

Hamiltonian, there must be a series of intermediates with mutual phase-space overlap leading connecting the two physical end states. The simplest choice for most transformations between two Hamiltonians  $H_0$  and  $H_1$  is the linear pathway:

$$H(\lambda, \mathbf{x}) = (1 - \lambda)H_0(\mathbf{x}) + \lambda H_1(\mathbf{x}). \quad (5.16)$$

A significant problem with this formulation is that equal spacing in  $\lambda$  does not actually lead to equal spacing in phase-space overlap. If a Lennard–Jones function is used to represent the atomic exclusion and dispersion interaction, as is typically the case in biomolecular force fields, then when  $\lambda = 0.1$ , nearly at the disappearing end state, the excluded volume (i.e., the volume with energy above 2–3  $kT$ ) will still occupy 60–70% of the original volume, depending on the original Lennard–Jones well depth.

Additionally, this choice of parameterization with an  $r^{-12}$  potential leads to a singularity in  $\langle dH/d\lambda \rangle$  at  $r = 0$ , which can be integrated formally but not numerically. By using a power of  $\lambda \geq 4$  instead of a strictly linear parameterization [such as  $H(\lambda) = (1 - \lambda)^4 H_0 + \lambda^4 H_1$ ],  $\langle dH/d\lambda \rangle$  can be numerically integrated correctly. However, it will still converge slowly.<sup>80,99</sup> For any nonzero  $\lambda$ , whatever the power, there will be small “fenceposts,” particles with a small impenetrable core.<sup>99</sup> One possible way to avoid issues with these fenceposts has been to shrink the entire molecular structure. However, this can create problems with nonbonded interactions as the molecular framework shrinks, causing instabilities in integration in molecular dynamics,<sup>99–101</sup> and is generally not practical for large numbers of bonds. A correction term must also be added for these bond length changes, which can be complicated if the bonds lengths are constrained.<sup>102</sup>

There are better ways to handle this transformation. The concept of a “soft core” was introduced around 1994,<sup>82,103</sup> with the infinity at  $r = 0$  in the  $r^{-12}$  interaction being “smoothed out” in a  $\lambda$ -dependent way. The most common-parameterizations for turning off the Lennard–Jones function are of the form

$$H(\lambda, r) = 4\epsilon\lambda^n \left\{ \left[ \alpha(1-\lambda)^m + \left(\frac{r}{\sigma}\right)^6 \right]^{-2} + \left[ \alpha(1-\lambda)^m + \left(\frac{r}{\sigma}\right)^6 \right]^{-1} \right\} \quad (5.17)$$

where  $\epsilon$  and  $\sigma$  are the standard Lennard–Jones parameters,  $\alpha$  is a constant (usually 0.5), with the original choice of  $n = 4$  and  $m = 2$ .<sup>82</sup> Further research has shown that using  $n = 1$  and  $m = 1$  noticeably improves the variance.<sup>80,99,104</sup> The more flexible the molecule, the more using a soft atomic core improves the efficiency of the free-energy calculation. Approaches using a soft core for the Coulombic term<sup>82,105</sup> or making all the interactions disappear into an imaginary fourth dimension have also been tried,<sup>106</sup> but it can be difficult to choose parameters for these approaches that are transferable between systems.

Recent studies have demonstrated that one highly reliable, relatively high efficiency pathway for an alchemical

change where an atomic site disappears is to turn off the charges linearly and then turn off the Lennard–Jones terms of the uncharged particles using a soft-core approach. The same pathway can be followed in reverse for atomic sites that are introduced.<sup>97,99</sup> It is the treatment of the singularities at the center of particles that is the real challenge; atomic sites that merely change atom type can be handled with linear interpolation of the potential energy, as the phase-space overlap changes are relatively small with respect to phase-space changes with introduction of particles. The variance due to changes in the bonding terms is not generally a problem; although the energy changes for these terms can be quite large, the time scale of the motions means that they converge quite quickly.

It is likely that further optimizations of pathways may lead to additional efficiency gains. But they will probably not increase efficiency by much more than a factor of 2 as there are limits to the lowest possible variance path.<sup>105</sup> Studies of optimal pathways have focused on minimizing the variance in TI,<sup>82,105</sup> but it appears that highly optimal pathways for TI work well for all other methods.

### Pulling methods

Another choice of pathway for determining the free energy of protein/ligand association is to physically pull the molecule away from the protein. If the final state is sufficiently far from the original protein, the free energy of this process will be the free energy of binding. This can be done either by nonequilibrium simulations, using the Jarzynski relationship, or by computing a PMF using umbrella sampling with different overlapping harmonic oscillators at specified distances from the binding site.<sup>107–110</sup>

There are a number of complications with pulling methods. Pulling a ligand out of a buried site can pose problems, and it can be difficult to pull the ligand sufficiently far away from the protein with a simulation box of computationally tractable size. In the latter case, some analytical or mean-field approximation must be applied for the free energy of pulling the ligand to infinity. However, it has been argued that pulling may be significantly more efficient for highly charged ligands.<sup>108</sup>

### Promising methods not yet routine

Researchers are experimenting with a number of intriguing methods that have significant potential to make ligand-binding calculations much more efficient but that are not yet routine. It is likely that many or all of these will become much more commonplace in the near future. We will give only brief a introduction along with references for further investigation.

### Using umbrella sampling for convergence

A general problem for any free-energy simulation method is sampling important configurations. One standard method