

Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria

William L. Jorgensen* and Laura L. Thomas

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received January 8, 2008

Background

The calculation of free energy changes is fundamental for the thermodynamic characterization of reaction pathways and chemical equilibria. The conventional approaches, free energy perturbation (FEP) and thermodynamic integration (TI), were predicated on Kirkwood's introduction of the continuous coupling parameter in his integral equation studies of fluid systems.^{1,2} Later work by Zwanzig further advanced the mathematical formalism of free energy perturbation theory,³ and this, in conjunction with molecular dynamics and Monte Carlo sampling techniques, has become a principal approach for calculating free-energy differences.^{2,4,5} Some applications have included computation of relative free energies of solvation, relative pK_a values, medium-effects on conformational equilibria, host–guest binding affinities, and free-energy surfaces for organic and biochemical reactions.

Free energy perturbation theory, as presented by Zwanzig in his 1954 paper,³ relates the free energy difference between an initial (reference) and a final (target) state of a system to an average of a function of their energy difference evaluated by sampling for the initial state (eq 1).

$$\Delta F = F_1 - F_0 = -kT \ln \left\langle \exp \left[\frac{-(E_1 - E_0)}{kT} \right] \right\rangle_0 \quad (1)$$

Zwanzig further went on to derive an expression for the free energy change as a power series. Defining the energy of the target state, E_1 , as the sum of the energy of a reference state, E_0 , and a small perturbing potential, V (eq 2), eq 3 was obtained by expanding eq 1 to second order.

$$E_1 = E_0 + V \quad (2)$$

$$\Delta F = F_1 - F_0 = \langle V \rangle - \frac{1}{2kT} (\langle V^2 \rangle - \langle V \rangle^2) \quad (3)$$

Here, the free energy difference between the reference and

target states is estimated to first order by the reference ensemble average of the perturbing potential and to second order by including the average squared deviation of the perturbing potential from its mean, or the variance.^{2,3} This approximation for calculating free energy differences has been recognized as especially relevant for systems with fluctuations governed by a Gaussian probability distribution,² and, for example, related protocols have been applied to the charging of ions in solution,⁶ the estimation of intrinsic pK_a s and pK_a shifts in proteins,⁷ and the calculation of solvent contributions to the excited states of tryptophan.⁸

Landau, Lifshitz, and Peierls.^{9–12} Perturbation theory in its various forms has long been of interest to the theoretical physics community.² Specifically, a section on “Thermodynamic Perturbation Theory” appeared in the 1951 Russian edition of Statistical Physics by Landau and Lifshitz,¹³ the fifth volume in their encyclopedic Course of Theoretical Physics. This section begins with a definition of the partition function and its expansion in powers of V to second-order as given by eq 4 (their eq 32.2). Here F is the target free energy (F_1 above) and analogously $V = E - E_0$.

$$\exp[-F/kT] = \int \exp[-(E_0(p, q) + V(p, q))/kT] d\Gamma \cong \int \exp[-E_0/kT] \left(1 - \frac{V}{kT} + \frac{V^2}{2(kT)^2} \right) d\Gamma \quad (4)$$

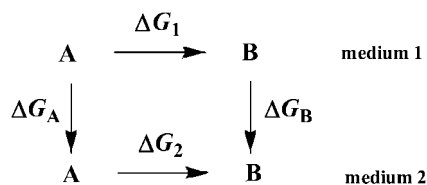
Then “taking the logarithm and again expanding in series to the same approximation” they obtain eq 5 (their eq 32.3). Eq 5 is identical to eq 3 using the equality, $\overline{V^2} - \overline{V}^2 = \langle V^2 \rangle - \langle V \rangle^2$, which they note.

$$F = F_0 + \overline{V} - \frac{1}{2kT} \overline{(V - \overline{V})^2} \quad (5)$$

While few citations appear in the 1951 edition of the book, due in part to Landau's preference for his own derivations, a rare attribution to Landau's colleague, R. E. Peierls, appears on p 93 at the beginning of the section on “Thermodynamic Perturbation Theory” in the 1958 English edition: “it is possible to develop a kind of ‘perturbation theory’ to

* Corresponding author e-mail: william.jorgensen@yale.edu.

Scheme 1



calculate thermodynamic quantities (R. E. Peierls, 1933).¹⁴ Except for the “R. E. Peierls, 1933”, the 1951 Russian and 1958 English versions of this section are essentially identical. The year was changed from 1933 to 1932 in later editions.¹⁵

This citation likely refers to Sir Rudolf E. Peierls’ paper “On the Theory of the Diamagnetism of Conduction Electrons”,¹⁶ submitted in 1932 and published in the original German in 1933. In this work, Peierls investigates the diamagnetic susceptibility of free electrons, and at the outset he provides an outline of thermodynamic perturbation theory. The paper includes a perturbative expansion of the partition function that encompasses eq 4 (eq 7 in ref 16). The connection between the material in Peierls’s 1933 paper and the section on “Thermodynamic Perturbation Theory” in the book by Landau and Lifshitz may have been strengthened by interactions between Landau and Peierls, as both were affiliated with Wolfgang Pauli’s laboratory around 1930; Landau was a visiting student, and Peierls was Pauli’s assistant at the ETH during 1929–1932. It is also possible that the “R. E. Peierls, 1933” reference was added by the translators for the 1958 English edition, E. Peierls and R. F. Peierls.¹⁴ Although their full appellations do not appear in the text, presumably the former is R. E. Peierls’s wife, Eugenia, who was also a classmate and friend of Landau’s at the Leningrad Physicotechnical Institute, and the latter is the Peierls’ son, Ronald, who received a Ph.D. in Physics from Cornell University in 1959. R. E. Peierls was a visiting lecturer in Leningrad in 1931, worked on the Manhattan Project in the early 1940s, and eventually became the Wykeman Professor of Theoretical Physics at Oxford during 1963–1974.¹² His contributions to the development of thermodynamic perturbation theory appear to be underappreciated; it would be reasonable to refer to eq 5 and equivalents such as eq 3 as the Peierls equation.

Sampling Methods. For applications of eq 1 to chemical equilibria, normally, states 0 and 1 are different molecules, A and B, and a change in medium is investigated by comparing the free energy change for the conversion of A to B in two environments.⁵ Thus, a thermodynamic cycle is considered as in Scheme 1, where ΔG_A and ΔG_B are the free energies of transfer of A and B from medium 1 to medium 2. ΔG_1 and ΔG_2 are computed, and the medium effect is given by $\Delta \Delta G = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A$. Gibbs free energies are shown here, which correspond to standard experimental conditions and to use of the isothermal–isobaric ensemble in the statistical mechanics simulations. The energetics for A and B can be represented by molecular mechanics (MM) or quantum mechanics (QM).¹⁷ The former has been the most used as it allows the treatment of large systems and the creation and deletion of atoms. When MM is used for the solutes, the solute–solvent

and solvent–solvent interactions are also represented with MM. With MM, raw energy changes between two different molecules, even isomers, are normally not meaningful by themselves, and it is essential to perform two perturbations, such as ΔG_1 and ΔG_2 , to cancel out artifacts of the arbitrary energy reference with force fields and conceptual omissions such as resonance effects.

The first FEP computations of the type represented by Scheme 1 were not carried out for a molecular system until 1985. The initial application was the computation of the relative free energies of hydration of ethane (A) and methanol (B) such that medium 1 was the gas phase and medium 2 was water; the aqueous FEP calculations were carried out in a periodic cube containing 125 TIP4P water molecules.⁵ This paper also introduced a straightforward scheme for the conversion of one molecule to another, which later became known as single-topology FEP calculations.² Specifically, a coupling parameter λ_i was used to linearly scale all geometrical and force-field parameters χ from those for A to those for B (eq 6). Owing to poor convergence for

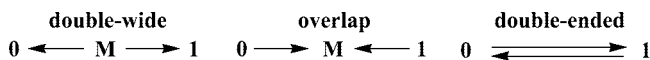
$$\chi_i = \lambda_i \chi_B + (1 - \lambda_i) \chi_A \quad (6)$$

eq 1 when the initial and final states are too different, it is necessary to carry out a series of simulations for multiple values of λ_i spanning between A ($\lambda_i = 0$) and B ($\lambda_i = 1$). The individual simulations are called “windows” or “stages”, and choices have to be made for the individual λ_i for a particular FEP series. The computed free-energy changes are increasingly precise as more windows are used; however, the total cost of a FEP calculation increases linearly with the number of windows. The resulting computational demands were sufficient that calculations like the ethane to methanol example were not attempted before 1985, and they only became routine with the advent of multiprocessor computer systems.

In order to minimize the computational demands, the 1985 paper also introduced “double-wide” sampling.⁵ In principle, the FEP calculations could be performed in the forward or reverse directions. E.g., if a constant $\Delta \lambda = 0.25$ were used, four windows could be executed to perturb $\lambda = 0 \rightarrow 0.25 \rightarrow 0.50 \rightarrow 0.75 \rightarrow 1.0$ or in reverse, $\lambda = 1 \rightarrow 0.75 \rightarrow 0.50 \rightarrow 0.25 \rightarrow 0.0$. This is sometimes referred to as “direct sampling”. However, it was recognized that only two simulations were actually required at $\lambda = 0.25$ and 0.75 , if forward and reverse perturbations were evaluated simultaneously using “double-wide” sampling, $0 \leftarrow 0.25 \rightarrow 0.50$ and $0.50 \leftarrow 0.75 \rightarrow 1.0$. If one does the FEP calculations on-the-fly using all configurations in the averaging rather than postprocessing saved configurations, more bookkeeping is required. Thus, in the BOSS program,¹⁸ coordinates for the reference (e.g., $\lambda = 0.25$) and two perturbed solutes (e.g., $\lambda = 0.0$ and 0.50) are maintained to evaluate both free-energy changes on-the-fly. Furthermore, it is expected that with use of a small number of windows, systematic errors arise such that the cumulative free-energy change for the forward FEP series is not the same as the negative of the result from the reverse series. This was shown to be the case for the ethane to methanol example and that double-wide sampling provides a more accurate estimate of the true free-energy change.

Alternatively, “double-ended sampling”¹⁹ can be employed, which refers to performing both the $\lambda_i \rightarrow \lambda_j$ perturbation and its reverse, $\lambda_j \rightarrow \lambda_i$, and taking $\Delta G(\lambda_i \rightarrow \lambda_j) = (\Delta G(\lambda_i \rightarrow \lambda_j) - \Delta G(\lambda_j \rightarrow \lambda_i))/2$; however, this again requires twice as many simulations as double-wide sampling and has no obvious advantages for accuracy.

Scheme 2



Overlap Sampling. A variant of these sampling choices is known as “overlap sampling”. It is similar to double-ended sampling but to an intermediate point M rather than the end points 0 and 1 (Scheme 2).^{2,20,21} From the Zwanzig expression (eq 1), the free-energy change for $0 \rightarrow 1$ from overlap sampling (OS) is then given by eq 7, while double-wide (DW) and double-ended (DE) sampling yield eqs 8 and 9.

$$\Delta G^{\text{OS}} = -kT \ln \left[\frac{\langle \exp(-(E_M - E_0)/kT) \rangle_0}{\langle \exp(-(E_M - E_1)/kT) \rangle_1} \right] \quad (7)$$

$$\Delta G^{\text{DW}} = kT \ln \left[\frac{\langle \exp(-(E_0 - E_M)/kT) \rangle_M}{\langle \exp(-(E_1 - E_M)/kT) \rangle_M} \right] \quad (8)$$

$$\Delta G^{\text{DE}} = -\frac{1}{2}kT \ln \left[\frac{\langle \exp(-(E_1 - E_0)/kT) \rangle_0}{\langle \exp(-(E_0 - E_1)/kT) \rangle_1} \right] \quad (9)$$

There appears to be little difference between eqs 7 and 8 except that the sampling is performed at M for DW and at the end points for OS and that only one simulation is required for DW versus two for OS. However, DW requires a geometrical definition of point M at, e.g., $\lambda = 0.5$, while M can be more conceptual for OS. Specifically, M can be defined by a parametric expression such as eq 10, and λ_M could be chosen to minimize the statistical noise (random

$$E_M = \lambda_M E_1 + (1 - \lambda_M) E_0 \quad (10)$$

error) in the results from application of eq 7. It is desirable for the energy distributions for M and the end points to overlap as much as possible, so a reasonable choice is $\lambda_M = 0.5$ and then $E_M = (E_0 + E_1)/2$. With this assignment, eq 7 becomes eq 11.

$$\Delta G^{\text{OS}} = -kT \ln \left[\frac{\langle \exp(-(E_1 - E_0)/2kT) \rangle_0}{\langle \exp(-(E_0 - E_1)/2kT) \rangle_1} \right] \quad (11)$$

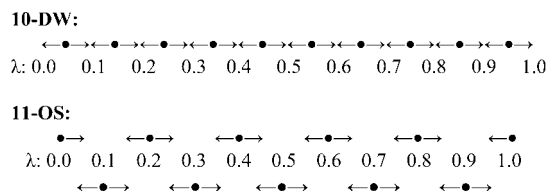
The advantage of eq 11 over DE (eq 9) is clear since the energy differences inside the averages are now divided by two. The advantage over eq 8 depends on the geometrical choice of M for the DW calculations; it is likely that the usual, convenient choice of $\lambda = 0.5$ is less optimal than the choice of $\lambda_M = 0.5$ in eq 10. Indeed, Lu et al. refer to use of eq 11 as “simple overlap sampling”^{20,21} and claims have been made that it is “usually markedly more accurate than simple sampling”,² i.e., direct averaging, that DW is just a “time-saving implementation of direct averaging”,²¹ and that “the direct averaging method can thus be seen as not reliable”.²¹ Though convergence of eq 11 (OS) is expected to be better

Table 1. Increments $\Delta\lambda$ Between Simulations^a

method	$\Delta\lambda$	method	$\Delta\lambda$
10-DW	0.10	6-OS	0.20
14-DW ^b	0.10	11-OS	0.10
25-DW	0.04	15-OS ^b	0.10
5-OS	0.25	21-OS	0.05

^a n -DW and n -OS are double-wide and overlap sampling using n simulations. For DW and OS, the first simulations occur at $\lambda = \Delta\lambda/2$ and $\lambda = 0$, respectively, and the spacing between simulations is $\Delta\lambda$. ^b $\Delta\lambda = 0.05$ is used for $\lambda = 0$ to 0.20 and $\lambda = 0.80$ to 1.0.

Scheme 3



than for eq 9 (DE), the case versus DW is less clear, particularly when a similar number of simulations is run. This has not been well tested previously, so results for a large set of comparisons are reported here. Specifically, FEP calculations have been performed with DW and OS alternatives for 30 conversions of substituted benzenes ($\text{PhX} \rightarrow \text{PhY}$) in the gas phase and in water, which yield relative free energies of hydration via Scheme 1.⁵ These are fundamental conversions that are used routinely, for example, in FEP-guided design of enzyme inhibitors,²² so establishment of optimal protocols is particularly important. In addition, the performance of the expansion given by the Peierls equation has also been tested to first and second order.

Computational Details

The substituted benzenes were represented by the OPLS-AA force field²³ and the water molecules by the TIP4P model.²⁴ Metropolis Monte Carlo statistical mechanics (MC) calculations were performed for a single solute in a periodic cube with 500 water molecules at 25 °C and 1 atm in the isothermal–isobaric ensemble using the BOSS program.¹⁸ The water–water cutoff was at 10 Å based on the OO distance, and the solute–water interactions were included if any non-hydrogen atom of the solute was within 10 Å of the water O. The interactions were quadratically smoothed to zero within 0.5 Å of the cutoff. For each FEP window, a MC simulation was performed with equilibration for 3 million configurations and averaging for a subsequent 10 million configurations; this requires ca. 35 min on a 3.0 GHz Pentium D processor. Our experience is that runs of this length are necessary to allow adequate sampling for different hydrogen-bonded states for a solute. For example, alcohols typically participate in one to three hydrogen bonds with water molecules, and these states do not rapidly interconvert. If the MC or molecular dynamics run is not long enough to sample well the different states, then significant errors can arise in computed free-energy changes independent of the sampling details, e.g., DW or OS.

The TIP4P water molecules underwent only rigid-body translations and rotations, while the sampling of the solutes

included all internal degrees of freedom as well as the total translations and rotations. Solute and volume moves were attempted every 120 and 3125 configurations, respectively. Acceptance rates of 30–50% for new configurations were obtained by use of ranges for translations and rotations of ± 0.06 Å and $\pm 6^\circ$ for the solutes and ± 0.15 Å and $\pm 15^\circ$ for the water molecules. For the gas-phase MC calculations, each window consisted of an equilibration phase of 0.15 million configurations followed by averaging over an additional 1 million configurations. Mutations involving different numbers of atoms for PhX and PhY were performed using dummy atoms that have no nonbonded parameters (Lennard-Jones and Coulomb); the r_0 values for bond lengths to dummy atoms are taken as 0.3 Å. The reported uncertainties in the averages ($\pm 1\sigma$) for the free energy changes were obtained from the fluctuation in separate averages over batches of 2 million configurations.^{4b,25} Eq 12 is used where m is the number of batches, θ_i is the average of property θ for the i th batch, and $\langle\theta\rangle$ is the overall average for θ . These are all standard procedures that have been followed in prior computations of relative free energies of hydration.²⁶ It may be noted that there are no “end-state problems when $\lambda = 0$ or 1”²⁰ with the MC simulations since calculations of forces are not required.

$$\sigma^2 = \sum_i^m (\theta_i - \langle\theta\rangle)^2 / m(m-1) \quad (12)$$

Eight FEP protocols were examined as summarized in Table 1. DW calculations were executed using 10, 14, and 25 windows. For the 10-window calculations, the first window is run at $\lambda = 0.05$ with perturbations to $\lambda = 0.0$ and 0.10; the second window is at $\lambda = 0.15$, etc. For the 25-window case, the first simulation is run at $\lambda = 0.02$, and the spacing is also even with $\Delta\lambda = 0.04$. The 14-window alternative has become a standard one in our inhibitor design work;²² this is the same as the 10-window case except that the spacing is reduced to $\Delta\lambda = 0.05$ for $\lambda = 0$ to 0.2 and 0.8 to 1.0. I.e., simulations are run at $\lambda = 0.025, 0.075, 0.125, 0.175, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.825, 0.875, 0.925$, and 0.975. This choice reflects the fact that the free-energy changes are most rapid near the end points as atoms are created or annihilated or atomic charges are changed when eq 6 is applied. OS calculations were executed using eq 11 and 5, 6, 11, 15, and 21 windows. The λ -spacing was uniform for all cases except 15-OS, which is the same as 11-OS with the spacing halved for $\lambda = 0$ to 0.2 and 0.8 to 1.0 as for 14-DW. For example, the simulations for 11-OS were run at $\lambda = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$, and 1.0, and two free-energy increments were computed at each λ as for DW (Scheme 3). In each case, the individual free-energy changes are summed to yield the total result.

An example of the distributions of total solute–water interaction energies from such a series of simulations is presented in Figure 1. There is clearly substantial overlap in the distributions for simulations at adjacent λ values even for a case like the illustrated one where the hydrogen-bonding abilities of the limiting solutes are significantly different. The illustrated example is, in fact, the worst case for the perturbations in Table 1 in that the average solute–water

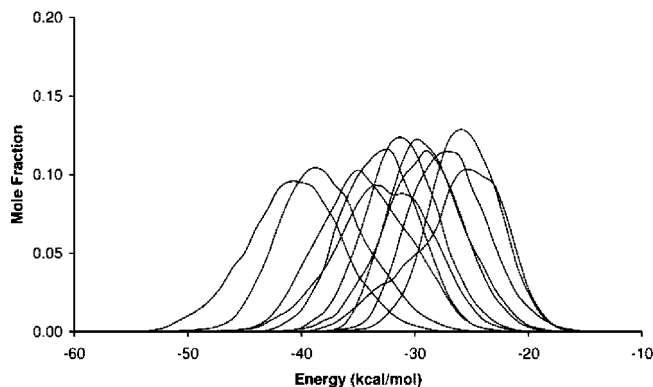


Figure 1. Distributions for the total solute–water interaction energies for the 11 windows in the 11-OS calculations for the PhCOMe → PhCONH₂ conversion. The distributions progress from that for $\lambda = 0.0$ (acetophenone) on the far right to $\lambda = 1.0$ (benzamide) on the far left. Units for the y-axis are mole fraction per kcal/mol.

interaction energy shifts the most from $\lambda = 0$ to $\lambda = 1$. For cases that do not involve large differences in hydrogen bonding, the centers of the distributions for $\lambda = 0$ and $\lambda = 1$ are only shifted by a few kcal/mol.

FEP Results

FEP calculations were executed for the 28 PhX → PhY changes listed in Table 2. The perturbations involve common substituents, and no more than one non-hydrogen atom is deleted or created in each case. Two forward and backward cases are considered for PhCl ↔ PhH and PhOH ↔ PhOMe; the results are not expected to be identical in magnitude because in the present protocol the simulations were executed sequentially starting from $\lambda = 0$ (PhX) and the last configuration from a completed window is used as the starting point for equilibration of the next window. The results in Table 2 are the differences in free energies of hydration for PhX and PhY, $\Delta\Delta G_{\text{hyd}}$, which are computed from the difference in the FEP results in water and in the gas phase, $\Delta G(\text{aq}) - \Delta G(\text{gas})$. There is essentially no statistical uncertainty in the gas-phase results; the uncertainties discussed below arise predominantly from the simulations in water.

Two degenerate cases were also studied in depth; PhCl was converted to itself by replacing the chlorine at C1 with hydrogen and replacing the hydrogen at C4 with chlorine. Similarly, phenol was converted to itself by analogous shrinking and growing of the hydroxyl group. The exact result is obviously $\Delta G = 0$ in these degenerate cases. The results for $\Delta G(\text{aq})$ are listed in Table 3 along with the statistical uncertainty ($\pm 1\sigma$) for the $\Delta G(\text{aq})$ as computed from eq 12. Each protocol was executed at least twice with small changes to the ranges for solute translations and rotations to yield independent estimates of the ΔG values. As first indications of the precision of the alternative protocols, the average absolute value of the ΔG results in Table 3 are listed in Table 4. The average statistical uncertainty for each method from the 28 results in Table 2 is also recorded in Table 4. Some salient points follow. (1)

Table 2. Computed and Experimental Differences in Free Energies of Hydration (kcal/mol) for Substituted Benzenes^a

X →	Y	10-DW	14-DW	25-DW	6-OS	11-OS	15-OS	21-OS	exptl ^b
Br	Cl	0.60	0.61	0.57	0.63	0.59	0.59	0.57	0.33
CH ₃	CF ₃	1.20	0.99	1.45	1.35	1.27	1.36	1.33	0.64
Cl	F	-0.13	-0.18	-0.10	0.00	0.00	-0.01	-0.12	0.31
Cl	H	-0.38	-0.20	-0.41	-0.77	-0.57	-0.58	-0.41	0.26
H	Cl	0.53	0.32	0.31	0.65	0.48	0.23	0.39	-0.26
CN	F	2.70	2.68	2.66	2.56	2.43	2.77	2.52	3.41
COMe	CONH ₂	-7.58	-6.45	-6.72	-7.31	-6.53	-6.42	-6.55	-6.42
COMe	NO ₂	0.78	0.80	0.48	0.89	0.72	0.66	0.95	0.46
Et	Me	-0.42	-0.50	-0.35	-0.35	-0.27	-0.32	-0.29	-0.10
Et	OMe	-0.42	-0.58	-0.58	-0.67	-0.84	-0.18	-0.51	-1.67
F	H	-0.23	-0.26	-0.40	-0.40	-0.21	-0.34	-0.36	-0.05
<i>i</i> -Pr	Et	-0.36	-0.13	-0.27	-0.28	-0.38	-0.59	-0.12	-0.49
Me	Cl	1.03	0.37	0.63	0.69	0.69	0.95	0.86	-0.23
Me	H	0.42	0.48	0.34	0.45	0.23	0.44	0.22	0.03
Me	NH ₂	-4.08	-4.48	-4.28	-4.09	-4.40	-4.32	-4.32	-4.60
MeC=CH ₂	COMe	-4.51	-3.55	-4.28	-4.16	-4.17	-3.64	-4.19	-3.34
NH ₂	OH	-0.54	-0.57	-0.48	-0.41	-0.18	-0.70	-0.38	-1.13
NHMe	NH ₂	-0.52	-0.92	-0.99	-0.91	-1.25	-1.12	-0.88	-0.80
NMe ₂	NHMe	-1.80	-0.85	-0.95	-0.70	-1.10	-1.22	-0.99	-1.24
NMe ₂	NO ₂	0.44	-0.06	0.46	0.33	0.77	0.32	0.13	-0.67
OH	H	5.45	5.22	5.35	5.20	5.03	5.29	4.97	5.76
OH	F	5.22	5.66	5.35	4.89	5.51	5.38	5.54	5.81
OH	OMe	4.41	4.36	4.82	4.55	4.46	4.70	4.29	4.16
OMe	OH	-4.67	-4.15	-4.70	-4.46	-4.35	-4.77	-4.77	-4.16
Pr	Et	-0.18	-0.65	-0.30	-0.43	-0.41	-0.48	-0.64	-0.49
SH	OH	-5.03	-5.22	-4.93	-5.06	-5.10	-5.20	-5.18	-4.07
SMe	OMe	0.30	0.32	0.05	0.03	0.03	0.07	0.09	0.27
Pr	CH ₂ OMe	-2.39	-2.44	-2.48	-2.44	-2.41	-2.64	-2.44	

^a Results for PhX → PhY; *n*-DW and *n*-SOS are FEP results using double-wide sampling and overlap sampling with *n* simulations.^b Reference 27.**Table 3.** Computed Free Energy Changes and Their Uncertainties^a in Water for Degenerate Mutations (kcal/mol)

	PhCl → ClPh		PhOH → HOPh	
	ΔG	σ	ΔG	σ
10-DW-1	0.11	0.09	-0.86	0.15
10-DW-2	-0.24	0.07	0.38	0.16
14-DW-1	0.13	0.08	-0.29	0.16
14-DW-2	-0.06	0.09	-0.11	0.14
25-DW-1	0.04	0.05	-0.40	0.09
25-DW-2	0.12	0.05	-0.04	0.10
5-OS-1	0.20	0.16	1.12	0.23
5-OS-2	-0.35	0.13	-0.44	0.25
6-OS-1	-0.16	0.10	-0.56	0.27
6-OS-2	-0.22	0.12	0.12	0.24
11-OS-1	-0.07	0.09	-0.18	0.20
11-OS-2	0.01	0.09	-0.27	0.21
11-OS-3	0.19	0.10	0.31	0.18
11-OS-4	-0.06	0.08	-0.56	0.24
15-OS-1	-0.32	0.09	0.11	0.15
15-OS-2	-0.03	0.09	0.04	0.16
21-OS-1	0.13	0.14	-0.15	0.07
21-OS-2	0.00	0.07	0.27	0.11

^a From eq 12.

The trends in the results are reasonable with the random error decreasing with the number of windows for both DW and OS. (2) The errors are larger for phenol since the removal/growth of a hydrogen-bonding group is a larger perturbation than interconversion of chlorine and hydrogen. (3) The uncertainty based on the fluctuations in the averages, σ , is somewhat less than the uncertainties indicated by the deviations of the ΔG values in Tables 3 and 4 from zero. (4) There is little difference in the quality of results from 14-DW, 15-OS, 21-OS, and 25-DW. The average error of

Table 4. Average Absolute ΔG for the Degenerate Mutations in Table 3 and Statistical Uncertainties for the 28 FEP Calculations in Table 2 (kcal/mol)

method	ΔG	σ	method	ΔG	σ
10-DW	0.40	0.12	6-OS	0.27	0.22
14-DW	0.15	0.12	11-OS	0.21	0.13
25-DW	0.15	0.07	15-OS	0.13	0.11
5-OS	0.53	0.25	21-OS	0.14	0.09

Table 5. Average Absolute and RMS Deviations for Computed and Experimental Free Energy Changes (kcal/mol)^a

	10-DW	14-DW	25-DW	6-OS	11-OS	15-OS	21-OS	exptl
10-DW	0	0.30	0.21	0.21	0.27	0.26	0.24	0.58
14-DW	0.43	0	0.23	0.25	0.24	0.23	0.18	0.39
25-DW	0.30	0.29	0	0.16	0.18	0.17	0.15	0.51
6-OS	0.28	0.34	0.23	0	0.18	0.23	0.18	0.58
11-OS	0.35	0.30	0.20	0.27	0	0.20	0.15	0.50
15-OS	0.36	0.28	0.22	0.32	0.27	0	0.18	0.48
21-OS	0.33	0.24	0.21	0.26	0.22	0.24	0	0.51
exptl	0.67	0.50	0.61	0.68	0.64	0.61	0.63	0

^a Using the data in Table 2. Average absolute deviations in upper triangle, rms deviations in lower triangle. Deviations with 25-DW and 21-OS are in bold.

ca. 0.15 kcal/mol in Table 4 appears to reflect the limit for the simulations of the present length, independent of the number of windows. (5) The errors with 5-OS and 10-DW are likely too large for these procedures to be generally useful, while 6-OS warrants further consideration.

Turning back to Table 2, the $\Delta\Delta G_{\text{hyd}}$ results are provided for all methods except 5-OS. Table 5 summarizes the results by providing the average unsigned and rms deviations

Table 6. Average Absolute and RMS Deviations between Computed Free Energy Changes (kcal/mol)^a

	6-OS-1	6-OS-2	11-OS-1	11-OS-2	11-OS-3	11-OS-4	25-DW	21-OS
6-OS-1	0	0.30	0.18	0.18	0.23	0.20	0.16	0.18
6-OS-2	0.36	0	0.26	0.30	0.30	0.24	0.26	0.27
11-OS-1	0.27	0.32	0	0.16	0.18	0.17	0.18	0.15
11-OS-2	0.26	0.35	0.22	0	0.16	0.19	0.17	0.13
11-OS-3	0.31	0.37	0.23	0.20	0	0.18	0.16	0.14
11-OS-4	0.29	0.30	0.22	0.22	0.24	0	0.16	0.15
25-DW	0.23	0.32	0.20	0.21	0.19	0.22	0	0.15
21-OS	0.26	0.33	0.22	0.16	0.18	0.18	0.21	0

^a *m*-OS-*n* refers to *n* independent *m*-OS evaluations for the 28 PhX → PhY FEP calculations in Table 2. Average absolute deviations in upper triangle, rms deviations in lower triangle.

between the seven sets of computed results and the experimental data. Agreement with experiment is not important for the present methodological testing, though it is noted that the average error of 0.5 kcal/mol for $\Delta\Delta G_{\text{hyd}}$ is typical with OPLS-AA²⁶ and it was also found previously in an extensive study by Price and Brooks on the performance of OPLS-AA for the properties of liquid substituted benzenes and their free energies of hydration.²⁸ Again, there are many points to highlight. (1) The average deviation between the 21-OS and 25-DW results is the smallest, 0.15 kcal/mol, which reinforces this figure as the precision limit for the present methods. The smallest rms deviation for 21-OS is also with 25-DW, 0.21 kcal/mol. Based on the results in Tables 2, 3, 4, and 5, 21-OS and 25-DW can be assumed to be the most accurate procedures. (2) The average deviations between 21-OS and 25-DW and the other procedures (10-DW, 14-DW, 6-OS, 11-OS, 15-OS) are all below 0.25 kcal/mol. The rms deviations are all below 0.30 kcal/mol except with 10-DW, which appears to be the least accurate of these alternatives. (3) The results in the two cases where the PhX to PhY and PhY to PhX FEP calculations were performed, chlorobenzene to benzene and phenol to anisole, are also near the precision limit with average discrepancies of 0.1–0.2 kcal/mol. (4) The largest single discrepancies with the 21-OS results are 1.03 (10-DW), 0.63 (14-DW), 0.53 (25-DW), 0.76 (6-OS), 0.65 (11-OS), and 0.55 kcal/mol (15-OS). The perturbations that show the largest oscillations in the results are acetophenone to benzamide, α -methylstyrene to acetophenone, and *N,N*-dimethylaniline to *N*-methylaniline; all involve substantial changes in hydrogen-bonding character and deletion of atoms. (5) The average deviations between the results with 5-OS (not shown) and 25-DW and 21-OS are 0.29 and 0.31 kcal/mol, and the rms deviations are 0.42 and 0.46 kcal/mol. The largest discrepancy with the 21-OS result is now 1.67 kcal/mol for the phenol to anisole conversion. Thus, 5-OS is reinforced as unacceptable for such FEP calculations. 10-DW is the next least acceptable. 6-OS performs notably well, while 14-DW, 11-OS, and 15-OS are comparably more reliable.

As a further check for consistency, the full set of 28 FEP calculations was repeated with the 11-OS procedure three more times with changes to the ranges for solute moves. The average unsigned and rms deviations for the four 11-OS data sets and the 25-DW and 21-OS results are presented in Table 6. The consistency is excellent with all rms values being close to 0.20 kcal/mol. The 6-OS calculations were also repeated; the rms discrepancies for this second set of the 28 results increased to 0.32 and 0.33 kcal/mol in comparison

to the results from 25-DW and 21-OS. Overall, in our view, 11-OS, 14-DW, and 15-OS are all viable procedures for routine FEP calculations of relevance to molecular design. The difference in performance between DW and OS is insignificant for the present typical FEP calculations with the number of windows in this range. Furthermore, the slightly greater accuracy obtainable with 21-OS or 25-DW does not justify the increased costs, while 6-OS may be useful for preliminary exploration of large numbers of mutations. An 8-OS alternative with the first and last window divided in half was also tried for the 28 perturbations and yielded rms discrepancies of 0.29 and 0.33 kcal/mol with the 25-DW and 21-OS results.

It may be noted that the present limit of precision of about 0.2 kcal/mol with 11-OS, 14-DW, and 15-OS is adequate in many situations. E.g., for computation of relative free energies of binding for two inhibitors, FEP calculations are required for the inhibitors by themselves in water and bound to the biomolecule. The total uncertainty would then be $(0.2^2 + 0.2^2)^{1/2} = 0.3$ kcal/mol corresponding to a factor of 1.6 in inhibition constants, which is at the limit of typical experimental assays. If improved precision is needed, then the length of the simulation for each window should be increased with the usual expectation that a factor of 2 improvement in precision requires at least a factor of 4 increase in sampling.^{2,4}

Results with the Peierls Equation

Application of eq 3 was also considered, though there is no practical advantage over use of the exact eq 1 since the MC or MD simulations still need to be run and the additional cost of averaging the quantities in either equation is insignificant. However, the performance of the expansion is of fundamental interest and has not been examined previously for a series of typical A → B FEP calculations. Thus, for each window in the 10-DW and 14-DW calculations, the energy changes between the reference and perturbed states and their squares were also averaged. The resulting differences between the aqueous- and gas-phase calculations are summarized in Table 7 for the 14-DW case; the comparisons are with the $\Delta\Delta G_{\text{hyd}}$ values calculated using 14-DW and eq 1.

The results from just the first-order term, the average energy changes, are very good with an average unsigned deviation of only 0.15 kcal/mol to the free-energy results from eq 1. However, in the case of phenol to anisole and its reverse the favorability of the hydration of phenol is

Table 7. Differences in Free Energies of Hydration (kcal/mol) for Substituted Benzenes Using Eq 1 and Its Expansion in Eq 3^a

X →	Y	ΔE	$\Delta\Delta G$ (eq 3)	$\Delta\Delta G$ (eq 1)
Br	Cl	0.61	0.61	0.61
CH ₃	CF ₃	1.00	0.98	0.99
Cl	F	-0.17	-0.16	-0.18
Cl	H	-0.22	-0.16	-0.20
H	Cl	0.34	0.28	0.32
CN	F	2.13	3.77	2.68
COMe	CONH ₂	-6.46	-6.42	-6.45
COMe	NO ₂	0.77	0.86	0.80
Et	Me	-0.59	-0.41	-0.50
Et	OMe	-0.61	-0.56	-0.58
F	H	-0.27	-0.26	-0.26
i-Pr	Et	-0.21	-0.05	-0.13
Me	Cl	0.35	0.38	0.37
Me	H	0.40	0.58	0.48
Me	NH ₂	-4.53	-4.42	-4.48
MeC=CH ₂	COMe	-3.57	-3.55	-3.55
NH ₂	OH	-0.57	-0.58	-0.57
NHMe	NH ₂	-1.33	-0.06	-0.92
NMe ₂	NHMe	-1.34	0.85	-0.85
NMe ₂	NO ₂	-0.15	0.09	-0.06
OH	H	5.20	5.22	5.22
OH	F	5.65	5.65	5.66
OH	OMe	5.40	3.94	4.36
OMe	OH	-5.22	-1.57	-4.15
Pr	Et	-0.73	-0.57	-0.65
SH	OH	-5.21	-5.20	-5.22
SMe	OMe	0.32	0.33	0.32
Pr	CH ₂ OMe	-2.46	-2.42	-2.44
	aad ^b	0.15	0.27	0
	rms	0.30	0.63	0

^a Results for PhX → PhY from the MC simulations with 14 windows. $\Delta\Delta G$ (eq 1) is the exact result using eq 1, while $\Delta E = \langle V \rangle$ in eq 3. ^b Average absolute deviation.

underestimated by 1.1 kcal/mol. The results get significantly worse when the full second-order treatment (eq 3) is used owing to poor convergence of the fluctuation term. The latter is basically a ΔC_V correction that only changes for successful solute moves. Thus, the convergence is expected to be significantly worse than for the computation of a pure liquid's heat capacity, which is well-known to be slow.^{4b,29}

There are now more cases with serious problems including PhCN → PhF, PhNMe₂ → PhNHMe, and PhOMe → PhOH; and, the rms deviation rises from 0.30 kcal/mol with the first-order estimate to 0.63 kcal/mol including both terms in eq 3. The situation further deteriorates when only 10 windows are used with the full eq 3; e.g., a 10 kcal/mol error was obtained for the PhNMe₂ → PhNHMe case, though the average error with only the first-order term remains acceptable at 0.19 kcal/mol. On the other hand, if changes that involve OH, NHR, and CN groups are excluded, it is apparent from Table 7 that there is negligible difference between the results from the Zwanzig equation (eq 1) and those from the first- or second-order expansion in the Peierls equation. The greater ranges of solute–water interactions for strongly hydrogen-bonding solutes are problematic for application of eq 3.

Conclusion

An overview has been provided on the history of free-energy perturbation (FEP) calculations and their application to

computation of relative free energies of solvation. In view of the expanding role of FEP calculations for pursuing a wide range of problems in chemistry and biochemistry including molecular design, it is important to establish reliable protocols for their execution. For 30 interconversions of substituted benzenes, which represent prototypical cases for molecular design, alternative sampling procedures and numbers of FEP windows were considered. Simple overlap sampling with 11 or 15 windows and double-wide sampling with 14 windows emerged as viable procedures for such routine FEP calculations; the present protocols with these sampling methods yield statistical uncertainties of 0.2 kcal/mol for free-energy differences. The slightly greater accuracy obtainable with 21 to 25 windows does not clearly justify the increased costs, while overlap sampling with 6 or 8 windows may be useful for preliminary exploration of large numbers of molecular changes. Application of the power-series expansion of the Zwanzig equation was also tested. Use of the first-order term is generally reliable, while inclusion of the slowly convergent, second-order fluctuation term causes deterioration in the results for strongly hydrogen-bonded solutes.

Acknowledgment. Gratitude is expressed to the National Science Foundation (CHE-0446920) and the National Institutes of Health (GM032136) for support of this work and to Drs. A. Alexandrova and B. Albrecht for assistance with translation of text.

References

- (1) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (2) For a recent review, see: Chipot, C.; Pohorille, A. In *Springer Series in Chemical Physics, Vol. 86: Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer-Verlag: Berlin, 2007; pp 33–75.
- (3) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (4) (a) Jorgensen, W. L. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: New York, 1998; Vol. 2, pp 1061–1070. (b) Jorgensen, W. L. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: New York, 1998; Vol. 2, pp 1754–1763. (c) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (5) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- (6) Levy, R. M.; Belhadj, M.; Kitchen, D. B. *J. Chem. Phys.* **1991**, *95*, 3627–3633.
- (7) (a) Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 85–97. (b) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.
- (8) Simonson, T.; Wong, C. F.; Brünger, A. T. *J. Phys. Chem. A* **1997**, *101*, 1935–1945.
- (9) For biographies of Landau and Peierls, see ref. 10–12.
- (10) Livanova, A. *Landau: A Great Physicist and Teacher*, translated by Sykes, J. B.; Pergamon Press: Elmsford, NY, 1980.

- (11) Khalatnikov, I. M. *Landau: The Physicist and the Man*, translated by Sykes, J. B.; Pergamon Press: Elmsford, NY, 1989.
- (12) Peierls, R. *Bird of Passage: Recollections of a Physicist*; Princeton University Press: Princeton, NJ, 1985.
- (13) Landau, L. D.; Lifshitz, E. M. *Statisticheskaya fizika, klassicheskaya i kvantovaya*; Gos. izd. tekhniko-teoret. lit-ry.; Moskva, 1951; pp 112–115.
- (14) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics, Vol. 5: Statistical Physics*; translated from Russian by Peierls, E., Peierls, R. F.; Pergamon Press: London, 1958; pp 93–95.
- (15) Landau, L. D.; Lifshitz, E. M., *Course of Theoretical Physics, Vol. 5: Statistical Physics*; translated from Russian by Sykes, J. B., Kearsley, M. J.; Pergamon Press: Elmsford, NY, 1969; pp 90–93.
- (16) (a) Peierls, R. *Z. Phys.* **1933**, *80*, 763–791. (b) *Selected Scientific Papers of Sir Rudolf Peierls: With Commentary (Series in 20th Century Physics)*; translated from German by Dalitz, R. H.; Peierls, R. E., Ed.; World Scientific Publishing: Singapore, 1997.
- (17) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **205**, *102*, 6665–6670.
- (18) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700.
- (19) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (20) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28–39.
- (21) Lu, N.; Woolf, T. B. In *Springer Series in Chemical Physics, Vol 86: Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer-Verlag: Berlin, 2007; pp 199–247.
- (22) (a) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basava-pathruni, A.; Anderson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663–667. (b) Thakur, V. V.; Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5664–5667. (c) Barreiro, G.; Kim, J. T.; Guimaraes, C. R. W.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *J. Med. Chem.* **2007**, *47*, 2416–2428.
- (23) (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. (b) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. *J. Comput. Chem.* **1993**, *14*, 206–215. (c) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264–16270.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (25) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Clarendon: Oxford, 1987; pp 191–198.
- (26) (a) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195–205. (b) Udier-Blagovic, M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322–1332.
- (27) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.
- (28) Price, D. J.; Brooks, C. L., III *J. Comput. Chem.* **2005**, *26*, 1529–1541.
- (29) Jorgensen, W. L. *Chem. Phys. Lett.* **1982**, *92*, 405–410.

CT800011M