

Free-energy calculations for fluid and solid phases by molecular simulation

Nandou Lu, C. Daniel Barnes, David A. Kofke*

Department of Chemical Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260-4200, USA

Received 15 April 2001; accepted 22 August 2001

Abstract

The ability to compute accurate free energies by molecular simulation is necessary for many applications. Methods for performing these calculations are well established, but are often misused or misinterpreted. We review ways to avoid and uncover systematic errors in free-energy calculations through simple and appropriate analysis of the simulation averages. One particularly difficult area of application of free-energy methods is in the analysis of solid-phase behaviors. Free energies cannot be computed as easily in solids as in liquids, owing to the inability to perform particle insertions into solids without introducing a defect. We describe our attempts to develop new and efficient methods to circumvent this problem, and show how the analysis of free-energy methods can help to understand the behavior of the newly proposed methods. We examine the method and analysis applied to a simple system of hard rods in one dimension, selected because all of the properties being measured by the simulation method can be evaluated analytically, including finite-size effects. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Molecular simulation; Solids; Free-energy perturbation

1. Introduction

Free-energy calculation is an important topic in molecular simulation, and has been widely used in many applications [1–3]. Free-energy perturbation (FEP) is one of the most important and well-established methods. However, it is often misused or misinterpreted. In our previous studies [4–9], we investigated this problem, discussed a guideline of how to choose an appropriate FEP scheme, and provided heuristics for estimating both the systematic and random errors in the calculation.

Free energies in solid phases are particularly difficult to measure by molecular simulation. The problem is that one cannot do a simple particle insertion or growth to measure the chemical potential, which for most intents and purposes is equivalent to measuring the free energy. Insertion of a new particle breaks the crystalline symmetry, so the free-energy measurement becomes that for a defective crystal. We have

* Corresponding author. Tel.: +1-716-645-2911x2209; fax: +1-716-645-3822.
E-mail address: kofke@acsu.buffalo.edu (D.A. Kofke).

recently introduced a new approach to evaluate solid-phase free energies, and we examine the method in light of the FEP analysis we have also advanced.

In Section 2 we review our FEP analysis and its results, and we outline a staged FEP method for a one-dimensional crystal in Section 3. In Section 4 we apply this analysis to the simulation results.

2. Free-energy perturbation analysis

In its simplest form, i.e. the single-stage calculation, the FEP method computes the free-energy difference between two systems while sampling configurations on only one of them [10]. We designate the system governing the sampling of configurations as the *reference*, and the other system as the *target*. Therefore, we have two choices for the direction of the perturbation, depending the system chosen as the reference. We identify the perturbation systems according to their relative values of entropy [6,7], denoting the system with higher entropy as H, and the system with lower entropy as L. For convenience, we always define the difference in a quantity between the H and L systems as the value of the low entropy system minus that of the high entropy system, i.e. (L) – (H). For example, the Helmholtz free energy, potential energy, and entropy differences are defined as $\Delta A = A_L - A_H$, $\Delta U = U_L - U_H$, and $\Delta S = S_L - S_H$, respectively. We also identify the FEP calculation direction according to the sign of the entropy change along the perturbation direction. We call it a ‘generalized insertion’ if the FEP calculation uses the H system as the reference, i.e. the entropy change along the perturbation direction ($S_L - S_H = \Delta S$) is negative. Otherwise, the FEP calculation is a ‘generalized deletion’ if the L system is used as the reference, i.e. the entropy change along the perturbation direction ($S_H - S_L = -\Delta S$) is positive.

Following this notation and definition, the working formula to compute the free-energy difference for an insertion calculation is

$$\exp(-\beta \Delta A) = \langle \exp(-\beta u) \rangle_H \quad (1)$$

and for a deletion calculation

$$\exp(+\beta \Delta A) = \langle \exp(+\beta u) \rangle_L \quad (2)$$

where u is the potential energy difference of a single perturbation trial, $\beta = 1/kT$ with k the Boltzmann’s constant and T is the absolute temperature. The angle brackets $\langle \dots \rangle$ indicate an ensemble average and the subscript, H or L, denotes the simulation reference. In the context of the chemical potential calculation, the perturbation systems differ in the presence of a single molecule, and these two directions of the FEP calculation are commonly known as the Widom insertion [11,12] and Widom deletion [13] methods.

It is well known that the error in insertion and deletion calculations have opposite signs—the insertion calculation overestimates the free-energy difference, while the deletion calculation underestimates it. In our work we have demonstrated why this is so, and we have argued further that the insertion and deletion inaccuracies will often be of very different magnitudes [4–7,9]. The insertion FEP calculation is much more accurate than deletion, and in fact a deletion FEP calculation is almost sure to provide an incorrect result and should be avoided if at all possible. It is worth emphasizing that one can not simply judge the performance of a FEP method by looking only at the *precision* of the simulation results (which is used to suggest confidence limits). Precision is a different statistic than accuracy. An FEP calculation may provide precise, yet inaccurate results. This is especially true in the deletion calculation,

where the free-energy results are typically precise due to the limited configurations being sampled in the simulation. In other words, in deletion FEP calculations, both the result and its confidence limits can be very inaccurate.

The conclusions above are based on a picture that the set of important configurations of the L system sit completely inside those sampled by the H system (consistent with, but not required by, their relative entropies). If this situation does not hold, it becomes necessary to introduce staging methods. In such approaches, intermediates are defined that bridge the two systems of interest. Depending on how the configurations of the H and L systems relate to one another, different staging strategies are appropriate. Ultimately the aim is to always perform FEP calculations on systems that do adhere to the “configuration-space subset” relation described above. For example, in a two stage FEP calculation, if the important configurations of the initial and final stages partially overlap, we should construct the intermediate inside the overlap region, and apply the Bennett’s method; when there is no overlap of the important configurations of the reference and target, we need to construct an intermediate with the important configurations completely covering both those of the initial and final systems, and use umbrella sampling as the suitable technique. These concepts are presented in more detail elsewhere [7].

Even when operating correctly, and performing the FEP calculation in the insertion direction, inaccurate results can still be obtained. The difference from the deletion calculation is that the inaccuracies vanish much more quickly with sampling size when performing the insertion FEP calculation. Our previous studies have found that the entropy difference between the perturbation systems, ΔS , can be used to quantify the inaccuracy of insertion-based free-energy difference calculations. The relevant group is formed with M , the sample size, and our studies suggest that there is a nearly-universal relationship between the FEP inaccuracy and the group $M \exp(\Delta S/k)$ [7,8]. In particular, we find that a value of 100 for $M \exp(\Delta S/k)$ indicates a threshold sample size for an acceptable accuracy level in the free-energy difference. This heuristic is useful in practice to judge the minimal simulation length for a good free-energy accuracy. The reason is that any inaccuracies in ΔS are typically such that $M \exp(\Delta S/k)$ is underestimated, and thus inaccuracies in the free energy will be overestimated.

Our previous study shows that for an optimal precision, the staging should be constructed in a way that the entropy differences in all stages are equal [6]. The findings of analysis above can be used to consider the question of the optimal number of intermediate stages for the most efficient simulation in a multistage FEP calculation. From the accuracy point of view, we find the number of stages should be chosen such that the entropy difference per stage satisfies $\Delta S/k = -1$; however, consideration based on the precision gives $\Delta S/k = -2$. Once the sample size M is reasonably large and the simulation reaches an acceptable level of accuracy, the precision becomes the primary concern and the latter criterion should then be used to guide selection of the number of intermediate stages.

3. A multistage FEP method for solids

The partition function is related to the free energy through the bridge equation of statistical mechanics. The multidimensional integral defining the classical partition function is written in terms of the phase-space coordinates describing the positions and momenta of all atoms in the system. We will consider a reformulation of this integral in terms of an alternative set of coordinates that are suitable for the description of a crystalline phase. We are interested here in only configurational degrees of freedom, so we will ignore any contributions due to the momenta.

Normal-mode coordinates can be defined by considering a particular lattice structure, and attaching harmonic potentials between neighboring lattice sites. Diagonalization of the expression for the energy in terms of the Cartesian coordinates yields a set of eigenvectors that define collective motions of the system of lattice points. It is possible to express the partition function of a crystalline system in terms of its normal-mode coordinates instead of the more typical Cartesian coordinate system. Our interest in doing this derives in part from the fact that the complete set of normal modes of an N -particle system can be expressed in terms of a subset of the normal modes of the $2N$ -particle system occupying twice the volume (the N even numbered modes in the one-dimensional case), making it possible to describe a free-energy path between the two systems. The idea is that the $2N$ -particle system constrained to move only according to the normal modes common the N -particle half-system has a free-energy identical to (or trivially related to) the N -particle system itself. In freeing the constrained normal modes, we traverse a path that effectively doubles the system size without actually requiring addition of new molecules into the system. In calculating the contribution due to the remaining modes unique to the $2N$ -particle system (the odd numbered modes in the one-dimensional case), we determine the free-energy difference between the two systems. Further, since the free energy is an extensive quantity, and presuming we can ignore finite-size effects (as is reasonable for sufficiently large systems), this free-energy difference is equal to the free energy of the N -particle system itself.

We examine this idea by performing a Monte Carlo simulation for an NVT ensemble of one-dimensional hard rods with periodic boundaries, though this method could certainly be extended to higher dimensions. The hard-rod system is convenient, because it has an analytical solution for the free energy, with complete characterization of finite-size effect, allowing for the direct comparison of this FEP method with the exact value. The harmonic system is one-dimensional with periodic boundaries containing $2N$ -particles. Because this approach requires that some of the normal-mode coordinates be frozen “off”, the method requires trial displacements in normal-mode space (phonon displacements) instead of the more common trial particle displacements in one-dimensional (Cartesian coordinate) space. Although the sampling occurs in normal-mode space, the test for accepting/rejecting a configuration is done conventionally, in the Cartesian coordinate space. The Jacobian for the normal-mode transformation is unity, so there is no inappropriate biasing to be removed.

3.1. Harmonic system

The internal energy of the nearest-neighbor one-dimensional harmonic system is given by Eq. (3), where γ_i is the deviation of the i th particle from its lattice site, W is the harmonic force constant. The diagonal elements of the matrix \mathbf{H} are $+2W$, and the secondary diagonal elements and the two elements in the upper right and lower left corners are $-W$.

$$U_{\text{harmonic}} = W \sum (\gamma_i - \gamma_j)^2 = \boldsymbol{\gamma}^T \mathbf{H} \boldsymbol{\gamma} \quad (3)$$

The matrix form of Eq. (3) can be diagonalized to express the harmonic internal energy in terms of its normal coordinates, which can be considered a collection of $2N$ independent oscillators

$$U_{\text{harmonic}} = \boldsymbol{\eta}^T \mathbf{A} \boldsymbol{\eta} \quad (4)$$

where $\boldsymbol{\eta}$ is the vector of normal-mode deviations (analogous to particle deviations in Cartesian coordinates), and \mathbf{A} is a diagonal matrix of the product of W with the eigenvalues of \mathbf{H} . The partition function

of each of the independent harmonic oscillator given is

$$q_m = \sqrt{\frac{2\pi}{W\lambda_m}} \tag{5}$$

where the eigenvalues [14], λ , are

$$\lambda_m = 2 \left[1 - \cos \left(\frac{2\pi m}{N} \right) \right], \quad m = 1, \dots, N \tag{6}$$

and the elements of the normalized eigenvectors, ϕ_{mk} , are

$$\phi_N(m, k) = N^{-(1/2)} \left[\sin \left(\frac{2\pi mk}{N} \right) + \cos \left(\frac{2\pi mk}{N} \right) \right], \quad m, k = 1, \dots, N \tag{7}$$

The feature of the normal modes that we exploit is that the complete set of eigenvectors for the N -particle system can be expressed in terms of the first N elements of the even numbered eigenvectors of the $2N$ -particle system, so that $\phi_N(m, k) = 2^{1/2}\phi_{2N}(2m, k)$. It should be noted that the second set of N elements of the even numbered eigenvector is identical to the first set of elements. Physically, this means that the system is symmetric with respect to a translation equal to half the system size. If all the odd normal-mode coordinates are zero, the system of $2N$ rods is practically an N -particle system.

3.2. Hard-rod system

The N and $2N$ hard-rod systems [15] have lengths of L and $2L$, respectively, with periodic boundaries, and the rods have a uniform width of size d . The partition function for the hard-rod system is

$$Q_N = \frac{L(L - Nd)^{N-1}}{N!} \tag{8}$$

The period boundaries cause the particles move into the system from one side as they move out of the system from the other side. The internal energy of the system is infinite if any two hard-rods overlap or move out of order, and is zero otherwise.

3.3. Method of perturbation

The free-energy difference between the reference and target systems can be expressed as

$$\Delta A = A_N + A_{\text{odd}} - A_{2N} = -\ln \left(\frac{Q_N Q_{\text{odd}}}{Q_{2N}} \right) \approx -A_N \tag{9}$$

where we presume A_{2N} to be the high-entropy system so that $A_{2N} = A_H$, and $A_N + A_{\text{odd}} = A_L$, and A_{odd} refers to the free energy due to the odd numbered modes of the $2N$ harmonic system and is given as $-\ln(Q_{\text{odd}}) = -\ln(\prod_{\text{mod } d} q_m)$. A_{2N} will be the high entropy system as long as the entropy of the harmonic system is not too large, i.e. as long as the harmonic force constant, W , is sufficiently large. But for small force constants, the combined entropy of the target N hard-rod and harmonic systems would become greater than that of the reference $2N$ hard-rod system, and we would have to reverse the indices L and H .

The ratio of partition functions in Eq. (9)

$$\frac{Q_N}{Q_{2N}} Q_{\text{odd}} = \frac{Q_{2N-1} q_1}{Q_{2N}} \frac{Q_{2N-2} q_2}{Q_{2N-1}}, \dots, \frac{Q_N q_N}{Q_{N+1}} \quad (10)$$

expresses a series of perturbations in terms of the normal-mode coordinates of the $2N$ system. For the first perturbation, the reference is the $2N$ hard-rod system with all the normal modes turned on, allowing for the greatest particle motion possible. The target is then the same $2N$ hard-rod system, but with one odd mode turned off (restricting the particle motion somewhat) plus the harmonic system with that same single mode turned on (increasing the motion allowed to the harmonic system). The target of this perturbation becomes the reference for the next following perturbation, or stage, and the target becomes the $2N$ hard-rod system with two odd modes turned off plus the same two modes turned on in the harmonic system. And so on. The last perturbation is between systems where the last odd mode is turned off. The partition function of the $2N$ system described only by its even modes is precisely equivalent to that of the N hard-rod system with a simple normalization. This series of perturbations describes a free-energy path from the $2N$ hard-rod system to the combination of the N hard-rod and odd-mode harmonic systems.

4. Results and discussion

In this section we apply the analysis of Section 2 to the simulation results recorded according to the methods of Section 3. We performed simulations of systems of hard rods of sizes ranging from $N = 4$ to 100. For each system size we performed normal-mode Monte Carlo calculations as described above, and evaluated the N -stage FEP for the process in which half the normal-mode coordinates are frozen at zero. The number of FEP trials per stage, M , was 10,000.

Fig. 1 shows a plot of the free energy per particle calculated by the simulation as a function of the harmonic system force constant, compared to the exact hard-rod free energy obtained from the analytic solution. Also shown on the plot is the free energy for the harmonic system of the same force constant, so the difference between that line and the exact free energy is the FEP calculated by the simulation.

Because we have an exact expression for the terms in Eq. (9), we can approximate the entropy change per stage Δs by dividing the total entropy difference by the number of stages.

$$\Delta s = \frac{\Delta S}{\text{stages}} = \frac{A_{2N} - A_N - A_{\text{odd}} + (N/2)}{\text{stages}} \quad (11)$$

The entropy change per stage is not constant for each of the FEP stages, but except near the end of the FEP path most of the single-stage entropy changes lie near this value. A plot of this “average” Δs is included in Fig. 1. The entropy difference is positive for the harmonic reference with the smallest force constant; therefore that simulation is not one of generalized insertion. In accord with the analysis of Section 2, the free energy calculated for this system is inaccurate. Even when Δs is near to zero it is difficult to characterize the simulation as strictly insertion. Although $\Delta s = 0$ indicates that the configuration spaces are comparable in size, their shapes may be different, so that a portion of the target may lie outside of that of the reference configuration space. This gives the calculation a deletion character, and the consequence is a continued systematic error in the results.

For an increasing force constant, Δs becomes increasingly negative. The configuration space of the target becomes correspondingly smaller, and is thus more likely to be completely contained within the

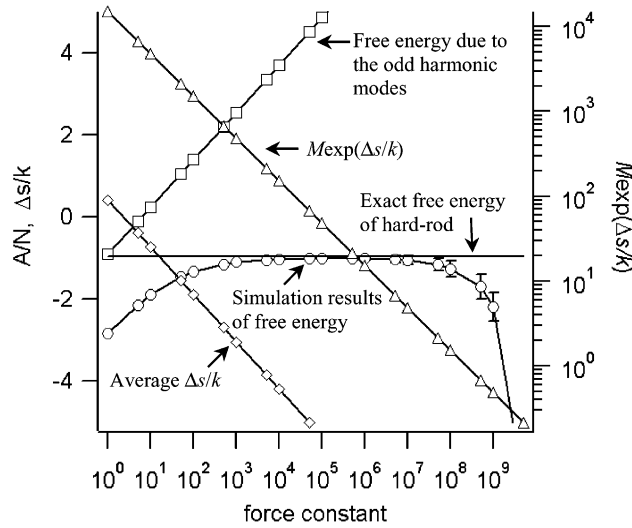


Fig. 1. Plot of free energy per particle (A/N) and entropy difference ($\Delta s/k$) as functions of the harmonic force constant from a simulation of 100 hard rods at a number density (Nd/L) of 0.5. The corresponding results of $M \exp(\Delta S/k)$ are also plotted using the right-hand y-axis. The number of trials per stage, M , is 1000.

reference, so that the simulation is of the generalized insertion type. In such a case the measures discussed in Section 2 can be invoked to gauge the accuracy of the calculation. Fig. 1 also plots $M \exp(\Delta S/k)$, which we have proposed as a gauge for estimating the accuracy of insertion FEP calculations. For force constants W of at least 10^3 , the simulation free-energy results come into agreement with the exact free-energy value, indicating that the target configuration space is completely contained within the reference. Also at $W = 10^3$, the value of $M \exp(\Delta S/k)$ is about 500, surpassing the threshold value of 100 needed to indicate an accurate result. We observe that the results remain good even up to a force constant of 10^7 , where $M \exp(\Delta S/k)$ is only about 5.0. However, beyond this point the simulation tends to significantly overestimate the correction to the harmonic system, as predicted by the error analysis for small values of $M \exp(\Delta S/k)$.

5. Conclusions

This study has demonstrated the basic concepts needed to perform FEP calculations in a manner that yields accurate results. We find in this test case that the deletion-type calculation arising for small harmonic spring constants yield systematically incorrect results, and that insertion-type calculations yield good results for sufficiently large values of $M \exp(\Delta S/k)$. Two new features of the analysis emerged in this study. First, we find that FEP calculations for which ΔS is approximately zero are also prone to inaccuracy, because the configurations important to both phases are unlikely to overlap well, even though their sizes are comparable. Second, we find that for the insertion calculation that the $M \exp(\Delta S/k) = 100$ threshold is overly conservative, and that good results can be obtained for values of this group of order 10 or smaller. This is not inconsistent with our original study. There we found that in some cases good insertion FEP results could be obtained for such small values of this group. However, we find that this is not always

the case, and safe practice—where one does not have the luxury of comparing to an pre-established exact result—should apply insertion FEP calculations in a way that has this group at the recommended threshold, or larger.

Acknowledgements

This work has been performed with the support from the Division of Chemical Sciences, Office of Basic Energy Sciences, Office of Energy Research of the US Department of Energy (Contract DE-FG02-96ER14677).

References

- [1] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [2] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, New York, 1996.
- [3] P. Kollman, *Chem. Rev.* 32 (1993) 2395–2417.
- [4] D.A. Kofke, P.T. Cummings, *Mol. Phys.* 92 (1997) 973–996.
- [5] D.A. Kofke, P.T. Cummings, *Fluid Phase Equilib.* 150 (1998) 41–49.
- [6] N. Lu, D.A. Kofke, *J. Chem. Phys.* 111 (1999) 4414–4423.
- [7] N. Lu, D.A. Kofke, *J. Chem. Phys.* 114 (2001) 7303–7311.
- [8] N. Lu, D.A. Kofke, *J. Chem. Phys.* 115 (2001) 6866–6875.
- [9] N. Lu, D.A. Kofke, in: P. Cummings, P. Westmoreland (Eds.), *Foundations of Molecular Modeling and Simulation*, AIChE Symp. Ser. 97 (2001) 146–149.
- [10] R.W. Zwanzig, *J. Chem. Phys.* 22 (1954) 1420.
- [11] J.L. Jackson, L.S. Klein, *Phys. Fluid* 7 (1964) 228–231.
- [12] B. Widom, *J. Chem. Phys.* 39 (1963) 2808–2812.
- [13] K.S. Shing, K.E. Gubbins, *Mol. Phys.* 46 (1982) 1109–1128.
- [14] N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, Saunders, London, 1976, p. 423 (Eq. 22.28).
- [15] L. Tonks, *Phys. Rev.* 50 (1936) 955.