



Calculation of temperature dependence of free energy caused by potential function changes

Mitsunori Ikeguchi^{*}, Seishi Shimizu, Koichi Tazaki, Shugo Nakamura,
Kentaro Shimizu

Department of Biotechnology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

Received 24 December 1997

Abstract

We propose a method to calculate the temperature dependence of free energy caused by potential function changes by combining λ dynamics with a method to obtain the potential of mean force (the weighted histogram analysis method) in which the free energy at any temperature can be evaluated from one or a few simulations. This method not only allows us to obtain the temperature dependence of free energy but it also has a built-in estimate for sampling errors. The temperature dependence of hydrophobic hydration at 20°C–80°C was successfully reproduced from two or four simulations. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

The temperature dependence of free energy is of great importance in the analysis of various molecular phenomena, such as protein folding, intermolecular binding, and hydrophobic hydration. For example, the temperature dependence of hydrophobic free energy has characteristics of hydrophobic hydration [1,2]: that is, the hydrophobic effect becomes stronger at higher temperatures. The temperature dependence of free energy of protein folding gives us insights into the protein folding mechanism [2].

Calculation of the temperature dependence of free energy is usually done by performing independent simulations at different temperatures, because the previously proposed methods to calculate the free energy were only for the simulation at constant

temperature [3–9]. However, the calculation of free energy is very expensive, and therefore, simulations at many temperatures are sometimes not practical. In this paper, we propose a method to calculate the free energy at any temperature from simulations at one or a few temperatures. We focus on the free energy caused by potential function changes (Such types of free energy are called “ λ type” free energy in this paper. λ is the coupling parameter which represents the pathway to change the potential from A to B .) The λ type free energy calculation is used for the solute-insertion process (pure solvent \rightarrow solute + solvent), amino-acid mutation in proteins [10,11] (e.g. alanine \rightarrow glycine) and so on.

Our method is an extension of the recently proposed free-energy calculation method, namely λ dynamics [12]. Previous λ dynamics calculation has been restricted to a single temperature and it has not been applied to calculating the temperature depen-

^{*} Corresponding author.

dence of free energy. We extended this method to calculate the temperature dependence of free energy. In λ dynamics, by extending the molecular coordinates to include the coupling parameter λ , λ type free energy calculation is transformed to a calculation of the potential of mean force (PMF). PMF is the free energy along the reaction coordinates derived from molecular coordinates. Examples of reaction coordinates are a dihedral angle, the length between two atoms, and the gyration radius of a given molecule. Recent developments in PMF methods including the multicanonical algorithm [13–21] and the weighted histogram analysis method (WHAM) [22–28] have provided the PMF at any temperature from one or a few simulations. Previous λ dynamics only utilized the constant-temperature PMF calculation and did not aim at the temperature dependence of free energy. We combined the multi-temperature facility of PMF methods with λ dynamics, which has enabled us to calculate the λ type free energy at any temperature from one or a few simulations.

This paper is organized as follows: In Section 2, we describe λ dynamics and a multi-temperature version of WHAM. In Section 3, as an application of our method, the solvation of methane into water is demonstrated and the utility of our method is discussed.

2. Methods

First, let us briefly review λ dynamics. In λ dynamics, λ is moved during simulation as well as molecular coordinates \mathbf{R} . In Monte Carlo simulation, when attempts are made to move a molecular coordinate, the coupling parameter is also changed. In molecular dynamics simulation, the coupling parameter λ is moved by the following equation of motion:

$$m_\lambda \ddot{\lambda} = - \frac{\partial U(\{\mathbf{R}, \lambda\})}{\partial \lambda} \quad (1)$$

where $U(\{\mathbf{R}, \lambda\})$ is the potential function in state λ , and m_λ is a pseudo weight of the coupling parameter λ . m_λ can be set to an arbitrary value because the momentum part in the partition function is analytically integrated in the canonical ensemble. Once the coupling parameter λ is considered to be one of the molecular coordinates, the free energy difference ΔF

between $\lambda = 0$ and $\lambda = 1$ can be given in the PMF form as,

$$\Delta F = -kT(\ln \rho_\lambda(1) - \ln \rho_\lambda(0)) \quad (2)$$

where $\rho_\lambda(\lambda)$ is the probability distribution of λ , k is the Boltzmann constant, and T is the absolute temperature. Only the probability distribution of λ needs to be calculated to obtain the free energy difference.

Various PMF methods can be used to calculate $\rho_\lambda(\lambda)$. Recently, some methods which provide the PMF at any temperature from one or a few simulations have been proposed. WHAM provides the best possible estimate of PMF by combining the results of multiple runs and it has a built-in estimate for sampling errors. Previous λ dynamics calculation has also been combined with WHAM, but it is limited to constant temperature simulation. The temperature dependence of free energy was not aimed at in the work. We propose a method to calculate the temperature dependence of free energy caused by potential function changes, by combining the multi-temperature facility of WHAM with λ dynamics.

A multi-temperature version of WHAM is described below. Refer the Kumar et al. for the detail [22].

Consider R simulations with the i th simulation carried out at temperature $T_i = 1/k\beta_i$ with the constraint potential $E_i^c(\lambda)$, which limits the distribution of λ within the range $[0,1]$. Let the number of sampling configurations from the i th simulation be n_i . Then obtain histogram $N_i(E, \lambda)$ at energy E and coupling parameter λ from the configurations. Then the probability histogram $P(\beta, E, \lambda)$ is given by

$$P(\beta, E, \lambda) = \frac{\sum_i^R N_i(E, \lambda) / g_i}{\sum_i^R n_i \exp[\beta E - \beta_i E - \beta_i E_i^c(\lambda)] / (Z_i g_i)} \quad (3)$$

$$Z_i = \sum_{E, \lambda} P(\beta, E, \lambda) \exp[\beta E - \beta_i E - \beta_i E_i^c(\lambda)] \quad (4)$$

where $T = 1/k\beta$ is any desired temperature, g_i is defined by $1 + 2\tau_i$ and τ_i is an integrated correlation time [22]. In WHAM, Eqs. (3) and (4) are iteratively calculated until $P(\beta, E, \lambda)$ and Z_i con-

verge. Then, from $P(\beta, E, \lambda)$, the distribution of λ can easily be obtained by

$$\rho_\lambda(\beta, \lambda) = \sum_E P(\beta, E, \lambda). \quad (5)$$

The free energy $\Delta F(\beta)$ at an arbitrary temperature is given by

$$\Delta F(\beta) = -kT(\ln \rho_\lambda(\beta, 1) - \ln \rho_\lambda(\beta, 0)). \quad (6)$$

The central importance of the multi-temperature facility of WHAM is the histogram of E , which enables us to reweight the probability of E at an arbitrary temperature.

The deviation $\delta P(\beta, E, \lambda)$ is given by

$$\frac{\delta P(\beta, E, \lambda)}{P(\beta, E, \lambda)} = \left[\sum_i N_i(E, \lambda) / g_i \right]^{-1/2}. \quad (7)$$

The deviation is used to estimate sampling errors.

The relationship with multicanonical simulation needs to be noted. The constraint potential $E_i^c(E)$ of system energy can be introduced into WHAM so that the distribution of E becomes constant. This is expected to improve the efficiency of sampling of energy. In such a case, equation,

$$P(\beta, E, \lambda) = \frac{\sum_i^R N_i(E, \lambda) / g_i}{\sum_i^R n_i \exp[\beta E - \beta_i E - \beta_i E_i^c(\lambda) - \beta_i E_i^c(E)] / (Z_i g_i)} \quad (8)$$

is used instead of Eq. (3). According to the results of multicanonical simulations [17–21], the constraint potential provides a more efficient sampling of configurational space. The introduction of the constraint potential $E_i^c(E)$ into our method will be examined in a forthcoming paper.

As an application of this method, we calculated the solvation free energy of methane into water using Monte Carlo simulations under the following conditions: We used the TIP3P water models [29] and OPLS parameters for methane [30]. The number of water molecules was 216. A periodic boundary condition was applied, and an NVT ensemble was generated. The methane–water potential function,

$$U_{\text{me-w}} = \lambda^2 \varepsilon_{\text{me-w}} \left\{ \left(\frac{\lambda r_{\text{me-w}}}{r} \right)^{12} - \left(\frac{\lambda r_{\text{me-w}}}{r} \right)^6 \right\} \quad (9)$$

was used for the coupling parameter λ . Parameters $\varepsilon_{\text{me-w}}$ and $r_{\text{me-w}}$ are OPLS parameters. The coefficient of $\varepsilon_{\text{me-w}}$ was λ^2 instead of λ is to avoid the large negative potential function. For the constraint potential $E_i^c(\lambda)$ in the i th simulation, the following harmonic function was used:

$$E_i^c(\lambda) = k_\lambda (\lambda - \lambda_i)^2 \quad (10)$$

where $k_\lambda = 300$ kcal/mol and $\lambda_i = i \times 0.1$. The constraint parameter k_λ was determined so that the distributions of λ overlapped the neighboring simulations. Thirteen simulations were done for one set of WHAM calculations. In each simulation, at least one million configurations for equilibration and four to eight million configurations for sampling were generated. Every 434 configurations, the energy and the value of λ were recorded for WHAM calculation.

3. Results and discussion

Fig. 1 shows the free energy profile $F(\lambda)$ at 25°C, which is defined by

$$F(\lambda) = -kT \ln \rho_\lambda(\lambda). \quad (11)$$

The solvation free energy of methane is the difference between $F(1)$ and $F(0)$, and the value is 1.93 ± 0.17 kcal/mol. This value agrees well with the experimental value of 1.93 kcal/mol [31]. In Fig. 1, the error-bar represents the deviation in free energy

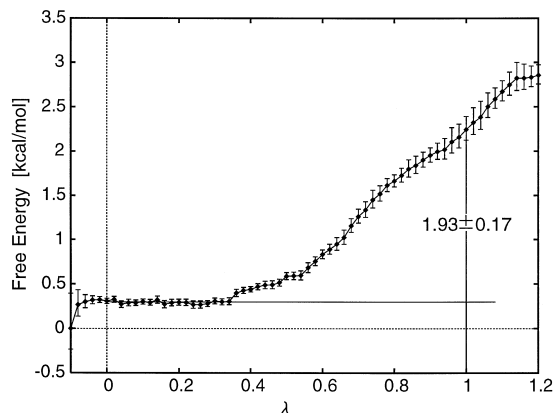


Fig. 1. The free energy profile in the solvation of methane into water at 25°C. The error-bars represent the deviation in the free energy. The free energy difference between the λ of 0 and 1 is the solvation free energy of methane.

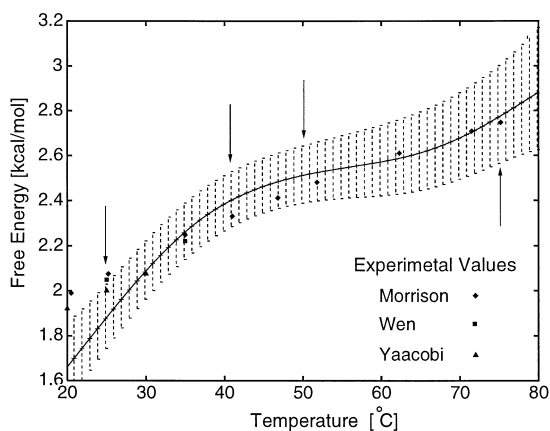


Fig. 2. The solvation free energies of methane at temperatures from 20°C to 80°C calculated from four simulations at 25°C, 40°C, 50°C and 75°C. The error-bars represent the deviation in the free energy. The arrows in the figure indicate the temperatures at which the simulations were performed.

at each λ as an estimate of the sampling error. One of advantages of this method is that it is possible to obtain such statistical error in free energy.

Fig. 2 shows the solvation free energies from 20°C to 80°C. The results were obtained from only four simulations at 25°C, 40°C, 50°C and 75°C. The experimental values by Morrison and Billett [32], Wen and Hung [33], Yaacobi and Ben-Naim [34] can be seen within the deviation in free energy. We emphasize that the generation of the temperature

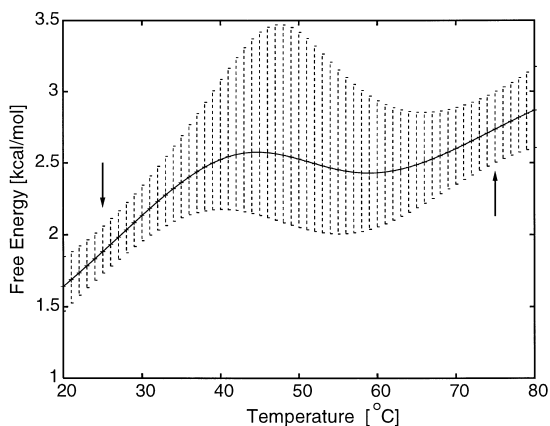


Fig. 3. The solvation free energies of methane at temperatures from 20°C to 80°C calculated from two simulations at 25°C and 75°C. The error-bars represent the deviation in the free energy. The arrows in the figure indicate the temperatures at which the simulations were performed.

dependence of free energy is based on statistical mechanics, not on the simple interpolation of data.

To test the multi-temperature facility of this method, we calculated the free energies at 20–80°C from only two simulations. Fig. 3 shows the results from simulations at 25°C and 75°C. Even from two simulations, the results qualitatively agree with the experiments. However, the deviations are large at temperatures from 40°C to 60°C. Comparing Fig. 2 with Fig. 3, the statistical errors at 40–60°C were greatly improved by adding 40°C and 50°C simulations. This is due to the improvement in the sampling of energy. In this way, the statistical error of free energy can be improved by adding simulations for a range in which the statistical error is large. We emphasize that this facility can give the guideline for efficient improvement of sampling by simulations.

4. Conclusion

We proposed a method to calculate the temperature dependence of free energy caused by potential function changes by combining λ dynamics with the PMF method (the weighted histogram analysis method) by which the free energy at any temperature can be evaluated from one or a few simulations. This method enables not only the temperature dependence of free energy to be obtained but also it has a built-in estimate for sampling errors. We demonstrated a calculation of the temperature dependence of hydrophobic hydration. We now plan to apply this method to estimating the stability of proteins or nucleic acids.

References

- [1] W. Kauzmann, *Adv. Protein Chem.* 14 (1959) 1.
- [2] P.L. Privalov, S.J. Gill, *Adv. Protein Chem.* 39 (1988) 191.
- [3] P.A. Kollman, *Chem. Rev.* 93 (1993) 2395.
- [4] W.L. Jorgensen, J.F. Blake, J.K. Buckner, *Chem. Phys.* 129 (1989) 193.
- [5] D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.* 90 (1989) 2460.
- [6] D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.* 91 (1989) 7831.
- [7] D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.* 94 (1991) 4532.
- [8] D.M. Ferguson, D.A. Pearlman, W.C. Swope, P.A. Kollman, *J. Comput. Chem.* 13 (1992) 362.

- [9] C. Chipot, P.A. Kollman, D.A. Pearlman, *J. Comput. Chem.* 17 (1996) 1112.
- [10] R. Tanimura, M. Saito, *Mol. Sim.* 16 (1996) 75.
- [11] P.A. Kollman, *Curr. Opin. Struct. Biol.* 4 (1994) 240.
- [12] X. Kong, C.L. Brooks III, *J. Chem. Phys.* 105 (1996) 2414.
- [13] B.A. Berg, T. Neuhaus, *Phys. Lett. B* 267 (1991) 249.
- [14] B.A. Berg, T. Neuhaus, *Phys. Rev. Lett.* 68 (1002) 9.
- [15] B.A. Berg, U. Hansmann, T. Neuhaus, *Phys. Rev. B* 47 (1993) 497.
- [16] J. Lee, *Phys. Rev. Lett.* 71 (1993) 211.
- [17] U.H.E. Hansmann, Y. Okamoto, *J. Comput. Chem.* 14 (1993) 1333.
- [18] Y. Okamoto, U.H.E. Hansmann, *J. Phys. Chem.* 99 (1995) 11276.
- [19] A. Kidera, *Proc. Natl. Acad. Sci. USA* 92 (1995) 9886.
- [20] U.H.E. Hansmann, Y. Okamoto, F. Eisenmenger, *Chem. Phys. Lett.* 259 (1996) 321.
- [21] N. Nakajima, H. Nakamura, A. Kidera, *J. Phys. Chem. B* 101 (1997) 817.
- [22] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, *J. Comput. Chem.* 13 (1992) 1011.
- [23] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, *J. Comput. Chem.* 16 (1995) 1339.
- [24] S. Kumar, P.W. Payne, M. Vasquez, *J. Comput. Chem.* 17 (1996) 1269.
- [25] A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* 61 (1988) 2635.
- [26] A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* 63 (1989) 1195.
- [27] E.M. Boczko, C.L. Brooks III, *J. Phys. Chem.* 97 (1993) 4509.
- [28] E.M. Boczko, C.L. Brooks III, *Science* 269 (1989) 393.
- [29] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *J. Chem. Phys.* 79 (1983) 926.
- [30] W.L. Jorgensen, J.D. Madura, C.J. Swenson, *J. Am. Chem. Soc.* 106 (1984) 6638.
- [31] A. Ben-Naim, Y. Marcus, *J. Chem. Phys.* 81 (1984) 2016.
- [32] T.J. Morrison, F. Billett, *J. Chem. Soc.* (1952) 3819.
- [33] W.-Y. Wen, J.H. Hung, *J. Phys. Chem.* 74 (1970) 170.
- [34] M. Yaacobi, A. Ben-Naim, *J. Phys. Chem.* 78 (1974) 175.