

# Thermodynamic Cycle without Turning Off Self-Interactions: Formal Discussion and a Numerical Example

Mauro L. Mugnai and Ron Elber\*

Department of Chemistry and Biochemistry, Institute for Computational Engineering and Sciences, University of Texas at Austin, 201 East 24th St. STOP C0200, Austin, Texas 78712-1229, United States

## Supporting Information

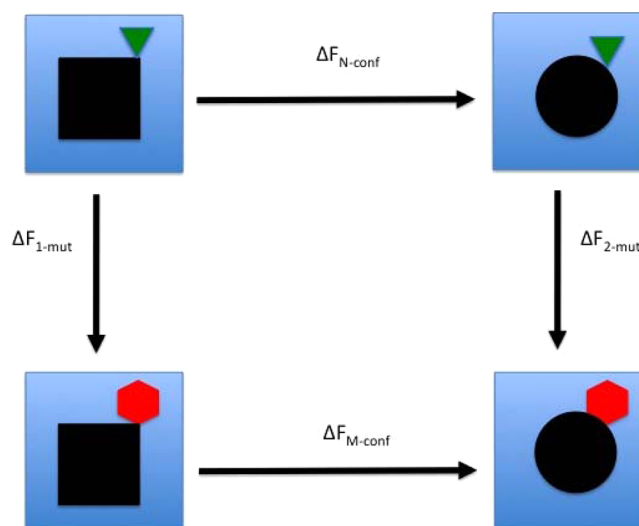
**ABSTRACT:** The efficiency and accuracy of thermodynamic cycle calculations are considered. It is rigorously shown that the energy of the mutated part (MP) need not be scaled in a thermodynamic cycle computed with dual topology. Hence, there is no need to scale to zero any of the self-interactions (i.e., the interactions involving only particles of the same MP) regardless of whether the MP is bound or not to the main system. This observation carries a promise to lower computational resources and increase accuracy. A numerical test of a complete thermodynamic cycle illustrates cost and accuracy considerations.

## INTRODUCTION

Free energy calculations make it possible to directly compare simulations to experiments, a necessary step in the validation of atomically detailed models. They also guide future experiments, by (for example) suggesting candidates for drug design.<sup>1</sup>

While the comparison between computed and measured free energy is direct, frequently the computational scheme does not “mimic” the experimental process. Indeed, experimental phenomena often occur at time scales inaccessible for direct molecular dynamics (MD) simulations. For example, it is costly to evaluate the dissociation constant of a ligand–enzyme complex by simulating binding and unbinding events. It is also challenging to compute the stability of the structure of a protein by letting the protein explore configurations that are accessible on long time scales. Specific reaction coordinates<sup>2</sup> or a set of sparsely sampled conformations (anchors)<sup>3,4</sup> can be used to compute the thermodynamics of large scale conformational transitions; however, they are typically very expensive and not always possible to converge.

A more efficient method to compute such free energy differences makes use of alchemical changes.<sup>5</sup> In alchemical changes, a molecule, or a small part of it, is mutated into another by modifying its external and self-interactions. We denote the mutated part MP. By external interactions we mean interactions between the MP and its environment, while self-interactions are within the MP. Such a mutation is usually performed according to two protocols:<sup>6</sup> a single topology protocol, in which the geometry of the native molecule/moiety is progressively changed into the geometry of the mutant, and a dual topology, in which the two molecules/moieties coexist. Along the substitution pathway, the interactions of the native part are gradually annihilated while those of the mutant are growing to their full strength. To connect the initial and final states, we often use multiple alchemical processes, presented on a thermodynamic cycle (TC;<sup>7</sup> see Figure 1). The experimental free energy difference is then obtained as a sum of the free energies of the computed alchemical steps. The sum can be used to compute “absolute”<sup>8,9</sup> or relative<sup>10–15</sup> free energy differences.



**Figure 1.** An example of a thermodynamic cycle for a conformational transition in a protein and a protein mutant. The black shapes represent the protein in two conformations (the square and the circle). The green triangle and the red hexagon represent native and mutant residues, respectively. The blue background represents solution. The horizontal arrows indicate conformational transitions. The vertical arrows indicate mutations. This cycle compares experimentally measured free energy differences (those that are related to the conformation transitions  $\Delta F_{N\text{-conf}}$  and  $\Delta F_{M\text{-conf}}$ ) to computed free energy differences (the mutations  $\Delta F_{1\text{-mut}}$  and  $\Delta F_{2\text{-mut}}$ ).

In the present paper, we examine the possibility of retaining the interaction of a MP throughout a mutation performed according to the dual topology protocol. Keeping the self-interactions is equivalent to a change in the end-state of the alchemical process from atomic to molecular ideal gas.<sup>16,17</sup> It was suggested that retaining the bonded self-interactions does not affect relative free energy differences,<sup>16,17</sup> and a proof of this

Received: May 14, 2012

Published: July 25, 2012

assertion followed.<sup>18</sup> In ref 19, a more general proof was reported, in which no distinction was made between bonded and nonbonded interactions. However, the proof given is less direct than the simple argument provided here.

We propose a dual topology alchemical pathway in which the self-interactions of the MP, bonded and nonbonded, are retained. We then prove that, in the context of a TC, the relative free energy difference computed according to this pathway is exact.

We also discuss the retention of some of the interactions between the decoupled fragment and the larger scaffold. The annihilation/creation of the external nonbonded interactions is required to compute the proper free energy difference. However, it is not necessary to annihilate/create all of the external bonded interactions connecting a fragment to a scaffold. It is possible to fix its six external degrees of freedom retaining six bonded interactions with the scaffold without affecting the internal degrees of freedom of the scaffold or of the fragment. All of the other external bonded interactions are annihilated/created in the MD simulation. This concept was introduced as the virtual bond algorithm (VBA),<sup>20</sup> which was proposed in the context of protein–ligand binding and is exact. Six bonded interactions between the protein and the ligand are added to restrain the overall motions of the ligand. The same VBA interactions may be retained in the process of substituting a fragment bonded to a scaffold. The idea of retaining external bonded interactions was examined also in refs 18 and 19. However, VBA makes it clearer which interactions can be safely retained in an alchemical substitution, and how to account for their contribution to the free energy.

The proof that we hereby propose and the use of VBA in the context of the alchemical substitution of a molecular fragment lead to an algorithm in which the number of interactions created/annihilated along a TC is reduced to the minimum. This is likely to benefit computations. Indeed, retaining the bonded self-interactions has improved convergence of the results since the overlap of distributions along the reaction coordinate increases, making the standard deviations smaller. It is likely that also retaining the nonbonded self-interactions will improve the convergence further.

This manuscript is organized as follows. We review the thermodynamic cycle and prove that the relative free energy difference computed along a thermodynamic path of our own design is in agreement with the formula for experimental free energy difference. We then review VBA and illustrate that in the context of a TC the restraining of the external degrees of freedom of the mutated part gives no contribution to the relative free energy difference. Finally, a numerical illustration of the theory is described in which the free energy of solvation of two amino acid side chain analogs is computed. In the numerical illustration, and in the Supporting Information, we discuss technical issues presented by the annihilation/creation of bonded terms in the force field and by the selective annihilation/creation of a subset of the nonbonded, long-range interactions.

## ■ FREE ENERGY DIFFERENCES

We consider the free energy difference between a physical system “*P,N*” which consists of “Protein” and “Native” parts, and a system “*P,M*” that consists of “Protein” and “Mutant” parts. “*P*” refers to all atoms that are not alchemically substituted. “*N*” and “*M*” are the two sets of atoms that are mutated to each other.

The two Hamiltonians are

$$H_N = K(P, N) + U(P) + U(P, N) + U(N)$$

$$H_M = K(P, M) + U(P) + U(P, M) + U(M) \quad (1)$$

Here, *K* is the kinetic energy and *U* the potential. The term *U*(*X*) includes interactions within “*X*” (self-interactions) and *U*(*X,Y*) only interactions between “*X*” and “*Y*” (external interactions). The phase space volume element is *dΓ* (e.g.,  $d\Gamma_{P,N} = d\Pi_P dQ_P d\Pi_N dQ_N$ , where  $\Pi_X$  denotes momentum and  $Q_X$  coordinate vectors of atoms of species “*X*”). The volume element of coordinate space is *dΓ'* (e.g.,  $d\Gamma'_{P,N} = dQ_P dQ_N$ ). Finally, we denote the number of “*X*” particles as *N<sub>X</sub>*, where “*X*” can be any combination of “*P*”, “*M*”, and “*N*”. The desired free energy difference,  $\Delta F_{M,N}$ , is given by

$$\begin{aligned} \Delta F_{M,N} &= F_{P,M} - F_{P,N} \\ &= -\beta^{-1} \ln \left[ \frac{\frac{1}{C_{PM} h^{3N_{PM}}}}{\frac{1}{C_{PN} h^{3N_{PN}}}} \right] \\ &\quad \times \frac{\int d\Gamma_{P,M} \exp\{-\beta[K(P, M) + U(P) + U(P, M) + U(M)]\}}{\int d\Gamma_{P,N} \exp\{-\beta[K(P, N) + U(P) + U(P, N) + U(N)]\}} \end{aligned} \quad (2)$$

The coefficients *C<sub>PM</sub>* and *C<sub>PN</sub>* are corrections for the permutation of identical particles<sup>21</sup>, and *h* is the Planck constant. Note that the number of particles of the two states can be different. The free energies are functions of the temperature, the volume, and the total number of particles (e.g., *N<sub>P</sub>* + *N<sub>N</sub>* or *N<sub>P</sub>* + *N<sub>M</sub>*).

In the dual topology approach of the alchemical methods, we mix the two Hamiltonians of eq 1 into a single Hamiltonian, which is a function of an order parameter  $\lambda$ :

$$\begin{aligned} H(P, M, N, \lambda) &= K(P, M, N) + U(P) \\ &\quad + (1 - \lambda)[U(P, N) + U(N)] + \lambda[U(P, M) + U(M)] \end{aligned} \quad (3)$$

The order parameter  $\lambda$  varies between 0 and 1. To ensure the stability of MD simulations, the kinetic energy is not scaled to zero, so the end state presents a gas of noninteracting particles. The free energies of noninteracting particles (ideal gas) are estimated analytically as explained in textbooks<sup>21</sup> and removed from the total. At  $\lambda = 0$ , the alchemical Hamiltonian includes interacting particles “*P,N*” and ideal gas particles “*M*”. At  $\lambda = 1$  the alchemical Hamiltonian describes interacting particles “*P,M*” and ideal gas particles “*N*”. This straightforward dual topology approach helps us illustrate the properties of a TC. On the other hand, it presents numerical difficulties that will be discussed later.

The free energy difference between the initial and final state according to the alchemical Hamiltonian is

$$\begin{aligned} f(\lambda = 1) - f(\lambda = 0) &= \Delta f = -\beta^{-1} \\ &\quad \ln \frac{\int d\Gamma_{P,N,M} \exp\{-\beta[K(P, M, N) + U(P) + U(P, M) + U(M)]\}}{\int d\Gamma_{P,N,M} \exp\{-\beta[K(P, M, N) + U(P) + U(P, N) + U(N)]\}} \end{aligned} \quad (4)$$

The integration over the momenta yields the same free energy contribution in the initial and final states and cancels out. In the initial (final) state, the “*M*” (“*N*”) particles do not interact with any other particle, so each of their configuration integrals is the volume *V*. Therefore, the integration over the *N<sub>N</sub>* decoupled atoms at the numerator and the *N<sub>M</sub>* decoupled atoms at the denominator contributes to the logarithm of the right-hand side of eq 4 with  $V^{N_N}/V^{N_M}$ . We have

$$\Delta f = -\beta^{-1} \ln \frac{V^{N_N}}{V^{N_M}} - \beta^{-1} \ln \frac{\int d\Gamma'_{P,M} \exp\{-\beta(U(P) + U(P, M) + U(M))\}}{\int d\Gamma'_{P,N} \exp\{-\beta(U(P) + U(P, N) + U(N))\}} \quad (5)$$

The alchemical Hamiltonian preserves the number of particles of the reactant and the product. Particles that are missing from the physical system are made into ideal gas atoms that bring a textbook contribution to the free energy.<sup>21</sup> The difference between the alchemical difference and the free energy of mutations (in which the number of particles changes) is summarized in the formula below:

$$\Delta F_{M,N} = \Delta f - \beta^{-1} \ln \left[ \frac{C_{PN}}{C_{PM}} \left( \frac{\prod_{i \in M} \frac{2\pi m_i}{h^2 \beta} V^{2/3}}{\prod_{i \in N} \frac{2\pi m_i}{h^2 \beta} V^{2/3}} \right)^{3/2} \right] = \Delta f - \Delta F_{\text{corr}} \quad (6)$$

Computing a single free energy difference between the native and the mutant using alchemical pathway requires the calculation of  $\Delta F_{\text{corr}}$ .

## THERMODYNAMIC CYCLE (TC)

Alchemical methods are often used in thermodynamic cycles (TC)<sup>22</sup> to compare an experimental measurement to an equivalent quantity that can be computed but is not accessible to direct experimental measurements. As an example, we refer to Figure 1.

The two different black shapes correspond to two different conformations of the protein, “P1” and “P2”. The green triangle is a native residue, “N”, and the red hexagon represents a mutated residue “M”. The free energy differences associated with transitions between the four states of the system are of conformational transitions (horizontal arrows) and of mutations of one residue (vertical arrows).

We are interested in the relative stability of the two conformational states (“P1” and “P2”) for the two mutants (“PN” and “PM”), i.e.,

$$\Delta \Delta F_{\text{exp}} = \Delta F_{N-\text{conf}} - \Delta F_{M-\text{conf}} \quad (7)$$

which is measured experimentally. The computation of the transition between the two conformations and the convergence of free energy differences may be difficult to obtain and/or expensive. For a complete TC, the free energy change is zero, i.e.,  $\Delta F_{N-\text{conf}} + \Delta F_{2-\text{mut}} - \Delta F_{M-\text{conf}} - \Delta F_{1-\text{mut}} = 0$ , which we exploit to write

$$\begin{aligned} \Delta \Delta F_{\text{exp}} &\equiv \Delta F_{N-\text{conf}} - \Delta F_{M-\text{conf}} \\ &= \Delta F_{1-\text{mut}} - \Delta F_{2-\text{mut}} \\ &\equiv \Delta \Delta F_{\text{mut}} \end{aligned} \quad (8)$$

We can therefore compute  $\Delta \Delta F_{\text{mut}}$  instead of  $\Delta \Delta F_{\text{exp}}$ , which is less costly.

The free energy differences of mutation computed by alchemical methods (see eq 5) are

$$\begin{aligned} \Delta f_{1-\text{mut}} &= -\beta^{-1} \ln \frac{V^{N_N}}{V^{N_M}} - \beta^{-1} \ln \frac{\int d\Gamma'_{P,M} \Theta_{P1} \exp\{-\beta(U(P) + U(P, M) + U(M))\}}{\int d\Gamma'_{P,N} \Theta_{P1} \exp\{-\beta(U(P) + U(P, N) + U(N))\}} \\ \Delta f_{2-\text{mut}} &= -\beta^{-1} \ln \frac{V^{N_N}}{V^{N_M}} - \beta^{-1} \ln \frac{\int d\Gamma'_{P,M} \Theta_{P2} \exp[-\beta(U(P) + U(P, M) + U(M))]}{\int d\Gamma'_{P,N} \Theta_{P2} \exp[-\beta(U(P) + U(P, N) + U(N))]} \end{aligned} \quad (9)$$

where  $\Theta_X$  is a Heaviside function which is equal to one if the coordinate vector belongs to conformation  $X$  and is zero otherwise. To adjust the free energy differences of the dual topology to the free energy of mutation, the correction must be added to each of these terms (eq 6). We have

$$\begin{aligned} \Delta F_{1-\text{mut}} &= \Delta f_{1-\text{mut}} - \Delta F_{\text{corr}} \\ \Delta F_{2-\text{mut}} &= \Delta f_{2-\text{mut}} - \Delta F_{\text{corr}} \end{aligned} \quad (10)$$

The correction to the free energy differences is independent of the protein conformation (eq 6). Therefore, the relative free energy difference computed by alchemical methods ( $\Delta \Delta f_{\text{mut}}$ ) is the same as the actual relative free energy ( $\Delta \Delta F_{\text{mut}}$ ):

$$\begin{aligned} \Delta \Delta F_{\text{mut}} &= \Delta F_{1-\text{mut}} - \Delta F_{2-\text{mut}} \\ &= \Delta f_{1-\text{mut}} - \Delta F_{\text{corr}} - \Delta f_{2-\text{mut}} + \Delta F_{\text{corr}} \\ &= \Delta \Delta f_{\text{mut}} \end{aligned} \quad (11)$$

The elimination of the correction terms suggests that the end points of the calculations can be manipulated to our advantage (as first suggested in ref 16) to minimize the cost of the calculations. As long as they cancel out, the correct relative free energy difference is obtained.

## RETAINING ALL THE SELF-INTERACTIONS

We illustrate an alchemical pathway that retains the self-interactions of the substituted fragment and provides the correct relative free energy difference. This pathway avoids the scaling of the self-interactions by the order parameter  $\lambda$  in the interpolating Hamiltonian:

$$\begin{aligned} H'(P, M, N, \lambda) &= K(P, M, N) + U(P) \\ &\quad + (1 - \lambda)U(P, N) + \lambda U(P, M) \\ &\quad + U(N) + U(M) \end{aligned} \quad (12)$$

If  $\lambda = 0$ , the system described by the Hamiltonian is composed by particles “P” and “N” and by a molecule “M” in the gas phase. When  $\lambda = 1$ , the “P” and “M” particles make the mutant protein, while “N” is a molecule in the gas phase.

The free energy difference between the initial and final state computed using this Hamiltonian is

$$\begin{aligned} \Delta f' &= -\beta^{-1} \ln \frac{\int d\Gamma'_{P,M,N} \exp\{-\beta[U(P) + U(P, M) + U(M) + U(N)]\}}{\int d\Gamma'_{P,M,N} \exp\{-\beta[U(P) + U(P, N) + U(M) + U(N)]\}} \end{aligned} \quad (13)$$

The key observation is that the “N” particles in the numerator and “M” particles in the denominator do not interact with the

protein or the solution; therefore, their contribution to the free energy difference can be isolated:

$$\Delta f' = -\beta^{-1} \ln \frac{\int d\Gamma'_{P,M} \exp\{-\beta[U(P) + U(P, M) + U(M)]\}}{\int d\Gamma'_{P,N} \exp\{-\beta[U(P) + U(P, N) + U(M)]\}} + -\beta^{-1} \ln \frac{\int dQ_N \exp[-\beta U(N)]}{\int dQ_M \exp[-\beta U(M)]} \quad (14)$$

The deviation of this free energy difference and the one obtained with alchemical methods that scale the self-interactions (eq 5) is

$$\Delta f = \Delta f' + \beta^{-1} \ln \frac{V^{N_M} \int dQ_N \exp(-\beta U(N))}{V^{N_N} \int dQ_M \exp(-\beta U(M))} \quad (15)$$

Similar to our previous argument about the difference between  $\Delta F$  and  $\Delta f$ , the difference in eq 15 is independent of the protein conformations, and we therefore have

$$\begin{aligned} \Delta \Delta f_{\text{mut}} &= \Delta f_{1-\text{mut}} - \Delta f_{2-\text{mut}} \\ &= \Delta f'_{1-\text{mut}} - \Delta f'_{2-\text{mut}} \\ &= \Delta \Delta f'_{\text{mut}} \end{aligned} \quad (16)$$

This proves that the  $\Delta \Delta F$  computed along this pathway that avoids the annihilation/creation of self-interactions of the substituted fragment is exact. This proof is similar to the one reported in ref 18, where only the nonbonded interactions are scaled. Our argument is more general: all the interactions involving only “M” or “N” particles are left “as are”. The gas phase molecule with nonbonded interactions on will have more restricted conformational space to explore and will have better overlap with neighboring distributions.

## ■ SCALING OF “P”–“M”(“N”) BONDED INTERACTIONS

The path discussed in the previous section requires that all of the external interactions (e.g., interactions involving “P” and “N” particles) be annihilated/created to bring the system to the correct end states. If the annihilated chemical group is decoupled from the rest of the environment, it explores the whole simulation box, which makes statistical convergence difficult.<sup>16</sup> A solution is to restrain the overall relative translations and rotations of the fragment with respect to the scaffold. An algorithm to that effect was presented for free energy calculation of binding of a ligand to an enzyme, and is called the Virtual Bond Algorithm (VBA).<sup>20</sup>

According to VBA, it is possible to retain a few bonded “P”–“M”(“N”) interactions by cross-linking the six external degrees of freedom of the annihilated particles to the “P” atoms. If there are no interactions between the annihilated particles (say “M”) and the “P” atoms, the partition function of the overall system is

$$Z(P, M) = Z(P) Z(M) = Z(P) Z_{\text{int}}(M) 8\pi^2 V \quad (17)$$

On the right-hand side of eq 17, we separated the partition function of the “M” molecule/fragment  $Z(M)$  into the partition function of its internal degrees of freedom  $Z_{\text{int}}(M)$  and the partition function of its external degrees of freedom, which is  $8\pi^2 V$ .<sup>20</sup>

The VBA algorithm makes it possible to find a transformation of coordinates such that we can isolate the six external degrees of freedom of the “M” species and restrain them to the “P” particles. This yields

$$Z'(P, M) = Z(P) Z_{\text{int}}(M) Z_{P-M} \quad (18)$$

Here,  $Z_{P-M}$  is the cross-linking partition function, i.e., the partition function of the six restraints on the external degrees of freedom of the “M” molecule/fragment that restrain its relative distance and orientation with respect to the “P” molecule/scaffold. It is important to highlight that  $Z_{P-M}$  does not depend on the coordinates of the “P” molecule/scaffold. The free energies of the system with the free “M” molecule/fragment (eq 17) and the one with the cross-linked “M” molecule/fragment (eq 18) are different. Their difference is

$$\begin{aligned} \Delta F_{P-M} &= -\beta^{-1} \ln Z'(P, M) + \beta^{-1} \ln Z(P, M) \\ &= -\beta^{-1} \ln \frac{Z_{P-M}}{8\pi^2 V} \end{aligned} \quad (19)$$

Since the six restraints in  $Z_{P-M}$  are independent,  $Z_{P-M}$  can be written as the product of six one-dimensional integrals. These integrals may be solved analytically or numerically, but no further MD simulations are required.

According to VBA, it is possible to retain one bond, two angles, and three torsions, chosen in such a way that they involve three “P” particles and three “M” or “N” particles.<sup>20</sup> Since we often have more bonded interactions between “P” and “M” or “N” particles, we still have to deal with the alchemical annihilation/creation of a few bonded interactions. In the methods section and in the Supporting Information, we propose a route to remove such interactions, avoiding problems in angles, torsions, and improper torsions due to the removal of bond angles.

In a thermodynamic cycle, these cross-linking restraints do not affect the relative free energy difference; i.e.,  $\Delta \Delta F$  computed with a cross-linked system is identical to  $\Delta \Delta F$  computed without the cross-linking interactions. To illustrate this, let us refer to Figure 1. The restraining potential appears only in the two mutations (vertical arrows). In the first mutation (right vertical arrow), the simulation is run with the cross-links “on” for both the mutant and the native fragments. To correct for this bias, we need to add the free energy contribution of restraining the mutant to the protein ( $\Delta F_{P-M}$ ) and remove the free energy difference of restraining the native ( $-\Delta F_{P-N}$ ). In the second mutation (left vertical arrow), the simulations are run again with the cross-links “on” for both the mutant and the native. However, in this case, we remove the free energy contribution of restraining the mutant ( $-\Delta F_{P-M}$ ) and add the free energy difference of restraining the native ( $\Delta F_{P-N}$ ). The restraining partition function is decoupled from the “P” partition function (see eq 18); therefore, it is not affected by the conformation of the protein. The contribution to the relative free energy difference due to the cross-linking  $\Delta \Delta F_{\text{CL}}$  is

$$\Delta \Delta F_{\text{CL}} = \Delta F_{P-M} - \Delta F_{P-N} + \Delta F_{P-N} - \Delta F_{P-M} = 0 \quad (20)$$

The possibility of retaining the external bonded interactions was discussed in ref 18, where the authors did not use VBA but considered examples to explain which bonded interactions between “P”–“M”(“N”) may be retained without introducing “spurious correlations” within the protein. In ref 19, it was highlighted that the contribution of dummy atoms (i.e., atoms that do not interact with the environment) cancels out exactly



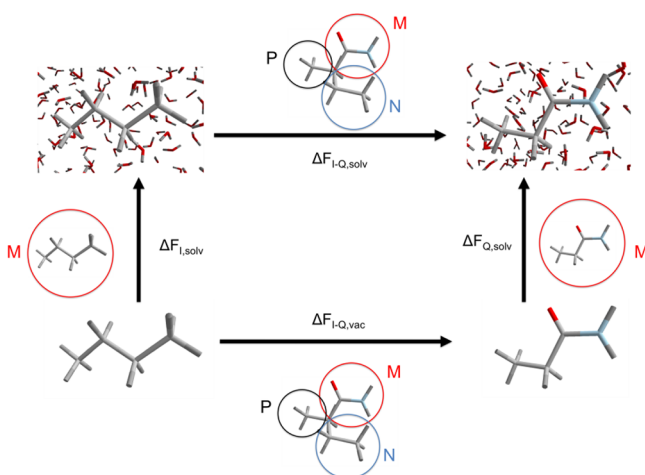
only if the set of bond angles and dihedrals between dummy atoms and the environment is nonredundant.

With respect to the arguments reported elsewhere,<sup>18,19</sup> the use of VBA theory clarifies how to choose “non-redundant” interactions between “P”–“M” (“N”), and so how to avoid “spurious correlations”.

A direct application of VBA is not possible if the created/annihilated fragment has more than one chemical bond with the scaffold. In this case, it is necessary to perform two simulations. First, all of the bonded interactions in excess of those required by VBA are removed. Since the fragment is still bound to the scaffold, the computational cost of this removal should not be too expensive. Then, the regular VBA can be applied.

## NUMERICAL EXAMPLE

To illustrate the theory, we study an alchemical substitution that retains all of the self-interactions. The difference in hydration free energies of ILE and GLN side chain analogs is computed. The cycle that we consider is presented in Figure 2.



**Figure 2.** The thermodynamic cycle that we used in our calculations. The horizontal arrows represent free energy differences upon mutation of the solute (ILE side chain analog and GLN side chain analog). The vertical arrows represent the solvation process; i.e., the solute is brought from a vacuum into solution. For each free energy calculation, the colored circles highlight which atoms are considered “P” (black), “N” (blue), or “M” (red).

$\Delta F_{I-Q,solv}$  is the free energy difference of mutating ILE to GLN side chain analog in solution, while  $\Delta F_{I-Q,vac}$  is the same quantity computed in a vacuum (gas phase).  $\Delta F_{I,solv}$  and  $\Delta F_{Q,solv}$  are the free energy differences of inserting ILE and GLN side chain analogs into water from the gas phase, respectively.

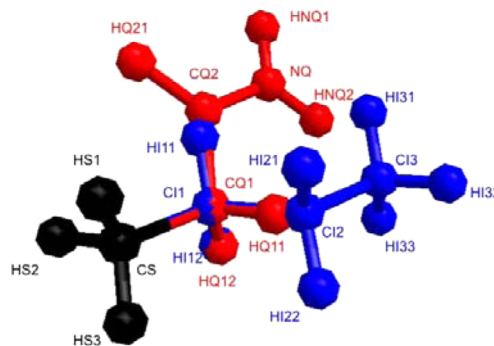
The free energy difference of the complete cycle is (the  $q$ -s are the partition functions)

$$\oint dF = \Delta F_{I-Q,solv} - \Delta F_{Q,solv} - \Delta F_{I-Q} + \Delta F_{I,solv} \\ = -\beta^{-1} \ln \left[ \frac{q_{solv,Q} q_{solv,Q} q_I q_{solv,I}}{q_{solv,I} q_{solv,Q} q_Q q_{solv,I}} \right] = 0 \quad (21)$$

as it should be. The numerical task and test is to reproduce the zero in simulations.

## MUTATIONS

To compute the free energy differences  $\Delta F_{I-Q,solv}$  and  $\Delta F_{I-Q}$  we used the system shown in Figure 3.



**Figure 3.** The alchemical intermediate that we use. The black methyl group is shared between the two side chain analogs. The blue atoms are the rest of the ILE side chain analog. The red atoms are the remainder of the GLN side chain analog. While the black atoms interact with all the other atoms, blue and red particles do not interact with each other.

The black methyl group is considered the “P” part of the system, i.e., the part that interacts with both mutants during TI. The blue atoms are the native “N”, i.e., the ILE side chain without its methyl group that is considered “P”. The red atoms are the mutant “M”, i.e., the GLN side chain analog without the methyl group which is part of “P”. To illustrate our procedure, the simulations are performed while retaining the full strength of the self-interactions. More details about the simulations follow.

## SOLVATION

To compute the free energy difference of the two side chains in solution and in the gas phase, the nonbonded interactions between the side chain analog and the solvent around it are turned off.<sup>23</sup> Following the nomenclature introduced in the previous sections, we consider water as “P” molecules and the side chain analog (ILE or GLN) as “M” particles. No “N” particles are involved in these simulations.

## ALCHEMICAL PATHWAY

In the previous sections, we described a Hamiltonian for the alchemical substitution that depends linearly on  $\lambda$  (see eq 3 and eq 12). This is the most straightforward choice, but not necessarily the best one from a numerical perspective. Indeed, the creation and the annihilation of particles with a finite van der Waals radius present (integrable) singularities.<sup>24</sup> Such singularities are removed if the van der Waals interactions are created/annihilated with scaling parameter  $\lambda^\alpha$ , with  $\alpha \geq 4$ .<sup>24</sup> Moreover, the use of a soft core Lennard-Jones potential avoids hard clashes during the creation/annihilation of interactions between particles.<sup>24</sup> The potential used is then

$$\lambda^4 U(r) = 4\epsilon \lambda^4 \left\{ \frac{\sigma^{12}}{[r^6 + \alpha_{LJ}(1 - \lambda)^2 \sigma^6]^2} - \frac{\sigma^6}{[r^6 + \alpha_{LJ}(1 - \lambda)^2 \sigma^6]} \right\} \quad (22)$$

It is clear that for  $\lambda = 0$  this potential disappears. For  $\lambda = 1$ , we recover the usual Lennard-Jones potential (in this way, the

Table 1. The Alchemical Pathway for the First Phase of the Mutation<sup>a</sup>

	MPH1									
	scaling					correction				
	P–P	P–M	P–N	M–M	N–N	P–P	P–M	P–N	M–M	N–N
Vdw	1	$\lambda^4$	1	1	1	soft core				
Ele	1	0	$(1 - \lambda)$	1	1					
Bond	1	1	1	1	1	Urey–Bradley				
Angle	1	$\lambda^2$	1	1	1					
Tors	1	0	$(1 - \lambda)^2$	1	1					
ImpT	1	0	$(1 - \lambda)^2$	1	1					

<sup>a</sup>The angular and van der Waals interactions between “P” and “M” particles are created, while torsions, improper torsions, and electrostatic interactions between “P” and “N” are progressively switched off. The right hand side of the table reports the corrections to the regular OPLS-AAL<sup>26</sup> potential that were used. Notice that the self-interactions, as well as the bonds, are never scaled (see details in the text).

Table 2. The Alchemical Pathway for the Second Phase of the Mutation<sup>a</sup>

	MPH2									
	scaling					correction				
	P–P	P–M	P–N	M–M	N–N	P–P	P–M	P–N	M–M	N–N
Vdw	1	1	$(1 - \lambda)^4$	1	1	soft core				
Ele	1	$\lambda$	0	1	1					
Bond	1	1	1	1	1	Urey–Bradley				
Angle	1	1	$(1 - \lambda)^2$	1	1					
Tors	1	$\lambda^2$	0	1	1					
ImpT	1	$\lambda^2$	0	1	1					

<sup>a</sup>The torsions, improper torsions, and electrostatic interactions between “P” and “M” particles are created, while angular and van der Waals interactions between “P” and “N” are progressively switched off. The right hand side of the table reports the corrections to the regular OPLS-AAL<sup>26</sup> potential that were used. Notice that the self-interactions, as well as the bonds, are never scaled (see details in the text).

correct end state is recovered). For  $0 < \lambda < 1$ , the hard core is softened; i.e., there is no divergence at  $r = 0$ . The larger the value of  $\alpha_{LJ}$ , the softer is the potential. In our simulations we used  $\alpha_{LJ} = 0.3$ .

To avoid the collapse of particles with opposite charges to the same point in space while the van der Waals repulsion is reduced, the free energy calculation is conducted in two steps: in the first step, the electrostatic interactions of the “N” particles are turned off, and in the second step, the van der Waals interactions are removed.<sup>25</sup> The opposite is done for “M” particles (first the van der Waals interactions are created, and then the electrostatic interactions are turned on).

While we retain (of course) all of the internal covalent interactions in “N” and “M”, some of the external covalent interactions need to be annihilated/created. Indeed, as we discussed in the section Scaling of “P”–“M”(“N”) Bonded Interactions, there is a maximal number of external bonded interactions that can be retained throughout the calculation. All of the other interactions have to be removed to properly decouple the partition function into separated terms. In the Supporting Information (SI), the computational implementation of this annihilation/creation is discussed. In particular, Urey–Bradley (UB) terms, instead of the angular harmonic potential, are used for the angles to be created/annihilated. UB terms are regular bonds in Cartesian space that are placed between two atoms sharing an angle but not directly bonded. The reason for the use of the UB terms is numerical stability: the removal/creation of a bond angle is problematic, as explained in the SI. In the solvation simulations, there are no angle terms annihilated/created; therefore, only the regular bond-angle potential is used. This inconsistency introduces a correction to the free energy (the difference between the UB and the usual harmonic angular

terms) that we need to compute explicitly. Such calculation is described in the SI. In this numerical illustration the resulting correction is negligible.

In what follows, given the small size of the “P” part of the solute (it is just the black methyl group in Figure 3), we decided to retain only the external bond between the CS and C11/CQ1 (see Figure 3). This avoids the drift of the decoupled fragment in the simulation box. All of the other external bonded interactions were treated in the same way and alchemically removed/created.

We further use an order parameter  $\lambda^\beta$  with  $\beta \geq 2$  for the scaling of the bond angles (see the SI).

Finally, we suggest that the torsion and improper torsion interactions should be removed before the bond angles (see the SI).

A summary of all these considerations is in Tables 1 and 2. We break the free energy calculation into two parts: Mutation-PHase1 (MPH1) and MutationPHase2 (MPH2). In MPH1, we remove electrostatic interactions, torsions, and improper torsions between “P” and “N” particles and create van der Waals interactions and angles between “P” and “M” particles. In MPH2, we do the opposite: we remove van der Waals interactions and angles between “P” and “N” particles and create electrostatic interactions, torsions, and improper torsions between “P” and “M” particles. Note that we do not create or remove the bond between particles “P” and “M” or the bond between “P” and “N.” Note also that all the self-interactions, bonded and nonbonded, are never created or annihilated.

To ensure accurate estimates of the electrostatic interaction, we use particle mesh Ewald.<sup>27</sup> The drawback of this method with respect to the alchemical substitution is that the reciprocal space contribution to the energy cannot be separated into the specific contributions of each pair of interacting particles. Therefore, it is

not obvious how to selectively annihilate/create only the interactions between “P” and “M” or “N” particles.

Recall the PME theory.<sup>27</sup> We define  $n$  as the lattice vectors, and  $C(i, j)$  are interaction parameters for pairs of particles. The potential energy is a sum of the interactions of all the pairs of particles with the exception of self. The exception is denoted by the prime in the summation symbol below

$$E(\vec{r}_1, \dots, \vec{r}_N) = \frac{1}{2} \sum_{\vec{n}}' \sum_i \sum_j \frac{C(i, j)}{|\vec{r}_i - \vec{r}_j + \vec{n}|} \quad (23)$$

the PME methodology is applicable if we can write (see Appendix of ref 27):

$$C(i, j) = \pm C(i)C(j) \quad (24)$$

If we want to retain the “self” electrostatic interactions, we cannot use eq 24. Indeed, let us suppose that particle  $i$  is an “M” particle, then we have

$$C(i, j) = \begin{cases} \lambda \frac{q_i q_j}{4\pi\epsilon_0} & j \in P \\ 0 & j \in N \\ \frac{q_i q_j}{4\pi\epsilon_0} & j \in M \end{cases} \quad (25)$$

where  $q$  are charges and  $\epsilon_0$  is the electric constant. Since the “P” particles interact according to the usual Coulomb law ( $q_i q_j / 4\pi\epsilon_0$ , without  $\lambda$ ) and the “N” particles may carry a charge different from zero, we cannot write eq 25 as eq 24.

To solve this problem, we compute the reciprocal space energy term between selected types of particles and scale the interactions with the proper switching parameter. A sum of these terms gives us an overall electrostatic potential energy in which the self-interactions are not scaled.

The recipe is summarized in Tables 3 and 4, which explain how to carry out the Ewald sums for MPH1 and MPH2.

**Table 3. Computing the Reciprocal Space Electrostatic Interactions by Ewald Sums in MPH1<sup>a</sup>**

MPH1			
#	type of particle interacting	scaling	result
1	P	$\lambda$	$\lambda U(P)$
2	P N	$(1 - \lambda)$	$(1 - \lambda)[U(P) + U(P, N) + U(N)]$
3	M	1	$U(M)$
4	N	$\lambda$	$\lambda U(N)$
OVERALL: $U(P) + (1 - \lambda)U(P, N) + U(N) + U(M)$			

<sup>a</sup>The reciprocal space is computed four times. In each of these calculations only a subset of charges is kept on, while the others are set to zero. The resulting potential and forces are then scaled by the switching parameter. In the first calculation, only the electrostatic interactions between the “P” particles are kept on, and the energy and forces are scaled by  $\lambda$ . Second, the electrostatic interactions of “P” and “N” particles are turned on, and the resulting energy and forces are scaled by  $1 - \lambda$ . Third, only the “M” particles have their charges on, and the result is not scaled. Finally, only the charges of the “N” particles are kept on, and the result is scaled by  $\lambda$ . The sum of these four contributions is shown at the bottom of the table. The self-interactions are not scaled, and the external interactions are scaled by  $1 - \lambda$ .

**Table 4. Computing the Reciprocal Space Electrostatic Interactions by Ewald Sums in MPH2<sup>a</sup>**

MPH2			
#	type of particle interacting	scaling	result
1	P	$(1 - \lambda)$	$(1 - \lambda)U(P)$
2	P M	$\lambda$	$\lambda[U(P) + U(P, M) + U(M)]$
3	M	$(1 - \lambda)$	$(1 - \lambda)U(M)$
4	N	1	$U(N)$
OVERALL: $U(P) + \lambda U(P, M) + U(M) + U(N)$			

<sup>a</sup>The reciprocal space is computed four times. In each of these calculations, only a subset of charges is kept on, while the others are set to zero. The resulting potential and forces are then scaled by the switching parameter. In the first calculation, only the electrostatic interactions between the “P” particles are kept on, and the energy and forces are scaled by  $1 - \lambda$ . Secondly, the electrostatic interactions of “P” and “M” particles are turned on, and the resulting energy and forces are scaled by  $\lambda$ . Third, only the “M” particles have their charges on, and the result is scaled by  $1 - \lambda$ . Finally, only the charges of the “N” particles are kept on, and the result is not scaled. The sum of these four contributions is shown at the bottom of the table. The self-interactions are not scaled, and the external interactions are scaled by  $\lambda$ .

This solution has two major disadvantages. First, using PME four times per time step is an expensive procedure. Second, the formulation is not exact. Indeed, we are neglecting a term in the Ewald summation that is proportional to the square of the dipole of the unit cell.<sup>27</sup> While this term is routinely neglected in solvated simulations, it is hard to argue that this term is zero in a vacuum calculation or when we use PME to evaluate  $U(M)$  or  $U(N)$ . From this perspective, further developments of the PME theory are needed to extend its applicability to the type of problem hereby described.

The use of PME in the context of calculation of the solvation free energy was debated.<sup>28</sup> The periodicity of the system adds some interactions between the solute in the unit cell and its periodic images. The solvation of an isolated molecule in liquid requires that all interactions between the solute in the unit cell and its periodic copies be removed. The advantage of PME is that the energy is conserved, providing well-defined statistical mechanic properties throughout the cycle. We use PME also in the simulations carried out in a vacuum to ensure consistency. The differences of free energies computed with and without PME in a vacuum were small ( $\sim 0.02$  kcal/mol).

## ■ THERMODYNAMIC INTEGRATION

To compute numerically the free energy differences defined in eq 21, we use thermodynamic integration (TI):<sup>29</sup>

$$\Delta f' = \int_0^1 d\lambda \frac{dF}{d\lambda} = \int_0^1 d\lambda \left\langle \frac{\partial H'}{\partial \lambda} \right\rangle_\lambda \quad (26)$$

where  $H'(\lambda)$  is the sum of kinetic and potential energies, and its dependence on the order parameter  $\lambda$  is defined in Tables 1 and 2.

The brackets in eq 26 denote an ensemble average performed at a fixed value of  $\lambda$ .

The simulations are performed at a finite set of  $\lambda$  states. The numerical integrations are performed according to the trapezoidal rule; i.e., the integral is approximated by

$$\Delta f' \approx \left\langle \frac{\partial H'}{\partial \lambda} \right\rangle_{\lambda_1} \lambda_1 + \frac{1}{2} \sum_{i=2}^{N_\lambda} \left[ \left\langle \frac{\partial H'}{\partial \lambda} \right\rangle_{\lambda_i} + \left\langle \frac{\partial H'}{\partial \lambda} \right\rangle_{\lambda_{i-1}} \right] (\lambda_i - \lambda_{i-1}) + \left\langle \frac{\partial H'}{\partial \lambda} \right\rangle_{\lambda_{N_\lambda}} (1 - \lambda_{N_\lambda}) \quad (27)$$

where  $N_\lambda$  is the number of  $\lambda$  states used.

The configurations are sampled with a MD simulation using a variant of the program MOIL<sup>30</sup> with the MUTA option (<https://wiki.ices.utexas.edu/clsb/wiki/GetSubversionMoil>).

## SIMULATION DETAILS

We used the OPLS-AAL force field<sup>26</sup> with a modification on the charges of the  $C_\beta$  of the residue analogs.<sup>23</sup> Also, the Lennard-Jones parameters for polar hydrogens belonging to the solutes were set to  $\sigma = 0.3 \text{ \AA}$  and  $\epsilon = 0.0498 \text{ kcal/mol}$ . All of the systems were solvated in a periodic cubic box of volume  $34.45 \text{ \AA}^3$  with 1355 TIP3P water molecules<sup>31</sup> at 300 K. The size of the water box was chosen in such a way to have an average water density at corners of the box near  $0.998 \text{ g/cm}^3$  with the ILE side chain analog solvated in the center of the water box. The rationale of this choice is that at the corners of the box we are far enough from the solute to be in the “bulk.”

The cutoff distance was 10  $\text{\AA}$  for van der Waals interactions and 14  $\text{\AA}$  for electrostatic forces. These real space cutoff distances gave very good energy conservation throughout the simulation on a 2 ns test case conducted in the NVE ensemble. Particle meshed Ewald (PME)<sup>27</sup> was used in all simulations with a grid of  $64 \times 64 \times 64$ . To be consistent with the solvated simulations, PME was used also in the simulations carried out in a vacuum.

The contributions of van der Waals interactions between particles that are beyond the cutoff distance are (of course) not considered explicitly in this simulation. Other authors have used an analytical correction for the free energy of solvation of small molecules.<sup>23</sup> The solvent distribution around each of the alchemically substituted particles of the solute is assumed homogeneous and isotropic. The cutoff is set large enough so there are no correlations between the particle at the origin and the particles beyond the cutoff. In this case, the pair correlation function between the alchemically substituted solute and the solvent particles is independent of  $\lambda$  and is equal to 1. The final correction formula is<sup>23,32</sup>

$$\Delta F_{\text{LRC}} = 4\pi \sum_{i \in M} \rho_O \int_{r_c}^{\infty} dr_{i,O} r_{i,O}^2 U_{\text{vdW}}(r_{i,O}) - 4\pi \sum_{i \in N} \rho_O \int_{r_c}^{\infty} dr_{i,O} r_{i,O}^2 U_{\text{vdW}}(r_{i,O}) \quad (28)$$

Here, the density  $\rho_O$  is the number density of oxygen in water molecules. The van der Waals interactions considered are only between the solute and oxygen, according to the TIP3P water model.

The results for this correction are reported in Table 6, together with all of the other results. This correction is not affecting the overall free energy cycle. In Figure 2, we notice that each correction is added (at the end of the solvation simulation for ILE and at the end of the mutation for GLN) and then removed (at the beginning of the mutation simulation for ILE and the end of the solvation simulation for GLN) per each solute. The net effect is zero.

In the two mutations, the geometric center of the “P” particles was restrained to the center of the box by a harmonic potential

with spring constant  $k = 1 \text{ kcal/mol \AA}^{-2}$ . In the two calculations in solvent, the geometric center of the analog was restrained to the center of the box by a harmonic potential with the same spring constant  $k$ . In both of the cases, the free energy contribution from this harmonic potential restrains an external degree of freedom, so it is separable. Since the two spring constants are the same, the free energy of the system is not affected by the choice of the subset of particles to restrain. Therefore, there is no overall contribution from this restraint to the total free energy difference.

The solvated simulations were performed in the NVT ensemble by rescaling the velocities at every time step to maintain the temperature of the thermal bath. The vacuum simulations were performed in the NVT ensemble using a Langevin thermostat<sup>32</sup> to enhance coupling between different degrees of freedom and ergodicity.

The TI was performed following the “multiconfiguration” approach.<sup>33</sup> The number of intermediate  $\lambda$  states at which the free energy differences were sampled and computed are listed in Table 5. They were chosen to allow a good description of the regions in  $dF/d\lambda$  characterized by a large curvature.

**Table 5. The Number of  $\lambda$  States Used in the Evaluation of the Free Energy Differences of Figure 2 for Each of the Mutation Phases MPH1 and MPH2<sup>a</sup>**

	MPH1	MPH2	total simulation time
$\Delta F_{\text{I-Q,solv}}$	33	33	132 ns
$\Delta F_{\text{I-Q,vac}}$	33	33	132 ns
$\Delta F_{\text{I,solv}}$	35	23	116 ns
$\Delta F_{\text{Q,solv}}$	35	23	116 ns

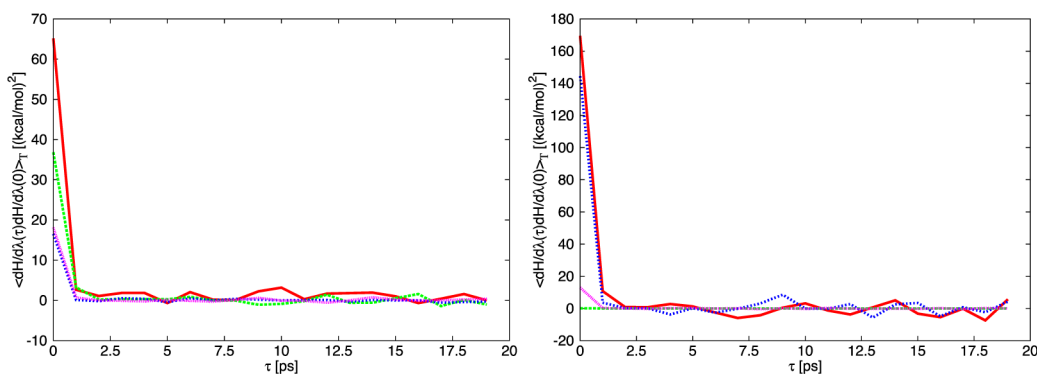
<sup>a</sup>More details on MPH1 and MPH2 are available in Tables 1 and 2. The last column shows the total simulation time for each stage in the cycle.

For the solvated simulations, at each  $\lambda$  state, two equilibrations of 100 ps were performed: one with the solute frozen and the other with the solute free to move. Then a 2-ns-long sampling of configurations was carried out for each  $\lambda$  state. For the simulations carried out in a vacuum, we sampled configurations for 2 ns using Langevin dynamics with friction coefficient  $\gamma = 60 \text{ ps}^{-1}$ . One configuration per picosecond was used to compute the average and the variance of  $dH/d\lambda$ . The configurations sampled in this way may be correlated. To account for this correlation, we computed the variance of  $dH/d\lambda$  using the following formula:<sup>32,34</sup>

$$\sigma^2 \left( \left\langle \frac{dH}{d\lambda} \right\rangle_T \right) = \frac{2}{T} \int_0^T d\tau \left( 1 - \frac{\tau}{T} \right) \left\langle \left( \frac{dH}{d\lambda}(\tau) - \left\langle \frac{dH}{d\lambda} \right\rangle_T \right) \left( \frac{dH}{d\lambda}(0) - \left\langle \frac{dH}{d\lambda} \right\rangle_T \right) \right\rangle_T \quad (29)$$

In this equation, the symbol “ $\langle \dots \rangle_T$ ” refers to a time average over the sampled configurations carried out for the length of simulation  $T$ . The argument of the integral contains the correlation function of  $dH/d\lambda$ . The correlation function decays to 0 within the first few picoseconds (see Figure 4). Therefore, it was computed for 19 ps and then set to 0 to avoid integration over a noisy tail that may introduce unphysical contributions to the integral. The integral in eq 29 was evaluated using the





**Figure 4.** Correlation function of  $dH/d\lambda$  as a function of time (in ps). The left figure reports calculations at  $\lambda = 0.5$  of the MPH1 (red) and MPH2 (green) phases of the mutation from the ILE side chain analog to the GLN side chain analog in solution. In the same figure the MPH1 (blue) and MPH2 (pink) phases of the same mutations in a vacuum are reported. The right figure shows the calculations at  $\lambda = 0.5$  of the MPH1 (red) and MPH2 (green) of the solvation of the ILE side chain. In the same figure, the MPH1 (blue) and the MPH2 (pink) phases of the solvation of GLN side chain analog are also reported.

trapezoidal rule. Examples of correlation functions are reported in Figure 4.

The variance of the integral was estimated from the variances of  $\langle \partial H / \partial \lambda \rangle_{\lambda_i}$  using the error propagation formula.

## RESULTS

The profiles for  $dF/d\lambda$  as a function of  $\lambda$  are reported in Figure 5.

In Table 6a, the results are reported for each part of the calculation.

In the second column of Table 6a, we report the total of the free energy calculation. The total is broken into three parts: the result for the first phase (MPH1, third column), the result for second phase (MPH2, fourth column), and the long range correction (LRC, fifth column) approximately accounting for the finite van der Waals cutoff used in the simulations.

In ref 23, the authors computed the solvation free energies of these two side chain analogs. The free energy calculation was performed following an alchemical pathway similar to ours: i.e., splitting the decoupling in two steps. First, the van der Waals interactions are turned on, while the electrostatic interactions are kept off. Second, the electrostatic interactions are created, while keeping the van der Waals interactions at their full strength. The first step is analogous to MPH1, and it is reported in the third column of Table 6b. The second step is analogous to MPH2, and it is reported in the fourth column of Table 6b. Our results are close to those in ref 23. This is particularly interesting since we used a significantly different sampling protocol. First, they sampled from the NPT ensemble, while we sampled configurations from the NVT ensemble. Second, they did not use the Ewald sum to compute long-range electrostatics. They adopt a neutral group-based, tapered, finite cutoff distance for both electrostatics and van der Waals interactions. The different cutoff scheme for van der Waals interactions results in a significant difference between our long-range van der Waals corrections (fifth column in Table 6a) and theirs (fifth column in Table 6b). Overall, our agreement with the simulations previously reported (second column of Table 6b)<sup>23</sup> and the experimental results to which they compare with<sup>23,35</sup> (Table 6c) are in a reasonable range.

The expected result for the free energy difference over the complete cycle is, of course, 0 kcal/mol. The free energy computed numerically over the complete cycle is

$$\begin{aligned} \Delta F_{I-Q, \text{solv}} - \Delta F_{Q, \text{solv}} - \Delta F_{I-Q, \text{vac}} + \Delta F_{I, \text{solv}} \\ = 0.02 \pm 0.12 \text{ kcal/mol} \end{aligned} \quad (30)$$

The correction for the use of Urey–Bradley term for angles over the cycle is negligible (see SI), and our final result is therefore

$$\oint dF = 0.02 \pm 0.12 \text{ kcal/mol} \quad (31)$$

To further illustrate that the result is converged, we report the free energy of the complete cycle for the last 1 ns of simulation in Figure 6.

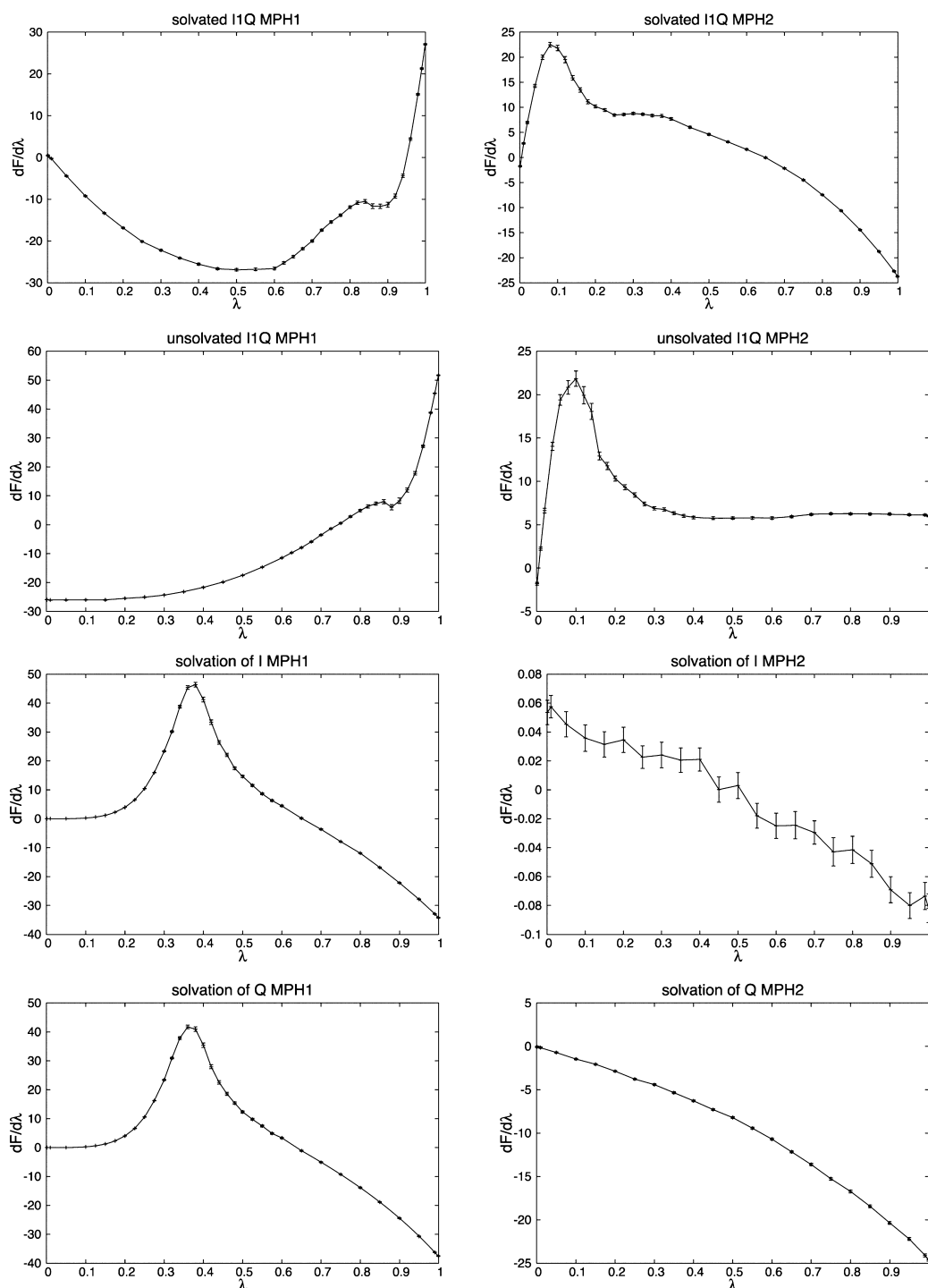
The numerical result is consistent with the analytical prediction of zero and is well within the statistical error bars. This calculation illustrates numerically the use of a thermodynamic cycle that retains the full strength of all the self-interactions.

## INFLUENCE OF THE SIZE OF THE SYSTEM

A possible dependence of the accuracy on the size of the periodic system was discovered while running the simulations in a smaller water box. The result of the cycle in this smaller box (29.35 Å size) turned out to be different from zero and of  $0.24 \pm 0.1$  kcal/mol. Inaccuracy in the treatment of long-range electrostatic interactions may be responsible for this anomalous behavior. Indeed, as we discussed in the section Alchemical Pathway, there is a term proportional to the dipole of the primitive cell that is neglected in the formulation of Ewald.<sup>27</sup> The error neglecting this term is inversely proportional to the volume of the system.<sup>27</sup> It is possible that such a term influences the accuracy of the calculation when the box is small. Further analysis is required to account for this term in free energy calculations.

## BAR VS TI

We require a relatively large number of  $\lambda$  states to accurately follow  $dF/d\lambda$  at domains of high curvature. A large number of intermediates reduces efficiency. As pointed out in ref 36, the method of Bennett acceptance ratio (BAR)<sup>37</sup> does not require as many intermediate states as TI. The equations for BAR are derived as a minimization of the variance of the estimate of the free energy difference in the limit of a large sample.<sup>37</sup> The free energy difference between two states defined by the values of the switching parameter  $\lambda_i$  and  $\lambda_{i+1}$  is computed according to the following formula:



**Figure 5.**  $dF/d\lambda$  profiles (in kcal/mol) as a function of  $\lambda$  of the phases MPH1 and MPH2 of the free energy differences of each part of the cycle of Figure 2. On the left, the MPH1 phase is reported and, on the right, the MPH2. From top to bottom, we have mutation from ILE to GLN in solution, mutation from ILE to GLN in a vacuum, solvation of ILE, and solvation of GLN. The error bars are shown in all figures.

$$\begin{aligned}
 \Delta f(\lambda_i \rightarrow \lambda_{i+1}) &= -k_B T \ln \frac{\left\langle \frac{1}{1 + \exp\{\beta[U(\lambda_{i+1}) - U(\lambda_i) - C]\}} \right\rangle_{\lambda_i}}{\left\langle \frac{1}{1 + \exp\{-\beta[U(\lambda_{i+1}) - U(\lambda_i) - C]\}} \right\rangle_{\lambda_{i+1}}} + C \\
 &\quad (32)
 \end{aligned}$$

The constant  $C$  is defined as

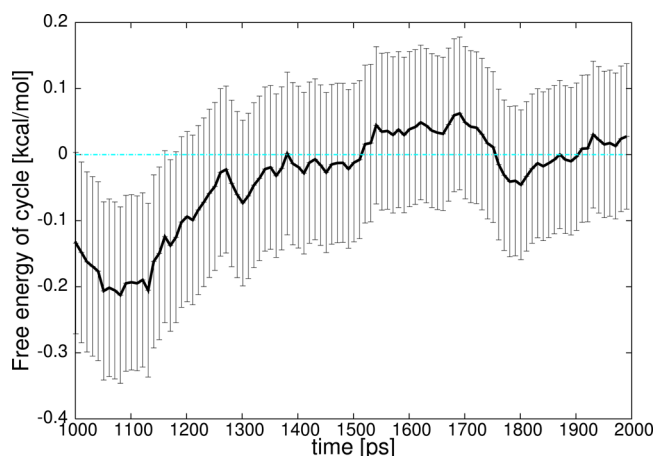
$$C = \Delta f(\lambda_i \rightarrow \lambda_{i+1}) + k_B T \ln \frac{n_{i+1}}{n_i} \quad (33)$$

where  $n_i$  and  $n_{i+1}$  are the number of configurations sampled with  $\lambda_i$  and  $\lambda_{i+1}$ , respectively. Using iteratively eq 32 and eq 33, we obtain the free energy difference between the two states. The overall free energy difference is then

Table 6. Free Energy Differences and Statistical Errors (All Values in kcal/mol)<sup>a</sup>

(a)				
	total	MPH1	MPH2	LRC
$\Delta F_{I \rightarrow Q, \text{solv}}$	$-14.073 \pm 0.066$	$-16.218 \pm 0.050$	$2.199 \pm 0.043$	$-0.051$
$\Delta F_{I \rightarrow Q, \text{vac}}$	$-2.847 \pm 0.068$	$-10.931 \pm 0.047$	$8.084 \pm 0.049$	
$\Delta F_{I, \text{solv}}$	$2.891 \pm 0.050$	$3.379 \pm 0.050$	$-0.0075 \pm 0.0019$	$-0.481$
$\Delta F_{Q, \text{solv}}$	$-8.354 \pm 0.054$	$1.858 \pm 0.047$	$-9.680 \pm 0.027$	$-0.532$
(b)				
	total <sup>23</sup>	Vdw <sup>23</sup>	Ele <sup>23</sup>	LRC <sup>23</sup>
$\Delta F_{I, \text{solv}}$	$2.73 \pm 0.03$	$3.10 \pm 0.03$	$-0.01 \pm 0.001$	$-0.37$
$\Delta F_{Q, \text{solv}}$	$-8.40 \pm 0.04$	$1.80 \pm 0.03$	$-9.79 \pm 0.02$	$-0.41$
(c)				
	experiment <sup>23,35</sup>			
$\Delta F_{I, \text{solv}}$	2.15			
$\Delta F_{Q, \text{solv}}$	-9.38			

<sup>a</sup>(a) Current simulations. The second column reports the total; the third, the free energy difference in MPH1; the fourth, the free energy difference in MPH2; and the fifth, the long-range correction (LRC). (b) Solvation free energies of the ILE and GLN side chain analogs as reported in ref 23. The second column is total; the third, the free energy difference while creating van der Waals interactions between the solute and the solvent with the electrostatic interactions off (analogous to MPH1); the fourth column reports the free energy of creating electrostatic interactions when the van der Waals are already “on” (analogous to MPH2). The last column is the long-range correction (LRC). The protocol used to carry out these calculations in ref 23 is significantly different from ours. The authors in ref 23 carried out NPT simulations, and their cutoff scheme is different (no Ewald summation for electrostatics, they used instead group-based and tapered interactions). This is reflected mostly in the LRC term, which depends on the scheme used for the van der Waals cutoff. (c) Experimental results for the solvation free energies of the two side chain analogs.<sup>35</sup> These results were used as a comparison with the simulations in ref 23.

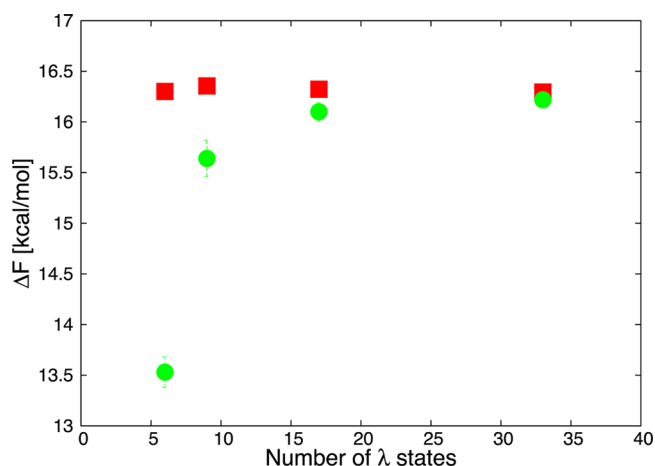


**Figure 6.** Free energy of the complete cycle as a function of time of the last 1 ns of simulation. The thickness of the black line provides the error bars. The light blue represents the exact value.

$$\Delta f' = \sum_{i=1}^{N_\lambda-1} \Delta f(\lambda_i \rightarrow \lambda_{i+1}) \quad (34)$$

To test if BAR is more efficient than TI in our example, we performed the calculation of the MPH1 phase of the mutation in solution from the ILE side chain analog to GLN using a series of sets of  $\lambda$  states decreasing in size. The largest set is the one used for the calculations reported in the results. The second one has been obtained by keeping the first and last  $\lambda$  states and removing every other  $\lambda$  state. Removal of every other state was used to create other sets with a smaller number of states. The result is reported in Figure 7.

Even when only seven  $\lambda$  states were used, BAR is within half a kcal/mol from the result obtained with all of the  $\lambda$  states, while the systematic error in TI grows much faster.



**Figure 7.** The value of the free energy difference obtained with BAR (red) and TI (green) is reported as a function of the number of  $\lambda$  states used. Missing error bars imply error smaller than the size of the symbol.

We repeated the cycle calculation with BAR for both the large and the smaller box, and the same values for the free energy (as in TI) were obtained.

## CONCLUSION

We examined and illustrated numerically a concrete dual topology pathway for a thermodynamic cycle that retains the self-interactions of the alchemical parts. We demonstrate that this path is exact, using straightforward argument based on statistical mechanics. No approximations are required. No distinction between bonded and nonbonded interactions was made. Numerical calculations illustrate high accuracy with a few nanosecond sampling at each  $\lambda$  using thermodynamic integration or BAR.

It is worth highlighting the following point: we are not proving that the self-interactions do not contribute to the free energy difference in a thermodynamic cycle. Different paths<sup>12,14,38</sup> show

different contributions of the self-interactions to the relative free energy difference. This is due to the fact that the free energy difference is a state function, but the single contributions of the different interactions to the free energy are not. So it is possible to engineer pathways in which the specific contribution of bonded interactions to the relative free energy difference are not the same. Instead, we propose an alchemical pathway that avoids their annihilation/creation. Avoiding their creation/annihilation reduces the terms that need to be turned on/off along the pathway, making, at least in principle, the calculation easier. The removal of interactions increases, in principle, the accessible space of conformations and is likely to require longer trajectories starting from a state with the full interactions on. This concept is intuitively clear when the bonded interactions are scaled down: in that case, all the atoms of the annihilated compound close to the end state are (almost) free to sample the whole simulation box,<sup>16,17</sup> and the overlap between the distributions at different  $\lambda$  values significantly diminishes. Therefore, retaining their covalent geometry reduces significantly the space to sample and increase the overlap of the intermediate distributions. The nonbonded interactions are likely to be similar. Indeed, hydrogen bonds and 1–4 interactions may help retain the conformation of a compound close to a certain structure enhancing useful overlap between distributions of sequential  $\lambda$ .

## ■ ASSOCIATED CONTENT

### ● Supporting Information

Additional discussions on concrete implementation of the thermodynamic cycle calculations are provided. We discuss alchemical changes of angles and torsions and Urey–Bradley terms. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: 512-232-5415. Fax: 512-471-8694. E-mail: [ron@ices.utexas.edu](mailto:ron@ices.utexas.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported by NIH grant GM59796, NSF grant CCF-0833162, and Welch grant F-1783 to R.E.

## ■ REFERENCES

- (1) Jorgensen, W. L. *Science* **2004**, 303, 1813.
- (2) Valleau, J. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Bruce, J., Berne, G. C., Coker, D. F., Eds.; World Scientific: Singapore, 1998.
- (3) West, A. M. A.; Elber, R.; Shalloway, D. *J. Chem. Phys.* **2007**, 126, 145104.
- (4) Kirmizialtin, S.; Elber, R. *J. Phys. Chem. A* **2011**, 115, 6137.
- (5) Straatsma, T. P.; Mccammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, 43, 407.
- (6) Pearlman, D. A. *J. Phys. Chem.* **1994**, 98, 1487.
- (7) Chipot, C.; Pohorille, A. *Free Energy Calculations. Theory and Applications in Chemistry and Biology*; Springer-Verlag: Berlin, 2007.
- (8) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, 89, 3742.
- (9) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, 72, 1047.
- (10) Wong, C. F.; Mccammon, J. A. *J. Am. Chem. Soc.* **1986**, 108, 3830.
- (11) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, 235, 574.
- (12) Tidor, B.; Karplus, M. *Biochemistry* **1991**, 30, 3217.
- (13) Chipot, C.; Rozanska, X.; Dixit, S. B. *J. Comput.-Aided Mol. Des.* **2005**, 19, 765.
- (14) Prevost, M.; Wodak, S. J.; Tidor, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88, 10880.
- (15) Sun, Y. C.; Veenstra, D. L.; Kollman, P. A. *Protein Eng.* **1996**, 9, 273.
- (16) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, 103, 103.
- (17) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, 103, 119.
- (18) Shobana, S.; Roux, B.; Andersen, O. S. *J. Phys. Chem. B* **2000**, 104, 5179.
- (19) Boresch, S. *Mol. Simul.* **2002**, 28, 13.
- (20) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, 107, 9535.
- (21) McQuarrie, D. *Statistical Mechanics*; University Science Books: Herndon, VA, 2000.
- (22) Tembe, B. L.; Mccammon, J. A. *Comput. Chem.* **1984**, 8, 281.
- (23) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, 119, 5740.
- (24) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, 222, 529.
- (25) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, 114, 10235.
- (26) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, 105, 6474.
- (27) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577.
- (28) Boresch, S.; Steinhauser, O. *J. Chem. Phys.* **2001**, 115, 10793.
- (29) Kirkwood, J. G. *J. Chem. Phys.* **1935**, 3, 14.
- (30) Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Li, H. Y.; Verkhivker, G.; Keasar, C.; Zhang, J.; Ulitsky, A. *Comput. Phys. Commun.* **1995**, 91, 159.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926.
- (32) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
- (33) Straatsma, T. P.; Mccammon, J. A. *J. Chem. Phys.* **1991**, 95, 1175.
- (34) Papoulis, A. *Probability, Random Variables, and Stochastic Processes*; Third ed.; McGraw-Hill Inc.: New York, 1991.
- (35) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, 20, 849.
- (36) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, 122, 144107.
- (37) Bennett, C. H. *J. Comput. Phys.* **1976**, 22, 245.
- (38) Jas, G. S.; Kuczera, K. *Proteins* **2002**, 48, 257.