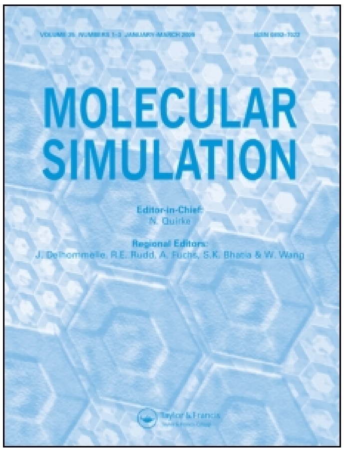

Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Determination of Free Energy from Chemical Potentials: Application of the Expanded Ensemble Method

A. P. Lyubartsev[a]; A. Laaksonen[a]; P. N. Vorontsov-velyaminov[b]
[a] Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden [b] Physics Faculty, St. Petersburg State University, St. Petersburg, Russia

To cite this Article Lyubartsev, A. P. , Laaksonen, A. and Vorontsov-velyaminov, P. N.(1996) 'Determination of Free Energy from Chemical Potentials: Application of the Expanded Ensemble Method', Molecular Simulation, 18: 1, 43 — 58

To link to this Article: DOI: 10.1080/08927029608022353

URL: http://dx.doi.org/10.1080/08927029608022353

Molecular Simulation, 1996, Vol. 18, pp. 43–58
Reprints available directly from the publisher
Photocopying permitted by license only

DETERMINATION OF FREE ENERGY FROM CHEMICAL POTENTIALS: APPLICATION OF THE EXPANDED ENSEMBLE METHOD

A. P. LYUBARTSEV[a], A. LAAKSONEN[a]
and P. N. VORONTSOV-VELYAMINOV[b]

[a]*Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University S-106 91, Stockholm, Sweden*
[b]*Physics Faculty, St. Petersburg State University, 198904, St. Petersburg, Russia*

(Received March 1996; accepted March 1996)

The expanded ensemble method, previously developed for free energy calculation [*J.Chem.Phys.*, **96**,1776 (1992)] is applied to calculate chemical potentials. The expanded ensemble is composed as a sum of canonical ensembles with gradually inserting the $(N+1)$:th particle. The probability distribution over the subsensembles is directly related to the ratio of the partition functions and, hence, to the free energy difference. The gradual insertion eliminates the difficulties arising in using the standard particle insertion method at high densities. The problem of an optimal choice of subsensembles is studied in detail. Since the chemical potential is defined as the Gibbs free energy per particle for macroscopic systems, the present method allows calculation of free energies in a convenient way. The method is applied to calculate chemical potentials and free energies for a Lennard-Jones system and for the flexible SPC water model. The results are compared with corresponding direct free energy calculations using the temperature expanded ensemble and the efficiency vs precision of the two approaches are evaluated.

Keywords: Free energy calculations; molecular dynamics; expanded ensemble; errors.

1. INTRODUCTION

The computation of free energy and other related quantities (chemical potential, entropy) by Monte Carlo (MC) or molecular dynamics (MD) computer simulations has become an area of considerable interest. Knowledge of free energy is a key quantity for understanding molecular phenomena, such as phase transitions, solvation, conformational transitions in macromolecules, chemical equilibria, to mention a few. The MC or MD simulation techniques are straightforward to use to obtain such characteristics of condensed molecular systems as internal energy, pressure, radial distribution functions, for which there exists an appropriate microscopic analogue – an estimator. However, free energy, chemical potential and other quantities involving contributions from entropy, cannot be obtained by a simple (direct) averaging of an appropriate estimator.

To calculate free energy within MC or MD simulations, a number of methods have been proposed. Examples of these are: thermodynamical integration [1–3], particle insertion [4,5], multistage sampling [6], various versions of umbrella sampling [7,8], perturbation methods [2,9,10], acceptance ratio method [11,12], and several others. The reader is referred to review articles [13,14] for an update. Recently, a new approach, the expanded ensemble method, was suggested [15] which allows us to calculate absolute free energies with a high precision for arbitrary systems using a relatively simple computational scheme within MC [15] or MD [16] simulations. The general idea of the method is to compose an expanded ensemble, which is a sum of conventional (e.g. canonical) ensembles with pre-weighting ("balancing") factors. A MC random walk, according to the Metropolis scheme [17], yields a probability distribution over subensembles, which is directly related to the ratio of corresponding partition functions and, hence, to the free energy differences. If the free energy of one of the subensembles is known (the reference system), we obtain the free energies for the other subensembles. It should be mentioned, that related approaches have been suggested in literature: semi-grand-canonical ensemble for chemical potential calculation [18], simulated tempering [19] and force balance method [20].

The expanded ensemble method has been applied for free energy calculations of different systems and found to be very accurate and effective. Examples of the applications so far are: primitive electrolyte model [15,21], Lennard-Jones fluids and liquid water [16], quantum Heisenberg model [21,22], ordered polyelectrolyte system [23], lattice polymers [24]. A Monte Carlo walk in the configurational space can be replaced by a constant-temperature MD trajectory [16,25].

The task of chemical potential calculations is closely related to the free energy calculations. For macroscopic systems the chemical potential μ is related to the Gibbs free energy G:

$$G = \mu N \tag{1a}$$

or to the Helmholtz free energy F:

$$F = \mu N - PV \tag{1b}$$

In order to calculate the free energies G or F, the corresponding chemical potential μ has to be determined for a sufficiently large periodic pattern. The standard way to proceed is to use the Widom formula [4]:

$$\exp(-\beta\mu) = \langle \exp(-\beta h_{N+1}) \rangle \tag{2}$$

where h_{N+1} is the energy of "random insertion" of the $(N+1)$:th particle. This method has been widely used for calculations of chemical potential [4,5]. It suffers, however, from a drawback of becoming inefficient at liquid densities. Various techniques, such as, *f-g* sampling [26], umbrella sampling [27] and biased simulations [28], have been used to improve the applicability of the Widom formula (2) at high densities.

An alternative technique is the application of the gradual particle insertion within the expanded ensemble method, in which the $(N+1)$:th particle is gradually mutated from a real particle into a "ghost" (non-interacting) particle. This method was suggested in Ref. [29], and it was, in fact, the first time when the expanded ensemble method was used, although expressed in slightly different terms. The method has proven to be a very efficient way of chemical potential calculations [18,29], including the case of dense polymer systems [30,31].

Clearly, the evaluation of chemical potential provides another route to obtain free energies. The attractive feature in doing so within the expanded ensemble method is the fact that only a single particle is needed to be inserted into the system, rather than converting the whole system into an ideal gas as it has been done using the direct method for obtaining free energy [15]. This is expected to offer some considerable computational savings. Previously, an analogous way was used in free energy calculations of a LJ system within the biased particle insertion method [32].

In this paper we have applied the expanded ensemble method to the chemical potential calculations for a Lennard-Jones system and liquid water. In addition, we have varied the number of intermediate subensembles in order to investigate characteristic features of the method and we have also performed an error analysis of the results. Finally, in the frame of the expanded ensemble method, we have compared the two ways to obtain free energies: the direct free energy calculations, as done in our previous work [16], and indirectly from the calculated chemical potentials according to Eq.(1).

2. EXPANDED ENSEMBLE METHOD FOR CHEMICAL POTENTIAL CALCULATIONS

Generally, the chemical potential μ can be expressed as a free energy difference:

$$\mu(V,T)=\left.\frac{\partial F}{\partial N}\right|_{V,T}\approx F_{N+1}(V,T)-F_N(V,T)=-\mathrm{kT}\ln\left(\frac{Z(N+1)}{Z(N)}\right) \tag{3}$$

where $Z(N)$ is the partition function and F_N is the Helmholtz free energy of the system of N indentical particles.

Another way is to calculate μ from the corresponding Gibbs free energy difference:

$$\mu(P,T)=\left.\frac{\partial G}{\partial N}\right|_{P,T}\approx G_{N+1}(P,T)-G_N(P,T)=-\mathrm{kT}\ln\left(\frac{Y(N+1)}{Y(N)}\right) \tag{4}$$

where $Y(N)$ is the partition function of the NPT-ensemble.

Consider an N-particle system into which an excess $(N+1)$:th particle is gradually inserted. At each insertion step we have a system with configurational partition function:

$$Z_m=\int_V\prod_{i=1}^{N+1}dq_i\exp(-\beta(H_N(q_i)+\alpha_m h_{N+1}(q_i)))$$

where H_N is the Hamiltonian of the system of N particles, h_{N+1} is the interaction energy of the $(N+1)$:th particle with the other N particles, α_m is a set of values, gradually changing from 0 to 1.

The partition function of the expanded ensemble can be given as:

$$Z=\sum_{m=0}^{M}Z_m\exp(\eta_m)=\sum_{m=0}^{M}\int_V\prod_{i=1}^{N+1}dq_i\exp(-\beta(H_N(q_i)+\alpha_m h_{N+1}(q_i))+\eta_m) \tag{5}$$

where η_m parameters are so called "balancing factors". The 0:th subensemble with $\alpha_0=0$ corresponds to the system with N interacting particles and a single "ghost" particle. Its partition function is related to the total partition function of the N particles system $Z(N)$ by:

$$Z_0\equiv\int_V\prod_{i=1}^{N+1}dqi\ \exp(-\beta(H_N(qi)))=N!Z(N)\,V\Lambda^{3N} \tag{6}$$

-where Λ is the de Broglie wavelength

Subensembles with the increasing m correspond to the gradual insertion of the $(N+1)$:th particle. The M:th subensemble with $\alpha_M = 1$ corresponds to the system with the true $(N+1)$ particle and the corresponding partition function:

$$Z_M = (N+1)!Z(N+1)\Lambda^{3(N+1)} \tag{7}$$

An MC walk in the expanded space of $(\{q_{i=1},\ldots,N+1\},m)$, combining two types of steps, particle displacements and changes of "m", yields the probability distribution over the subensembles p_m, which satisfies:

$$\begin{aligned}\frac{p_M}{p_0} &= \frac{Z_M}{Z_0}\exp(\eta_M - \eta_0) = \frac{(N+1)Z(N+1)\Lambda^3}{Z(N)V}\exp(\eta_M - \eta_0)\\ &= \frac{(N+1)\Lambda^3}{V}\exp(-\beta(F(N+1) - F(N)) + \eta_M - \eta_0)\end{aligned} \tag{8}$$

and from Eq. (3) above, (6) and (7) we have for the chemical potential:

$$\beta\mu = -\ln\left(\frac{p_M}{p_0}\right) + \eta_M - \eta_0 - \ln\left(\frac{V}{\Lambda^3(N+1)}\right) \tag{9}$$

Usually, only the "excess" part of the chemical potential is of interest. For μ_{ex} we have:

$$\beta\mu_{\text{ex}} = -\ln\left(\frac{p_M}{p_0}\right) + \eta_M - \eta_0 \tag{9a}$$

For macroscopic systems the chemical potential is related to the Gibbs or Helmholtz free energy by (1a) or (1b), respectively, thus for the corresponding Helmholtz excess free energy we have:

$$\begin{aligned}\beta F_{\text{ex}} &= \beta F + \ln\frac{(V/\Lambda^3)^N}{N!}\\ &= N\left(-\ln\left(\frac{p_M}{p_0}\right) + \eta_M - \eta_0 - \ln\left(\frac{V}{\Lambda^3(N+1)}\right)\right) + \ln\frac{(V/\Lambda^3)^N}{N!} - \beta PV\\ &= N\left(-\ln\left(\frac{p_M}{p_0}\right) + \eta_M - \eta_0 - 1\right) - \beta PV = N(\mu_{\text{ex}} + 1) - \beta PV\end{aligned} \tag{10}$$

The pressure can be calculated in a conventional way using the virial expression:

$$PV = NkT - \sum_{i,j} \left\langle \frac{\partial H_N}{\partial r_{ij}} r_{ij} \right\rangle \tag{11}$$

In the expanded ensemble we can calculate the pressure at the two "end points", corresponding to N and $N+1$ particles, and take the mean value of the two.

An alternative way to obtain chemical potentials is calculating of the Gibbs free energy difference at constant pressure according to:

$$\mu = [G(N+1) - G(N)]_P$$

The Gibbs free energy difference can be obtained in an analogous way from the expanded NPT ensemble with gradually inserting the $(N+1)$:th particle. The partition function in this case is given as

$$Y = \sum_{m=0}^{M} Y_m \exp(\eta_m) = \sum_{m=0}^{M} \int_0^{\infty} dV \int_V \prod_{i=1}^{N+1} dq_i \exp(-\beta(H_N(q_i) \\ + \alpha_m h_{N+1}(q_i) + PV)) + \eta_m) \tag{12}$$

where Y_m are the configurational NPT partition functions of the systems with partially inserted $(N+1)$:particle. The MC procedure now consists of three types of steps: *i*) particle displacements, *ii*) change of the volume according to an appropriate scheme to create an NPT ensemble [17] and *iii*) steps with gradual particle insertion or deletion (change of "m"). It is straightforward to show that the chemical potential and Helmholtz free energy can be obtained from the probability distribution over subsensembles p_m using the same formulas (8) and (10) as in the case of NVT ensemble. One can apply constant-temperature/constant-pressure molecular dynamics [33] for moves in the coordinate space, combined with Monte Carlo steps for transitions among the subensembles.

It is also possible to implement the expanded ensemble method in terms of purely dynamical simulations, by changing a discrete set of α_m-points by a continuous α-interval and declaring α as a dynamical variable. The algorithm obtained in this way would be a complete analogue to grand canonical molecular dynamics simulations [34,35] with the only exception that the number of particles fluctuates between N and N + 1.

3. COMPUTATIONAL DETAILS AND THE NUMBER OF SUBENSEMBLES

The use of the expanded ensemble method for each particular system or a specific application requires a number of parameters to be chosen in advance. First, the number of subensembles, M, must be specified together with the points α_m, distributed within the interval [0,1]. In addition, the balancing factors, η_m, have to be set. The problem of the optimum choice of the balancing factors has been discussed in details in our previous works [15,16]. Here we will analyze the proper choice of subensembles.

The intermediate subensembles are introduced to provide an effective sampling of the configurational space between the two end states (creating "an umbrella"). The expanded ensemble method resembles the "umbrella sampling" [7] in the sense that in both methods the probability distribution covers all the relevant regions of the configurational space in a single run. The difference is that in the expanded ensemble no special non-Boltzmann distribution is needed, an "umbrella" is automatically generated by MC transitions between subensembles. For an effective exploring of the configurational space the intermediate α_m points should be chosen in such a way that a reasobably high probability is provided for transitions between subensembles. A large number of subensembles and small differences between the α_m values may yield a good transition probability. However, using a small step size requires consequently more time for the system to walk all the way between the two extreme subensembles at the both ends. Larger intervals between the α_m points, on the contrary, result in a poor acceptance probability. Apparently there exists a certain optimum number of subensembles with an optimum distribution of points, which yield a minimum statistical error in the calculated free energy. There is also an analogue to be found with the problem of choosing a maximum particle displacement in the conventional MC method: a general rule of thumb being that the acceptance ratio should be about 50%. However in the case of free energy calculations based on the expanded ensembles method, one is only interested in probabilities of the two extreme points, P_0 and P_M, whereas the intermediate points and the corresponding canonical averages over them are of marginal importance. Therefore, the optimum number of subensembles can be less than which follows from the 50% acceptance ratio criteria.

We have performed a series of calculations of the chemical potential of a Lennard-Jones system using the expanded ensemble method with different number of intermediate subensembles in order to analyze the influence of this parameter. The LJ system consisted of 256 particles with the following reduced parameters to specify the physical state: $T^* = kT/\varepsilon = 0.72$, density $\rho^* = N\sigma^3/V = 0.82$ in the NVT ensemble or pressure $P^* = P\sigma^3/\varepsilon = 0.0024$ in the NPT ensemble that corresponds to liquid argon system ($\sigma = 3.41$ Å and $\varepsilon = 0.99$ kJ/M) at $T = 86$ K, $\rho = 1.364\,\text{g/cm}^3$ and $p =$ 1atm. The molecular dynamics version of the expanded ensemble method was used [16]. The time step was 1 fs and attempts to change the subensemble were made after each 5 fs. The balancing factors were optimized in a number of initial trial runs with total length of 2.10^5 (200 ps) MD steps. The length of the final runs was extended to 2 ns. The statistical error was calculated from the variance of averages over intervals of 200 ps each over the entire MD trajectory.

First, we carried out simulations with 20 subensembles at constant volume and at constant pressure, respectively. The intermediate α_m points were chosen condensing close to 0, because the effective size of LJ particles decreases very slowly upon decreasing the α parameter. The results of these simulations are presented in Table I for the NVT and in Table II for the NPT ensemble. These both results nearly coincide and there is no difference observed whether one of the two ensembles is more effective to use

TABLE I Results of chemical potential calculations in the expanded NVT ensemble for a Lennard-Jones system, $T^* = 0.72$, $\rho^* = 0.82$. h is the mean interaction energy of the inserted particle with the other particles, $P^* = P\sigma^3/\varepsilon$ is the reduced pressure, $p-$ and $p+$ are the acceptance ratios of transitions to a subensemble with a smaller or a larger m, respectively. For more details, see the text

λ_m	η_m	p_m	βF	$\beta\alpha_m h$	P^*	p^-	p^+
1.	−5.2	0.064	−5.76	−11.35	0.004	–	0.92
0.9	−3.6	0.066	−4.21	−10.04	0.006	0.91	0.88
0.8	−2.1	0.063	−2.65	−8.75	0.011	0.92	0.86
0.7	−0.7	0.058	−1.17	−7.47	0.013	0.92	0.91
0.6	0.8	0.060	0.28	−6.22	0.014	0.88	0.89
0.5	2.2	0.060	1.69	−4.95	0.016	0.89	0.85
0.4	3.5	0.057	3.04	−3.72	0.019	0.90	0.79
0.3	4.6	0.051	4.26	−2.51	0.021	0.87	0.68
0.2	5.6	0.047	5.32	−1.29	0.025	0.81	0.69
0.1	6.2	0.047	5.93	−0.09	0.024	0.69	0.59
0.04	5.8	0.044	5.58	0.55	0.020	0.61	0.68
0.02	5.1	0.041	4.95	0.75	0.011	0.71	0.75
0.01	4.4	0.043	4.21	0.79	0.004	0.72	0.62
0.003	3.2	0.046	2.96	0.67	−0.009	0.58	0.69
0.001	2.3	0.046	2.05	0.49	−0.015	0.70	0.69
0.0003	1.5	0.042	1.36	0.34	−0.021	0.75	0.83
0.0001	1.1	0.042	0.92	0.23	−0.023	0.81	0.84
0.00003	0.8	0.044	0.59	0.15	−0.026	0.84	0.86
0.00001	0.6	0.043	0.41	0.09	−0.027	0.88	0.62
0.	0.	0.037	0.	0.	−0.028	0.82	–

TABLE II Results of chemical potential calculations in the expanded NPT ensemble for a Lennard-Jones system, $T^* = 0.72$, $\rho^* = 0.0024$. Parameters and notations are the same as in Table I

λ_m	η_m	p_m	βG	$\beta\alpha_m h$	ρ^*	p^-	p^+
1.	−5.2	0.064	−5.80	−11.40	0.8187	–	0.90
0.9	−3.6	0.066	−4.20	−10.12	0.8183	0.90	0.88
0.8	−2.1	0.063	−2.65	−8.80	0.8182	0.91	0.83
0.7	−0.7	0.057	−1.14	−7.51	0.8178	0.92	0.88
0.6	0.8	0.056	0.37	−6.26	0.8177	0.90	0.88
0.5	2.2	0.055	1.79	−5.00	0.8178	0.90	0.84
0.4	3.5	0.053	3.11	−3.75	0.8178	0.86	0.79
0.3	4.6	0.046	4.35	−2.51	0.8178	0.86	0.77
0.2	5.6	0.045	5.37	−1.30	0.8181	0.80	0.70
0.1	6.2	0.046	5.95	−0.09	0.8188	0.68	0.60
0.04	5.8	0.046	5.56	0.58	0.8199	0.60	0.69
0.02	5.1	0.044	4.90	0.72	0.8204	0.71	0.60
0.01	4.4	0.045	4.18	0.77	0.8209	0.72	0.60
0.003	3.2	0.046	2.96	0.64	0.8210	0.60	0.69
0.001	2.3	0.046	2.07	0.49	0.8213	0.70	0.73
0.0003	1.5	0.044	1.3	0.35	0.8218	0.74	0.82
0.0001	1.1	0.046	0.86	0.24	0.822	0.79	0.86
0.00003	0.8	0.046	0.56	0.15	0.8221	0.86	0.87
0.00001	0.6	0.045	0.38	0.09	0.8224	0.88	0.62
0.	0.	0.036	0.	0.	0.8227	0.74	–

than the other. The acceptance ratio for transitions between the subensembles are also the same and relatively high (about 80%) throughout the whole interval. In both simulations, the system appears to have walked about 80 times between the two end subensembles, corresponding to N and N+1 particles, respectively. The statistical error of the probabilities P_m is about 0.004, corresponding to a statistical error in chemical potential and free energy of about 0.12 kT.

Thereafter we repeated the calculations with a smaller number of α_m points. These results are shown in Table III. The distributions of the α_m points within the interval [0,1] were chosen in each case to provide a nearly uniform acceptance ratio for transitions between the subensembles. The specific sets of the used $\{\alpha_m\}$ parameters are also given in Table III. The optimum number of subensembles, which yields minimum statistical error, is 5 for the studied system, and it corresponds to an average acceptance ratio of about 30%. A smaller number of subensembles leads to a fast decay in the acceptance ratio values and also to a larger statistical error. The case with only two subensembles corresponds to the standard particle insertion method and we can see from Table III that it has a very low acceptance ratio for the transitions, and consequently, it gives a large statistical uncertainty it the resulting value. A very large number of subensembles means a

TABLE III Comparison of some computational parameters and results from a series of chemical potential calculations for a LJ system. Parameters are the same as in Table I

number of subensembles	*set of* $\{\alpha_m\}$	μ_{ex}/kT	*statistical error*	*acceptance ratio of* α*-transitions*	*average number of steps to walk from one end point to another and back*
2	{0,1}	−5.2	1	0.0003	40000
3	{0,0.02,1}	−5.8	0.3	0.05	19000
4	{0,0.003,0.2,1}	−5.61	0.12	0.15	11000
5	{0,0.001,0.02,0.4,1}	−5.72	0.08	0.32	8000
7	{0,0.0003,0.01,0.04,0.1,0.4,1}	−5.75	0.09	0.5	12000
10	{0,0.0001,0.001,0.01, 0.04,0.02,0.4,0.6,1}	−5.68	0.10	0.67	17000
20	see Table 1,2	−5.78	0.12	0.82	23000

longer path between the ends, that results in increasing of the statistical error despite of more frequent transitions between the intermediate subensembles. However, a specific choice of the number of subensembles in reasonable limits is not very critical for the calculations. For example, the variation of M between 4 and 20 leads to no more than 50% increase of the statistical error compared to the optimum value. One further observation is that at the optimum number of subensembles, the average time for the system to walk from the 0:th to the M:th subensemble and back is minimized. This may be considered as another important criteria in choosing the optimum number of subensembles.

4. THE PRESENT METHOD AND THE DIRECT FREE ENERGY CALCULATIONS

4.1. Lennard-Jones System

The excess free energy can be calculated from the excess chemical potential using Eq. (10), yielding for the LJ system with the specifications given in the previous section: $F_{ex}/\mathrm{NkT} = -4.72 \pm 0.08$. This result can be compared with the corresponding free energy, obtained from the direct calculations (converting the whole system into an ideal gas using the method described in our previous work [16]). The corresponding result of the direct method was $F_{ex}/\mathrm{NkT} = -4.697 \pm 0.005$. The computational parameters and the characteristic results from the calculation are given in Table IV. The results from using the both methods are within the statistical error. Comparing the computational efficiency, it can be concluded, on the one hand, that the direct free energy calculations are more accurate with a much lower statistical error than arises from the corresponding chemical potential calculations. On the other hand, the direct free energy calculations require much more computational efforts for determination of suitable balancing factors.

4.2. Liquid Water

A comparison of the two ways of free energy calculations, from chemical potential using Eq. (10) and direct calculation, was repeated for the flexible SPC water model [36]. The number of molecules was 256, periodic boundary conditions and the Ewald method [17] for the treatment of the electrostatic interactions were applied. We used the multiple time step algorithm by Tuckerman *et al.*[37] with the small time step of 0.2 fs for the fast fluctuating

TABLE IV Comparison of computational parameters and the results from two alternative ways to calculate free energy for a LJ system. Parameters are the same as in Table I

method	*from chem.pot. and formula (10)*	*direct free energy calculations*
-F/NkT	-4.72	-4.697
statistical error	0.08	0.005
number of subensembles		
number of trial runs to fit the balancing factors	3	6
total length of the trial runs (MD steps)	2.10^5	3.10^6
total length of the final run with averaging	2.10^6	2.10^6
average acceptance ratio of transitions between subensembles	0.32	0.11
average number of steps to walk from one end point to another and back	8000	2.10^6

intramolecular degrees of freedom and for intermolecular interactions between the nearest neighbours and the large time step 1.0 fs for the slower intermolecular movements. Attempts to make transitions among the subensembles were performed after each 5 fs. In the case of direct free energy calculations, the initial choice of balancing factors was made on the basis of our previous free energy calculations for the system of 108 water molecules. It appeared, however, that a few additional simulations for optimization of balancing factors for the larger system of 256 molecules were necessary.

The final results for liquid water are shown in Table V. One can observe, that as in the case for the Lennard-Jones system, the direct free energy calculation is more accurate with much lower statistical error. However, it requires much more computing time to adjust the balancing factors so that the whole range of subensembles can be convered in one single run. As in the case of Lennard-Jones system, no differences were found between calculations using the NVT and NPT ensembles, respectively.

5. DISCUSSION

The results of the comparative study of the two ways of free energy calculations are understandable. In the expanded ensemble, the statistical error in the probabilities P_m is related to the statistical error of the free energy

TABLE V Comparison of computational parameters and results from two alternative ways of free energy calculations for the flexible SPC water model. For more details, see the text

method	*from chem.pot. and formula (10)*		*direct free energy calculations*
	NVT	*NPT*	
F/N (kJ/M)	−27.3	−26.5	−26.75
statistical error	0.7	0.7	0.02
number of subensembles	20	20	50
number of trial runs to fit the balancing factors	3	*	8
total length of the trial runs (MD steps)	2.10^{5*}		4.10^{6}
total length of the final run with averaging	5.10^{5}	5.10^{5}	3.10^{6}
average acceptance ratio of transistions between subensembles	0.38	0.38	0.22
average number of steps to walk from one end point to another and back	3.10^{4}	3.10^{4}	2.10^{6}

change for the whole system [15]:

$$\Delta\beta F = \ln\left(\frac{p_M}{p_0}\right) + \eta_M - \eta_0 \tag{13}$$

The calculated quantity in the temperature expanded ensemble [15] is the free energy difference between the system and the ideal gas. Consequently, for a specific free energy (per molecule), the statistical error is divided by the number of particles, N. Even in the case where the uncertainty in the ratio of the end point probabilities, P_0 and P_M, is of an order of a magnitude, the statistical error in specific free energy F/N is still very small. For example, for N = 256 it will be about 0.01 kT, and it will further decrease with increasing the number of particles. The cost of the high precision is the necessity to carry out a very delicate tuning of the balancing factors. If it is not done and deviations of the balancing factors from their optimum values exceed (per particle) several kT/N, some of the probabilities p_m would be vanishingly small and could not be determined in a finite simulation run. Another reason to a rapid increase of the computing time within the temperature expanded ensemble with increasing N is the fact that the energy distributions become narrower. A good acceptance ratio of transitions between subensembles can be achieved only if their shifted on η_m energy distributions overlap [15,24]. That is why, with a rise of N, more intermediate subensembles are needed to provide overlapping energy distributions and, consequently, a reasonable acceptance ratio for the transitions between subensembles. The growing number of subensembles is responsible for the increase of the time both to optimize the balancing factors and to make the complete walk between the two extreme subensembles. Naturally, for any evaluation of probabilities P_0 and P_M, the system must accomplish at least one returned walk between the two extreme subensembles. This can be considered as a minimum requirement for free energy calculations using the direct expanded ensemble method [15]. For systems with very large number of particles, this may be a limitation and using the present method may still be beyond the available computer resources.

In contrary to the latter, the calculations of chemical potentials are stable with respect to the number of particles. An increase of the number of particles N leads in this case only to an increase of the CPU time to move the particles in the coordinate space, while the requirements for the number of subensembles and the accuracy in fitting the balancing factors are kept unchanged. Our simulations show that the chemical potential and the corresponding specific free energy can be determined within a precision of about 0.5 kT for systems like water, after about 200–300 ps of simulations, including also the time needed for the optimization of the balancing factors. Longer simulations will produce more accurate results with the statistical error scaling in usual way, as $t^{-1/2}$, t being the duration of the simulation.

As a conclusion of the present study we find that if we need to obtain very accurate values of free energy, the temperature expanded ensemble with the accompanied calculation of the absolute free energy is the best choice. But if such a high precision in the free energy is not important or if the task is beyond the available computing resources, the now presented scheme for the chemical potential calculations, by gradual insertion of a particle and subsequent determination of the free energy should be a good alternative method in hands.

Finally, the relation between the free energy and the chemical potential can be used in opposite direction. If we are interested in a very accurate value of the chemical potential, the best way to obtain one, is to carry out a direct free energy calculation and to use Eq. (10) to calculate an accurate value of the chemical potential.

Acknowledgements

This work has been supported by the Swedish Natural Science Research Council, NFR (A.L.) and by the Wenner-Gren Foundation (A.P.L.)

References

[1] Straatsma, T. P., Berendsen, H. J. C. and Postma, J. P. M. (1986) "Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water", *J. Chem. Phys.*, **85**, 6720.

[2] Pearlman, D. A. (1994) "A comparison of alternative approaches to free energy calculations" *J. Phys. Chem.*, **98**, 1487; Pearlman, D. A. (1994) "Free energy derivatives: A new method for probing the convergence problem in free energy calculations", *J. Comp. Chem.*, **15**, 105.

[3] Mezei, M. (1993) "Calculation of solvation free energy differences for large solute change from computer simulations with quadrature-based nearly linear thermodynamic integration", *Mol. Simulations*, **10**, 225.

[4] Widom, B. (1963) "Some topics in the theory of fluids", *J. Chem. Phys.*, **39**, 2808.

[5] Adams, D. J. (1974) "Chemical Potential of hard-sphere fluids by Monte Carlo methods", *Mol. Phys.*, **28**, 1241.

[6] Valleau, J. P. and Card, D. N. (1972) "Monte Carlo estimation of the free energy by multistage sampling", *J. Chem. Phy.*, **57**, 5457.

[7] Torrie, G. M. and Vallueau, J. P. (1977) "Nonphysical sampling distributions in Monte Carlo free energy estimation: umbrella sampling", *J. Comput. Phys.*, **23**, 187; Valleau, J. P. (1991) "Density-scaling: A new Monte Carlo technique in statistical mechanics", *J. Comput. Phys.*, **96**, 193.

[8] Mezei, M. (1987) "Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias", *J. Comput. Phys.*, **68**, 237.

[9] Pearlman, D. E. and Kollman, P. A. (1989) "The lag between Hamiltonians and the system configuration in free energy perturbation calculations", *J. Chem. Phy.*, **91**, 7831; Kollman, P. A. (1993) "Free energy calculations: Application to chemical and biochemical phenomena", *Chem. Rev.*, **93**, 2395.

[10] Reynolds, C. A., King, P. M. and Richard, W. G. (1992) "Free energy calculations in molecular biophysics", *Mol. Phys.*, **76**, 251.

[11] Bennet, C. H. (1976) "Efficient estimation of free energy differences from Monte Carlo data", *J. Comput. Phys.*, **22**, 245.

[12] Rittger, E. (1993) "Free energy of systems with several local minima by the acceptance ratio method", *Mol. Phys.*, **79**, 1073.

[13] Frenkel, D. "Free energy computations and first-order phase transitions", in "*Molecular dynamics simulation of statistical-mechanical systems*", edited by Giccotti G. and Hoover W. G. (Amsterdam: North-Holland, 1986), p. 151

[14] Beveridge, D. L. and DiCapua, F. M. (1989) "Free energy via molecular simulations: Application to chemical and biochemical systems", *Ann. Rev. Biophys. Chem.*, **18**, 431.

[15] Lyubartsev, A. P., Martsinovskii, A. A., Shevkunov, S. V. and Vorontsov-Velyaminov, P. N. (1992) "New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles", *J. Chem. Phys.*, **96**, 1776.

[16] Lyubartsev, A. P., Laaksonen, A. and Vorontsov-Velyaminov, P. N. (1994) "Free energy calculations for Lennard-Jones systems and water using the expanded ensemble method. A Monte Carlo and molecular dynamics simulation study", *Mol. Phys.*, **82**, 455.

[17] Alllen, M. P. and Tildesley,, D. J. *The Computer simulation of liquids* (Oxford, Clarendon, 1987)

[18] Nezbeda, I. and Kolafa, J. A. (1991) "A new version of the insertion particle method for determining the chemical potential by Monte Carlo simulations", *Mol. Simulations*, **5**, 391.

[19] Marinari, E. and Parisi, G. (1992) "Simulated tempering: A new Monte Carlo scheme", *Europhys. Lett.*, **19**, 451.

[20] Attard, P. (1993) "Simulation of the chemical potential and the cavity free energy of dense hard-sphere fluids", *J. Chem. Phys.*, **98**, 2225.

[21] Lyubartsev, A. P., Martsinovskii, A. A., Vorontsov-Velyaminov, P. N. and Kuznetsova, T. V. (1993) "A new approach to the Monte Carlo calculation of free energy with the use of expanded ensembles", *Rus. J. Phys. Chem.*, **66**, 230.

[22] Kuznetsova, T. V. and Voronsov-Velyaminov, P. N. (1993) "Monte Carlo computation of the free energy in quantum two-dimensional Heinsenberg ferromagnets using the expanded ensemble method", *J. Phys.: Condens. Matter*, **5**, 717.

[23] Lyubartsev, A. P. and Nordenskiöld, L. (1995) "Monte Carlo simulation study of ion distribution and osmotic pressure in hexagonally oriented DNA", *J. Phys. Chem.*, **99**, 10373.

[24] Vorontsov-Velyaminov, P. N., Broukhno, A. V., Kuznetsova, T. V. and Lyubartsev, A. P. (1996) "Free energy calculations by expanded ensemble method for lattice and continuous polymers", *J. Phys. Chem.*, **100**, 1153.

[25] Swope, W. S. and Andersen, H. C. (1995) "A computer simulation method for the calculation of chemical potentials of liquids and solids using the bicanonical ensemble", *J. Chem. Phys.*, **102**, 2851.

[26] Shing, K. E. and Gubbins, K. E. (1981) "The chemical potential from computer simulations: Test particle method with umbrella sampling", *Mol. Phys.*, **43**, 717.

[27] Ding, K. and Valleau, J. P. (1993) "Umbrella-sampling realization of "Widom" chemical potential estimation", *J. Chem. Phy.*, **98**, 3306.

[28] Frenkel, D. and Smit, B. (1992) "Unexpected length dependence of the solubility of chain molecules", *Mol. Phys.*, **75**, 983.

[29] Shevkunov, S. V., Martsinovskii, A. A. and Vorontsov-Velyaminov, P. N. (1988) "Calculation of the critical size and properties of microdrops by Monte Carlo method in the generalized ensemble", *High Temp. (USSR)*, **26**, 246; Shevkunov, S. V., Vorontsov-Velyaminov, P. N. and Martsinovskii, A. A. (1991) "A new method for direct calculations of the critical size and the formation work of a microdrop", *Mol. Simulations*, **5**, 119.

[30] Wilding, N. B. and Muller, M. (1994) " Accurate measurements of the chemical Potential of polymeric systems by Monte Carlo simulations", *J. Chem. Phys.*, **101**, 4324.

[31] Escobedo, F. A. and dePablo, J. J. (1995) "Monte Carlo simulation of the chemical potential of polymers in an expanded ensemble", *J. Chem. Phys.*, **103**, 2703.

[32] Mezei, M. (1987) "Grand canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water", *Mol. Phys.*, **61**, 565.

[33] Melchionna, S., Ciccotti, G. and Holian, B. L. (1993) "Hoover NPT dynamics for systems varying in shape and size", *Mol. Phys.*, **78**, 533.

[34] Cagin, T. and Pettitt, B. M. (1991) "Molecular dynamics with a variable number of molecules", *Mol. Phys.*, **72**, 169.

[35] Ji, J., Cagin, T. and Pettitt, B. M. (1992) "Dynamic simulations of water at constant chemical potential", *J. Chem. Phys.*, **96**, 1333.

[36] Toukan, K. and Rahman, A. (1985) "Molecular dynamics study of atomic motions in water", *Phys. Rev. B*, **31**, 2643.

[37] Tuckerman, M., Berne, B. J. and Martyna, G. J. (1992) "Reversible multiple time scale molecular dynamics", *J. Chem. Phys.*, **97**, 1990.