GENESIS Hands-on Part 3:

Generalized-ensemble simulations using GENESIS

Shingo Ito
RIKEN Center for Computational Science
2024/07/02

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

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

06/30 Part 1	
13:30 – 15:00	GENESIS basics and GENESIS on Fugaku (Kobayashi)
	Lecture
	Hands-on tutorial on Fugaku

07/01 Part 2	
14:30 – 15:30	Coarse-grained simulations in GENESIS (Tan)
15:30 – 16:30	High-performance computation with GENESIS (Jung)

07/02 Part 3	
13:30 – 15:00	**Generalized-ensemble simulations using GENESIS (Ito)**

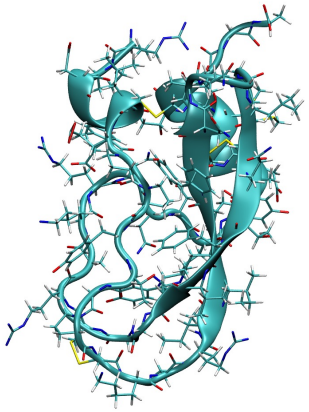
Contents

- What are Generalized-Ensemble simulations?
- Demonstration of Temperature Replica-Exchange MD

Molecular dynamics (MD) simulation

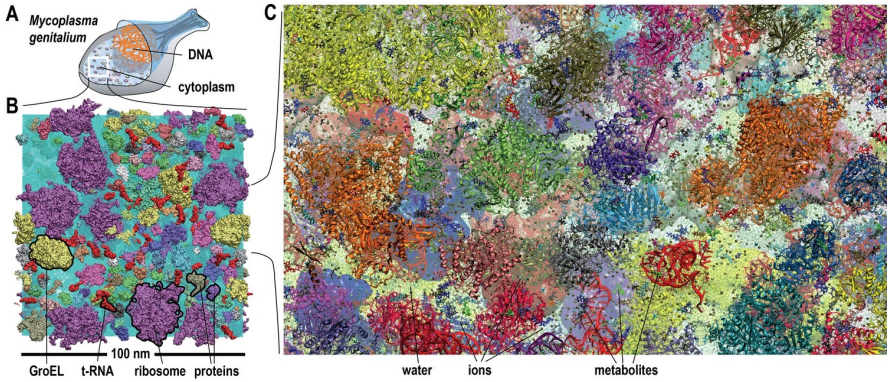
- Molecular dynamics (MD) simulation is useful for understanding biological functions such as protein folding, membrane transport, ligand binding, etc.

PTI (1977)
58 res: < 1000 atoms



40 years later...

Bacterial cytoplasm (2016)
> 100 M atoms

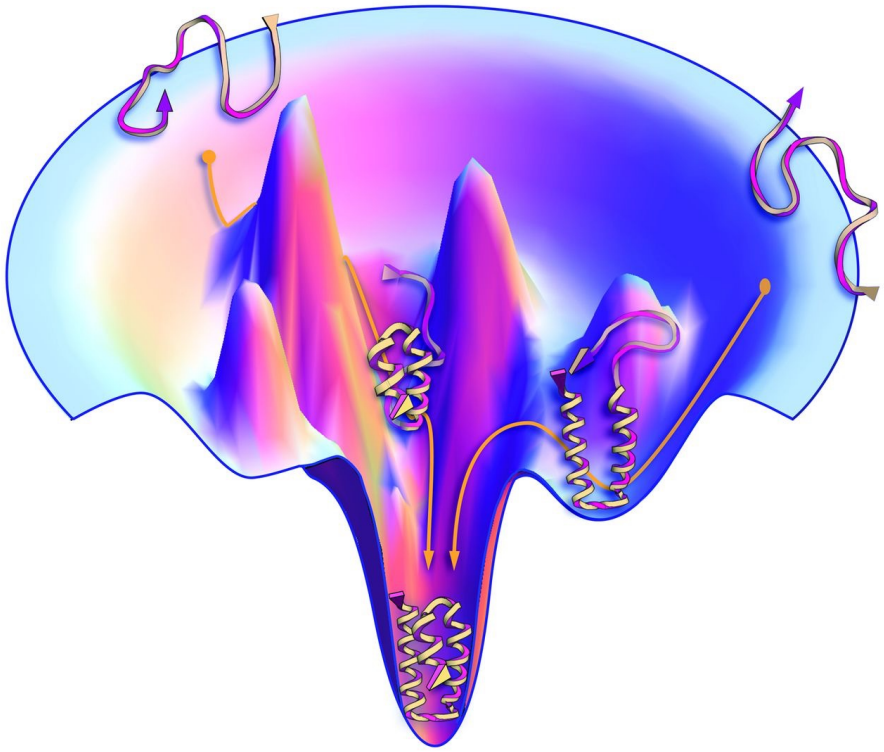


J. A. McCammon, et. al., *Nature* **267**, 585–590 (1977)

I. Yu, et. al., *eLife* **5**, e19274 (2016)

Common issues of MD simulation

Folding Funnel



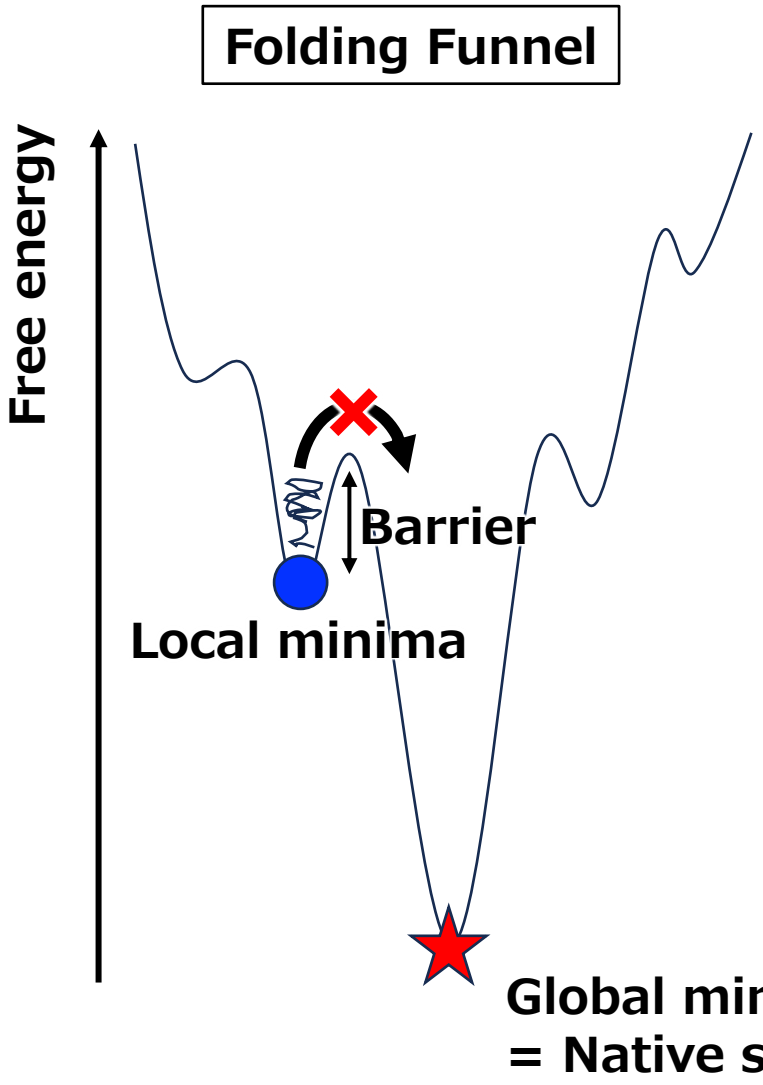
- The folding funnel hypothesis suggests that protein conformation reaches the global minima (= thermal equilibrium) after infinite-time scale MD simulations.

⬇ But...

- We can perform only finite-time (ns–μs) scale MD simulations.

K. E. Dill, et al., *Science.* **338**, 1042-1046 (2012)

Common issues of MD simulation



- Sometimes, finite time-scale MD simulations get trapped at one of the local minima.
- The sampling in the limited conformational space does not satisfy the ergodicity.

⬇

- We can not estimate physical quantities.

How can the finite time-scale MD simulation overcome the barrier?

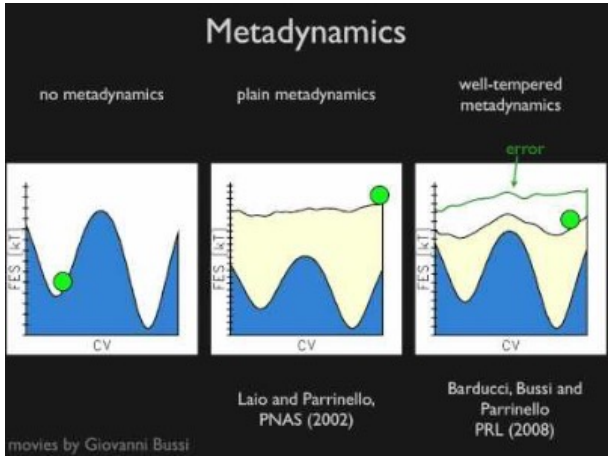
Generalized-ensemble simulations

- Generalized-ensemble simulations is one of the solutions

  = Conventional MD + enhanced sampling method

- Generally, there are two types of enhanced sampling methods

1. Using thermal fluctuations at high temperatures (> 300.0 K)
   - Simulated tempering
   - Temperature replica exchange
2. Smoothing the potential/free energy surface using an additional potential
   - Umbrella sampling
   - Metadynamics



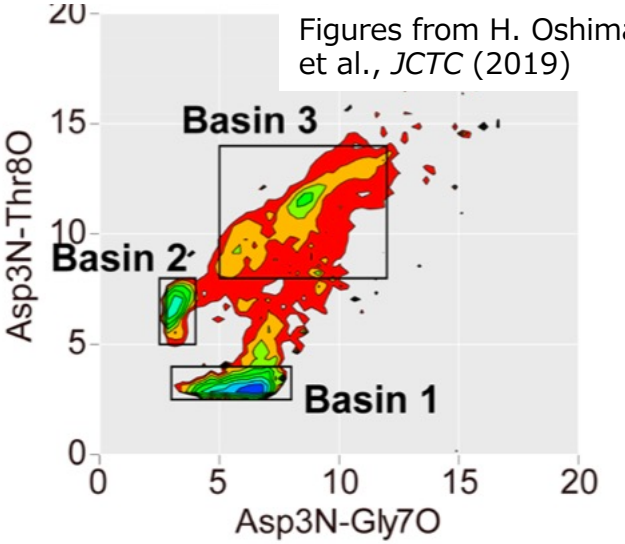
YouTube movie by Giovanni Bussi
https://www.youtube.com/watch?v=IzEBpQ0c8TA

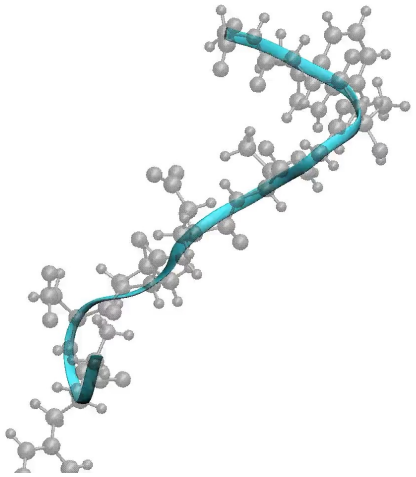
Simulated Tempering MD

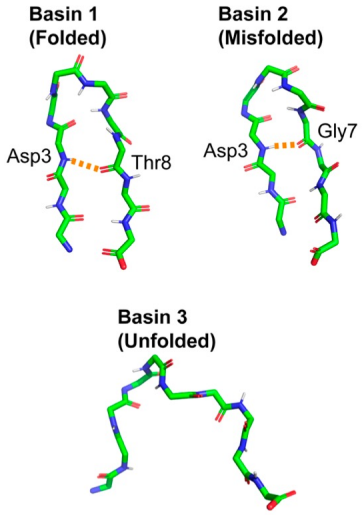
- In the Simulated Tempering (ST) MD simulation, the temperature changes from T_m to $T_{m\pm1}$.
- The protein, RNA/DNA etc., become more flexible at the high temperature.

Folding simulation of Chignolin

Figures from H. Oshima, et al., *JCTC* (2019)

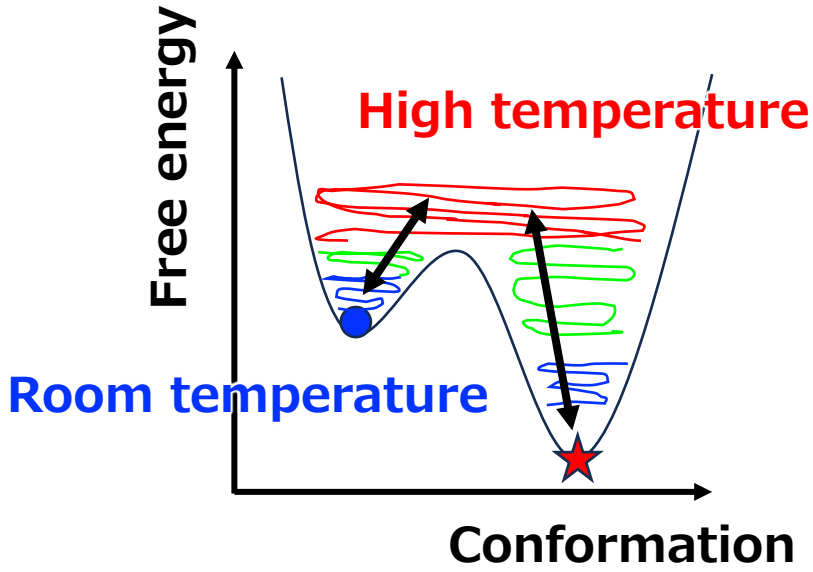






A. Irbäck, et al., *J. Chem. Phys.* **103**, 10298 (1995)
U.H.E. Hansmann, et al., *J. Comput. Chem*. **18**, 920 (1997)

Simulated Tempering MD



- Simulated tempering (ST) performs a random walk in temperature space.

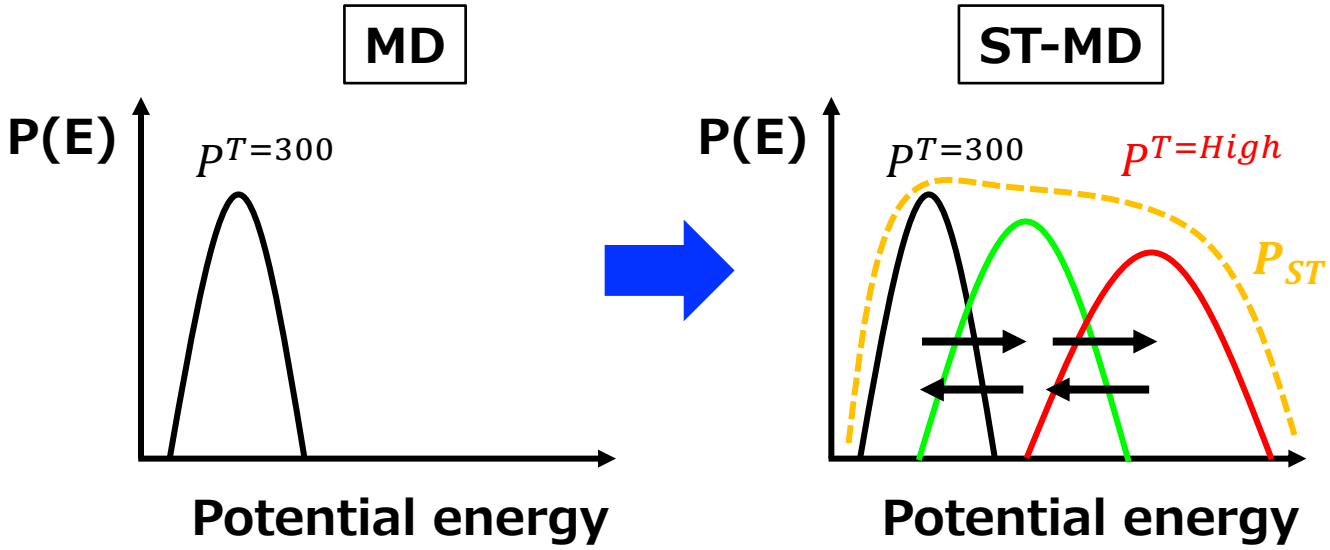
$$W_{canonical}(E) = \exp(-\beta E)$$

discretization

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

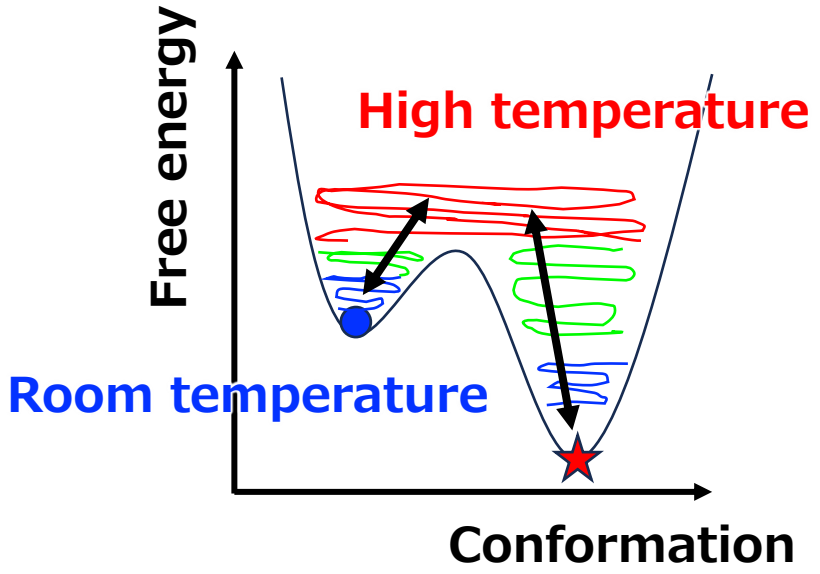
$$W_{ST}(E;T_m) = \exp(-\beta_m E + \alpha(T_m))$$

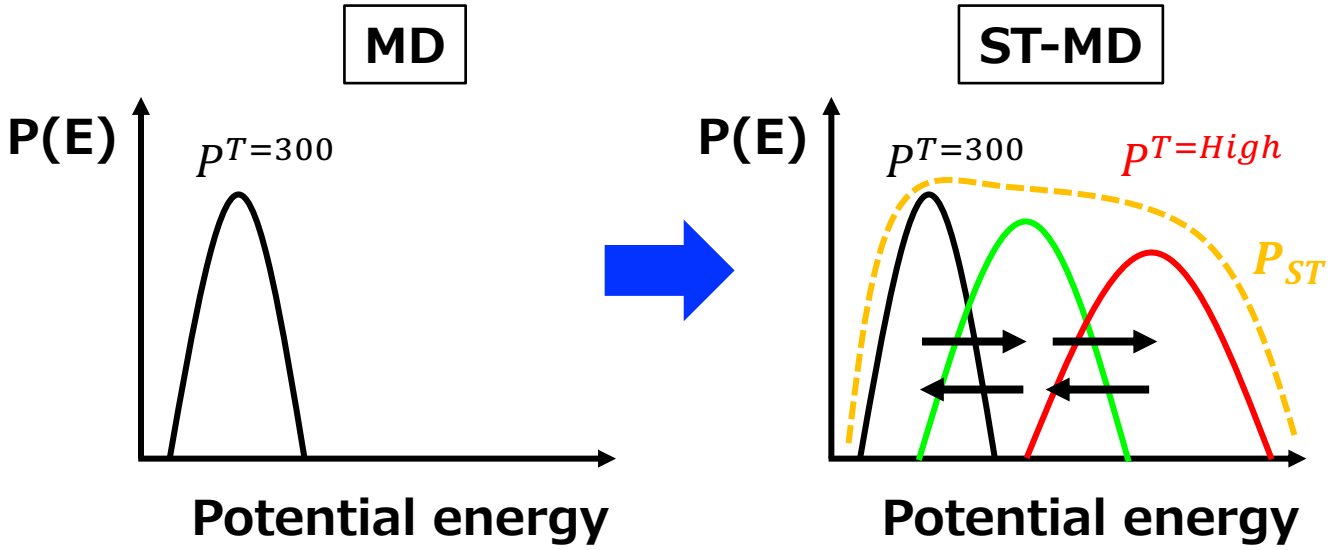
$$P_{ST}(T) = \int dE\, n(E) W_{ST}(E;T) = \text{constant}$$



A. Irbäck, et al., *J. Chem. Phys.* **103**, 10298 (1995)
U.H.E. Hansmann, et al., *J. Comput. Chem*. **18**, 920 (1997)

Simulated Tempering MD





- This is a procedure of the ST-MD simulation.
- α_m is decided by the short MD simulations.
  1. Perform MD at T_m $(m = 1, \ldots, M)$
  2. Change T_m to $T_{m\pm1}$ each fixed MD step with following probability, *w*.

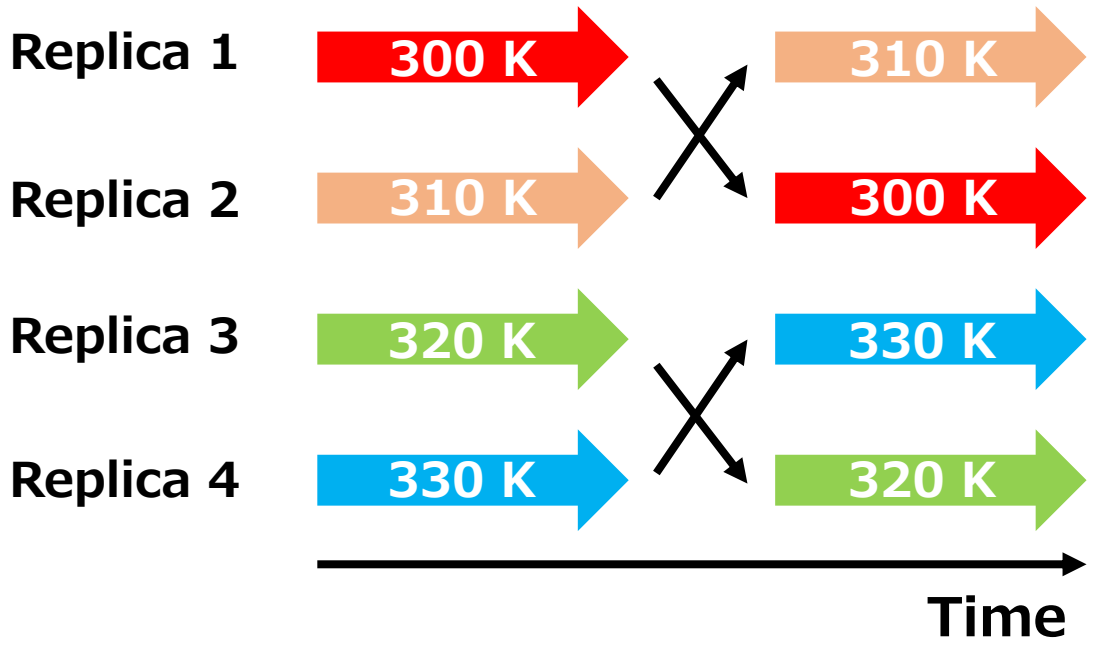
$$w(T_m \to T_{m\pm1}) = \begin{cases} 1, & \text{if } \Delta \leq 0 \\ \exp(-\Delta), & \text{if } \Delta > 0 \end{cases}$$

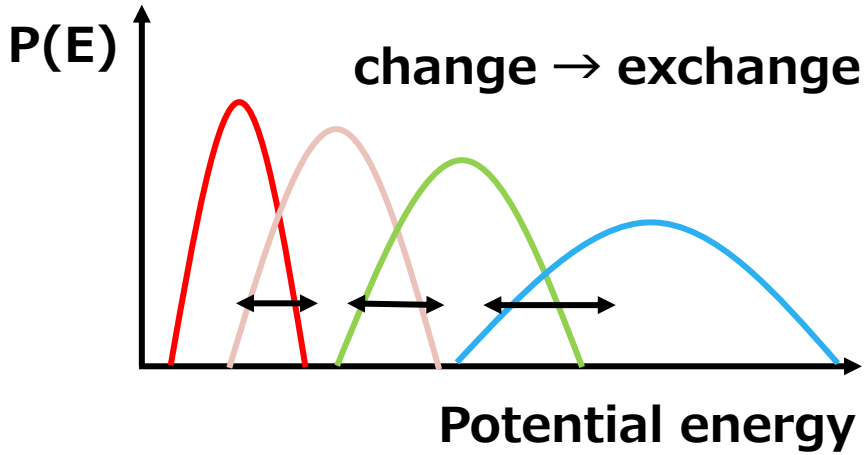
$$\Delta = (\beta_{m\pm1} - \beta_m)E - (\alpha_{m\pm1} - \alpha_m), \quad \beta_m = {}^{1}/_{k_B T_m}$$

- It is difficult to define hyperparameter, α_m, for large/complicated system.

A. Irbäck, et al., *J. Chem. Phys.* **103**, 10298 (1995)
U.H.E. Hansmann, et al., *J. Comput. Chem*. **18**, 920 (1997)

Temperature replica-exchange MD (T-REMD)





- T-REMD is a parallelization model of simulated tempering (ST) MD (= parallel tempering MD).
- T-REMD consists of *M* non-interacting copies ($i = 1, \ldots, M$) at *M* different temperatures, T_m $(m = 1, \ldots, M)$.

$$\begin{cases} i = i(m) \equiv f(m) \\ m = m(i) \equiv f^{-1}(i) \end{cases}$$

f: permutation function

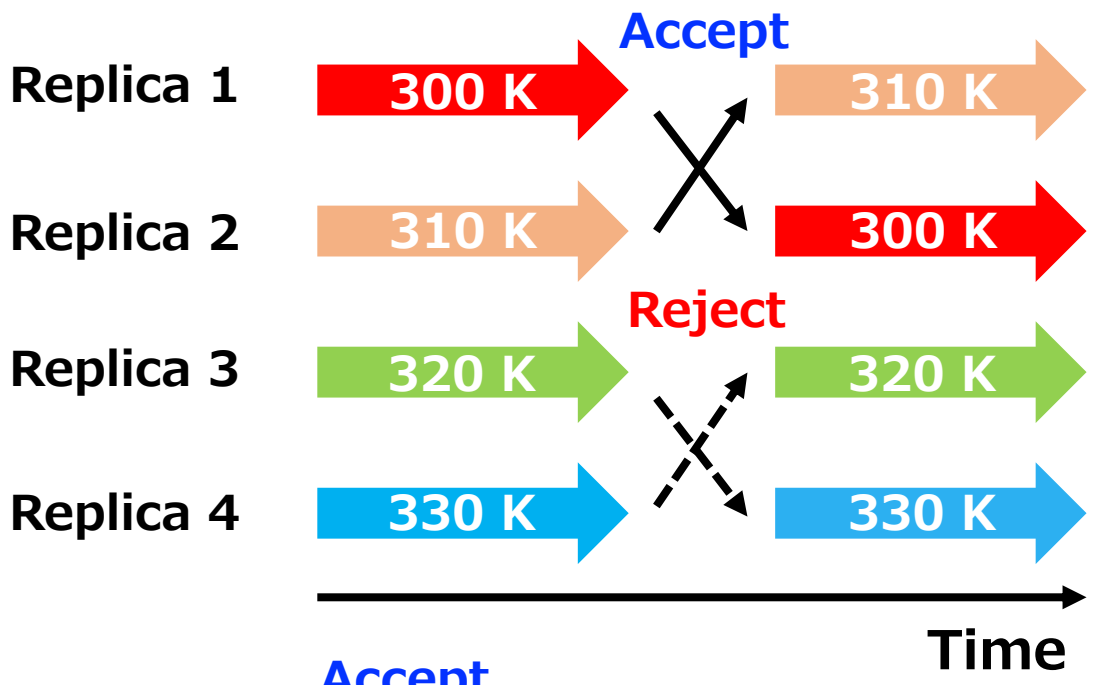
- The weight factor for the state *X*, *W(X)*, is given by

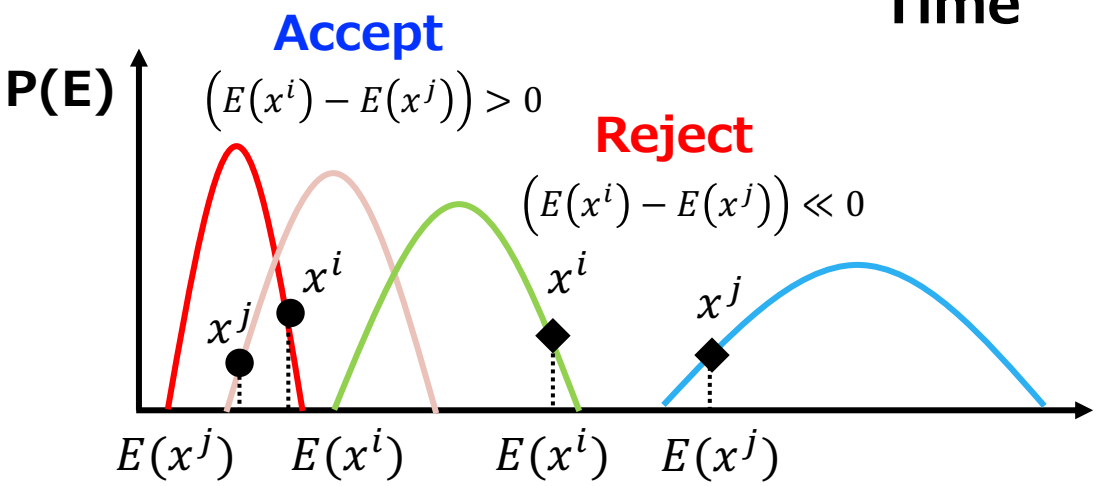
$$W_{REM}(x) = \exp\left\{-\sum_{m=1}^{M} \beta_m H(q^{[i(m)]}, p^{[i(m)]})\right\}$$

H: Hamiltonian

K. Hukushima, et al., *J. Phys. Soc. Jpn*. **65**, 1604 (1996)
Y. Sugita, et al., *Chem. Phys. Lett*. **314**, 141 (1999).

Temperature replica-exchange MD (T-REMD)





- Every fixed number of MD steps, a pair of replicas, *i* and *j*, corresponding to neighboring temperatures, T_m, and T_n ($T_m < T_n$), are exchanged.

$$X = \left(\dots, x_m^{[i]}, \dots, x_n^{[j]}\right) \rightarrow X' = \left(\dots, x_m^{[j]'}, \dots, x_n^{[i]'}\right)$$

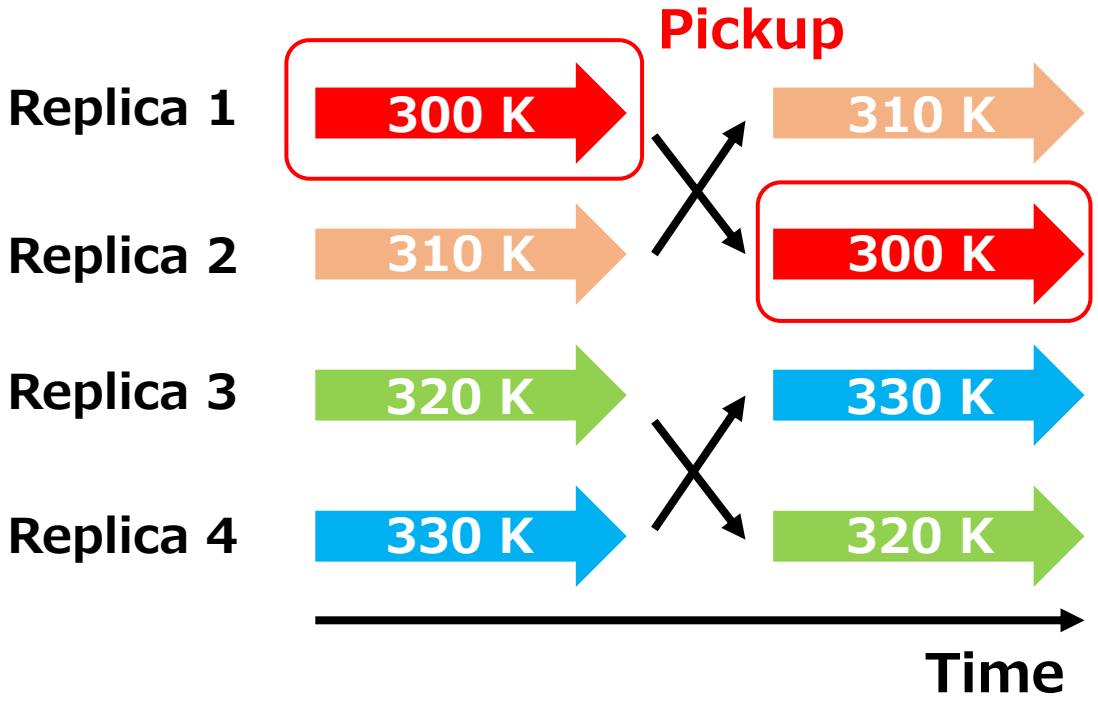
- According to the Metropolis criteria, satisfying the detailed balance, the replica-exchange is “accepted” or “rejected”.

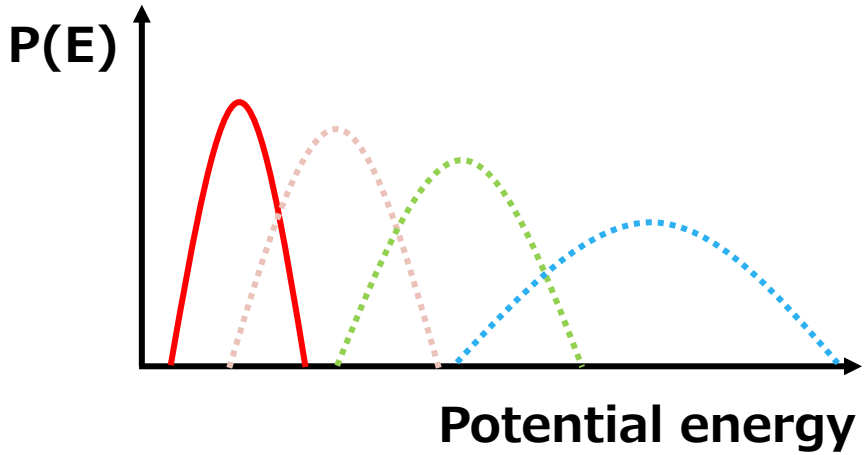
$$w(X \rightarrow X') = \begin{cases} 1, & \text{if } \Delta \leq 0 \\ \exp(-\Delta), & \text{if } \Delta > 0 \end{cases}$$

$$\Delta = (\beta_n - \beta_m)(E(x^i) - E(x^j))$$

- Succeeding replica exchanges must have overlapping distributions.

Estimating unbiased physical quantities





- Physical quantities are obtained by following procedure.
  1. Collect data at “300 K” from each replica
  2. Estimate time averages at 300 K by following

$$\langle A \rangle_{time} = \sum_{t=1}^{N_{\mathrm{sample}}} A_{T=300K}(t)$$

- It means that we waste all the data except for the data at T=300 K.

Wasteful!!

Multistate Bennett Acceptance Ratio (MBAR)

M. R. Shirts, et al., *J. Chem. Phys.* **129**, 124105 (2008)

- MBAR can estimate dimensionless free energy differences by the following equation.

$$\widehat{f}_i = \ln \sum_{j=1}^{M} \sum_{n=1}^{N_j} \frac{\exp\left[-\beta_i U(x_{jn})\right]}{\sum_{k=1}^{M} N_k \exp\left[\widehat{f}_k - \beta_k U(x_{jn})\right]}$$

- Once $\widehat{f}_i$ is obtained by using MBAR equation, average of physical quantities at target temperature, $\langle A \rangle_{\mathrm{target}}$, can be derived as follows,

$$\hat{f}_{\mathrm{target}} = \ln \sum_{j=1}^{M} \sum_{n=1}^{N_j} \frac{\exp\left[-U_{\mathrm{target}}(x_{jn})\right]}{\sum_{k=1}^{M} N_k \exp\left[\widehat{f}_k - U_k(x_{jn})\right]}$$

$$\hat{f}_{\mathrm{target,\,A}} = \ln \sum_{j=1}^{M} \sum_{n=1}^{N_j} \frac{\mathrm{A}(x_{jn}) \exp\left[-U_{\mathrm{target}}(x_{jn})\right]}{\sum_{k=1}^{M} N_k \exp\left[\widehat{f}_k - U_k(x_{jn})\right]}$$

$$\langle A \rangle_{\mathrm{target}} = \frac{\exp\left[-f_{\mathrm{target,\,A}}\right]}{\exp\left[-f_{\mathrm{target}}\right]} \sim \frac{\exp\left[-\hat{f}_{\mathrm{target,\,A}}\right]}{\exp\left[-\hat{f}_{\mathrm{target}}\right]}$$

$f_{i/k}$: Dimensionless free energy of temperature label i/k
M : Total # of replicas
$N_{j/k}$: Total # of snapshot of temperature label j/k
U : Potential energy
x : Coordinate
A : Physical quantities

Multistate Bennett Acceptance Ratio (MBAR)

- $\hat{f}_i$ is obtained by a simple self-consistent iteration as follows.

$$\hat{f}_i = \ln \sum_{k=1}^{M} \sum_{n=1}^{N_k} \frac{\exp[-\beta_i U(x_{kn})]}{\sum_{l=1}^{M+1} N_l \exp\left[\hat{f}_l - \beta_l U(x_{kn})\right]}$$

unknow value

1. Set initial (t=0) $\hat{f}_i^{t=0}$ to 0.0 from i=1 to M
2. Calculate each $\hat{f}_i^{t+1}$ by using MBAR equation
3. Check $\Delta\hat{f}_i = \hat{f}_i^{t+1} - \hat{f}_i^{t}$
   - If $\Delta\hat{f}_i >$ threshold ($\sim 1.0\times10^{-8}$), go back step 2
   - If $\Delta\hat{f}_i \leq$ threshold, go to step 4
4. Estimate weight factors at each snapshot and temperature

Estimating free energy differences

- Weight factor, w_{kn}, is obtained as follows.

$$\langle A \rangle_{\text{target}} = \frac{\exp\left[-f_{\text{target, A}}\right]}{\exp\left[-f_{\text{target}}\right]} \sim \frac{\exp\left[-\hat{f}_{\text{target, A}}\right]}{\exp\left[-\hat{f}_{\text{target}}\right]} = \sum_{k=1}^{M} \sum_{n}^{N_k} w_{kn} A(x_{kn})$$

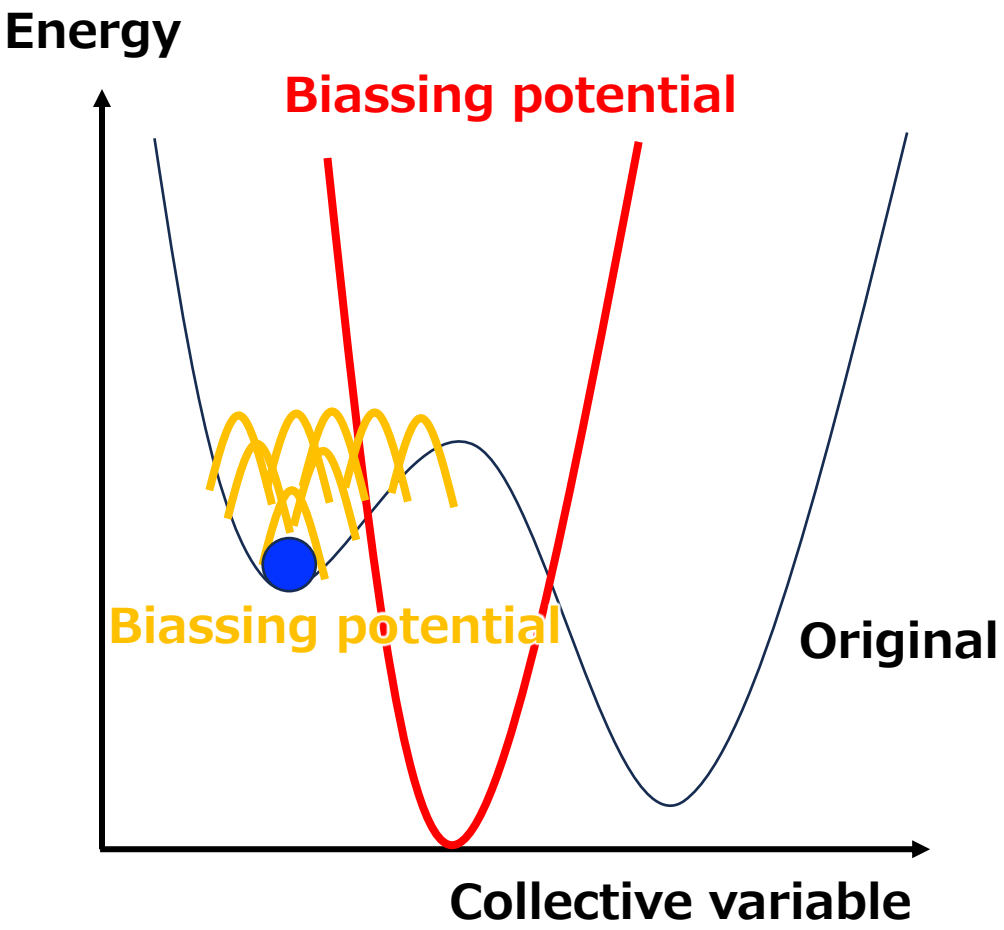
- Since w_{kn} is independent of *A*, it is available for estimating other quantities, *B*, *C*, etc.
- We can estimate the free energy differences along a reaction coordinate, ξ, named “potential of mean force (PMF)”, $F(\xi)$, using w_{kn}.

$$P_{target}(x_{kn}) = w_{kn} P_{biased}(x_{kn})$$

$$F(\xi) = -k_B T_{target} \ln P_{target}(\xi) = -k_B T_{target} \ln w P_{biased}(\xi)$$

Appendix

Potential smoothing



- Potential smoothing is useful if we already know which collective variables (CVs) are important.
  - CVs are quantities that describe the behavior or state of a system.

1. Approach 1: Decrease the potential energy barrier
   - Multicanonical
   - Umbrella sampling
2. Approach 2: Fill the free energy surface
   - Metadynamics