JCTC Journal of Chemical Theory and Computation

Accurate Estimates of Free Energy Changes in Charge Mutations

Brittany R. Morgan[†] and Francesca Massi*[,‡]

Department of Physics, Clark University, 950 Main Street, Worcester, Massachusetts 01610 and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts, 55 Lake Avenue North, Worcester, Massachusetts 01655

Received October 23, 2009

Abstract: The ability to determine the effect of charge changes on the free energy is necessary for fundamental studies of the electrostatic contribution to protein binding and stability. Currently, calculations of differences in free energy for charge mutations carried out in systems with periodic boundary conditions must include an approximate self-energy correction that can be on the same order of magnitude as the calculated free energy change. Here, a new method for accurately calculating the free energy change associated with any alchemical mutation, regardless of charge, is presented. In this method, paired mutations of opposite charge exactly cancel the self-energy term because of its quadratic charge dependence. Since the self-energy term implicitly cancels within the method, a correction never needs to be applied, and the statistical uncertainty associated is thereby removed. An implementation procedure is described and applied to the free energy of ionic hydration and a charged amino acid mutation.

I. Introduction

The favorability of a chemical reaction is determined by the net change in free energy. Experimentally, the detailed mechanism by which these changes occur is not always apparent. The calculation of free energy differences using molecular simulations is a powerful tool to explore the atomistic basis of changes in free energy.[1−3] Free energy methods can be used to determine properties such as the free energy of solvation[4−9] or to explain and predict relative affinities in protein−ligand binding.[10−16] The physical principles that affect the free energy of binding are crucial for rational drug design and understanding the fundamental mechanisms of selectivity and affinity in binding. The importance of understanding the source of free energy differences emphasizes the need for accurate and widely applicable free energy calculation methods.

In an experiment, the free energies of binding for proteins A and B (ΔG_1 and ΔG_2 from Figure 1) are measured to determine the free energy difference ($\Delta\Delta G$) of binding. The time required to compute $\Delta\Delta G$ from spontaneous binding events is prohibitively long, however. Instead, $\Delta\Delta G$ is calculated in free energy methods via the mutation of a protein A into B in the bound state and in the free state (ΔG_3 and ΔG_4 from Figure 1), where proteins A and B usually differ by a single amino acid.[1−3,17−20] This approach is possible due to the path independence of thermodynamic quantities, since

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

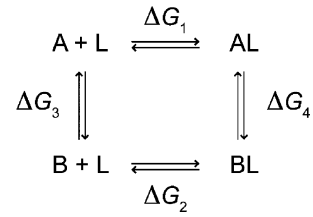

Figure 1. Example of a thermodynamic cycle. ΔG_1 is the free energy of protein A binding a ligand (L), and ΔG_2 is the free energy of protein B binding L. ΔG_3 is the free energy of mutating protein A into protein B in the free state, and ΔG_4 is the free energy of mutating protein A into protein B in the bound state.

* Corresponding author phone: 508-856-4501; fax: 508-856-6464; e-mail: Francesca.Massi@umassmed.edu.

[†] Clark University.

[‡] University of Massachusetts.

10.1021/ct900565e
Published on Web 05/17/2010

Current free energy methods are limited by the need for potentially large corrections to the free energy change when mutations change the overall system charge; such corrections arise from the treatment of electrostatic potentials.[21−26] The correction can be on the same order of magnitude as the calculated free energy change, which makes these methods unattractive for non-neutral mutations due to concerns about accuracy.

The ability to determine the effect of charge changes on the free energy is necessary for fundamental studies of the electrostatic contribution to protein binding and stability. The method proposed here eliminates the need to explicitly include a correction, which removes the uncertainty in estimating the correction. This advancement gives free energy calculations a similar level of accuracy for any mutation, regardless of change in charge.

II. Background

Free energy perturbation (FEP) and thermodynamic integration (TI)[1−3,19,20] are used in molecular dynamics (MD) or Monte Carlo (MC) simulations to determine free energy differences. The formalism for TI will be discussed, but the following arguments are equally applicable to FEP as well.

In TI, a scaling parameter λ is varied between 0 and 1, such that, given two states A and B, state A vanishes and state B grows as λ increases from 0 to 1. The free energy change of the mutation of state A into state B is calculated as

$$\Delta G = G_{\mathrm{B}} - G_{\mathrm{A}} \tag{1}$$

$$\Delta G = \int_0^1 \langle \Delta V \rangle_\lambda \, \mathrm{d}\lambda \cong \sum_i \langle \Delta V \rangle_{\lambda_i} \Delta\lambda \tag{2}$$

where the energy difference $\Delta V = V_{\mathrm{B}} - V_{\mathrm{A}}$ and λ satisfies $V_\lambda = (1 - \lambda)V_{\mathrm{A}} + \lambda V_{\mathrm{B}}$.[19] MD or MC is used to calculate the thermodynamic average $\langle e^{-(H_{\mathrm{B}} - H_\lambda)} \rangle_{\mathrm{A}}$ for each value of λ. The values of λ must be chosen carefully, to correctly evaluate $\langle \Delta V \rangle_\lambda$, as the accuracy of ΔG depends on the accuracy of the thermodynamic average $\langle \Delta V \rangle_\lambda$.[1−3]

The potential energy calculated in an MD or MC simulation has both bonded and nonbonded contributions, but the electrostatic potential is particularly difficult to model both accurately and efficiently.[21] The long-range nature of the electrostatic potential makes boundary conditions and system size important factors in free energy calculations.[27,28] Over the past 20 years, calculations of the solvation free energy for atomic ions using finite or periodic boundary conditions have produced accurate results, independent of system size.[4,28−35] An important difference between the use of finite versus periodic boundary conditions lies in the fact that, while finite boundaries, such as finite droplets or semifinite slab systems, have a liquid−vacuum interface, infinite periodic boundary conditions (tin-foil) do not. For this reason, calculations of hydration free energies of atomic ions performed with finite boundary conditions include the contribution of the surface potential of this interface where bulk solvent simulations performed using periodic boundary conditions do not.[25,27,36] This interface potential is an important contribution to the solvation free energy of charged solutes that arises when an ion crosses the liquid−vacuum boundary. Its contribution to the free energy of solvation is $q\phi_{l\to v}$, where q is the charge of the solute and $\phi_{l\to v}$ is the electrostatic potential jump due to the liquid−vacuum interface, which depends upon the solvent model used in the simulation.[37] Free energies obtained in the absence of the interface potential are also called intrinsic free energies. The electrostatic interface potential is essential for obtaining accurate estimates of the free energy of hydration for ionic molecules, and its contribution must be added to the intrinsic free energy of hydration obtained from simulations that use tin-foil boundary conditions.[25,36] The magnitude of the interface potential depends on the particular electrostatic summation methodology employed, for example, whether the P-sum (particle-based) or M-sum (molecule-based) convention is used.[25]

For free energy calculations where the difference in the free energy change, $\Delta\Delta G$, is the desired quantity (Figure 1), the free energy change associated with crossing the liquid−vacuum interface is the same for ΔG_3 and ΔG_4; because the charge and the solvent model used are the same, the contributions that the interface potential makes will cancel in $\Delta\Delta G$. For this reason, intrinsic free energy changes, which do not include the interface potential, can be used to evaluate $\Delta\Delta G$. Calculations of $\Delta\Delta G$ associated with the binding of two proteins, wild-type and mutant, to the same ligand (Figure 1) often use Ewald summation with tin-foil boundary conditions.[38−40]

The electrostatic potential in the Ewald summation is ill-defined if the system is not neutral. For charged solutes, the Ewald formulation implicitly neutralizes the charge with a homogeneous background charge.[41] This neutralizing plasma (NP) is a uniform charge distribution which is present everywhere in the system, including inside the solute. An alternative to the NP is to add explicit counterions to restore neutrality. Adding counterions is preferable because the NP is not an accurate physical representation of the charge density in solution.[21]

There is, however, an artifact introduced when using Ewald summation. Due to the periodicity of the Ewald implementation, charged atoms in the central cell have an electrostatic interaction with their periodic images, creating an artificial self-energy term which has a system size-dependent contribution to the free energy change.[21,22] In neutral systems, the contribution of the self-energy term to the free energy change is zero.

In TI or FEP, if states A and B have different charges, the self-energy term is different for each value of the scaling parameter, λ, during the mutation of state A into state B. For this reason, the self-energy term must then be accounted for in the final estimate of the free energy change. Hence, the intrinsic free energy change is equal to the calculated free energy change (ΔG^{calc}) plus a correction term,

$$\Delta G = \Delta G^{\mathrm{calc}} + \text{correction} \tag{3}$$

The Ewald summation with tinfoil boundaries has no interface and is always neutral; thus, there is no contribution from the interface potential in the free energy change

calculated by TI or FEP.[36] The intrinsic free energy change calculated by TI or FEP using tin-foil boundary conditions can be used to obtain an estimate of $\Delta\Delta G$ (Figure 1), noting that if the calculation involves a charged species, the interface potential has the same contribution to ΔG_3 and ΔG_4. However, if we are interested in accurately determining the solvation free energy of a single ion, the contribution of the interface potential must be added to the intrinsic free energy change.[25,36] The interface potential is a physical quantity present in the experimental solvation free energies of charged solutes which must be added (in the case of Ewald summation with tin-foil boundaries) to the intrinsic free energy for accurate calculations of solvation free energies.

In contrast, the self-energy is an artifact of the infinite periodicity in the Ewald implementation, which is always present in non-neutral systems.[21,22] For accurate estimates of any intrinsic free energy change for charge mutations, the self-energy must be removed from the free energy change calculated by TI or FEP (eq 3). Self-energy artifacts in free energy calculations arising from mutations of charged solutes during TI or FEP have been well studied by calculating the free energy of hydration of atomic ions.[23,24,31,32,42−45]

A correction has been derived for the self-energy contribution to the free energy change when using Ewald summation with periodic boundary conditions.[41] The electrostatic potential (U_{Coul}) of an infinitely periodic system is defined as

$$U_{\mathrm{Coul}} = \sum_{1\le i\le j\le N} q_i q_j \phi_{\mathrm{EW}}(\mathbf{r}_{ij}) + \frac{1}{2}\sum_{1\le i\le N} q_i^2 \xi_{\mathrm{EW}} \qquad (4)$$

where r_{ij} is the distance between atoms i and j with charges q_i and q_j, N is the total number of atoms, and $\phi_{\mathrm{EW}}(\mathbf{r})$ is the position-dependence of the electrostatic potential. ξ_{EW} is the self-energy term defined by the lattice summation used in Ewald summation,[46−48] which has the form

$$\xi_{\mathrm{EW}} = \lim_{r\to 0}\left(\phi_{\mathrm{EW}}(\mathbf{r}) - \frac{1}{r}\right) \qquad (5)$$

ξ_{EW} is the electrostatic potential from the periodic images and the neutralizing background charge in a Wigner lattice.[46−48] From eq 4, the energy difference between an initial state with charge q_0 and a final state with charge q_1 and can be calculated as

$$\Delta U_{\mathrm{Coul}} = \Delta q \phi_{\mathrm{EW}}(\mathbf{r}) + \frac{1}{2}\xi_{\mathrm{EW}}(q_1^2 - q_0^2), \quad \text{where } \xi_{\mathrm{EW}} = \frac{-2.837}{4\pi\varepsilon_{\mathrm{o}} L} \qquad (6)$$

for Ewald summations in a cubic lattice with side length L and $\Delta q = q_1 - q_0$.[47,48]

III. Previous Work

The second term of eq 6 is the self-energy correction (C_{EW}) which is added to the intrinsic free energy change calculated via TI or FEP (ΔG^{calc}) in eq 3, such that

$$\Delta G = \Delta G^{\mathrm{calc}} + C_{\mathrm{EW}} \qquad (7)$$

Ideally, C_{EW} is equal to and opposite of the true self-energy (E_{self}), but this approximation may not be accurate. Charge mutations are problematic even for single atom mutations,[25,26] and there are far fewer examples of charge mutations[10−12,49] than neutral mutations in biomolecular simulations. While charge mutations of atomic ions have been well studied, little has been done to advance the application to more complex systems.

The approach often used for charge mutations is to avoid the self-energy problem by performing two simultaneous mutations which render the system neutral in both initial and final states.[10,12] For neutral systems, the self-energy is zero, and no correction is needed ($\Delta G = \Delta G^{\mathrm{calc}}$). Either a residue far from the area of interest or a neutral dummy atom in the system can be mutated to compensate for the change in charge. Using this approach, the free energy change of the dual mutation is calculated ($\Delta G_1 = \Delta G_1^{\mathrm{calc}}$); hence the free energy change of the counter mutation must be known in order to obtain the free energy change of the charge mutation. Assuming the mutations are independent, ΔG_1 can be written as

$$\Delta G_1 = \Delta G_{\mathrm{A}} + \Delta G_{\mathrm{B}} \qquad (8)$$

where ΔG_{A} is the free energy change of the charge mutation and ΔG_{B} is the free energy change of the counter mutation; ΔG_{A} is the quantity of interest.

To find ΔG_{A}, we must calculate ΔG_{B} separately, but the counter mutation is also a charge mutation, and the same concerns regarding the self-energy apply in determining its free energy change. However, the self-energy is dependent on system size as well as charge (eq 6). If the system size is sufficiently large such that the self-interactions are minimal, then the self-energy vanishes and $\Delta G_{\mathrm{B}} = \Delta G_{\mathrm{B}}^{\mathrm{calc}}$, but at what point the system size is sufficiently large can be difficult to quantify a priori.

To determine whether the self-interactions are minimal, the neutral dummy atom can be mutated into an atomic ion for the counter mutation (mutation B), and ΔG_{B} can be compared to experimental free energies of hydration. As these free energy changes can be on the order of 100 kcal/mol, accuracy is an important concern. In order to reproduce experimental values, self-energy corrections to ΔG_{B} dependent on the ionic charge, ionic radius, boundary conditions, system size, electrostatic scheme, and solvent model are needed,[25,26] and the liquid−vacuum interface potential must also be included.[25,26,36] The solvation free energy can be accurately calculated for atomic ions when these corrections are included.[21−26] However, the free energy of ionic hydration cannot be measured directly, and extrathermodynamic assumptions[37] introduce inconsistencies in the experimental values,[50−56] which makes it difficult to assess the accuracy of ΔG_{B}.

Another serious concern exists with no current solution. For the self-energy to vanish, the charge mutation and the counter mutation must be performed simultaneously to preserve electrostatic neutrality. It is assumed that the free energy changes of the mutations are independent (eq 8). If they are not independent, then the free energy change of the dual mutation becomes

$$\Delta G_1^* = \Delta G_A + \Delta G_B + \Delta G_{AB} \quad (9)$$

where ΔG_{AB} is the free energy change due to the interaction, and it is impossible to find ΔG_A even with an accurate estimate of ΔG_B.

The method of dual mutations is not widely used because of concerns regarding these assumptions. In this work, we present a method which also uses dual mutations to remove the self-energy but does not suffer from these limitations. In section IV, we present the theoretical rationale and three cases for which the self-energy correction is zero. These cases can be used to accurately find the free energy change of both the charge mutation and the counter mutation and, additionally, confirm the independence of the dual mutations. The assumption of a sufficiently large system size is not needed; in fact, the method presented in this work can directly quantify at what point the system size becomes sufficiently large to make the self-energy negligible. More importantly, this method is directly applicable to complex systems, and independence can be verified in any system. It is not limited to the basic examples used to illustrate the method in section V.C. A detailed procedure for applying the method is presented in section V.A and demonstrated by two examples in section V.C.

IV. Theoretical Basis

The self-energy correction (C_{EW}) from eq 6 can be rewritten as

$$C_{EW} = \frac{1}{2}\xi_{EW}(q_1^2 - q_0^2) = \frac{1}{2}\xi_{EW}(q_1 - q_0)(q_1 + q_0) \quad (10)$$

In this form, it becomes obvious that $C_{EW} = 0$ not only for $(q_0,q_1) = (0,0)$, but also for $(a,-a)$ and (a,a). Therefore, for any mutation where $|q_1| = |q_0|$, the self-energy correction C_{EW} vanishes. In the remainder of this section, we will discuss explicitly how C_{EW} depends on q_0 and q_1.

Case 1A. Consider the case when $(q_0,q_1) = (0, 0)$ as the mutation $A^0 \rightarrow B^0$, where initial state A^{q_0}, with charge $q_0 = 0$, mutates into final state B^{q_1}, with charge $q_1 = 0$. Substitution of q_0 and q_1 into eq 10 yields the expected result of

$$C_{EW} = \frac{1}{2}\xi_{EW}(q_1^2 - q_0^2) = \frac{1}{2}\xi_{EW}((0)^2 - (0)^2) = 0 \quad (11)$$

Case 1B. The dual mutation $A^0 + B^0 \rightarrow A^+ + B^-$ also satisfies $(q_0,q_1) = (0,0)$; hence, $C_{EW} = 0$. If these mutations are separated into two single mutations, (1) $A^0 \rightarrow A^+$ and (2) $B^0 \rightarrow B^-$, with $C_{EW} = C_{EW}^{(1)} + C_{EW}^{(2)}$, then

$$C_{EW}^{(1)} = \frac{1}{2}\xi_{EW}((1)^2 - (0)^2) = \frac{1}{2}\xi_{EW} \quad (12)$$

and

$$C_{EW}^{(2)} = \frac{1}{2}\xi_{EW}((-1)^2 - (0)^2) = \frac{1}{2}\xi_{EW} \quad (13)$$

C_{EW} is no longer equal to zero if the two mutations are not performed simultaneously.

Case 2A. As an example of the case $(q_0,q_1) = (a,-a)$, consider the mutation $A^+ \rightarrow B^-$. Substitution of $q_0 = +1$ and $q_1 = -1$ into eq 10 yields

$$C_{EW} = \frac{1}{2}\xi_{EW}(q_1^2 - q_0^2) = \frac{1}{2}\xi_{EW}((-1)^2 - (1)^2) = \frac{1}{2}\xi_{EW}(1 - 1) = 0 \quad (14)$$

Thus, the Ewald implementation considers the self-energy term of the mutation $A^+ \rightarrow B^-$ equivalent to $A^0 \rightarrow B^0$.

Thermodynamic cycles are path-independent, which makes it possible to decompose the mutation $A^+ \rightarrow B^-$ into two serial mutations, (1) $A^+ \rightarrow N^0$ and (2) $N^0 \rightarrow B^-$, where N^0 is some neutral species. For the first mutation

$$C_{EW}^{(1)} = \frac{1}{2}\xi_{EW}((0)^2 - (1)^2) = \frac{1}{2}\xi_{EW} \quad (15)$$

and for the second mutation

$$C_{EW}^{(2)} = \frac{1}{2}\xi_{EW}((-1)^2 - (0)^2) = -\frac{1}{2}\xi_{EW} \quad (16)$$

$C_{EW}^{(1)}$ is equal and opposite to $C_{EW}^{(2)}$, demonstrating that $C_{EW} = C_{EW}^{(1)} + C_{EW}^{(2)} = 0$ as well after decomposition.

Case 2B. The dual mutation $A^+ + B^0 \rightarrow A^0 + B^-$ has the same charge at the initial and final states as the mutation $A^+ \rightarrow B^-$, which was shown to have $C_{EW} = 0$ in eq 14. The mutation $A^+ + B^0 \rightarrow A^0 + B^-$ can be separated into two single mutations (1) $A^+ \rightarrow A^0$ and (2) $B^0 \rightarrow B^-$, as in case 1B, but here $C_{EW} = C_{EW}^{(1)} + C_{EW}^{(2)} = 0$ as in eqs 15 and 16.

Case 3A. The mutation $A^+ \rightarrow B^+$ corresponds to the case $(q_0,q_1) = (a,a)$. Substituting into eq 7,

$$C_{EW} = \frac{1}{2}\xi_{EW}(q_1^2 - q_0^2) = \frac{1}{2}\xi_{EW}((1)^2 - (1)^2) = 0 \quad (17)$$

Case 3B. Separation of $A^+ + B^0 \rightarrow A^0 + B^+$ into (1) $A^+ \rightarrow A^0$ and (2) $B^0 \rightarrow B^+$ results in the same cancellation of $C_{EW}^{(1)}$ and $C_{EW}^{(2)}$ as in case 2B.

Cancellation of C_{EW} occurs for cases 2B and 3B, but not 1B, which leads to the conclusion that the *direction* of the change in charge is the basis of the cancellation, while the sign of the individual charges is irrelevant. This property is akin to the ionic strength, as both depend on the magnitude of the charge in the system, not the sign.

The cancellation of C_{EW} results from the quadratic dependence of C_{EW}, which arises because the Coulomb potential is quadratic in the charge. Any energy derived from a point-charge Coulomb potential will be quadratic in the charge, including the self-energy; hence, the self-energy will also cancel according to the preceding argument, provided that other dependencies in the self-energy have a negligible effect compared to the charge dependence.[25] This cancellation does not occur, however, for the interface potential (discussed in section II), which is linear in the charge.

The problems inherent to charge mutations are well understood for atomic ions.[21−26,36,57] Unfortunately, the translation of these finding to more complex systems is not straightforward. The work on atomic ions shows dependencies in the free energy change on ionic charge, ionic radius, boundary conditions, system size, electrostatic scheme, and the solvent model.[25] The method presented here addresses the charge dependence, but self-energy corrections for infinitely periodic systems still include a dependence on the

solute radius.[21,26,58] We will demonstrate in section V.C that the dependence on solute radius is weak and does not reduce the applicability of this method. The effect of system size will also be discussed.

V. Application to Charge Mutations in Proteins

The cancellation demonstrated for C_{EW} in section IV, which applies to the self-energy (E_{self}) as well, can be used to accurately find the free energy change of both the charge mutation and the counter mutation and, additionally, confirm the independence of the mutations. We will use FEP to outline and demonstrate the method, using the notation $\Delta G^{FEP} = \Delta G^{calc}$.

Consider the change in free energy for protein P_A which begins neutral and becomes protein P_B with a charge of -1 through the mutation of a residue ($P_A^0 \rightarrow P_B^-$). Concurrently performing the mutation $N^0 \rightarrow K^+$, where N^0 is a neutral dummy atom and K^+ is a potassium ion, renders the system charge neutral continuously from the initial to the final state, as in section III. This is described by the dual mutation

$$P_A^0 + N^0 \rightarrow P_B^- + K^+ \quad (18)$$

with free energy change ΔG_1.

As shown in cases 1A and 1B, $C_{EW} = 0$ only when the two mutations are performed simultaneously, and thus $\Delta G_1 = \Delta G_1^{FEP}$(eq 7). Assuming that the two concurrent mutations are independent, the calculated free energy change of the dual mutation in eq 18 (ΔG_1^{FEP}) is the combined free energy change of the two single mutations without a contribution from E_{self} such that $\Delta G_1^{FEP} = \Delta G_{P_A^0 \rightarrow P_B^-} + \Delta G_{N^0 \rightarrow K^+}$ (eq 8), where $\Delta G_{P_A^0 \rightarrow P_B^-}$ is the free energy change of the mutation $P_A^0 \rightarrow P_B^-$ and $\Delta G_{N^0 \rightarrow K^+}$ is the free energy change of the mutation $N^0 \rightarrow K^+$. From this calculation, we only obtain the sum of $\Delta G_{P_A^0 \rightarrow P_B^-}$ and $\Delta G_{N^0 \rightarrow K^+}$.

However, we can avoid the problems outlined in section III by additionally performing the mutation

$$P_A^0 + K^+ \rightarrow P_B^- + N^0 \quad (19)$$

with free energy change ΔG_2, which results in $C_{EW} = 0$ whether performed concurrently in the same FEP or separately as shown in cases 2A and 2B. The direction of the mutation of the potassium ion is reversed ($K^+ \rightarrow N^0$), which has the opposite free energy change ($\Delta G_{K^+ \rightarrow N^0}^{FEP} = -\Delta G_{N^0 \rightarrow K^+}^{FEP}$). The advantage is that we can calculate ΔG_2^{FEP} of the dual mutation in eq 19, which is also free of E_{self} (i.e., $\Delta G_2 = \Delta G_2^{FEP}$), but in this case, $\Delta G_2^{FEP} = \Delta G_{P_A^0 \rightarrow P_B^-} - \Delta G_{N^0 \rightarrow K^+}$. Combining ΔG_1^{FEP} and ΔG_2^{FEP}, we find the individual free energy changes $\Delta G_{P_A^0 \rightarrow P_B^-}$ and $\Delta G_{N^0 \rightarrow K^+}$ as

$$\Delta G_{P_A^0 \rightarrow P_B^-} = \frac{\Delta G_1^{FEP} + \Delta G_2^{FEP}}{2} \text{ and}$$

$$\Delta G_{N^0 \rightarrow K^+} = \frac{\Delta G_1^{FEP} - \Delta G_2^{FEP}}{2} \quad (20)$$

As already discussed in case 2B, the second mutation (eq 16) can also be performed as two separate mutations, (1) $P_A^0 \rightarrow P_B^-$ and (2) $K^+ \rightarrow N^0$, to give $\Delta G_{P_A^0 \rightarrow P_B^-}^{FEP}$ and $-\Delta G_{N^0 \rightarrow K^+}^{FEP}$; we can add $\Delta G_{P_A^0 \rightarrow P_B^-}^{FEP}$ and $-\Delta G_{N^0 \rightarrow K^+}^{FEP}$ to obtain ΔG_2. If the two mutations are independent, this ΔG_2 should be equal to ΔG_2^{FEP} calculated from the FEP of eq 19.

By the method presented here, we can obtain $\Delta G_{P_A^0 \rightarrow P_B^-}$ without using a self-energy correction or needing to calculate $\Delta G_{N^0 \rightarrow K^+}$ separately, as has previously been necessary. The self-energy is a simulation artifact that needs to be removed. We do this by performing the free energy calculation on a modified system that includes an additional ion, as illustrated in eqs 18 and 19, to cancel the self-energy. The undesired quantity ($\Delta G_{N^0 \rightarrow K^+}$) can be determined and removed from the calculated ΔG (as shown in eq 20), whereas the self-energy term associated with the calculated free energy for the mutation of the charged species of interest alone could not. In this method, only self-consistency is necessary for the atomic ion mutations; i.e., force field, solvent model, electrostatic convention, and boundary conditions should be identical.

The method uses the equality $\Delta G_{K^+ \rightarrow N^0}^{FEP} = -\Delta G_{N^0 \rightarrow K^+}^{FEP}$, which is true if the hysteresis of the forward and reverse FEP is negligible;[59] this is a reasonable assumption for atomic ion mutations, as used here, and can be easily verified. Moreover, the crucial assumption of independence of the two mutations can now be confirmed, as detailed below.

In order to guarantee the independence of the two concurrent mutations, it is necessary to position the counterion at an adequate distance from the protein in order to minimize the electrostatic interaction between the protein and the ion. The Bjerrum length is defined as the distance at which the Coulomb interaction between two monovalent ions in a uniform dielectric is equal to the thermal energy, k_BT.[60] In water at 300 K, the Bjerrum length is ~7 Å. This distance can be used to approximately determine how far apart the protein and the counterion should be placed in the simulation. For example, for free energy calculations of proteins in water at 300 K, a distance of ~7 Å or greater between the surface of the protein and the counterion will be adequate to ensure that the electrostatic interaction between the protein and the ion is effectively screened.

A. Procedure to Calculate the Free Energy Change of Any Charge Mutation. In the example above, the FEP of the protein mutation was performed three times: as the mutations $P_A^0 + N^0 \rightarrow P_B^- + K^+$, $P_A^0 + K^+ \rightarrow P_B^- + N^0$, and $P_A^0 \rightarrow P_B^-$. To reduce the computational demand of this method, we propose the following generalized procedure to calculate the free energy change for any charge mutation.

Step 1: Determine ΔG of a counter mutation such as an atomic ion (K^+, in this example).

Step 2: Combine $\Delta G_{N^0 \rightarrow K^+}$ with $\Delta G_2^{FEP} = \Delta G_{P_A^0 \rightarrow P_B^-} - \Delta G_{N^0 \rightarrow K^+}$ of the dual mutation $P_A^0 + K^+ \rightarrow P_B^- + N^0$ to find $\Delta G_{P_A^0 \rightarrow P_B^-}$.

Step 3: Use the single mutations $P_A^0 \rightarrow P_B^-$ and $K^+ \rightarrow N^0$ to confirm the independence of the mutations.

Although only the first two steps are necessary to find the free energy change, we perform all three steps to check the accuracy of the method and the independence of the two mutations.

Table 1. Free Energy Change of the Mutations $N^0 \rightarrow Cl^-$ and $N^0 \rightarrow K^+$

step[a]		ΔG^{FEP}	mutation
1−1: $\Delta G_1^{FEP} = \Delta G_{N^0\rightarrow Cl^-} + \Delta G_{N^0\rightarrow K^+}$		−162.8 ± 0.7	$N^0 + N^0 \rightarrow Cl^- + K^+$
1−2: $\Delta G_2^{FEP} = \Delta G_{N^0\rightarrow Cl^-} - \Delta G_{N^0\rightarrow K^+}$		−21.5 ± 0.6	$N^0 + K^+ \rightarrow Cl^- + N^0$
3−3: $\Delta G_{N^0\rightarrow Cl^-}^{FEP} - \Delta G_{N^0\rightarrow K^+}^{FEP}$		−21.3 ± 0.4	$N^0 \rightarrow Cl^-$ $K^+ \rightarrow N^0$
$\Delta G_{N^0\rightarrow Cl^-}^{FEP}$ [b]	−91.9 ± 0.3	$\Delta G_{N^0\rightarrow K^+}^{FEP}$ [b]	−70.6 ± 0.1
$\Delta G_{N^0\rightarrow Cl^-}$ [c]	−92.2 ± 1.3	$\Delta G_{N^0\rightarrow K^+}$ [c]	−70.7 ± 1.3

[a] Mean and error of free energy change from FEP calculations (ΔG^{FEP}) in kcal/mol. The mutations used to obtain each ΔG^{FEP} are listed, see section V for details. The values of the independent mutations used to calculate the free energy change for step 3−3 are reported in the next line (b). [b] Mean and error of the hydration free energy computed directly using FEP on the individual ions. [c] Mean and error of final free energy change (ΔG) in kcal/mol for the mutation of a neutral dummy atom (N^0) into chloride (Cl^-) and potassium (K^+) ions calculated from steps 1−1 and 1−2 according to eq 20.

Step 1: Calculate the Free Energy Change of Counter Mutations. The free energy change of the counter mutation is a constant value and only needs to be determined once for a given water model.[61] Atomic ions are convenient for the counter mutation because the FEP can be performed in a short time compared to the mutation of a residue in a solvated protein. The system size used in the simulations should be sufficiently large to minimize interactions between the mutating ions, which is verified in step 3.

As an example, we will calculate the free energy change of chloride and potassium ions associated with the mutations $N^0 \rightarrow Cl^-$ ($\Delta G_{N^0\rightarrow Cl^-}$) and $N^0 \rightarrow K^+$ ($\Delta G_{N^0\rightarrow K^+}$), using the following three step strategy. In each step, E_{self} is identically zero. The numerical results are presented in Table 1.

1−1. Calculate $\Delta G_1^{FEP} = \Delta G_{N^0\rightarrow Cl^-} + \Delta G_{N^0\rightarrow K^+}$ of the mutation $N^0 + N^0 \rightarrow Cl^- + K^+$.

1−2. Calculate $\Delta G_2^{FEP} = \Delta G_{N^0\rightarrow Cl^-} - \Delta G_{N^0\rightarrow K^+}$ of the mutation $N^0 + K^+ \rightarrow Cl^- + N^0$.

1−3. The free energy changes $\Delta G_{N^0\rightarrow Cl^-}$ and $\Delta G_{N^0\rightarrow K^+}$ can be determined from ΔG_1^{FEP} and ΔG_2^{FEP} as in eq 20.

Step 2: Calculate the Free Energy Change of Any Mutation. The free energy change of the mutation $P_A^0 \rightarrow P_B^-$ ($\Delta G_{P_A^0\rightarrow P_B^-}$) can now be calculated using $\Delta G_{N^0\rightarrow K^+}$ from step 1−3. The mutation $P_A^0 + K^+ \rightarrow P_B^- + N^0$ is used here because, while it would be equally valid mathematically to use the mutation $P_A^0 + N^0 \rightarrow P_B^- + K^+$, the first mutation reduces the interactions between the two charged species.

2−1. Calculate $\Delta G_2^{FEP} = \Delta G_{P_A^0\rightarrow P_B^-} - \Delta G_{N^0\rightarrow K^+}$ of the dual mutation $P_A^0 + K^+ \rightarrow P_B^- + N^0$.

2−2. Use ΔG_2^{FEP} and $\Delta G_{N^0\rightarrow K^+}$ (from step 1−3) to obtain $\Delta G_{P_A^0\rightarrow P_B^-}$ using the relationship $\Delta G_{P_A^0\rightarrow P_B^-} = \Delta G_2^{FEP} + \Delta G_{N^0\rightarrow K^+}$.

To find the free energy change for a positive protein mutation, use Cl^- instead of K^+. No special techniques are needed to calculate the free energy change of a continuous $P_A^- \rightarrow P_B^+$ mutation, since the self-energy exactly cancels during the mutation (as in case 2A).

Step 3: Confirm Independence of Concurrent Mutations. We need to verify that, for the dual mutation $P_A^0 + K^+ \rightarrow P_B^- + N^0$, $\Delta G_2^{FEP} = \Delta G_{P_A^0\rightarrow P_B^-} - \Delta G_{N^0\rightarrow K^+}$ with no additional contribution from interactions between the two mutations. To do this, we will use the FEP of (1) $P_A^0 \rightarrow P_B^-$ and (2) $K^+ \rightarrow N^0$ in two separate simulations of the same system size (ξ_{EW} depends on L, eq 6). Steps 3−1 and 3−2 only need to be performed once per mutation because they are independent.

3−1. Calculate $\Delta G_{P_A^0\rightarrow P_B^-}^{FEP} = \Delta G_{P_A^0\rightarrow P_B^-} + E_{self}$ of the single mutation $P_A^0 \rightarrow P_B^-$.

3−2. Calculate $\Delta G_{N^0\rightarrow K^+}^{FEP} = \Delta G_{N^0\rightarrow K^+} + E_{self}$ of the single mutation $N^0 \rightarrow K^+$.

3−3. If $\Delta G_{P_A^0\rightarrow P_B^-}^{FEP} - \Delta G_{N^0\rightarrow K^+} = \Delta G_2^{FEP}$ (from step 2−1), then the self-energy terms correctly cancel, and the two mutations are independent.

Furthermore, we can determine E_{self} using $\Delta G_{N^0\rightarrow K^+}$ from step 1−3, since $E_{self} = \Delta G_{N^0\rightarrow K^+}^{FEP} - \Delta G_{N^0\rightarrow K^+}$ if the mutations used to determine $\Delta G_{N^0\rightarrow K^+}$ are independent. Knowledge of E_{self} can provide a metric to assess the amount of self-interactions in the simulation, as discussed below.

B. Advantages of the Approach

1. Increased Accuracy of the Free Energy Calculation. The self-energy is implicitly canceled in the method presented here, which removes the need for a correction. The self-energy correction is no longer a crucial component in obtaining an accurate estimate of the free energy change for charge mutations.

2. Knowledge of the Self-Energy. E_{self} is a measure of the self-interactions between periodic images which depends on system size. For accurate MD simulations using PME, the system size should be sufficiently large so as to minimize these interactions, but what constitutes "sufficiently large" for long-range electrostatic interactions is not well-defined. E_{self} could be used as a quantitative metric to evaluate the system size.

3. Adaptable to Changes of Larger Magnitude. In the examples, the change in charge is limited to ±1. While it is possible to perform a mutation such as $A^{2-} \rightarrow A^{2+}$ directly, as the extent of the change between the initial and final states increases, it becomes more difficult to obtain an accurate thermodynamic average for each value of λ (eq 2).[1] Decreasing the incremental change in λ can provide greater accuracy; however, the charge mutation can also be divided into smaller steps. A mutation $A^{q-} \rightarrow A^{q+}$ can be decomposed into N steps, where $2q/N$ can be less than 1 if fractional charge increments are used. As long as the incremental charge changes are identical in each direction, the self-energy still vanishes in the final free energy. Any staging method that can be applied to λ to increase the accuracy of the calculation[62] can be applied to q as well, but as q is not necessarily linear in λ, it may be easier to apply an efficient staging scheme as a function of q rather than λ.

4. Identity of the Counter Mutation Is Irrelevant. Since the contribution of the counter mutation is removed, the

Table 2. Free Energy Change of the Mutations $D^- \rightarrow S^0$ and $N^0 \rightarrow K^+$

step[a]		ΔG^{FEP}	mutation
1−1: $\Delta G_1^{FEP} = \Delta G^{FEP}_{D^-\rightarrow S^0} - \Delta G^{FEP}_{N^0\rightarrow K^+}$		82.5 ± 1.3	$D^- + N^0 \rightarrow S^0 + K^+$
1−2: $\Delta G_2^{FEP} = \Delta G^{FEP}_{D^-\rightarrow S^0} + \Delta G^{FEP}_{N^0\rightarrow K^+}$		224.2 ± 0.2	$D^- + K^+ \rightarrow S^0 + N^0$
3−3: $\Delta G^{FEP}_{D^-\rightarrow S^0} - \Delta G^{FEP}_{N^0\rightarrow K^+}$		81.8 ± 0.7	$D^- \rightarrow S^0$ $N^0 \rightarrow K^+$
$\Delta G_{D^-\rightarrow S^0}$ [b]	153.4 ± 1.5	$\Delta G_{N^0\rightarrow K^+}$ [b]	−70.9 ± 1.5

[a] Mean and error of free energy change from FEP calculations (ΔG^{FEP}) in kcal/mol. The mutations used to obtain each ΔG^{FEP} are listed, see section V for details. [b] Mean and error of final free energy change (ΔG) in kcal/mol for the mutation of aspartic acid (D^-) into serine (S^0) and a neutral dummy atom (N^0) into a potassium ion (K^+).

Table 3. Self-Energy and Calculated Corrections Terms (kcal/mol)[a]

L (Å)	ΔG^{FEP}	E_{self}	C_{EW}[b]	$\Delta\Delta G_{solv}$[c]
		K^+		
29.16	−70.6 ± 0.1	0.1 ± 1.4	−16.2	−0.6
18.93	−69.9 ± 0.2	0.9 ± 1.5	−24.9	−1.6
11.13	−65.4 ± 0.2	5.3 ± 1.5	−42.3	−6.6
		Cl^-		
29.11	−91.9 ± 0.3	0.3 ± 1.6	−16.2	−0.8
18.97	−90.8 ± 0.1	1.4 ± 1.4	−24.8	−2.4
10.78	−88.0 ± 0.2	4.2 ± 1.5	−43.7	−11.2

[a] Mean and errors for the free energy calculated via FEP (ΔG^{FEP}) and the self energy (E_{self}) for the mutation of a neutral dummy atom (N^0) into a potassium (K^+) or chloride (Cl^-) ion for three system sizes (L). The error of E_{self} is the error of ΔG^{FEP} and ΔG from Table 1. [b] C_{EW} is dependent on system size and charge (eq 10).[41] [c] $\Delta\Delta G_{solv}$ is dependent on system size, charge, and solute radius (eq 21).[21]

accuracy of the estimated free energy change of the counter mutation does not affect the overall accuracy of the method. Moreover, unphysical mutations, such as fractional charges or artificial ions, can be used.

C. Demonstration of the Method. The method described above is illustrated here by several examples. FEP calculations were done with the molecular dynamics package NAMD 2.7[63] using the CHARMM 27[64] force field and explicitly solvated with the TIP3P water model. Each FEP was repeated five times to give the results in Tables 1−3. All simulations use PME with tin-foil boundary conditions. The free energy changes reported below are the intrinsic free energy change. The particular choice of boundary conditions affects the results, as discussed in section II. While the free energy change associated with the dual mutation reaction illustrated in eq 18 has no contribution from the surface potential term, the free energy change associated with the reaction illustrated in eq 19 does. If single ion solvation free energies are the goal of the calculation, then the contribution of the interface potential jump must be included.[26,36,57]

1. Free Energy Change of Atomic Ion Mutations. The free energy changes for $N^0 \rightarrow Cl^-$ and $N^0 \rightarrow K^+$ were calculated according to the procedure in step 1. N^0 is a neutral dummy atom with the following Lennard-Jones parameters: $\varepsilon = 0$ and $\sigma = 0$. As a practical matter, before charging and alchemical mutation to the target species, the N^0 state was transformed into an argon-like particle with the CHARMM 27[64] Lennard-Jones parameters for sodium (and zero charge). The free energy associated with insertion of this argon-like particle was 2.0 ± 0.2 kcal/mol. For concurrent dual mutations, the two ions were fixed in the simulated water box at a distance $L/2$ apart in x, y, and z (L = 30 Å). When performing single mutations, the ion was fixed at the center of the simulation box (L = 30 Å). The results from each step are listed in Table 1.

The free energy changes obtained from the concurrent dual mutations of steps 1−1 and 1−2 using eq 20 are $\Delta G_{N^0\rightarrow K^+} = -70.7 \pm 1.3$ kcal/mol and $\Delta G_{N^0\rightarrow Cl^-} = -92.2 \pm 1.3$ kcal/mol (Table 1, step 1−3). $\Delta G_2^{FEP} = -21.5 \pm 0.6$ kcal/mol from the concurrent dual mutations (Table 1, step 1−2), which agrees well with $\Delta G^{FEP}_{N^0\rightarrow Cl^-} - \Delta G^{FEP}_{N^0\rightarrow K^+} = -21.3 \pm 0.4$ kcal/mol from the hydration free energies computed directly using FEP on the individual ions (Table 1, step 3−3), confirming that the mutations are independent.

The hydration free energy of the neutral salt, computed from concurrent dual mutations, $\Delta G_1^{FEP} = \Delta G_{N^0\rightarrow Cl^-} + \Delta G_{N^0\rightarrow K^+} = -162.8 \pm 0.7$ kcal/mol is in excellent agreement with that computed using separate FEP computations, $\Delta G_1^{FEP} = \Delta G^{FEP}_{N^0\rightarrow Cl^-} + \Delta G^{FEP}_{N^0\rightarrow K^+} = -162.5 \pm 0.4$ kcal/mol. These hydration free energies agree with the results for $\Delta G_{N^0\rightarrow K^+}$ and $\Delta G_{N^0\rightarrow Cl^-}$ from previous work using the CHARMM 27 force field and the TIP3P water model.[42,57,61] The good agreement between the computed and experimental values for the hydration free energy of the neutral salt (potassium chloride) of −160.4 kcal/mol by Tissandier et al.[53] (corrected for the typographical errors detailed by Kelly et al.[65]) is a measure of the quality of, and compromises inherent in, the ion parametrization with the TIP3P model; any discrepancy can be attributed to the limitations of the model and/or parametrization. Although the concurrent dual mutation method is intended for applications where theoretical treatments of the self-energy correction are not as easily applied, such as mutations in biomolecules, this exercise serves as proof-of-concept.

2. Effect of Radius Dependence on Free Energy Change. Analytical self-energy corrections for infinitely periodic systems also include a dependence on solvent permittivity and solute radius.[21,26,58] The radius-dependent correction is intended to account for finite solute size and finite solvent permittivity in a periodic system.[21,58] It is defined as

$$\Delta\Delta G_{solv} = -\frac{1}{2}\frac{q^2}{4\pi\varepsilon_0 L}(\varepsilon_i^{-1} - \varepsilon_s^{-1})\left\{\xi'_{EW} + \frac{4\pi}{3}\left(\frac{R}{L}\right)^2 - \frac{16\pi^2}{45}\left(\frac{R}{L}\right)^5\right\} \text{ for } R \leq \frac{1}{2}L \quad (21)$$

Given $\xi'_{EW} = -2.837$ and $\varepsilon_i = 1$ (internal permittivity), in the limit of $\varepsilon_s \rightarrow \infty$ (solvent permittivity) and $R \rightarrow 0$ (hard-sphere radius), $\Delta\Delta G_{solv} \rightarrow C_{EW}$. The radius for atomic ions is defined by the excluded volume, which can be approximated as the Lennard-Jones radius or calculated empiri-

cally from radial distribution functions.[57] Solvent permittivity is a constant for a given water model.

The radius dependence in eq 21 could pose an issue for the cancellation of E_{self} in cases where the sizes of the two mutations are dissimilar, such as an atomic ion and a residue within a protein. As a test, step 1 was applied to the mutation $D^- + N^0 \rightarrow S^0 + K^+$, where the mutation of aspartic acid (D, $q = -1$) to serine (S, $q = 0$) is denoted as $D^- \rightarrow S^0$ with free energy change $\Delta G_{D^- \rightarrow S^0}$. Note that the direction of the change in charge is reversed for the negative species from the example ($N^0 \rightarrow Cl^-$), which changes the procedure slightly. The mutation $N^0 \rightarrow K^+$ has the free energy change $\Delta G_{N^0 \rightarrow K^+}$, as defined previously. The zwitterionic forms of D^- and S^0 were used, and only the position of K^+ was fixed in the simulation box ($L = 40$ Å). The results of each step are listed in Table 2.

We find $\Delta G_{D^- \rightarrow S^0} = 153.4 \pm 1.5$ kcal/mol and $\Delta G_{N^0 \rightarrow K^+} = -70.9 \pm 1.5$ kcal/mol (Table 2, step 1−3). $\Delta G_1^{FEP} = 82.5 \pm 1.3$ kcal/mol of the concurrent dual mutation (Table 2, step 1−1) is in good agreement with $\Delta G_{D^- \rightarrow S^0}^{FEP} + \Delta G_{N^0 \rightarrow K^+}^{FEP} = 81.8 \pm 0.7$ kcal/mol from the combined single mutations (Table 2, step 3−3); therefore the mutations are independent. To calculate $\Delta G_{D^- \rightarrow S^0}$ using step 2, subtract $\Delta G_{N^0 \rightarrow K^+}$ (Table 1, step 1−3) from ΔG_1^{FEP} (Table 2, step 1−1) to obtain $\Delta G_{D^- \rightarrow S^0} = 153.2 \pm 2.6$ kcal/mol. This result is in agreement with step 1−3, where $\Delta G_{D^- \rightarrow S^0} = 153.4 \pm 1.5$ kcal/mol (Table 2, step 1−3).

$\Delta G_{N^0 \rightarrow K^+}$ obtained from simulations where the second mutation is either an atomic ion of similar radius (Cl^-) or a larger amino acid are nearly identical (-70.7 ± 1.3 versus -70.9 ± 1.5 kcal/mol, respectively). The self-energy contributions to $\Delta G_{N^0 \rightarrow Cl^-}$ and $\Delta G_{D^- \rightarrow S^0}$ must be the same to produce such similar results; therefore, the self-energy depends on neither the radius nor the identity of the ionic species, to within the error. This result highlights a strength of the method presented in this work. $\Delta\Delta G_{solv}$ depends on R/L; hence, the effect of the difference in the radii should be minimal for sufficiently large L. However, the radius of a complex molecule is not usually well-defined. In addition, eq 21 was derived for a hard sphere, as is appropriate for atomic ions, but the accuracy of the correction for other solutes is unknown. Fortunately, the self-energy cancels implicitly in this method, and the radius is not needed to calculate the free energy change.

3. Comparison of E_{self} to Correction Terms. The FEP of the single mutations calculated for step 3 yields $\Delta G + E_{self}$. Having previously obtained the free energy change, we can calculate the self-energy and compare it to theoretical corrections such as C_{EW} and $\Delta\Delta G_{solv}$ (Table 3).

The self-energy was calculated for $N^0 \rightarrow K^+$ and $N^0 \rightarrow Cl^-$ in system sizes of $L = 30$, 20, and 11 Å (exact values are given in Table 3). E_{self} was calculated by subtracting the free energy changes (Table 1) from ΔG^{FEP} of the single ion mutations. C_{EW} is defined by eq 10, and $\Delta\Delta G_{solv}$ was calculated from eq 21 using $\varepsilon_s = 78$ and $R = 3.53$ and 4.54 Å for K^+ and Cl^-, respectively, as defined by the CHARMM 27 force field.[64]

The system sizes are not identical as the FEP calculations are performed in the isobaric−isothermal ensemble; however, E_{self} and the theoretical correction terms do not appear to be overly sensitive to minor differences in the system size. The self-energy terms of the two ions are comparable for each system size, confirming that E_{self} does cancel.

Table 4. Evaluation of System Size Dependence via Free Energy Changes (kcal/mol)[a]

L (Å)	$\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + C_{EW}$ [b]	$\Delta G_{N^0 \rightarrow Cl^-} + \Delta\Delta G_{solv}$ [c]	$\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP}$ [d]
29.11	-108.1 ± 0.3	-92.7 ± 0.3	-21.3 ± 0.4
18.97	-115.6 ± 0.1	-93.2 ± 0.1	-20.9 ± 0.3
10.78	-131.7 ± 0.2	-99.2 ± 0.2	-22.6 ± 0.4

[a] All values are derived from Table 3. The errors are from the free energy calculated via FEP (ΔG^{FEP}) for a neutral dummy atom (N^0) into a potassium ion (K^+) or chloride ion (Cl^-) for three system sizes (L). [b] $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + C_{EW}$ is the final free energy change as estimated by the theoretical correction terms from eq 10.[41] No contribution of C_{EW} and $\Delta\Delta G_{solv}$ to the error is included. [c] $\Delta G_{N^0 \rightarrow Cl^-} + \Delta\Delta G_{solv}$ is the final free energy change as estimated by the theoretical correction terms from and eq 21.[21] No contribution of C_{EW} and $\Delta\Delta G_{solv}$ to the error is included. [d] $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP}$ is step 3−3 from Table 1 recalculated using ΔG^{FEP} from each system size.

C_{EW} greatly overestimates the value of the self-energy term. For $L = 30$ Å, $E_{self} = 0.1$ and 0.3 kcal/mol for $N^0 \rightarrow K^+$ and $N^0 \rightarrow Cl^-$, respectively, while $C_{EW} = -16.2$ kcal/mol for both ions (Table 3). This discrepancy worsens as the system size decreases. $\Delta\Delta G_{solv}$ is slightly higher than E_{self}, but within the error, except for $N^0 \rightarrow Cl^-$ when $L = 11$ Å (Table 3). The similarity of E_{self} and $\Delta\Delta G_{solv}$ indicates that additional dependencies are likely small contributions to the self-energy. At $L = 30$ Å, we consider E_{self} to be a negligible contribution to the free energy change for these mutations.

The relative effect of these corrections on the final free energy change can be seen by adding C_{EW} or $\Delta\Delta G_{solv}$ to ΔG^{FEP} (Table 4). For $\Delta G_{N^0 \rightarrow Cl^-}^{FEP}$ with $L = 30$ Å, the final free energy changes are $\Delta G_{N^0 \rightarrow Cl^-} = -108.1$ and -92.7 kcal/mol for $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + C_{EW}$ and $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + \Delta\Delta G_{solv}$, respectively. Table 4 shows a system size dependence in $\Delta G_{N^0 \rightarrow Cl^-}$ calculated using the theoretical correction terms C_{EW} and $\Delta\Delta G_{solv}$. For $L = 11$ Å, $\Delta G_{N^0 \rightarrow Cl^-} = -131.7$ and -99.2 kcal/mol for $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + C_{EW}$ and $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} + \Delta\Delta G_{solv}$, respectively. Residual dependence on system size is an indication of inaccuracy in these correction terms.

The system size dependence of the method presented here can be evaluated by recalculating step 3−3 in Table 1 using different system sizes. In this step, the independence of the mutations is evaluated by comparing the difference in the free energy changes calculated either concurrently or individually. Above, we found $\Delta G_2^{FEP} = -21.5 \pm 0.6$ kcal/mol for the concurrent dual mutation and $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP} = -21.3 \pm 0.4$ kcal/mol for the combined single mutations ($L = 30$ Å). If instead we calculate $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP}$ using ΔG^{FEP} of $L = 20$ Å and 11 Å, we find $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP} = -20.9 \pm 0.3$ and -22.6 ± 0.4 kcal/mol, respectively (Table 4). The cancellation of the self-energy in $\Delta G_{N^0 \rightarrow Cl^-}^{FEP} - \Delta G_{N^0 \rightarrow K^+}^{FEP}$ requires that the two system sizes be equal. From Table 3, it is evident that this is not always the case in the isothermal−isobaric ensemble, which is an additional source of error not included in Table

4. For large system sizes, this error is minimal, but for $L = 11$ Å, the difference is likely not negligible. Nevertheless, there is very little dependence on system size in the results from the method presented in this paper.

4. Efficiency. The efficiency of the method presented in this work is similar to current calculations for neutral mutations, as FEP of atomic ions (step 1) is computationally inexpensive compared to FEP in a biomolecule required for any method (step 2). If parallel computing resources are available, this method adds little additional wall-clock time. All of the simulations, including the optional calculations for verifying independence (step 3), can be run in parallel.

In FEP calculations performed with biomolecules, the system sizes are often much larger than those used here. The two single mutation FEP simulations must be performed using the same system size, which should be chosen to minimize nonbonded interactions between the larger solute and its periodic images. However, since the verification step (step 3) is independent of system size, the single mutation FEP simulations are not required to use the same system size as the dual mutation FEP simulation. This could be of particular use if interactions between the two mutations are an issue; if the dual mutation FEP needs to be rerun using a larger system size, the smaller single mutation FEP simulations can still be used to check for independence. In general, smaller system sizes can be used for the single mutation FEP simulations to make the verification of independence less computationally costly, provided that the system size is reasonable.

The ability to divide large charge mutations into smaller steps allows the staging methods often implemented for λ to improve accuracy[62] to be applied to q as well. While this does not increase the computational efficiency of the final calculation, if q is not linear in λ, it may be easier to find an optimal staging scheme for q than for λ.

We have shown that the method presented in this work allows the free energy change for charge mutations to be determined without the limitations of current approaches. The mutations were confirmed as independent in the cases demonstrated here, validating the use of concurrent dual mutations. We find that the self-energy is primarily dependent on charge; other dependences in the self-energy, such as radius, are seen to have little affect. We also find that, at $L = 30$ Å, the system is sufficiently large to minimize the self-interactions for these mutations.

VI. Summary and Conclusions

The dependence of the self-energy on both charge and system size has been a barrier to accurately calculating the free energy change for charge mutations by FEP or TI using the Ewald summation for electrostatics. The method presented here removes the need for precise knowledge of the self-energy contribution to the free energy introduced by the periodicity in the Ewald implementation.

The method presented in this work translates the results from studies of charge mutations in atomic ions to more complex systems. The solvation free energy of atomic ions can be accurately obtained, but it requires analytical corrections which are not well-defined, or well studied, for multiatom mutations. The method presented in this work is general in its implementation but will be most useful for calculations of binding free energy differences, $\Delta\Delta G$, of biomolecules, as illustrated in Figure 1. Unlike current methods used for charge mutations in biomolecules,[10−12,49] it does not require experimental data to confirm the accuracy of the self-energy correction. It is self-contained and requires only self-consistency in the simulation parameters. Moreover, the assumptions made here can be tested for any system.

The uncertainty of the self-energy correction is removed because the self-energy terms implicitly cancel with the method presented here. Furthermore, the self-energy can be determined using the free energy change, and self-energy dependencies can be tested for any complex system. The self-energy can provide a quantitative metric for the extent of self-interactions in MD simulations that use the Ewald implementation.

Electrostatic interactions are the source of all nonbonded interactions in the physical world. The ability to calculate the free energy of perturbation associated with a charge differential without the uncertainty associated with large corrections has many important applications. Mutation studies can be done computationally for any residue, regardless of charge. The effect of ions on DNA, RNA, and ion channels can be directly studied, and the effect of charge on ligand binding can be isolated. The ability to decompose a thermodynamic reaction into intermediate steps allows small increments of charge to be used, which enables more general studies to explore the effect of electrostatic interactions on protein dynamics and binding as a function of charge.

References

(1) Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.

(2) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(3) Chipot, C.; Pearlman, D. A. *Mol. Simul.* **2002**, *28*, 1–12.

(4) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564–568.

(5) Dixit, S. B.; Bhasin, R.; Rajasekaran, E.; Jayaram, B. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 1105–1113.

(6) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508−134508–13.

(7) Bren, U.; Martinek, V.; Florian, J. *J. Phys. Chem. B* **2006**, *110*, 12782–12788.

(8) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.

(9) Shivakumar, D.; Deng, Y. Q.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.

(10) Zhou, R. H.; Das, P.; Royyuru, A. K. *J. Phys. Chem. B* **2008**, *112*, 15813–15820.

(11) Donnini, S.; Mark, A. E.; Juffer, A. H.; Villa, A. *J. Comput. Chem.* **2005**, *26*, 115–122.

(12) Das, P.; Li, J. Y.; Royyuru, A. K.; Zhou, R. H. *J. Comput. Chem.* **2009**, *30*, 1654–1663.

(13) Michielin, O.; Karplus, M. *J. Mol. Biol.* **2002**, *324*, 547–569.

(14) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.

(15) Pan, Y. M.; Gao, D. Q.; Yang, W. C.; Cho, H.; Zhan, C. G. *J. Am. Chem. Soc.* **2007**, *129*, 13537–13543.

(16) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

(17) Frenkel, D.; Smit, B. Free energy calculations. In *Understanding molecular simulation: from algorithms to applications*, 2nd ed.; Elsevier: San Diego, CA, 1996; pp 167−171.

(18) Allen, M. P.; Tildesley, D. J. Free energy estimation. In *Computer simulation of liquids*; Oxford University Press: New York, 1987; pp 213−219.

(19) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069–1072.

(20) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.

(21) Hunenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.

(22) Figueirido, F.; Delbuono, G. S.; Levy, R. M. *J. Chem. Phys.* **1995**, *103*, 6133–6142.

(23) Sakane, S.; Ashbaugh, H. S.; Wood, R. H. *J. Phys. Chem. B* **1998**, *102*, 5673–5682.

(24) Bogusz, S.; Cheatham, T. E.; Brooks, B. R. *J. Chem. Phys.* **1998**, *108*, 7070–7084.

(25) Kastenholz, M. A.; Hunenberger, P. H. *J. Chem. Phys.* **2006**, *124*, 124106−124106–27.

(26) Kastenholz, M. A.; Hunenberger, P. H. *J. Chem. Phys.* **2006**, *124*, 224501−224501–20.

(27) Aqvist, J.; Hansson, T. *J. Phys. Chem. B* **1998**, *102*, 3837–3840.

(28) Aqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587–1597.

(29) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2009**, *113*, 10255–10260.

(30) Carlsson, J.; Aqvist, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385–5395.

(31) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876–5886.

(32) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.

(33) Haranczyk, M.; Gutowski, M.; Warshel, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4442–4448.

(34) Jorgensen, W.; Ulmschneider, J.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264–17270.

(35) Almlof, M.; Carlsson, J.; Aqvist, J. *J. Chem. Theory Comput.* **2007**, *3*, 2162–2175.

(36) Harder, E.; Roux, B. *J. Chem. Phys.* **2008**, *129*, 234706−234706–9.

(37) Lamoureux, G.; Roux, B. *J. Phys. Chem. B* **2006**, *110*, 3308–3322.

(38) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(39) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(40) Petersen, H. G. *J. Chem. Phys.* **1995**, *103*, 3668–3679.

(41) Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, *100*, 1206–1215.

(42) Grossfield, A.; Ren, P. Y.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(43) Figueirido, F.; DelBueno, G. S.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 5622–5623.

(44) Rozanska, X.; Chipot, C. *J. Chem. Phys.* **2000**, *112*, 9691–9694.

(45) Marrone, T. J.; Merz, K. M. *J. Phys. Chem.* **1993**, *97*, 6524–6529.

(46) Deleeuw, S. W.; Perram, J. W.; Smith, E. R. *Annu. Rev. Phys. Chem.* **1986**, *37*, 245–270.

(47) Cichocki, B.; Felderhof, B. U.; Hinsen, K. *Phys. Rev. A* **1989**, *39*, 5350–5358.

(48) Nijboer, B. R. A.; Ruijgrok, T. W. *J. Stat. Phys.* **1988**, *53*, 361–382.

(49) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.

(50) Gomer, R.; Tryson, G. *J. Chem. Phys.* **1977**, *66*, 4413–4424.

(51) Klots, C. E. *J. Phys. Chem.* **1981**, *85*, 3585–3588.

(52) Schmid, R.; Miah, A. M.; Sapunov, V. N. *Phys. Chem. Chem. Phys.* **2000**, *2*, 97–102.

(53) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.

(54) Noyes, R. M. *J. Am. Chem. Soc.* **1962**, *84*, 513–522.

(55) Marcus, Y. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2995–2999.

(56) Randles, J. E. B. *J. Trans. Faraday Soc.* **1956**, *52*, 1573–1581.

(57) Warren, G. L.; Patel, S. *J. Chem. Phys.* **2007**, *127*, 064509−064509−19.

(58) Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Chem. Phys.* **1997**, *107*, 9275–9277.

(59) *Free energy calculations*; Chipot, C., Pohorille, A., Eds.; Springer: New York, 2007.

(60) Zwanikken, J.; van Roij, R. *J. Phys.: Condens. Matter* **2009**, *21*, 424102−424102–6.

(61) Rajamani, S.; Ghosh, T.; Garde, S. *J. Chem. Phys.* **2004**, *120*, 4457–4466.

(62) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Phys. Chem. B* **2003**, *107*, 5598–5611.

(63) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(64) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(65) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066–16081.

CT900565E