Free energy calculation in MD
Enhanced sampling

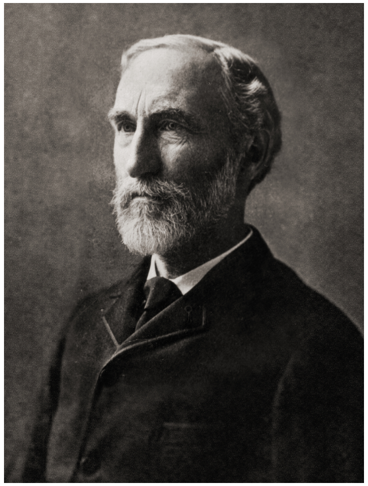
Yongle Li

2020

Free energy calculation: Meaning

$$G(N,P,T) = U - TS + pV$$

Josiah Willard Gibbs (February 11, 1839 – April 28, 1903)



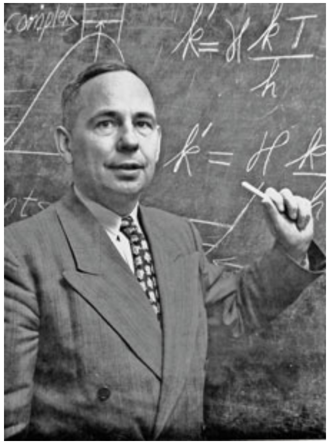
$$F(N,V,T) = U - TS$$

Hermann Ludwig Ferdinand von Helmholtz (August 31, 1821 – September 8, 1894)

Free energy calculation: meaning

$$k(T) \equiv \kappa\left(\frac{k_{\mathrm{B}}T}{h}\right)\exp\left(-\frac{\Delta G}{k_{\mathrm{B}}T}\right)$$

Transition state theory



Henry Eyring
(February 20, 1901 -
December 26, 1981)

Measurement of free energy：

1. ratio of probability

$$F = -k_{\mathrm{B}}T\ln\frac{p_{\mathrm{A}}}{p_{\mathrm{B}}}$$

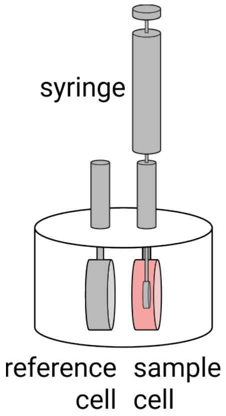
2. reversible work

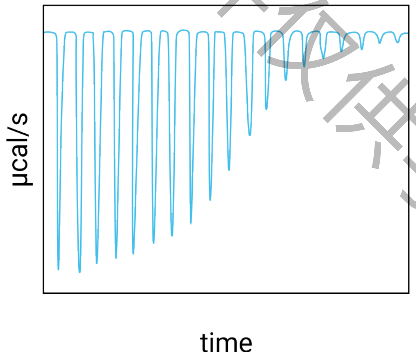
$$\Delta G \approx W^{\mathrm{reversible}}$$

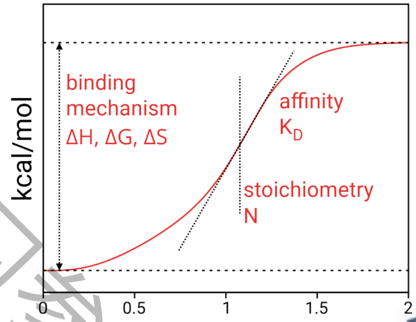
Free energy: measurement

X-ray crystallography
Neutron Scattering
fluorescence spectroscopy
Hydrogen-deuterium exchange mass spectrometry (HX-MS)
time-resolved HX-MS
surface plasmon resonance (SPR)
Terahertz spectroscopy
NMR
differential scanning calorimetry (DSC)
isothermal titration calorimetry
isothermal titration calorimetry (ITC)

Free energy: measurement







ITC: Isothermal Titrational Calorimetry

MicroCal

Free energy calculation: difficulty

$$F = -k_{\mathrm{B}} T \ln Q$$

$$Q = \int d\Gamma \exp\left(-\beta U(\mathbf{r}^N)\right)$$

$$F = k_{\mathrm{B}} T \ln\left(\frac{1}{Q}\right) \boxed{+C}$$

An analytical term, ignorable

$$= k_{\mathrm{B}} T \ln\left(\frac{1}{\int d\Gamma \exp\left(-\beta U(\mathbf{r}^N)\right)}\right)$$

$$= k_{\mathrm{B}} T \ln\left(\frac{\int d\Gamma \exp\left(+\beta U(\mathbf{r}^N)\right) \exp\left(-\beta U(\mathbf{r}^N)\right)}{\int d\Gamma \exp\left(-\beta U(\mathbf{r}^N)\right)}\right)$$

$$= k_{\mathrm{B}} T \ln\left\langle \exp\left(+\beta U(\mathbf{r}^N)\right)\right\rangle$$

Here *V* denotes **volume**, and *U* denotes **potential energy**.

$$Z(N,V,T)=\frac{1}{N!h^{3N}}\int e^{-\beta H}\,\mathrm{d}\Gamma$$

$$=\frac{1}{N!h^{3N}}\int \exp\left[-\beta\left(\sum_i \frac{p_i^2}{2m}\right)-\beta U(\vec{r}_1,\vec{r}_2,\cdots,\vec{r}_N)\right]\mathrm{d}\Gamma$$

$$=\frac{1}{N!\lambda^{3N}}\boxed{\int \exp\left[-\beta U(\vec{r}_1,\vec{r}_2,\cdots,\vec{r}_N)\right]\mathrm{d}^{3N}r}$$

$$=\frac{1}{N!\lambda^{3N}}Q$$

Configurational integral

$$\lambda=\sqrt{\frac{h^2}{2\pi m k_\mathrm{B}T}}=\sqrt{\frac{2\pi\hbar^2}{mk_\mathrm{B}T}}$$

de Broglie thermal wavelength

https://www.natur.cuni.cz/chemie/fyzchem/nachtig/prednaskst/Lecture_8.pdf

http://puccini.che.pitt.edu/~karlj/Classes/CHE2101/l14.pdf

Three major frameworks

- Core idea:

$$\Delta F = -k_\mathrm{B}T\ln\frac{Q_\mathrm{final}}{Q_\mathrm{strart}}$$

$$= -k_\mathrm{B}T\ln\frac{Q_1}{Q_0}$$

1. Use different parameters to describe states, such as temperature
2. Use different Hamiltonians: H_1, H_2
3. Introduce a parameter, and construct a parameterized Hamiltonian: $H(\lambda)$, leaving: $H(\lambda=0)=H_1$, $H(\lambda=1)=H_2$

Chapter 1

FREE ENERGY PERTURBATION

FEP(1)

$$\Delta F = -k_B T \ln \frac{Q_{\text{final}}}{Q_{\text{strart}}}$$

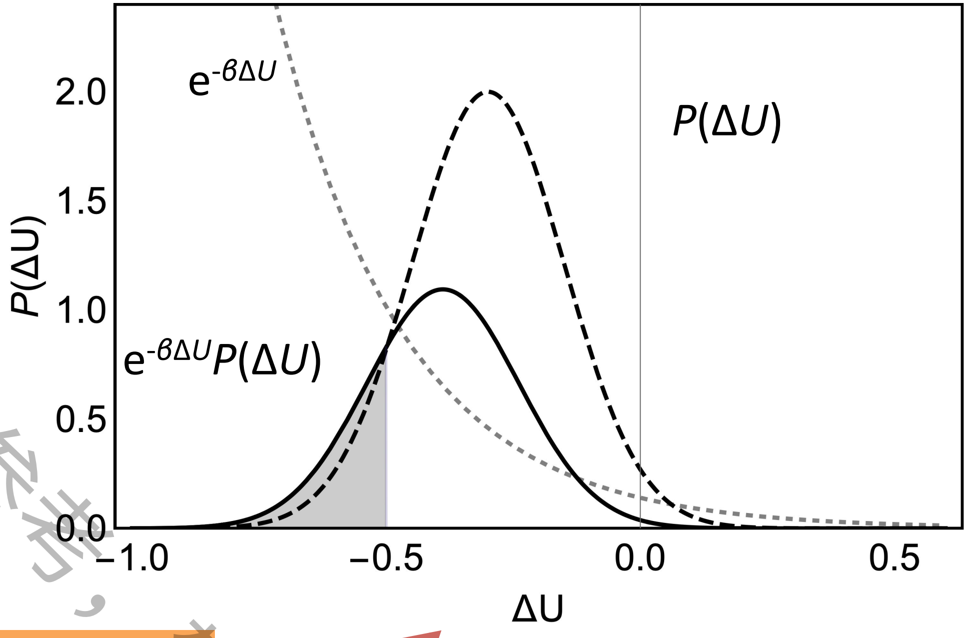
$$= -k_B T \ln \frac{Q_1}{Q_0}$$

$$\Delta F = -k_B T \ln \frac{\int \exp\left(-\beta U_1(\mathbf{r}^N)\right) \mathrm{d}\Gamma}{\int \exp\left(-\beta U_0(\mathbf{r}^N)\right) \mathrm{d}\Gamma}$$

$$= -k_B T \ln \frac{\int \exp\left(-\beta [U_1(\mathbf{r}^N) - U_0(\mathbf{r}^N)]\right) \exp\left(-\beta U_0(\mathbf{r}^N)\right) \mathrm{d}\Gamma}{\int \exp\left(-\beta U_0(\mathbf{r}^N)\right) \mathrm{d}\Gamma}$$

$$= -k_B T \ln \left\langle \exp\left(-\beta [U_1(\mathbf{r}^N) - U_0(\mathbf{r}^N)]\right) \right\rangle_0$$

$$\boxed{= -k_B T \ln \int \exp\left(-\beta \Delta U\right) P_0(\Delta U) \mathrm{d}\Delta U}$$

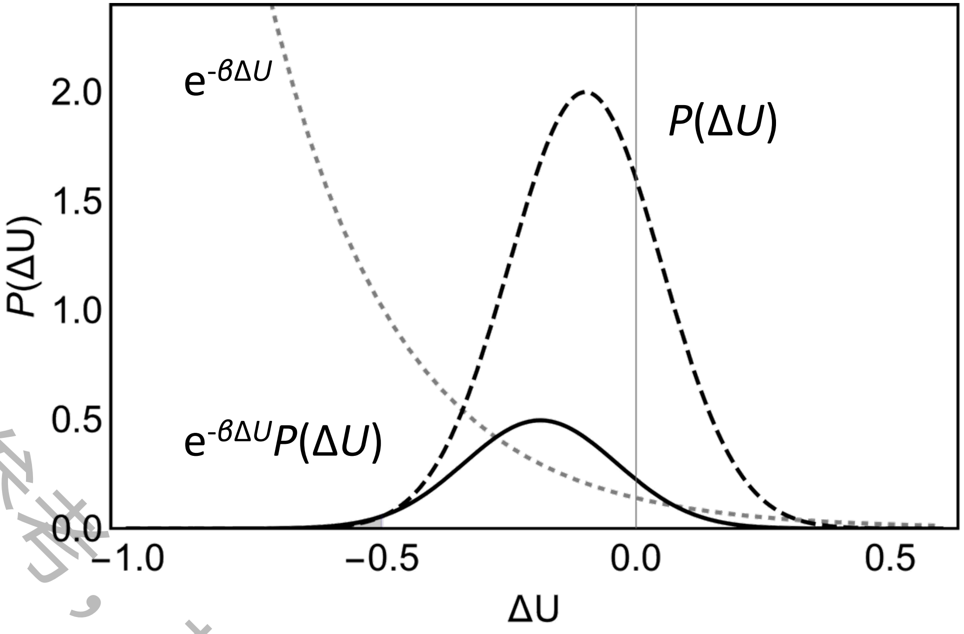


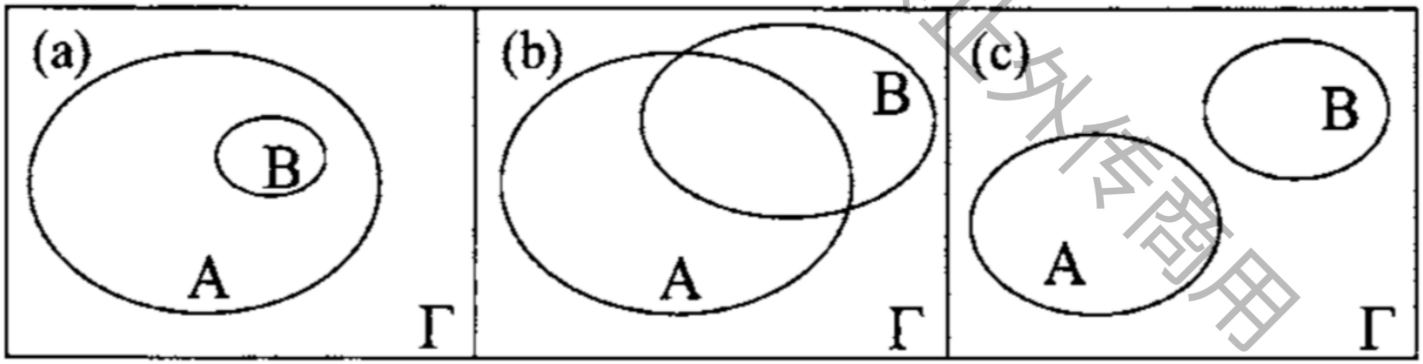
Calculate the difference between two phase spaces in on single phase space, actually this is a perturbation method.

FEP(2)

If the difference between two potential energies is too large, the integrant becomes very small, the simulation thus fails!

$$\Delta F = -k_B T \ln \int \exp\left(-\beta \Delta U\right) P_0(\Delta U)\, \mathrm{d}\Delta U$$





Di Wu and David A. Kofke, *J. Chem. Phys*. **2005**, *123*, 054103

FEP(3)

$$\Delta F = -k_{\mathrm{B}}T\ln\left\langle \exp\left(-\beta[U_1(\boldsymbol{r}^N)-U_0(\boldsymbol{r}^N)]\right)\right\rangle_0$$

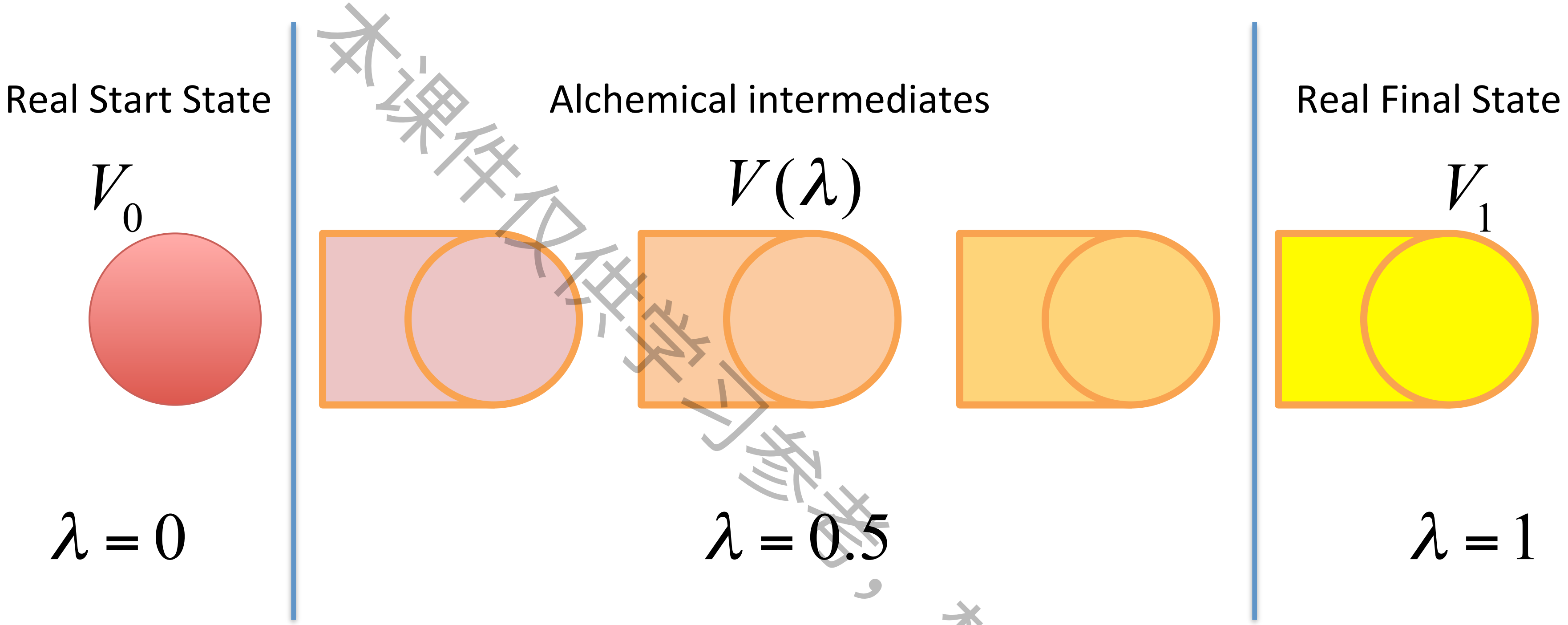
$$\Delta U = U_1(\boldsymbol{r}^N)-U_0(\boldsymbol{r}^N)$$

$$\approx \left\langle \Delta U\right\rangle_0 - \frac{\beta}{2}\left(\left\langle \Delta U^2\right\rangle_0 - \left\langle \Delta U\right\rangle_0^2\right)$$

$$\Delta F = \left\langle \Delta U\right\rangle_0 - \frac{1}{2}\beta\sigma^2$$

From *cumulant expansion*, one can use even simpler formula to practice.

1. Zwanzig, R. W., High-temperature equation of state by a perturbation method.
I. Nonpolar gases, *J. Chem. Phys.* **1954**, 22, 1420–1426
2. Kenney, J. F.; Keeping, E. S., *Mathematics of Statistics*, [2nd edition], Van Nostrand: Princeton, NJ, 1951



Real Start State

V_0

$\lambda = 0$

Alchemical intermediates

$V(\lambda)$

$\lambda = 0.5$

Real Final State

V_1

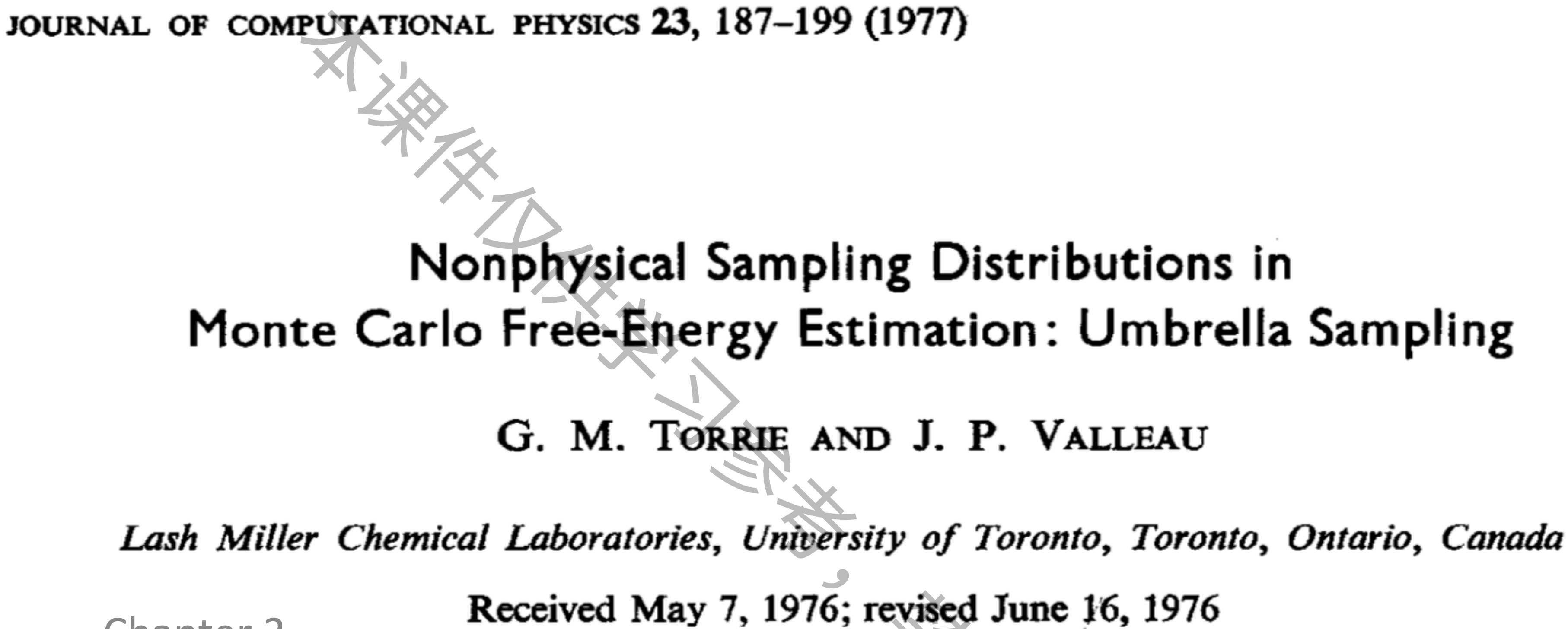
$\lambda = 1$

$$\Delta G = G(1) - G(0) = \left[G(1) - G(\lambda = 0.9)\right] +$$
$$\left[G(\lambda = 0.9) - G(\lambda = 0.8)\right] + \cdots + \left[G(\lambda = 0.1) - G(0)\right]$$

(Suppose we added 8 intermediate states)

For more sophisticated method to obtain ΔG accurately, one can use BAR (Bennet’s acceptance Ratio).

https://github.com/samuelymei/freeenergy/blob/master/notes/Bennett_Acceptance_Ratio.ipynb



JOURNAL OF COMPUTATIONAL PHYSICS 23, 187–199 (1977)

Nonphysical Sampling Distributions in
Monte Carlo Free-Energy Estimation: Umbrella Sampling

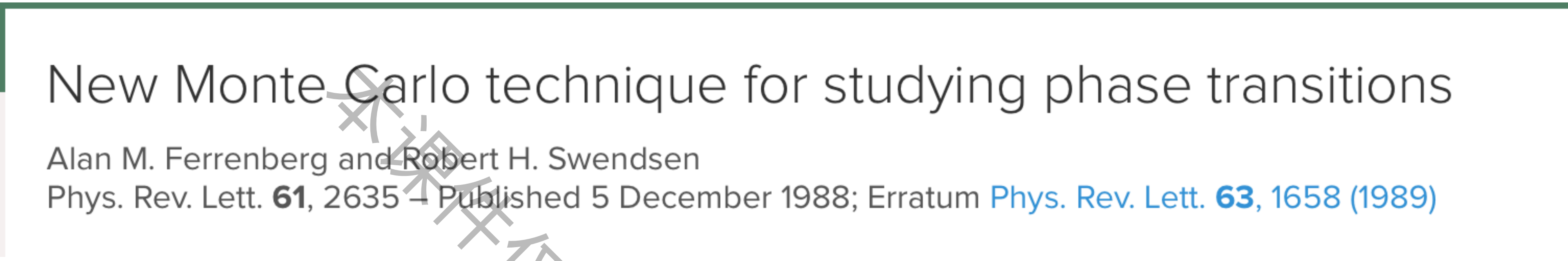
G. M. TORRIE AND J. P. VALLEAU

Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario, Canada

Received May 7, 1976; revised June 16, 1976

Chapter 2

PROBABILITY METHOD

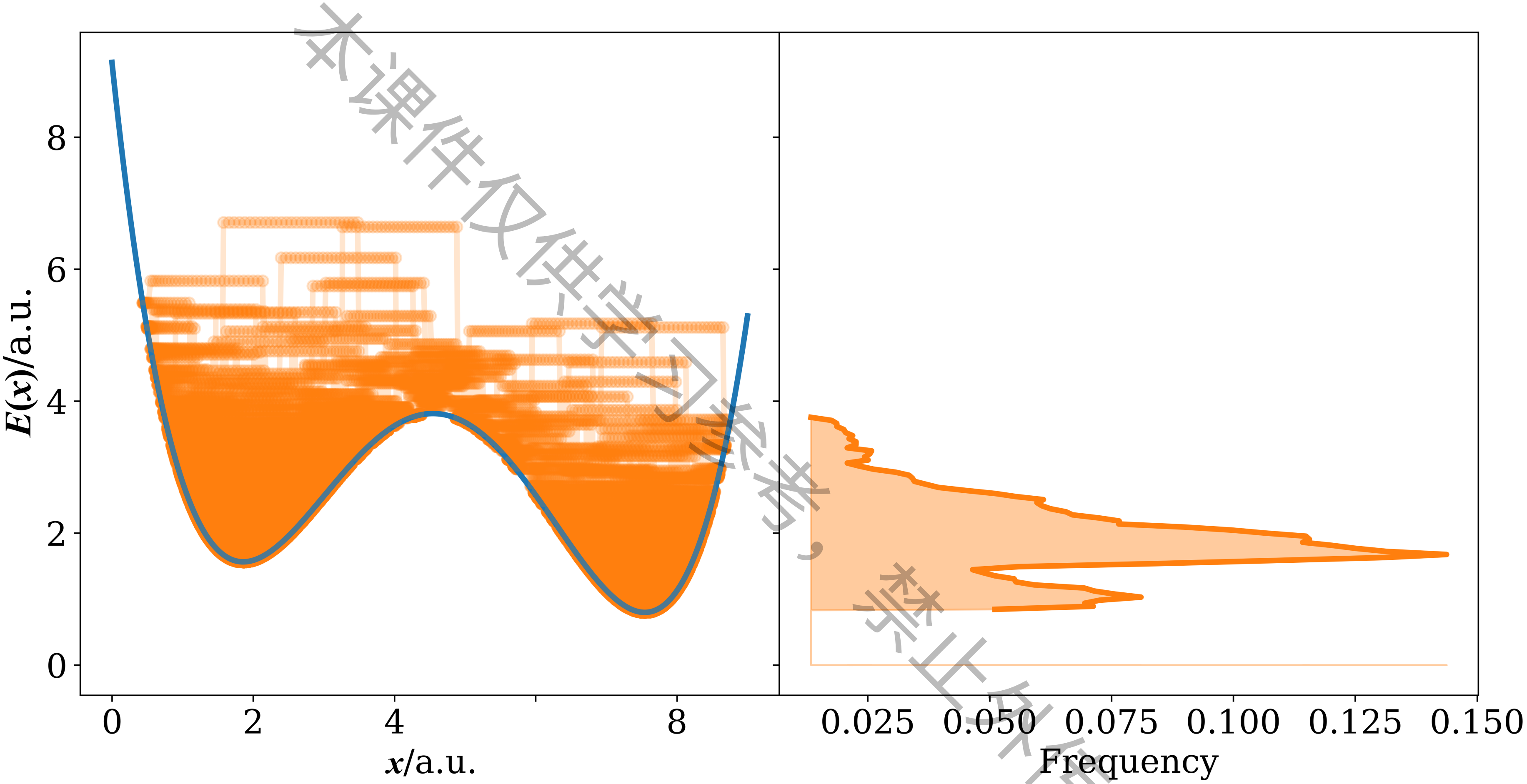


New Monte Carlo technique for studying phase transitions

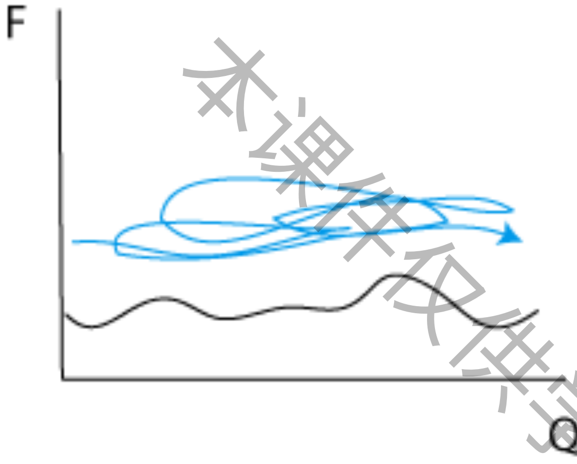
Alan M. Ferrenberg and Robert H. Swendsen
Phys. Rev. Lett. **61**, 2635 – Published 5 December 1988; Erratum Phys. Rev. Lett. **63**, 1658 (1989)

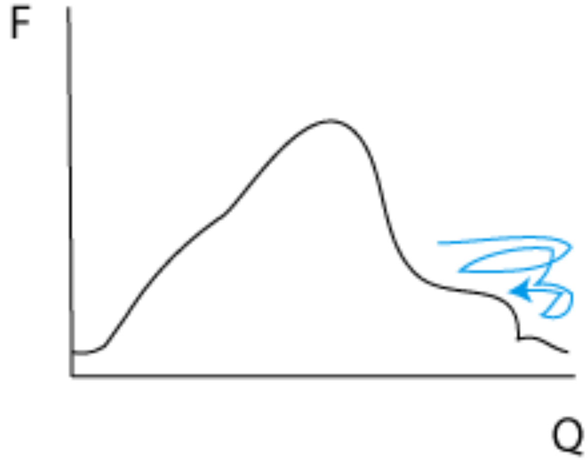
Optimized Monte Carlo data analysis

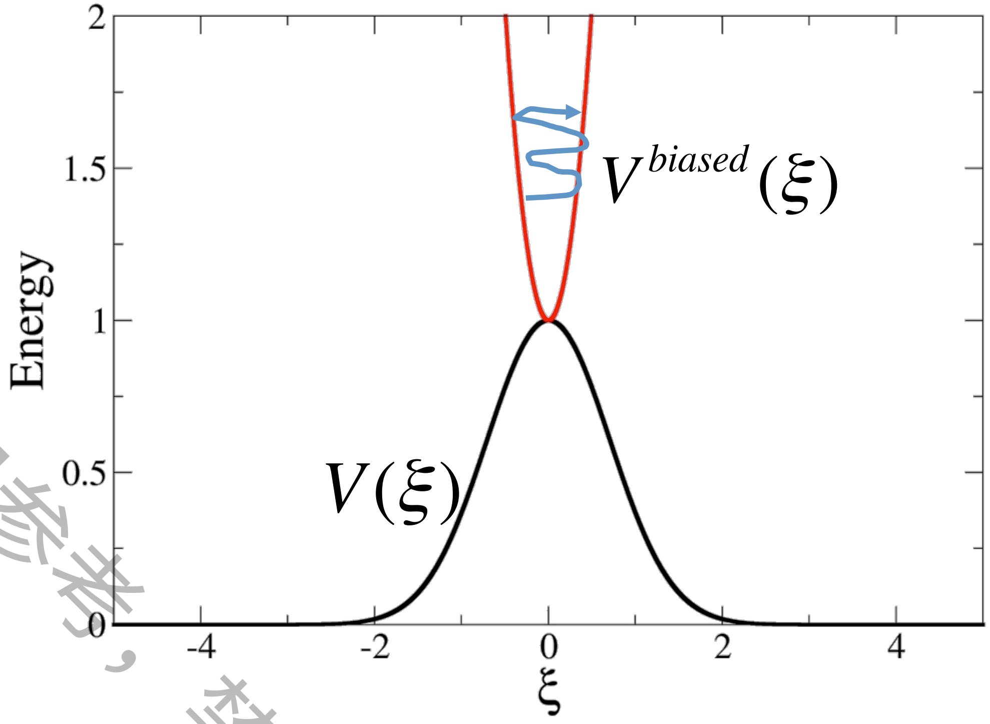
Alan M. Ferrenberg and Robert H. Swendsen
Phys. Rev. Lett. **63**, 1195 – Published 18 September 1989



T=80000.0 K


F
Q


F
Q


Energy
2
1.5
1
0.5
0
-4
-2
0
2
4
ξ
$V^{biased}(\xi)$
$V(\xi)$

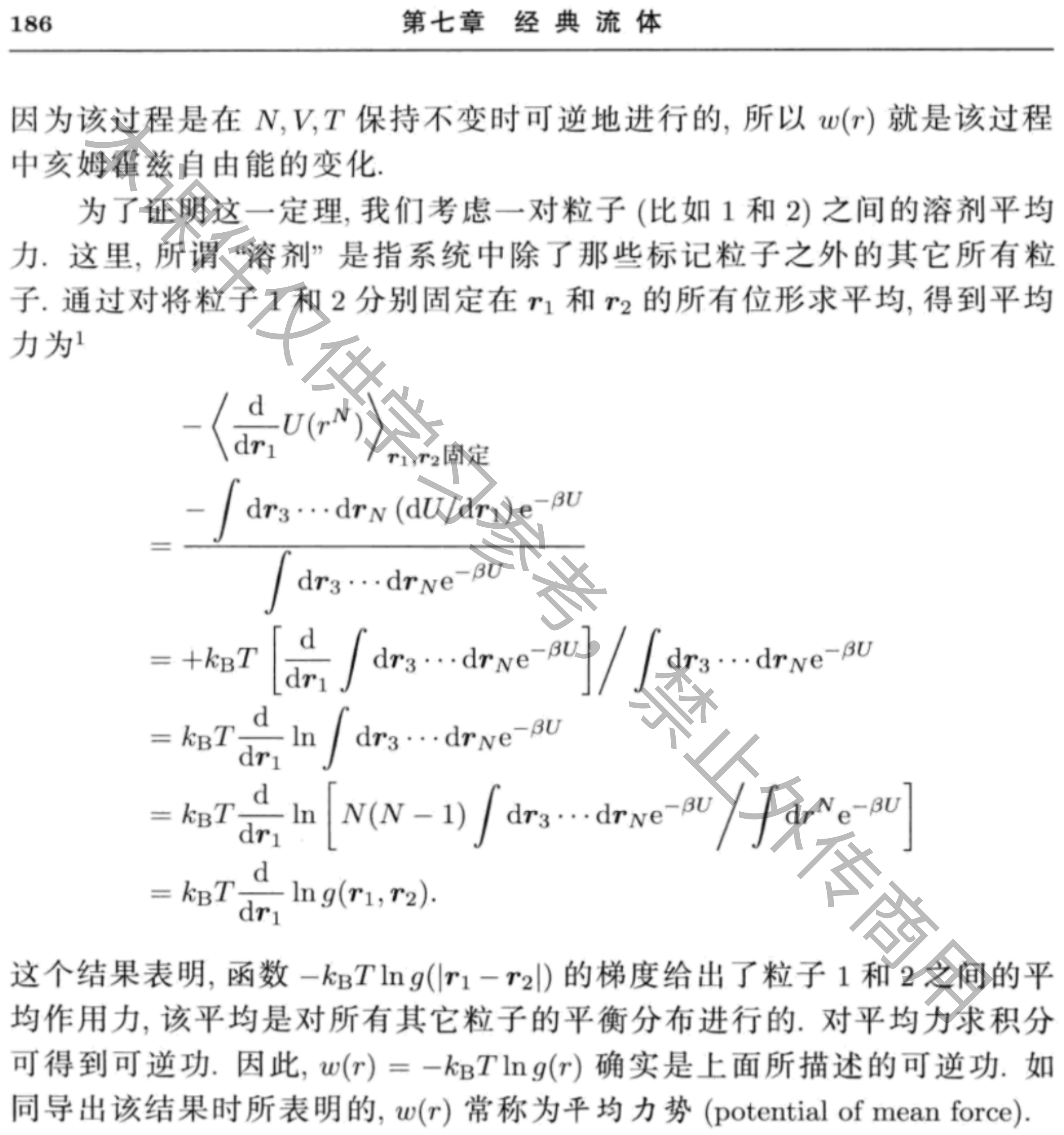
Probability

$$\Delta F_{1\leftarrow 0} = -k_{\mathrm{B}} T \ln \frac{P_1}{P_0} \qquad P(\xi) = \frac{\int e^{-\beta U}\delta(\xi - \vec{x})\,\mathrm{d}^{3N}x}{Q}$$

$$P(U) = \frac{e^{-\beta U} g(U)}{Q} = \frac{\int \delta(U-U')e^{-\beta U'} g(U')\,\mathrm{d}U'}{Q}$$

$$\mathrm{d}^{3N}x = g(U)\,\mathrm{d}U$$

$$P(\lambda) = \frac{Q(\lambda)}{N_{\mathrm{norm}}} = \frac{\int e^{-\beta H}\delta(\lambda' - \lambda)\,\mathrm{d}^{3N}x_{\lambda'}}{N_{\mathrm{norm}}}$$

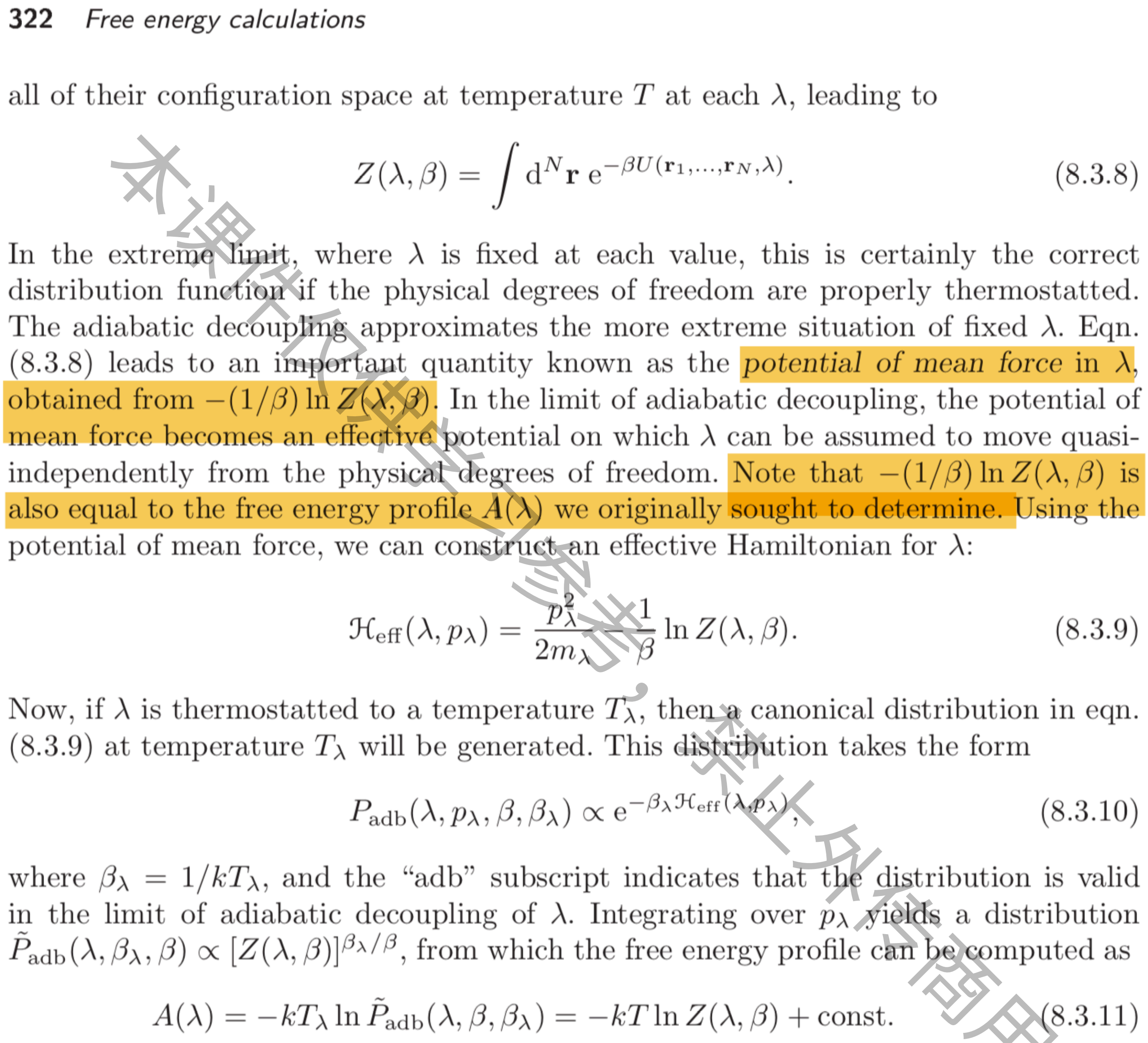


因为该过程是在 N, V, T 保持不变时可逆地进行的，所以 $w(r)$ 就是该过程中亥姆霍兹自由能的变化.

为了证明这一定理，我们考虑一对粒子 (比如 1 和 2) 之间的溶剂平均力. 这里，所谓 “溶剂” 是指系统中除了那些标记粒子之外的其它所有粒子. 通过对将粒子 1 和 2 分别固定在 $\boldsymbol{r}_1$ 和 $\boldsymbol{r}_2$ 的所有位形求平均，得到平均力为[1]

$$
\begin{aligned}
&-\left\langle \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}_1} U(r^N) \right\rangle_{\boldsymbol{r}_1,\boldsymbol{r}_2\text{固定}} \\
&= \frac{-\int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \, (\mathrm{d}U/\mathrm{d}\boldsymbol{r}_1) \mathrm{e}^{-\beta U}}{\int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \mathrm{e}^{-\beta U}} \\
&= +k_{\mathrm{B}}T \left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}_1} \int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \mathrm{e}^{-\beta U} \right] \Big/ \int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \mathrm{e}^{-\beta U} \\
&= k_{\mathrm{B}}T \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}_1} \ln \int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \mathrm{e}^{-\beta U} \\
&= k_{\mathrm{B}}T \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}_1} \ln \left[N(N-1) \int \mathrm{d}\boldsymbol{r}_3 \cdots \mathrm{d}\boldsymbol{r}_N \mathrm{e}^{-\beta U} \Big/ \int \mathrm{d}\boldsymbol{r}^N \mathrm{e}^{-\beta U} \right] \\
&= k_{\mathrm{B}}T \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}_1} \ln g(\boldsymbol{r}_1, \boldsymbol{r}_2).
\end{aligned}
$$

这个结果表明，函数 $-k_{\mathrm{B}}T \ln g(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)$ 的梯度给出了粒子 1 和 2 之间的平均作用力，该平均是对所有其它粒子的平衡分布进行的. 对平均力求积分可得到可逆功. 因此，$w(r) = -k_{\mathrm{B}}T \ln g(r)$ 确实是上面所描述的可逆功. 如同导出该结果时所表明的，$w(r)$ 常称为平均力势 (potential of mean force).



all of their configuration space at temperature T at each λ, leading to

$$Z(\lambda, \beta) = \int d^N\mathbf{r}\; e^{-\beta U(\mathbf{r}_1,\ldots,\mathbf{r}_N,\lambda)}. \tag{8.3.8}$$

In the extreme limit, where λ is fixed at each value, this is certainly the correct distribution function if the physical degrees of freedom are properly thermostatted. The adiabatic decoupling approximates the more extreme situation of fixed λ. Eqn. (8.3.8) leads to an important quantity known as the *potential of mean force* in λ, obtained from $-(1/\beta)\ln Z(\lambda, \beta)$. In the limit of adiabatic decoupling, the potential of mean force becomes an effective potential on which λ can be assumed to move quasi-independently from the physical degrees of freedom. Note that $-(1/\beta)\ln Z(\lambda, \beta)$ is also equal to the free energy profile $A(\lambda)$ we originally sought to determine. Using the potential of mean force, we can construct an effective Hamiltonian for λ:

$$\mathcal{H}_{\text{eff}}(\lambda, p_\lambda) = \frac{p_\lambda^2}{2m_\lambda} - \frac{1}{\beta}\ln Z(\lambda, \beta). \tag{8.3.9}$$

Now, if λ is thermostatted to a temperature T_λ, then a canonical distribution in eqn. (8.3.9) at temperature T_λ will be generated. This distribution takes the form

$$P_{\text{adb}}(\lambda, p_\lambda, \beta, \beta_\lambda) \propto e^{-\beta_\lambda \mathcal{H}_{\text{eff}}(\lambda, p_\lambda)}, \tag{8.3.10}$$

where $\beta_\lambda = 1/kT_\lambda$, and the "adb" subscript indicates that the distribution is valid in the limit of adiabatic decoupling of λ. Integrating over p_λ yields a distribution $\tilde{P}_{\text{adb}}(\lambda, \beta_\lambda, \beta) \propto [Z(\lambda, \beta)]^{\beta_\lambda/\beta}$, from which the free energy profile can be computed as

$$A(\lambda) = -kT_\lambda \ln \tilde{P}_{\text{adb}}(\lambda, \beta, \beta_\lambda) = -kT\ln Z(\lambda, \beta) + \text{const.} \tag{8.3.11}$$

Mark Tuckerman, *Statistical Mechanics*, Oxford University Press, 2010

Density of state

$$\boxed{p(U;N,V,T)} = \frac{e^{-\beta U}\boxed{g(N,V,U)}}{\boxed{Q(N,V,T)}}$$

Probability density

Configurational Integral

In discrete case:

Number of times the system is in $[U, U+\Delta U]$

Tilde '~' means in ***practical***.

$$\tilde{p}(U;N,V,T) = \frac{f(U)}{\Delta U \sum_{U'} f(U')}$$

Can be obtained from MD directly!

In continuous limit:

$$\Delta U \to 0 \quad \sum_{U'} f(U') \to \infty$$

$$\tilde{p}(U;N,V,T) \to p(U;N,V,T)$$

Can be obtained by MD

$$\tilde{g}(N,V,U) = \boxed{\tilde{p}(U;N,V,T_0)}\boxed{e^{\beta_0 U} Q(N,V,T_0)}$$

$Q(N,V,T_0)$ is a *constant*, and is *irrelevant* to *U*.

And looks we can choose *any* temperature T_0 to obtain $g(N,V,U)$.

But in MD it’s *hard* to obtain $g(U)$ out of its equilibrium value.

$$\langle U\rangle(T)=\frac{\int U e^{-\beta U}\,\tilde{g}(U)\,\mathrm{d}U}{\int e^{-\beta U}\,\tilde{g}(U)\,\mathrm{d}U}$$

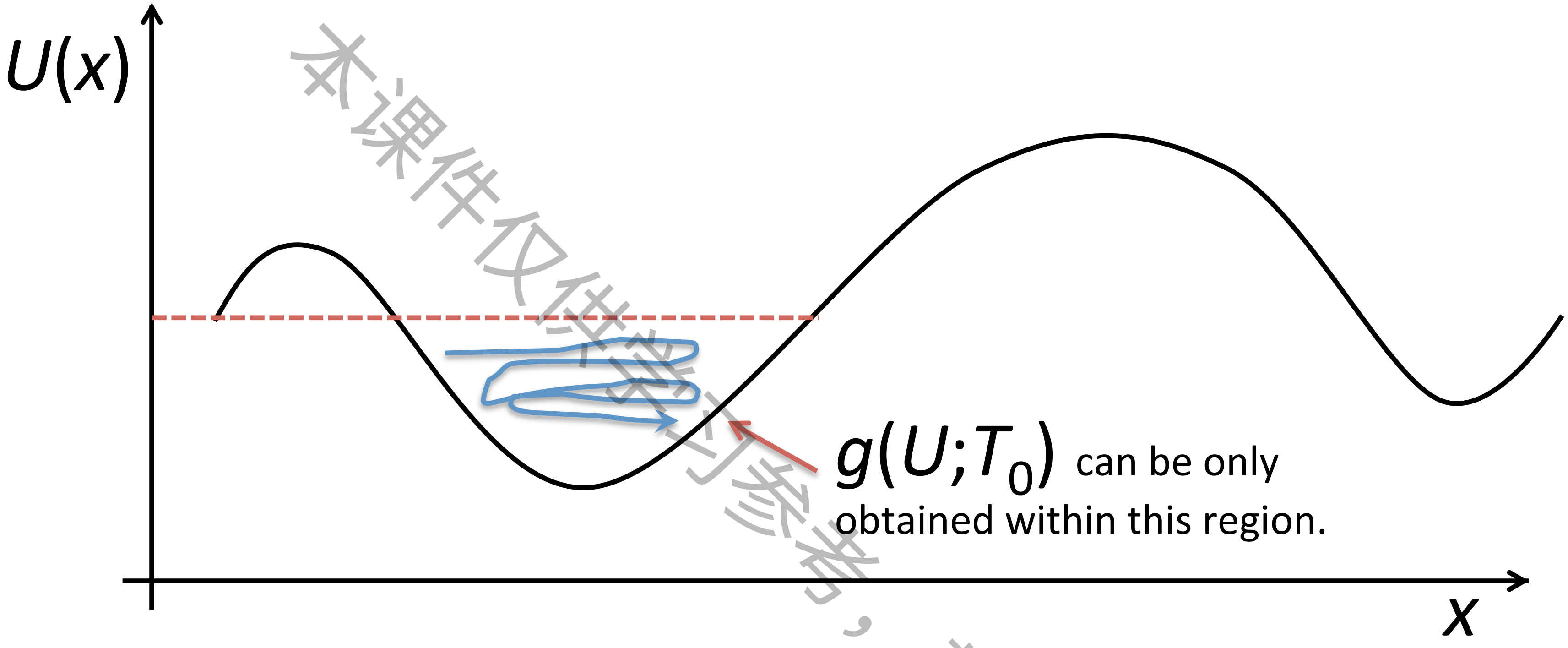
$$=\frac{\int U e^{-(\beta-\beta_0)U}\,\tilde{p}(U;T_0)\,\mathrm{d}U\,\cancel{Q(N,V,T_0)}}{\int e^{-(\beta-\beta_0)U}\,\tilde{p}(U;T_0)\,\mathrm{d}U\,\cancel{Q(N,V,T_0)}}$$

What we have done:

measured distributions at one set of state conditions, (*N*, *V*, T_0) here, are projected onto another, (*N*, *V*, *T*) here, is termed as '***histogram reweighting***'

$$\left\langle \xi \right\rangle(T) = \frac{\int \tilde{\xi}(U) e^{-(\beta-\beta_0)U} \tilde{p}(U;T_0)\,\mathrm{d}U}{\int e^{-(\beta-\beta_0)U} \tilde{p}(U;T_0)\,\mathrm{d}U}$$

$$\xi(U') = \frac{\int_{V^N} \xi(\boldsymbol{q})\delta(U(\boldsymbol{q}) - U')\,\mathrm{d}\boldsymbol{q}}{\int_{V^N} \delta(U(\boldsymbol{q}) - U')\,\mathrm{d}\boldsymbol{q}}$$



$g(U;T_0)$ can be only obtained within this region.

$$\beta_1 A(T_1) - \beta_2 A(T_2) = \ln\left[\frac{\tilde{p}(U;T_1)}{\tilde{p}(U;T_2)}\right] + (\beta_1 - \beta_2)U$$

And there is a best *U*, which can maximize this equation.

Proof:

$$\tilde{p}(U;N,V,T_1) = \frac{e^{-\beta_1 U}\tilde{g}(N,V,U)}{Q(N,V,T_1)}$$

$$\tilde{p}(U;N,V,T) = \frac{f(U)}{\Delta U \sum_{U'} f(U')}$$

$$\tilde{p}(U;N,V,T_2) = \frac{e^{-\beta_2 U}\tilde{g}(N,V,U)}{Q(N,V,T_2)}$$

$$\beta_1 A(T_1) = -\ln[Z(N,V,T_1)] = -\ln[Q(N,V,T_1)] + C(T_1)$$

$$= -\ln[e^{-\beta_1 U}\tilde{g}(N,V,U)] + \ln[\tilde{p}(U;N,V,T_1)] + C(T_1)$$

Analytical

$$\beta_1 A(T_1) - \beta_2 A(T_2) = \ln[\frac{\tilde{p}(U;T_1)}{\tilde{p}(U;T_2)}] - \ln[e^{-\beta_1 U}e^{+\beta_2 U}] + \boxed{C(T_1) - C(T_2)}$$

$\sim 3N\ln(T_1/T_2)/2$

Application of reweighting: WHAM

- weighted histogram analysis method: WHAM
- The variable wanted is $g(U)$, which contains no temperature. But the potential energy U wanted would be hard to sample at some temperature. So the actual usage is to simulate at different temperature several times, to obtain multiple trajectories, and at last use a weight based on the error estimation to construct the $g(U)$.

From definition:

$$\tilde{p}_{r,i}(U;T) = \frac{\tilde{p}(U;T_i)e^{-(\beta-\beta_i)U}}{\int \tilde{p}(U;T)e^{-(\beta-\beta_i)U}\,\mathrm{d}U}$$

r means reweighted.

It's obtained from the i_{th} simulation.

Here T is a target temperature, while T_i is the temperature MD is performed. $T<T_\text{i}$

And from the original formula:

$$p(U;N,V,T) = \frac{e^{-\beta U}g(N,V,U)}{Q(N,V,T)}$$

And consider the $g(U)$ is unchanged, we can derive:

$$\tilde{p}_{r,i}(U;T) = \frac{Q_i}{Q}\tilde{p}(U;T_i)e^{-(\beta-\beta_i)U}$$

Ferrenberg & Swendsen’s WHAM

$$\tilde{p}_r(U;T) = \sum_i w_i(U)\tilde{p}_{r,i}(U;T) = \sum_i w_i(U)\frac{Q_i}{Q}\tilde{p}(U;T_i)e^{-(\beta-\beta_i)U}$$

With a constraint:

$$\sum_i w_i(U) = 1$$

Now the problem at hand is to optimize the set of W_i with the constraint above.
But we can go further to transform the problem even easier.

That is, to minimize the sampling error.

Final result

$$\tilde{p}_r^*(U;T) = \frac{\sum_i f_i(U)\exp(-\beta U)}{\sum_i f_{tot,i}\exp(\beta_i \tilde{A}_i - \beta_i U)}$$

Number of counts in potential energy *U* and in the *i*$_{\text{th}}$ MD run.

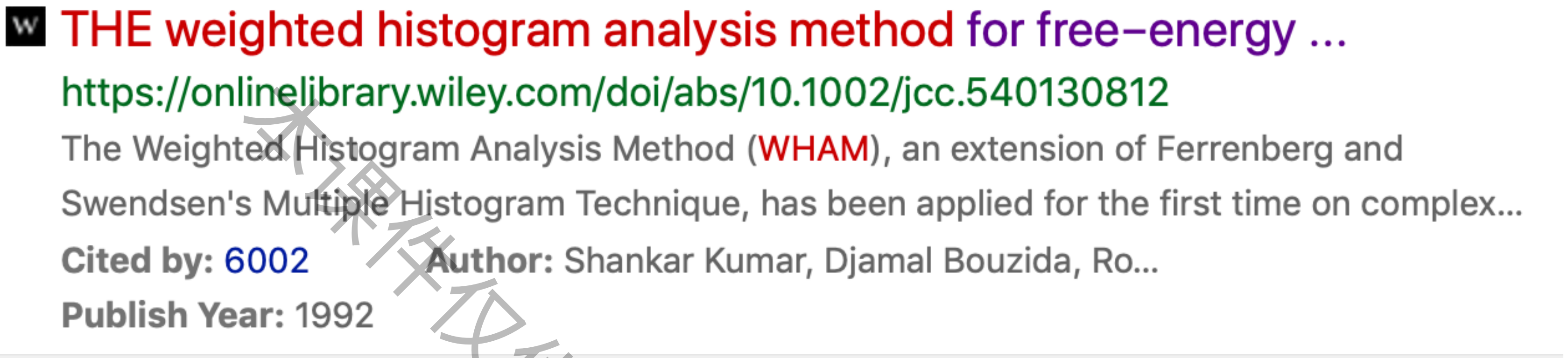
Total number of counts in the *i*$_{\text{th}}$ MD run.

$$\boxed{\exp(-\beta_i \tilde{A}_i) = \sum_U \tilde{p}_r^*(U;T_i)}$$

$$\tilde{p}_r(U;T) = \frac{\tilde{p}_r^*(U;T)}{\sum_U \tilde{p}_r^*(U;T)}$$

Homework:

Derive this set of formula from previous equations.



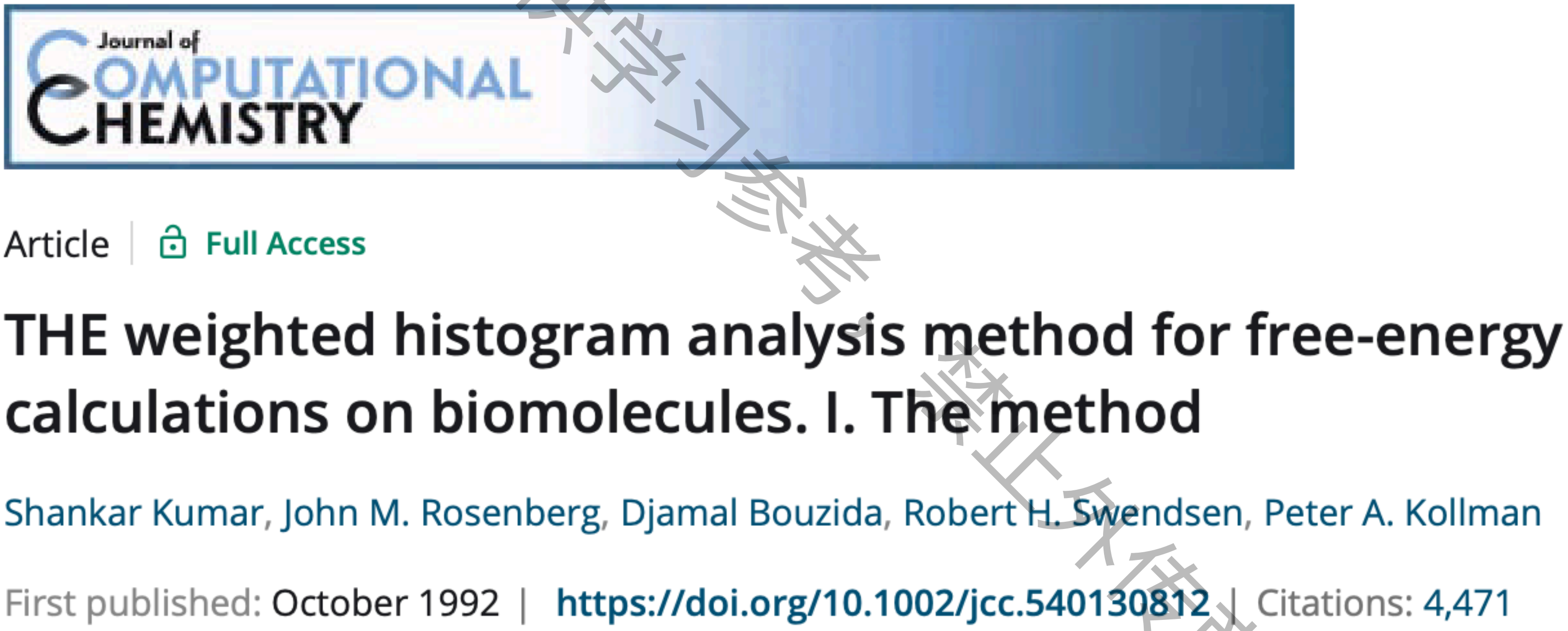
THE weighted histogram analysis method for free–energy ...

https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130812

The Weighted Histogram Analysis Method (WHAM), an extension of Ferrenberg and Swendsen's Multiple Histogram Technique, has been applied for the first time on complex...

Cited by: 6002 **Author:** Shankar Kumar, Djamal Bouzida, Ro...

Publish Year: 1992



Journal of COMPUTATIONAL CHEMISTRY

Article | Full Access

THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method

Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, Peter A. Kollman

First published: October 1992 | https://doi.org/10.1002/jcc.540130812 | Citations: 4,471

When *U* has parameters

$$U = \sum_j \lambda_j V_j; \quad U_i = \sum_j \lambda_{i,j} V_j$$

i: the *i*th simulation run. (window)
j: the *j*th parameter, λ_j

$\vec{V}$ means all possible V

$$\tilde{p}_r^*(U;T) = \frac{\sum_i f_i(\vec{V},\xi)\exp(-\beta\sum_j \lambda_j V_j)}{\sum_i f_{tot,i}\exp(\beta_i \tilde{A}_i - \beta_i \sum_j \lambda_{i,j} V_j)}$$

$$\exp(-\beta_i \tilde{A}_i) = \sum_U \tilde{p}_r^*(\vec{V},\xi;T)$$

$$\tilde{p}_r(\vec{V},\xi;T) = \frac{\tilde{p}_r^*(\vec{V},\xi;T)}{\sum_U \tilde{p}_r^*(\vec{V},\xi;T)}$$

Stratification (Window)

Stratification (Window)

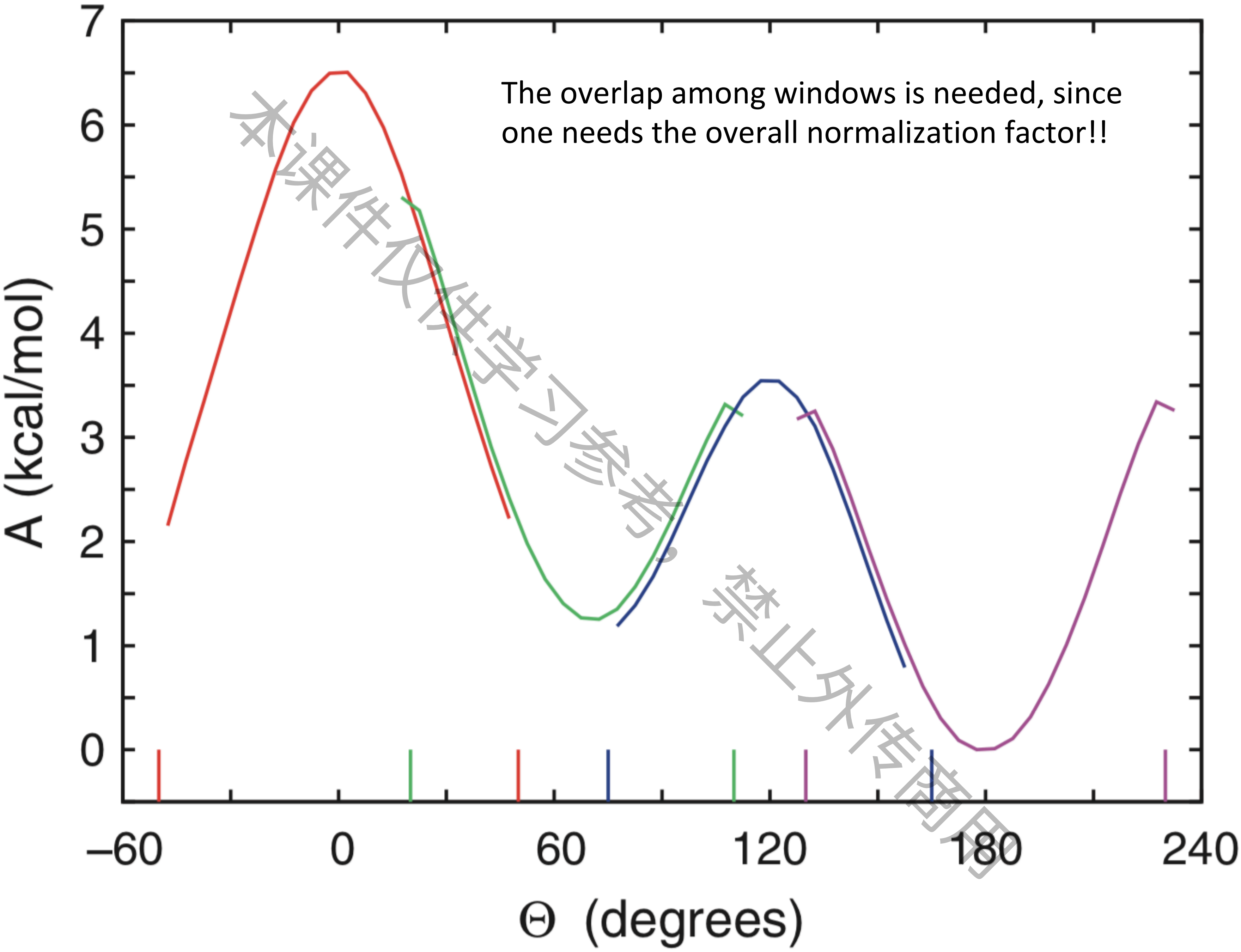
Intuitive thought:

$$\Delta A(\xi) = A(\xi) - A(\xi_0) = -k_{\mathrm{B}} T \ln\left[\frac{P(\xi)}{P(\xi_0)}\right]$$

$$P(\xi_0 + (i - 0.5)\xi) = \frac{f_i}{\sum_j f_j}$$

$$\Delta A(\xi_0 + (i - 0.5)\xi) = -k_{\mathrm{B}} T \ln\left[\frac{f_i}{f_0}\right]$$

So, one needs to obtain *P*(*ξ*) stepwise!



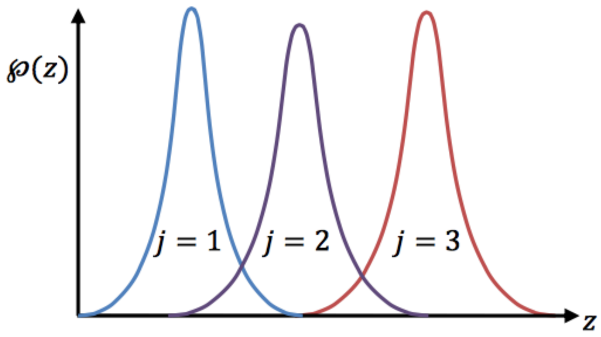
The overlap among windows is needed, since one needs the overall normalization factor!!

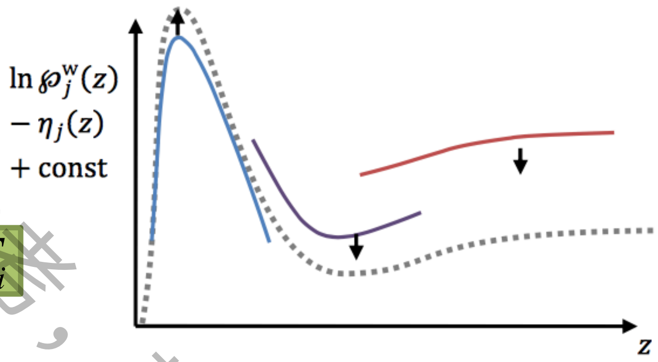
Umbrella sampling for rare event:

$$V_i^{biased}(\xi) = V(\xi) + \frac{1}{2}k_i(\xi - \xi_{0,i})^2$$

$$\boxed{\Delta G_i(\xi) = -\frac{1}{\beta}\ln P_i^{unbiased}(\xi)}$$

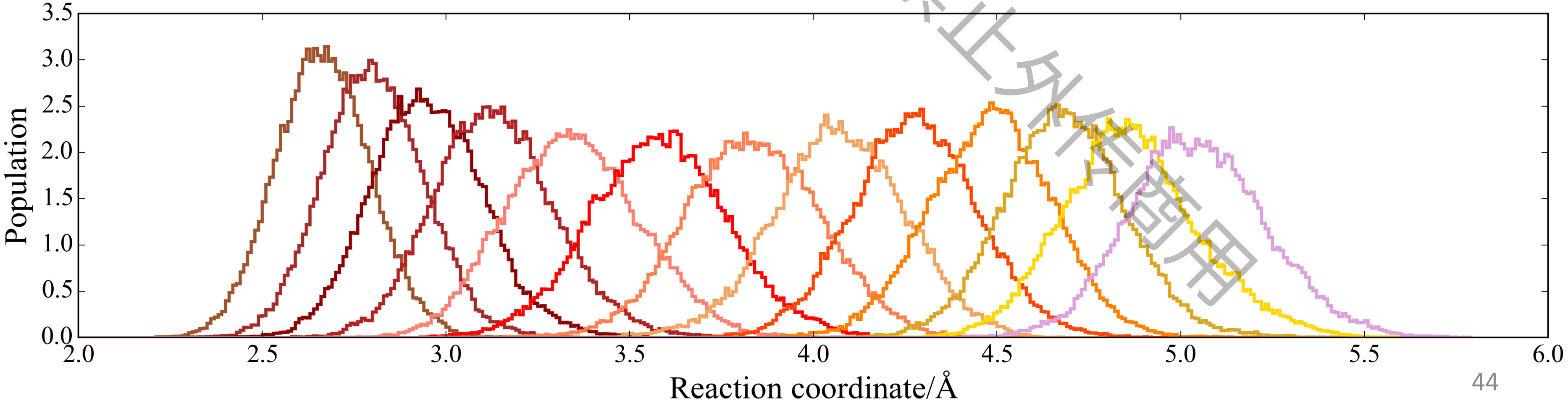
$$= -\frac{1}{\beta}\ln P_i^{biased}(\xi) - \frac{1}{2}k_i(\xi - \xi_{0,i})^2 + F_i$$





作业：

推导这个公式

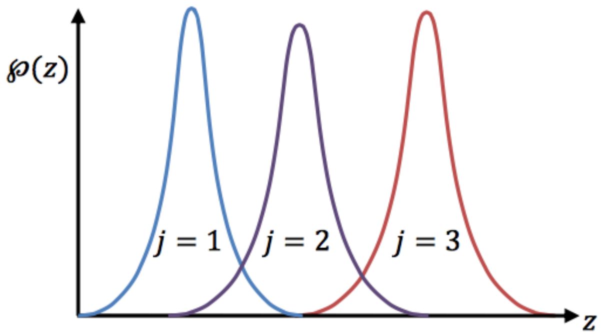


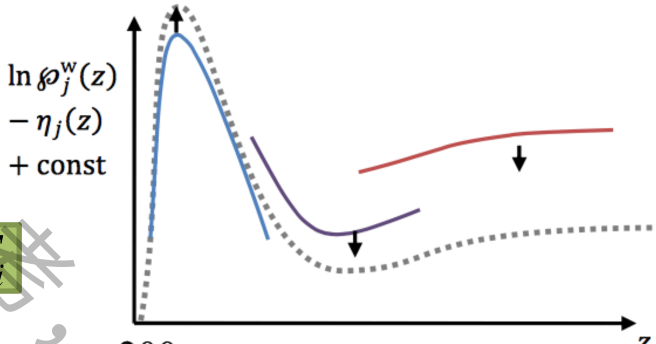
Umbrella sampling for rare event:

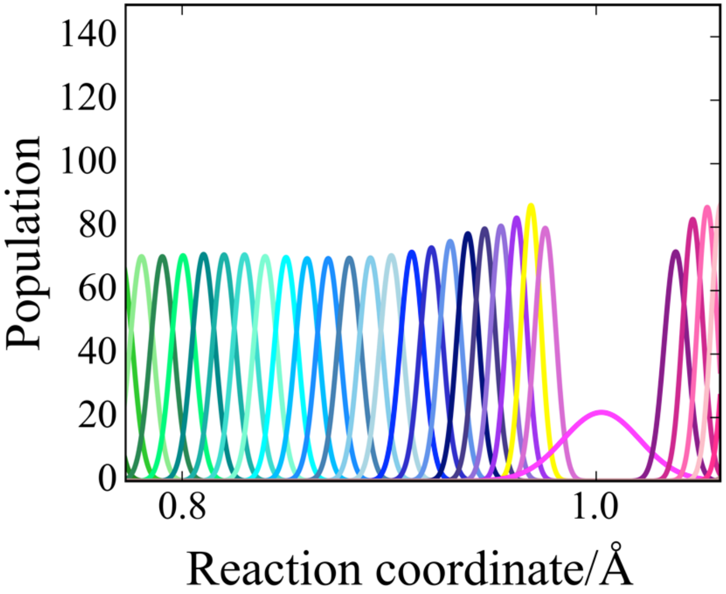
$$V_i^{biased}(\xi) = V(\xi) + \frac{1}{2}k_i(\xi - \xi_{0,i})^2$$

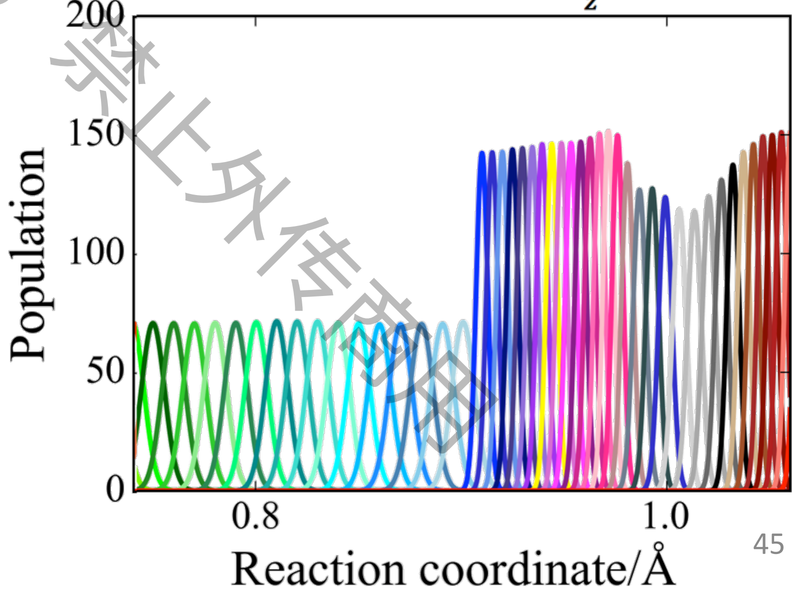
$$G_i(\xi) = -\frac{1}{\beta}\ln P_i^{unbiased}(\xi)$$

$$= -\frac{1}{\beta}\ln P_i^{biased}(\xi) - \frac{1}{2}k_i(\xi - \xi_{0,i})^2 + F_i$$

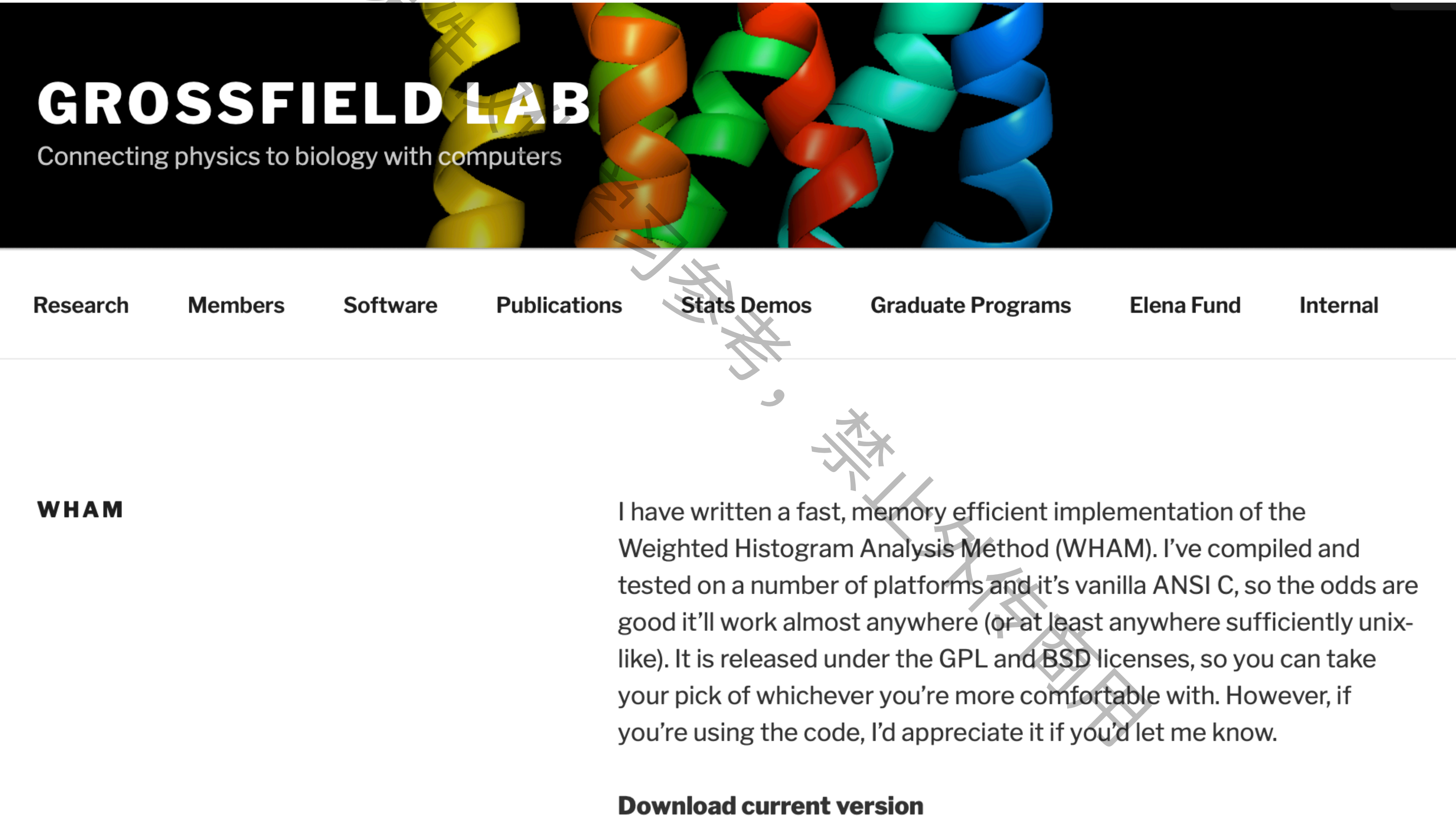








http://membrane.urmc.rochester.edu/?page_id=126



GROSSFIELD LAB

Connecting physics to biology with computers

WHAM

I have written a fast, memory efficient implementation of the Weighted Histogram Analysis Method (WHAM). I've compiled and tested on a number of platforms and it's vanilla ANSI C, so the odds are good it'll work almost anywhere (or at least anywhere sufficiently unix-like). It is released under the GPL and BSD licenses, so you can take your pick of whichever you're more comfortable with. However, if you're using the code, I'd appreciate it if you'd let me know.

Download current version

Chapter 3

THERMODYNAMIC INTEGRATION

Free energy calculation-TI

Thermodynamic Integration

$$\Delta F = F(\lambda_b) - F(\lambda_a) = \int_{\lambda_a}^{\lambda_b} \frac{dF(\lambda)}{d\lambda} d\lambda$$

$$\frac{dF(\lambda)}{d\lambda} = -k_{\mathrm{B}} T \frac{d}{d\lambda} \ln Q$$

$$= -k_{\mathrm{B}} T \frac{1}{Q} \frac{dQ}{d\lambda}$$

$$= -k_{\mathrm{B}} T \frac{1}{Q} \frac{d \int d\Gamma \exp\left(-\beta U(\boldsymbol{r}^N)\right)}{d\lambda}$$

$$= -k_{\mathrm{B}} T \frac{1}{Q} \int d\Gamma \left(-\frac{1}{k_{\mathrm{B}} T}\right) \frac{dU(\boldsymbol{r}^N)}{d\lambda} \exp\left(-\beta U(\boldsymbol{r}^N)\right)$$

$$= \left\langle \frac{dU(\boldsymbol{r}^N)}{d\lambda} \right\rangle$$

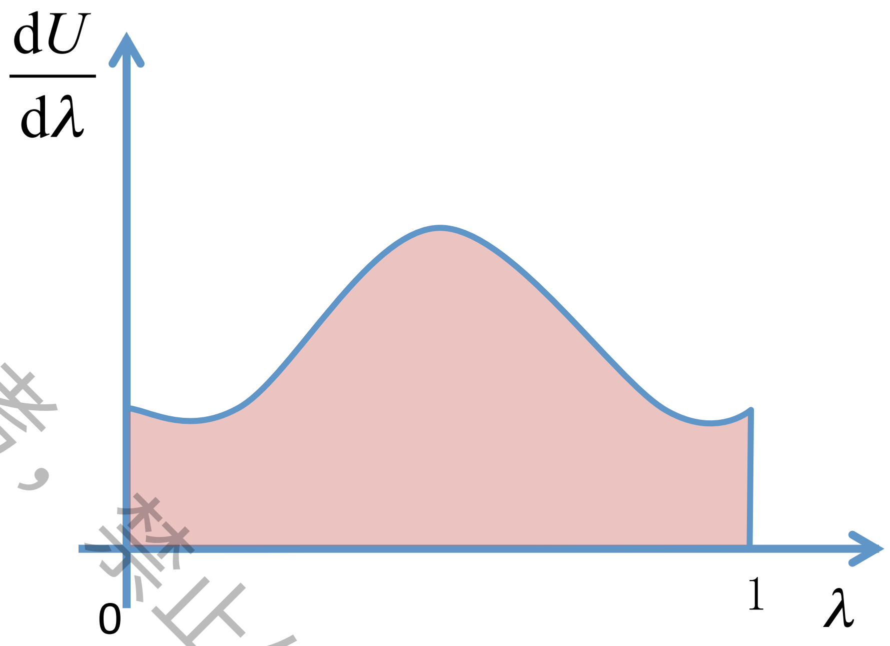
$$A = \sum_n A_n P_n = \sum_n A_n \frac{1}{Z_\lambda} e^{-\beta E_n}$$

$$= \frac{1}{Z_\lambda} \sum_n \left(\frac{\partial E_n}{\partial \lambda}\right) e^{-\beta E_n} = \sum_n \left\langle \frac{\partial E_n}{\partial \lambda} \right\rangle$$

$$U(\lambda) = \lambda U_{\mathrm{N}} + (1-\lambda)U_{\mathrm{O}}$$

$$\frac{\mathrm{d}U}{\mathrm{d}\lambda} = U_{\mathrm{N}} - U_{\mathrm{O}}$$

$$\frac{dF(\lambda)}{d\lambda} = \left\langle \frac{dU(\mathbf{r}^N)}{d\lambda} \right\rangle = \langle U_{\mathrm{N}} - U_{\mathrm{O}} \rangle$$



$$\Delta F = F(\lambda_b) - F(\lambda_a) = \int_{\lambda_a}^{\lambda_b} \frac{dF(\lambda)}{d\lambda} d\lambda$$

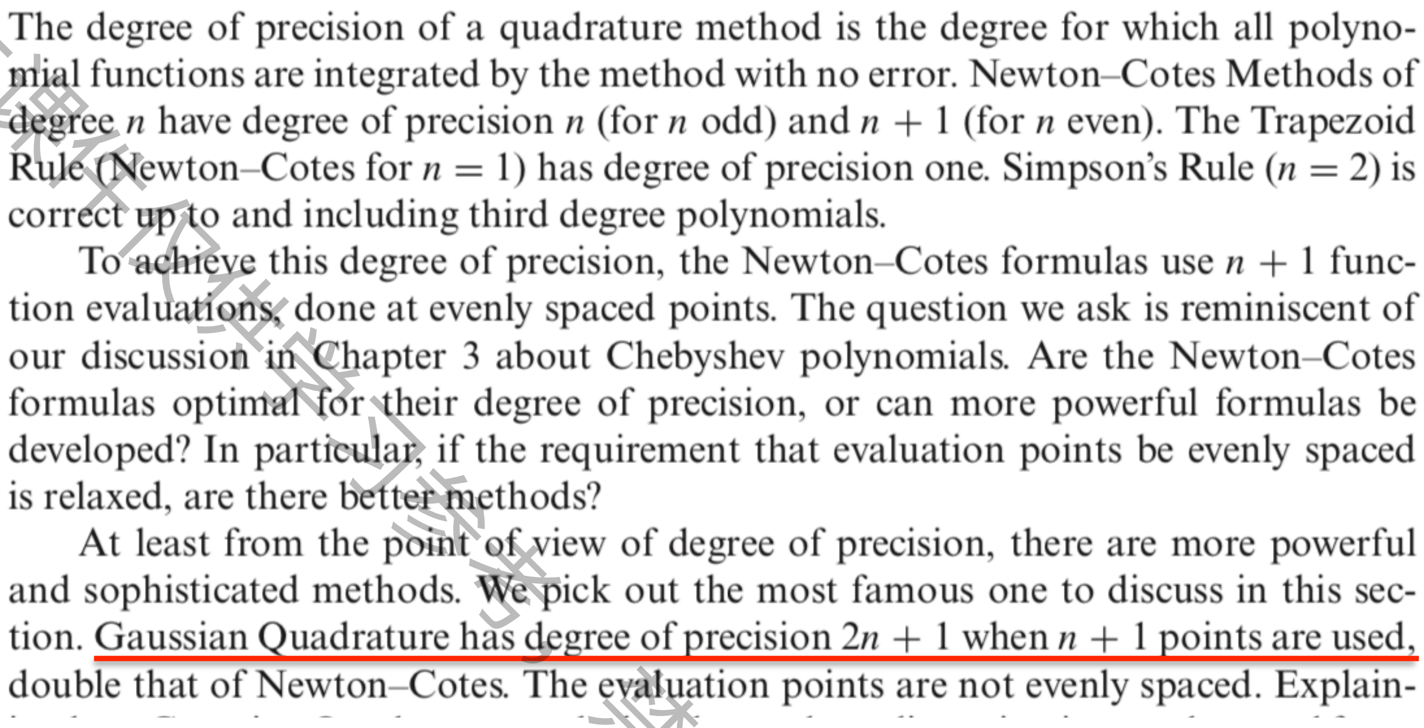
$$= \int_{\lambda_a}^{\lambda_b} \langle U_{\mathrm{N}} - U_{\mathrm{O}} \rangle d\lambda$$

Other example of dU/dl:

$$U(\lambda) = \lambda^2 U_{\mathrm{N}} + (1-\lambda)^2 U_{\mathrm{O}}$$

$$\frac{\mathrm{d}U}{\mathrm{d}\lambda} = 2\lambda U_{\mathrm{N}} - 2(1-\lambda)U_{\mathrm{O}}$$

5.5 GAUSSIAN QUADRATURE



The degree of precision of a quadrature method is the degree for which all polynomial functions are integrated by the method with no error. Newton–Cotes Methods of degree n have degree of precision n (for n odd) and $n + 1$ (for n even). The Trapezoid Rule (Newton–Cotes for $n = 1$) has degree of precision one. Simpson's Rule ($n = 2$) is correct up to and including third degree polynomials.

To achieve this degree of precision, the Newton–Cotes formulas use $n + 1$ function evaluations, done at evenly spaced points. The question we ask is reminiscent of our discussion in Chapter 3 about Chebyshev polynomials. Are the Newton–Cotes formulas optimal for their degree of precision, or can more powerful formulas be developed? In particular, if the requirement that evaluation points be evenly spaced is relaxed, are there better methods?

At least from the point of view of degree of precision, there are more powerful and sophisticated methods. We pick out the most famous one to discuss in this section. Gaussian Quadrature has degree of precision $2n + 1$ when $n + 1$ points are used, double that of Newton–Cotes. The evaluation points are not evenly spaced. Explain-

Tim Sauer, *Numerical Analysis*, 3rd Edition, 2018

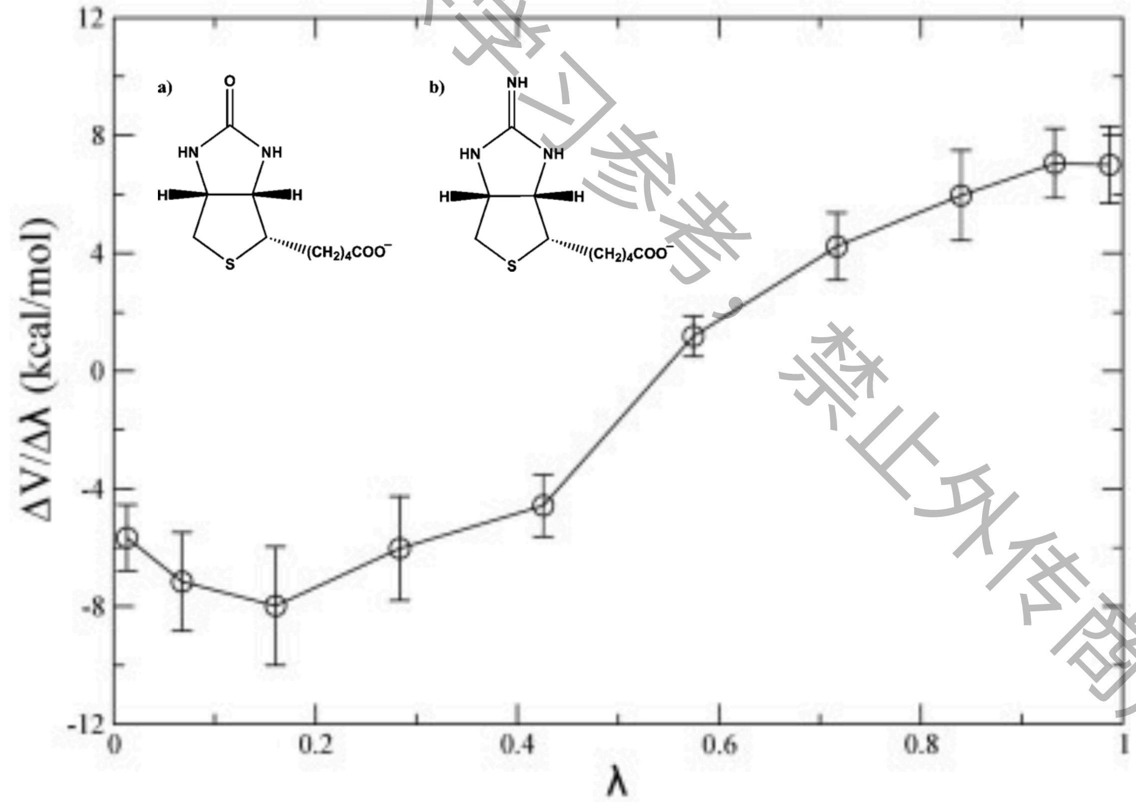
TI

$$\begin{tikzcd}
\text{protein} + \text{iminobiotin} \arrow[r, "\Delta G_a^{imi}"] \arrow[d, "\Delta G_{mut}^{wat}"] & \text{protein/iminobiotin} \arrow[d, "\Delta G_{mut}^{pro}"] \\
\text{protein} + \text{biotin} \arrow[r, "\Delta G_a^{btn}"] & \text{protein/biotin}
\end{tikzcd}$$

TI

Thermodynamic Integration

Pengyu Ren
asolute free
energy change



J. AM. CHEM. SOC. 132(14):9 2010
Journal of Computational Chemistry 33:1374–1382 2012

Chapter 4

OTHER METHODS

Umbrella Integration

How to choose the force constant and the interval?

$$P_i^{biased}(\xi) \overset{\text{assume}}{=} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{(\xi-\xi_{0,i})^2}{2\sigma_i^2}\right);$$

$$\frac{dG_i(\xi)}{d\xi} = -\frac{1}{\beta}\frac{1}{P_i^{biased}(\xi)}\frac{dP_i^{biased}(\xi)}{d\xi} - k(\xi-\xi_{0,i})$$

$$= (\xi-\xi_{0,i})\left[\frac{1}{\beta\sigma_i^2} - k_i\right]$$

$$\Delta G(\xi) = \sum_{i=1}^{N}\int \frac{dG_i(\xi)}{d\xi}d\xi = \sum_{i=1}^{N}\frac{1}{2}(\xi-\xi_{0,i})^2\left[\frac{1}{\beta\sigma_i^2} - k_i\right]$$

Kästner, J. 2011. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1:932-942.

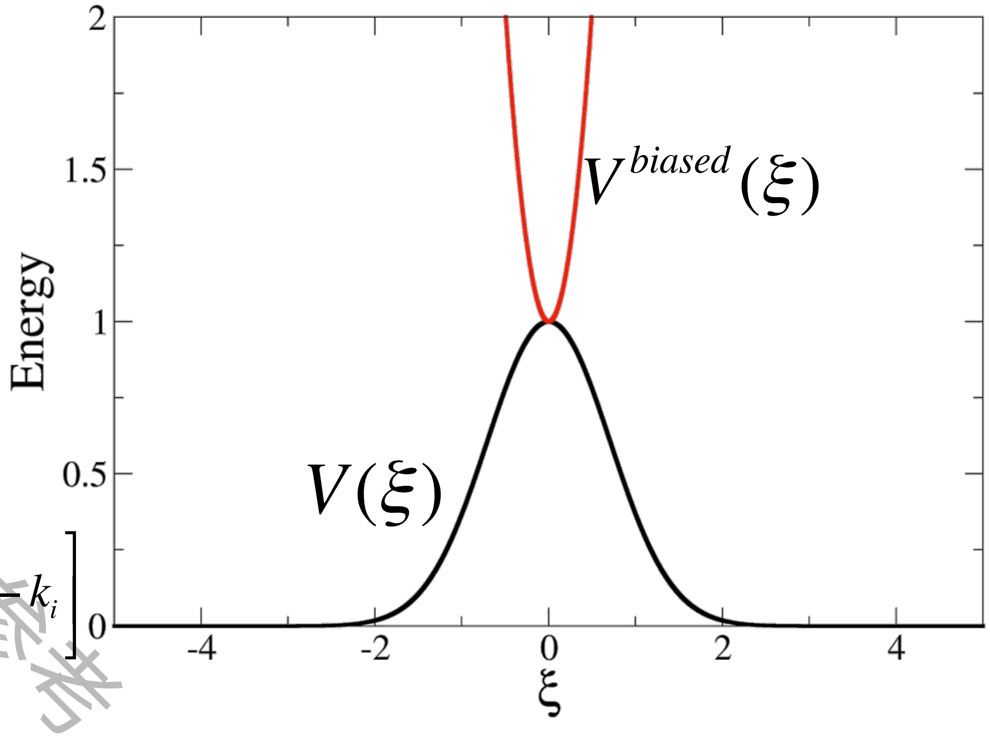
Kastner, J. and W. Thiel. (2005)*J. Phys. Chem*. 123:144104-144105.

US vs. UI

$$G_i^{US}(\xi) = -\frac{1}{\beta}\ln P_i^{unbiased}(\xi)$$

$$= -\frac{1}{\beta}\ln P_i^{biased}(\xi) - \frac{1}{2}k_i(\xi-\xi_{0,i})^2 + F_i$$

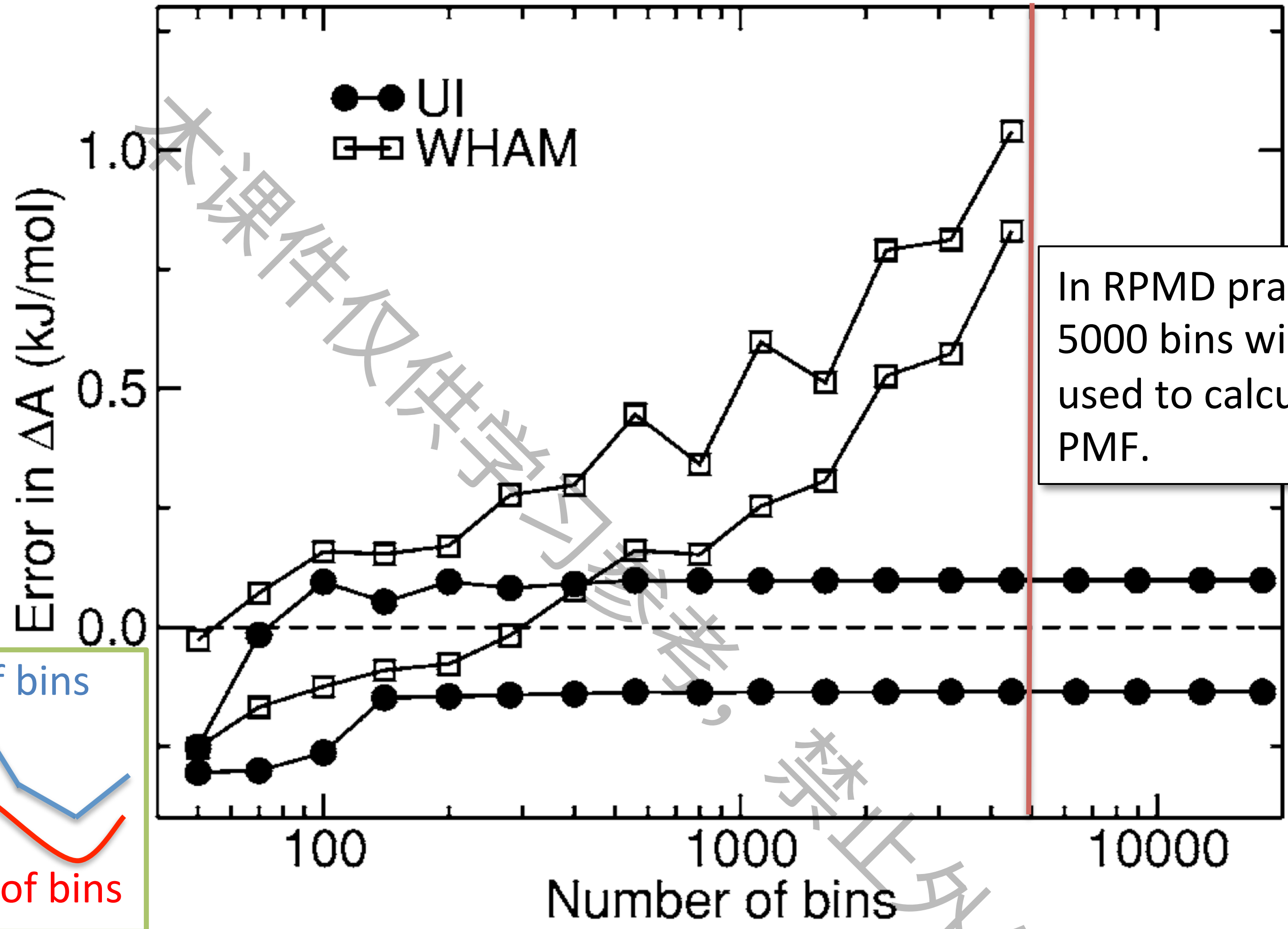
$$\Delta G^{UI}(\xi) = \sum_{i=1}^{N}\int \frac{dG_i(\xi)}{d\xi}d\xi = \sum_{i=1}^{N}\frac{1}{2}(\xi-\xi_{0,i})^2\left[\frac{1}{\beta\sigma_i^2} - k_i\right]$$



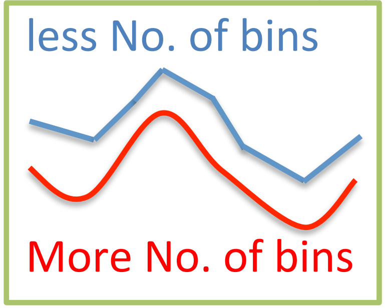
There is a constant F_i in each window.

In traditional method, weighted histogram analysis method (**WHAM**), this constant is determined iteratively.

In **UI**, $dG_i/d\xi$ is calculated, then integrate with windows, avoid the determining of F_i.



In RPMD practice, 5000 bins will be used to calculate PMF.

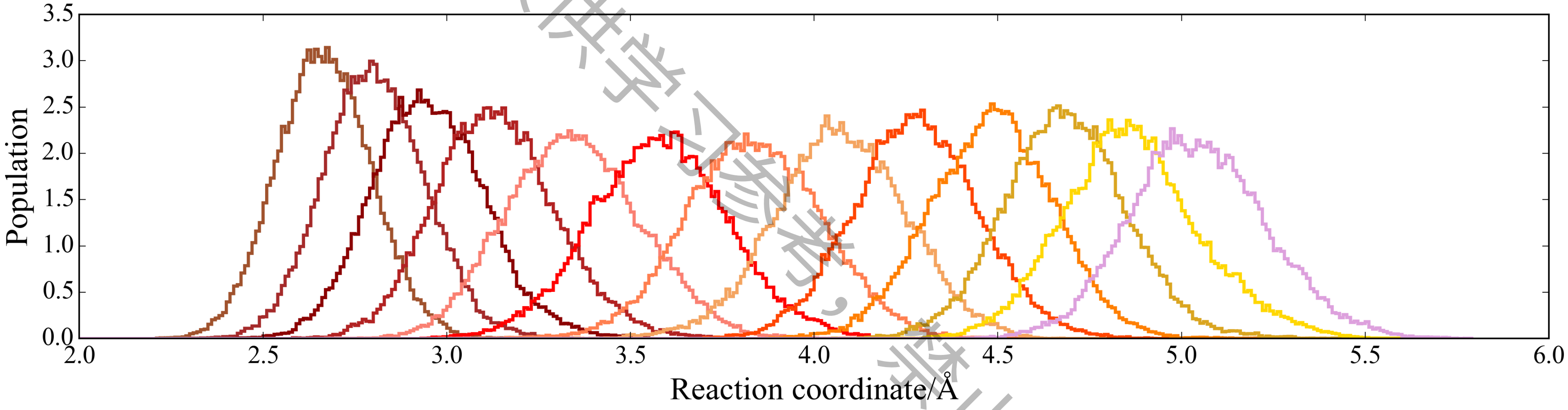


Kastner, J. and W. Thiel. (2005)*J. Phys. Chem*. 123:144104-144105.

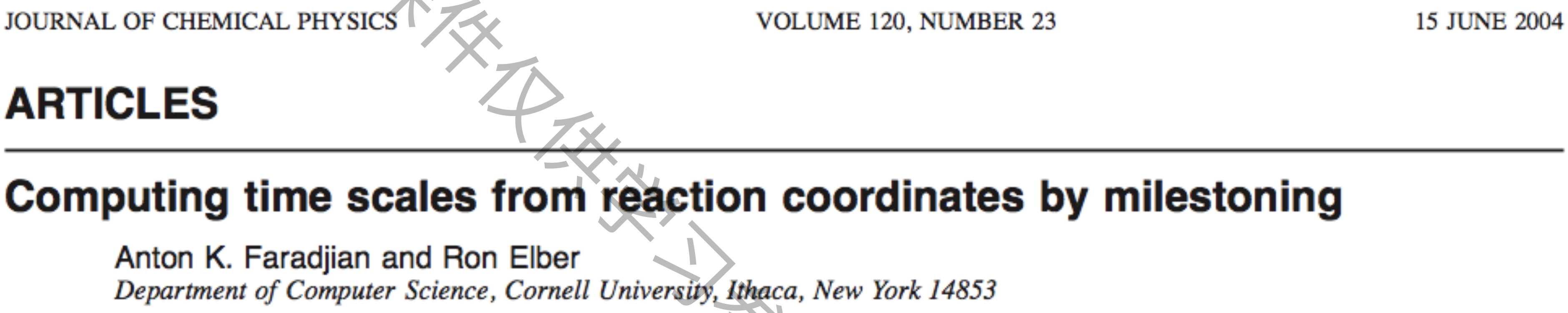
In reaction dynamics, lots of bins are needed to ensure accurate energy. (energy change 1kcal/mol, the rate changes about 10 times at room temperature.)

Bins: the number of bins used to construct histograms

The confidence of the UI assumption



Anything can help to reduce computing time?



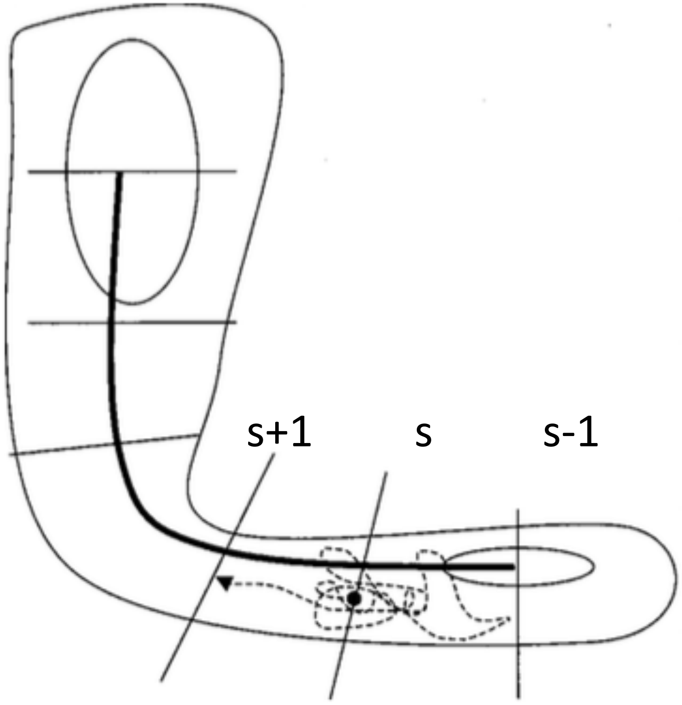
JOURNAL OF CHEMICAL PHYSICS VOLUME 120, NUMBER 23 15 JUNE 2004

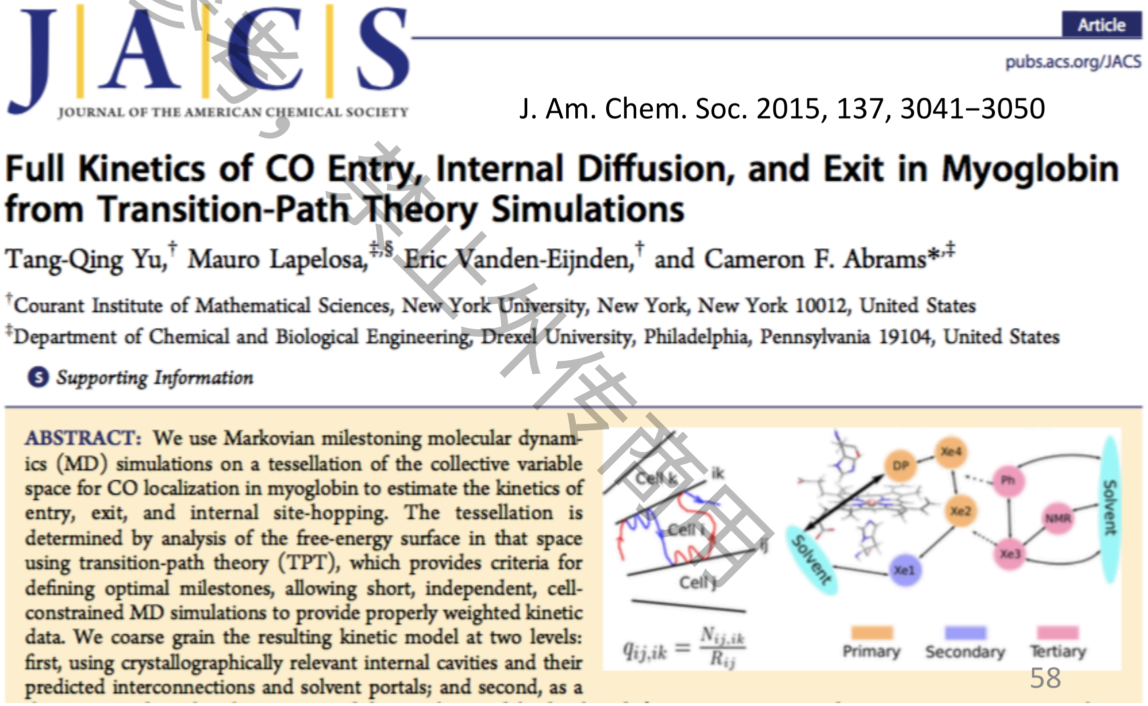
ARTICLES

Computing time scales from reaction coordinates by milestoning

Anton K. Faradjian and Ron Elber

Department of Computer Science, Cornell University, Ithaca, New York 14853





J|A|C|S JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

J. Am. Chem. Soc. 2015, 137, 3041−3050

Full Kinetics of CO Entry, Internal Diffusion, and Exit in Myoglobin from Transition-Path Theory Simulations

Tang-Qing Yu,[†] Mauro Lapelosa,[‡,§] Eric Vanden-Eijnden,[†] and Cameron F. Abrams*[,‡]

[†]Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States

[‡]Department of Chemical and Biological Engineering, Drexel University, Philadelphia, Pennsylvania 19104, United States

Supporting Information

ABSTRACT: We use Markovian milestoning molecular dynamics (MD) simulations on a tessellation of the collective variable space for CO localization in myoglobin to estimate the kinetics of entry, exit, and internal site-hopping. The tessellation is determined by analysis of the free-energy surface in that space using transition-path theory (TPT), which provides criteria for defining optimal milestones, allowing short, independent, cell-constrained MD simulations to provide properly weighted kinetic data. We coarse grain the resulting kinetic model at two levels: first, using crystallographically relevant internal cavities and their predicted interconnections and solvent portals; and second, as a

Other methods for Free energy

ABF

Metadynamics

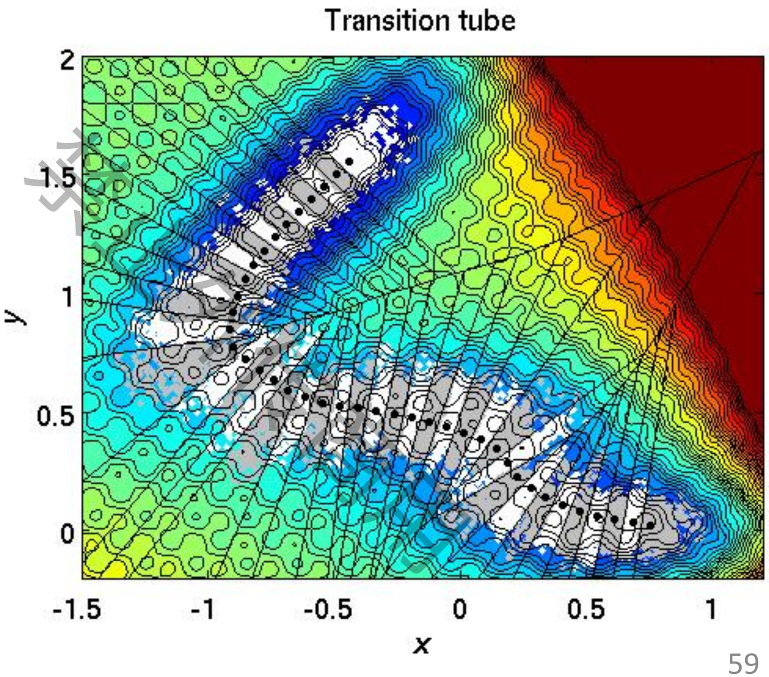
String method: Eric Vanden-Eijnden, Benoit Roux, Martin Karplus, Weinan E

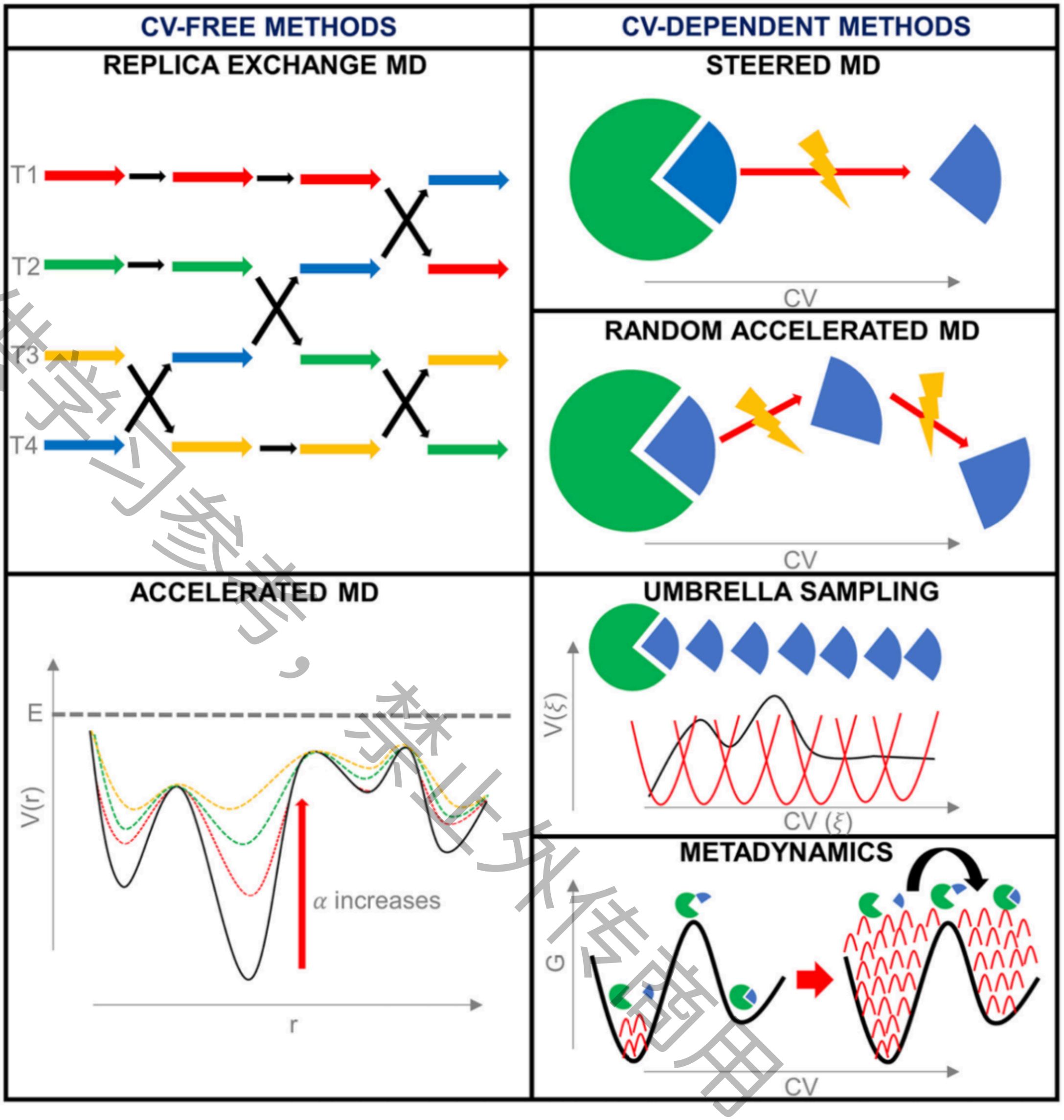
Nudged Elastic Band

ITS

Transition Path Theory

Replica Exchange





Frontiers in Pharmacology, **2018**, *9*, 923

aMD (accelerated MD)

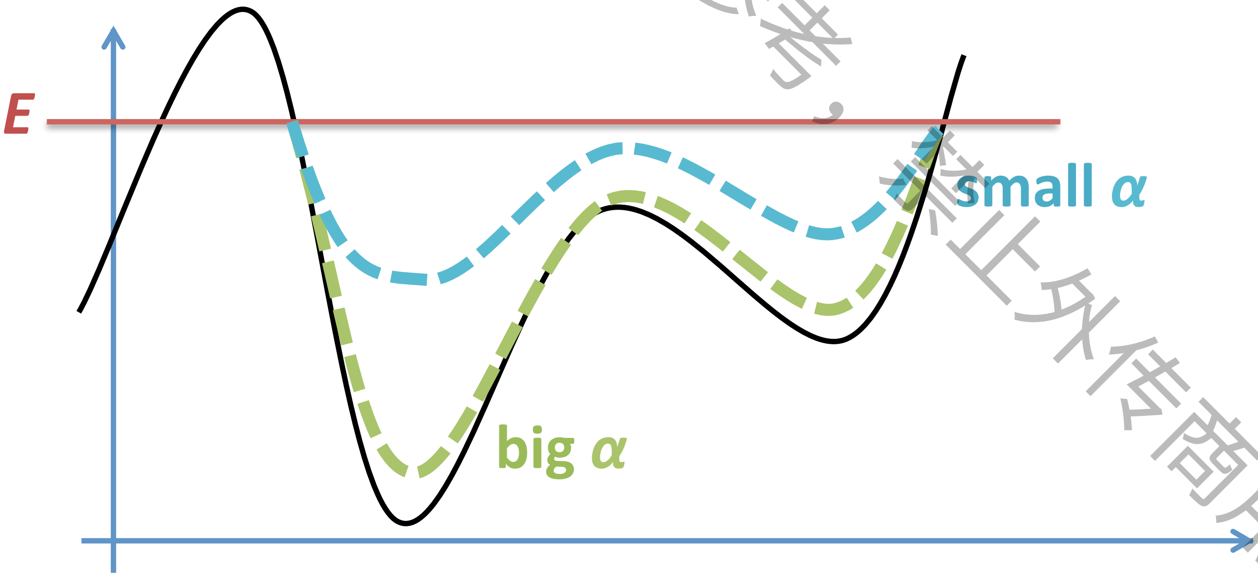
$$V(x;\alpha,E)=\begin{cases} V(x) & V(x)\geq E \\ V(x)+\dfrac{(E-V(x))^2}{\alpha+(E-V(x))} & V(x)<E \end{cases}$$

$$F=-k_{\mathrm{B}}T\ln\left[P^0(x)\right]$$

$$P^0(x)\propto e^{-\beta V(x)}$$

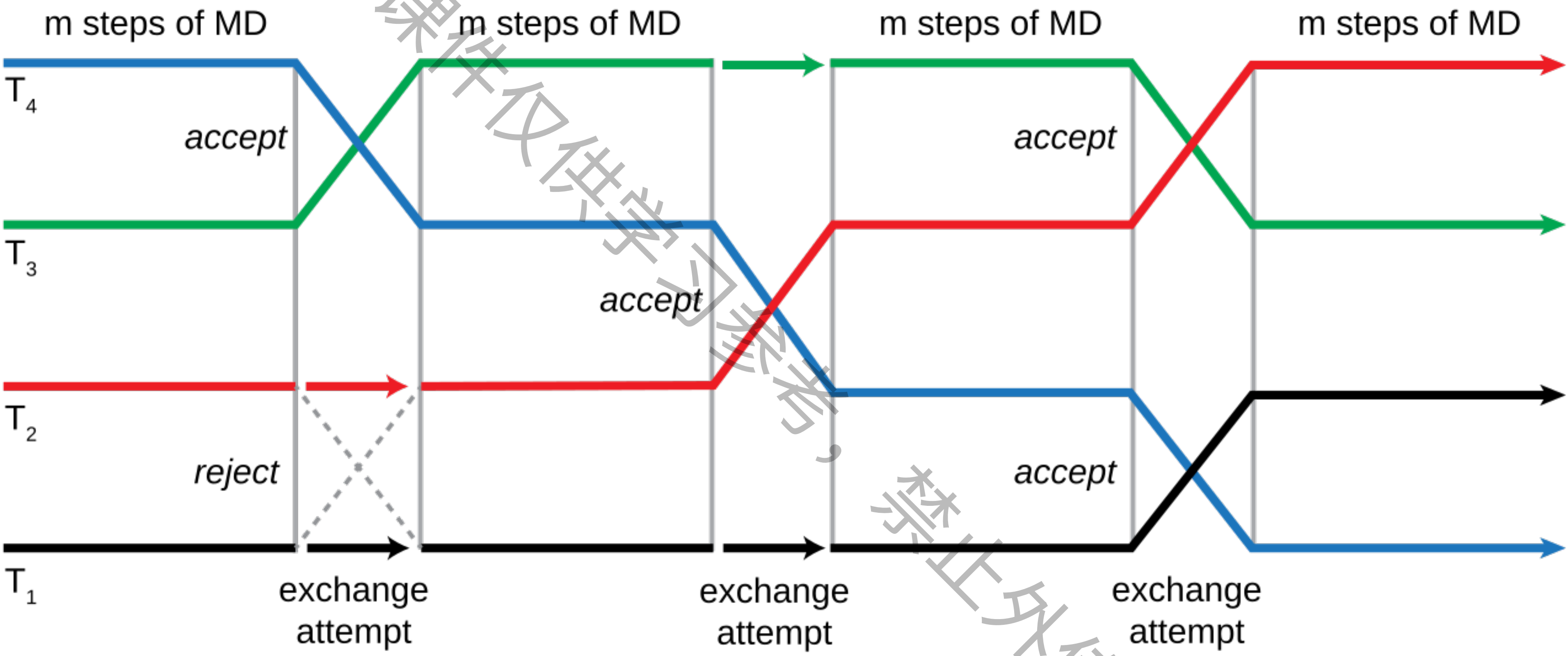
$$P(x)\propto e^{-\beta V(x;\alpha,E)}$$

$$\Rightarrow P^0(x)=P(x)e^{\beta\left[V(x;\alpha,E)-V(x)\right]}$$



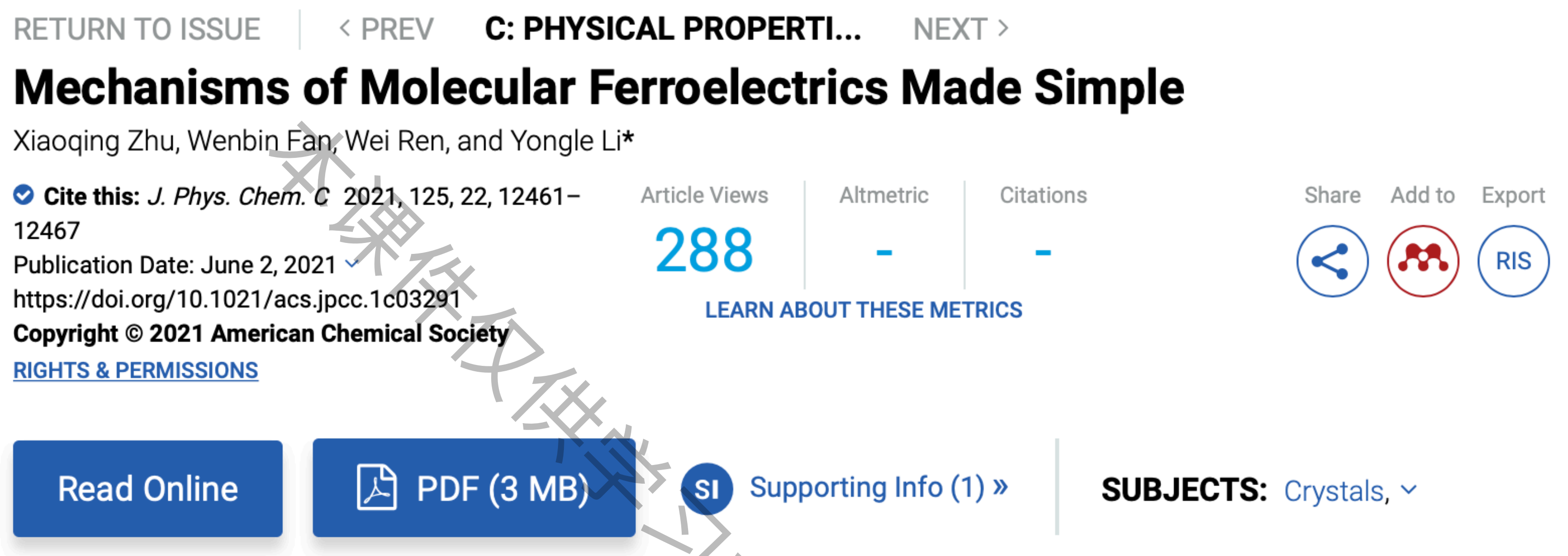
Song, J., Li, Y., Ji, C. *et al*. *Sci Rep* **2015**, *5*, 7906. https://doi.org/10.1038/srep07906

REMD



$$w = \begin{cases} 1, & \Delta \le 0 \\ e^{-\Delta}, & \Delta \ge 0 \end{cases}$$

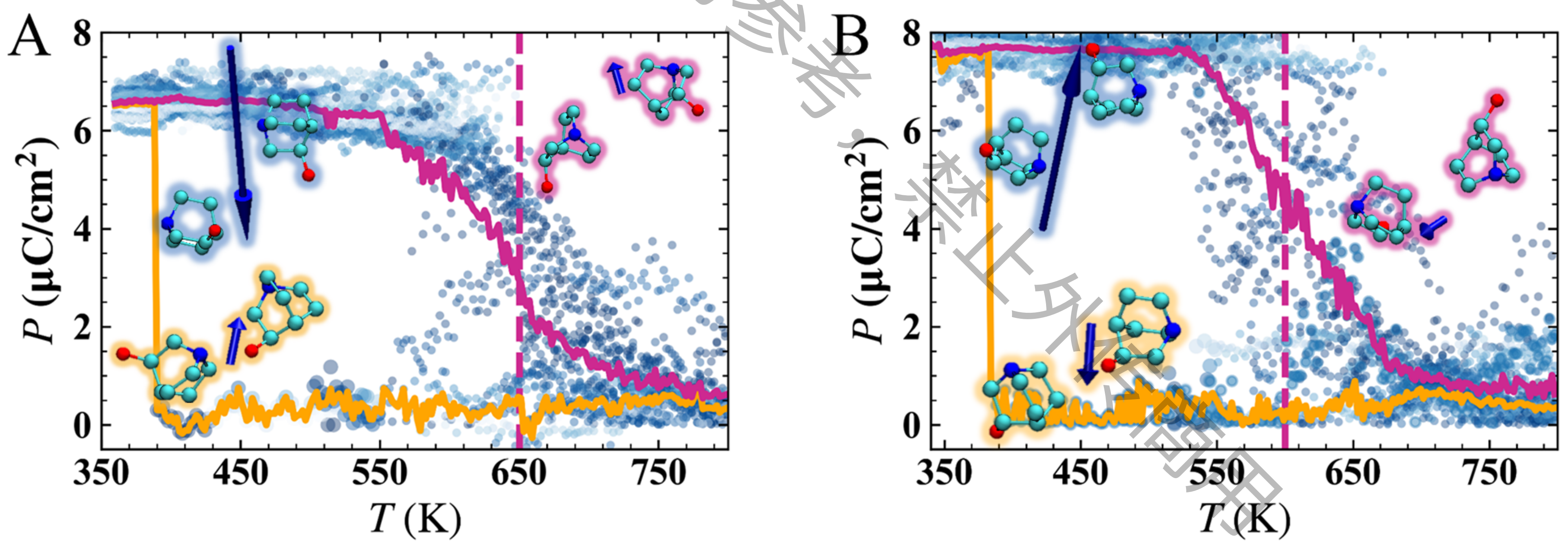
$$\Delta = \left(\frac{1}{k_B T_j} - \frac{1}{k_B T_i} \right) \left(E(q^{[i]}) - E(q^{[j]}) \right)$$



Mechanisms of Molecular Ferroelectrics Made Simple

Xiaoqing Zhu, Wenbin Fan, Wei Ren, and Yongle Li*

Cite this: *J. Phys. Chem. C* 2021, 125, 22, 12461–12467
Publication Date: June 2, 2021
https://doi.org/10.1021/acs.jpcc.1c03291



铁电相变机理

- 机器学习方法：主成分分析(PCA)

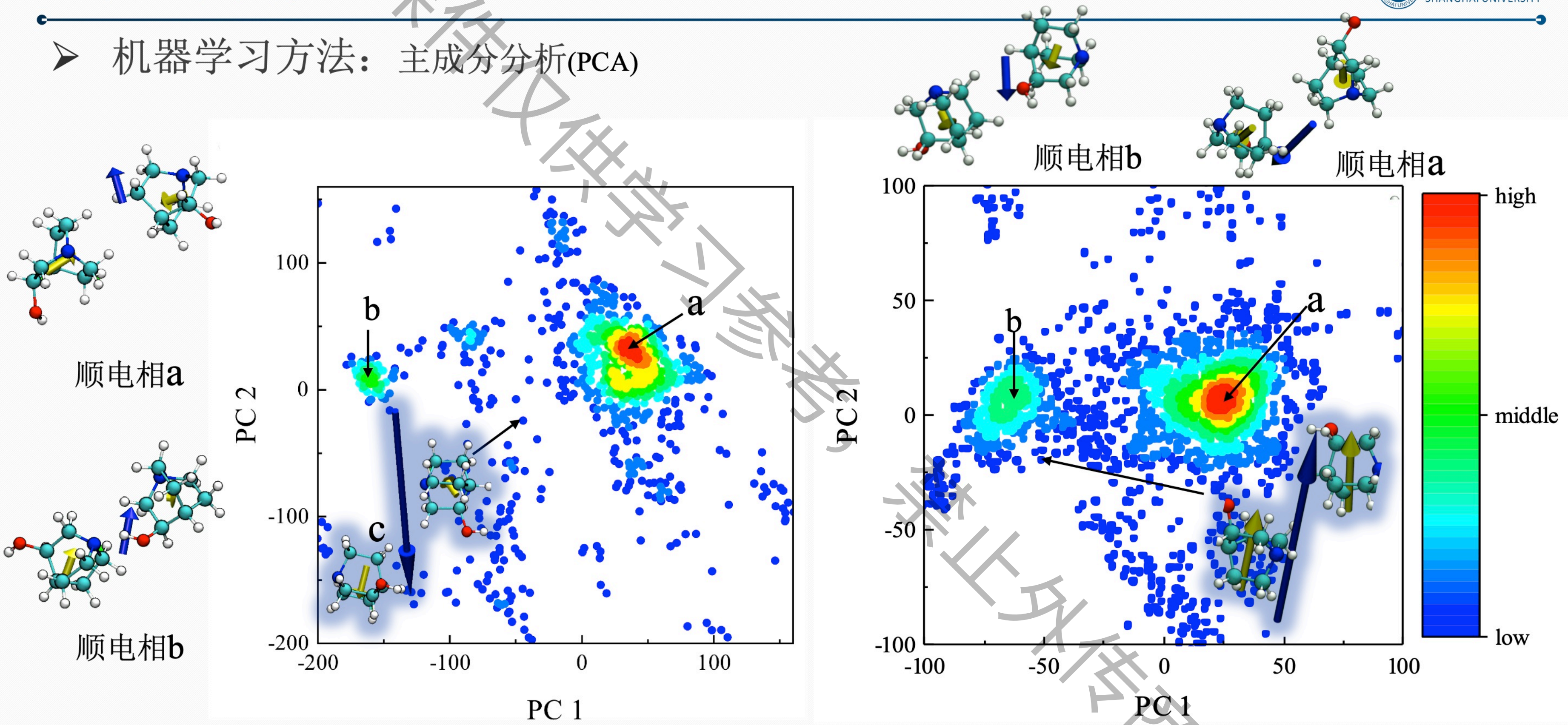


图 11：(*R*)/(*S*)-3-奎宁环醇晶体的相变主成分分析。(a)和(b)为两顺电相， (c)为铁电相。

- 自由能计算: 多态Bennett接受率法(multistate Bennett Acceptance ratio, MBAR)方法

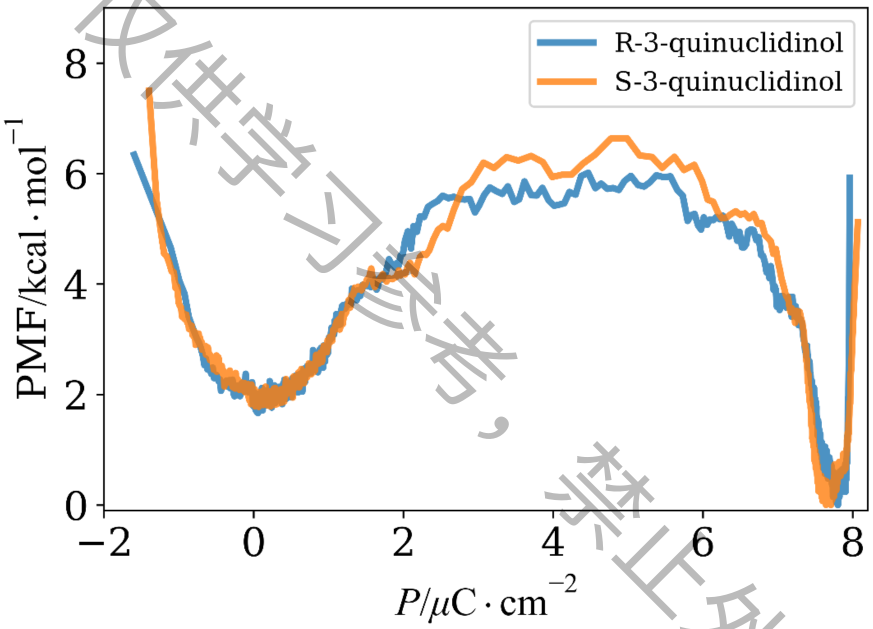


图　13：在(*R*)/(*S*)-3-奎宁环醇的第一个相变居里温度380 K(*R*)-/385 K(*S*)-下，通过多态Bennett接受率法(MBAR)计算得到的平均力势

1. Liu, W.; Li, P.; Mei, Y., *Journal of Molecular Modeling* **2019**, *25*, 187.
2. Shirts, M. R.; Chodera, J. D., *The Journal of Chemical Physics* **2008**, *129*, 124105.

Metadynamics

nature reviews physics

Explore our content ⌄ Journal information ⌄

nature › nature reviews physics › technical review › article

Technical Review | Published: 06 March 2020

Using metadynamics to explore complex free-energy landscapes

Giovanni Bussi & Alessandro Laio ✉

$$V(x,t)=\int_0^t d\tau\,\omega e^{-[x-x(\tau)]^2/(2\sigma^2)}$$

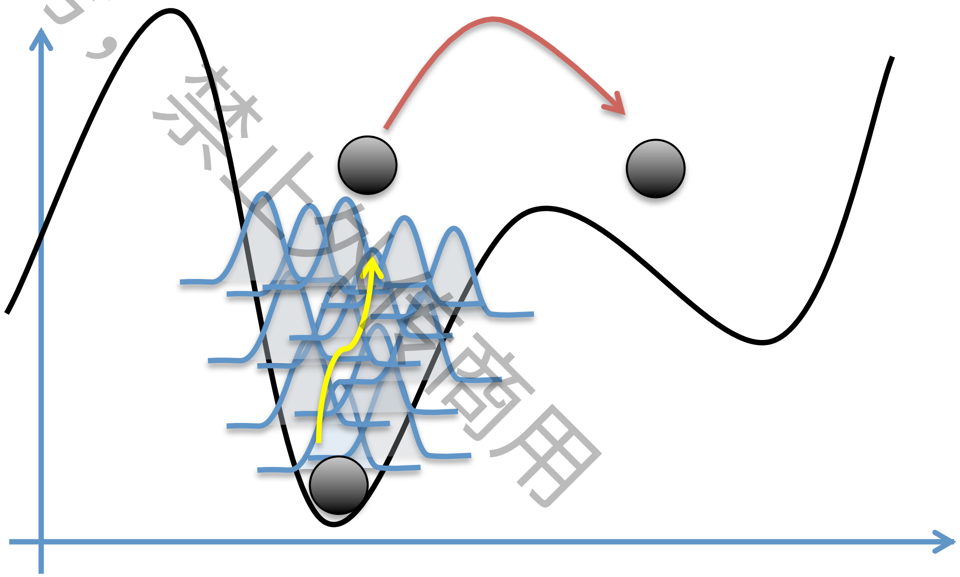


Table 2 | **Availability of metadynamics in commonly used MD codes**

MD code	Native	PLUMED	COLVARS	SSAGES
ACEMD[68]	No	Yes	No	No
AMBER[69]	No	Yes	No	No
CP2K[70]	Yes	Yes	No	No
DLPOLY[71]	No	Yes	No	No
DESMOND[72]	Yes	No	No	No
GROMACS[73]	No	Yes	Yes	Yes
i-Pi[74]	No	Yes	No	No
HOOMD[75]	No	No	No	Yes
LAMMPS[76]	Yes[a]	Yes	Yes	Yes
NAMD[77]	Yes[b]	Yes	Yes	No
OPENMD[78]	No	No	No	Yes
OPENMM[79]	Yes	Yes	No	No
ORAC[80]	Yes	No	No	No
PINY-MD[81]	No	Yes	No	No
QUANTUM-ESPRESSO[82]	No	Yes	No	No
QBOX[83]	No	No	No	Yes

Metadynamics (2)

Reweighting

Histogram: $N(s,f)$

$t_0 \le t < t_1$ $N_t(s,f)$

$$t == t_1$$

$$\dot{V}(s,t) = \omega(t) e^{-(s-s_1)^2/(2\sigma^2)}$$

$$\omega(t) = \omega_0 e^{-V(s_1,t)/\Delta T}$$

$$\dot{c}(t) = -\left\langle \dot{V}(s,t) \right\rangle$$

Here *f* is a variable *f*=*f*(***R***). The *f*(***R***)'s unbiased distribution *P*(*f*) is what we want.

$$N_t(s) = \frac{\sum_f N_t(s,f)}{\sum_{s,f} N_t(s,f)}$$

$$N_{t+\Delta t}(s) = e^{-\beta(\dot{V}(s,t)+\dot{c}(t))\Delta t} N_t(s)$$

$$t_1 \le t < t_2$$

$$P(f) = \frac{\sum_s e^{\beta V(s,t)} N_t(s,f)}{\sum_{s,f} e^{\beta V(s,t)} N_t(s,f)}$$

J. Comput. Chem. **2009**, *30*, 1615–1621

Metadynamics (3)

$$c(t) = \int ds \left[e^{\beta V(s,t+\Delta t)} - e^{\beta V(s,t)} \right]$$

$$\left\langle O(\boldsymbol{R}) \right\rangle_0 = \left\langle O(\boldsymbol{R}) e^{\beta \left[V(s[\boldsymbol{R}],t) - c(t) \right]} \right\rangle$$

Once set:

$$O(\boldsymbol{R}) = \delta(s[\boldsymbol{R}] - s^*)$$

we obtain $F(s^*)$

J. Phys. Chem. B **2015**, *119*, 736−742