The Development of Replica-Exchange-Based Free-Energy Methods

Christopher J. Woods and Jonathan W. Essex*

Department of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

Michael A. King

Celltech Group plc, 208 Bath Road Slough, SL1 3WE, United Kingdom

Received: June 11, 2003; In Final Form: October 3, 2003

The calculation of relative free energies that involve large reorganizations of the environment is one of the great challenges of condensed-phase simulation. Such calculations are of particular importance in protein−ligand free-energy calculations. To meet this challenge, we have developed new free-energy techniques that combine the advantages of the replica-exchange method with free-energy perturbation (FEP) and finite-difference thermodynamic integration (FDTI). These new techniques are tested and compared with FEP, FDTI, and the adaptive umbrella weighted histogram analysis method (AdUmWHAM) on the challenging calculation of the relative hydration free energy of methane and water. This calculation involves a large solvent configurational change. Through the use of replica-exchange moves along the λ-coordinate, the configurations sampled along λ are allowed to mix, which leads to dramatic improvements in solvent configurational sampling, an efficient reduction of random sampling error, and a reduction of general simulation error. This is achieved at effectively no extra computational cost, relative to standard FEP or FDTI.

1. Introduction

The calculation of relative free energies is still one of the great challenges of condensed-phase simulation. Of particular difficulty is the calculation of relative free energies that involve substantial reorganization of the environment, as in the case of the binding of different ligands to a protein. This leads to restricted sampling, and the system can become locked in local minima. Recently, replica-exchange[1,2] (RE) and parallel tempering[3,4] (PT) methods have been proposed, each of which improves sampling. These methods achieve this feat by running multiple replicas of the system at different temperatures and pressures, or with different Hamiltonians, and by periodically swapping the coordinates of selected pairs. The use of these techniques has been shown to improve the quality of conformational sampling of protein and polypeptide systems.[1−7] It is the purpose of this paper to apply these techniques to the calculation of relative free energies, and to demonstrate how their use can significantly improve these calculations when they involve extensive reorganization of the environment. The challenging calculation of the relative hydration free energy of methane and water will be used to illustrate these methods.

2. Theory

2.1. Calculation Methods. The three main methods of calculating relative free energies are free-energy perturbation (FEP),[8−11] thermodynamic integration (TI),[12−14] and adaptive umbrella sampling.[15−19] These methods have been extensively reviewed elsewhere.[20,21] A fourth method calculates the relative free energies via nonequilibrium sampling using Jarzynski's equality.[22−25] Comparisons between this method and replica-exchange-based free-energy methods are a subject for future study.

2.1.1. Free-Energy Perturbation. Free-energy perturbation (FEP)[8−11] calculates the relative free energy of two systems, A and B, via the Zwanzig equation:[8]

$$\Delta G_{A\rightarrow B} = -kT \ln \left\langle \exp\left(-\frac{\Delta E_{AB}}{kT}\right)\right\rangle_A \qquad (1)$$

where the difference in free energy between systems A and B, $\Delta G_{A\rightarrow B}$, is given as the ensemble average of the exponential of the difference in energy of the two systems, ΔE_{AB}, divided by the quantity of the Boltzmann's constant k multiplied by the temperature T. The average is formed over the ensemble generated by system A, which is known as the reference state. System B is known as the perturbed state.

The average formed in this equation will only converge satisfactorily if the reference and perturbed states are similar, and thus the overlap between them is good.[26] To ensure that this is the case, the two systems are connected by a reaction coordinate, λ. This coordinate is typically used to scale the force field of the system, such that (i) at $\lambda = 0.0$, the force field represents system A; (ii) at $\lambda = 1.0$ the force field represents system B, and (iii) at λ values between 0.0 and 1.0, the system is a nonphysical hybrid of A and B. λ does not need to be a scaling coordinate; it could be any reaction coordinate, e.g., an intermolecular distance or dihedral angle. In practice, the reaction coordinate λ is divided into a sequence of windows, and FEP is applied between adjacent windows. The technique of double-wide sampling[27] may be used to calculate the difference in free energy to the next (forward) window and previous (backward) window simultaneously. The sum of all the individual forward free energies should then equal the negative of the sum of all the individual backward free energies. Any difference between the two is a sign of error. This error is known as hysteresis, and it may be caused by insufficient overlap between the reference and perturbed states or by

* Author to whom correspondence should be addressed. E-mail: J.W.Essex@soton.ac.uk.

10.1021/jp0356620 CCC: $25.00
Published on Web 11/14/2003

inadequate sampling of the reference state.[26] Although this error may be reduced by increasing the amount of sampling per window, and increasing the number of windows, such approaches may lead to prohibitively expensive simulations.

2.1.2. Thermodynamic Integration. Thermodynamic integration (TI)[12−14] uses the gradient of the free energy, with respect to λ (denoted as $(\partial G/\partial\lambda)_\lambda$), to construct the potential of mean force (PMF) across λ. This is achieved by calculating the gradients at various points along λ and then numerically integrating them to return the complete PMF. This numerical integration is typically performed via the trapezium rule.[12] The free-energy gradients may be calculated analytically throughout a simulation as the ensemble average of the gradient of the force field.[12] However, this calculation requires significant support from the program used to perform the simulation. As an alternative, the free-energy gradients may be calculated numerically by approximating $(\partial G/\partial\lambda)_\lambda$ by the finite difference, $(\Delta G/\Delta\lambda)_\lambda$. This may be calculated via the Zwanzig equation, by placing the reference state at λ and the perturbed state at $\lambda + \Delta\lambda$. To verify the accuracy of the approximation, both the forward gradient (λ to $\lambda + \Delta\lambda$) and backward gradient (λ to $\lambda - \Delta\lambda$) may be calculated. Assuming that $\Delta\lambda$ is small enough, the forward gradient then should equal the negative of the backward gradient, and thus would be a good estimate of the analytic gradient. This approach is typically called finite-difference thermodynamic integration[28−32] (FDTI). The benefit of FDTI over FEP is that the overlap between the reference and perturbed states is significantly improved through the use of such a small value of $\Delta\lambda$, and, thus, the Zwanzig equation is expected to converge well. The advantage of FDTI over TI is that it may be performed in any FEP-capable simulation package. The disadvantage of TI and FDTI is that the final numerical integration can magnify any random errors in the gradients and, thus, lead to a loss of precision in the results.

2.1.3. Adaptive Umbrella Sampling. Adaptive umbrella sampling, and the related adaptive umbrella weighted histogram analysis method (AdUmWHAM), are described elsewhere.[17−19,33] Both methods treat λ as a dynamic coordinate and perform moves in λ throughout the simulation. An umbrella potential[15] is iteratively refined to encourage even sampling of the λ-coordinate. After this is achieved, the umbrella equals the negative of the potential of mean force along λ, from which the relative free energy may be obtained. AdUmWHAM[18,19] is a modification of adaptive umbrella sampling that uses the weighted histogram analysis method[34] (WHAM) to combine the statistics of all previous iterations efficiently when refining the umbrella. The primary disadvantage of these methods is that the moves along λ are driven by the umbrella potential and, thus, can progress too quickly for the rest of the system to respond, which leads to so-called "Hamiltonian lag". This problem is particularly acute when the relative free-energy calculation involves a significant system configurational change, as is the case for the calculation of the relative hydration free energy of water and methane. The solvent structure for water-in-water is significantly different to that of methane-in-water. Unless care is taken, the umbrella potential may drive the perturbation of water into methane too quickly for the rest of the solvent to respond. This may lead to the sampling of unphysical states, e.g., the sampling of methane hydrated by solvent that is in the water-in-water configuration. To ensure that this is not the case, the rate of change of λ must be sufficiently small to allow the system to relax fully between λ moves. Unfortunately, the relaxation time of the system is difficult to know before the start of a simulation. In addition, the scale of the relaxation time may be such that it would not be possible to perform AdUmWHAM simulations with a low enough rate of λ moves to avoid Hamiltonian lag.

2.2. Replica-Exchange-Based Free-Energy Methods. Three new free-energy methods were developed as part of this work. These combined either parallel tempering (PT)[3,4] or Hamiltonian replica-exchange[1,2] with FEP or FDTI.

2.2.1. Parallel Tempering Thermodynamic Integration. Parallel tempering (PT)[3,4] enhances sampling through the use of multiple replicas of the system, each running at a different temperature (Figure 1).

Periodically, neighboring temperatures are tested according to a PT Monte Carlo (MC) test. If this test is passed, then the coordinates of the system at the two temperatures are swapped. At the end of the simulation, the set of trajectories at each temperature form a correct ensemble distribution at that temperature. The PT test has different forms, depending on the type of the individual sub-ensembles. For example, if replica i, with temperature T_A and energy $E(i)$, was swapping with replica j, with temperature T_B and energy $E(j)$, then the test for a set of NVT ensembles would be[3]

$$\exp\left[\left(\frac{1}{kT_B} - \frac{1}{kT_A}\right)(E(j) - E(i))\right] \geq \text{rand}(0, 1) \qquad (2)$$

where rand(0, 1) represents a uniform random number generated from $0.0 \leq \text{rand}(0, 1) \leq 1.0$. For the NPT ensemble, the test also depends on the pressure, P, and volumes, $V(i)$ and $V(j)$ of the two replicas:[35]

$$\exp\left[\left(\frac{1}{kT_B} - \frac{1}{kT_A}\right)(E(j) - E(i) + P(V(j) - V(i)))\right] \geq \text{rand}(0, 1) \qquad (3)$$

Parallel tempering thermodynamic integration (PTTI) was developed as a version of FDTI that used PT at each of the n_λ values of λ. Thus, there are n_T temperature replicas at each value of λ. The potential advantages of this method should be increased sampling at each λ value, and the calculation of the temperature dependence of the free energy. A disadvantage of this method is that it is very expensive, requiring $n_\lambda \times n_T$ replicas to perform the complete simulation.

2.2.2. Replica Exchange Thermodynamic Integration. The second newly developed free-energy method combines Hamiltonian replica-exchange with FDTI. The use of the λ-coordinate to scale the force field leads to the system having a different Hamiltonian at each value of λ. This means that Hamiltonian replica-exchange may be used within a normal FDTI simulation to test and swap the coordinates of neighboring λ values periodically (Figure 2).

The FDTI simulation may be conducted in the standard manner, with a replica of the system at each value of λ. Unlike standard FDTI, there is the addition of a periodic λ-swap move, whereby neighboring λ-replicas are tested according to the Hamiltonian replica-exchange test. This test is the same for the NVT and NPT ensembles. For replica i at $\lambda = \text{A}$ and energy $E_A(i)$, and replica j at $\lambda = \text{B}$ and energy $E_B(j)$, the test is given by[2]

$$\exp\left[\frac{1}{kT}(E_B(j) - E_B(i) - E_A(j) + E_A(i))\right] \geq \text{rand}(0, 1) \qquad (4)$$

If this test is passed, then the coordinates of the pair of λ values are swapped. The derivation of this test[2] is similar to that used for PT.[3,35] It is useful to note that this test is just the product of

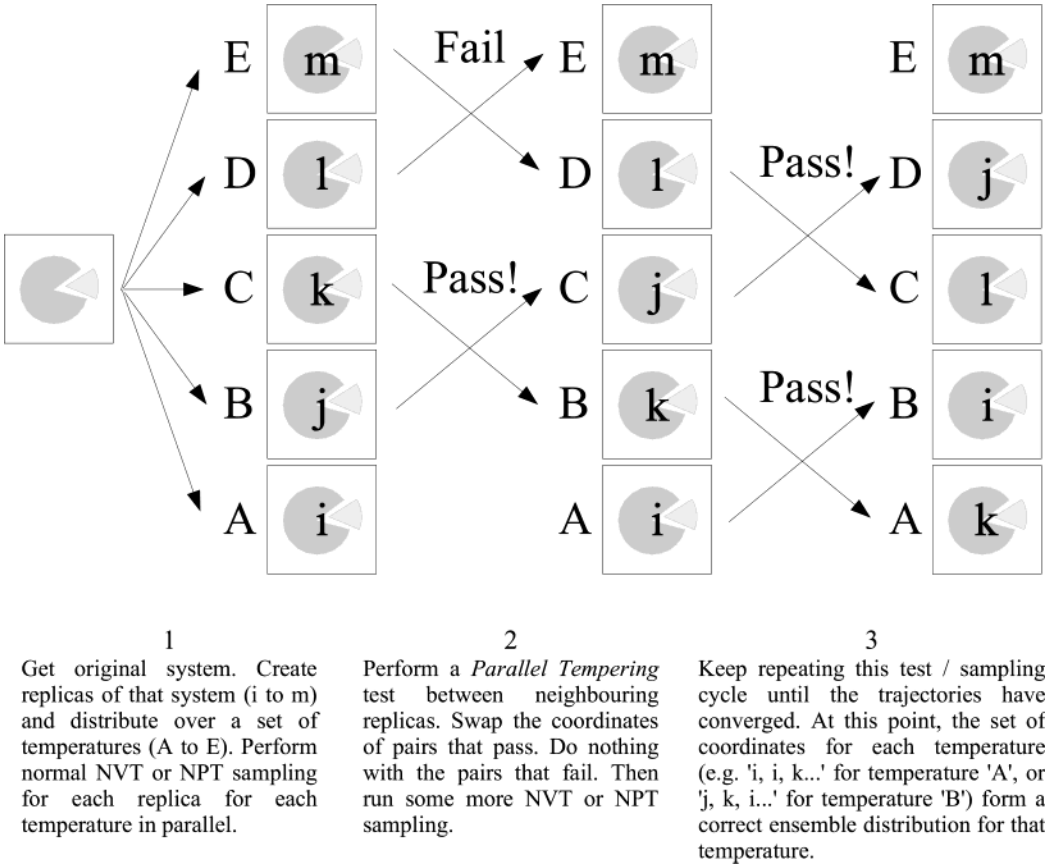

Figure 1. Schematic of the parallel tempering (PT) algorithm.

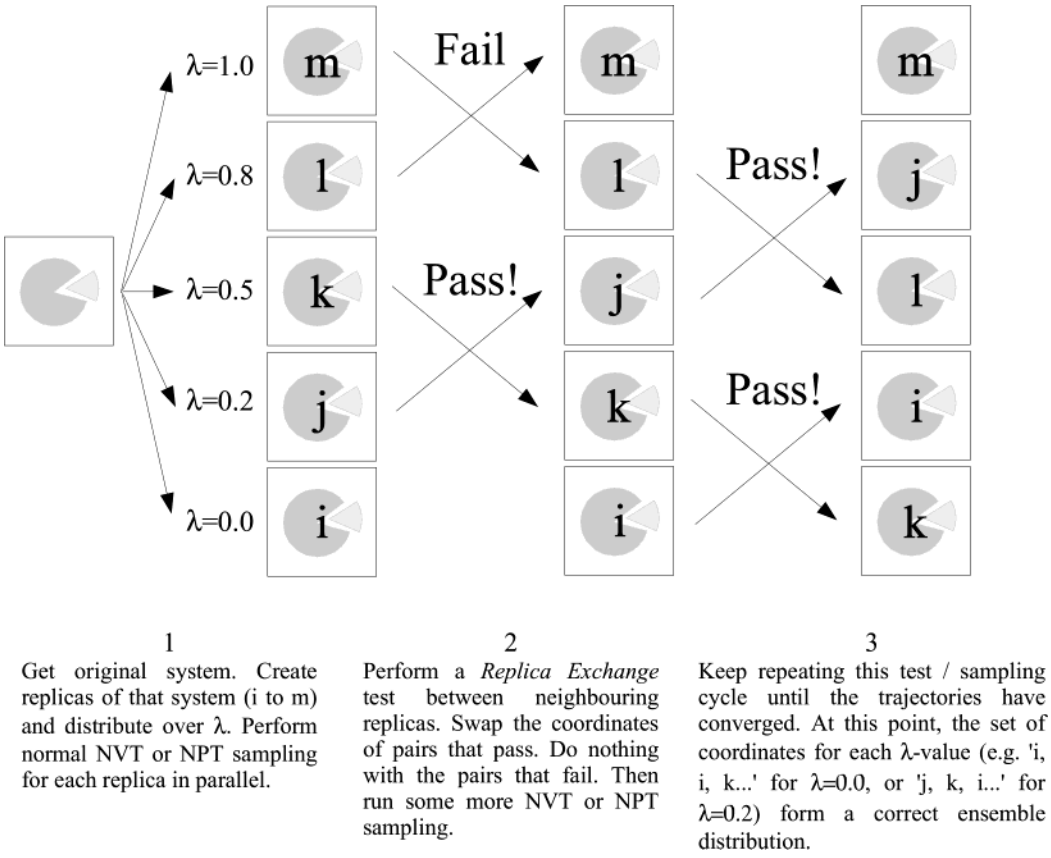

Figure 2. Schematic of the replica-exchange thermodynamic integration (RETI) algorithm.

two normal Metropolis tests: one going from configuration i to j at λ = A, and the other going from j to i at λ = B. The replica-exchange test is the product of these tests, because both moves must occur simultaneously.

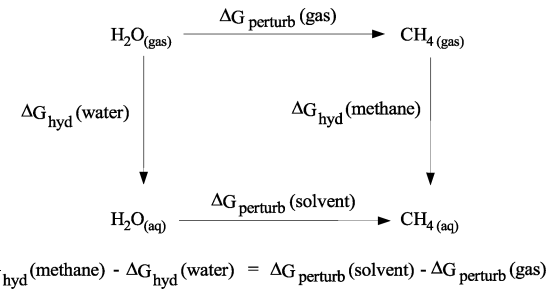

Figure 3. Free-energy cycle used to calculate the relative hydration free energy of water and methane.

This sampling-test iteration is continued until convergence. At this point, the set of trajectories at each value of λ form a correct ensemble distribution for that value of λ, and, thus, the free-energy average is formed correctly. The advantages of this method are that it is almost trivial to implement, with the addition of an inexpensive λ-swap move and test, and the additional bookkeeping needed to keep track of which trajectories correspond to what λ values. In addition, this scheme allows each trajectory to move freely across λ, as in the case of AdUmWHAM, but without the problems of Hamiltonian lag, because an umbrella potential is not involved. Also, the probability of accepting a λ-swap move will be based on the amount of overlap between neighboring λ values. The FDTI integration requires good overlap to converge well; therefore, it is expected that the acceptance ratio of the λ-swap moves should be high. This combination of replica-exchange with FDTI shall be known as replica-exchange thermodynamic integration (RETI).

2.2.3. Replica-Exchange Free-Energy Perturbation. RETI combines the replica-exchange method with FDTI through the facile addition of periodic λ-swap moves. A similar technique may be used to combine replica-exchange with FEP, resulting in replica-exchange FEP (REFEP). As in the case of RETI, it is expected that the acceptance ratio of λ-swap moves should be high, because FEP already requires good overlap between neighboring λ values.

3. Experimental Section

Three new free-energy methods have been proposed to improve sampling during free-energy calculations that involve large solvent reorganizations. To test these methods, the calculation of the relative hydration free energy of water and methane was attempted. This calculation is deceptively complicated, because water-in-water has a completely different solvent structure than methane-in-water. Thus, the calculation involves large solvent reorganization, because a dipolar, hydrophilic water molecule is perturbed into a neutral, hydrophobic methane molecule. Because a large solvent box was used during the simulation, there were more than 5000 degrees of freedom that required sampling. The TIP4P water model[36] was used, and it was perturbed into an OPLS united atom methane.[37] The cycle used to calculate the relative hydration free energy is shown in Figure 3.

A rigid water model and a united atom methane model were used; therefore, the free-energy change of the gas-phase perturbation is equal to zero. Thus, the entire relative hydration free energy can be calculated from a single solvent-phase perturbation and compared to the experimental value of 8.31 kcal mol^{-1} (the hydration free energy of water is -6.31 kcal mol^{-1}, whereas the hydration free energy of methane is 2.00 kcal mol^{-1}).[38]

3.1. Simulation Conditions. The perturbation was designed to morph the O atom of the water into the united-atom CH_4 particle. The remaining H and TIP4P "M" atom were morphed into dummy atoms. To improve convergence, and prevent abrupt changes as the H atoms disappeared, the H atoms were gradually pulled into the O atom as the perturbation progressed.[39] The O$-$H bond lengths started at 0.96 Å and were linearly scaled down to 0.2 Å by $\lambda = 1.0$. A single water$-$methane molecule was placed in an orthorhombic box of 1679 TIP4P water molecules, with initial dimensions of 37.3 Å $\times$ 37.9 Å $\times$ 37.4 Å. This system was then equilibrated at $\lambda = 0.0$ for 2 million (2 M) MC steps, under NPT periodic boundary conditions, at a temperature of 25 °C and a pressure of 1 atm. Nonbonded cutoffs were set to 15 Å, and one solute move was made for every 1600 solvent moves, whereas changes in volume were attempted every 10 375 moves. Solute and solvent moves consisted of rigid-body translations and rotations, with a maximum translation of 0.1 Å, and a maximum rotation of 2.5° for the solvent and 5.0° for the solute. The volume moves changed the volume of the solvent box by a maximum of 830 Å^3. Preferential sampling of the solvent,[40,41] as implemented in the MCPRO program,[42] was used, with a preferential sampling constant of 200.0 Å.

The purpose of this equilibration was to remove any bad contacts in the constructed solvent box, and its quality was ensured by monitoring the total energy of the system. The equilibration, and all subsequent simulations, were performed using a modified version of MCPRO 1.5.[42] The modifications were made to allow the code to run all of the free-energy methods under investigation, and they were checked to ensure that they did not interfere with the normal running of the code.

The final configuration from the equilibration was used as the starting structure for each value of λ, for each of the free-energy simulations. Four copies of each free-energy simulation were performed, each starting from the same starting point, but each using a different random number seed. For the fixed λ methods (FEP, FDTI, PTTI, RETI, and REFEP), λ was divided into 21 evenly spaced windows. Ten million (10 M) steps of MC sampling were run at each window; these were divided into 3 M steps of equilibration and 7 M steps of data collection. This division was based on the stabilization of the relative free energy, which was calculated every 500 000 (500 K) steps. Double-wide sampling[27] was used to obtain both the forward and backward estimates of the free energy. FDTI, PTTI, and RETI all used a value of $\Delta\lambda = 0.001$. This was sufficiently small to ensure the quality of the numerical gradients, while being large enough to prevent problems due to numerical imprecision. The resulting gradients were integrated via the trapezium rule. λ-swap moves were attempted every 50 K MC steps during the RETI and REFEP simulations, whereas temperature swap moves were attempted every 50 K steps during the PTTI simulations. The PTTI simulations used 16 temperature replicas at each of the 21 values of λ. To investigate the effect of temperature spacing, two of the four PTTI simulations used temperatures that were evenly spaced every 5 °C, from 10 °C to 85 °C. The remaining two simulations used replicas that were spaced evenly every 2.5 °C, from 20 °C to 55 °C, with a final replica at 60 °C. To enable each replica to equilibrate to each temperature, another 0.9 M steps of equilibration were performed before the start of each trajectory. No equilibration was necessary either before or after each temperature swap move. An identical scheme was used for each replica of the REFEP and RETI simulations, to allow them to equilibrate to each λ value correctly before any λ-swap moves

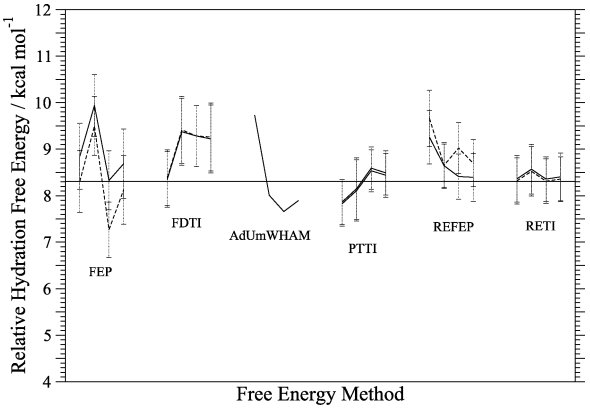

Figure 4. Relative hydration free energy of water and methane, as predicted by four repeated applications of the FEP, FDTI, AdUmWHAM, PTTI, RETI, and REFEP methods. The experimental value of 8.31 kcal mol^{-1} is shown as a horizontal line. The forward (solid line) and backward (dashed line) free energies are shown, where applicable, in addition to the standard errors.

were attempted. No equilibration was necessary, either before or after each λ-swap move.

The four AdUmWHAM simulations were designed to be as similar to the fixed-λ simulations as possible. The AdUmWHAM protocol used has been described elsewhere,[33] although it was found that, to improve the quality of the convergence of the predicted free energy, the value of λ had to be preserved between each WHAM iteration, rather than be assigned to a new random value, as was previously described.[33] A λ-move with a maximum change of 0.05 was made every 500 MC moves. These values were chosen to minimize the effect of Hamiltonian lag, while allowing the simulation to converge within a similar time to the fixed-λ methods. Each simulation used two trajectories running in parallel. Both trajectories were generated simultaneously, and when each one finished, it submitted its λ-sampling statistics to a custom program that performed the WHAM analysis.[33] The resulting umbrella was processed as described previously.[33] Each trajectory ran in blocks of 200 K steps, with the first 40 K steps being used to equilibrate the trajectory to the new umbrella, and the remaining 160 K steps being used for data collection. The simulation ran for 500 completed WHAM iterations, corresponding to 100 M MC steps. Although this is a fraction of the 210 M steps that are used in the fixed-λ methods, or the 3.3 billion (3.3 B) steps used for PTTI, the fixed-λ methods were able to run each λ value and replica in parallel, with each trajectory consisting of only 10 M steps. In contrast, each AdUmWHAM trajectory ran for 50 M steps; thus, the AdUmWHAM simulations took five times longer to run than the fixed-λ methods.

4. Results

The free-energy results are shown in Figure 4. The error bars for all methods other than AdUmWHAM were determined by calculating batch averages of the relative free energy at each λ value every 500 K steps. The standard error of these averages was then integrated (FDTI, PTTI, RETI) or summed (FEP, REFEP) across the entire λ-coordinate to yield the maximum error. This error analysis will overestimate the error but should be sufficiently sensitive to allow good comparison between the free-energy methods. The AdUmWHAM simulations failed to achieve an even sampling of λ, so an error analysis was not performed.

5. Analysis

5.1. Simulations. The FEP results show large hysteresis ($\sim$0.4$-$1.0 kcal mol^{-1}), which suggests that there was insufficient overlap between adjacent windows. Although these results may be improved through the use of additional windows, a disadvantage of FEP is that it is is difficult to know the optimum number and position of windows that are required at the start of the simulation.

The FEP results show a wide variation in the relative hydration free energies ($\sim$2 kcal mol^{-1}). This variation is despite each simulation starting from the same configuration and running under identical conditions. The only difference between the four simulations was that each used a different random number seed and, thus, generated a different subset of configurations. Examination of the individual free-energy averages at each λ value shows that each subset of configurations resulted in a slightly different free energy. When these are summed across λ, they produce four different relative hydration free energies. This variation is a result of the slightly different sampling of the four "identical" simulations and shall be called the random sampling error.

The FDTI results show very little hysteresis, because of the small difference in λ between the reference and perturbed states. Although the hysteresis is low, the standard errors are almost identical to those calculated from FEP, which suggests that the hysteresis is a poor measure of error in the FDTI method. In addition, the FDTI results also show a large random sampling error ($\sim$1.0 kcal mol^{-1}).

The AdUmWHAM simulations failed to converge fully, despite running for five times longer than the amount of time of the other free-energy methods. Although the simulations took five times longer than those of the fixed-λ methods, the AdUmWHAM simulations were not as parallizable and, therefore, only generated half the number of configurations. Although the shape and magnitude of the umbrella seemed to converge after $\sim$200 of the 500 WHAM iterations, the underlying λ-sampling did not become totally even. Each simulation converged onto a slightly different umbrella, and, thus, the random sampling error for the four simulations was significant, at 2.0 kcal mol^{-1}.

Because the PTTI simulations were based on FDTI, they also showed very low hysteresis. The standard errors are smaller than those of FEP or FDTI (0.5 kcal mol^{-1}, compared to 0.7 kcal mol^{-1}). The improvement in these results is due to the averaging of multiple trajectories at each λ value, because multiple replicas are able to swap into the 25 °C level. This averaging results in a lower random sampling error, because the average free energy at each λ value is comprised of multiple subsets of configurational sampling. The temperature-swap acceptance ratios of the first two PTTI simulations were significantly less than those for the last two simulations (40%, compared to 65%), because of the smaller difference in temperature between replicas in the last two simulations. Despite this observation, the results from all four simulations are very similar. The PTTI simulations each required 16 times more CPU resources than the other fixed-λ methods. Some of the justification for this increase was that the method should have been able to return the relative free energy, with respect to temperature, and, hence, yield relative entropies from the gradient. Unfortunately, this was not realized, because each of the last two PTTI simulations showed almost opposite trends in free energy, with respect to temperature (Figure 5).

The REFEP results show a lower random sampling error and standard error than those from FEP. The use of a λ-swap move

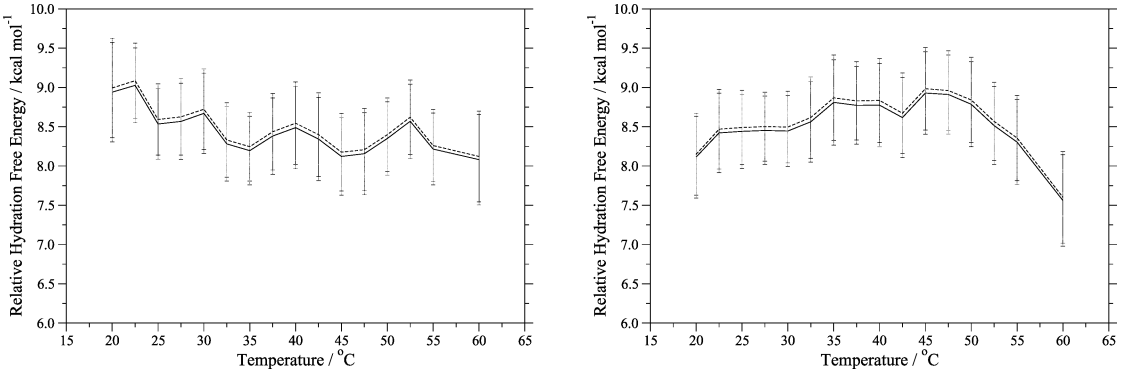

Figure 5. Relative hydration free energy, with respect to temperature, as calculated by the last two PTTI simulations. Both the forward (solid) and backward (dashed) relative free energies are shown, in addition to the standard errors.

has allowed free motion of the trajectories across the λ-coordinate, with an acceptance ratio of >80%. This replica-exchange methodology has allowed each value of λ to experience replicas that have visited each of the other values of λ. Although this averaging of multiple trajectories at each value of λ has reduced random sampling error, it has not improved the overlap between the reference and perturbed states, and, thus, there is still a significant hysteresis error. In contrast, the RETI results show very little hysteresis. This, combined with the low random sampling error, leads to a set of results that are of significantly higher quality than those from the other methods. Indeed, each of the four RETI simulations produced relative hydration free energies that differed by, at most, 0.26 kcal mol^{-1}. In addition, the four RETI results show excellent agreement with experiment. This is achieved at little or no extra computational cost, relative to FEP or FDTI.

5.2. Radial Distribution Functions. We can plot the radial distribution function (RDF) around the central solute atom as a function of λ. At $\lambda = 0.0$, this returns g_{O-O} and g_{O-H}, whereas, at $\lambda = 1.0$ this returns g_{C-O} and g_{C-H}. The RDFs for $\lambda = 0.0-1.0$ should show a general progression from the water-in-water solvent structure to the methane-in-water solvent structure.

The RDFs for the fixed-λ methods were generated from simulation snapshots that were taken every 50 K MC steps during data collection and were histogrammed with a bin width of 0.1 Å. The RDFs for the AdUmWHAM simulations were generated by taking snapshots every 50 K steps, and each RDF was then histogrammed according to the λ value of the corresponding snapshot. Using a λ-histogram bin size of 0.05 produced plots that were comparable to those from the fixed-λ methods. However, care must be taken with any comparison, because these plots are constructed from fewer snapshots than those from the fixed-λ methods.

FEP and FDTI have the same reference states; therefore, all the FEP and FDTI RDFs are very similar. Thus, only those from the first FEP simulation are shown. REFEP and RETI also share the same reference states, so only the RDFs from the first RETI simulation are shown. The third PTTI simulation was the first that used the smaller temperature spacings, so its RDF is shown. The RDFs from the first AdUmWHAM simulation are also shown (see Figure 6).

The RDFs produced show key features that agree with other simulation studies[43] and experiment.[44,45] They clearly reveal the improvement in sampling that is due to the addition of the λ-swap move in the REFEP and RETI simulations. The main features in the RETI and REFEP RDFs morph smoothly between the water-in-water and methane-in-water solvent structures. In contrast, the transition between solvent structures for the FEP and FDTI RDFs is less smooth, with the exact location of the first and second solvation shell peaks varying by as much as 0.5 Å between neighboring λ values.

The reason for the improvement in the RDFs is that the addition of the λ-swap move has connected each λ value and allowed information to be shared across the entire λ-coordinate. In the case of FEP, FDTI, or PTTI, the subset of configurations sampled at each λ value was independent of that sampled at each of the other λ values. In contrast, the configurational sampling at each λ value from the REFEP or RETI simulations were connected together and formed a single, consistent subset across λ. This was also the case for the AdUmWHAM simulations, which used a dynamic λ value to ensure consistent sampling across the entire λ-coordinate. Although RETI and REFEP have demonstrated the same improvements in sampling that have been achieved by the AdUmWHAM method,[33] they do so without the need to use or generate an umbrella potential. RETI or REFEP do not thus require a tedious iteration, and they do not suffer from problems that are associated with Hamiltonian lag. These problems are apparant in the hydrogen RDFs plotted in Figure 6. Those produced from the AdUmWHAM simulations clearly show the effects of Hamiltonian lag, as the double peak in the hydrogen RDF exists all the way up to $\lambda = 0.4$, compared to only $\lambda = 0.25$ for the fixed-λ methods. This lag has occurred because AdUmWHAM uses an umbrella potential to drive sampling across the λ-coordinate. In contrast, the λ-swap move of RETI and REFEP allows free motion along the λ-coordinate and does not drive the sampling through potentially nonequilibrium structures.

6. Conclusion

Three new free-energy methods have been developed that improve configurational sampling through the combination of replica-exchange moves with established free-energy methods. These new methods were applied to the challenging water−methane system and compared to three established free-energy methods. The results of this application show the following:

(1) Finite-difference thermodynamic integration (FDTI) has little hysteresis, although its standard errors and random sampling error are comparable to those of free-energy perturbation (FEP).

(2) The combination of parallel tempering with FDTI (parallel tempering thermodynamic intergration, PTTI) reduces the errors in the results, although at significant extra cost, compared to standard FDTI. PTTI failed to produce reliable free-energy trends, with respect to temperature.

(3) The development of a replica-exchange move over λ has significantly reduced the errors in the results, compared to standard FEP or FDTI.

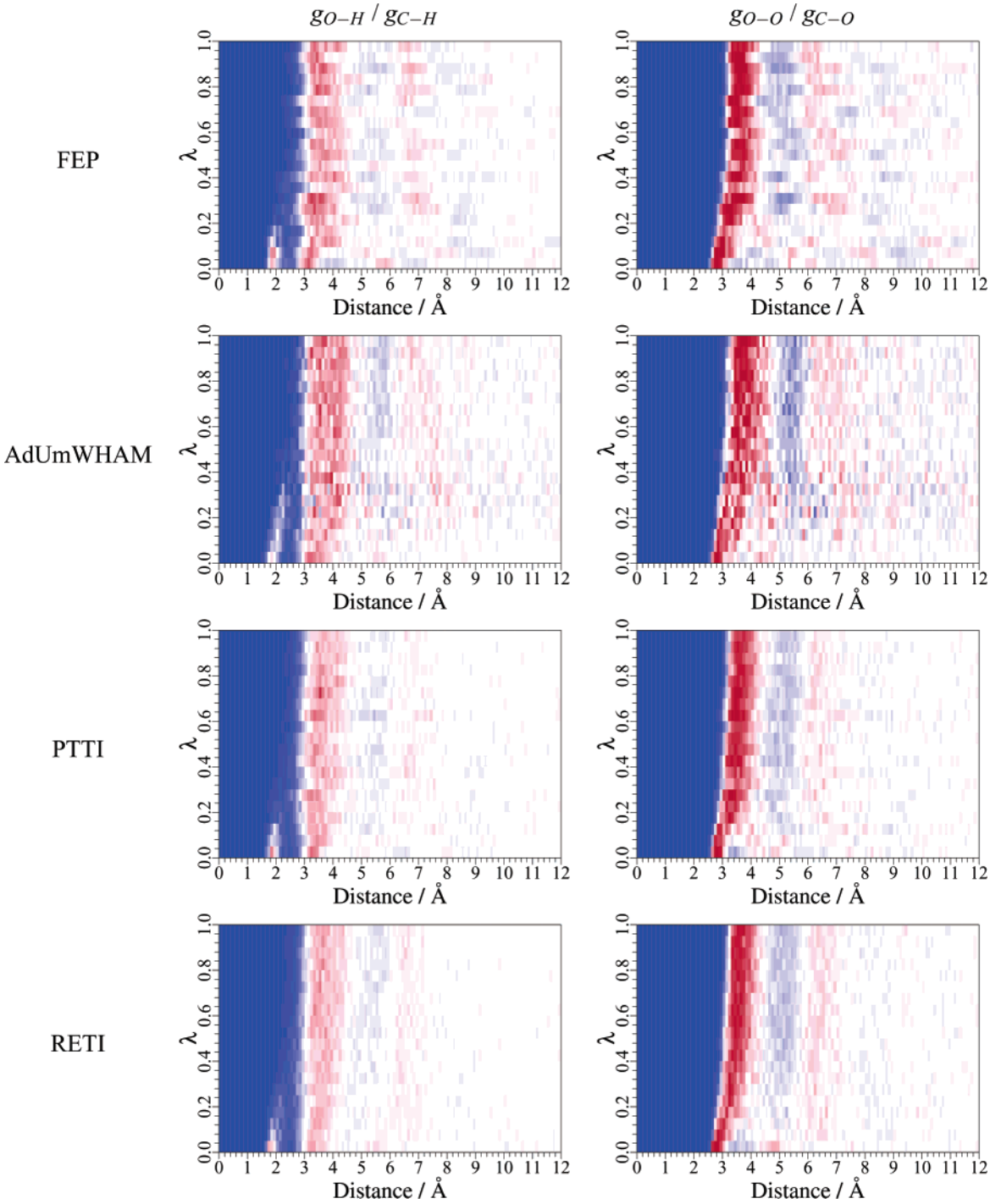

Figure 6. Radial distribution function (RDF), with respect to λ, between the central solute and the surrounding water solvent; the left column shows the RDF calculated between the central solute and solvent H atoms, whereas the right column shows the RDF between the central solute and the solvent O atoms. From top to bottom, the RDFs were generated via FEP, AdUmWHAM, PTTI, and RETI methods. The color scale goes from blue (0.0), through white (1.0), to red ($\geq$2.0).

(4) The combination of the λ-swap move with FDTI (replica-exchange thermodynamic intergration, RETI) produces results with very little hysteresis and a very small random sampling error. The results produced are in excellent agreement with experiment.

(5) The sampling of the RETI and replica-exchange free-energy perturbation (REFEP) simulations is much improved over FEP or FDTI. The λ-swap move allows each trajectory to move freely throughout the λ-coordinate, which brings the same advantages as those observed in the adaptive umbrella weighted histogram analysis method (AdUmWHAM) method. RETI and REFEP achieve this improvement without the need for an umbrella potential; thus, they do not require a tedious iteration, nor do they suffer from problems of Hamiltonian lag.

The newly developed method of RETI has been shown to perform significantly better than any of the other free-energy methods when applied to the water−methane system. The method allows consistent, high-quality sampling of a complex solvent configurational change. This achievement occurs through the facile addition of a λ-swap move to standard FDTI, thus effectively adding no extra expense to the standard calculation.

Acknowledgment. We thank the BBSRC, EPSRC, Celltech, and the University of Southampton for funding this work, and

Prof. W. L. Jorgensen for the provision of the MCPRO program and associated source code.

References and Notes

(1) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(2) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058−9067.
(3) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(5) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002**, *46*, 225−234.
(6) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12777−12782.
(7) Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. *Chem. Phys. Lett.* **2001**, *337*, 181−189.
(8) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(9) Price, M. L. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 9455−9466.
(10) Price, D. J.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2067−2070.
(11) Price, D. J.; Jorgensen, W. L. *J. Comput. Aid. Mol. Des.* **2001**, *15*, 681−695.
(12) Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(13) Oostenbrink, B. C.; Pitera, J. W.; Lipzig, M. M. H.; Meerman, J. H. N.; Van Gunsteren, W. F. *J. Med. Chem.* **2000**, *43*, 4594−4605.
(14) Barril, X.; Orozco, M.; Luque, F. J. *J. Med. Chem.* **1999**, *42*, 5110−5119.
(15) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(16) Hooft, R. W. W.; Vaneijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*, 6690−6694.
(17) Engkvist, O.; Karlstrom, G. *Chem. Phys.* **1996**, *213*, 63−76.
(18) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450−1462.
(19) Kumar, S.; Payne, P. W.; Vasquez, M. *J. Comput. Chem.* **1996**, *17*, 1269−1275.
(20) Reynolds, C. A.; King, P. M.; Richards, W. G. *Mol. Phys.* **1992**, *76*, 251−275.
(21) Chipot, C.; Pearlman, D. A. *Mol. Simulat.* **2002**, *28*, 1−12.
(22) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018−5035.
(23) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(24) Jarzynski, C. *Proc. Natl. Acad. Sci.* **2001**, *98*, 3636−3638.
(25) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330−7337.
(26) Wood, R. H.; Muhlbauer, W. C. F.; Thompson, P. T. *J. Phys. Chem.* **1991**, *95*, 6670−6675.
(27) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(28) Mezei, M. *J. Chem. Phys.* **1987**, *86*, 7084−7088.
(29) Guimaraes, C. R. W.; Alencastro, R. B. *Int. J. Quantum Chem.* **2001**, *85*, 713−726.
(30) Guimaraes, C. R. W.; Alencastro, R. B. *J. Med. Chem.* **2002**, *45*, 4995−5004.
(31) Guimaraes, C. R. W.; Alencastro, R. B. *J. Phys. Chem. B* **2002**, *106*, 466−476.
(32) Kamath, S.; Coutinho, E.; Desai, P. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1239−1244.
(33) Woods, C. J.; Camiolo, S.; Light, M. E.; Coles, S. J.; Hursthouse, M. B.; King, M. A.; Gale, P. A.; Essex, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 8644−8652.
(34) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(35) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435−439.
(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(37) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638−6646.
(38) Zhu, T. H.; Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117−9133.
(39) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487−1493.
(40) Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600−602.
(41) Jorgensen, W. L. *J. Phys. Chem.* **1983**, *87*, 5304−5314.
(42) Jorgensen, W. L. MCPRO, Version 1.5; Yale University: New Haven, CT, 1996.
(43) Hernandez-Cobos, J.; Mackie, A. D.; Vega, L. F. *J. Chem. Phys.* **2001**, *114*, 7527−7535.
(44) Dejong, P. H. K.; Wilson, J. E.; Neilson, G. W.; Buckingham, A. D. *Mol. Phys.* **1997**, *91*, 99−103.
(45) Head-Gordon, T.; Hura, G. *Chem. Rev.* **2002**, *102*, 2651−2669.