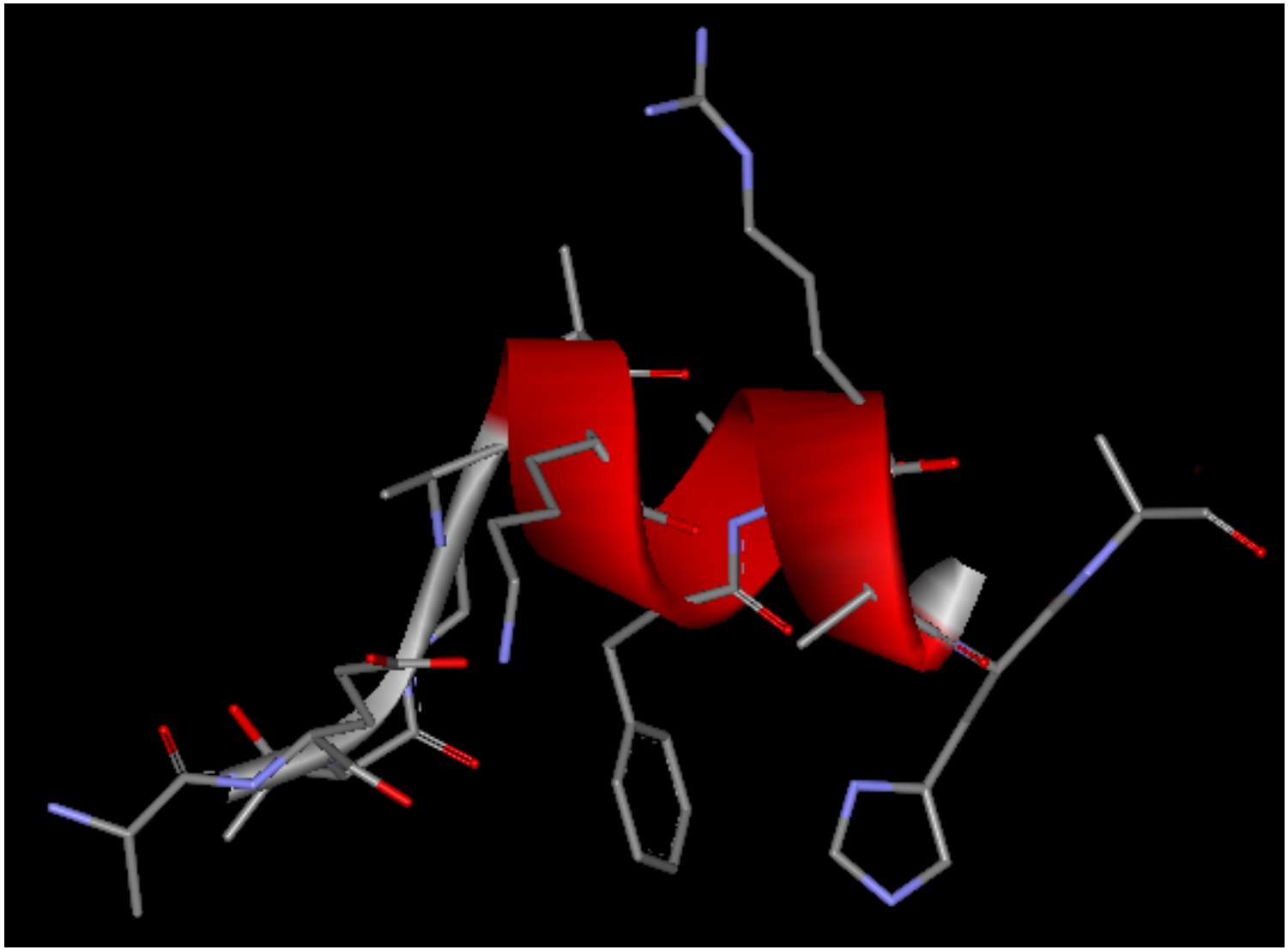
M. Masuya & Y.O., unpublished.

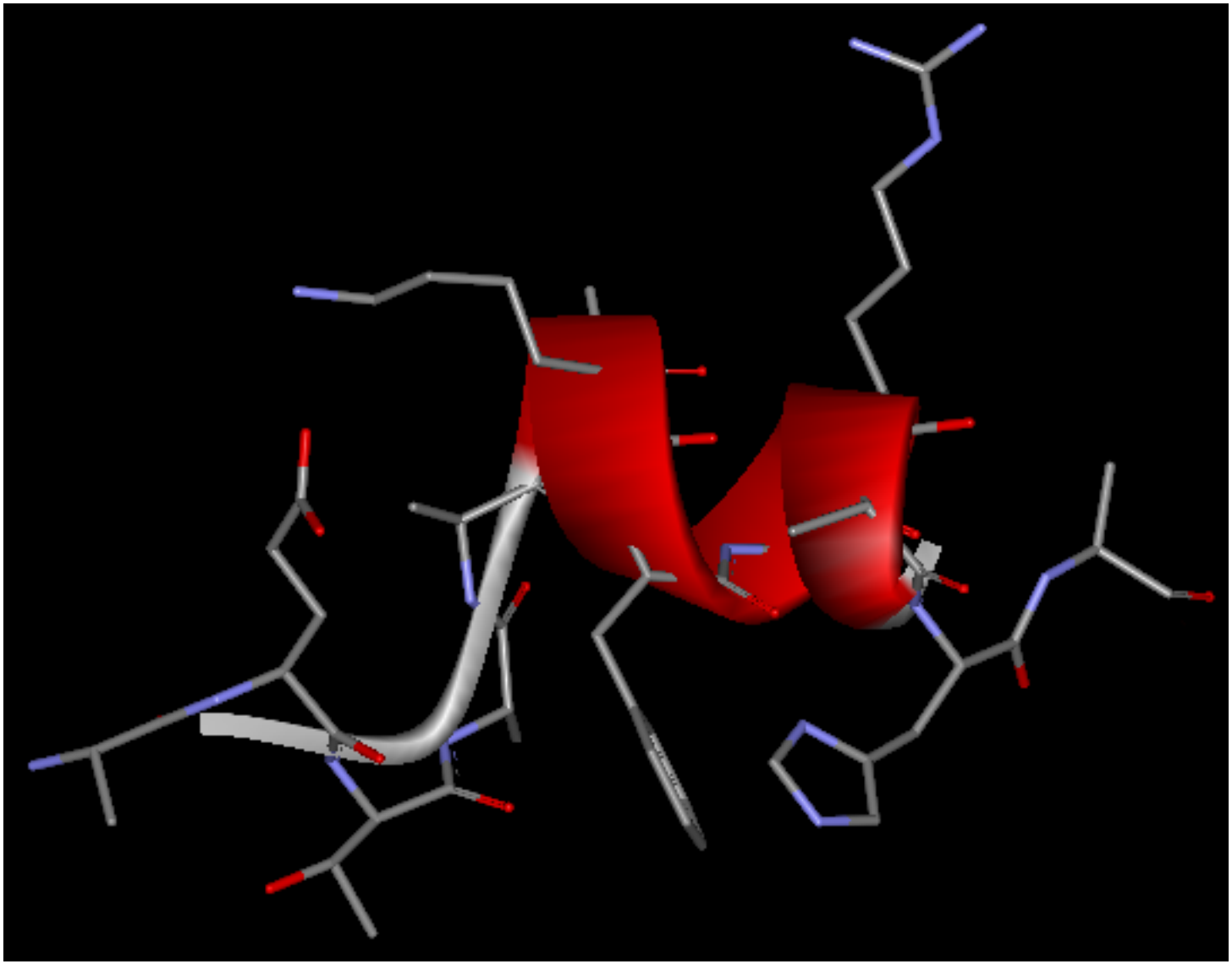


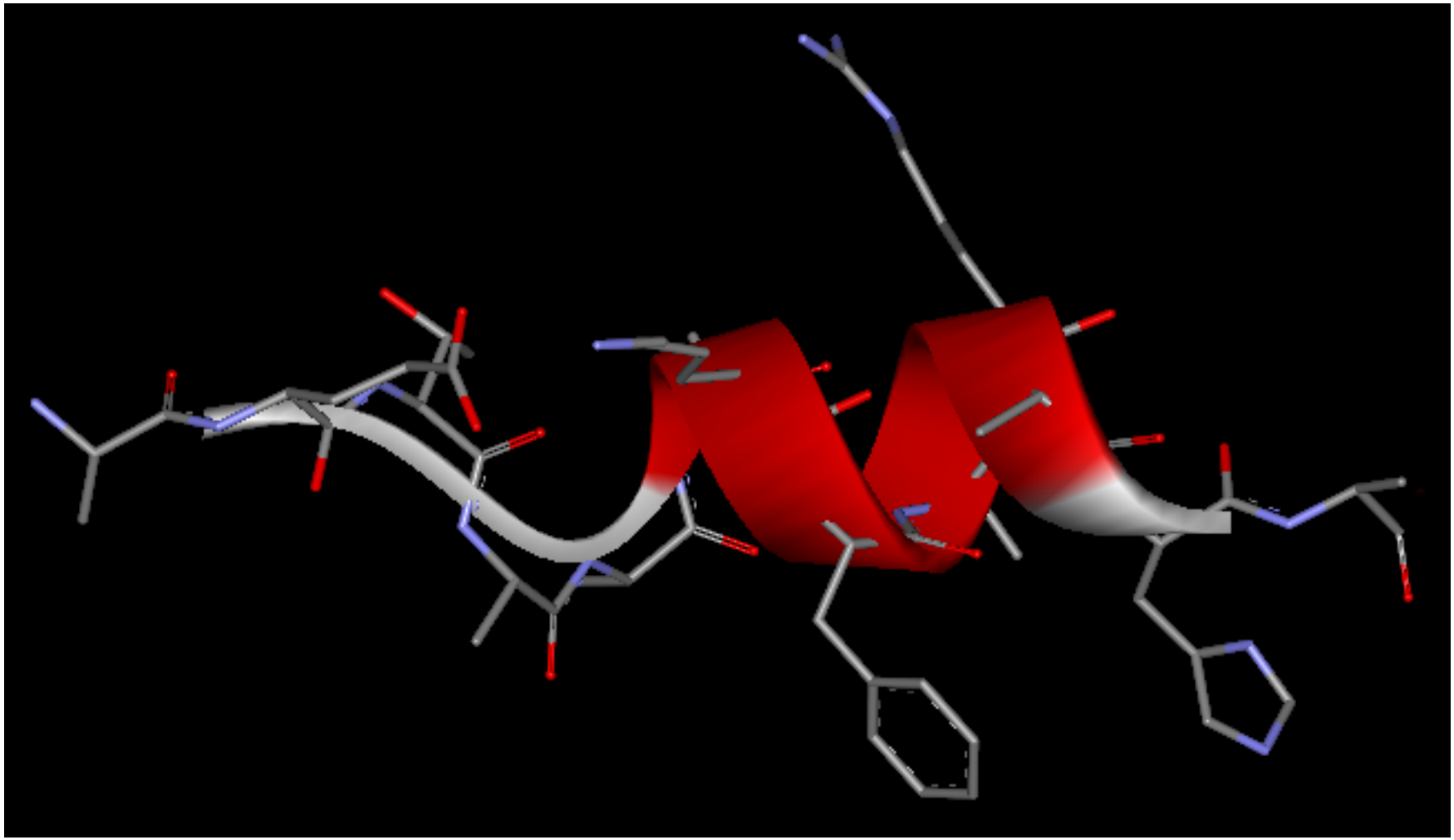


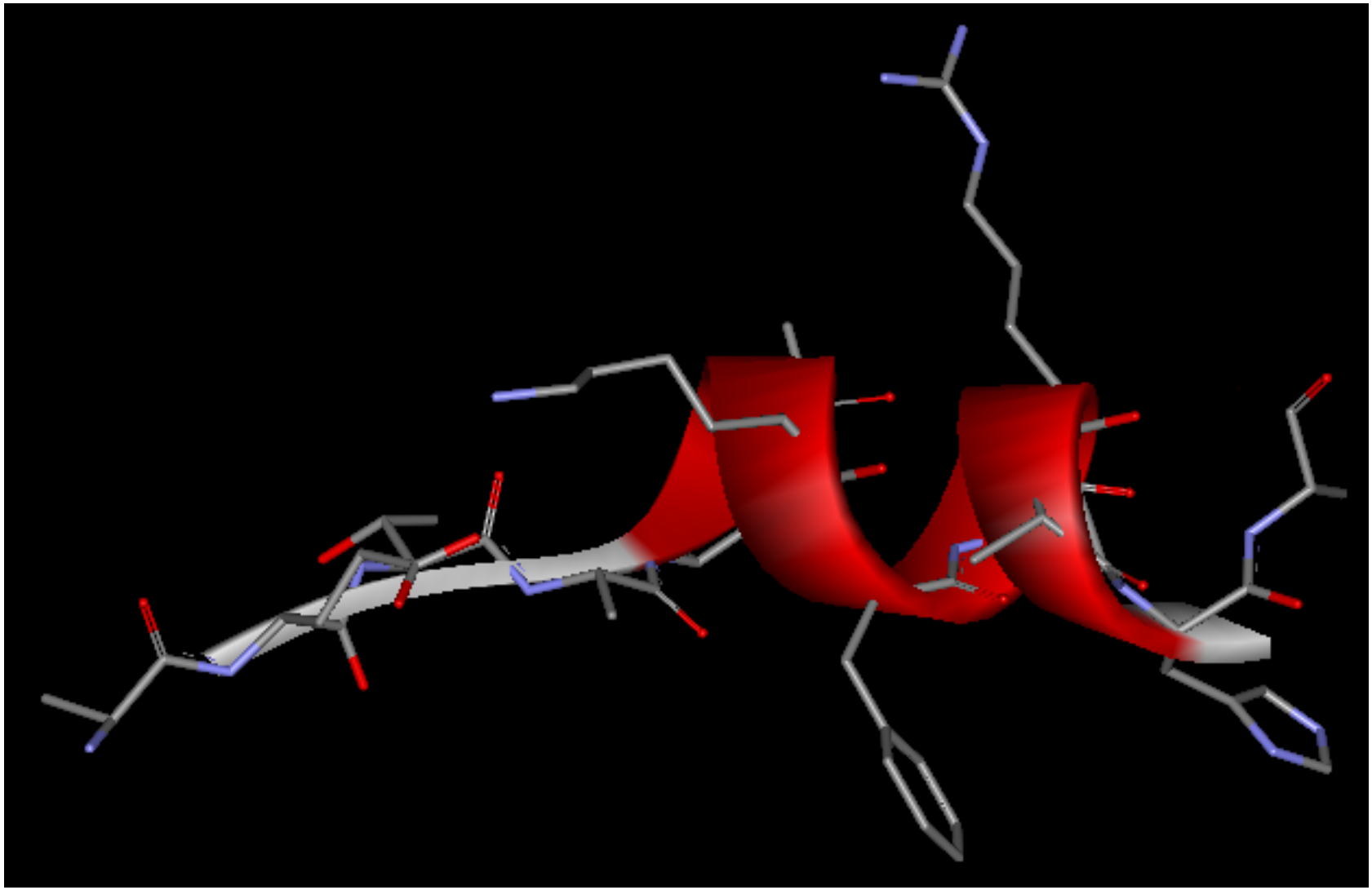


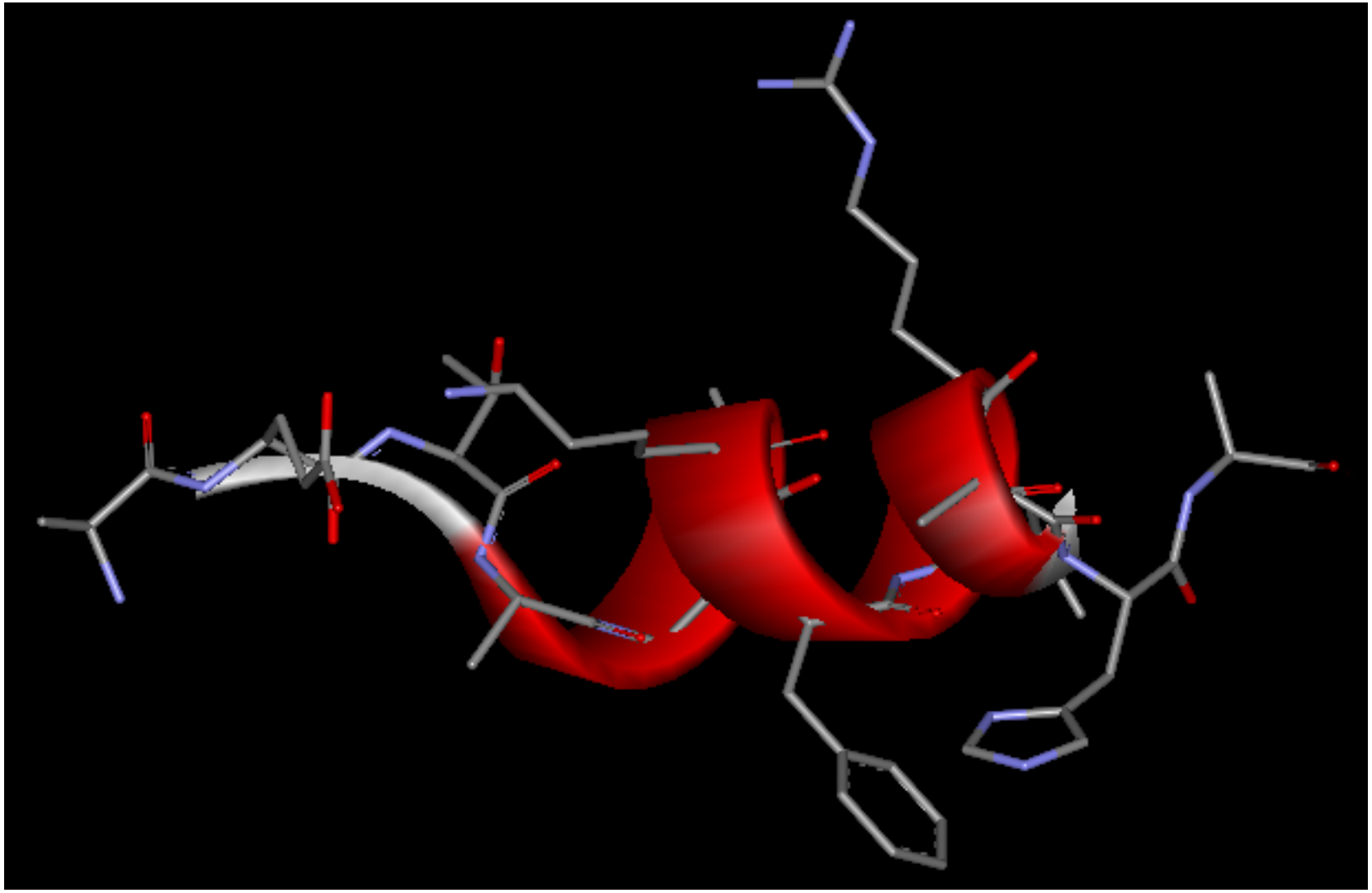


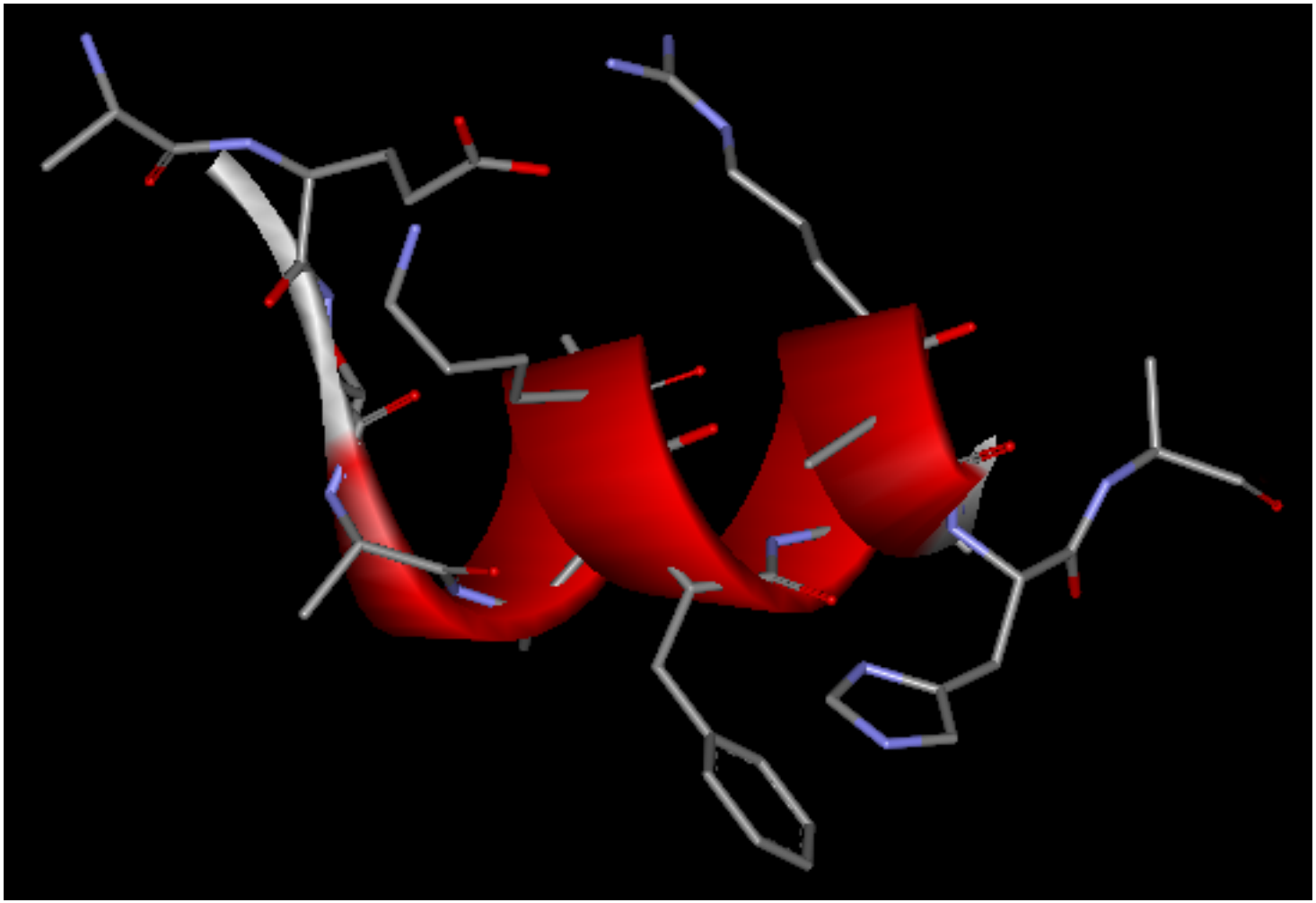




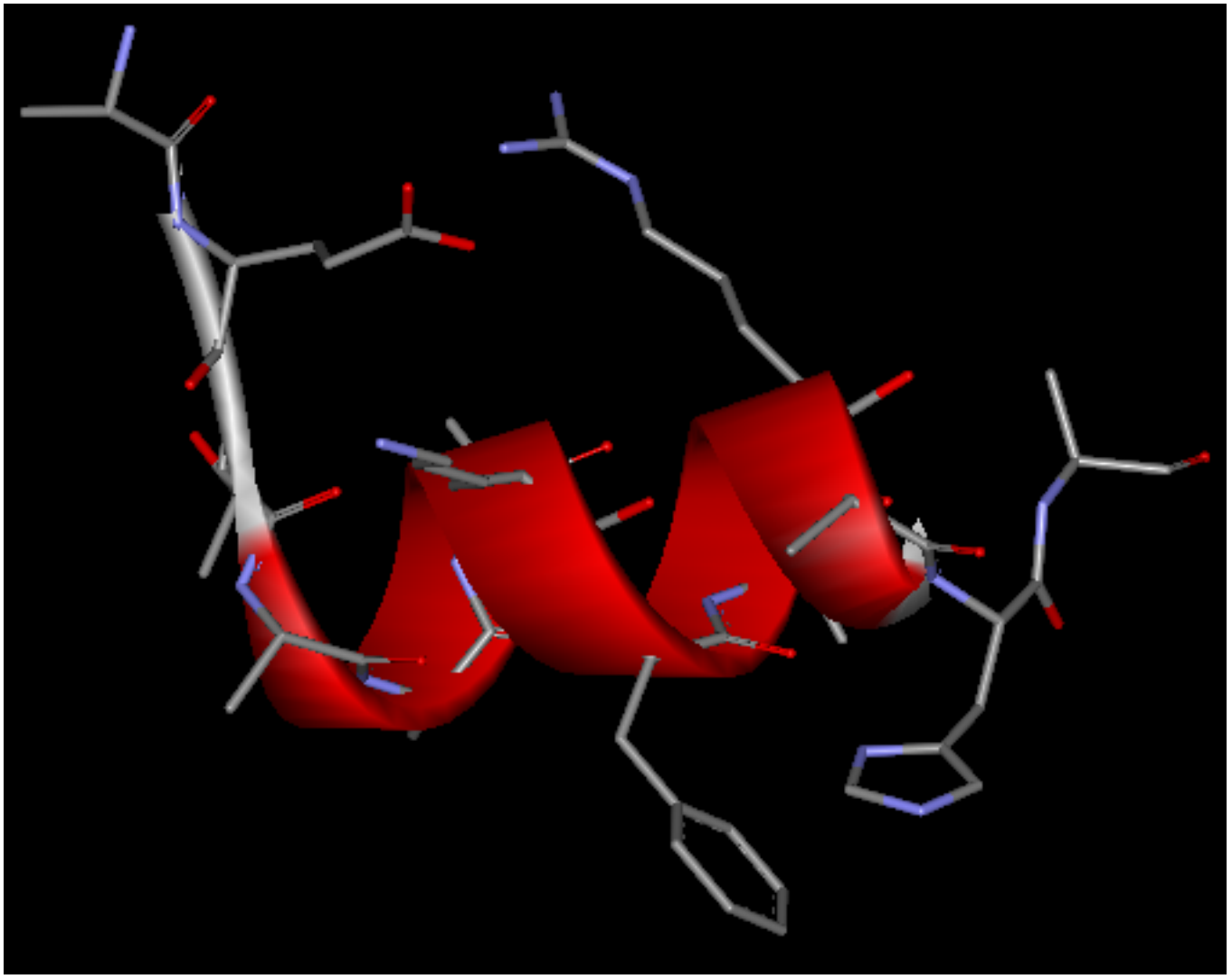




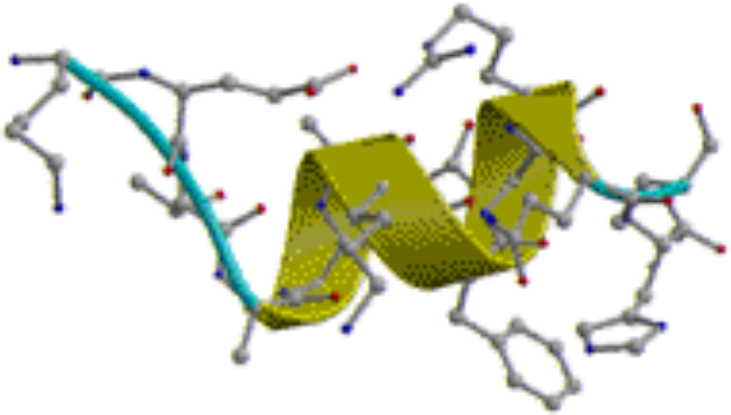




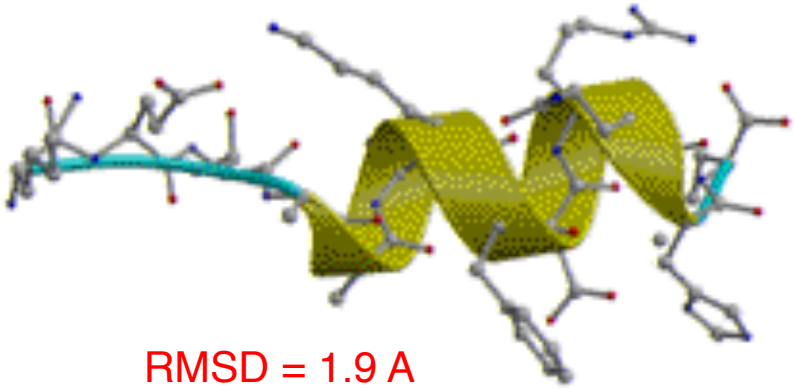
M. Masuya & Y.O., unpublished.



C-Peptide (N = 13): X-Ray

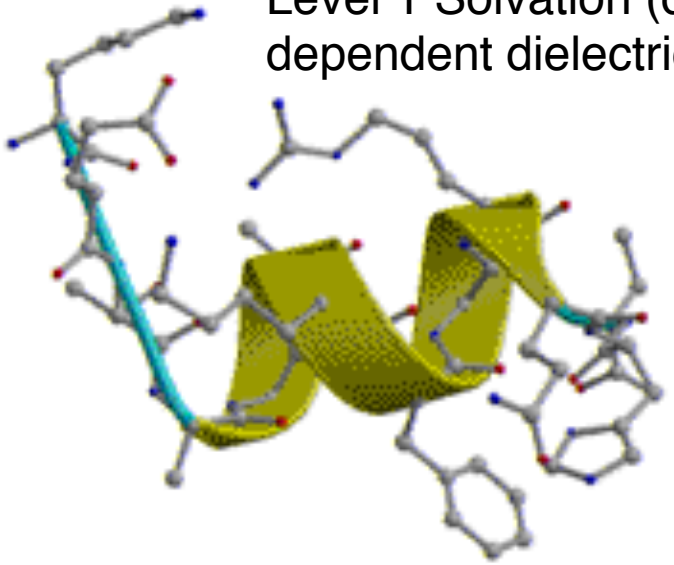


Level 0 Solvation (Gas Phase)



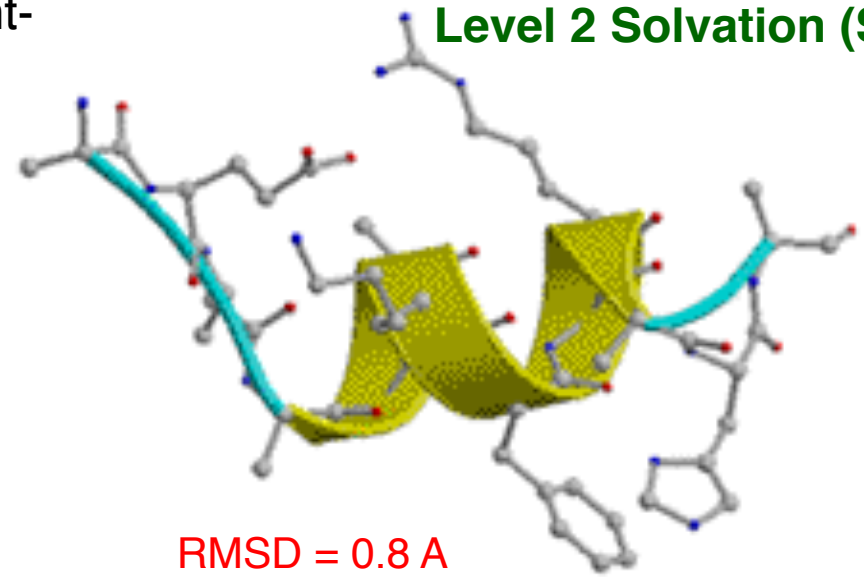
RMSD = 1.9 A

Level 1 Solvation (distant-dependent dielectric)



RMSD = 1.4 A

Level 2 Solvation (SASA)



RMSD = 0.8 A

M. Masuya & Y.O., unpublished.

Problems of Simulated Annealing (SA)

Because we lower the temperature during the simulation, we are always breaking thermal equilibrium (and detailed balance conditions).

Hence, thermodynamic quantities obtained from SA are not reliable.

Generalized-Ensemble Algorithms (MUCA, ST, and REM)

Generalized-Ensemble Algorithm（拡張アンサンブル法）

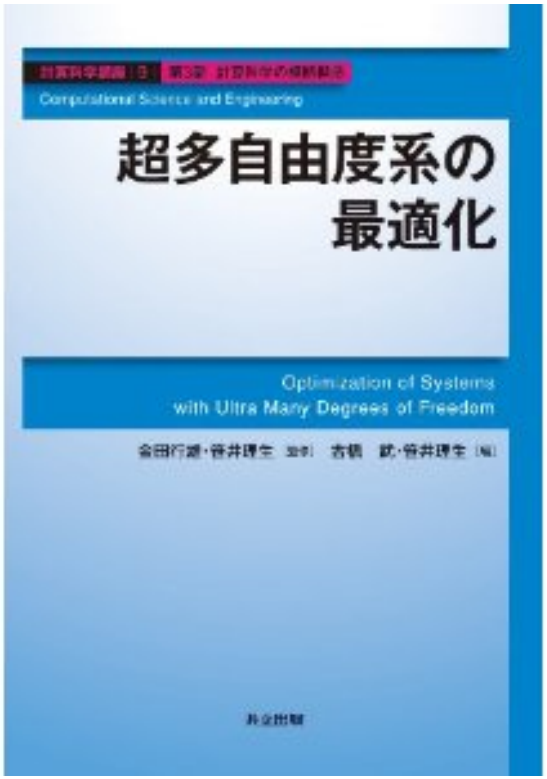
- **Generic Term for Simulation Methods that Greatly Enhance Conformational Sampling** [e.g., **Multicanonical Algorithm, Wang-Landau, Simulated Tempering, Replica-Exchange Method**, etc.]
- **Based on Non-Boltzmann Weight Factors**

  Realize random walks in potential energy and/or any other physical quantities (OR their conjugate parameters)
- **Histogram Reweighting Techniques**

  Can obtain thermodynamic quantities for a wide range of temperature and/or other parameter values from a single simulation run

REVIEWS: U.H.E. Hansmann & Y.O., *Curr. Opin. Struct. Biol.* **9**, 177 (1999);
A. Mitsutake, Y. Sugita, & Y.O., *Biopolymers* **60**, 96 (2001);
Y.O., *J. Mol. Graphics Modell.* **22**, 425 (2004);
Y. Sugita, A. Mitsutake, & Y.O., in *Lecture Notes in Physics*, W. Janke (ed.) (Springer-Verlag, Berlin, 2008) pp. 369-407;
H. Okumura, S.G. Itoh, & Y.O., in *Practical Aspects of Computational Chemistry II: An Overview of the Last Two Decades and Current Trends,* J. Leszczynski and M.K. Shukla (eds.) (Springer, Dordrecht, 2012) pp. 69-101;
A. Mitsutake, Y. Mori, and Y.O., in *Biomolecular Simulations: Methods and Protocols*, L. Monticelli and E. Salonen (eds.) (Humana Press, New York, 2012) pp. 153-195;
岡本祐幸、「拡張アンサンブル法」、古橋武・笹井理生 編、*計算科学講座* 第9巻「超多自由度系の最適化」第2章（共立出版, 2013) pp. 119-241.

岡本祐幸、「拡張アンサンブル法」、
計算科学講座 第9巻「超多自由度系の最適化」第2章
（古橋武、笹井理生編、共立出版、2013）pp.
119-241.

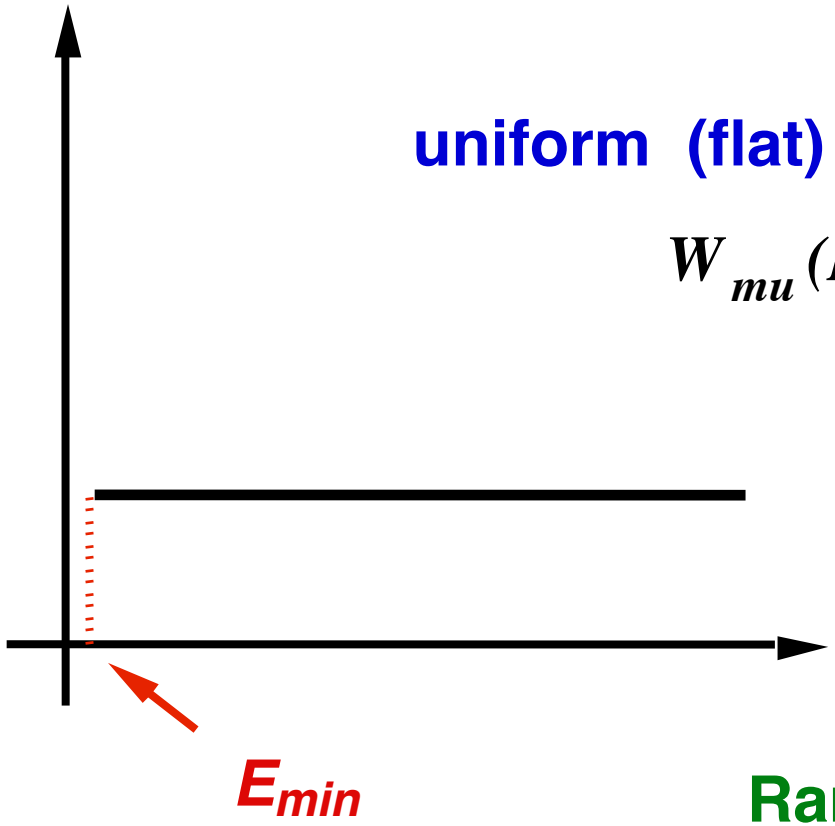


Multicanonical Algorithm

B. Berg & T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).

B. Berg & T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).

$$P_{mu}(E) = n(E)W_{mu}(E) = \text{const}$$



uniform (flat) distribution in energy

$$W_{mu}(E) = n(E)^{-1}$$

Random Walk in Energy Space

Canonical Ensemble

$$W_{\mathrm{B}}(x;T) = e^{-\beta E(x)}$$

MC version:

$$w(x \to x') = \min\left(1, \frac{W_{\mathrm{B}}(E';T)}{W_{\mathrm{B}}(E;T)}\right) = \min\left(1, \exp(-\beta \Delta E)\right)$$

Multicanonical Ensemble

$$W_{\mathrm{mu}}(E) = \frac{1}{n(E)}$$

MC version:

$$w(x \to x') = \min\left(1, \frac{W_{\mathrm{mu}}(E')}{W_{\mathrm{mu}}(E)}\right) = \min\left(1, \frac{n(E)}{n(E')}\right)$$

Generalized-Ensemble Algorithms have been developed in MC algorithms

Canonical Ensemble

MD version:

$$\begin{cases} m\ddot{\boldsymbol{q}}_i = -\dfrac{\partial E}{\partial \boldsymbol{q}_i} - \dfrac{\dot{s}}{s} m\dot{\boldsymbol{q}}_i = \boldsymbol{f}_i - \dfrac{\dot{s}}{s} m\dot{\boldsymbol{q}}_i \\ Q\ddot{s} = s\left[\sum_i m\dot{\boldsymbol{q}}_i^2 - 3Nk_BT\right] + Q\dfrac{\dot{s}^2}{s} \end{cases}$$

$$W_B(x;T) = e^{-\beta E(x)}$$

Multicanonical Ensemble

MD version:

$$W_{mu}(E) = \frac{1}{n(E)} = \exp(-\beta_0 E_{mu}(E))$$

$$\begin{cases} m\ddot{\boldsymbol{q}}_i = -\dfrac{\partial E_{mu}}{\partial \boldsymbol{q}_i} - \dfrac{\dot{s}}{s} m\dot{\boldsymbol{q}}_i = \dfrac{\partial E_{mu}}{\partial E}\boldsymbol{f}_i - \dfrac{\dot{s}}{s} m\dot{\boldsymbol{q}}_i \\ Q\ddot{s} = s\left[\sum_i m\dot{\boldsymbol{q}}_i^2 - 3Nk_BT_0\right] + Q\dfrac{\dot{s}^2}{s} \end{cases}$$

U. Hansmann, Y.O. & F. Eisenmenger, *Chem. Phys. Lett.* **259**, 321 (1996);
N. Nakajima, H. Nakamura & A. Kidera, *J. Phys. Chem. B* **101**, 817 (1997).

MULTICANONICAL ALGORITHM

B. Berg & T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).

B. Berg & T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).

Step 1: Iterations of Short Preliminary Runs to Determine the Multicanonical Weight Factor $W_{mu}(E)$

Step 2: One Long Production Run

Step 3: Analyze the Data to Obtain:

* Global-Minimum Energy Configuration

* Thermodynamic Quantities for Desired Temperatures (by Ferrenberg-Swendsen Single-Histogram Reweighting Techniques)

$$P_C(E;T) \propto P_{mu}(E) \frac{W_B(E;T)}{W_{mu}(E)}$$

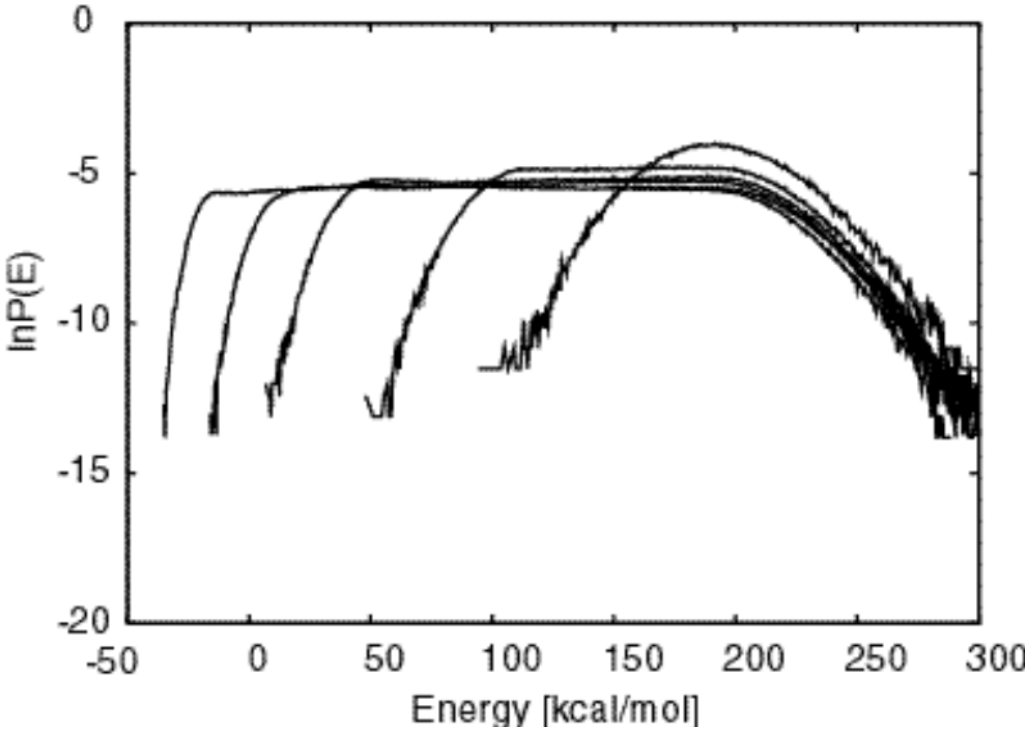
MULTICANONICAL ALGORITHM

Step 1: Determination of Multicanonical Weight Factor *Wmu (E)*

$$W(E) = \ln n(E) = \beta_0 E + \ln P_c(E, \beta_0)$$

$$W^{i+1}(E) = W^i(E) + \ln P^i_{mc}(E)$$

This step becomes non-trivial as the system becomes complex.



Single-Histogram Reweighting Techniques

A. Ferrenberg & R. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).

$$< A >_T = \frac{\sum_E A(E) P_C\left(E;T\right)}{\sum_E P_C\left(E;T\right)} = \frac{\sum_E A(E) n\left(E\right) e^{-\beta E}}{\sum_E n\left(E\right) e^{-\beta E}}$$

Here, the density of states *n(E)* is obtained from the histogram of the energy distribution $N_{mu}(E)$ that was obtained from the production run of the multicanonical simulation:

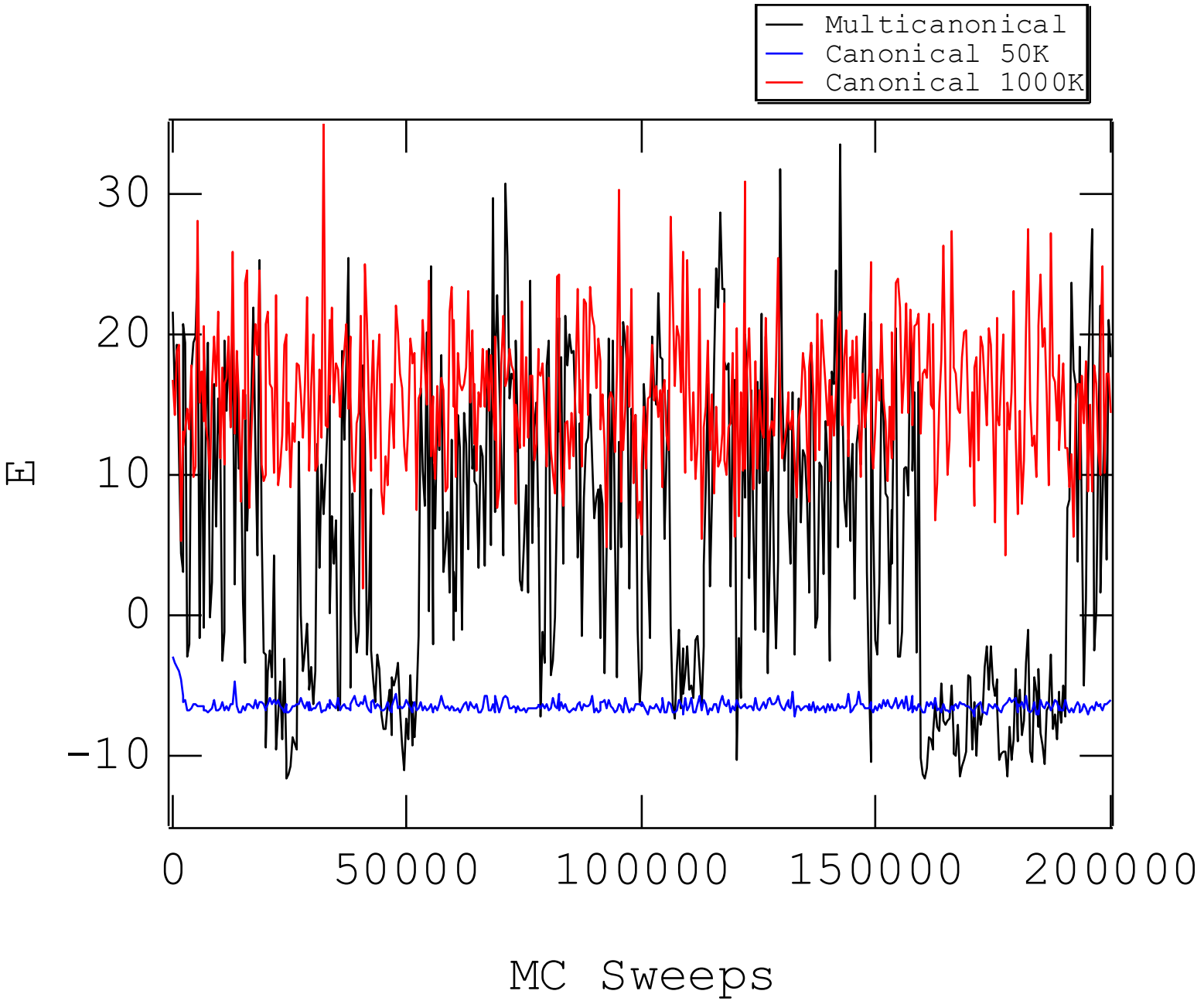
$$n\left(E\right) = \frac{N_{mu}(E)}{W_{mu}(E)}, \quad \text{where} \quad N_{mu}(E) = n\left(E\right) W_{mu}(E) \ .$$

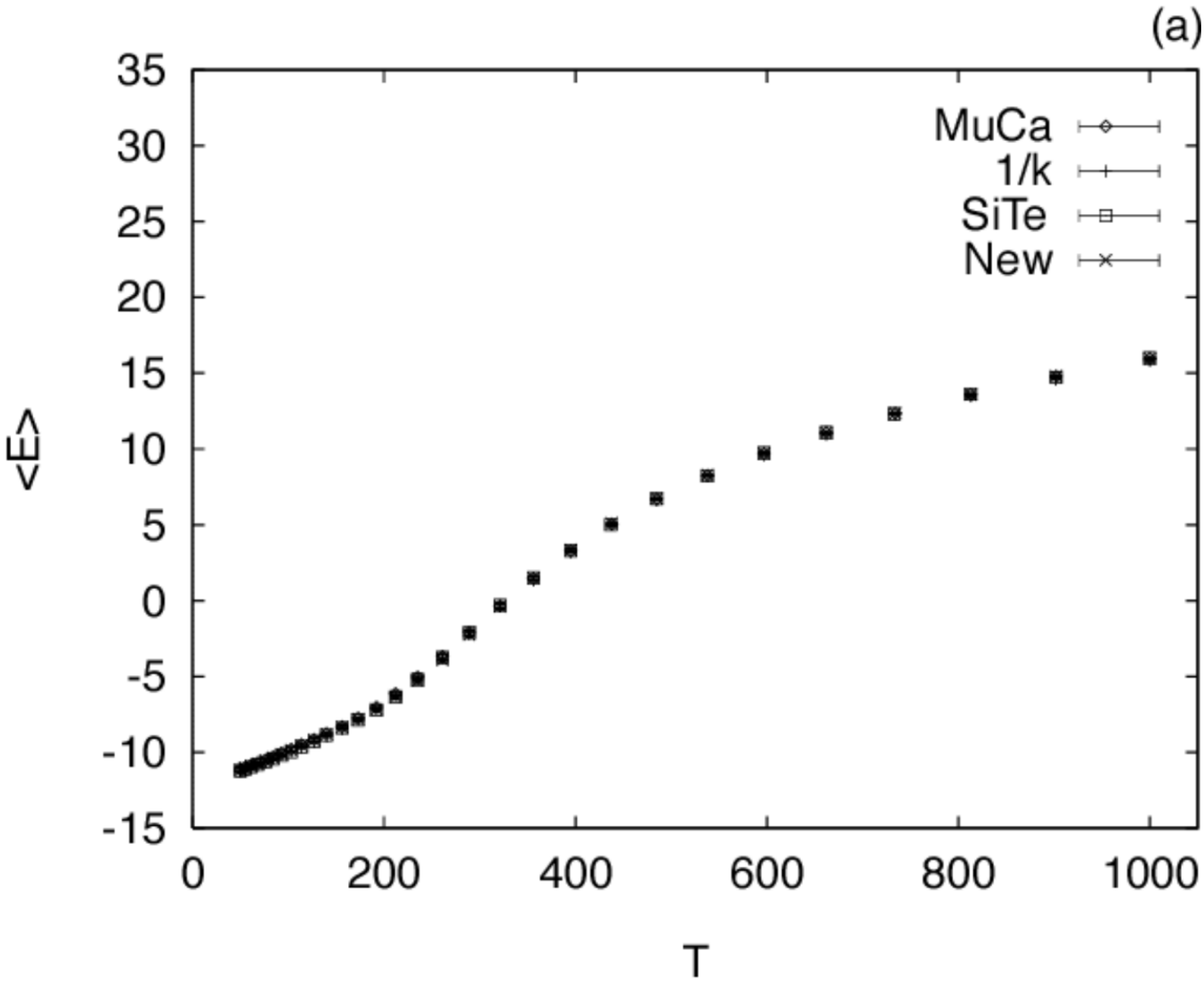
Single-Histogram Reweighting Techniques

A. Mitsutake, Y. Sugita & Y.O., *J. Chem. Phys.* **118**, 6664 (2003).

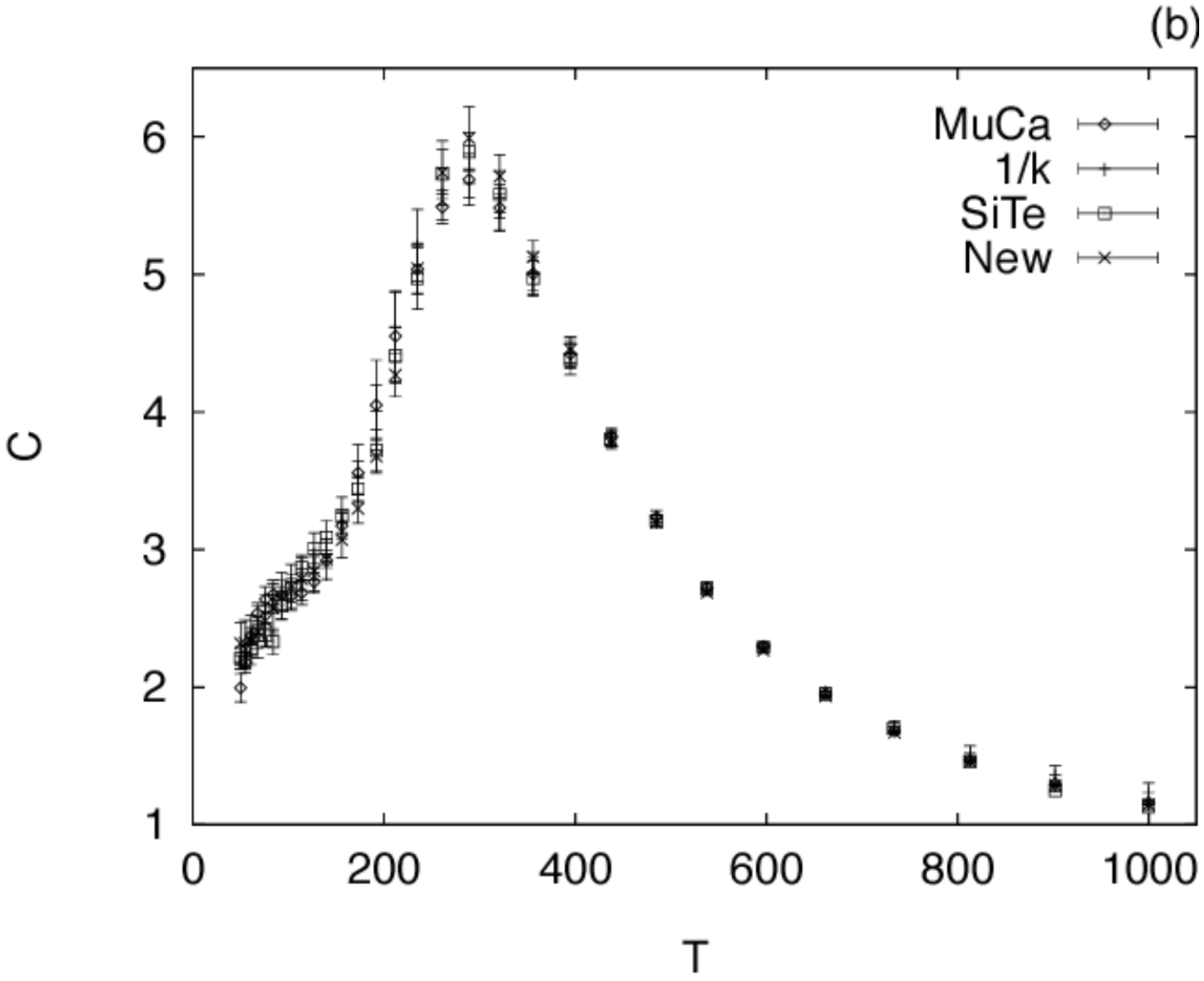
When the physical quantity A cannot be written as a function of E, we use the following equation:

$$\langle A \rangle_T = \frac{\sum_{x_k} A(x_k) W_{mu}^{-1}(E(x_k)) \exp(-\beta E(x_k))}{\sum_{x_k} W_{mu}^{-1}(E(x_k)) \exp(-\beta E(x_k))}$$


Multicanonical
Canonical 50K
Canonical 1000K
30
20
10
0
-10
E
0
50000
100000
150000
200000
MC Sweeps

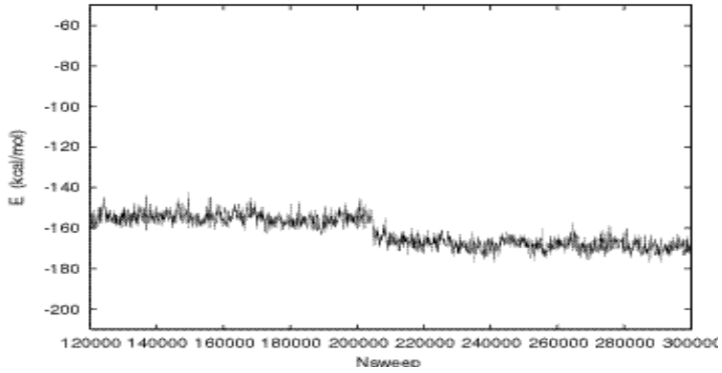

(a)
MuCa
1/k
SiTe
New
<E>
T
35
30
25
20
15
10
5
0
-5
-10
-15
0
200
400
600
800
1000

Enk-s

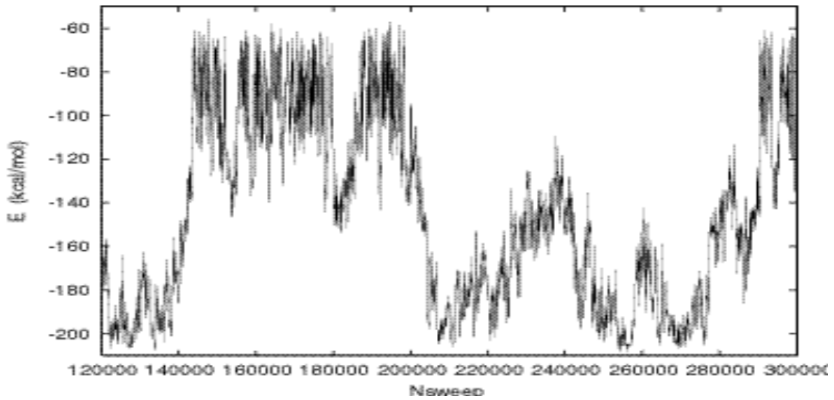


17-Residue Helical Peptide (120000-300000 MC Sweeps)

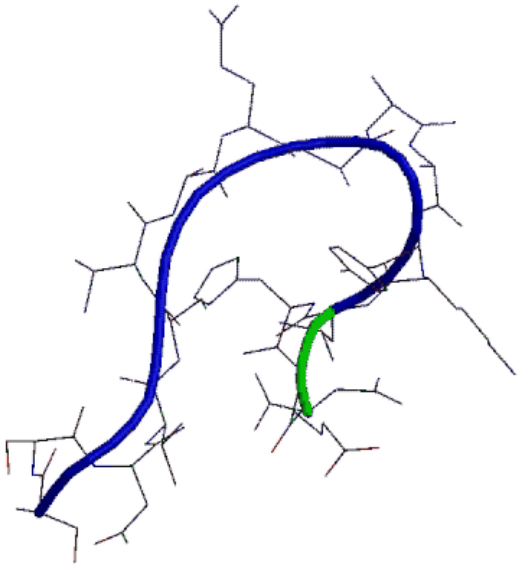
Simulation and movie by A. Mitsutake

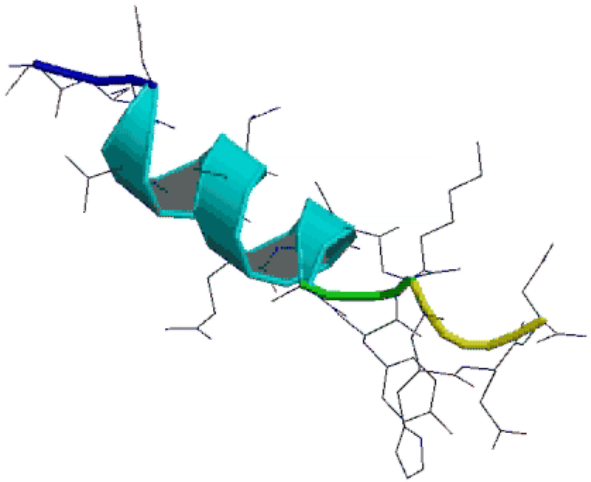


Canonical MC: T = 200 K



Multicanonical MC





Simulated Tempering (ST)

A.P. Lyubartsev, *et al.*, *J. Chem. Phys.* **96**, 1776 (1992).
E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).

Temperature becomes a dynamical variable: Sample temperature uniformly

$$W_{ST}(E;T) = \exp(-\beta E + a(T))$$

$$P_{ST}(T) = \int dE n(E) \exp(-\beta E + a(T)) = const$$

Random Walk in Temperature Space
→ Random Walk in Energy Space

Discretize Temperature:

$T_m \quad (m = 1, \ \ldots \ , M)$

$$W_{ST}(E;T_m) = \exp(-\beta_m E + a_m)$$

$$\exp(-a_m) \propto \int dE n(E) \exp(-\beta_m E)$$

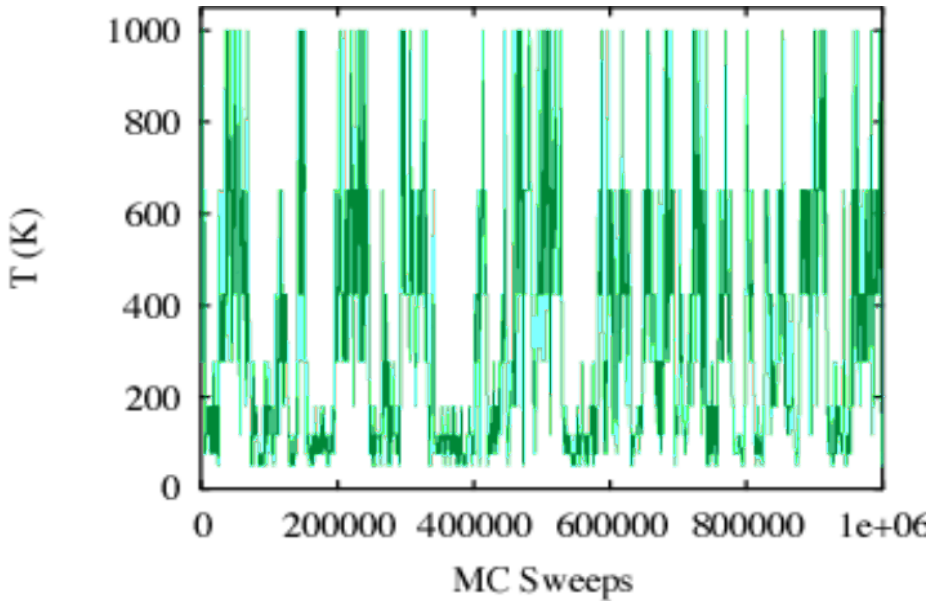
a_m: Dimensionless Helmholtz free energy at temperature T_m

a_m is determinded by iteration of short ST runs

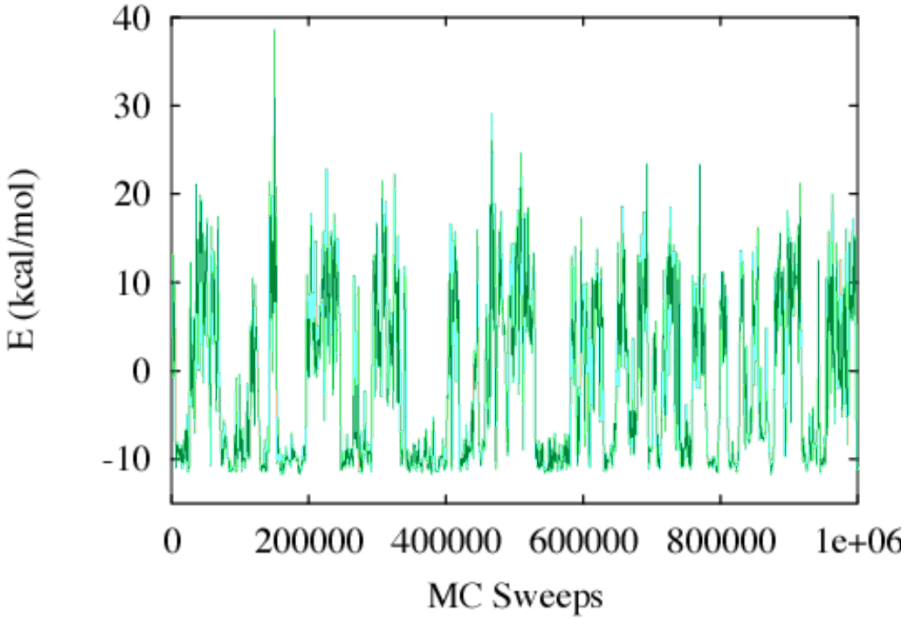
This weight determining process can be very tedious and time-consuming.

Simulated Tempering (ST)

Time Series of Temperature



Time Series of Potential Energy



(Cf. Global Minimum: -12.2 kcal/mol)

Random Walk in Temperature

Random Walk in Energy

A. Mitsutake & Y.O., *Chem. Phys. Lett.* **332**, 131 (2000).

Expectation Values of a Physical Quantity *A* at Temperatures

$$T_m \quad (m = 1, \ \ldots \ , M)$$

$$\langle A \rangle_{T_m} = \frac{1}{n_m} \sum_{k=1}^{n_m} A(x_m(k))$$

where n_m are the total number of samples obtained at T_m.

Multiple-Histogram Reweighting Techniques (Weighted Histogram Analysis Method: WHAM)

A. Ferrenberg & R. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989); S. Kumar, D. Bouzida, R. Swendsen, P. Kollman & J. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).

$$< A >_T = \frac{\sum_E A(E) n(E) e^{-\beta E}}{\sum_E n(E) e^{-\beta E}}$$

Given *M* set of histograms $N_m(E)$, which were obtained at T_m, the following WHAM equations are solved iteratively for density of states $n(E)$ and dimensionless Helmholtz free energy f_m : (n_m are the total number of samples obtained at T_m)

$$n(E) = \frac{\sum_{m=1}^{M} N_m(E)}{\sum_{m=1}^{M} n_m e^{f_m - \beta_m E}}, \text{ where } e^{-f_m} = \sum_E n(E) e^{-\beta_m E}.$$

Multiple-Histogram Reweighting Techniques (Weighted Histogram Analysis Method: WHAM)

A. Mitsutake, Y. Sugita & Y.O., *J. Chem. Phys.* **118**, 6664 (2003).

When the physical quantity A cannot be written as a function of E, we first obtain the dimensionless Helmholtz free energy f_m ($m = 1, \ldots, M$) by solving the WHAM equations. We then use the following equation:

$$< A >_T = \frac{\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{n_m} A(x_m(k)) \frac{1}{\displaystyle\sum_{\ell=1}^{M} n_\ell \exp\left[f_\ell - \beta_\ell E(x_m(k))\right]} \exp\left[-\beta E(x_m(k))\right]}{\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{n_m} \frac{1}{\displaystyle\sum_{\ell=1}^{M} n_\ell \exp\left[f_\ell - \beta_\ell E(x_m(k))\right]} \exp\left[-\beta E(x_m(k))\right]}$$

Replica-Exchange Method

MC: K. Hukushima & K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
MD: Y. Sugita & Y.O., *Chem. Phys. Lett.* **314**, 141 (1999).

1. System

M *Non-Interacting* Replicas of the Original System at *M* Different Temperatures

2. Replica-Exchange

Step 1: Independent Canonical Simulations Performed for Each Replica

Step 2: A Pair of Replicas (*i* and *j*) Corresponding to Neighboring Temperatures (T_m and T_n) (i.e., $n=m+1$) are Exchanged *a la* Metropolis

$$w\left(x_m^{[i]} \mid x_n^{[j]}\right) = \min\left(1, \frac{W_{\mathrm{REM}}(X')}{W_{\mathrm{REM}}(X)}\right) = \min\left(1, \exp\left(-\Delta\right)\right)$$

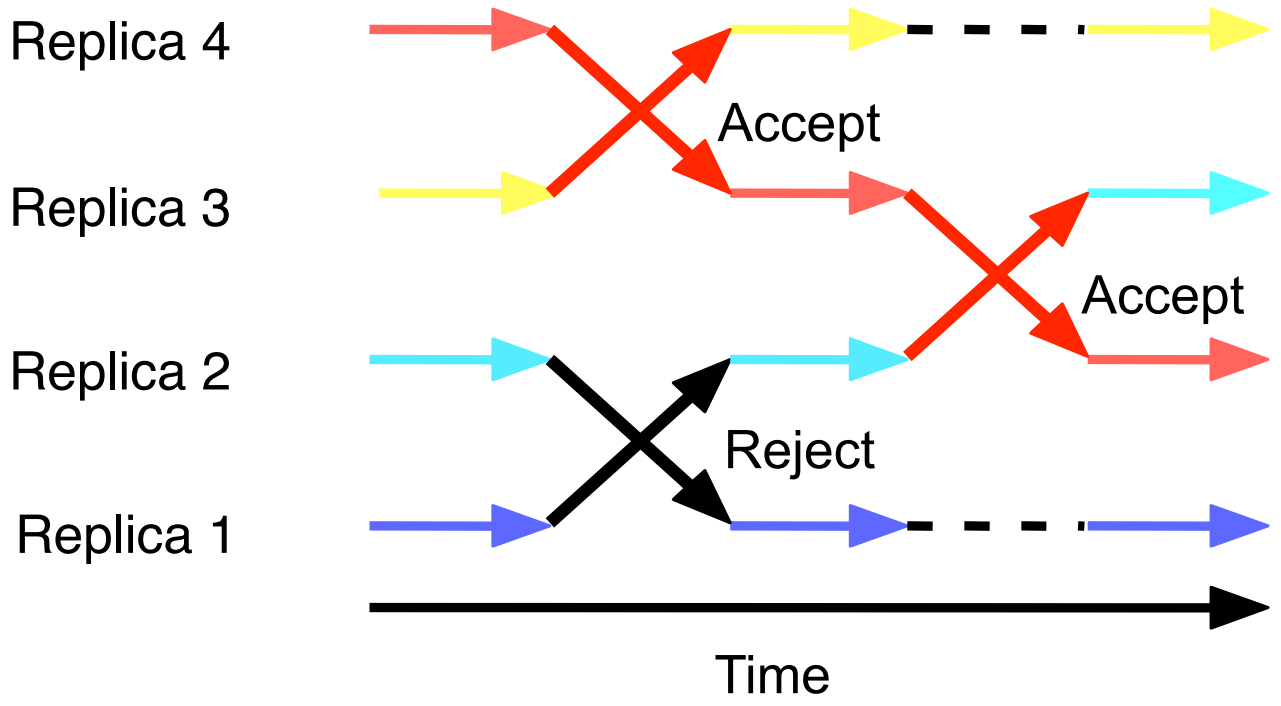
$$\Delta = \left(\beta_m - \beta_n\right)\left(E\left(q^{[j]}\right) - E\left(q^{[i]}\right)\right)$$

Repeat These 2 Steps

3. Canonical Distribution at Any Temperature

by Multiple Histogram Reweighting Techniques (WHAM)

Replica-Exchange Method



Particularly Suitable for Parallel Computers

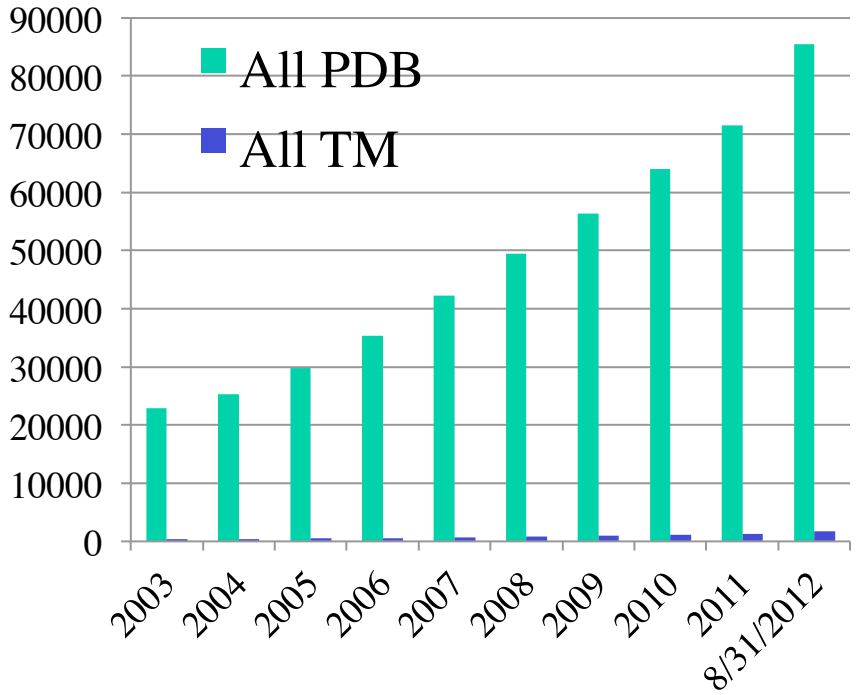
膜タンパク質の立体構造予測
Prediction of Membrane Protein Structures by Replica-Exchange Monte Carlo Simulations

H. Kokubo & Y.O., *Chem. Phys. Lett.* **383**, 397-402 (2004).
H. Kokubo & Y.O., *J. Chem. Phys.* **120**, 10837-10847 (2004).
H. Kokubo & Y.O., *J. Phys. Soc. Jpn.* **73**, 2571-2585 (2004).
H. Kokubo & Y.O., *Chem. Phys. Lett.* **392**, 168-175 (2004).
H. Kokubo & Y.O., *Biophys. J.* **96**, 765-776 (2009).
R. Urano, H. Kokubo & Y.O., submitted.
R. Urano & Y.O., submitted.

Motivation:

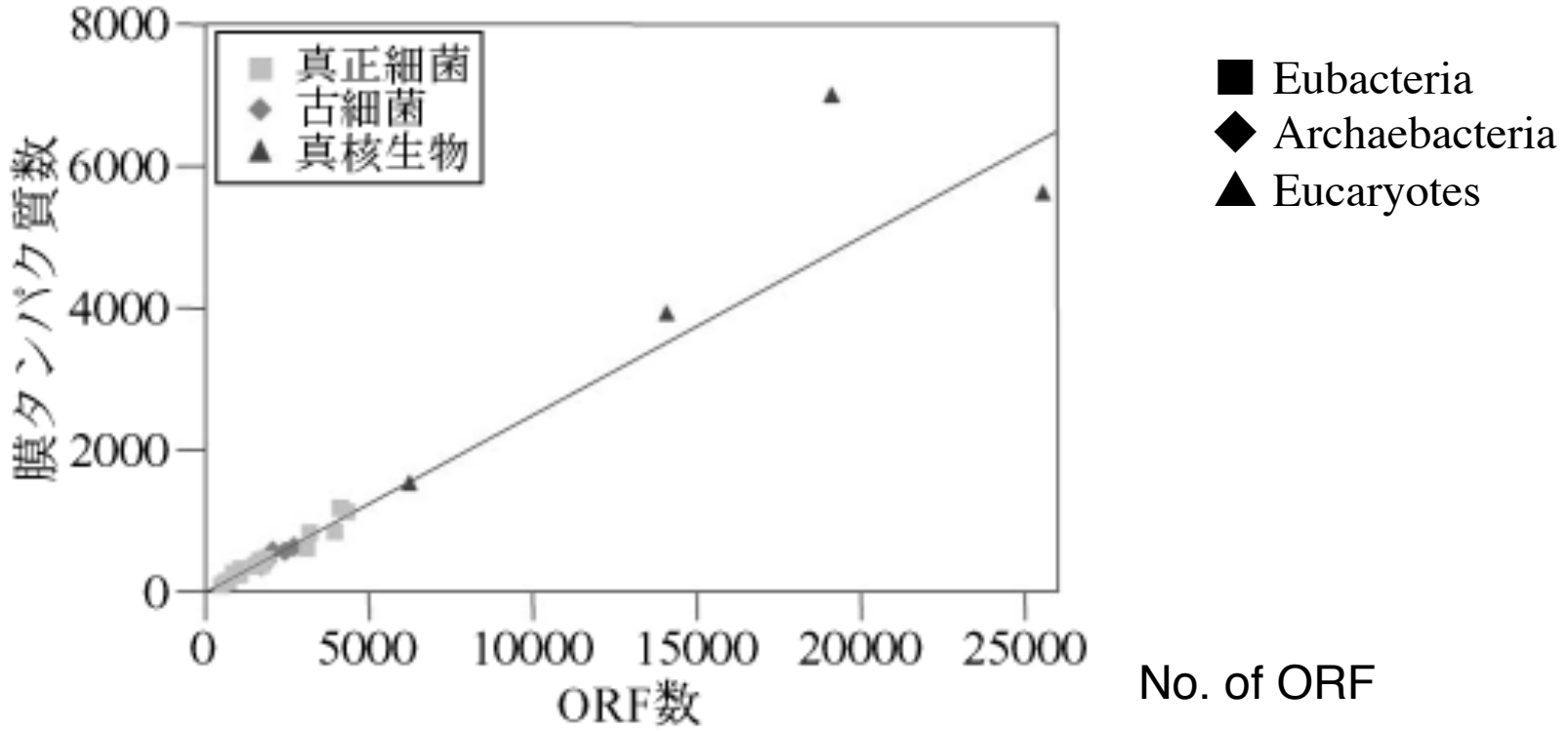
Protein Data Bank (PDB) :

105,465 entries as of January 3, 2015, but only about 2 % of them are membrane protein structures .



It is estimated that 20-30 % of all genes in most genomes encode membrane proteins. However, only a small number of detailed structures have been obtained for membrane proteins because of technical difficulties such as high quality crystal growth.

No. of Membrane Proteins



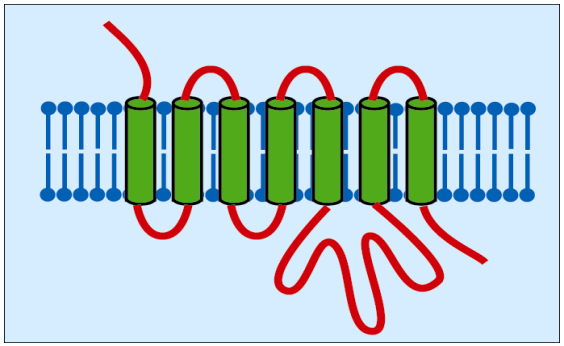
S. Mitaku, *Biophysics* **42**, (2002) 104-109 (in Japanese).
A. Krogh et al., *J. Mol. Biol.* **305** (2001) 567-580.

We can know **Membrane Spanning Regions** by **Prediction Tool on the Web:** SOSUI,HMMTOP,TMHMM..etc.

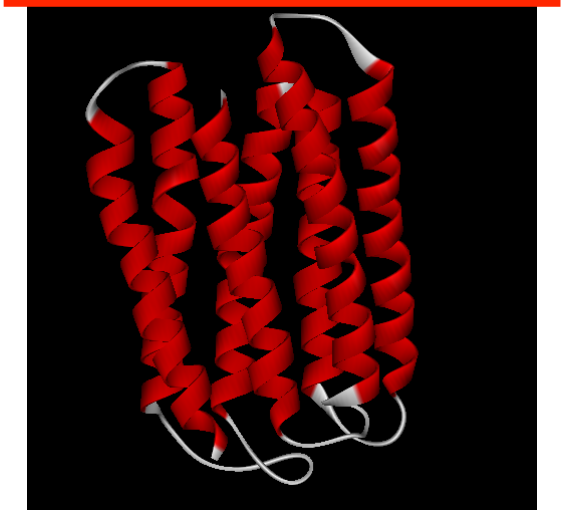
Predict transmembrane helix configurations by molecular simulations

Sequence

MNGTEGPNFYVP、、、、



No.	N terminal	transmembrane region	C terminal	type	length
1	10	IWLALGTALMGLGTLYFLVKGMG	32	SECONDARY	23
2	45	TTLVPAIAFTMYLSMLLGYGLTM	67	SECONDARY	23
3	77	IYWARYADWLFTTPLLLLDLALL	99	SECONDARY	23
4	106	TILALVGADGIMIGTGLVGALTK	128	PRIMARY	23
5	136	WWAISTAAMLYILYVLFFGFTSK	158	PRIMARY	23
6	176	VTVVLWSAYPVVWLIGSEGAGIV	198	PRIMARY	23
7	203	ETLLFMVLDVSAKVGFGLILLR	224	PRIMARY	22

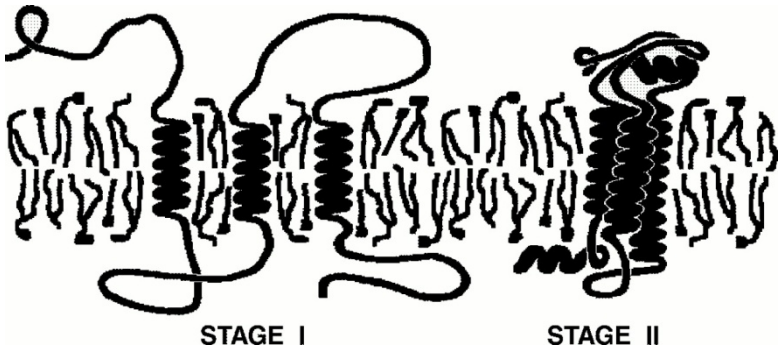


Tertiary Structure

Two-stage model

Individual helices of a protein are postulated to be stable separately as domains in a lipid bilayer. And then side-to-side helix association is driven, resulting in a functional protein.

J.-L. Popot and D.M. Engelman, *Annu. Rev. Biochem.* **69**, 881 (2000).

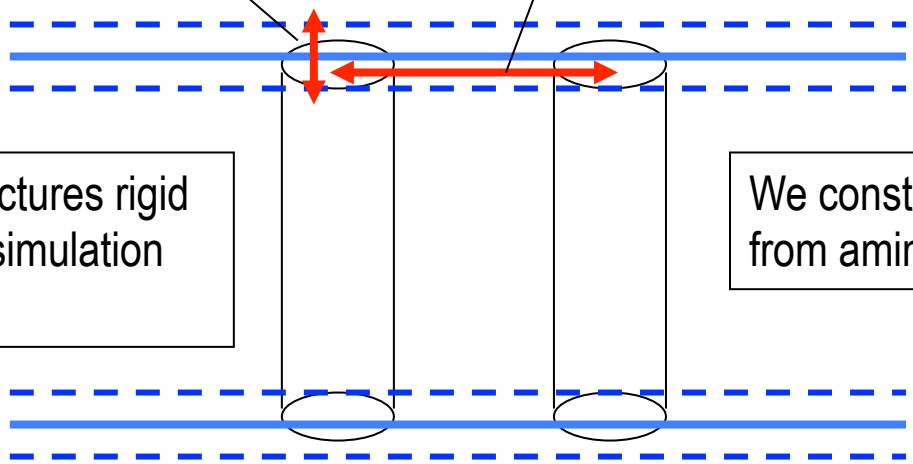


- Individual helix stability as domains is a consequence of the hydrophobic effect and main-chain hydrogen bonding.
- Specific folding energy is provided mainly by packing of the preformed helices with each other, by loop structures, and by interactions with prosthetic groups.
- Additionally, ion pairs and hydrogen bonds between helices are sometimes found, and general contributions are made by interactions with the lipid environment.

Simulation Conditions

Harmonic constraint are set only when helices are apart at some distance from membrane boundary planes.

Harmonic constraint are set only when helices are apart at some distance.



We keep backbone structures rigid in order to perform the simulation with helix structures.

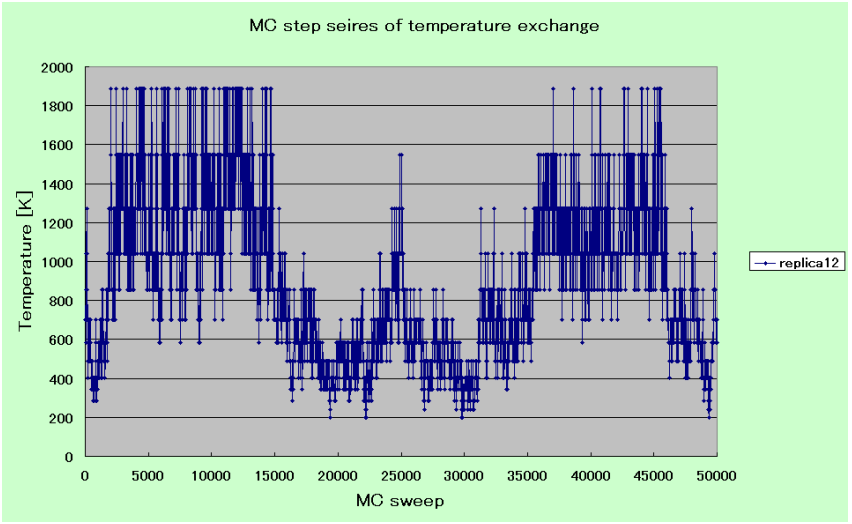
We construct ideal α-helices from amino-acid sequence.

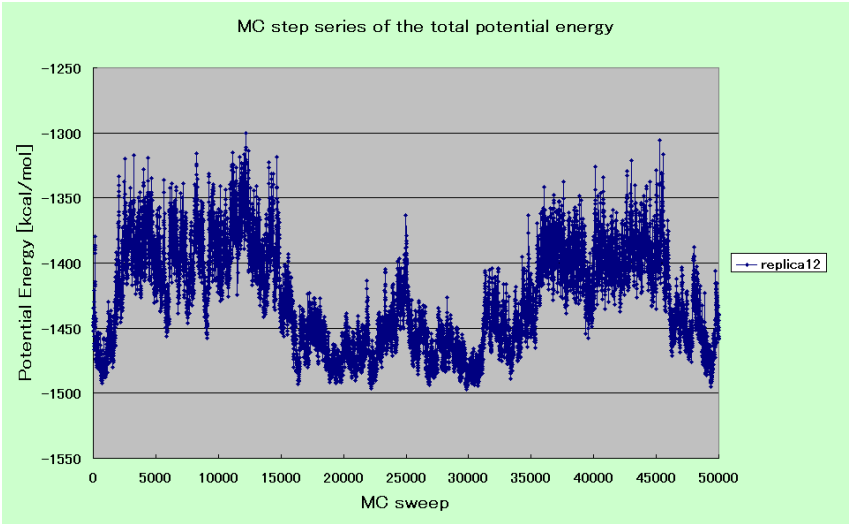
We neglect loop region amino acids, lipids, and water.

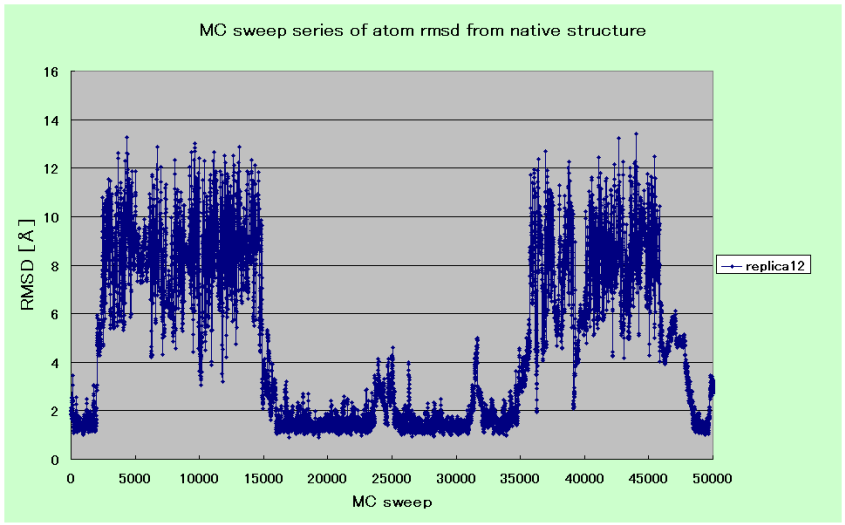
CHARMM force-field (ver. 19)
Replica-exchange Monte Carlo simulation (rigid translation, rigid rotation of each helix and torsion angle rotation of side chains)

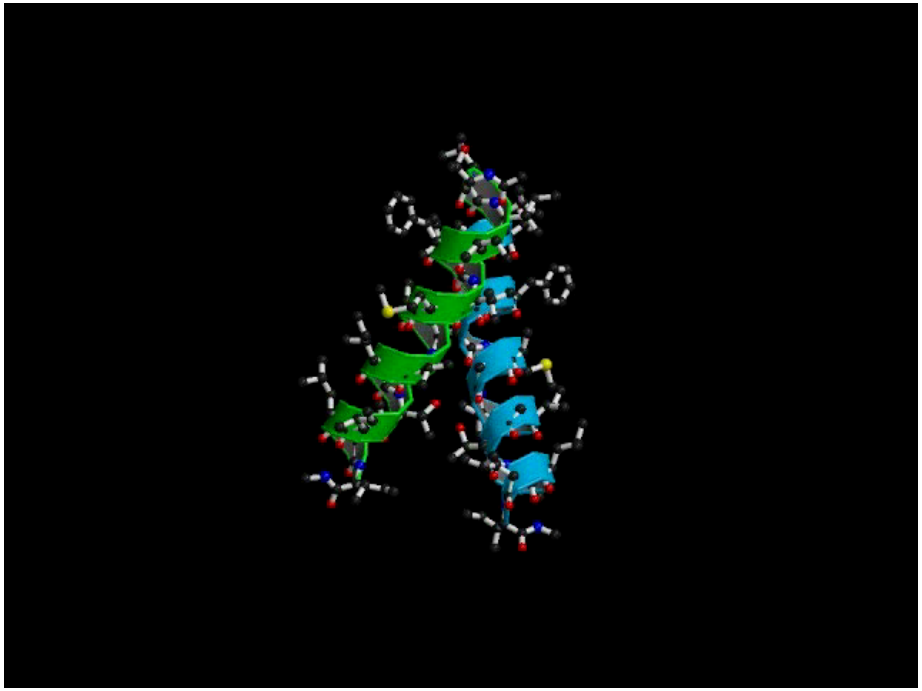
DIMERIC TRANSMEMBRANE DOMAIN OF HUMAN GLYCOPHORIN A

Simulation and movie by H. Kokubo

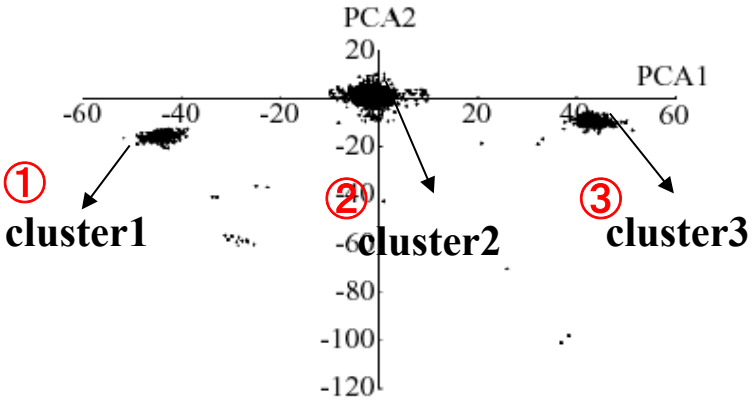




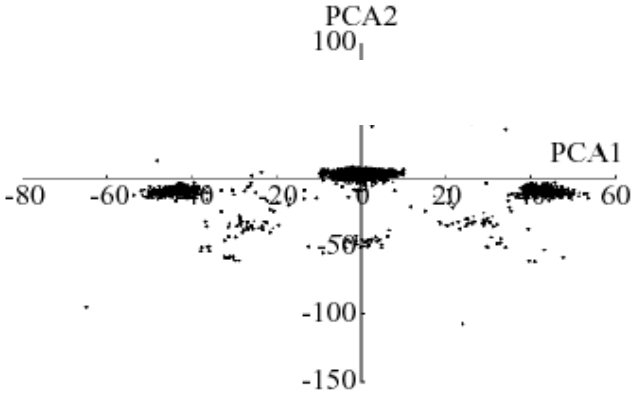




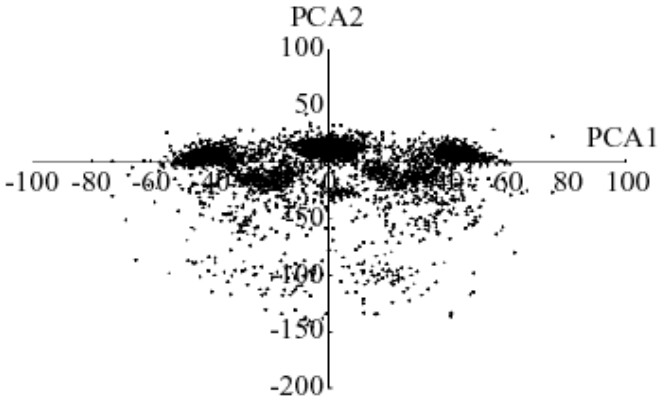
Principal Component Analysis: case for ε=1.0



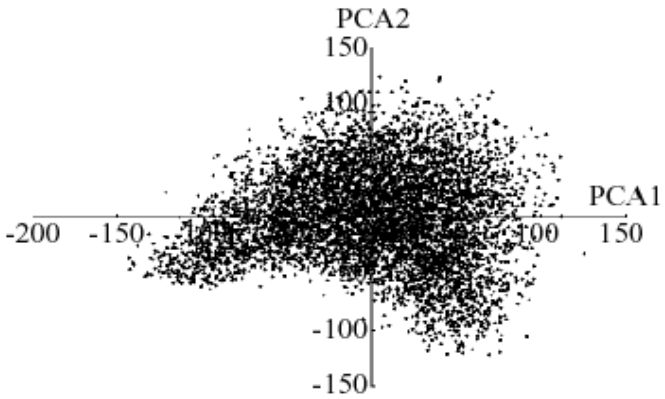
200 K



342 K

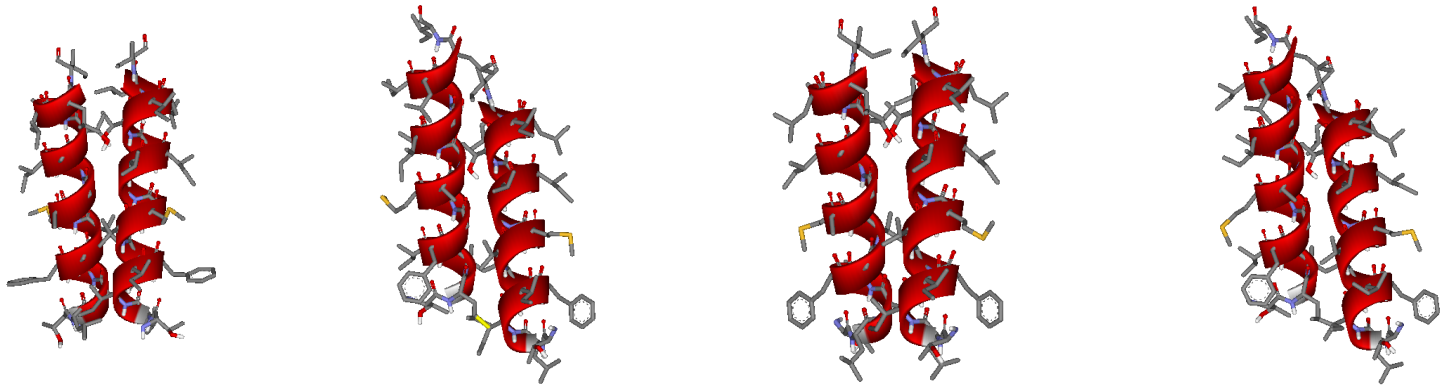


585 K

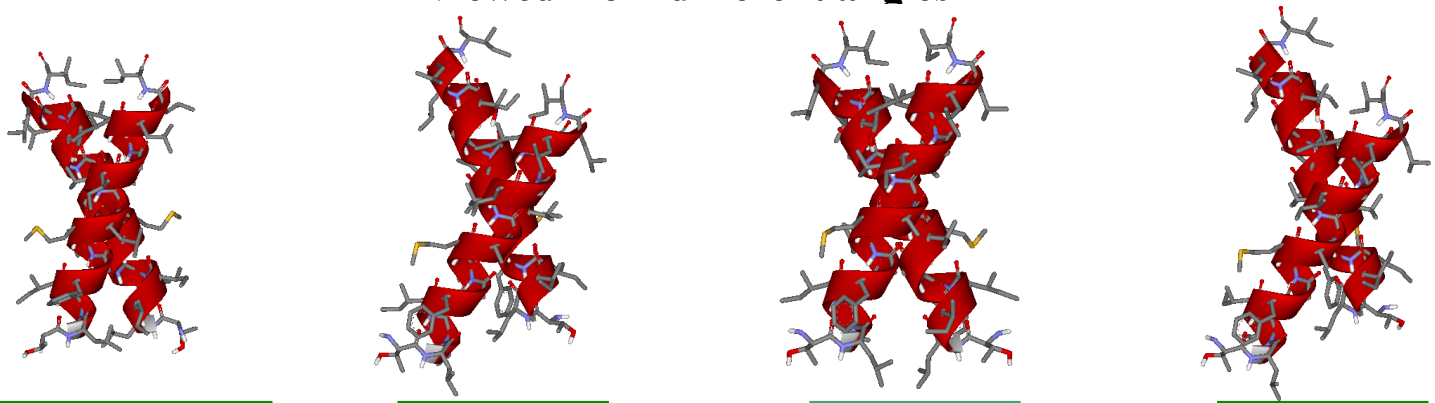


1888 K

Structure of Each Cluster: case for ε=1.0



Viewed from different angles

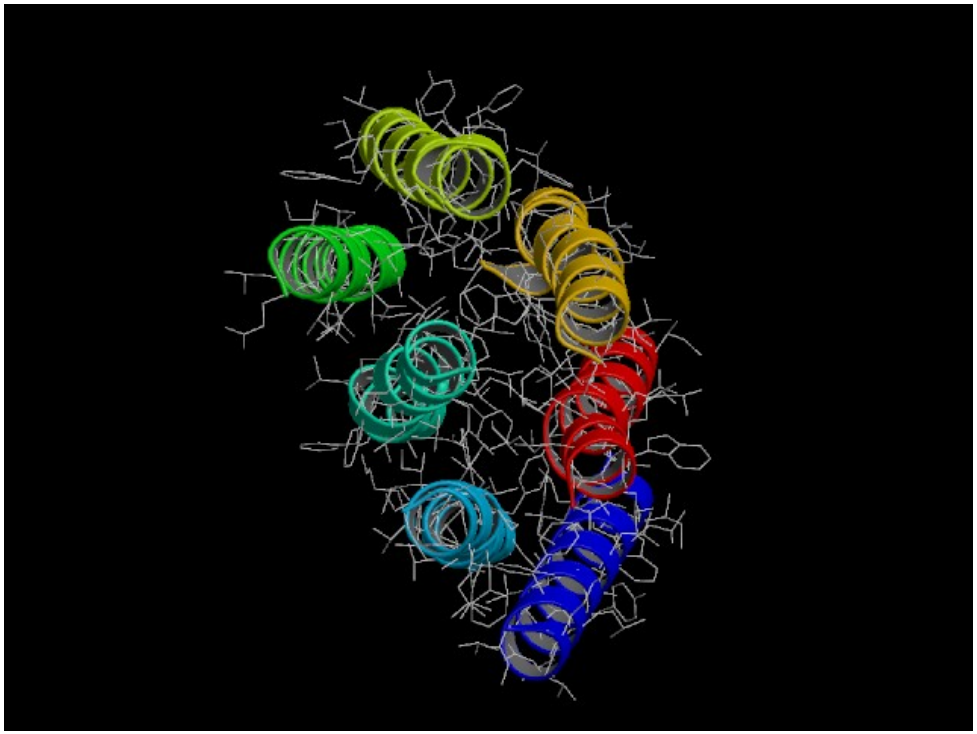


Native(NMR) | **Cluster1** | **Cluster2** | **Cluster3**

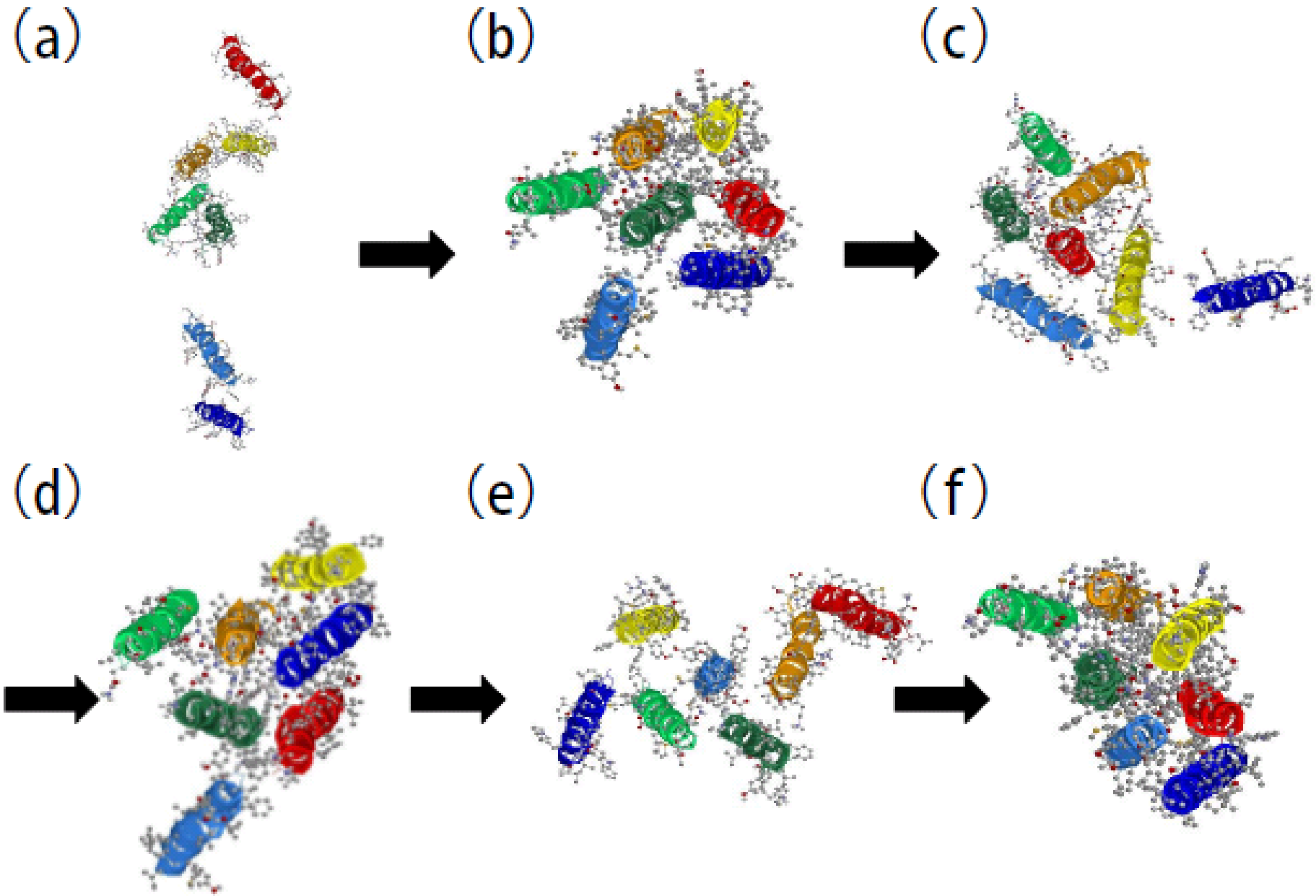
H. Kokubo & Y.O., *Chem. Phys. Lett.* **392**, 168-175 (2004).
H. Kokubo & Y.O., *Biophys. J.* **96**, 765-776 (2009).

Bacteriorhodopsin（Case of 7 Helices）

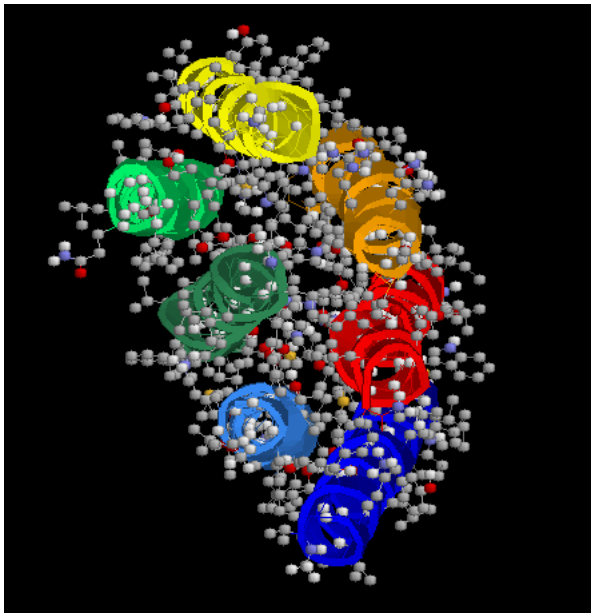
Native Structure

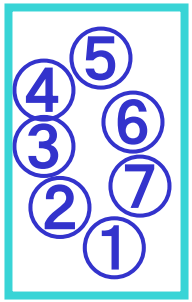


Snapshots (Replica 16)

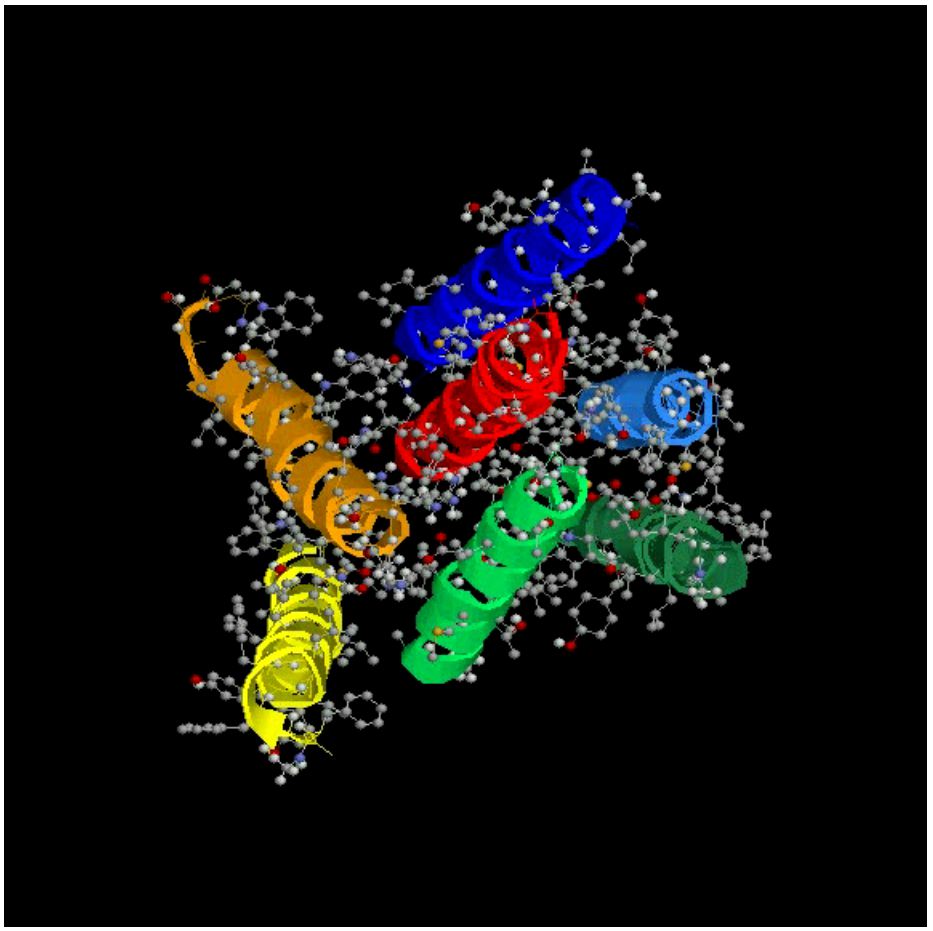


REMC
(from above)



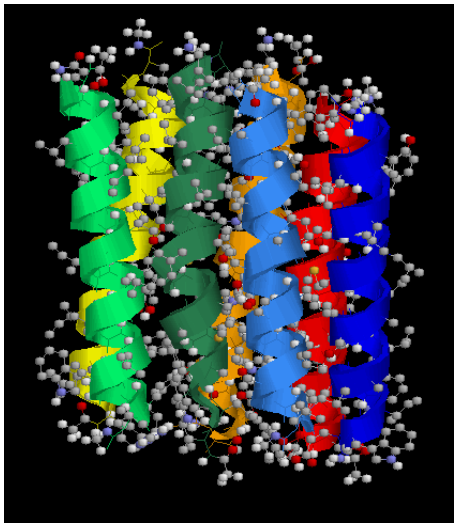


Replica 14



Native structure

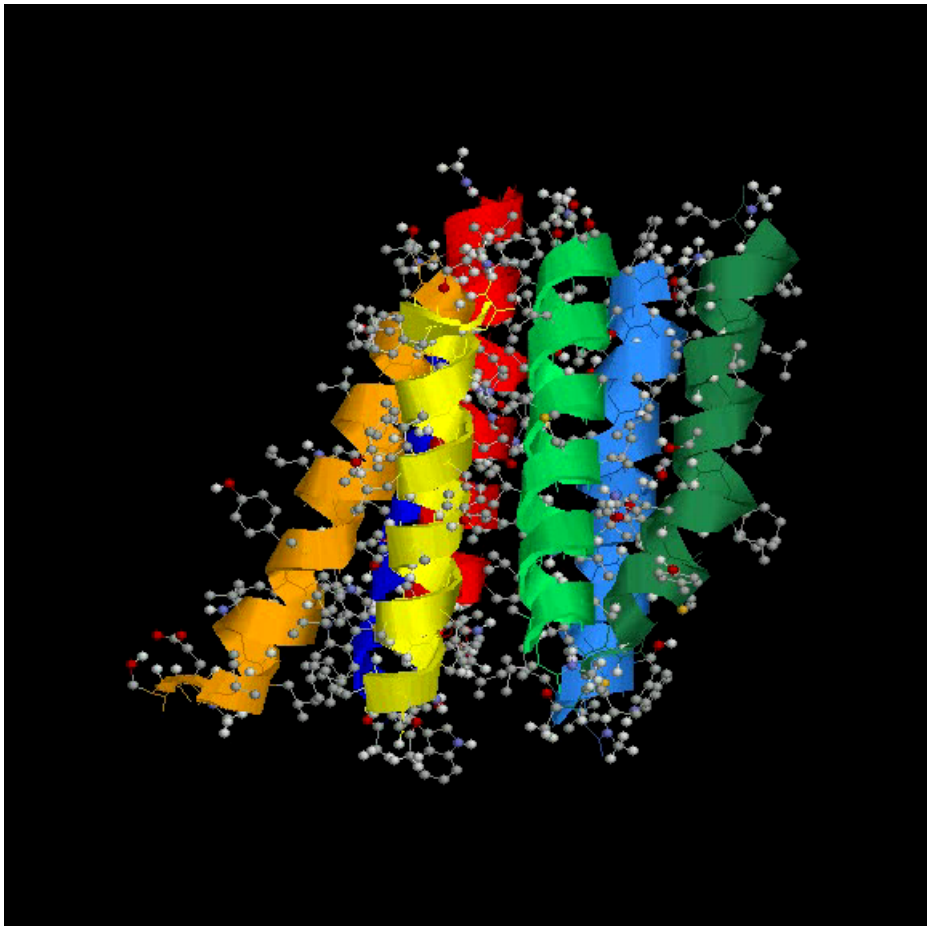
REMC
(from above)



5 6 7
4 3 2 1

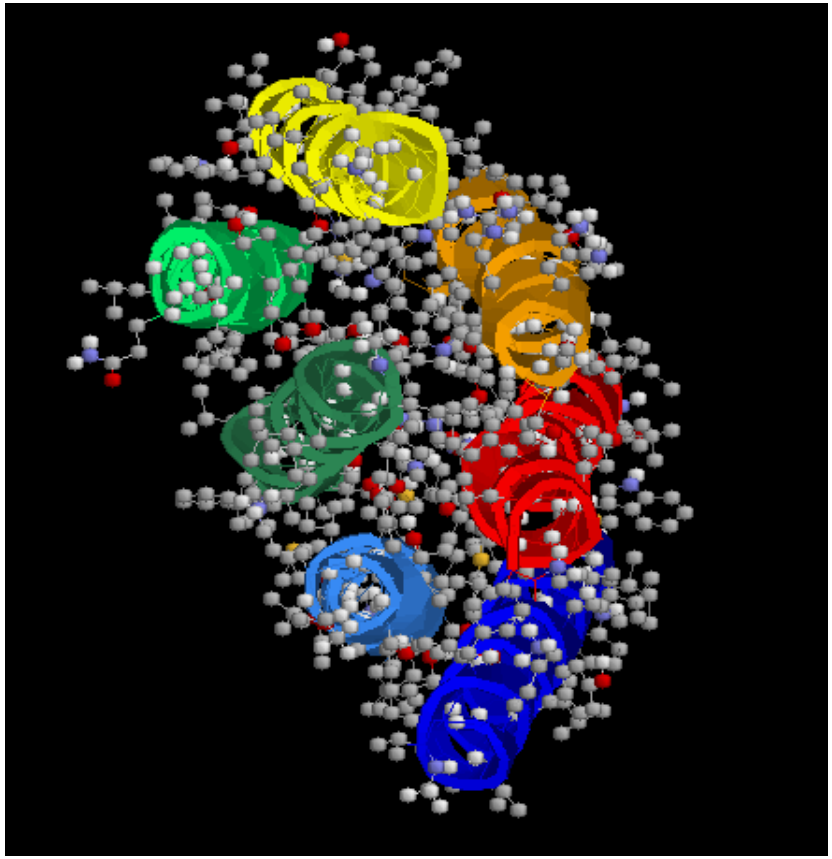
Native structure

Replica 14

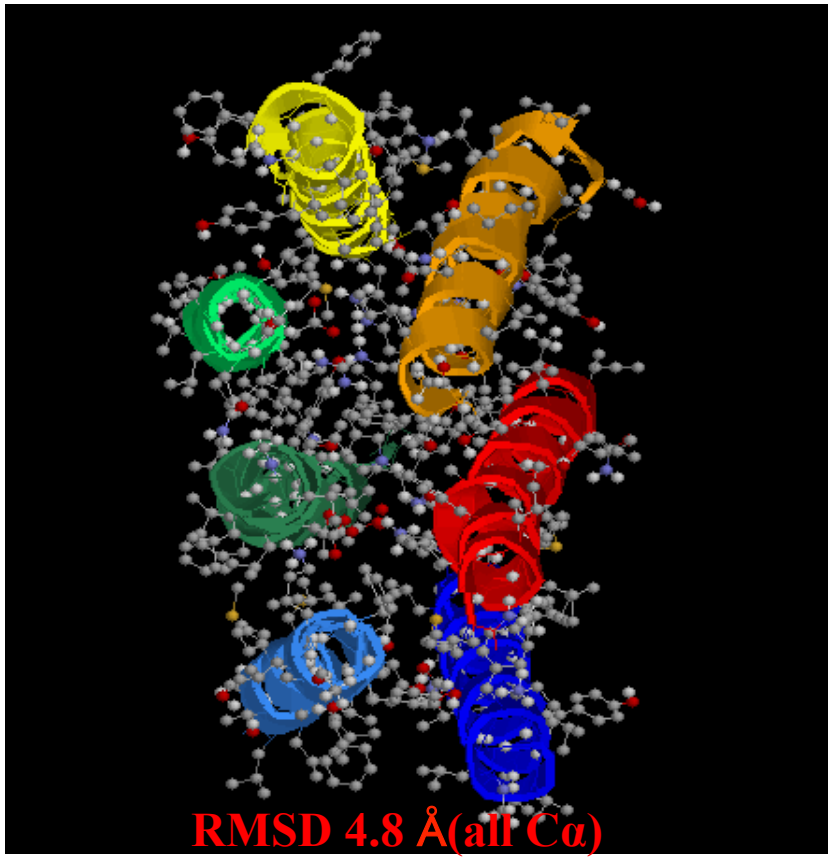


H. Kokubo & Y.O., *Biophys. J.* **96**, 765-776 (2009).

Local Minimum Structure



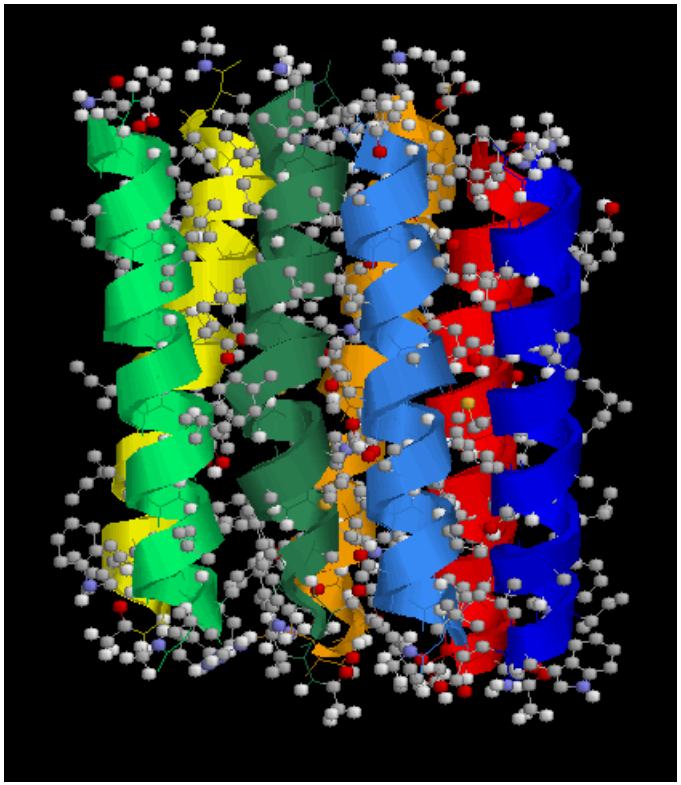
Native Structure



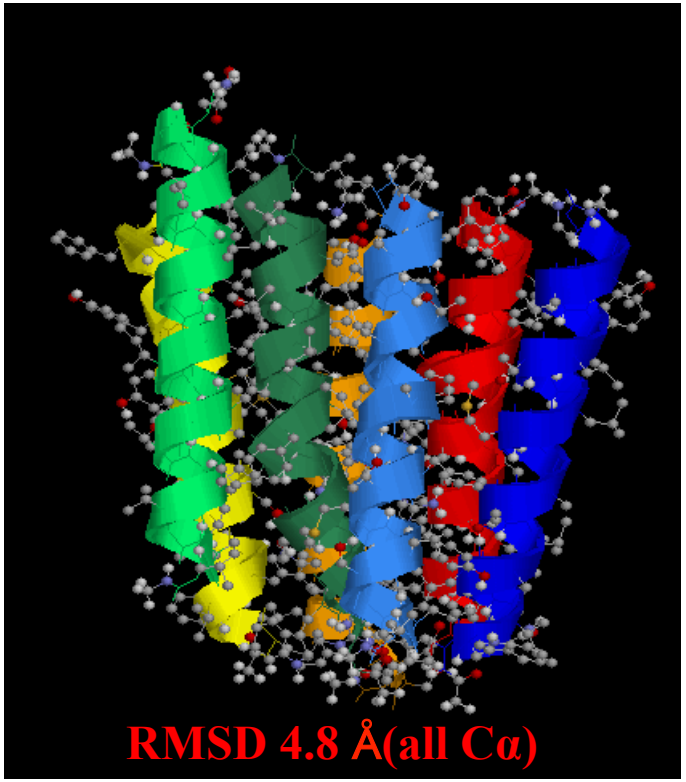
Replica 14

H. Kokubo & Y.O., *Biophys. J.* **96**, 765-776 (2009).

Local Minimum Structure



Native Structure



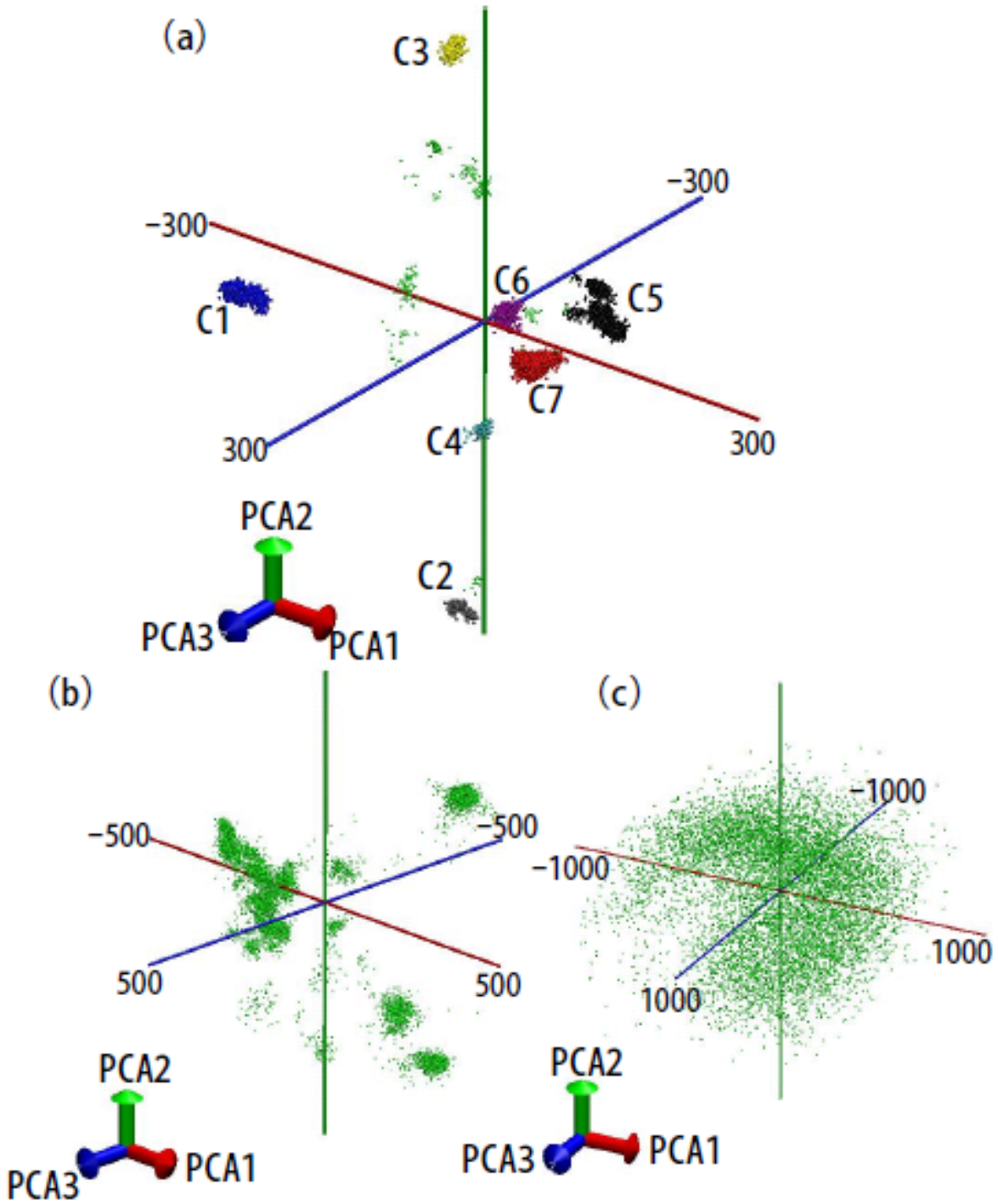
Replica 14

Principal Component Analysis at

(a) $T = 500$ K

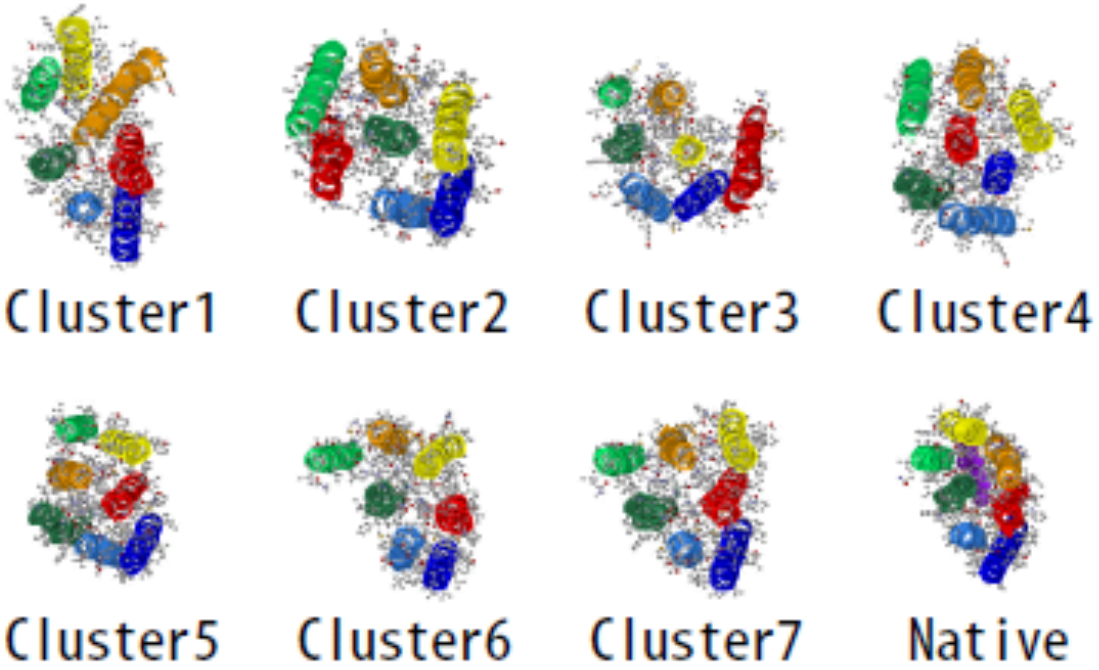
(b) $T = 976$ K

(c) $T = 5000$ K

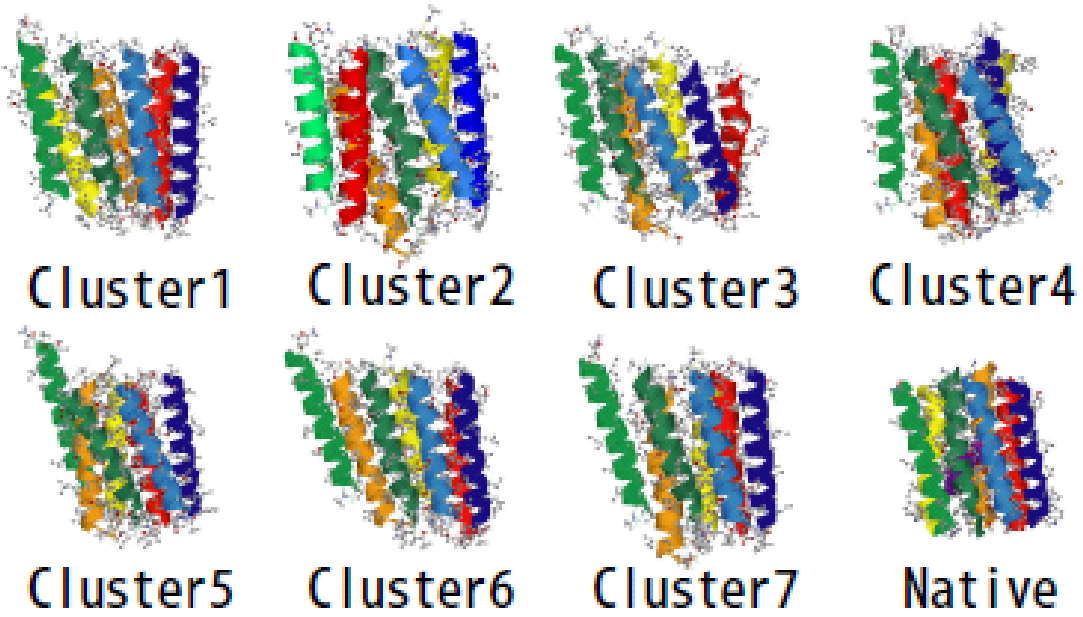


Typical Structures of the 7 Clusters

Top view



Side view



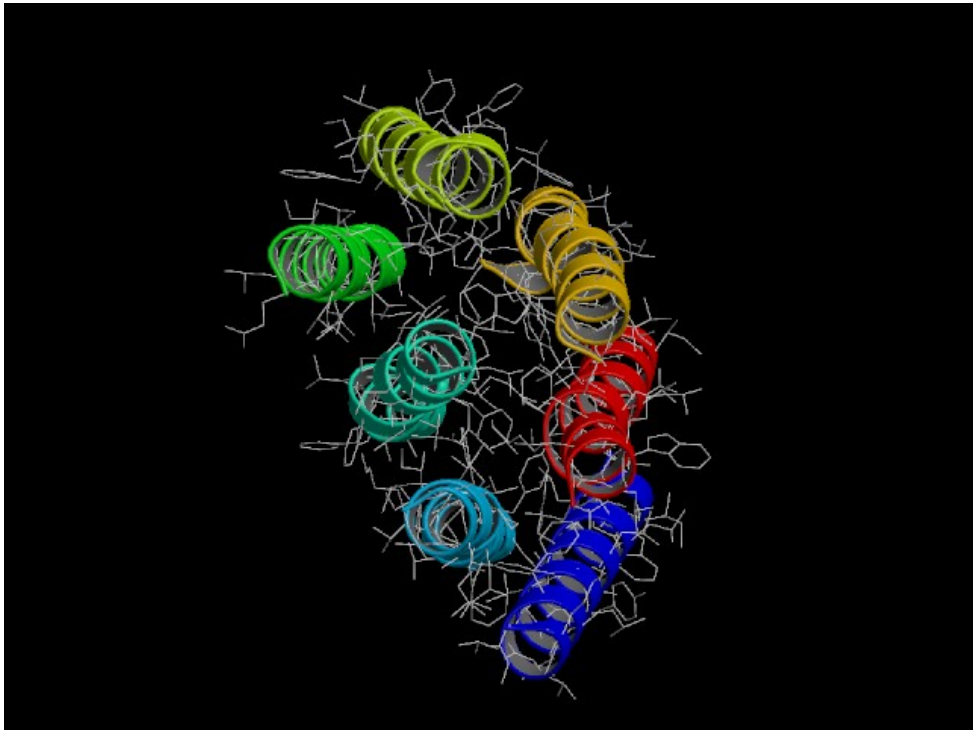
H. Kokubo & Y.O.,
Biophys. J. **96**, 765-776 (2009).

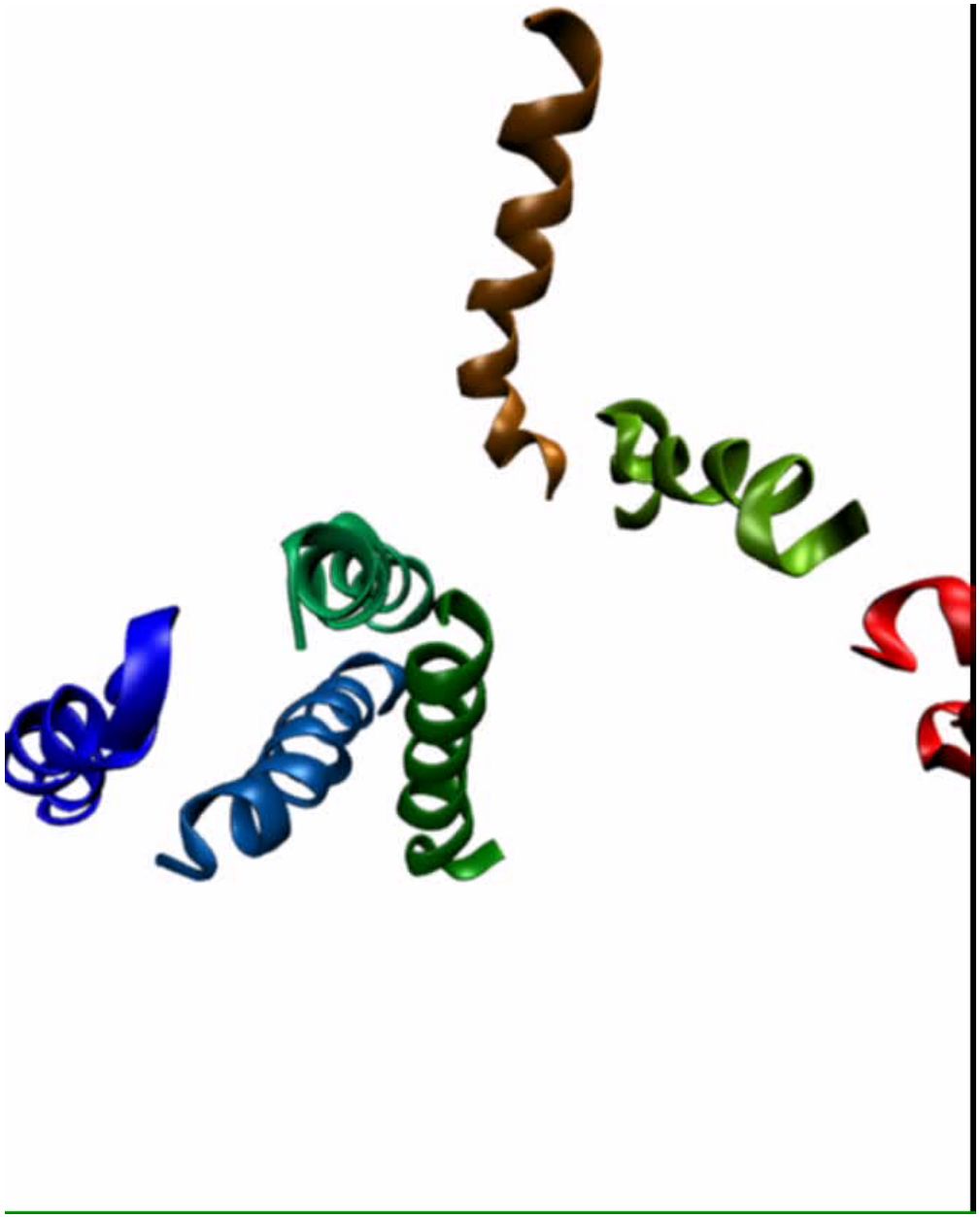
Bacteriorhodopsin（Case of 7 Helices）

R. Urano & Y.O., submitted.

Introduced bending of TM helices

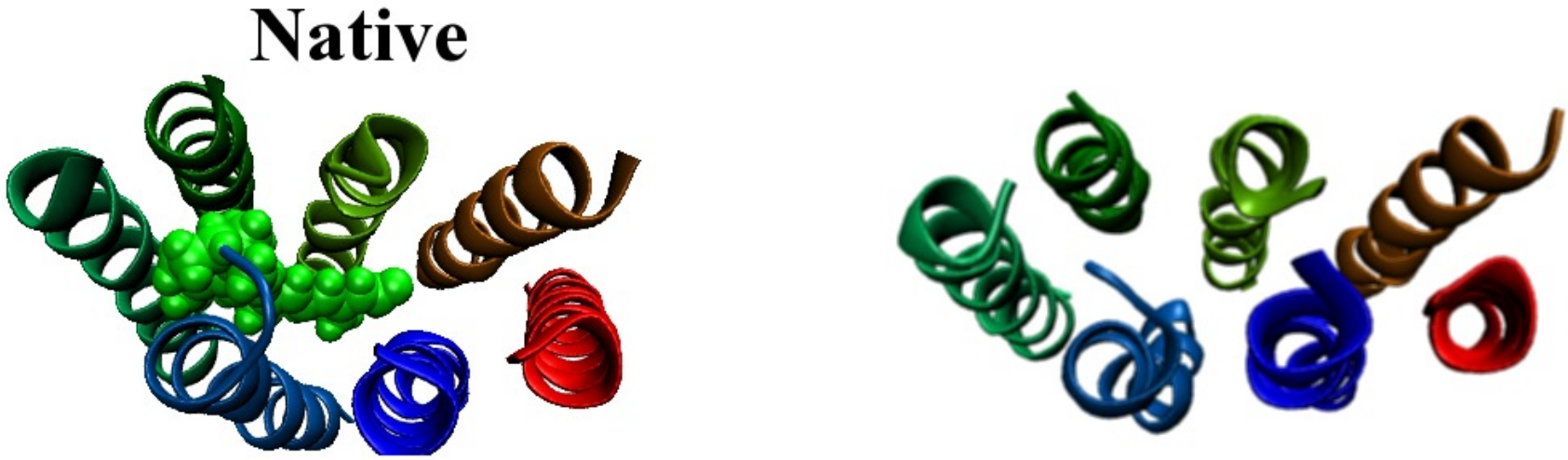
Native Structure





Simulation and movie by R. Urano

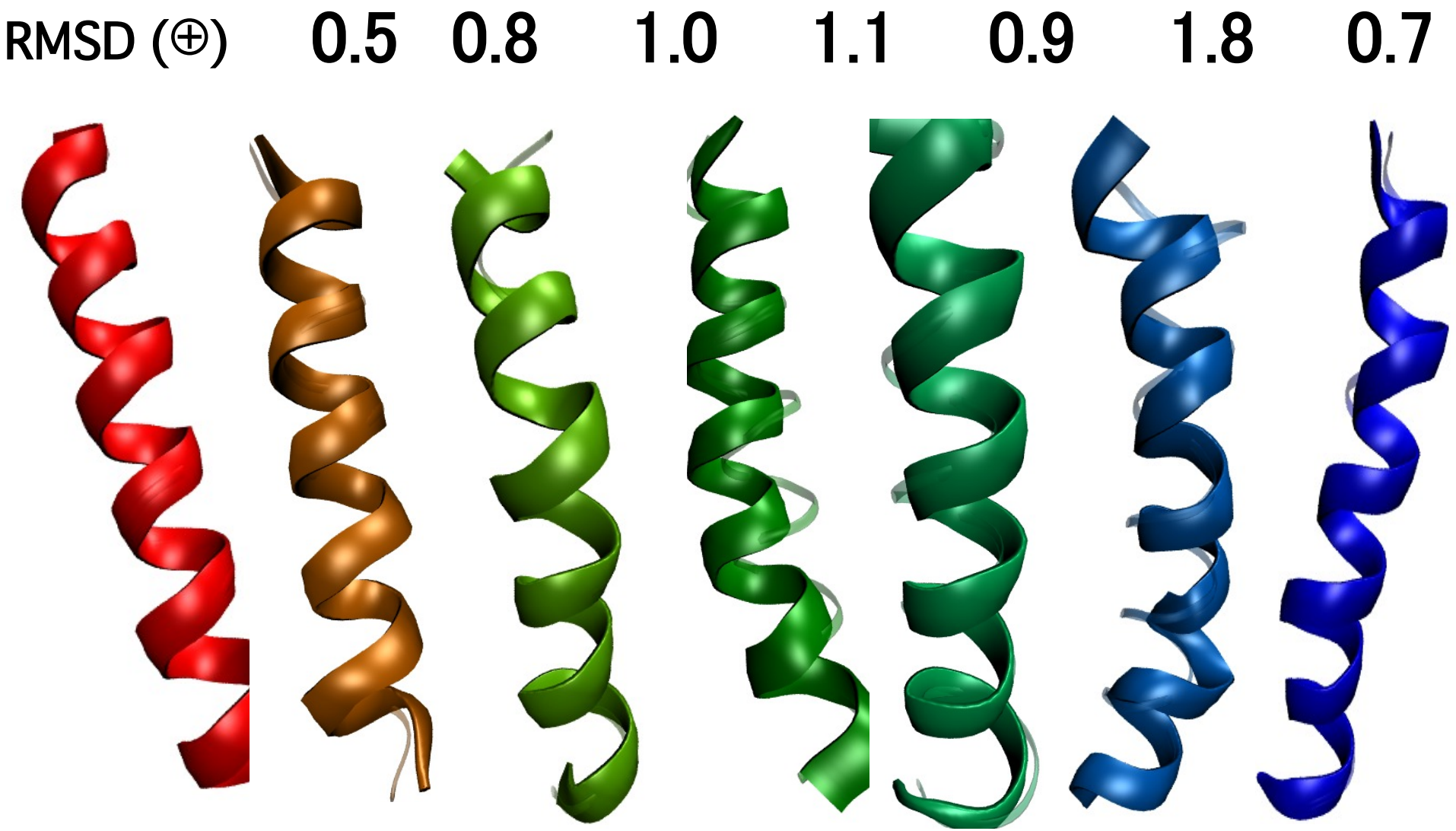
Native Structure (left) and Global-Minimum Free-Energy Structure (right) from REM Simulations



R. Urano & Y.O., submitted.

Bending of TM helices reproduced.

R. Urano & Y.O., submitted.



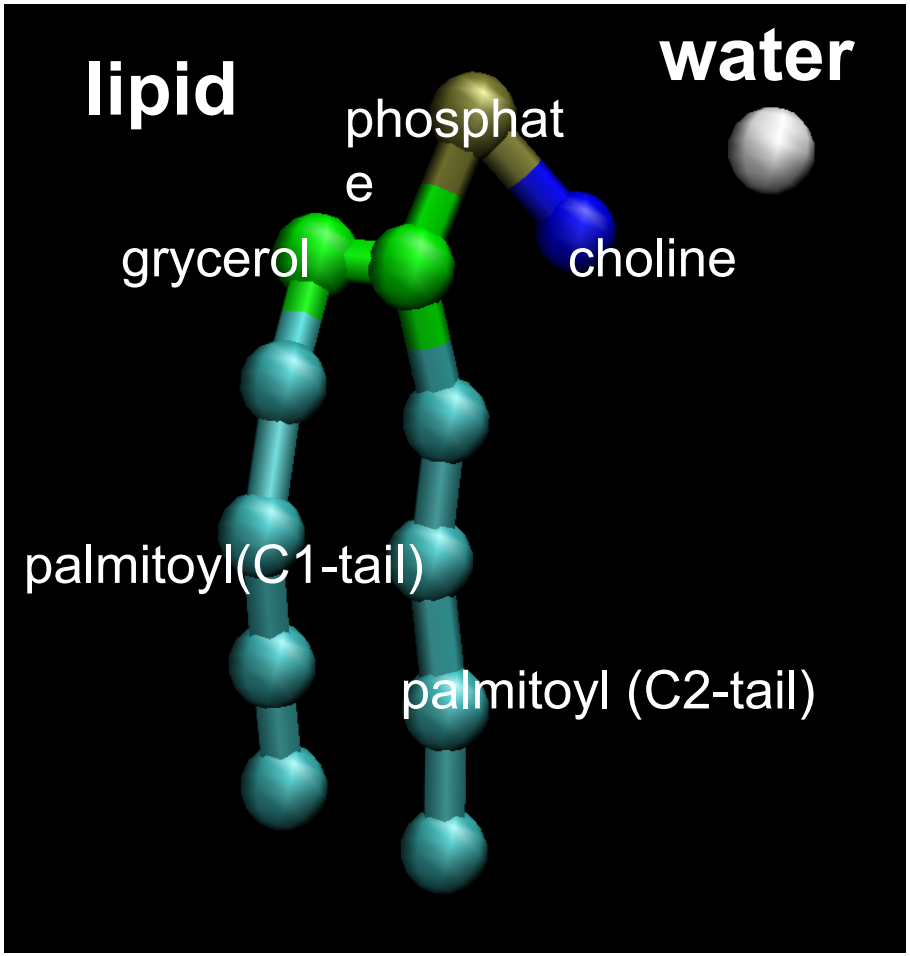
T. Nagai, R. Ueoka & Y.O., *J. Phys. Soc. Jpn.* **81**, 024002 (2012).

Phase behavior of DPPC bilayer studied by REMD with a coarse-grained model

MARTINI 2.0

a coarse-grained model.

four atoms are treated as one site, so that larger system size and/or time scale can be studied.

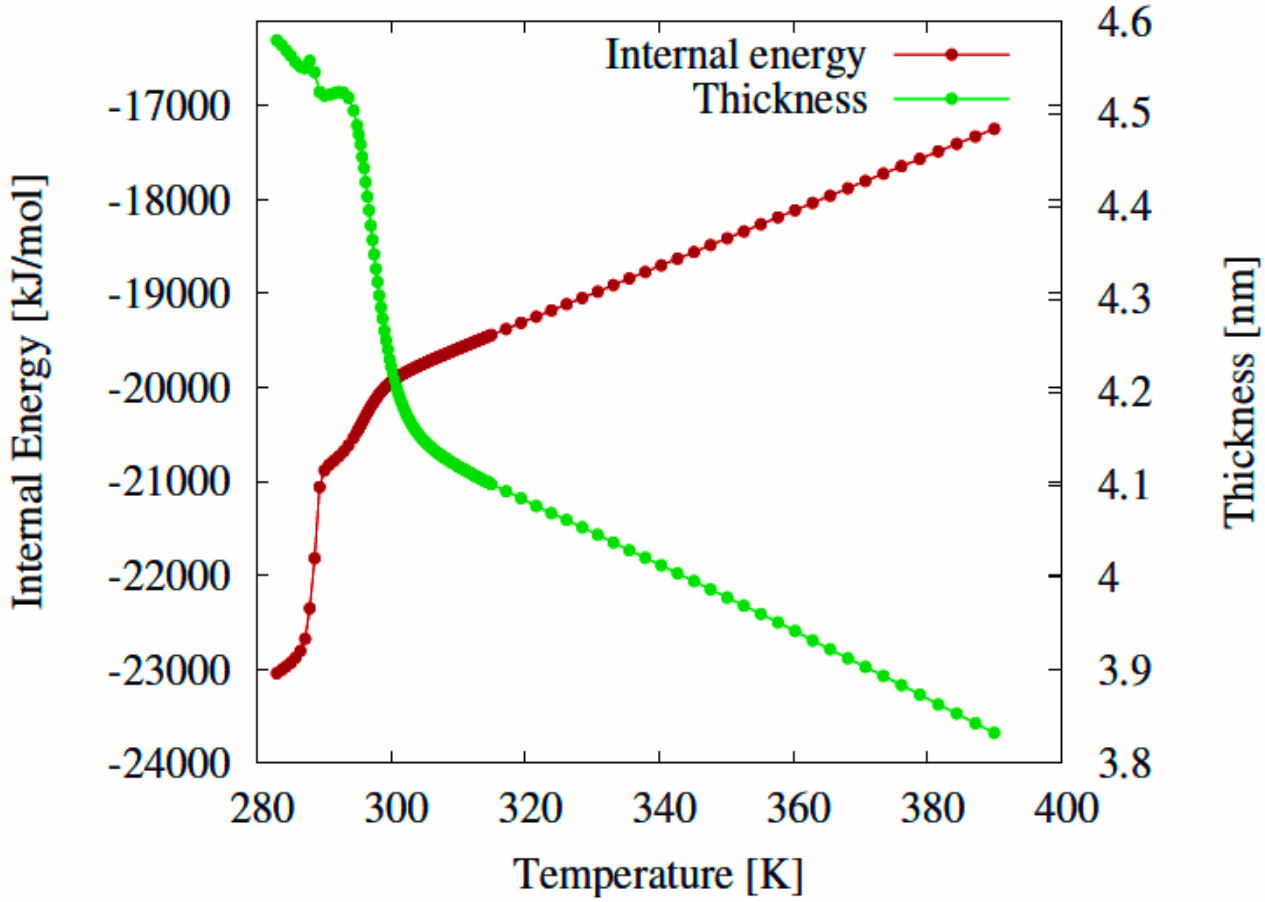


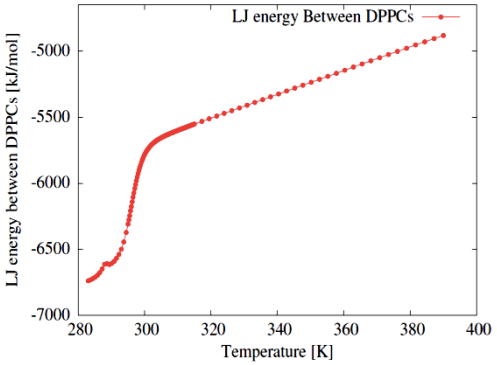
Simulation of a 32-lipid system

- system
  - 32 DPPC and 500 water particles (including 50 anti-freeze particles)
- temperature
  - Thermostat: Nose-Hoover method in the production run
  - Temperature distribution: 127 points between283 K and 390 K
- pressure
  - Reference pressure: 1atom
  - Barostat: Parrinello-Raman method in the production run
- Frequency of replica exchange
  - Every 100 step
- Time step
  - 20 fs

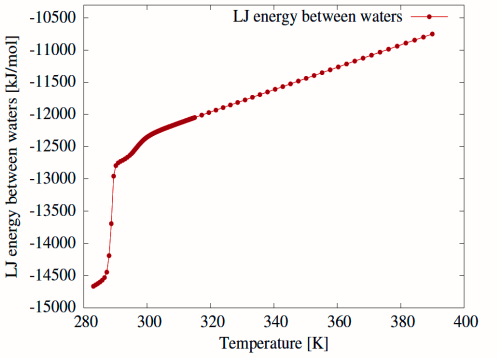
Results

Temperature dependence.





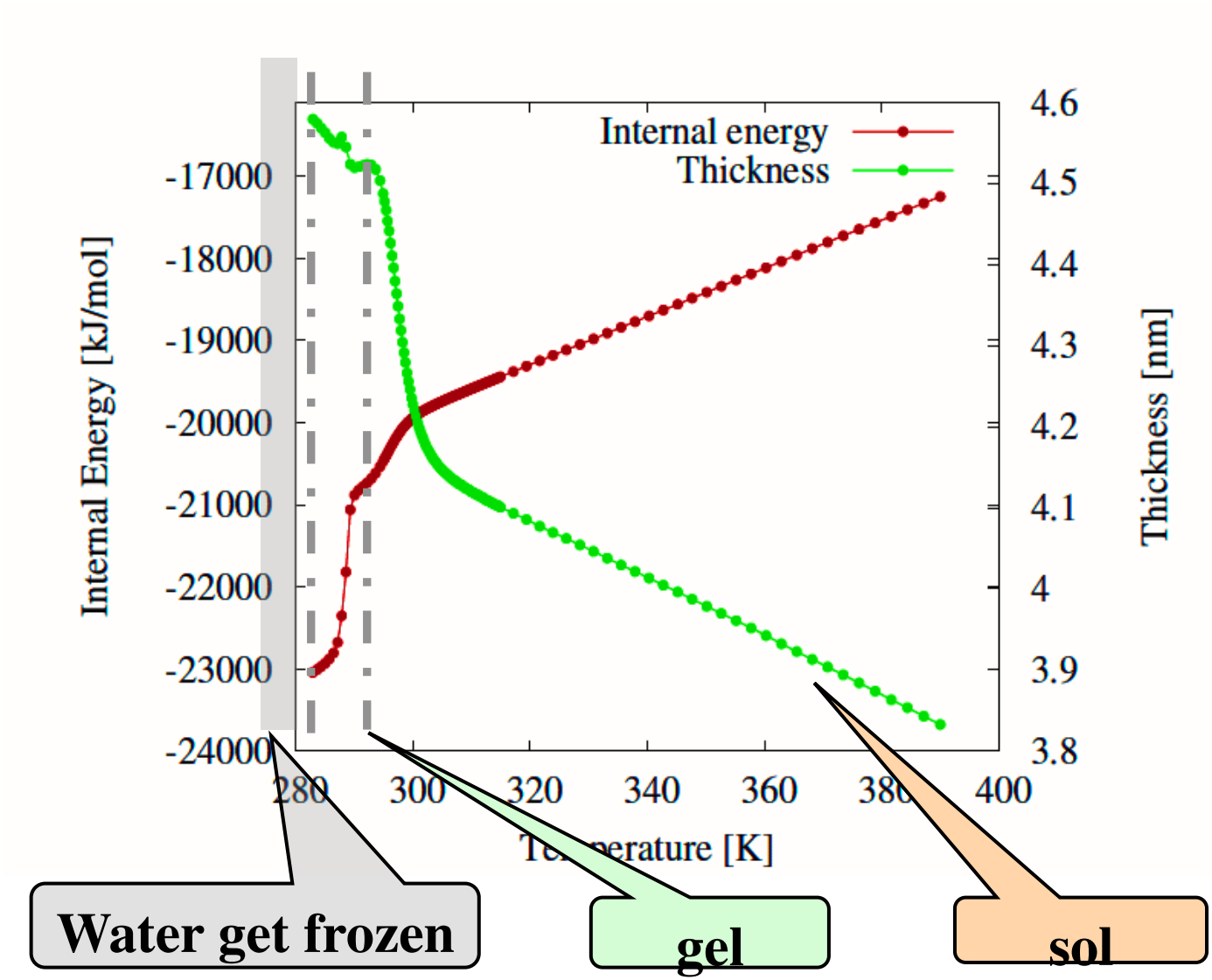
LJ between Lipids

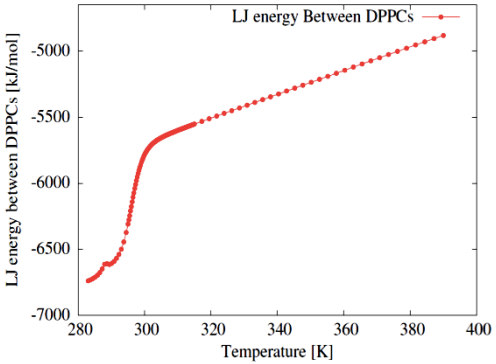


LJ between Water

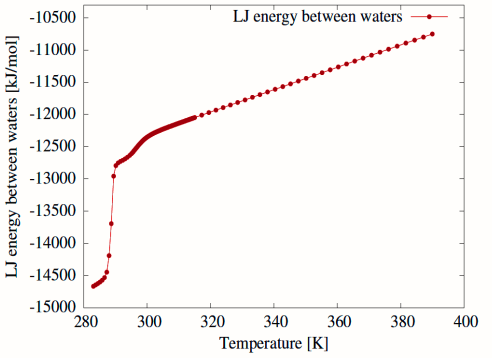
Results

Temperature dependence.



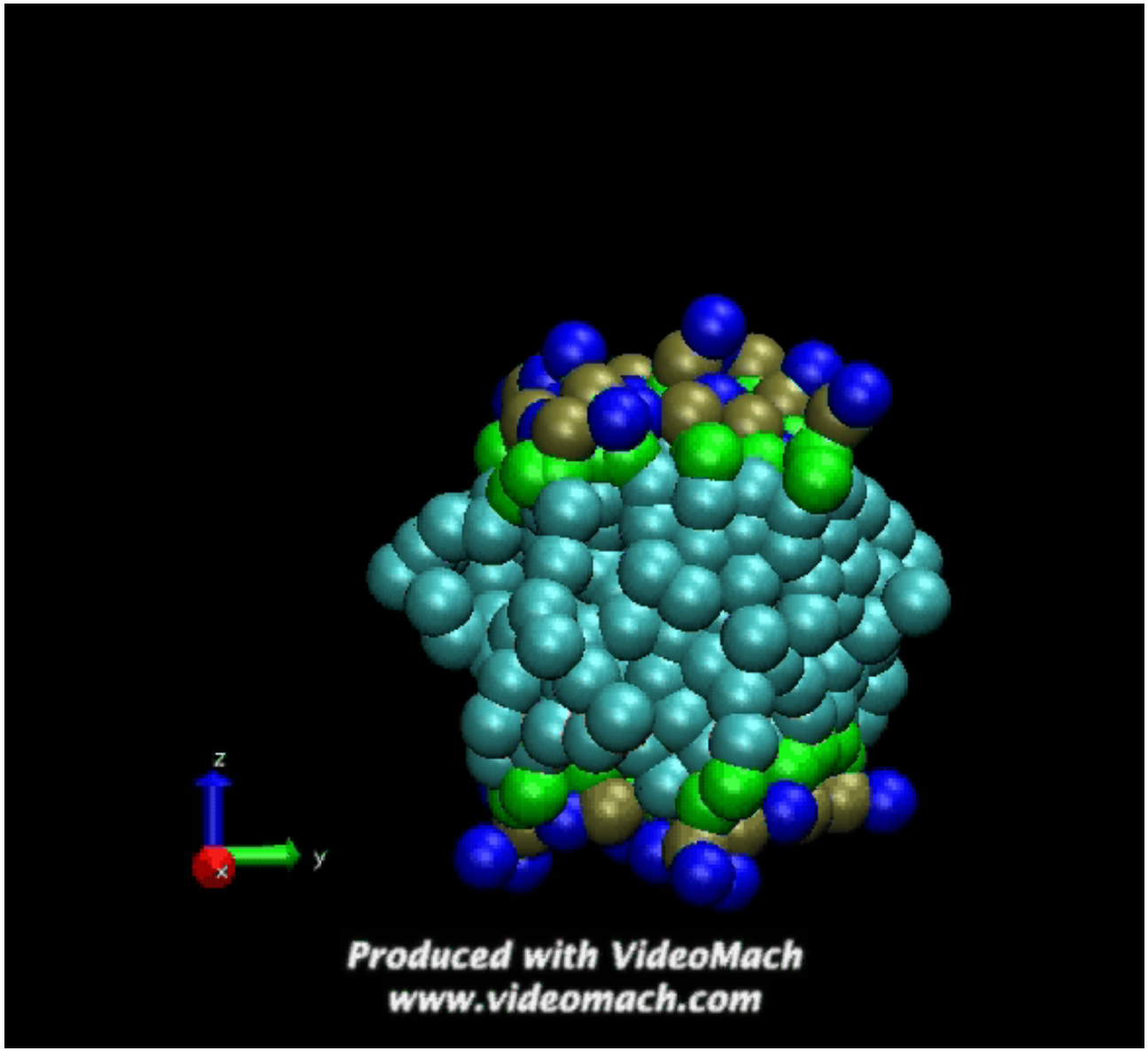


LJ between Lipids



LJ between Water

Results
Animation



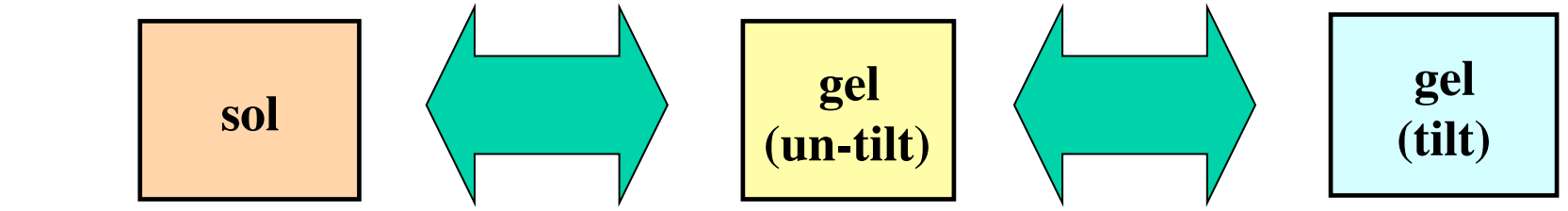
A trajectory of one of the replicas.

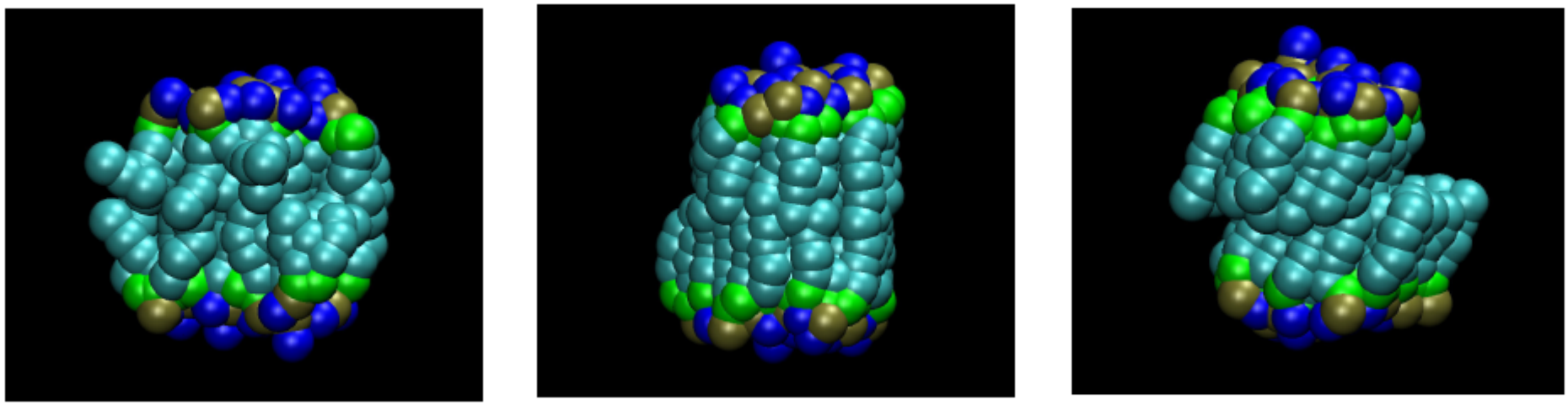
Water is omitted for clarity.

T. Nagai, R. Ueoka & Y.O., *J. Phys. Soc. Jpn.* **81**, 024002 (2012).

Results

This is a possible pathway of the phase transitions (32 lipids)

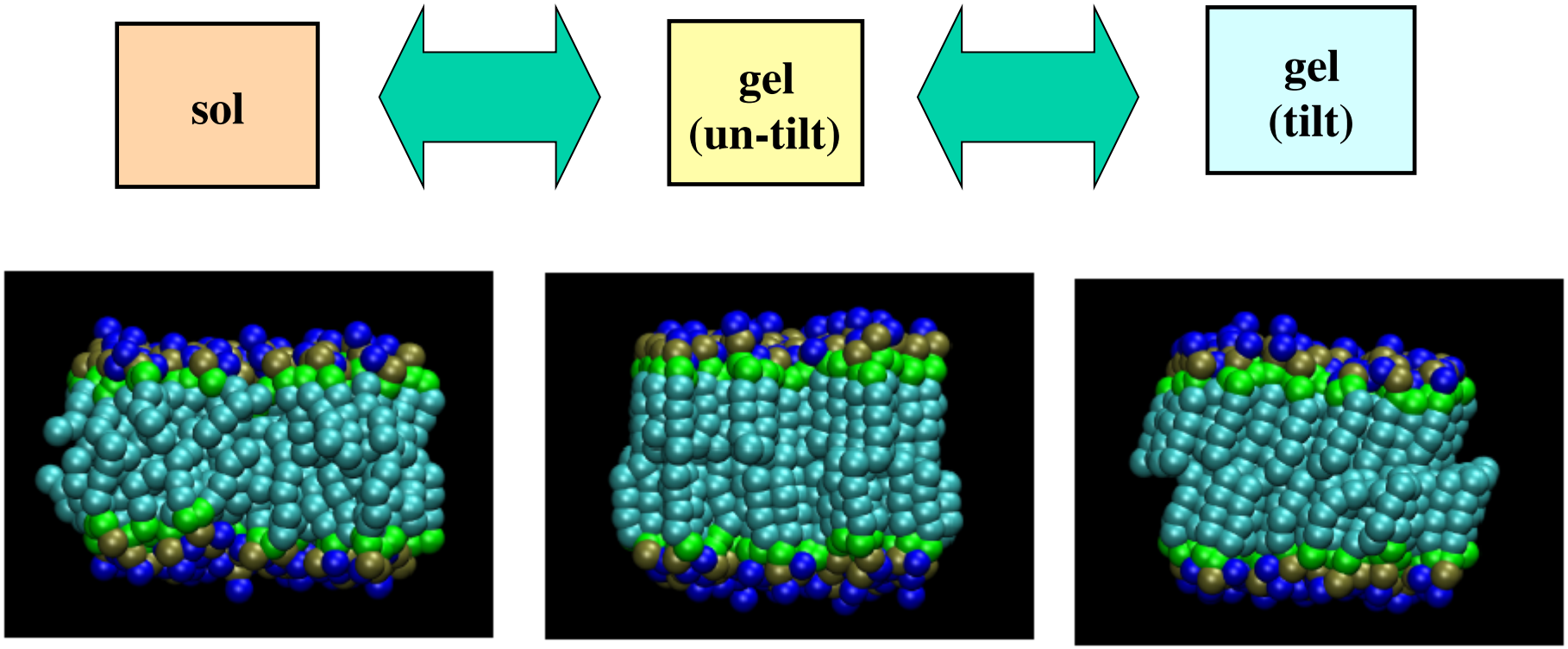




T. Nagai, R. Ueoka & Y.O., *J. Phys. Soc. Jpn.* **81**, 024002 (2012).

Results

Case for a bigger system: 128 lipids



Development of New Generalized-Ensemble Algorithms

REVIEWS: A. Mitsutake, Y. Sugita, & Y.O., *Biopolymers* **60**, 96 (2001);
Y.O., *J. Mol. Graphics Modell.* **22**, 425 (2004);
Y. Sugita, A. Mitsutake, & Y.O., in *Lecture Notes in Physics*,
W. Janke (ed.) (Springer-Verlag, 2008) pp. 369-407.

Combination of Replica-Exchange Method (REM) and
Multicanonical Algorithm (MUCA)

1. **Replica-Exchange Multicanonical Algorithm (REMUCA)**

Y. Sugita & Y.O., *Chem. Phys. Lett.* 329, 261 (2000).

A. Mitsutake, Y. Sugita, & Y.O., *J. Chem. Phys.* 118, 6664; 6676 (2003).

- Multicanonical weight factor is determined from a short REM simulation by multiple histogram techniques

2. **Multicanonical Replica-Exchange Method (MUCAREM)**

Y. Sugita & Y.O., *Chem. Phys. Lett.* 329, 261 (2000).

A. Mitsutake, Y. Sugita, & Y.O., *J. Chem. Phys.* 118, 6664; 6676 (2003).

- Multicanonical simulations are performed for each replica and a pair of replicas are exchanged

Development of New Generalized-Ensemble Algorithms

REVIEWS: A. Mitsutake, Y. Sugita, & Y.O., *Biopolymers* **60**, 96 (2001);
Y.O., *J. Mol. Graphics Modell.* **22**, 425 (2004);
Y. Sugita, A. Mitsutake, & Y.O., in *Lecture Notes in Physics*,
W. Janke (ed.) (Springer-Verlag, 2008) pp. 369-407.

Combination of Replica-Exchange Method (REM) and
Simulated Tempering (ST)

3. **Replica-Exchange Simulated Tempering (REST)**

A. Mitsutake & Y.O., *Chem. Phys. Lett.* 332, 131 (2000).

- Simulated tempering weight factor is determined from a short REM simulation by multiple histogram techniques

4. **Simulated Tempering Replica-Exchange Method (STREM)**

A. Mitsutake & Y.O., *J. Chem. Phys.* 121, 2491 (2004).

- Simulated tempering simulations are performed for each replica and a pair of replicas are exchanged

See also:

M.K. Fenwick & F.A. Escobedo, *J. Chem. Phys.* 119, 11998 (2003).

KEY ELEMENT: Multiple-Histogram Reweighting Techniques (WHAM)

A. Ferrenberg & R. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989); S. Kumar, D. Bouzida, R. Swendsen, P. Kollman & J. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).

Given M set of histograms $N_m(E)$, the following WHAM equations are solved iteratively for density of states $n(E)$ and dimensionless Helmholtz free energy f_m :

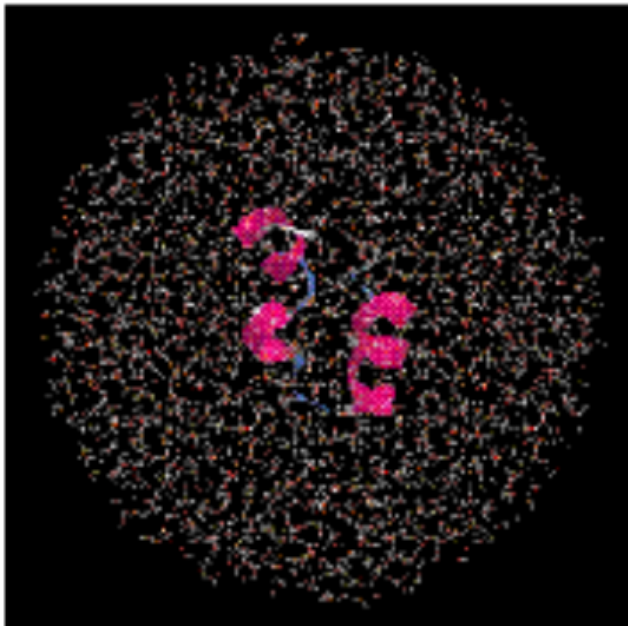
$$n(E) = \frac{\sum_{m=1}^{M} N_m(E)}{\sum_{m=1}^{M} n_m e^{f_m - \beta_m E}}, \text{ where } e^{-f_m} = \sum_E n(E) e^{-\beta_m E}.$$

Folding of a Small Globular Protein

T. Yoda, Y. Sugita & Y.O., *Biophys. J.* **99**, 1637 (2010).
T. Yoda, Y. Sugita & Y.O., *Proteins* **82**, 933-943 (2014).

Challenging the prediction of the 3-dimensional structure of a small protein by MUCAREM.

Villin headpiece subdomain
(36 amino acids; 596 atoms)
sphere of water with radius 30 Å
(3513 water molecules);
Total number of atoms = 11,135



Primary Sequence of HP-36

```
1         10        20        30
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
```

helix1 helix2 helix3

Computational Details

T. Yoda, Y. Sugita & Y.O., *Biophys. J.* **99**, 1637 (2010).

(Force Field = CHARMM22/CMAP for protein
& TIP3P for water)

(1) REMD with 96 replicas in implicit solvent (GB/SA); initial conformation was fully extended

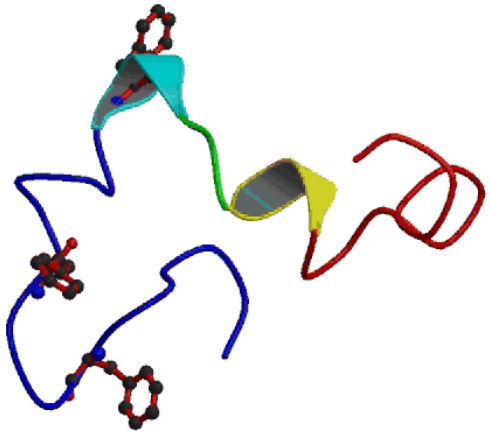
(2) Unfolded protein w/o any secondary structures was soaked in a sphere of radius 30Å (with 3513 TIP3P water molecules)

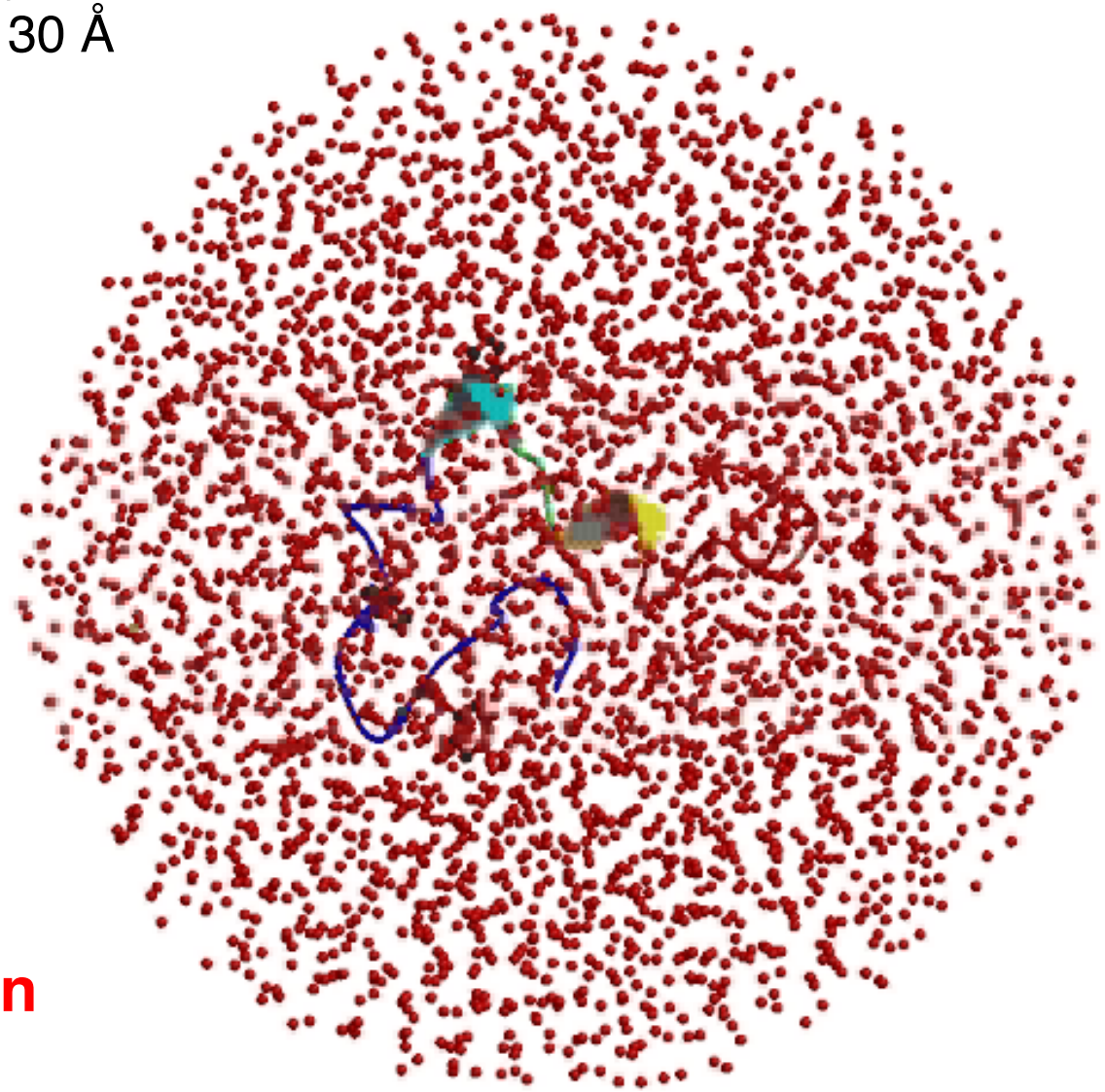
(3) REMD with 128 replicas (T = 250 K ∼ 700 K)

(4) Determine multicanonical weight factors by WHAM (iterate several times to refine weight)

(5) Two production runs of MUCAREM with 8 replicas (MUCAREM1: 1.127 μs in total covering T = 269 K ∼ 699 K
MUCAREM2: 1.157 μs in total covering T = 289 K ∼ 699 K)

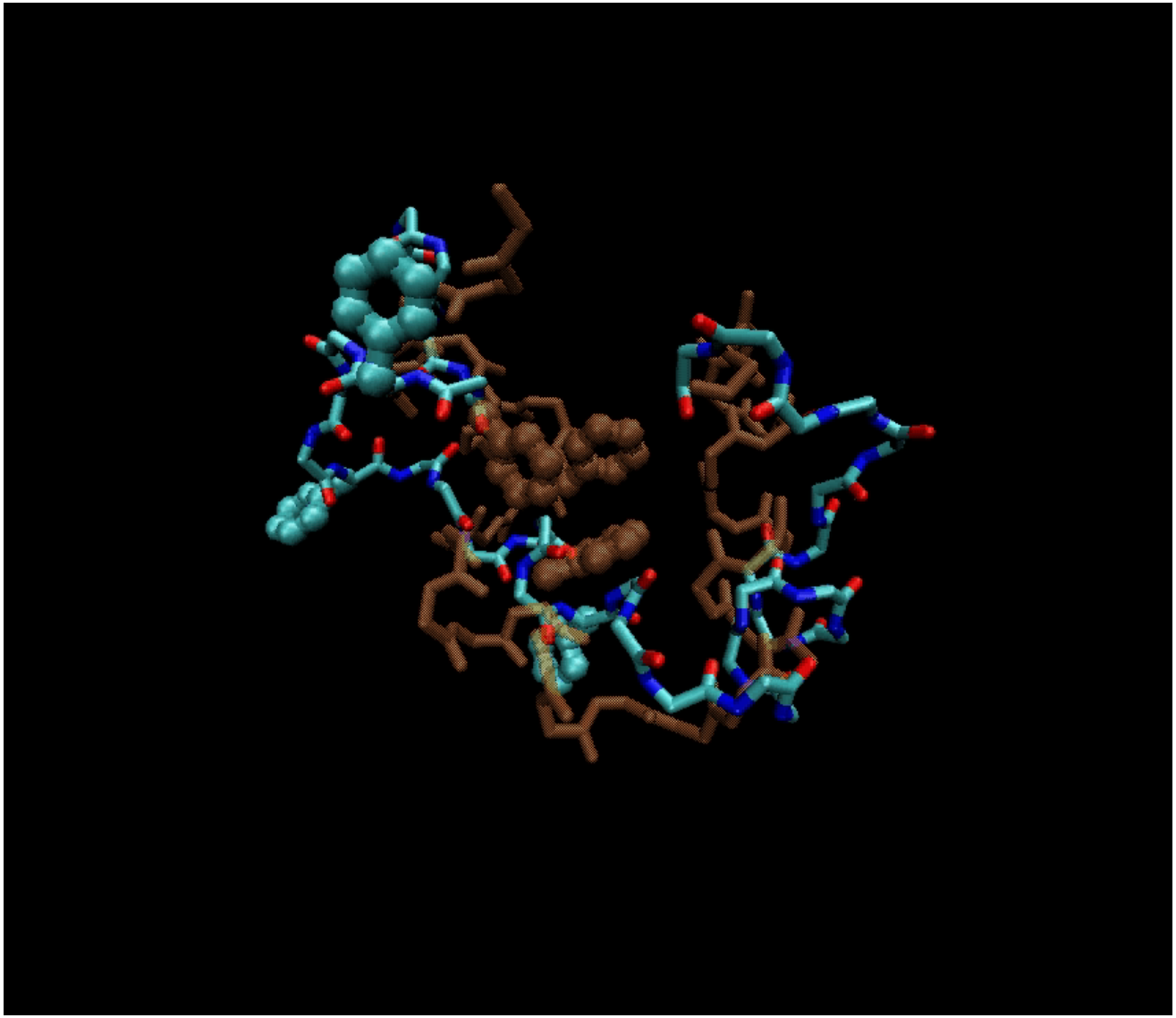
Villin headpiece subdomain
(36 amino acids; 596 atoms)
in sphere of water of radius 30 Å
(3513 water molecules);
altogether 11,135 atoms





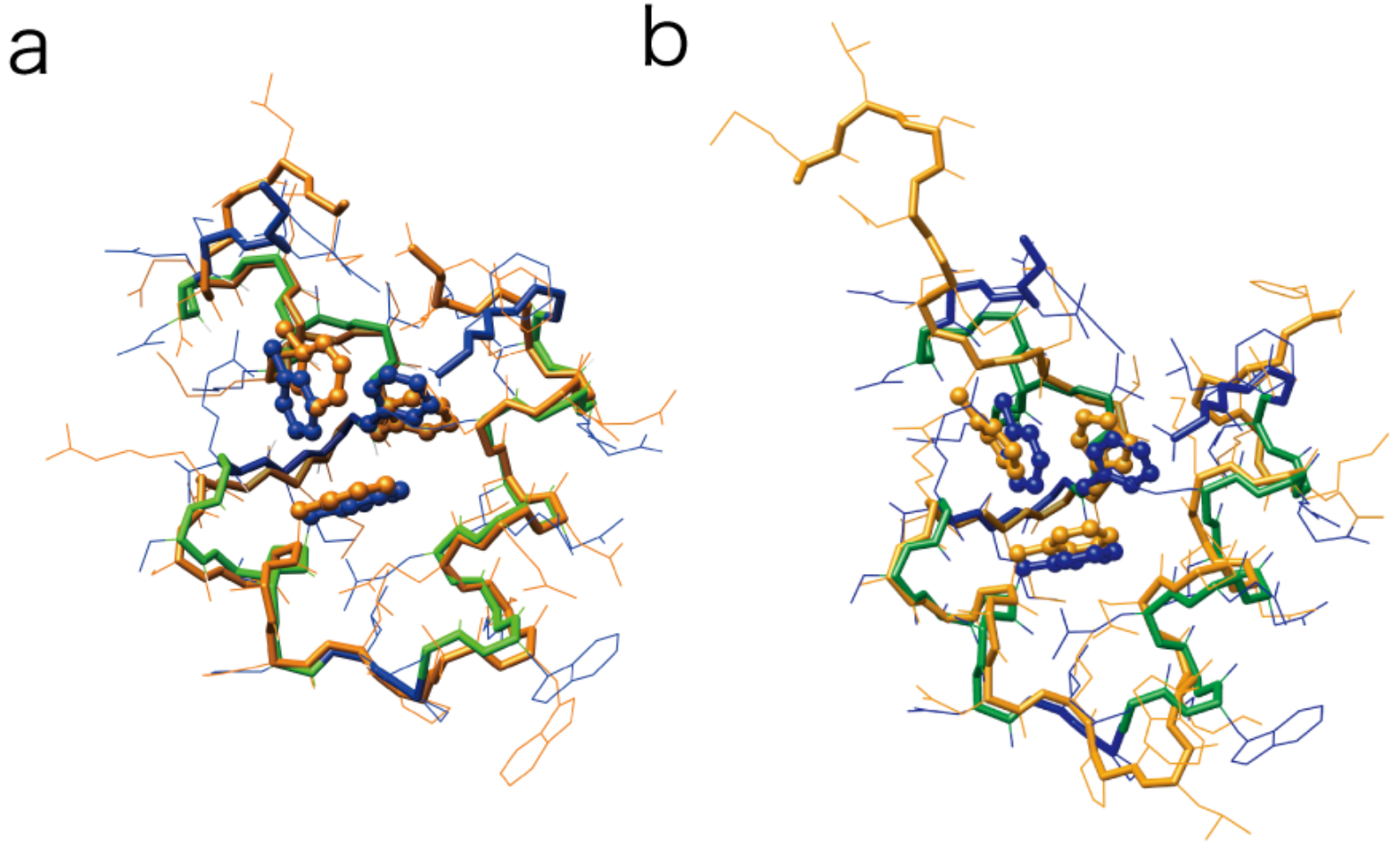
MUCAREM simulation

MUCAREM2 (Replica 5)



Simulation and movie by T. Yoda

Native-Like Structures Obtained from MUCAREM2 (Left) and MUCAREM1 (Right)

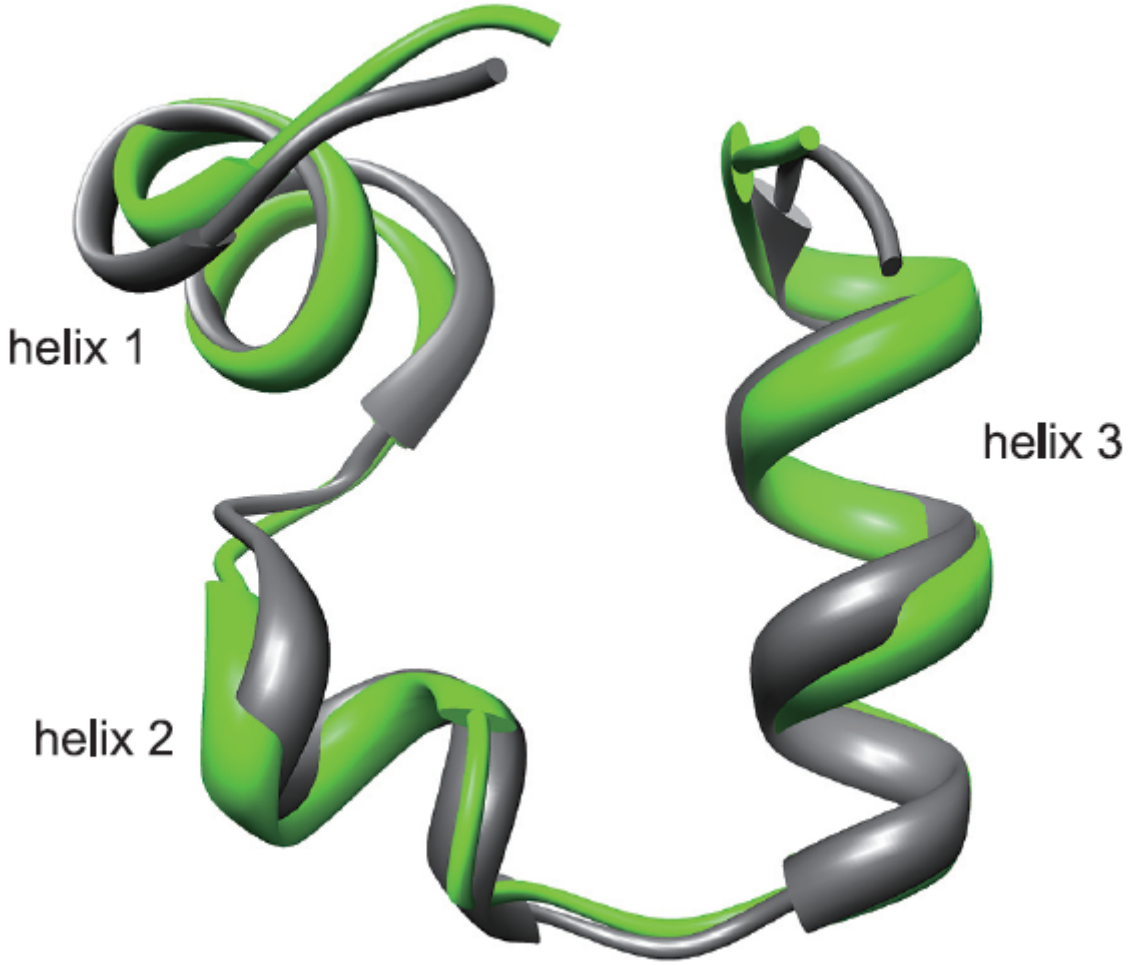


Main-Chain RMSD = 1.1 ⊕ (residues 2 to 35) [Replica 5]

Main-Chain RMSD = 3.3 Å (residues 2 to 35) [Replica 8] RMSD = 1.2 Å (residues 9 to 32)

Native-Like Structures Obtained from MUCAREM

T. Yoda, Y. Sugita & Y.O., *Biophys. J.* **99**, 1637 (2010).



Main-Chain RMSD = 1.1 ⊕ (residues 2 to 35) [Replica 5]
灰色：自然の構造（PDB ID: 1YRF）、緑色：シミュレーションの結果

α-Helix Formation

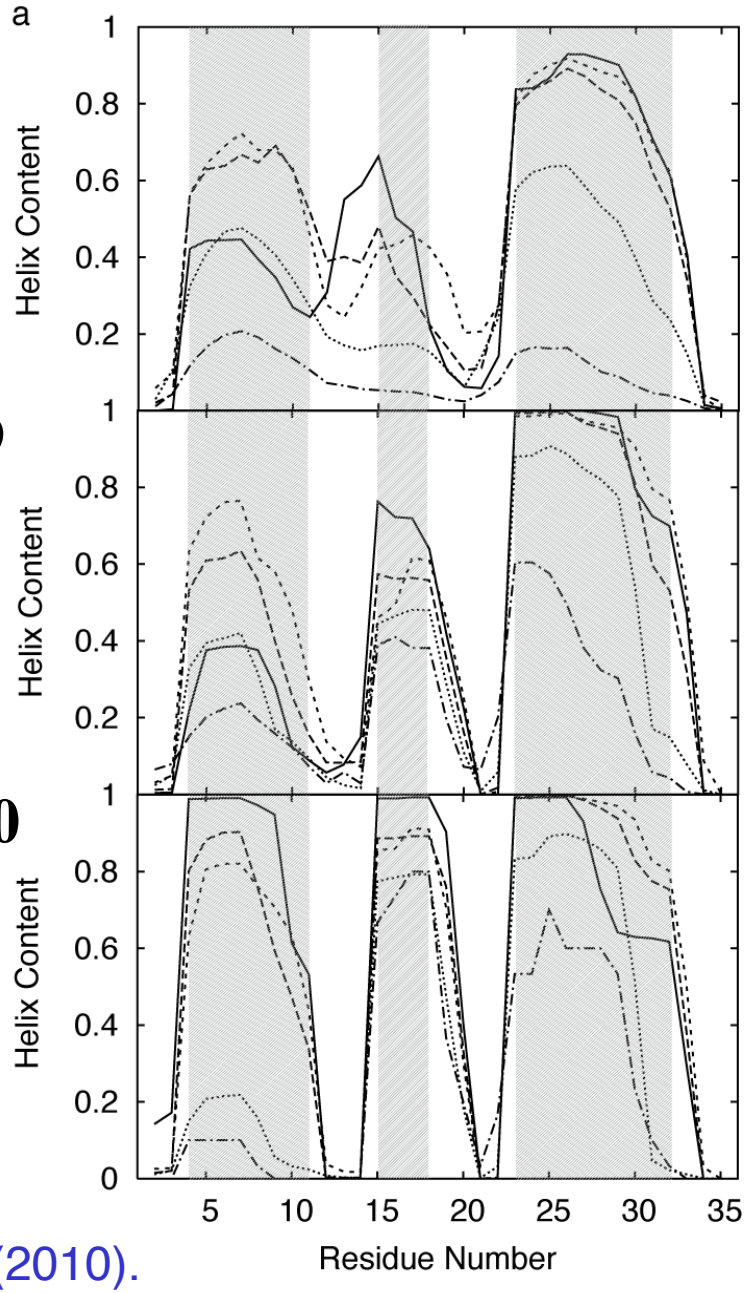
$N_n \leq 7$

N_n : no. of non-local side-chain native contacts (=18 for X-ray structure)

$N_n = 8, 9$

Solid: 300 K,
Dashed: 350 K,
Short Dashed: 400 K,
Dotted: 500 K,
Dotted Dash: 600 K

$N_n \geq 10$



T. Yoda, Y. Sugita & Y.O., *Biophys. J.* **99**, 1637 (2010).

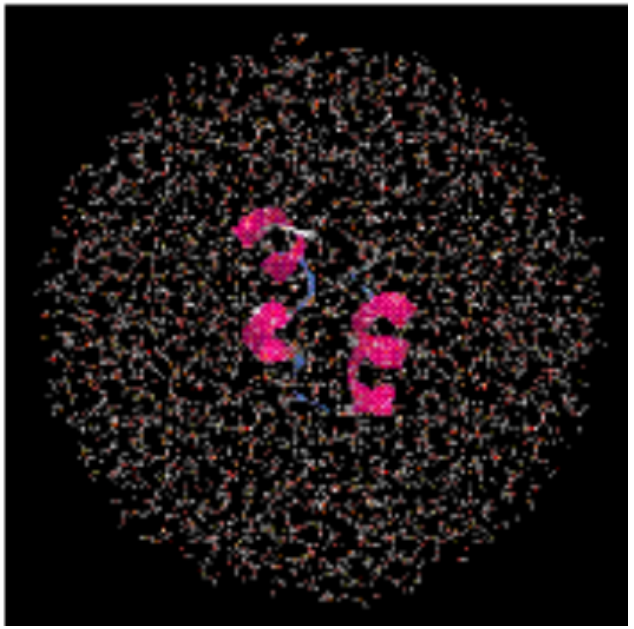
Salt Effects on Folding of a Small Globular Protein

T. Yoda, Y. Sugita & Y.O., *Proteins* **82**, 933-943 (2014).

Challenging the prediction of the 3-dimensional structure of a small protein by MUCAREM.

R = 30A, 3550wat, 11246atom

Villin headpiece subdomain
(36 amino acids; 596 atoms)
sphere of salted water with radius 30 Å
(3494 water molecules,
11 K^+, 13 Cl^- ≈ 0.2 M KCl);

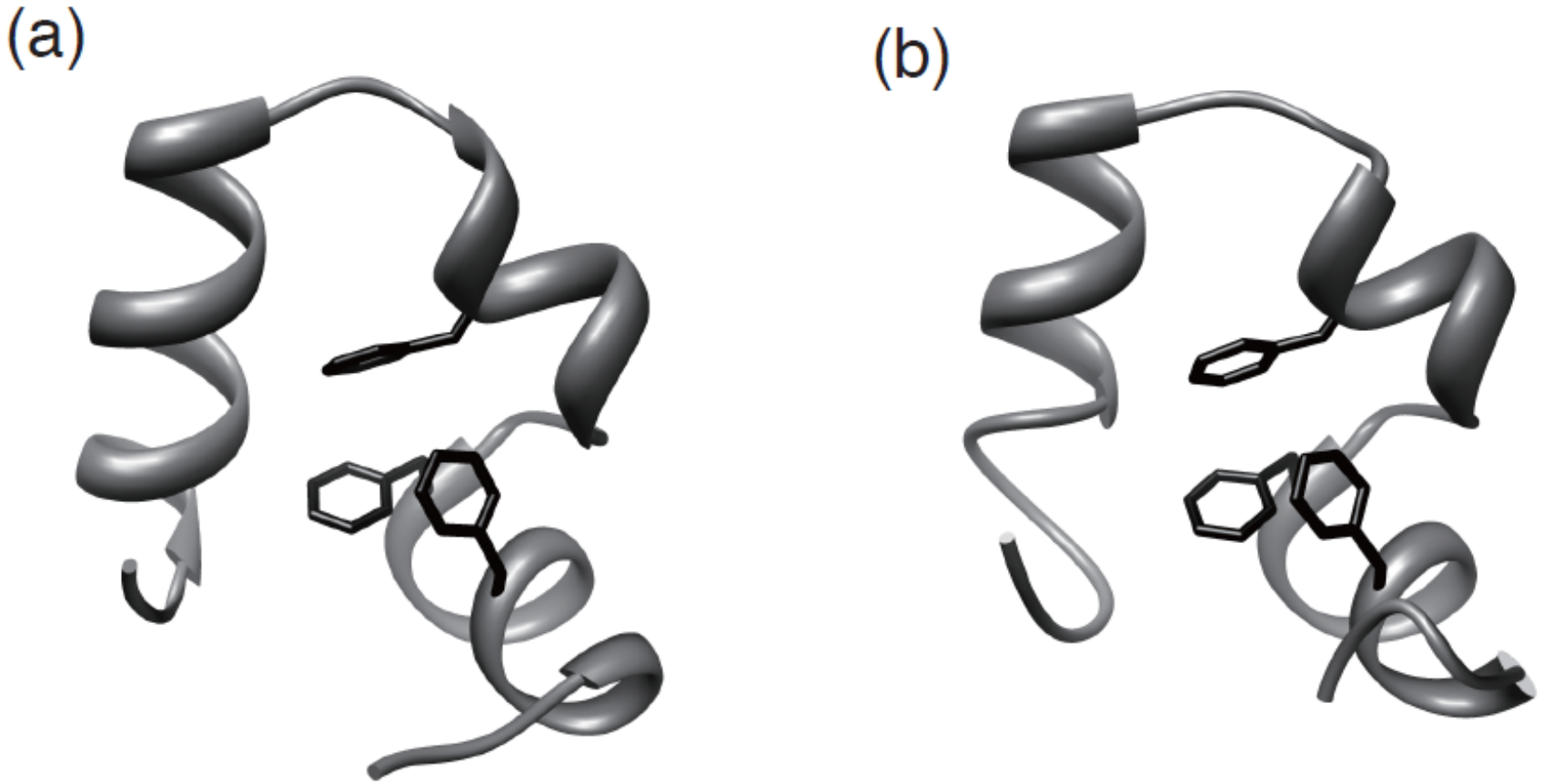


Primary Sequence of HP-36

```
1                   10                  20                  30
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
 helix1              helix2              helix3
```

Native-Like Structure (Global Minimum in Free Energy) Obtained from MUCAREM Simulation (Left)

T. Yoda, Y. Sugita & Y.O., *Proteins* **82**, 933-943 (2014).

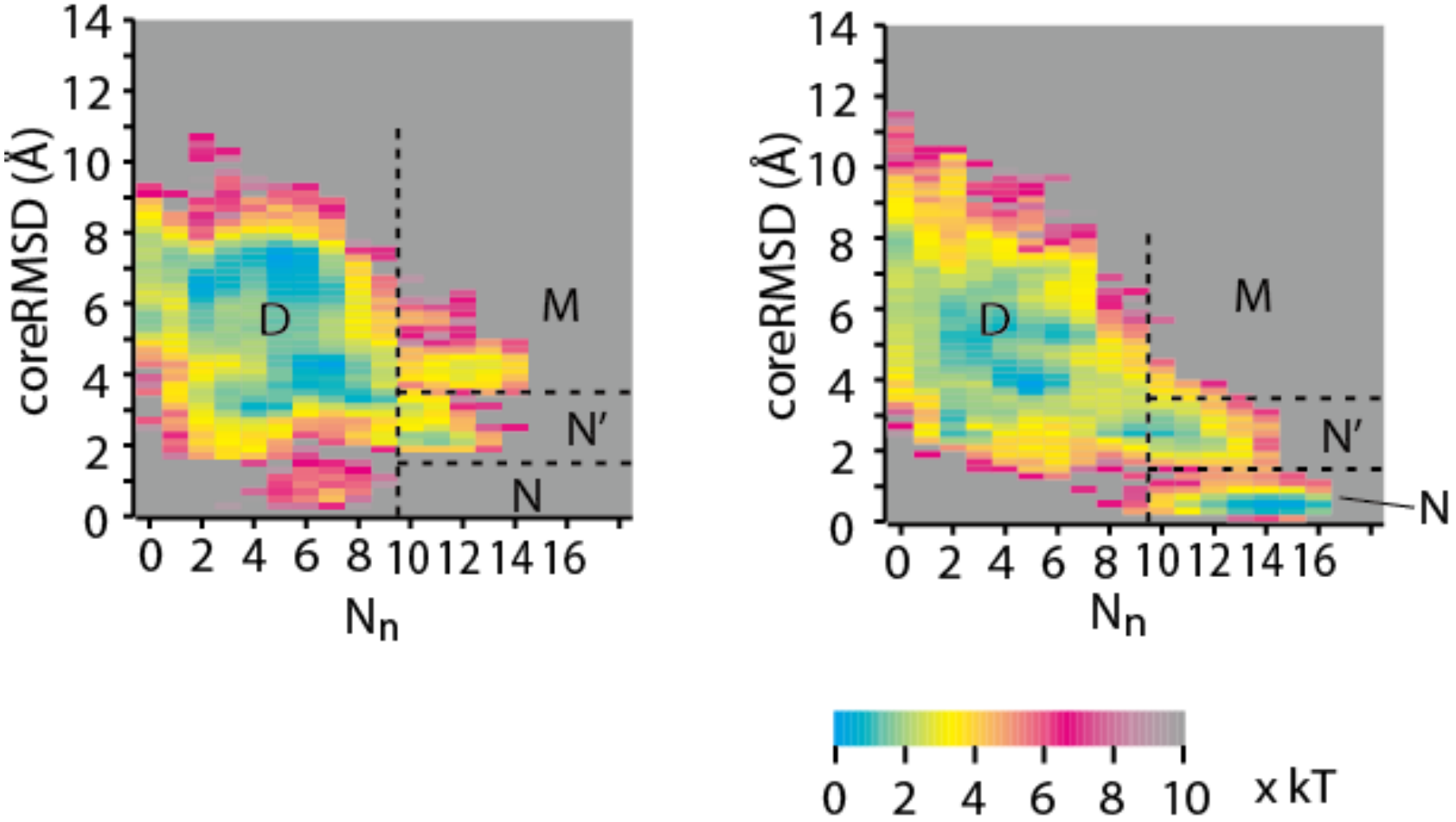


Main-Chain RMSD = 1.25 ⊕

Experimental Structure (PDB ID: 1YRF)

Free Energy Landscape

T. Yoda, Y. Sugita & Y.O., *Proteins* **82**, 933-943 (2014).



in Pure Water

in 0.2 M Salted Water

Multidimensional Replica-Exchange Method (MREM)

Y. Sugita, A. Kitao & Y.O., *J. Chem. Phys.* **113**, 6042 (2000).

1. System

M *Non-Interacting* Replicas of the Original System at *M* Different Sets of Temperatures and Parameters

2. Replica-Exchange

Step 1: Independent Canonical Simulations Performed for Each Replica

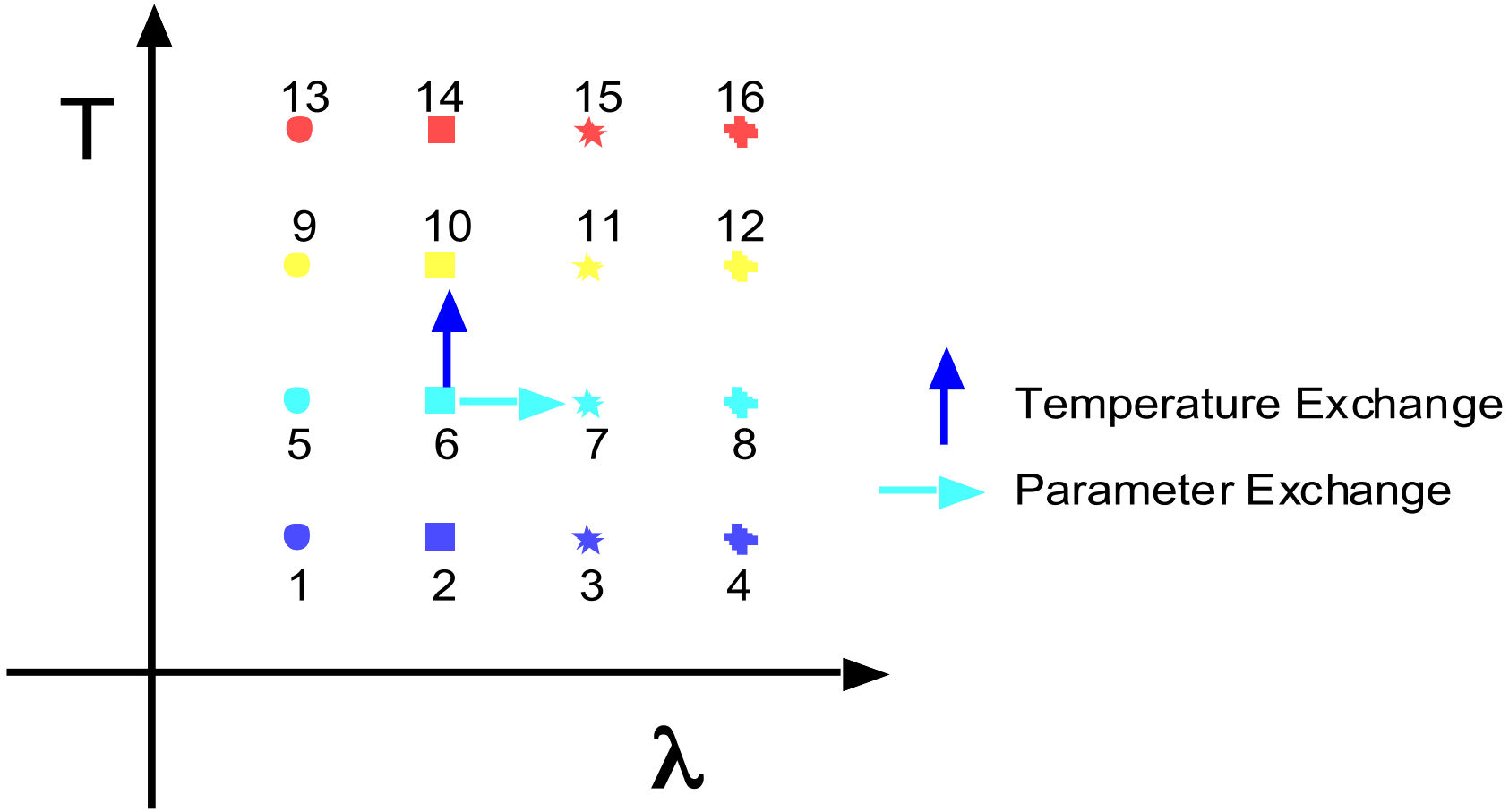
Step 2: A Pair of Replicas Corresponding to Neighboring Temperatures or Parameters are Exchanged

Repeat These 2 Steps

3. Canonical Distribution at Any Temperature

by Multiple Histogram Reweighting Techniques (WHAM)

Multidimensional Replica-Exchange Method (MREM)



From Multidimensional REM to Multidimensional MUCA and ST

A. Mitsutake & Y.O., *Phys. Rev. E* **79**, 047701 (2009);
J.Chem. Phys. **130**, 214105 (2009);
A. Mitsutake, *J.Chem. Phys.* **131**, 094105 (2009).

MMUCA: random walk in multidimensional energy

MST: random walk in multidimensional parameter

MREM: random walk in multidimensional parameter

e.g., $E_\lambda = E + \lambda V$

$$W_{mu}(E,V) = \frac{1}{n(E,V)}$$

$$W_{ST}(E,V;T_m,\lambda_n) = \exp(-\beta_m(E+\lambda_n V) + f_{m,n})$$

WHAM eqns.

$$n(E,V) = \frac{\sum_{m,n=1}^{M} N_{m,n}(E,V)}{\sum_{m,n=1}^{M} n_{m,n} e^{f_{m,n} - \beta_m(E+\lambda_n V)}}, \text{ where } e^{-f_{m,n}} = \sum_{E,V} n(E,V) e^{-\beta_m(E+\lambda_n V)}.$$

Examples of Multidimensional REM, MUCA, and ST

A. Mitsutake & Y.O., *Phys. Rev. E* **79**, 047701 (2009);
J. Chem. Phys. **130**, 214105 (2009);
A. Mitsutake, *J. Chem. Phys.* **131**, 094105 (2009).

$$E_\lambda = E + \lambda V$$

$\lambda = h =$ external field
$V = M =$ magnetization

1. Simulated Tempering and Magnetizing

random walk in temperature *T* and external field *h*

∗Ising Model

T. Nagai & Y.O., *Phys. Rev. E* **86**, 056705 (2012).

$$H = E - hM$$

$$E = -\sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

$$M = \sum_{i}^{N} \sigma_i \,,$$

∗3-state Potts Model

T. Nagai, Y.O., & W. Janke,
J. Stat. Mech. (2013) P02039.

$$H = E - hM \,,$$

$$E = -\sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j} \,,$$

$$M = \sum_{i=1}^{N} \delta_{0,\sigma_i} \,,$$

Examples of Multidimensional REM, MUCA, and ST

A. Mitsutake & Y.O., *Phys. Rev. E* **79**, 047701 (2009);
J. Chem. Phys. **130**, 214105 (2009);
A. Mitsutake, *J. Chem. Phys.* **131**, 094105 (2009).

$$E_\lambda = E + \lambda V$$

2. Isobaric-Isothermal Ensemble（定圧定温アンサンブル）

$$\lambda = P = \text{pressure} \qquad V = \text{volume}$$

∗ **MUCA: Multibaric-Multithermal Algorithm (MUBATH)**

random walk in potential energy *E* and volume *V*

H. Okumura & Y.O., *Chem. Phys. Lett.* **383**, 391 (2004). (MC version)
H. Okumura & Y.O., *Chem. Phys. Lett.* **391**, 248 (2004). (MD version)

∗ **REM: random walk in temperature *T* and pressure *P***

Y. Sugita & Y.O., in *Lect. Notes in Computational Science & Engineering*,
ed. by T. Schlick and H.Gun (2002) pp. 304-332; cond-mat/0102296.
T. Okabe, M. Kawata, Y.O. & M. Mikami, *Chem. Phys. Lett.* **335**, 435 (2001).
Also, see D. Paschek & A. Garcia, *Phys. Rev. Lett.* **93**, 238105 (2004).

∗ **ST: random walk in temperature *T* and pressure *P***

Y. Mori & Y.O., *J. Phys. Soc. Jpn.* **79**, 074003 (2010).

Examples of Multidimensional REM, MUCA, and ST

A. Mitsutake & Y.O., *Phys. Rev. E* **79**, 047701 (2009);
J. Chem. Phys. **130**, 214105 (2009);
A. Mitsutake, *J. Chem. Phys.* **131**, 094105 (2009).

$$E_\lambda = E + \lambda V$$

3. Umbrella Sampling

$$H_k(q,p) = K(p) + E_0(q) + \sum_{l=1}^{L} \lambda^l V_l(q), \text{ where } V_l(\xi) = K_l\left(\xi(q) - d_l\right)^2.$$

∗ **REM: Replica-Exchange Umbrella Sampling (REUS)**
random walk in reaction coordinate ξ

Y. Sugita, A. Kitao & Y.O., *J. Chem. Phys.* **113**, 6042 (2000).

∗ **ST: Simulated Tempering Umbrella Sampling (STUS)**
random walk in reaction coordinate ξ

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).

Cf.

∗ **MUCA: Metadynamics (Wang-Landau in reaction coordinate)**
random walk in reaction coordinate ξ

A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002).

拡張アンサンブル法によるタンパク質と リガンドの結合構造予測

Prediction of protein-ligand binding structures by generalized-ensemble simulations

Y.O., T. Tanaka & H. Kokubo, *J. Comp.-Aided Mol. Design* **24**, 699-712 (2010).

H. Kokubo, T. Tanaka & Y.O., *J. Comput. Chem*. **32**, 2810-2821 (2011).

H. Kokubo, T. Tanaka & Y.O., *J. Chem. Theory Comput.* **9**, 4660-4671 (2013).

H. Kokubo, T. Tanaka & Y.O., *J. Comput. Chem*. **34**, 2601-2614 (2013).

Y.O., H. Kokubo & T. Tanaka, *J. Chem. Theory Comput.* **10**, 3563-3569 (2014).

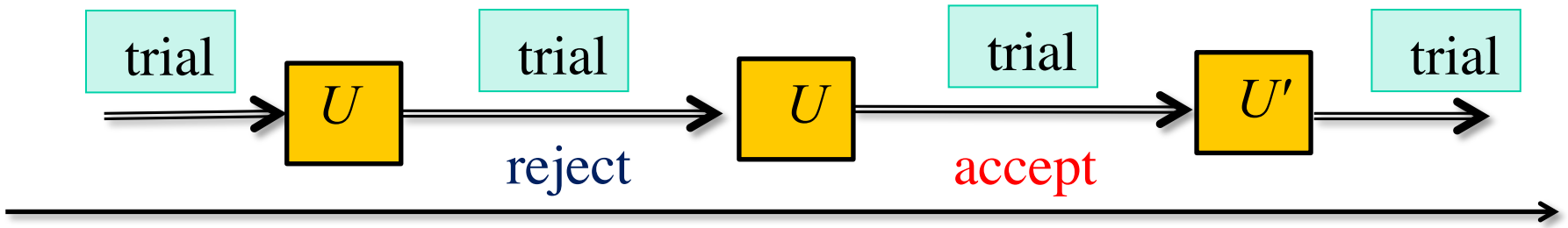
Review:

H. Kokubo, T. Tanaka, & Y.O., in *Advances in Protein Chemistry and Structural Biology*, T. Karabencheva-Christova (ed.) (Elsevier, Amsterdam, 2013) pp. 63-91.

焼き戻し傘サンプル法

Simulated Tempering Umbrella Sampling (STUS)

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).



アンブレラポテンシャル Umbrella Potential

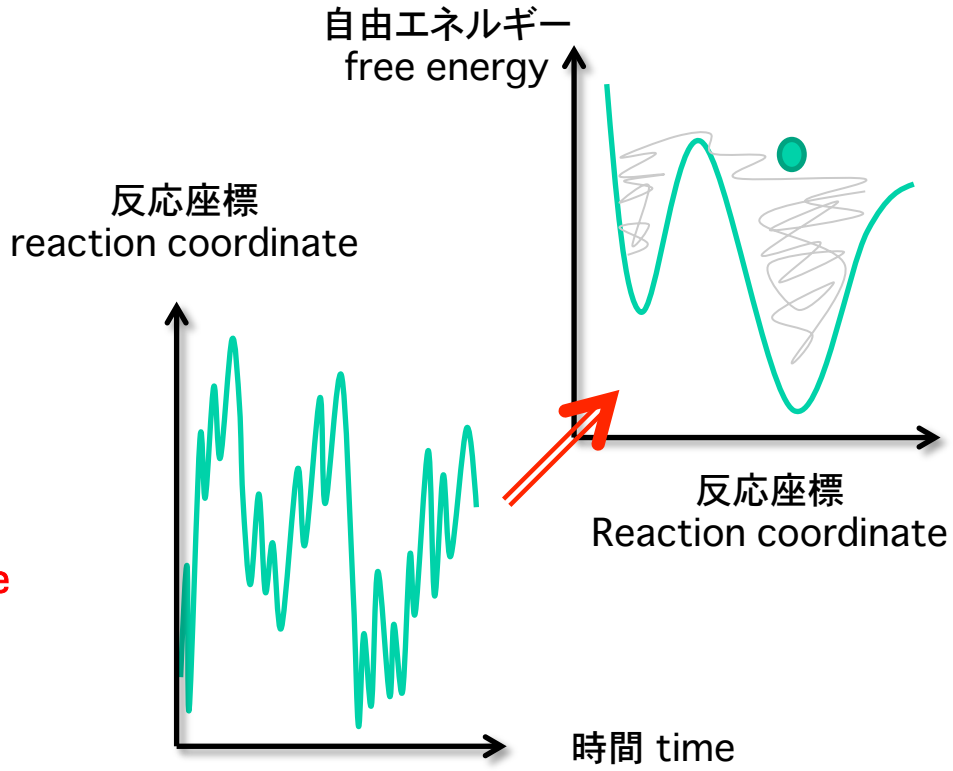
$$U[\xi(r)] = k\left[\xi(r) - \xi_0\right]^2$$

遷移確率 Transition Probability

$$w(X \to X') = \min\left[1, \exp(-\Delta)\right]$$
$$\Delta = \beta(U' - U) - (f' - f)$$

無次元化された自由エネルギー Dimensionless Free Energy

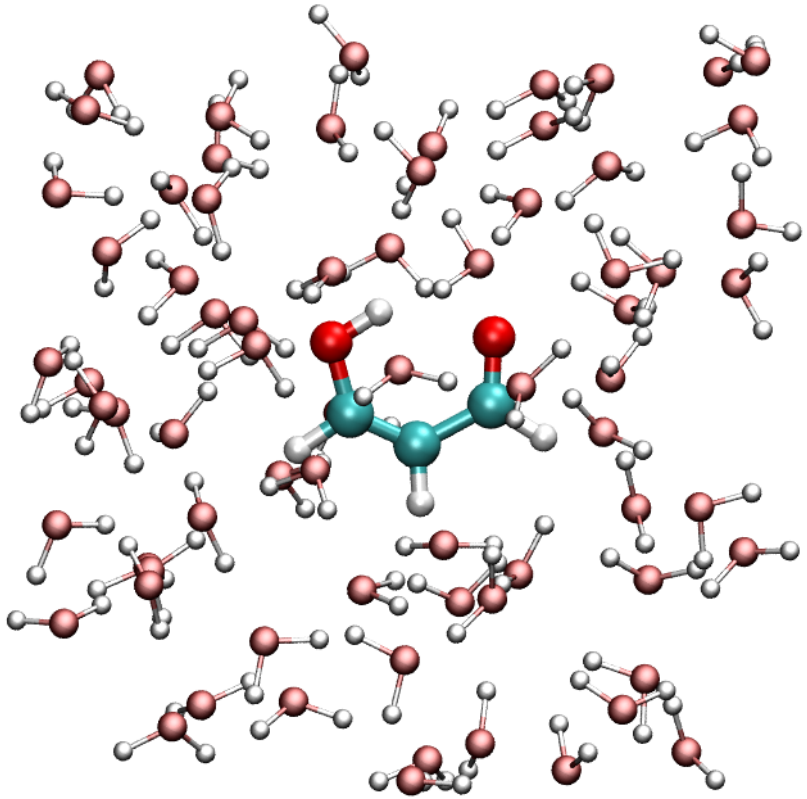
$$f = -\ln\left\{\int_{-\infty}^{\infty} dr \exp\left\{-\beta\left[E(r) + U(r)\right]\right\}\right\}$$



マロンアルデヒドの分子内プロトン移動

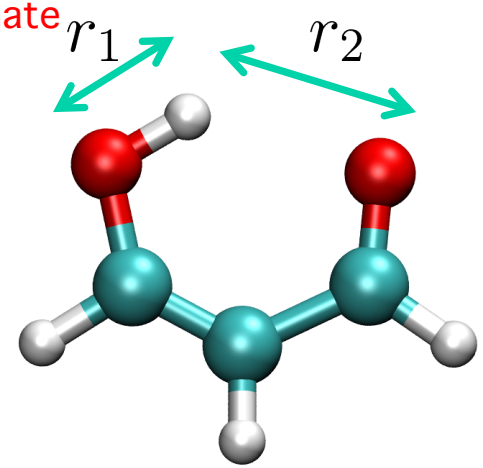
Proton Transfer in Malonaldehyde

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).

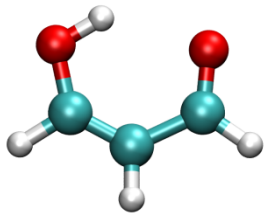


反応座標の定義
Definition of reaction coordinate

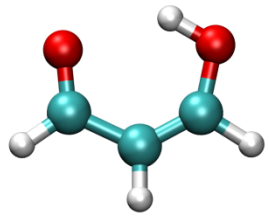
$$\xi = r_2 - r_1$$



$\xi > 0$



$\xi < 0$



マロンアルデヒド malonaldehyde: 1 分子 molecule
水 water: 71 分子 molecules
総原子数 Total No. of Atoms: 222

Simulation Conditions (シミュレーション条件)

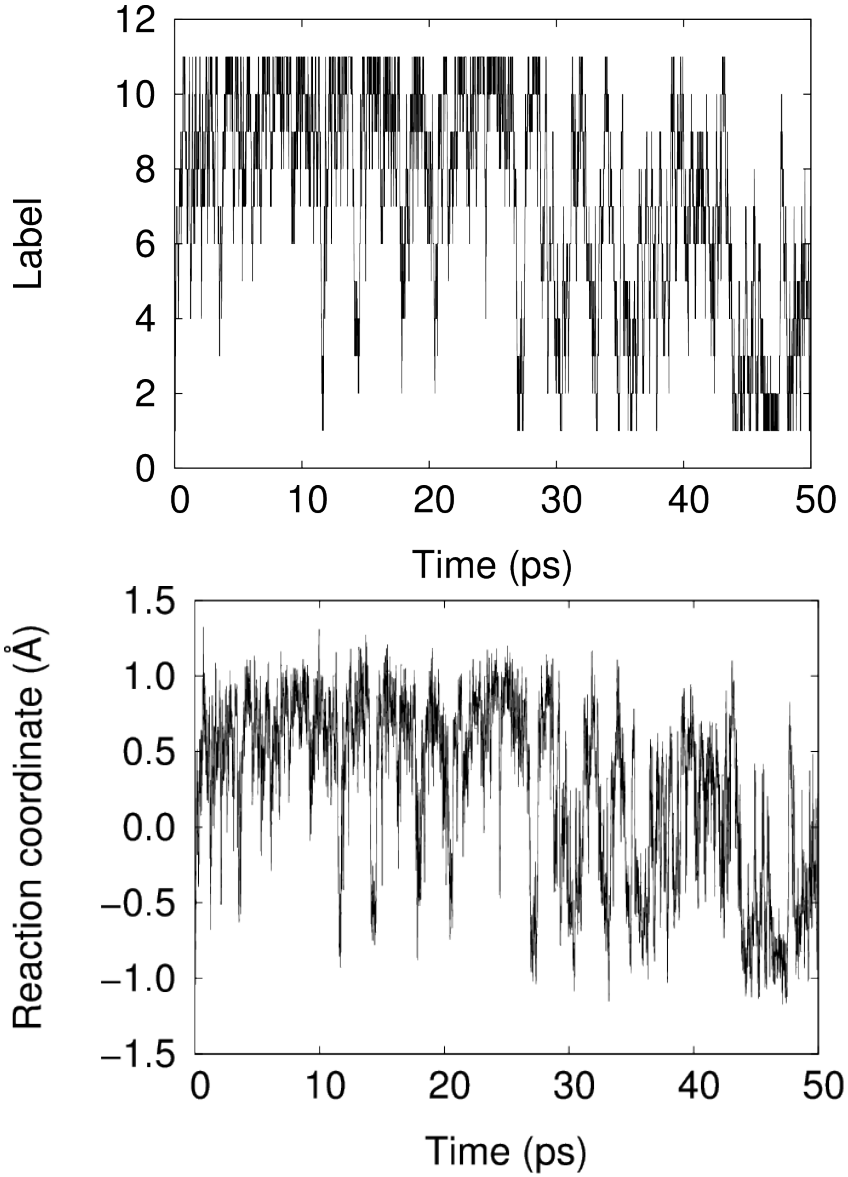
Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).

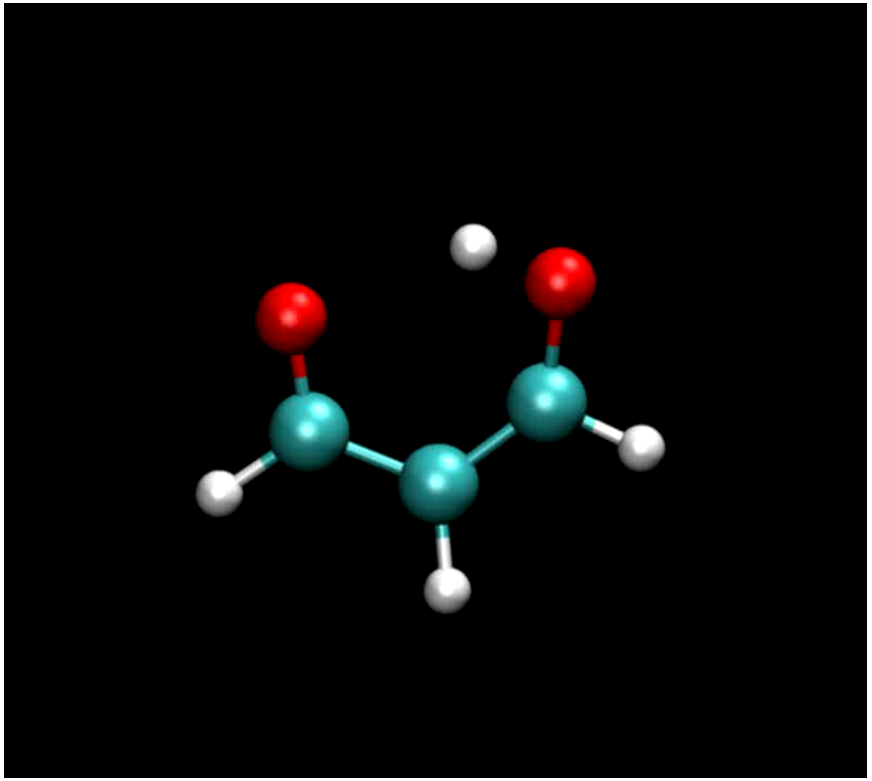
計算条件	
プログラム Program	CP2K (ver. 2.1) *
交換相関汎関数 Functional	BLYP
基底 Bases	Gaussian and plane wave
擬ポテンシャル Pseudo Potential	Goedecker-Teter-Hutter
アンサンブル Statistical Ensemble	*NVT*
温度制御法 Thermostat	Nosé-Hoover chains (# of chains = 3)
温度 Temperature	300 K
体積 Volume	13.82×13.82×13.82 (Å^3)
時間ステップ Time Step	0.5 fs
総シミュレーション時間 Total Sim. Time	200 ps (50 ps × 4 runs)
Umbrella Potentials	
アンブレラポテンシャル(k)	0.01 (hartree・bohr^{-2})
アンブレラポテンシャル(ξ_0) 11個のパラメータを使用	-1.0, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0 (Å)

* http://cp2k.berlios.de/

シミュレーション結果 Results

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).

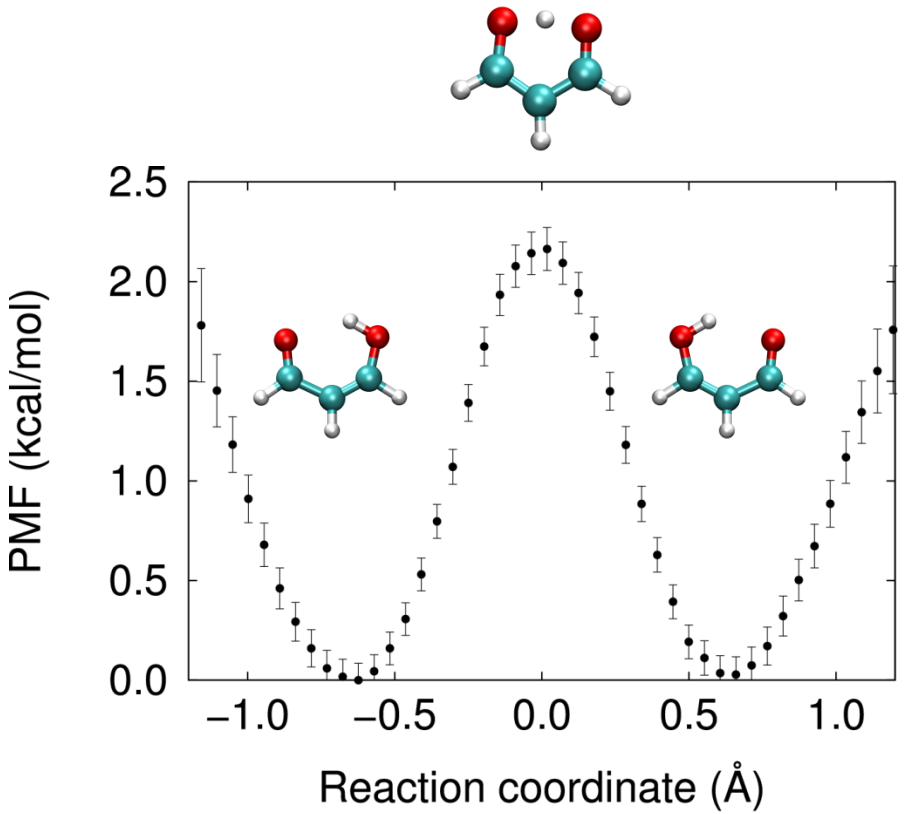




Simulation and movie by Y. Mori

Potential of Mean Force（平均力ポテンシャル）

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).



平均力ポテンシャル
Potential of Mean Force: PMF

$$\mathcal{F}(\xi) \propto -k_{\mathrm{B}}T \ln \left\{ \frac{1}{Q} \int dr \; \delta\left[\xi'(r) - \xi\right] e^{-\beta E(r)} \right\}$$

平均力ポテンシャルを再重法により計算
Reweighting by Multistate Bennett Acceptance Ratio estimator: MBAR*

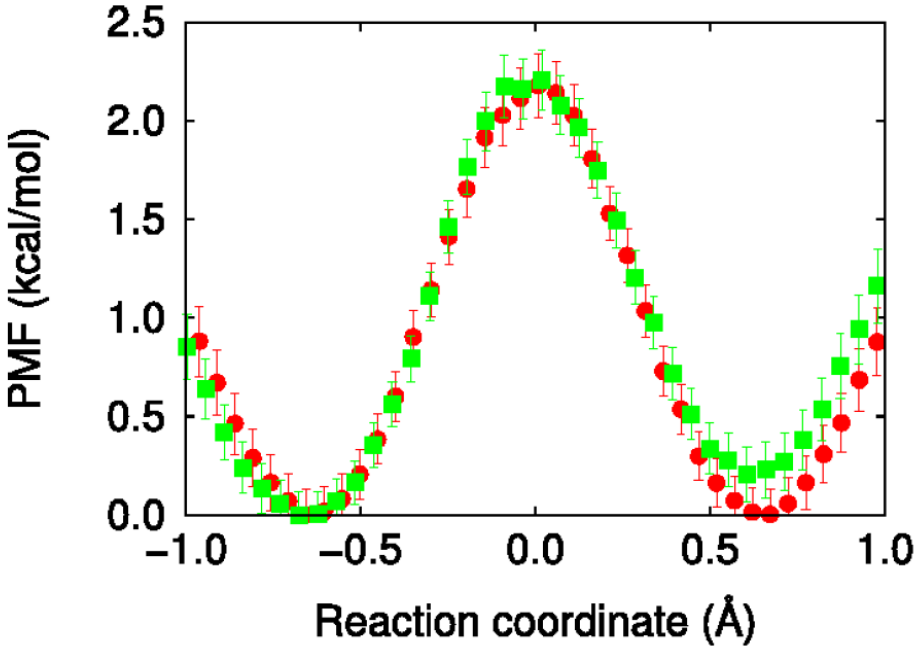
$$\hat{A}_a = \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{A(\boldsymbol{x}_{jn}) \exp[\hat{f}_a - \beta E(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp\{\hat{f}_k - \beta[E(\boldsymbol{x}_{jn}) + U_k(\boldsymbol{x}_{jn})]\}}$$

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-\beta E(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp\{\hat{f}_k - \beta[E(\boldsymbol{x}_{jn}) + U_k(\boldsymbol{x}_{jn})]\}}$$

* M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).

Potential of Mean Force（平均力ポテンシャル）

Y. Mori & Y.O., *Phys. Rev. E* **87**, 023301 (2013).



red：STUS
green：
conventional US

Proton transfer energy barriers for malonaldehyde

Method	Geometry	Barrier (kcal/mol)
MP2	MP2	2.8
CCSD (T)	MP2	3.9
CCSD (T)	VSXC	3.9
BLYP	BLYP	2.0
PBE	PBE	0.9
B3LYP	B3LYP	3.0
B1B95	B1B95	2.8
PBE1PBE	PBE1PBE	2.0
VSXC	VSXC	3.7

S. Sadhukhan, D. Munoz, C. Adamo, & G.E. Scuseria, *Chem. Phys. Lett.* **306** (2008) 83.

2-Dimensional ST Simulation in Isobaric-Isothermal Ensemble

Y. Mori & Y.O., *J. Phys. Soc. Jpn.* **79**, 074003 (2010);
in preparation.

temperature and pressure become dynamical variables.

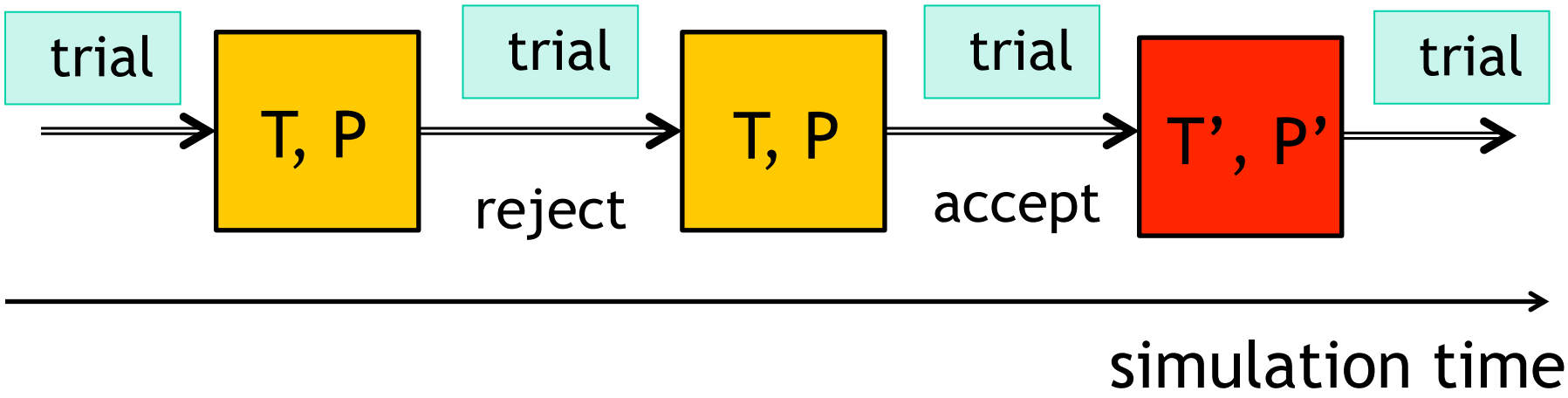
$$\lambda = P = \text{pressure} \qquad V = \text{volume}$$

2-dimensional random walk

in temperature and pressure

$$E_\lambda = E + \lambda V$$

Simulated tempering in the isobaric-isothermal ensemble

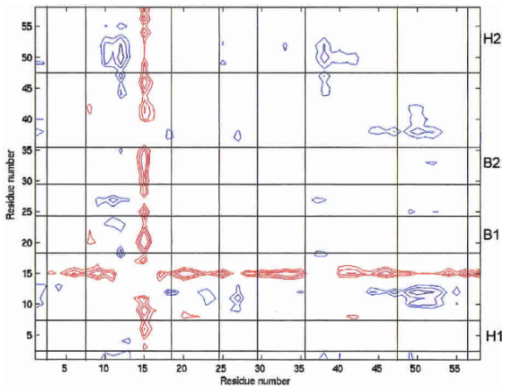


- In addition to temperature T, pressure P is also treated as a dynamical variable.
- The transition probability at every trial is determined so that the detailed balance condition is satisfied.
- Normal constant temperature and pressure simulations are performed between the temperature and/or pressure update trials.

Pressure-Induced Unfolding: NMR Experiments

BPTI

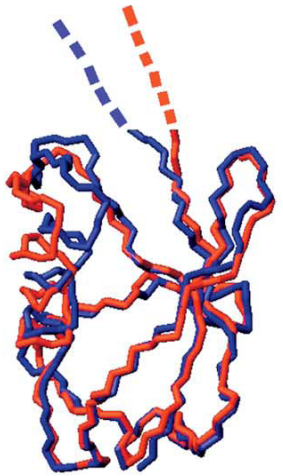
30 bar – 2000 bar

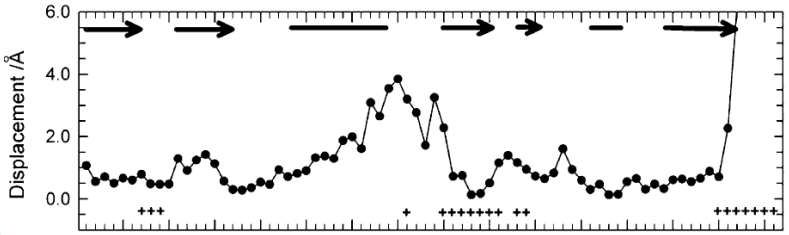


M. P. Williamson *et al*. (2003)

Ubiquitin

30 bar – 3000 bar



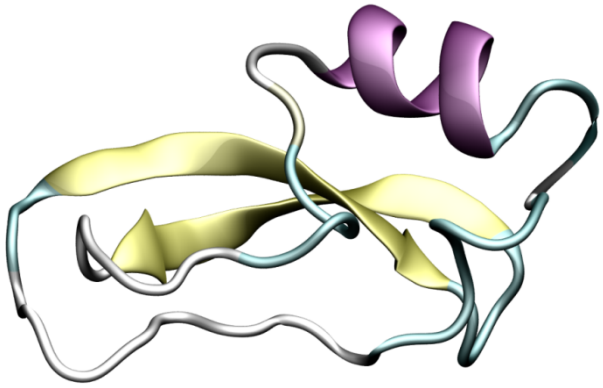


R. Kitahara *et al*. (2005)

Simulated Systems

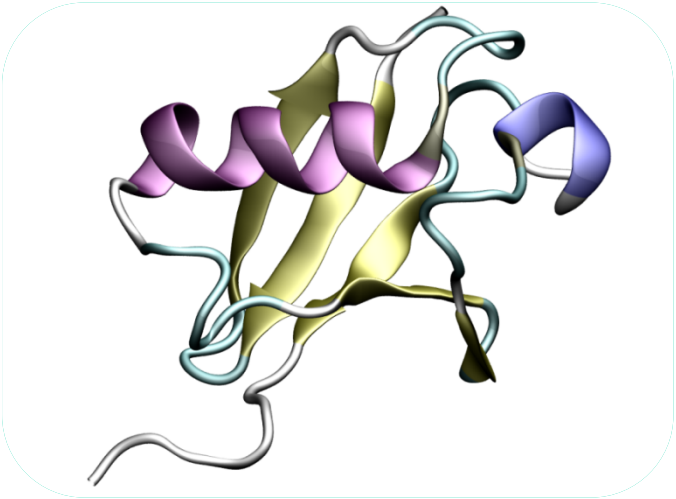
Y. Mori & Y.O., in preparation.

- BPTI
- 58 amino acids
- 6363 water molecules
- 20041 atoms



PDB: 1OA5

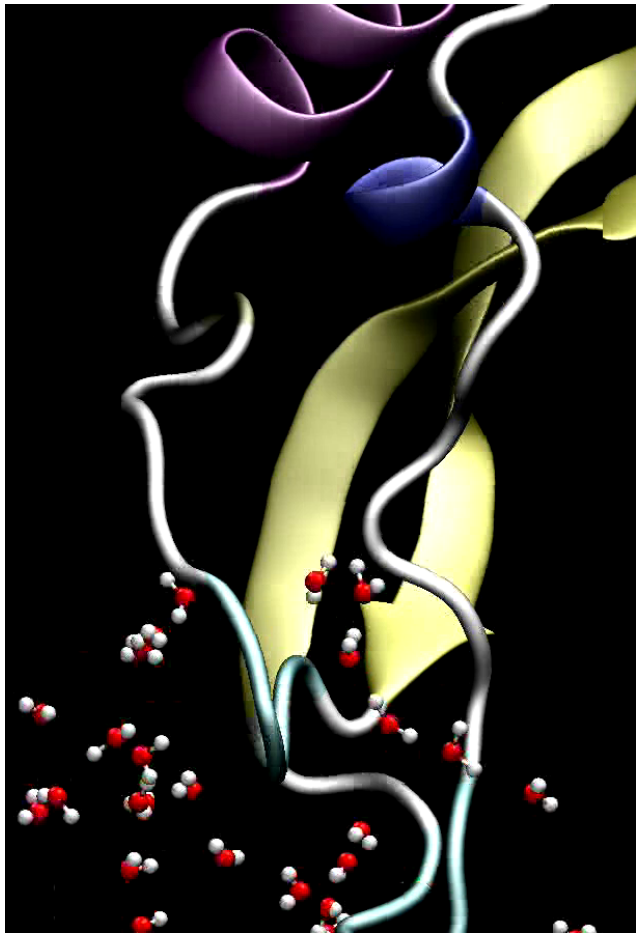
- Ubiquitin
- 76 amino acids
- 6232 water molecules
- 19985 atoms



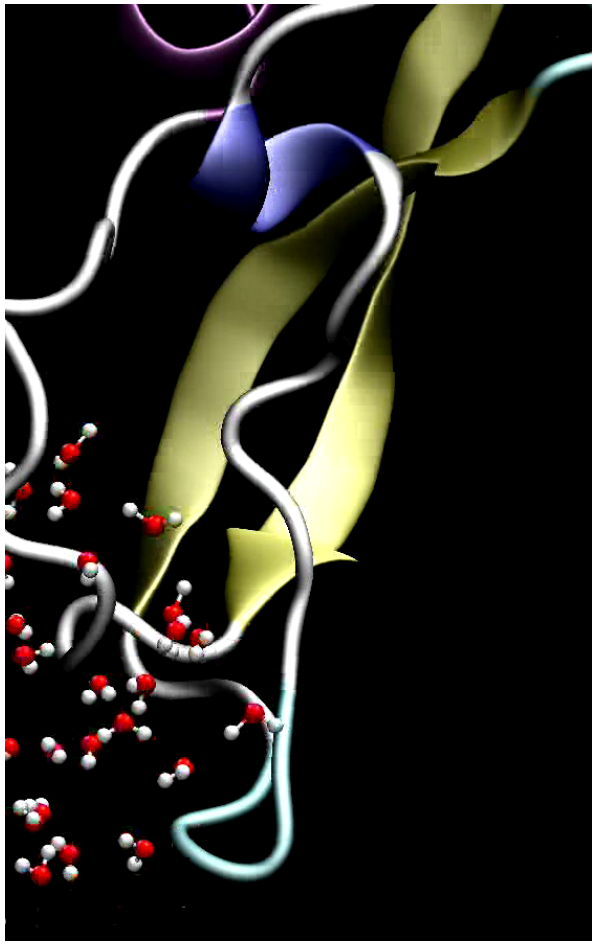
PDB: 1V80

Water Molecules (BPTI)

Low Pressure

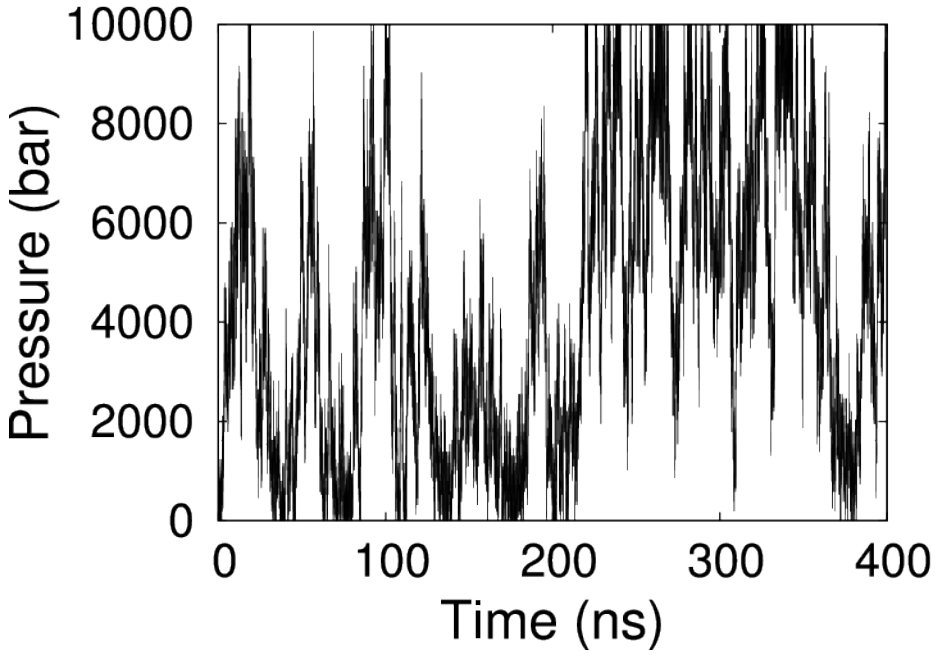


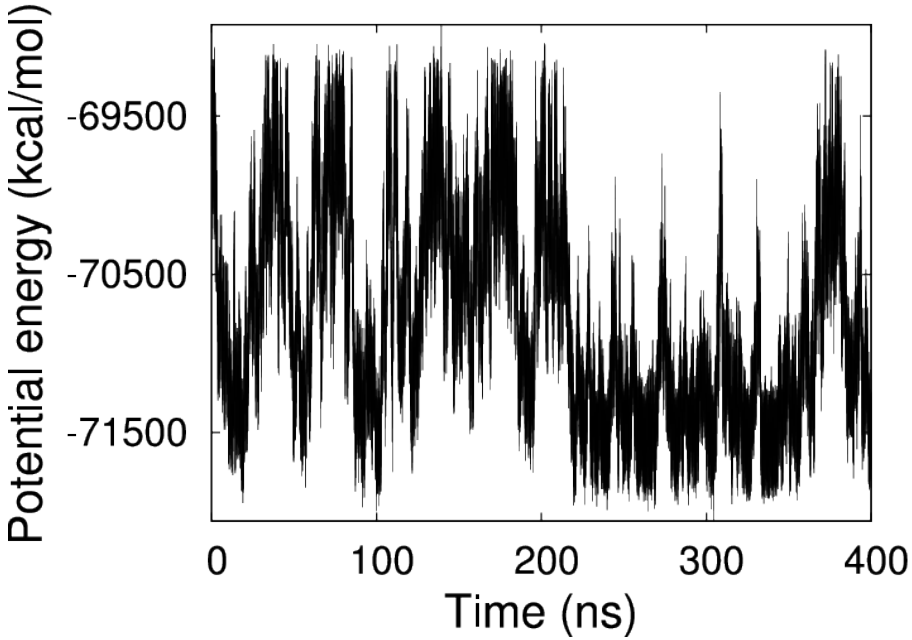
High Pressure

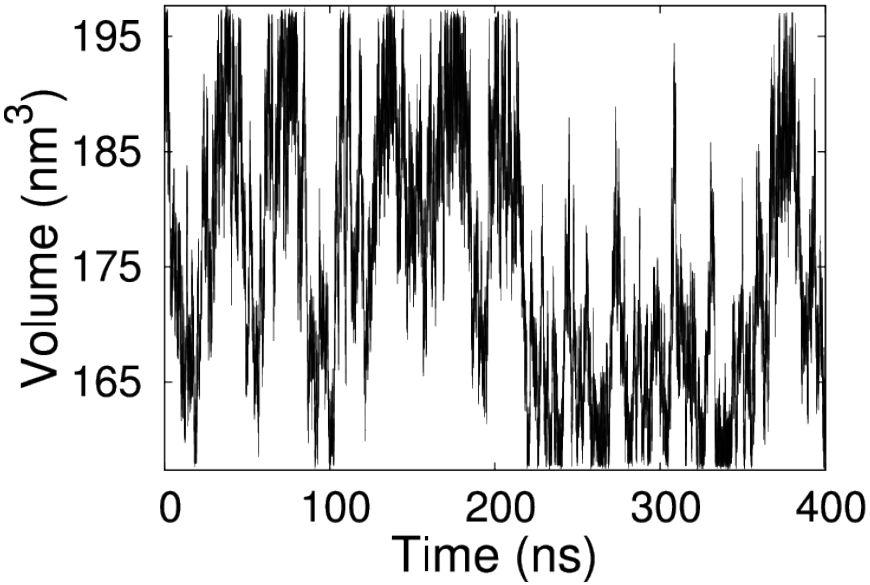


Simulation and movie by Y. Mori

Time series of pressure *P*, potential energy *E*, and volume *V* for ubiquitin





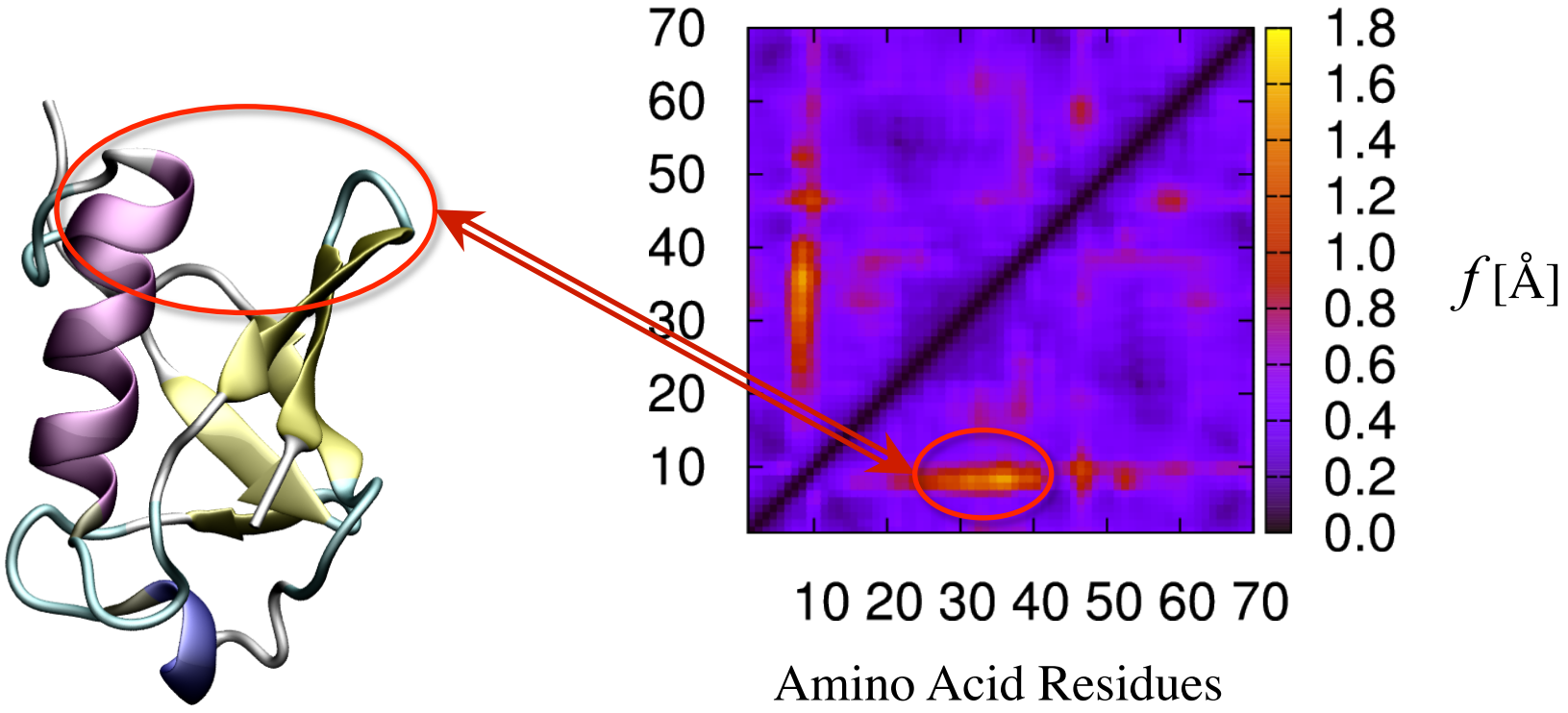


Y. Mori & Y.O., in preparation.

Large structural fluctuations

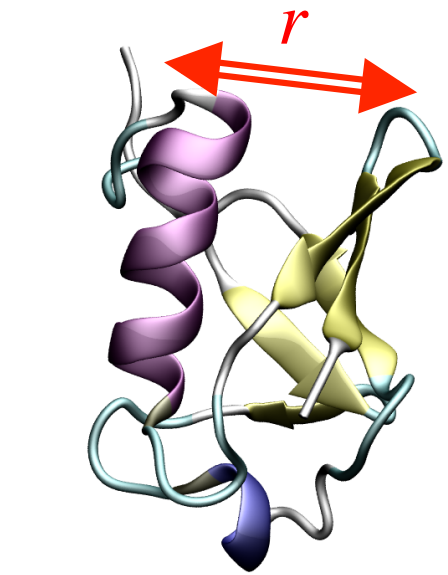
- Fluctuations of distance d between pairs of Cα atoms.

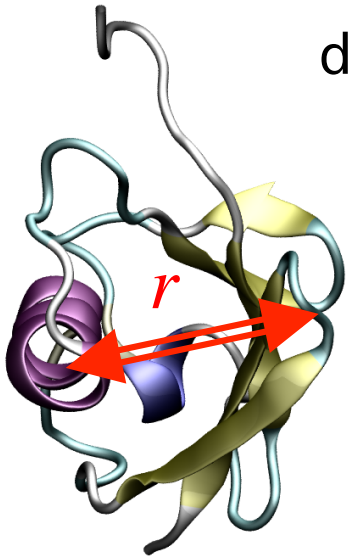
$$f \equiv \sqrt{\langle d^2 \rangle - \langle d \rangle^2}$$



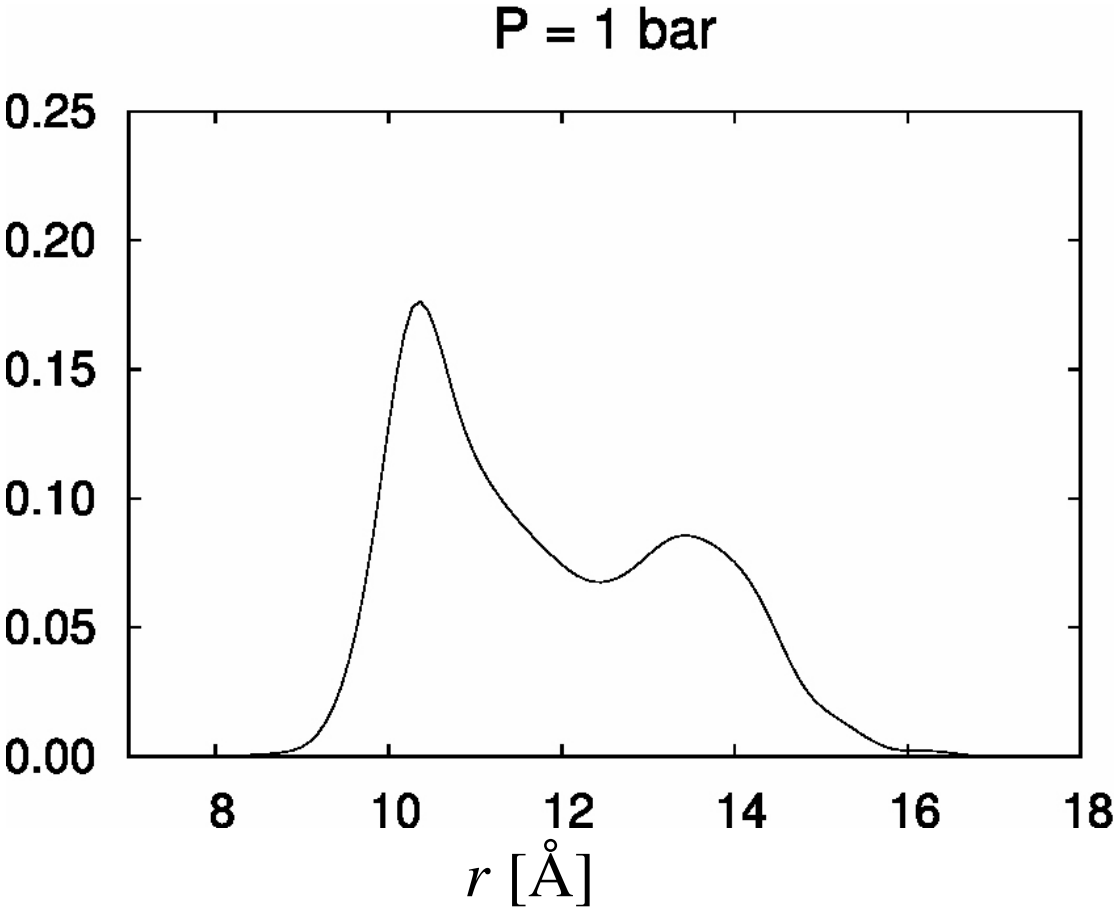
Large fluctuations observed in agreement with experiments.

Structural changes under high pressure

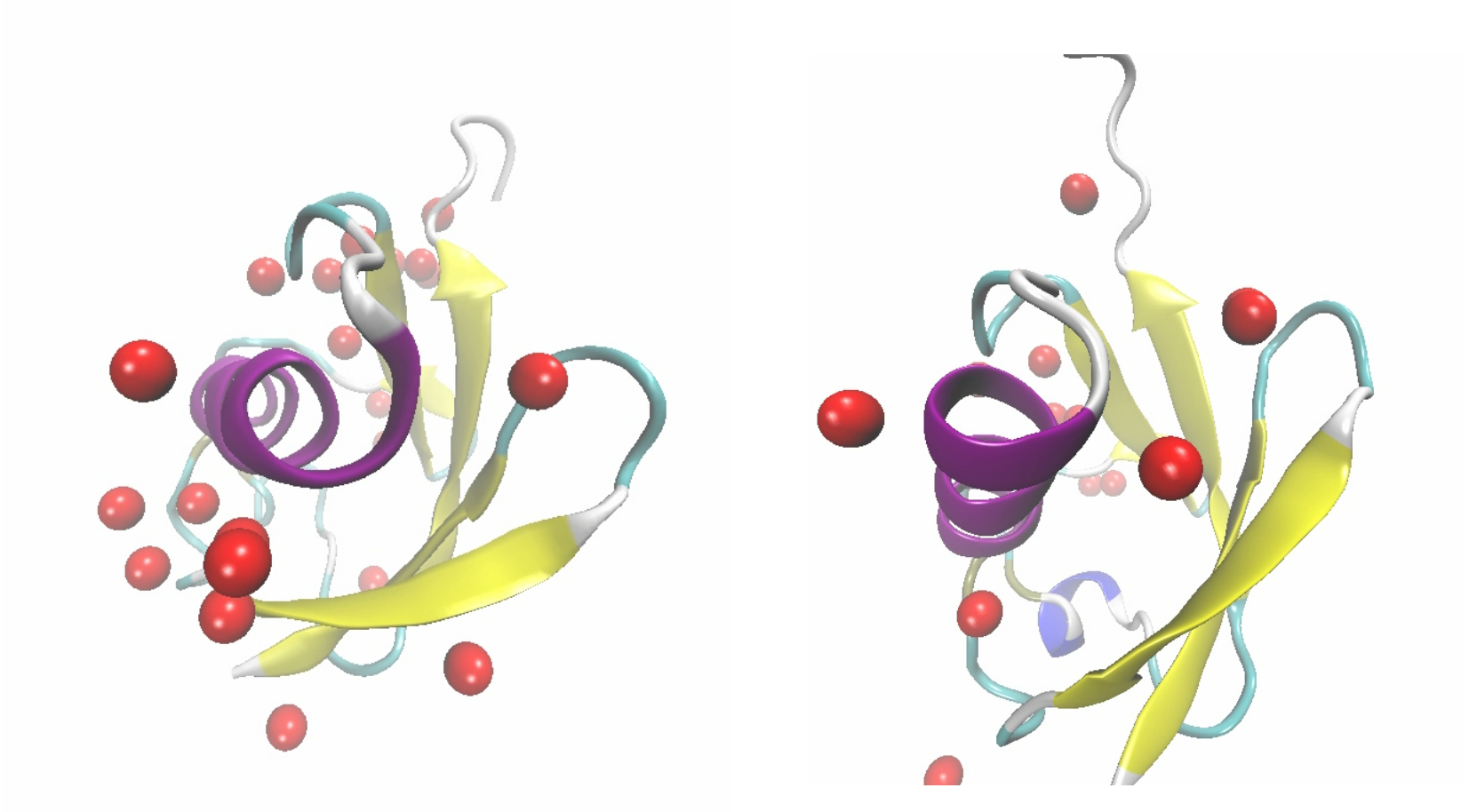




distribution
[Å^{-1}]



Ubiquitin and water molecules



at low pressure

at high pressure

Simulation and movie by Y. Mori

SUMMARY

We have shown that generalized-ensemble algorithms are particularly suitable for biomolecular simulations.

COLLABORATORS

Ayori MITSUTAKE [IMS → Keio Univ.]
Yuji SUGITA [IMS → Univ. Tokyo → RIKEN]
Takao YODA [IMS → Nagahama Inst. Bio-Science]
Masato MASUYA [IMS → Kagoshima Univ.]
Ryo URANO [Nagoya Univ.]
Yoshiharu MORI [Nagoya Univ. → IMS]
Tetsuro NAGAI [Nagoya Univ. → Ritsumeikan Univ.]
Ryuichi UEOKA [Sojo Univ.]
Hironori KOKUBO [IMS → Univ. of Houston → Takeda Pharm.]
Toshimasa TANAKA [Takeda Pharm.]