total free energy can simply be written as a sum of the individual free energies between intermediate states, which can be completely nonphysical. We will first assume the existence of these intermediates, and the ability to perform simulations at these intermediates and then discuss the best choice of intermediate states.

If phase-space overlap between consecutive intermediates is very high, then EXP can work well. For example, if an ether is changed to a thioether, there is relatively little change in phase-space, and EXP will be effective with a small number of intermediates. However, if an entire heavy atom is disappearing or appearing or if the charge of an atom is changing significantly, phase-space overlap will not be significant, and EXP is almost guaranteed to do poorly without a large number of intermediates. Because of the large number of intermediates required, computations with EXP can be very inefficient, requiring the simulation of many states for computing the free energy of a single alchemical transformation.

Double-wide sampling is a commonly used technique that consists of simulating only at every other intermediate state and computing EXP in both directions from these intermediates.[78] The biases of free energy computed from EXP in different directions have opposite signs, so alternating directions will tend to cancel bias somewhat. This method nominally reduces the number of simulations necessary by half, but because the variance in the direction of increasing entropy is usually lower, this twofold gain in efficiency is rarely obtained. Fortunately, there are a number of alternatives that are more efficient than EXP in most cases.

Thermodynamic integration

By taking the derivative of the free energy with respect to some continuous parameter λ describing a series of intermediate alchemical states, we can see that

$$\begin{aligned} dG/d\lambda &= \frac{d}{d\lambda}\int e^{-\beta H(\lambda,\boldsymbol{x})}\,d\boldsymbol{x} = \left\langle \frac{dH(\lambda,\boldsymbol{x})}{d\lambda}\right\rangle_{\lambda} \\ \Delta G &= \int_0^1 \left\langle \frac{dH(\lambda,\boldsymbol{x})}{d\lambda}\right\rangle_{\lambda}, \end{aligned} \tag{5.11}$$

where the pathway of intermediates between the states of interest is parameterized between $\lambda = 0$ and $\lambda = 1$. This formula can also be obtained by expanding the Zwanzig relationship in a Taylor series. Computing free energies using this formula is typically called thermodynamic integration (TI). In the rest of the discussion, we will denote $H(\lambda, \boldsymbol{x})$ by simply $H(\lambda)$. Note that when the end states have different masses, the momenta will have λ dependence as well, which must also be included in the derivative, but we omit this detail for clarity in the discussion.

Thermodynamic integration essentially trades variance for bias. Averaging over $\langle \frac{dH}{d\lambda}\rangle$ will require fewer uncorrelated samples to reach a given level of relative error than averaging $e^{-\beta\Delta H(x)}$, as as long as $\langle \frac{dH}{d\lambda}\rangle$ is well behaved, an important condition we will address later in the section "Choice of Alchemical Pathways." However, to compute the total free energy from a series of individual simulations, we must use some type of numerical integration of the integral, which by definition introduces bias. A number of different numerical techniques have been applied.[79,80] A simple trapezoidal rule is usually used or, occasionally, Simpson's rule. Higher order integration methods converge more quickly in the distance between integration points, but this error term is proportional to the derivatives of the function, which can become large in some situations, such as when repulsive atomic centers are removed from the system entirely. Other techniques such as Gaussian integration have been used[79] but require knowledge about the variance to determine the Gaussian weighting and so become cumbersome to use.

For alchemical changes that result in smooth, monotonic curves for $\langle dH/d\lambda\rangle$, TI can be quite accurate using a relatively small number of points. However, if the curvature becomes large, as can frequently be the case in alchemical simulations where Lennard–Jones potentials are turned on or off, then the bias introduced by discretization of the integral can become large.[73,81,82] Even in the case of small curvature (i.e., charging of SPC water in water) reasonably large errors can be introduced (i.e., 5–10% of the total free energy with 5 λ values).[83]

Many early free-energy calculations approximated the integral by varying λ throughout the simulation, called "slow growth." The total free energy is then estimated as

$$\Delta G \approx \int_{t=t_0}^{t_1} \left\langle \frac{dH}{d\lambda}\right\rangle_{\lambda(t)} \frac{d\lambda}{dt}\,dt. \tag{5.12}$$

This, however, proved to be a very bad approximation in most molecular simulations, introducing large speed-dependent biases even for relatively long simulations. Forward and reverse simulations show significant hysteresis.[84,85] This method should always be avoided, except when used in the context of Jarzynski's relationship, which we will now discuss.

Jarzynski's relationship

If we have a physical or alchemical process that takes place in finite time, the amount of work performed will not be reversible and hence will not be equal to the free energy. Equation (5.12) can then be identified as the nonequilibrium work W for this transformation, not the equilibrium free energy ΔG. Jarzynski noticed that the free energy of the transformation can be written as the average of the nonequilibrium trajectories that start from an equilibrium ensemble:

$$\Delta G = \beta^{-1}\ln\left\langle e^{-\beta W}\right\rangle_0 . \tag{5.13}$$

If the switching is instantaneous, then Equation (5.13) becomes identical to EXP because the instantaneous work is simply the change in potential energy. A number of studies have compared nonequilibrium pathways to single-step perturbations. However, in both theory and practice it