

Supporting Text

Potential of Mean Effective Force. For a given separation r between two solutes, the potential of mean force (PMF) describes an ensemble average over directions between the solutes and over all the conformations of the surrounding solvent molecules. We will show that the free energy change of the system *does not* equal the change in the PMF, but they can be related by a simple formula.

In a liquid of spherically symmetric molecules, let \mathbf{r}_1 and \mathbf{r}_2 denote the positions of the solutes, and $\mathbf{r}_3, \dots, \mathbf{r}_N$ denote the positions of solvent molecules. Let $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$, $\beta = (k_B T)^{-1}$, and the Hamiltonian be $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) = \mathcal{K}(\mathbf{p}^N) + \Phi(\mathbf{r}^N)$. Note that r_{12} is a reaction coordinate (often denoted by $\xi(\mathbf{r}^N)$ (1)), which is itself a function of the configuration coordinates \mathbf{r}^N . The PMF is given by $w(r) = -k_B T \ln g(r)$, where $g(r)$ is the radial distribution function,

$$g(r) = g(\mathbf{r}_1, \mathbf{r}_2) = g(|\mathbf{r}_1 - \mathbf{r}_2|). \quad (10)$$

$g(\mathbf{r}_1, \mathbf{r}_2)$ is defined as (2)

$$P(\mathbf{r}_1, \mathbf{r}_2) = \rho^2 g(\mathbf{r}_1, \mathbf{r}_2) \quad (11)$$

where ρ is the solute density. $P(\mathbf{r}_1, \mathbf{r}_2)$ is the probability that any solute is in $d\mathbf{r}_1$ and another is in $d\mathbf{r}_2$ (hence a factor of 2 in the next equation).

$$g(\mathbf{r}_1, \mathbf{r}_2) = \rho^{-2} \cdot 2 \cdot \frac{\int e^{-\beta\Phi(\mathbf{r}^N)} d\mathbf{r}_3 \cdots d\mathbf{r}_N}{\int e^{-\beta\Phi(\mathbf{r}^N)} d\mathbf{r}_1 \cdots d\mathbf{r}_N}. \quad (12)$$

Of crucial importance is the normalization condition for $g(r)$. By definition,

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = 1. \quad (13)$$

We can also define a probability density function $P(r)$ which has a different normalization condition,

$$P(r) = \frac{\rho^2}{2} \int g(\mathbf{r}_1, \mathbf{r}_2) \delta(|\mathbf{r}_1 - \mathbf{r}_2| - r) d\mathbf{r}_1 d\mathbf{r}_2. \quad (14)$$

The prefactor is chosen such that $\int_0^\infty P(r) dr = 1$. The probability density is used in the constrained molecular dynamics method by Sprik and Ciccotti (1), where they denote the free energy W for a given value $\xi = \xi'$ as

$$W(\xi') = -k_B T \ln P_\xi(\xi') \quad (15)$$

$$P_\xi(\xi') = \langle \delta(\xi(\mathbf{r}) - \xi') \rangle. \quad (16)$$

For the special case here, we can equate $\xi' = r$, $\xi(\mathbf{r}) = r_{12}$, and the free energy for a given value $r_{12} = r$ is given by the function $W(r)$. In contrast, the PMF, $w(r)$, is different from $W(r)$ because $P(r) = \rho \cdot 4\pi r^2 g(r)$.

$W(r)$ and $w(r)$ can be related as follows. First, we start with the potential $W(r)$,

$$W(r) = -k_B T \ln P(r) \quad (17)$$

$$P(r) = \mathcal{N} \int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r), \quad (18)$$

where \mathcal{N} is a normalization factor.

The derivative of $W(r)$ can be expressed as an ensemble average:

$$\frac{dW(r)}{dr} = -\beta^{-1} \frac{\frac{\partial}{\partial r} \int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r)}{\int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r)} \quad (19)$$

$$= -\beta^{-1} \frac{\frac{\partial}{\partial r} \langle \delta(r_{12} - r) \rangle}{\langle \delta(r_{12} - r) \rangle}. \quad (20)$$

We can change variables to relative and center of mass coordinates of two

solutes, $d\mathbf{r}_1 d\mathbf{r}_2 = r_{12}^2 dr_{12} d\Omega d\mathbf{R}$.

$$\begin{aligned} \mathcal{N}^{-1}P(r) &= \int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r) \\ &= \int d\mathbf{r}_3 \cdots d\mathbf{r}_N \int d\Omega d\mathbf{R} r^2 e^{-\beta\Phi(r, \Omega, \mathbf{R}, \mathbf{r}_3, \dots, \mathbf{r}_N)} \end{aligned}$$

Now we can evaluate the numerator in Eq. (20),

$$\begin{aligned} & -\beta^{-1} \frac{\partial}{\partial r} \int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r) \\ &= \int d\mathbf{r}_3 \cdots d\mathbf{r}_N d\Omega d\mathbf{R} \left(\frac{-2}{r\beta} + \frac{\partial\Phi}{\partial r} \right) r^2 e^{-\beta\Phi(r, \Omega, \mathbf{R}, \mathbf{r}_3, \dots, \mathbf{r}_N)} \\ &= \frac{-2}{r\beta} \int d\mathbf{r}^N e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r) + \\ & \quad \int d\mathbf{r}_3 \cdots d\mathbf{r}_N d\Omega d\mathbf{R} \frac{\partial\Phi}{\partial r} r^2 e^{-\beta\Phi(r, \Omega, \mathbf{R}, \mathbf{r}_3, \dots, \mathbf{r}_N)}. \end{aligned}$$

The last term on the RHS can be recast as a *conditional* average:

$$\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \frac{\partial\Phi(r_{12}, \Omega, \mathbf{R}, \mathbf{r}_3, \dots, \mathbf{r}_N)}{\partial r_{12}} e^{-\beta\Phi(\mathbf{r}^N)} \delta(r_{12} - r) \quad (21)$$

Finally, a major result

$$\frac{dW(r)}{dr} = \frac{\frac{-2k_B T}{r} \langle \delta(r_{12} - r) \rangle + \langle \frac{\partial\Phi}{\partial r_{12}} \delta(r_{12} - r) \rangle}{\langle \delta(r_{12} - r) \rangle} \quad (22)$$

$$= \frac{-2k_B T}{r} + \left\langle \frac{\partial\Phi}{\partial r_{12}} \right\rangle_{r_{12}=r}^{\text{cond}}. \quad (23)$$

In Sprik and Ciccoti's treatment of constrained molecular dynamics, they identified the free energy W as the potential of mean force (1). However, the probabilistic definition in Eq. (15) is different from the convention of PMF in classical liquid theory (2). From the definition of $w(r)$, the mean force acting on a solute is

$$-\frac{\partial w(\mathbf{r}_1, \mathbf{r}_2)}{\partial \mathbf{r}_1} = \frac{\int e^{-\beta\Phi} \left(-\frac{\partial\Phi}{\partial \mathbf{r}_1} \right) d\mathbf{r}_3 \cdots d\mathbf{r}_N}{\int e^{-\beta\Phi} d\mathbf{r}_3 \cdots d\mathbf{r}_N}. \quad (24)$$

We can compute $w(r)$ by using Eq. (23):

$$\frac{dw(r)}{dr} = -k_B T \frac{d}{dr} \ln g(r) \quad (25)$$

$$= -k_B T \frac{d}{dr} \ln \frac{P(r)}{4\pi r^2 \rho} \quad (26)$$

$$= -k_B T \left(\frac{d}{dr} \ln P(r) - \frac{2}{r} \right) \quad (27)$$

$$= \frac{dW(r)}{dr} + \frac{2k_B T}{r} \quad (28)$$

$$= \left\langle \frac{\partial \Phi}{\partial r_{12}} \right\rangle_{r_{12}=r}^{\text{cond}}. \quad (29)$$

The constraint force between the solutes is related to the change of the Helmholtz free energy $F(r; T, V, N)$,

$$\bar{f}(r) = -\frac{\partial}{\partial r} F(r; T, V, N) = k_B T \frac{d}{dr} \ln Q(r), \quad (30)$$

where the conditional partition function $Q(r)$ is given by

$$Q(r) = \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)} \delta(r_{12} - r). \quad (31)$$

The conditional partition function $Q(r)$ is proportional to the probability density function $P(r)$, so

$$\bar{f}(r) = k_B T \frac{d}{dr} \ln Q(r) = k_B T \frac{d}{dr} \ln P(r) = -\frac{dW(r)}{dr}, \quad (32)$$

or equivalently,

$$\bar{f}(r) = \frac{2k_B T}{r} - \frac{dw(r)}{dr}. \quad (33)$$

Finally, the PMF (up to a constant) is given by

$$w(r) = \int_{r_0}^r -\bar{f}(r') dr' + \int_{r_0}^r \frac{2k_B T}{r'} dr' \quad (34)$$

$$= W(r) + 2k_B T \ln r + \text{const}, \quad (35)$$

which is Eq. (6) in the paper.

The difference between the PMF and PMEF can be understood qualitatively as follows. When two solutes are constrained to a fixed distance, the entropy S of this two-particle system is reduced. The contribution $T\Delta S$ to the free energy when the two solutes are brought from r_0 to r in an isothermal process is

$$\Delta F(r_0 \rightarrow r) = -T\Delta S(r_0 \rightarrow r) = -k_B T \ln \frac{r^2}{r_0^2}, \quad (36)$$

hence the appearance of $2k_B T \ln r$ in Eq. (35).

Data Collection and Error Analysis. In the constrained MD method, one calculates the constraint force directly, and then integrates the mean force to obtain the PMEF or PMF. Here we examine the statistical error involved in obtaining the mean force. As noted in the main text, the hydration structure around a methane pair is less stable at certain methane-methane separations. This is reflected in the slow convergence of the constraint force at those methane-methane separations. Fig. 8 shows the convergence of the constraint force. Whereas at a separation $r = 5.6 \text{ \AA}$, the time average of the constraint force converges in about 3 ps, the mean force at $r = 4.4 \text{ \AA}$ exhibits larger fluctuations over a longer time scale (Fig. 8(b)). A longer MD simulation shows the mean force at $r = 4.4 \text{ \AA}$ can drift from being repulsive to attractive (Fig. 8(c)). The last column in Table 1 is the RMS constraint force fluctuation, analyzed by the method of the statistical inefficiency (see M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (1989)). This gives a measure of the error in the calculated mean force. The difference in the average force for total simulation time 6 ps and 10 ps is also given in Table 1 as an additional measure of statistical

error.

Solubility Data and Transfer Free Energy. The transfer free energy of a hydrophobic solute from liquid n -alkanes to water can be calculated by Eq. (9) of the paper. Fig. 9 shows the transfer free energy of methane, ethane, ..., decane at room temperature, compiled by Sharp *et al.* (3). The molar volume ratio term in Eq. (9) of the paper is critical to obtain the almost perfect linear relationship shown in Fig. 9. We are interested in the temperature dependence of the transfer free energy. At higher temperatures, due to the low boiling temperature of small n -alkanes (from methane to butane), partition-coefficient data is only available for transfer between gas phase and water (5; 6). Fig. 10 shows that solubility decreases with increasing temperature. Note that the temperature dependence of molar fraction solubility of methane behaves differently from other small n -alkanes. Fig. 11 shows the transfer free energy from gas phase to water at 298 K and 343 K. When temperature increases from 298 K to 343 K, the transfer free energy per unit area increases by 39%.

Force between Two Methanes. In a recent paper, Ikeda and Terakura provided some additional support for the use of the PBE functional in the simulation of water/methane systems (7). They found some of the weak van der Waals forces for methane-water and methane-methane complexes are reproduced within DFT/PBE. Fig. 12 shows the force between 2 methanes in vacuum calculated by PBE-GGA and Lennard-Jones potentials. The force derived from DFT depends on the relative orientations of the two methanes. On the other hand, the force derived from Lennard-Jones potentials can be regarded

as the average over all relative orientations of the two molecules. Since the hydrophobic attractive force in Fig. 1(a) of the paper is of the order of 0.001 au, its magnitude is much larger than the force between 2 methanes in vacuum when $r > 4 \text{ \AA}$.

As shown in Eq. (8), the hydrophobic force between two methanes can be estimated via the surface-tension model. A derivation is given here. Denote the radius of water as r_W , the radius of methane as r_M , and $\rho_0 = r_W + r_M$. When the distance between two methanes r is such that $2r_M < r < 2\rho_0$, the solvent-accessible area is

$$\begin{aligned}\Delta A &= 8\pi\rho_0^2 - 2 \cdot \rho_0^2 \int_0^\theta \sin\theta d\theta \int_0^{2\pi} d\phi \\ &= 8\pi\rho_0^2 - 4\pi\rho_0^2 \cdot (1 - \cos\theta) \\ &= 8\pi\rho_0^2 - 4\pi\rho_0^2 \left(1 - \frac{r/2}{\rho_0}\right).\end{aligned}$$

From the last expression, it can be readily seen that the total available solvent-accessible area is $8\pi\rho_0^2$, the maximum buried area is $4\pi\rho_0^2(1-r_M/\rho_0)$, or 20.8955% of the total available solvent-accessible area. Now, from Eq. (7), the free energy change in the surface-tension model is

$$\Delta G = \sigma \cdot \Delta A,$$

so we can differentiate ΔG with respect to the separation r and obtain the hydrophobic force

$$f = -\sigma \cdot \frac{dA}{dr} = -\sigma \cdot 2\pi\rho_0.$$

References

1. Sprik, M. & Ciccotti, G. (1998) *J. Chem. Phys.* **109**, 7737–7744.
2. McQuarrie, D. A. (1976) *Statistical Mechanics*. (Harper and Row, New York).
3. Sharp, K. A, Nicholls, A, Friedman, R, & Honig, B. (1991) *Biochemistry* **30**, 9686–9697.
4. McAuliffe, C. (1966) *J. Phys. C* **70**, 1267–1275.
5. Clever, H. L & Young, C. L, eds. (1987) *IUPAC Solubility Data Series*. (Pergamon Press, Oxford, England) Vol. 27/28.
6. Lide, D. R, ed. (2003) *CRC Handbook of Chemistry and Physics*. (CRC Press), 84 edition.
7. Ikeda, T. & Terakura, K. (2003) *J. Chem. Phys.* **119**, 6784–6788.