JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

An Overview of Electrostatic Free Energy Computations for Solutions and Proteins

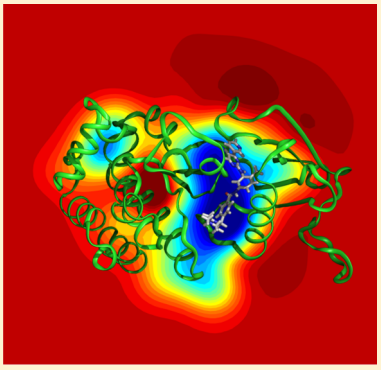
Yen-Lin Lin,[†] Alexey Aleksandrov,[‡] Thomas Simonson,*[,‡] and Benoît Roux*[,†,§]

[†]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, United States

[‡]Department of Biology, Laboratoire de Biochimie (CNRS UMR7654), Ecole Polytechnique, 91128 Palaiseau, France

[§]Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, United States

ABSTRACT: Free energy simulations for electrostatic and charging processes in complex molecular systems encounter specific difficulties owing to the long-range, $1/r$ Coulomb interaction. To calculate the solvation free energy of a simple ion, it is essential to take into account the polarization of nearby solvent but also the electrostatic potential drop across the liquid−gas boundary, however distant. The latter does not exist in a simulation model based on periodic boundary conditions because there is no physical boundary to the system. An important consequence is that the reference value of the electrostatic potential is not an ion in a vacuum. Also, in an infinite system, the electrostatic potential felt by a perturbing charge is conditionally convergent and dependent on the choice of computational conventions. Furthermore, with Ewald lattice summation and tinfoil conducting boundary conditions, the charges experience a spurious shift in the potential that depends on the details of the simulation system such as the volume fraction occupied by the solvent. All these issues can be handled with established computational protocols, as reviewed here and illustrated for several small ions and three solvated proteins.



1. INTRODUCTION

Ever since the first applications to biomolecular systems in the 1980s,[1−4] one major goal of free energy simulations has been to enable a quantitative characterization of complex processes occurring in a solution environment. For most biomolecular systems, long-range electrostatic interactions play a key role, and to obtain meaningful results, a number of issues must be handled with great care. Such issues are encountered when considering problems that range from the very fundamental, such as the solvation of small monatomic ions,[5−9] to the extremely complex, such as acid/base equilibria, redox potentials, catalysis, and binding of charged ligands.[1,10−14]

To calculate charging free energies, one must determine the electrostatic potentials at atomic sites and average them over multiple statistical configurations of the system of interest. In practice, this is carried out using either molecular dynamics (MD) or Monte Carlo (MC) simulations.[15−22] In the real world, systems are large but finite. However, in practice, it is most common to use periodic simulation models that that are meant to represent infinite systems in order to approach the thermodynamic limit and reduce undesired edge effects. Thus, the specific assumptions underlying a chosen computational treatment must be considered carefully for meaningful comparison with experiments. One strategy consists of treating a finite inner region with atomic details while accounting for the influence of the remaining distant particles implicitly through a potential of mean force (PMF).[23] This simulation strategy is widely used, with the PMF usually approximated through a dielectric continuum model or a dipole lattice,[24−28] yielding a finite cluster embedded within an infinite dielectric (e.g., the "Spherical Solvent Boundary Potential," or "SSBP," method in CHARMM[27]). Another widely used strategy consists of replicating a finite simulation cell through periodic boundary conditions (PBC).[29] The long-range interactions can then be computed, without truncation, via the method of Ewald or one of its variants,[30−34] or alternatively, by assuming a cutoff combined with a continuum reaction field.[35,36] All of these methods have been used to study a wide variety of problems, with various levels of theory for the energy function, from first-principles quantum mechanics to force fields, either additive or polarizable.

In this article, we first begin with a Theoretical Background section aimed at reviewing some earlier work and discussing the key conceptual aspects in detail. We first review an analysis of the liquid−gas (vacuum) interfacial potential, which reveals the role of the solvent quadrupole tensor.[37] We then introduce the main issues with simulating long-range electrostatics, first in a finite cluster and then in infinite models. We recall how to absorb the influence of the distant solvent molecules into a PMF, and we describe so-called "charge sorting" effects: the different ways of summing up the contributions of atoms and molecules to the electrostatic potential that can lead to different

Special Issue: Free Energy Calculations: Three Decades of Adventure in Chemistry and Biophysics

Received: March 6, 2014

   dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

values, due to conditional convergence.[17−21] We show that charge sorting affects the charging free energy in a simple way, which scales with the charge that is introduced. Therefore, differences between commonly used summation methods cancel out for processes that conserve the total charge, such as charging a neutral pair of ions. We review briefly the Ewald method for infinite, periodic systems. We show that with tinfoil boundary conditions, a perturbing charge is accompanied by a uniform canceling background charge, or "gellium," which makes the simulation cell of the periodic system neutral. As a result, charges experience a potential that is spuriously shifted, such that the spatial average over the simulation box is constrained to be zero. In this sense, the potential "floats" with respect to the usual reference potential (an ion in vacuum). Lastly, we consider the problem of single ion transfer between two media (e.g., gas and solution) and discuss how to relate such computations to experiments.

In the Results section, we present a number of illustrative free energy calculations for several small ions and three proteins, simulated either using a spherical model (a water cluster surrounded by a dielectric continuum)[27] or using PBC with Particle Mesh Ewald summation (PME) and tinfoil boundary conditions. With these calculations (details of the computational methods are given at the end), our objective is not to examine each test system in complete detail but rather to illustrate and help clarify some of the theoretical and practical considerations mentioned above. The first system is a simple toy model, where small ions (K^+ or Cl^-) are solvated within a small spherical water cluster, which itself is placed in a periodic lattice and simulated with PME. By varying the edge length L of the cubic box, we illustrate with FEP calculations the shifting of the value taken to be the electrostatic potential's zero or origin. A second toy model with a planar slab is examined with a similar strategy. Then, for the small ions sodium, aspartate, two histidine forms, and Gleevec, we perform charging free energy calculations under a variety of conditions, to probe potential offset effects with PME, charge sorting effects, and sensitivity to model size and other details. Finally, for three proteins, we consider respectively side chain deprotonation, ligand binding, and the spatial variation of the electrostatic potential. We obtain estimates for the system-dependent potential offset with PME and for the polarization artifacts with PBC, which are small with our 60−80 Å boxes, though not always negligible.

The article ends with a final section (Concluding Discussion) where we provide overall guiding principles for carrying out reliable free energy calculations in light of the present analysis. In particular, it is emphasized that good agreement with experiments has been obtained for protein systems, here and earlier, for a wide variety of problems, including acid/base pK_a shifts, redox potential shifts, and ligand binding free energies.

2. THEORETICAL BACKGROUND

2.1. Liquid-Vacuum Interfacial Potential. Many problems of interest involve moving charge from one place to another within the same solvent medium; e.g., proton transfer from one molecule to another, or protein:ligand binding/unbinding, so that boundaries and interface potentials are not important. In contrast, for problems where an ion is transferred from one medium or solvent to another, the potential drop across the interface can be very large and must be taken into account. This can be done by explicitly dragging the ion through a real interface. But the presence of the ion constitutes a perturbation of the interface, which complicates the analysis. A simpler problem is to consider the intrinsic potential at the liquid−vacuum interface, in the absence of any perturbation (n.b., throughout this article, we refer to the region of vanishingly low solvent density as "vacuum" rather than vapor or gas for the sake of simplicity). The average electrostatic potential along the z axis taken as the normal to the interface can be computed from the planar charge density, as a solution to the Poisson equation in one dimension, $d^2\langle\phi(z)\rangle/dz^2 = -4\pi\rho(z)$:

$$\langle\phi(z)\rangle = -2\pi \int_{-\infty}^{z'} \mathrm{d}z' \left[\int_{-\infty}^{z'} \mathrm{d}z'\langle\rho(z'')\rangle - \int_{z'}^{\infty} \mathrm{d}z''\langle\rho(z'')\rangle\right] \quad (1)$$

where $\langle\rho(z)\rangle$ is the average charge density of the liquid projected along the z axis. The quantity $\Phi_{lv} \equiv \langle\phi(z_l)\rangle - \langle\phi(z_v)\rangle$ is the liquid−vacuum (lv) interfacial Galvani potential. A similar formulation could be done for a spherical cluster, though the potential drop across the boundary is expected to converge to the value for a planar interface for reasonably large clusters (a few nanometers).[38] The magnitude of the Galvani potential of a physical liquid−vacuum interface calculated from eq 1 for the TIP3 water model is −520 mV[39,40] and is −545 mV for the SWM4-NDP polarizable water model based on Drude oscillators.[41] It is interesting to note that, after integration by parts, eq 1 can be rewritten as

$$\Phi_{lv} = +4\pi \int_{z_v}^{z_l} \mathrm{d}z' z'\langle\rho(z')\rangle \quad (2)$$

showing that the potential arises from the effective "dipole" moment of the interface. But the issue is considerably more subtle than just averaging the dipole of the molecules located at the interface. In this regard, the formal analysis introduced by Wilson et al.[37] provided great insights by showing how higher moments contribute to the interfacial potential. It is briefly summarized in the following for the sake of completeness. Following Jackson,[42] we write the charge density in the form:

$$\langle\rho(\mathbf{r})\rangle = \langle\rho_{net}(\mathbf{r})\rangle - \sum_{\alpha} \frac{\partial}{\partial x_\alpha}\langle P_\alpha(\mathbf{r})\rangle + \sum_{\alpha,\gamma} \frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\gamma}\partial\langle Q_{\alpha\gamma}(\mathbf{r})\rangle + \ldots \quad (3)$$

where $\langle\rho_{net}(\mathbf{r})\rangle$ is the distribution of the net (free) charge associated with the molecular groups (or mobile ions) and $\langle\mathbf{P}_\alpha(\mathbf{r})\rangle$ and $\langle\mathbf{Q}_{\alpha\gamma}(\mathbf{r})\rangle$ are the distribution of the dipole and quadrupoles of the molecules, respectively (α and $\gamma \equiv x,y,z$). The contribution from higher multipoles represented by +... will be neglected in the following development (but we will return to this later). Ignoring free net charges, we assume that the system is homogeneous in directions orthogonal to the interface, leaving only the dependence along z:

$$\langle\rho(z)\rangle = -\frac{d}{dz}\langle\mathbf{P}_z(\mathbf{r})\rangle + \frac{1}{6}\frac{d^2}{dz^2}\langle\mathbf{Q}_{zz}(z)\rangle \quad (4)$$

Finally, the interfacial Galvani potential can be rewritten as[37]

$$\Phi_{lv} = 4\pi \int_{z_v}^{z_l} \mathrm{d}z'\mathbf{P}_z(z') - \frac{4\pi}{6}[\mathbf{Q}_{zz}(z_l) - \mathbf{Q}_{zz}(z_v)] \quad (5)$$

Equation 5 can be simplified further if the interface is artificially constructed from the configurations generated from a bulk

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

simulation of the liquid, by assuming that there are no solvent molecules beyond the edge of the PBC box. In practice, this can be done by rereading configurations of the bulk simulation after setting the periodic lattice repeat distance to infinity in one direction (the coordinates of the solvent molecules must be first be wrapped to lie inside the PBC box without breaking molecules at the edge of the box). This procedure yields a "virtual interface,"[40] while retaining the isotropic orientation of the molecules in the liquid phase all the way to the edge of the PBC box. When this is done, the first term involving $\mathbf{P}_z$ is zero, yielding

$$\Phi_{lv} = -\frac{2\pi}{3}\rho_l Tr\{\mathbf{Q}\} \tag{6}$$

where ρ_l is the bulk density of the liquid, and $Tr\{\mathbf{Q}\} = \sum_j q_j \mathbf{r}_j^2$ is the trace of the quadrupole of the molecule.[37] Equation 6 is equivalent to the so-called "orientational disorder limit" (ODL), though $\Phi_{lv}^{(ODL)}$ was recovered here without making drastic approximations about the orientational order within the liquid as is often the case.[20]

According to eq 6, an important component of the interfacial Galvani potential, $Tr\{\mathbf{Q}\}$, is a property of the bulk liquid that has nothing to do with the orientation of the solvent molecules at the liquid−vacuum interface. Strikingly, the interfacial Galvani potential is not zero, even when the molecules of an interface are isotropically oriented. This is the critical insight provided by the analysis of Wilson et al.[37] The quadrupolar contribution to the interfacial potential is negative when $Tr\{\mathbf{Q}\} > 0$. This is the case with standard water potentials like SPC or TIP3, because the electronegative oxygen is surrounded by positively charged sites, though the precise value depends on the specific choice of the molecular center within the solvent molecule as shown in refs 20 and 40. The "center" can be chosen in different ways (e.g., geometric center, center of mass, oxygen, hydrogens, edge), and $\Phi_{lv}^{(ODL)}$ depends on this choice. For the SPC water model[43] with the center of mass as the molecular center, for example, it is roughly −550 mV. But the value of $\Phi_{lv}^{(ODL)}$ can vary by hundreds of millivolts depending on the choice of center or the details of the models.[20,40]

It is important to note that while eq 1 is exact, eq 5 was obtained by ignoring all contributions beyond the molecular quadrupole in eq 3. Equation 5 is thus an approximation that will not be quantitatively exact for all situations, depending on the various length scales in the problem. For example, it is not expected to adequately reflect the interfacial Galvani potential of a liquid formed by large and flexible polar molecules, or the potential arising from the solvent at the center of a neutral spherical particle. Furthermore, the Galvani potential for a physical water−vacuum interface based on eq 1 does not depend on the arbitrary choice of molecular center, whereas the quadrupolar term in eq 5 has a clear dependence on such a choice. We will discuss below ODL estimates of the Galvani potential based on eq 6 for a *virtual* interface, which depend critically on the choice of molecular center.[40] Unless stated otherwise, when we refer to the Galvani potential in the following, we will mean the potential difference at a physical water−vacuum interface based on eq 1.

A few important points warrant further remarks. Mathematically, the interfacial Galvani potential Φ_{lv} arising from the unperturbed charge distribution at the liquid−vacuum interface is a quantity that is unambiguously defined via eq 1. Φ_{lv} can be calculated with a variety of models, whether they are *ab initio* or based on simplified potential functions with point charges. However, Φ_{lv} is not a physically measurable quantity in the real world; it is the potential measured by an infinitesimally sharp nonperturbing test charge that probes all of space, including the interior of the solvent molecules. This, obviously, cannot be done in reality. In that sense, the interfacial Galvani potential Φ_{lv} defined by eq 1 is a formal construct that only has relevance to computational models. Furthermore, it should be noted that the actual value of Φ_{lv} depends upon the convention used to define it. In the construct above, the potential at all points in space arises from the superposition of the total charge density from all the particles.[44,45] This is the "internal Galvani potential." Strikingly, even if two models yield identical intermolecular forces and liquid structure, their internal Galvani potentials could differ substantially due to the different charge distribution near the nuclei.[46,47] For this reason, simulations results that are directly affected by the internal Galvani potential such as the absolute solvation free energy of ions must be compared with great caution across different models (i.e., force field, *ab initio*, density functional theory, etc.). Alternatively, one may construct an "external" Galvani potential, where the potential from each water molecule contributes only to points in space that are outside their repulsive core.[44,45] While both the internal and external Galvani potentials can be defined mathematically without ambiguity by specifying which convention has been used, neither can be measured experimentally by a physical process. The internal Galvani potential is the quantity that must be used in conjunction with charging free energy calculation carried out with PBC/PME (see below). In the remainder of this article, only the "internal Galvani potential" will be considered, and it shall be referred to simply as the "Galvani potential."

2.2. Finite Simulation Models. *2.2.1. Charge Sorting Effects in Finite Clusters.* Conditional convergence, in the mathematical sense, of the electrostatic potential arises with infinite simulation models.[17−21] As mentioned, different ways of summing over the contributions of atoms and molecules can give different results. For example, schemes that use a radial cutoff that is molecule-based ("M-summation") will differ from ones that are particle-based ("P-summation"). Each scheme can be seen as a particular way of "sorting" the atomic charges. But before we consider infinite models, it is instructive to examine how "charge sorting" effects can manifest themselves in a finite water cluster. Finite cluster models offer practical strategies to study the solvation of small ions[48,49] or even macromolecules.[50,51] The system is illustrated schematically in Figure 1. Let us first introduce the average potential $\langle\phi(\mathbf{r})\rangle$

$$\langle\phi(\mathbf{r})\rangle = \int d\mathbf{r}'\langle\rho(\mathbf{r}')\rangle\frac{1}{|\mathbf{r}-\mathbf{r}'|} \tag{7}$$

where $\langle\rho(\mathbf{r})\rangle$ is the total atomic charge density at point $\mathbf{r}$. As discussed in section 2.1 above, there is a sharp variation in $\langle\phi(\mathbf{r})\rangle$ at the physical boundary of the cluster corresponding to the internal water−vacuum (wv) Galvani interfacial potential $\Phi_{wv} = \langle\phi(\mathbf{r}_w)\rangle - \langle\phi(\mathbf{r}_v)\rangle$. As long as a physical boundary exists, this is an offset that does not vanish, even in the limit where $R_{cluster} \to \infty$. Now, let us imagine that we formally partition the cluster into two regions: an inner region of radius R_i containing one set of waters and an outer region containing the others (see Figure 1). The average electrostatic potential arising from the inner water molecules is

$$\langle\phi_i(\mathbf{r})\rangle = \int d\mathbf{r}'\langle\rho_i(\mathbf{r}')\rangle\frac{1}{|\mathbf{r}-\mathbf{r}'|} \tag{8}$$

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

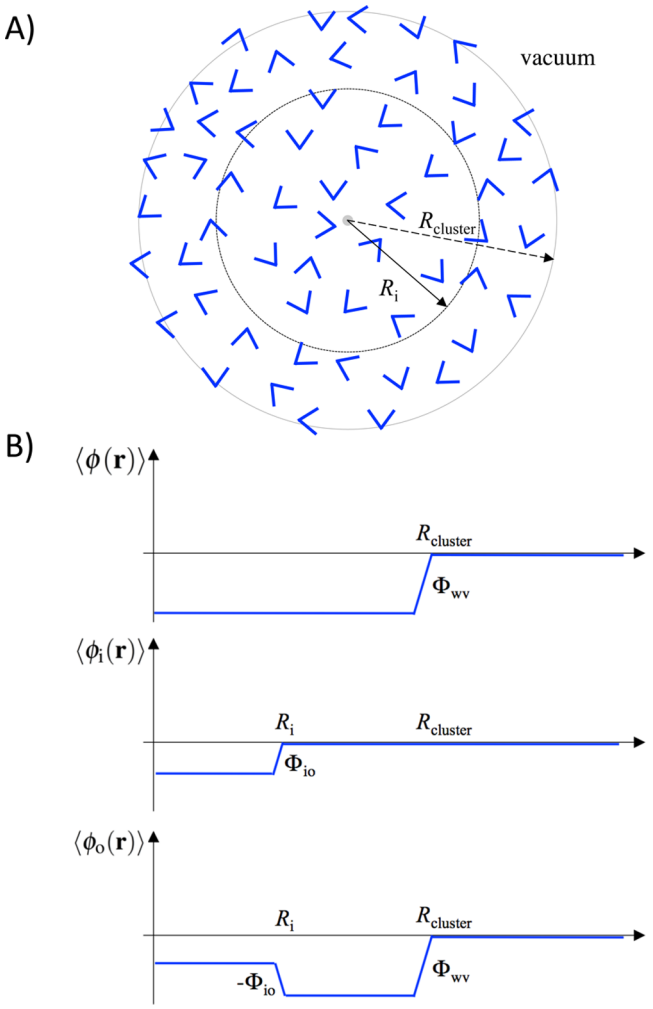


Figure 1. Properties of a finite spherical cluster in a vacuum. (A) Spherical water droplet of radius $R_{cluster}$ and definition of the inner region of radius R_i. (B) Average potential arising within the cluster, with a molecule-based partition into inner/outer regions according to the definition $\Phi_{wv} = \langle\phi(\mathbf{r}_w)\rangle - \langle\phi(\mathbf{r}_v)\rangle$ and $\Phi_{io} = \langle\phi_i(\mathbf{r}_i)\rangle - \langle\phi_i(\mathbf{r}_o)\rangle$.

while that from the outer molecules is

$$\langle\phi_o(\mathbf{r})\rangle = \int d\mathbf{r}'\langle\rho_o(\mathbf{r}')\rangle\frac{1}{|\mathbf{r}-\mathbf{r}'|} \tag{9}$$

By construction, the total average potential $\langle\phi(\mathbf{r})\rangle$ is equal to $\langle\phi_i(\mathbf{r})\rangle + \langle\phi_o(\mathbf{r})\rangle$.

As illustrated in Figure 1, electrostatic potential differences are expected to occur across the inner−outer (io) boundary. The values of $\langle\phi_i(\mathbf{r})\rangle$ will usually reach a plateau for large R_i, as long as we stay within the cluster ($R_i < R_{cluster}$).[17−21] We define the inner−outer (io) potential difference arising from the inner molecules as $\Phi_{io}^{(inner)} \equiv \langle\phi_i(\mathbf{r}_i)\rangle - \langle\phi_i(\mathbf{r}_0)\rangle$, where $\mathbf{r}_i$ is some point inside the inner region and $\mathbf{r}_o$ is some point inside the outer region. Likewise, we could define a similar potential difference $\Phi_{io}^{(outer)}$ arising from the outer molecules between some point inside the inner region and another point inside the outer region. However, because there is no discontinuity in the total potential defined by eq 7, $\langle\phi(\mathbf{r}_i)\rangle = \langle\phi(\mathbf{r}_o)\rangle$, we have the constraint that $\Phi_{io}^{(inner)} = -\Phi_{io}^{(outer)}$; in the following, we will refer to $\Phi_{io}^{(inner)}$ as simply Φ_{io}. Obviously, the inner−outer separation imposes a fictitious (virtual) boundary; this is not a real physical interface. Here, the radius R_i plays a role similar to a cutoff distance (see the discussion of infinite models, below). Near the inner−outer boundary, the molecules are essentially in an isotropic bulk phase, where they can adopt any orientation. As a result, the potential $\langle\phi_i(\mathbf{r})\rangle$ is formally equivalent to the "orientation disorder limit," or ODL. In this limit, the potential can be written as[17,20]

$$\Phi_{io}^{(ODL)} = -\frac{2\pi\rho_1}{3}Tr\{\mathbf{Q}\} \tag{10}$$

This is effectively eq 6, already derived above for a planar interface. We emphasize that, here, the quadrupole of the molecule is computed *relative to the definition of the molecular center* which has been used to "sort" the molecules, i.e., to determine which molecules are inner ones. For this reason, the magnitude of $\Phi_{lv}^{(ODL)}$ has an arbitrary character, which depends on the precise way the boundary is defined. An alternative route that avoids the ODL interfacial potential is to define the inner region on the basis of particles (P-summation) rather than whole molecules (M-summation). Sorting by particles between the inner and outer regions does not yield a potential jump at R_i in the quantity $\langle\phi_i(\mathbf{r})\rangle$, because the atomic charge density is uniform within the isotropic bulk liquid phase.[22] Notice that regardless of whether a molecule- or a particle-based sorting is used at the inner-outer boundary, the Galvani potential Φ_{wv} across the real, physical boundary of the cluster (the water−vacuum interface) must still be taken into account if we want to compute the free energy to solvate an ion within the cluster, relative to a vacuum.

2.2.2. Finite Cluster Embedded in a Dielectric Medium. A physically intuitive strategy to calculate the solvation free energy of charged solutes is to simulate a finite water inner region atomistically and treat the influence of the surrounding solvent implicitly. Strictly speaking, a finite cluster embedded in a dielectric medium is designed to represent an infinite system. However, many of the issues that pertain to truly infinite systems do not apply directly because the sums are finite. Formally, creating a finite representation with statistical properties similar to those of of an infinite system is achieved by integrating the degrees of freedom associated with the surrounding solvent out of the partition function and absorbing them into a PMF.[23,27] In the case of the Spherical Solvent Boundary Potential (SSBP) implemented in CHARMM,[27] the inner region is defined to contain a *fixed* number of water molecules, while the radius R_i is dynamically variable. The PMF for the inner region, W, can be defined by[27]

$$e^{-\beta W(\mathbf{r}_1,\ldots,\mathbf{r}_n)} \propto \int' d\mathbf{r}_{n+1}\ldots d\mathbf{r}_N\, e^{-\beta U(\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_{n+1},\ldots,\mathbf{r}_N)} \tag{11}$$

where the prime on the integral implies that the outer water molecules must lie farther than the inner water molecules, $\beta = 1/k_BT$. An alternative formulation would be to consider a fixed radius and let the number of water molecules in the inner region fluctuate according to a grand canonical ensemble.[52] Such PMFs are meant to incorporate all the effects from the water molecules in the outer region. Accordingly, the PMF must account for long-range dielectric polarization, through a reaction field method.[24,27,53] Once the PMF is defined, it is conceptually possible to extend the radius of the cluster $R_{cluster}$ toward infinity, until the contribution from the dielectric polarization is converged. It is customary to write the PMF as

$$W(\mathbf{r}_1, \ldots, \mathbf{r}_n) = U(\mathbf{r}_1, \ldots, \mathbf{r}_n) + \Delta W(\mathbf{r}_1, \ldots, \mathbf{r}_n) \tag{12}$$

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

where U is the potential energy of the n water molecules and ΔW represents the influence of all the water molecules in the outer region. It was previously shown that ΔW can be understood as the solvation free energy of the system embedded within a hard-sphere restriction.[23,27] If the PMF ΔW of such a hybrid explicit/implicit scheme was perfect, the (explicit) inner water molecules would be distributed isotropically (as in the bulk phase) up to the inner−outer boundary. As a result, the explicit water molecules of the inner would give rise to an ODL interface offset potential $\Phi_{io}^{(ODL)}$, which would be canceled by an opposite contribution from the solvent molecules in the outer region, $-\Phi_{io}^{(ODL)}$. In addition, the PMF ΔW would incorporate the effect of the physical water−vacuum interface, Φ_{wv}, thereby recovering exactly the Galvani potential of the physical water−vacuum interface. However, in practical implementations, the ΔW is necessarily imperfect and the inner region will not yield an ODL-like interfacial potential. In fact, it is observed that the average structure and orientation of the water near the edge of the inner region simulated with the PMF method resembles the liquid−vacuum interface,[27,38] which will give rise to an interfacial potential Φ_{PMF} very close to Φ_{wv} (a correction can be calibrated empirically). For example, charging free energy computations carried out with SSBP[27] approximately simulate the effect of Φ_{wv} without additional corrections.

2.3. Infinite Simulation Models. While experimental measurements are always carried out on large but finite systems, theoretical formulations typically consider properties of interest in the thermodynamic limit of infinite models. It is important to realize that infinite simulation models are abstract theoretical constructs and that charging free energy calculations carried out with these models can be limited in very specific ways. We distinguish four particularly important problems: (1) a lack of proper vacuum reference for the electrostatic potential, (2) conditional convergence of the electrostatic potential, (3) constraints imposed by the commonly used tinfoil boundary conditions in Ewald lattice sums, and (4) constraints imposed by PBC on the solvent polarization. The first two are relevant for any infinite model, while the third is more specific. All three are related to dielectric boundary effects. The fourth is relevant to any computations based on periodic models.

The first and most obvious limitation of an infinite system is the lack of a proper vacuum reference. If a system is infinite, then there is no distant vacuum and no macroscopic interface because all of space is filled by solvent. Strictly speaking, a charged solute cannot be brought from a distant position in a vacuum into the system through a physical liquid−vapor interface, and so the choice of origin for the potential (the state where its value is equal to zero) is not automatically defined. This is not an important issue when the charges are kept within the same medium (e.g., charge transfer between molecules in solution) or when the process being considered involves a neutral *pair* of charges. But in the most general situation, the absence of a well-defined origin for the potential must be handled carefully.

Another problem stems from the conditional convergence of the electrostatic potential, which now takes the form of an infinite sum. Because of the long-range nature of the $1/r$ Coulomb potential, this sum is only semiconvergent, and the result is not uniquely defined.[54] As with any alternating series, different results are obtained when the contributions of the individual charges are added up in different orders, leading to artifactual charge sorting effects.[15−22] In addition to Ewald's summation method, we have already mentioned schemes that are particle-based (P-sum) or molecule-based (M-sum). With these last schemes, a fictitious boundary associated with a cutoff distance is introduced to help order the infinite summation. It is noteworthy that the final potential depends on the details of this fictitious boundary, much like it would with a *real*, physical boundary (and the two are sometimes confused). Conditional convergence appears even if the system is not infinite in all three spatial directions: it could be an infinite planar slab, for instance. Fortunately, when we compute the free energy to introduce charges onto solutes, the resulting systematic differences or offsets in the electrostatic potential behave in a very simple way. Their contribution to the charging free energies scales linearly with the total charge introduced, independently of the solute structure. As a result, charge sorting effects vanish completely in situations where we introduce a neutral ion/counterion pair. This includes any applications that move a charge from one molecule to another within the same solvent (e.g., proton transfer or ligand binding), and any applications that move a neutral pair of charges across a vacuum−liquid or liquid−liquid interface. In the following, the conditional convergence due to charge sorting and fictitious boundaries is discussed in some detail, and it is shown with accurate numerical examples that these effects do indeed cancel out, subject to reasonable conditions on model size.

A third, more specific artifact is introduced when treating the long-range electrostatics with PBC using Ewald summation with tinfoil boundary conditions,[31−34] perhaps the most widely used approach in MD simulations. To maintain a neutral system overall, one assumes that the array of boxes, though infinite, is nevertheless embedded in a conducting medium. Whenever a new charge is introduced, there is then an accompanying background charge density that exactly compensates the new charge.[31−35] The background charge is often referred to as a "gellium," as mentioned above. It is structureless and spatially uniform, so it does not contribute to the electric field or forces. However, it has another effect: it effectively shifts the mean potential in the simulation box to zero and maintains this value at all times. Thus, not only does the infinite simulation model fill all of space, preventing the traditional choice for the potential origin or reference value, it actually imposes a specific choice. In fact, the origin (the zero) of the potential in Ewald summation with tinfoil boundary conditions "floats," depending on the solute, the solvent, their relative volumes in the simulation box, and the stage of the charging process. Fortunately, illustrative computations will show that, with a sufficiently large box size L and solvent fraction, the mean potential will approach the value for the pure solvent (as L^{-3}) and become independent of the solute.

Finally, with PBC, a fourth artifact arises from the constraint on the solvent polarization imposed by the periodicity. This affects both dielectric response and the organization of the counterion atmosphere. Thus, instead of a spherically symmetric polarization density that decreases as $1/r^2$ for a single ion q in solution, we have a periodic array of charges, boxes, and polarization.[56−60] This problem can become especially important when electric fields are not sufficiently shielded by a high dielectric solvent. Problematical situations can arise, for example, when the surrounding solvent is fairly nonpolar, but also when a macromolecule occupies a large fraction of the simulation box, or when the system includes a planar bilayer membrane with its low dielectric hydrocarbon

core simulated as an infinite two-dimensional periodic array.[61−63]

2.3.1. Charge Sorting Effects and Conditional Convergence of the Potential. We now consider charge sorting effects in more detail. With infinite systems, the electrostatic potential takes the form of an infinite sum. There are several different strategies to compute the electrostatic potential; for the moment, we consider three of them. The first is to treat an inner region atomistically and integrate out the degrees of freedom of the more distant particles, replacing them with a potential of mean force or PMF.[23] The second is to use PBC and compute all the long-range interactions directly (in Fourier space), with the method of Ewald or one of its variants, like Particle Mesh Ewald (PME).[30−34] The third strategy considers the effects associated with a virtual or fictitious cutoff, which is used as a conceptual tool to examine the conditional convergence of the electrostatic potential.

Let us consider an infinite system S, which may or may not be periodic. In practice, the system cannot be simulated directly without PBC, but it could be described by a mathematical model (quasicrystal, glass). We assume that statistical configurations are known for this system, and we consider one of them. Although S is formally infinite, we can still compute the potential by introducing a virtual or fictitious cutoff, in the following way. For the potential $\phi(\mathbf{r})$ at a particular point $\mathbf{r}$, we first do a calculation with a cutoff R_c then let R_c go progressively to infinity. Specifically, we define the potential $\phi_c(\mathbf{r};R_c)$ produced at $\mathbf{r}$ by solvent atoms within a cutoff region, using a specific rule to decide which atoms are inside or outside the region (such as a particle- or molecule-based cutoff),

$$\phi_c(\mathbf{r};\, R_c) = \sum_{i\in R_c} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} \tag{13}$$

In favorable situations, $\phi_c(\mathbf{r};R_c)$ will reach a plateau for large R_c. In that case, we can infer the limit when $R_c \to \infty$. This favorable situation is verified for models of liquid water, for example.[22,64] Notice that the limiting process used above does *not* involve letting the system size go to infinity: S is infinite from the outset and has no gas−liquid boundary. Notice also that in eq 13, we have taken as the zero of potential the value at inifinity, *inside* our infinite medium; this choice does not affect the differences between potential summation methods.

Although this conceptual protocol might not be explicitly carried out in applications, and the final result is obtained by taking a limit, this strategy models an infinite system with all its electrostatic interactions, just like the second, PBC/Ewald strategy. Why introduce such a protocol when PME works well? In fact, it was shown during the 1990s that when a cutoff R_c is used, a single simulation model can give different potential values, depending not only on the cutoff distance but on the precise cutoff method.[15−22] The differences persist even in the limit $R_c \to \infty$, which is relevant here. Åqvist and Hansson pointed out such a dependency for the potential at an uncharged particle in water.[15] Hummer and collaborators as well as Ashbaugh and Wood showed that in this case, the dependency arises from different ways of summing over the individual water atoms that give rise to the potential,[17−21] which can be atom-based or molecule-based (two possible methods, but not the only ones).[17−22,44,45,59,65−67] The differences affect the plateau value of $\phi_c(\mathbf{r};R_c)$ and persist when we take the limit $R_c \to \infty$. This means that for the infinite system, the potential $\phi(\mathbf{r})$ is not uniquely defined but depends on the precise summation method. If the system is periodic, we can use Ewald summation, which will give a unique result, but we can obtain different results for the same physical system if we use a different summation method.

Mathematically, these differences arise because the potential $\phi(\mathbf{r})$ has the form of an infinite series that is only conditionally convergent.[31−33,54] Physically, they can be traced to the different ways of defining the boundary between the inner, cutoff region and the rest of the system. The differences arise even though the cutoff boundary is *fictitious*, a tool used to schedule the calculation, *not* a real boundary. Just as with a real, physical boundary, the fictitious boundary effects cancel out from the free energy (as shown below) in applications where the total charge is conserved: the transfer of a neutral pair of ions from one solvent to another and the transfer of a proton from one molecule to another; numerical examples are given below. As with a real boundary, they do not cancel out in applications where a single ion is moved from one medium, such as a vacuum, to another, such as a bulk liquid. However, even in such cases, they can easily be eliminated, either by transferring the ion of interest along with a counterion, or by simulating a finite cluster with a real boundary, or both. Therefore, while conditional convergence and charge sorting are conceptually important, they are easy to deal with in practice.

2.3.2. Charge Sorting and Solute Structure. We show now that in a charging process, any potential offsets due to a particular charge sorting method will scale with the net perturbing charge but will not depend on the solute structure (if the simulation model is big enough). For concreteness, we consider the uncharging of an ionic ligand in a protein binding site, while recharging an identical ligand in solution; we assume a PBC setup. We start with the charged ligand bound to the protein, and another, uncharged ligand far from the protein, either in the same simulation box or a different box. We treat the ligand partial charges one at a time. For each ligand charge in the protein complex, we begin by (a) "growing in" a soft cavity that surrounds the partial charge. Then (b) we drag the charge and cavity into a bulk region of the box. Finally, (c) we decouple the charge and delete the cavity. The "PMF" step (b) is sensitive to the potential drop between the binding site and the bulk region, but not to any potential offset, choice of origin, or charge sorting effects on the potential. Indeed, different potential summation schemes differ by a constant offset, which cancels out of the field and forces. With this protocol, charge decoupling is always done for small spherical charges in bulk. Therefore, charge sorting effects only depend on the net value of the charges, *not* the structure of the protein or ligand. They cancel out if we now recharge the ligand in solution by doing steps a−c in reverse.

A simpler protocol is to uncharge, or "decouple," the ligand in the binding pocket, then recouple it in solution,[14,68] either in the bulk-like region of the same simulation box or in its own box. In either case, we have the exact same end points as with the "PMF" protocol, above. Since we use the same simulation box(es), the same end points, and the same energy function (including the specific form of the electrostatic energy term, the potential summation method, and any specific cutoff rules), the charging free energy will be the same (free energy is a state function). We may conclude that with the second, "decoupling" protocol, any charge sorting effects will also cancel out. The two protocols are illustrated in Figure 2.

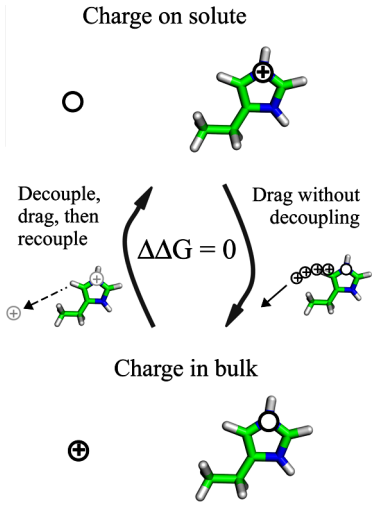


Figure 2. Thermodynamic cycle to uncharge a solute by moving its partial charges into bulk one by one. Only one charge is shown. The righthand leg drags it into the bulk without decoupling it (the series of charges and the arrow represent the motion). The lefthand leg first decouples the charge, then drags it into bulk, then recouples it. We show in light gray the decoupled charge both before and after it is dragged.

On the basis of this analysis, charge sorting effects vanish for problems that conserve the total charge, if the model is big enough (the potential drop between the ligand binding site and the edge of the box should be close to the large box limit; solvent far from the protein should be bulk-like). The cancellation of charge sorting effects was verified numerically in three examples reported below.

2.3.3. Ewald Summation and Potential Shifting. With periodic boundary conditions (PBC), one considers a finite box and an infinite array of periodic images. This is in fact the only practical computational method to formally simulate an infinite system. As mentioned above, there are several strategies to compute the electrostatic interactions. One of the most straightforward is to use Ewald lattice summation[30−33] (or its PME approximation[34,69]), which explicitly accounts for all the particles in the box and all their images. With conducting, or tinfoil boundary conditions, one considers a large array of simulation boxes, itself embedded in a conducting medium.[31−33] For each charge q, there is a background charge, or "gellium," spread uniformly throughout the box and totaling $-q$.[31−33] The Ewald sum with tinfoil boundary conditions converges to a unique result, and there is no need to choose a molecular center. The tinfoil boundary conditions do, however, impose a particular choice of origin for the potential (the box average). In effect, the potential is "shifted" so as to have a zero mean.

The neutralizing gellium allows us to simulate charging processes and molecules that are non-neutral, while maintaining a neutral periodic system. Being uniform, the gellium does not contribute to the electric field or the atomic forces. An alternative strategy to maintain neutrality in a charging calculation is to create both the ion and a counterion at the same time. With a large enough box, the two ions can be well separated and should not affect each other's solvation. In fact, it can be helpful to think of the gellium itself as a very large and diffuse "counterion" filling the box. We present examples below where the two strategies (gellium and counterion) are compared and give good agreement.

The potential produced at a point $\mathbf{r}_0$ by a charge q, along with all its images and associated gellium, is known as the Ewald potential, $\phi_{EW}(\mathbf{r})$, where $\mathbf{r}$ is the vector from q to $\mathbf{r}_0$. As above (for a cluster), we can define $\phi_P(R_c)$ as the potential at $\mathbf{r}_0$ produced by all the solvent atoms within a distance R_c. However, following Hummer et al.,[64] we now use the Ewald potential $\phi_{EW}(\mathbf{r})$ instead of the Coulomb potential q/r. Thus, for each charge q included in the sum $\phi_P(R_c)$, we also include its images and neutralizing gellium. When the cutoff distance R_c is increased, $\phi_P(R_c)$ reaches a plateau, as above. The plateau matches the value obtained by summing over the whole simulation box.[64] In this sense, Ewald summation matches (or "imposes") P-summation. A calculation using a molecule-based cutoff (M-summation) would give a different plateau ϕ_M.

When charging an ion with tinfoil boundary conditions, it is important to note that the charging free energy depends explicitly on the "shifted" box potential, which in turn depends (implicitly) on the size and content of the box. Indeed, consider a perturbing charge density added to an existing system; for simplicity, assume it is made up of a few point charges q_i, whose instantaneous positions $\mathbf{r}_i$ fluctuate: $\rho(\mathbf{r}) = \sum_k \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i^k)$, where the first sum is over all boxes k and $\mathbf{r}_i^k$ is the position of q_i in box k. With tinfoil boundary conditions, there is an associated gellium, so that the overall perturbing density has the form $\rho(\mathbf{r}) = \sum_k \sum_i q_i(\delta(\mathbf{r} - \mathbf{r}_i^k) - L^{-3})$, where L is the edge length of the box (cubic for simplicity). Integrating over one box, we note that the total perturbing charge is zero: $\int_{box} d\mathbf{r}\, \rho(\mathbf{r}) = 0$. We denote $\phi(\mathbf{r})$ as the instantaneous electrostatic potential in the simulation box; $\bar{\phi}$ is its spatial average: $\bar{\phi} = L^{-3} \int_{box} d\mathbf{r}\, \phi(\mathbf{r})$, and we write $\phi'(\mathbf{r}) = (\phi(\mathbf{r}) - \bar{\phi})$ the shifted potential. $\phi(\mathbf{r})$ includes contributions from charges in the primary box and all its images, as well as gellium contributions. In contrast, the (spatially uniform) gellium does not contribute to $\phi'(\mathbf{r})$. The perturbing charges contribute a term ΔU to the total energy per box. ΔU is the quantity that determines the perturbation free energy; it is often referred to as the energy gap.[6] It has the form

$$\Delta U = \int_{box} d\mathbf{r} \rho(\mathbf{r})\, \phi'(\mathbf{r}) = \sum_i q_i \phi'(\mathbf{r}_i) \tag{14}$$

where the first equality uses the special properties of the gellium and $\bar{\phi}$. In the bottom-most sum, $\mathbf{r}_i$ represents the instantaneous position of q_i in the parent box. Thus, the energy gap couples the perturbing charge density ρ and the *shifted* potential ϕ'. The mean potential $\bar{\phi}$ has dropped out through cancellation of $\rho(\mathbf{r})$ by the gellium. If the perturbing charge is a single infinitesimal charge increment dq at $\mathbf{r}_0$, the corresponding free energy derivative is

$$\frac{dG}{dq} = \langle \phi'(\mathbf{r}_0) \rangle \tag{15}$$

The brackets represent a time (or ensemble) average. Charges in the box thus experience an instantaneous potential ϕ' whose spatial average is zero—compare eq 15 to eq 17, further on. In effect, tinfoil boundary conditions impose the spatial average of $\phi(\mathbf{r})$, rather than a vacuum, as the potential origin.

Consider now the spherical cluster of the earlier section, with an uncharged particle P_0 at its center $\mathbf{r}_0$. Suppose we put the cluster in a large box and replicate it through PBC. The potential within the cluster in the parent box will not depend much on the edge length L of the box; the potential outside the cluster will be small, and the mean box potential will scale

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

approximately as L^{-3}. If L is only slightly larger than the cluster, the unshifted potential in the box will have a mean close to the cluster mean and close to the phase potential ϕ of liquid water; the potential shift $\phi-\phi'$ will be large, on the order of ϕ. If L is large, the potential averaged over the box will be close to zero, and ϕ will approach ϕ'. The energy gap and the free energy derivative to charge P_0 thus depend artificially on L, with the dependency disappearing as L^{-3} in the limit of large box sizes. This situation is modeled in detail in a later section (Results). In general situations, where the box contains a solute and solvent, there is also a dependency on box size due to potential shifting, with the potential of the phase converging to the Galvani potential Φ_{wv} in the limit of a large box, mostly filled by the solvent. In applications, one should check numerically that the individual derivatives and the total charging free energy have all converged to the large box limit. In this limit, the effect of potential shifting becomes solute-independent and cancels out when we compare different solutes (such as a protein and a small model compound).

2.3.4. Polarization Artifacts with PBC. So far, we have mainly discussed issues due to charge sorting, potential shifting with PME, and interface potentials, which are all inter-related. With PBC, there is another, potentially important artifact of a simpler and more direct nature, namely, the constraint imposed on the polarization density due to the periodicity. While the problem is clearly worse in the case of a charged solute, polarization artifacts in the periodic system exist even in the case of a neutral polar solute. Similar artifacts occur also when the solvent includes counterions, in which the ionic atmosphere is constrained to be periodic.

For a single ion of charge q in an infinite bulk solution, the polarization density is spherically symmetric and decreases as $1/r^2$. In the lattice, we instead have a periodic array of ions, boxes, and polarization.[56−60] Methods to correct for this have been proposed, especially for the simplest case of a small spherical ion in a large box.[33,56−58,70] Expanding the electrostatic charging free energy as a function of inverse box size, L^{-1}, Figueirido et al. observed[56] that the term in L^{-1} can be interpreted as the free energy of a point ion q, interacting with its images and neutralizing gellium through the Ewald potential, shielded by the uniform, solvent, dielectric medium. For a cubic box, this free energy is simply $-q^2\zeta/(2\varepsilon_w L)$, where ε_w is the dielectric constant of the water solvent and ζ is a constant, −2.837297. Thus, the electrostatic charging free energy $\Delta G_{elec}(L\to\infty)$ in the infinite box limit is related to the result obtained with a box size of L as follows:

$$\Delta G_{elec}(L\to\infty) = \Delta G_{elec}(L) + \frac{q^2\zeta}{2\varepsilon_w L} + O(L^{-2}) \tag{16}$$

The first-order L^{-1} correction is scaled by $1/\varepsilon_w$; for boxes of L = 60 Å (Gleevec, thioredoxin, below) and ε_w = 80, it is very small, while the last L^{-2} term is negligible. This implies that $\Delta G_{elec}(L)$ has already converged to the infinite box limit. Notice that $\Delta G_{elec}(L)$ includes two contributions: the first corresponds to the interactions of the new charge with solvent molecules; the second corresponds to its interactions with its images and the neutralizing gellium. For a single charge and a cubic box, this second term is $-q^2\zeta/(2L)$, which is about 8 kcal/mol for a 60 Å box, far from negligible (but it is included in a standard PME treatment) . Since the total $\Delta G_{elec}(L)$ has already converged with respect to L^{-1}, this implies that the first contribution to $\Delta G_{elec}(L)$ has a comparable but opposing dependency on L. This dependency can be thought of as a shielding of the ion−gellium interaction by the solvent molecules.

Correction schemes based on eq 16 are accurate for small solutes and highly polar solvents like water. For more complicated situations, numerical schemes have been proposed, which compute the charging free energy with a PB dielectric continuum model.[58,61−63,65] Essentially, these methods consist in carrying out PB calculations with and without PBC to determine the bias introduced by the constraint of periodicity. The effects of box size and shape were investigated for sodium, Gleevec, and the proteins thioredoxin, Crk, and Abl kinase. Even for the proteins, with box sizes of 60 Å or more, polarization artifacts appear to be small.

2.4. Charging Process: Transfer Free Energy of Simple Ions in Water. Now that we have surveyed many of the formal issues associated with electrostatic interactions in finite and infinite systems, it is worth returning to a fundamental process that illustrates many of the main issues: the transfer of a small monatomic ion from vacuum to bulk water. Conceptually, if we consider a macroscopically large but finite cluster model, we can imagine simulating the physical process of bringing the ion from a distant position in the vacuum into the cluster. By starting from this distant position in the vacuum, we effectively (and sensibly) choose it as the zero of potential. When the ion moves from the vacuum to the interior of the cluster, it experiences a large potential drop across the water−vacuum interface, Φ_{wv}, due to the Galvani interfacial potential.[22,40,47] Once inside the cluster, the average electrostatic potential $\langle\phi(\mathbf{r})\rangle$ from the solvent is expected to be fairly constant and uniform. Bringing the charge q as a series of small increments to the ion's center $\mathbf{r}$ = 0, the electrostatic charging free energy has the form[55]

$$\Delta G_{elec} = \int_0^q dq'\,\langle\phi(\mathbf{r}=0;\,q')\rangle \tag{17}$$

where $\langle\phi(\mathbf{r}=0;q')\rangle$ is the average electrostatic potential arising from the solvent when the ion carries the charge q'. It is important to note that, at the outset when q' = 0, there is a constant "offset" potential $\langle\phi(\mathbf{r}=0;q'=0)\rangle$ relative to the vacuum. If the radius of the ion is vanishingly small, the offset potential should be approaching the interfacial Galvani potential Φ_{wv} of an unperturbed bulk water phase. In contrast, if the radius of the ion is very large, $\langle\phi(\mathbf{r}=0;q=0)\rangle$ should actually be close to zero—the value of the potential in a vacuum—since the solute recreates a second macroscopic liquid−vacuum interface that gives an identical potential drop but with the opposite sign. For an ion of finite size, the true offset potential may be expected to be between zero and Φ_{wv} because the Galvani potential from the liquid−vacuum interface will not be completely canceled by the potential at the interface between the uncharged ion and the surrounding solvent.

If we adopt an alchemical decoupling protocol in which the charge of the ion is turned off, a number of issues arise, especially if we use PBC. With an infinite simulation model, the interface potential contribution is not directly represented, thus there is no vacuum to set the zero of the potential as a reference. The ion charge formally never crosses the solvent−vacuum interface; it simply vanishes from one medium and reappears in the other. If we use PBC with PME and tinfoil boundary conditions, the spatial average of the potential in the box is set to zero, whereas the physical reference for the potential origin should correspond to the ion in a vacuum. We

Table 1. Free Energy Derivatives $\partial G/\partial\lambda$ from PME and Cluster (SSBP) Simulations (kcal/mol)[a]

	sodium				His[b]			Asp		
λ	PME	SSBP	diff.	λ	PME	SSBP	diff.	PME	SSBP	diff.
0.00	9.1 ± 0.1	−0.2 ± 0.1	9.3	0.00	70.1	62.8	7.3	11.0 ± 1.4	15.9 ± 0.9	−4.9
0.20	−28.0 ± 0.4	−39.0 ± 0.4	11.0	0.25	21.3	13.2	8.1		−16.3 ± 1.1	
0.40	−68.9 ± 0.6	−78.6 ± 0.5	9.7	0.50	−17.1	−25.2	8.1	−58.4 ± 0.8	−50.2 ± 1.0	−8.2
0.60	−114.9 ± 0.1	−120.8 ± 0.7	5.9	0.75	−51.7	−58.6	6.9		−88.0 ± 5.0	
0.80	−160.3 ± 0.1	−165.0 ± 0.6	4.7	1.00	−87.4	−92.2	4.8	−138.5 ± 1.4	−132.3 ± 1.0	−6.2
1.00	−204.9 ± 0.8	−208.5 ± 1.6	3.6							
ΔG	−93.9 ± 0.3	−101.6 ± 0.6	**7.7**	ΔG	−14.0 ± 0.3	−21.3 ± 0.1	**7.3**[b] **7.9**[c]	−59.9 ± 0.3	−52.6 ± 0.7	**−7.3**
size	40 Å	26/32 Å			45 Å	48 Å		49 Å	38−56 Å	

[a]PME indicates PBC with PME, tinfoil boundary conditions, and a cubic box. SSBP indicates a spherical cluster surrounded by a dielectric continuum with a dielectric of 80. "size" indicates the box edge or sphere diameter. The PME/SSBP differences are shown. Results are averaged over at least two runs; uncertainties correspond to the deviation between runs. [b]Using the Hsd singly protonated form. [c]Using the Hse singly protonated form. The self-interactions of the perturbing charges do not contribute to the His solvation free energy and have been subtracted out for clarity.

must therefore also properly treat the origin for the electrostatic potential by adding the interfacial potential drop back to the decoupling free energy.

With PBC, conditional convergence also becomes an issue. If one computes the electrostatic potential acting on the ion with an M-sum method, as recommended sometimes, this creates a fictitious interface at the cutoff distance and its associated boundary potential $\Phi_{lv}^{(ODL)}$. But this ODL estimate is a poor substitute for the Galvani potential Φ_{wv} arising from a physical water−vacuum interface. Notice that when a cutoff is combined with PBC, whether molecule- or particle-based, a method must also be defined for choosing which image molecules to include in the potential calculation. Hummer et al. showed[17] that with P-summation and an atom-based minimum image convention, the mean potential on the central particle converges to the same plateau as in a cluster. In a related approach, Darden et al.[22] studied the solvation of a small spherical ion at the center of a large spherical water cluster of radius $R_{cluster}$. The potential at the center was obtained by integrating the field between the center and a cutoff distance R_c that is large, but less than the cluster radius R. The field at large distances from the center was approximated by continuum electrostatics. Because the method relies on Poisson's equation applied to a spherical cutoff sphere, it yields a mean electrostatic potential on the central ion that is the same as P-summation. The interfacial Galvani potential Φ_{wv} is not taken into account in these last two treatments.

In discussions of the hydration free energy of ions, one may consider the *real* physical value, which includes the contribution of the reversible work from crossing the physical liquid−vacuum interface, or the *intrinsic* bulk phase value, which is independent of the interfacial potential.[38,40,49,71−73] The relationship between *real* and *intrinsic* hydration free energies is defined as $\Delta G^{real} \equiv \Delta G^{intr} + q\Phi_{wv}$. Without any additional manipulations, ΔG^{intr} is the quantity that naturally comes out of a charging free energy calculation carried out with PBC/PME. Thus, it may be tempting to view the intrinsic free energy as reflecting more faithfully the true hydration free energy of the ion in the heart of the bulk phase and disentangled from the influence of a physical liquid−vacuum interfacial potential. However, this is a false impression because Φ_{wv} remains a formal quantity that cannot be measured in the real physical world and which can be shifted arbitrarily while leaving the physical properties of the model unchanged. For example, two water potentials could be engineered by embedding an isotropic quadrupole inside the oxygen core (like a point charge surrounded by a spherical shell of opposite charge) to yield arbitrarily large differences in their internal Galvani potentials while retaining completely identical intermolecular forces and liquid structure. The intrinsic solvation free energy of an ion in these two models calculated within a PBC/PME scheme would be very different, though the disparity would disappear by considering the real free energy, $\Delta G^{intr} + q\Phi_{wv}$. Alternatively, the real free energy can be obtained directly by carrying out the calculation with a finite cluster model embedded in a dielectric such as in the SSBP method implemented in CHARMM.[27] It follows that the concept of *intrinsic* free energy is essentially formal and subservient to the rule used to define Φ_{wv}. For this reason, it is probably safer to use the *real* solvation free energy when comparing different microscopic models based on force fields or *ab initio* treatments. A number of computational studies have avoided these issues by focusing on the total solvation free energy of neutral salts.[38,71,74]

Correction for polarization artifacts due to periodicity can be treated via eq 16, though the latter tend to be rather small in the case of a high dielectric solvent such as water. Several of these issues are circumvented if the alchemical charging process is simulated with a finite system in which the dielectric influence of the distant solvent molecules is incorporated implicitly via a PMF.[23,27] Ultimately, there are practical difficulties with any of the decoupling protocols described above. To avoid many of the issues associated with charged ions, a practical alternative consists of using the coupling method to introduce a neutralizing counterion at the same time, either in its own box or in another part of the simulation box (see Asp+sodium example below). If the theoretical transfer free energy of the counterion is known (for example from a PMF calculation), it can then simply be subtracted out of the computed charging free energy and the experimental value added back in. This protocol is closely analogous to the well-established use of model compounds to help estimate absolute pK_a's in proteins: one subtracts out a model compound simulation then adds back in a model compound experiment.[2,75] This protocol shifts the problem from single ion transfer to neutral pair transfer, with the help of a reference ion. It eliminates most if not all of the uncertainty and difficulties associated with interface potentials, box potential offsets, and charge sorting effects. All of these effects will cancel out of the computed free energy, so that their discussion becomes moot.

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

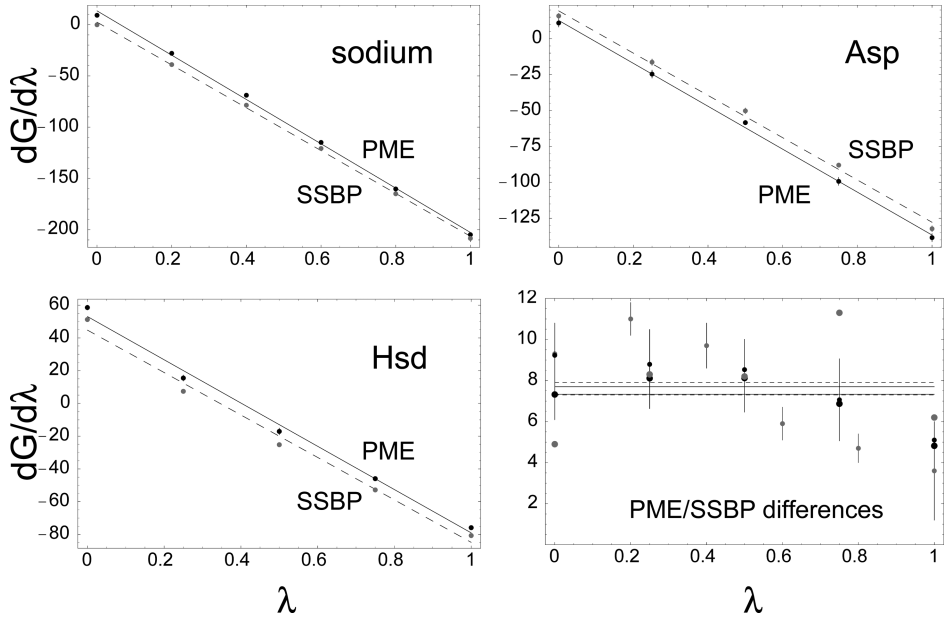


Figure 3. Free energy derivatives (kcal/mol) vs λ for sodium, Asp, and His (Hsd form), with PME (black dots) and SSBP (gray dots). Lines are linear fits. The bottom right panel shows the PME/SSBP differences (or SSBP/PME for the negative Asp); horizontal lines are the overall free energy differences. Results are shown for sodium (small gray dots), Asp (large gray dots), Hsd, and also Hse.

3. RESULTS FROM ILLUSTRATIVE CALCULATIONS

3.1. Charge Sorting and Potential Shifting Effects for Small Ions. We consider a set of small ions and compare calculations with PBC/PME and calculations with a cluster model embedded in a dielectric (SSBP method in CHARMM[27]). This allows us to illustrate not only potential shifting effects with PME but also the charge sorting effects discussed above. Uncharging simulations were performed for sodium and for aspartate and histidine analogues with neutral blocking backbone groups (see Methods). For His, we modeled two distinct forms, switching doubly protonated His to a singly protonated form with the proton on either the δ or ε nitrogen atom. We refer to these as Hsd and Hse, respectively. All four ions were simulated with PME and cubic boxes, using tinfoil boundary conditions. For sodium, we compared several box sizes, with and without a chloride (spectator) counterion. Each ion was also simulated in a water cluster, surrounded by a dielectric continuum representing bulk water (SSBP method).[27] For Asp and sodium, we compared several cluster sizes. The position of each (counter)ion in each box/sphere was harmonically restrained. Table 1 lists the free energy derivatives as a function of the coupling parameter λ for selected runs, also shown in Figure 3. Notice that for the insertion of a single charge, the free energy derivative, $\partial\Delta G_{elec}/\partial\lambda$, is equal to the mean electrostatic potential at the charge, $\langle\phi(\lambda)\rangle$,[76] so that it reports directly on potential shifting and charge sorting effects. Free energies are summarized in Table 2.

Several results should be noted. First, the simulations mostly display good convergence and robustness with respect to model details, including box or cluster size and the presence or absence of a counterion. The largest deviations are for Asp with SSBP (results in a ±1 kcal/mol range). Linear response is verified to a good approximation (though not perfectly; some dielectric saturation is evident in Figure 3). For the Asp/sodium pair, with PME, the total charging free energy is very similar whether the two ions are simulated in separate boxes or in the same box. In the latter case, the system is neutral at all λ values, and there is no gellium. Second, because the PME free energies for sodium are practically independent of box size over the range 20−50 Å, we conclude that the potential shift applied by PME is the same for all these boxes. This means that the contribution of the solute to the mean box potential is small, on the order of the statistical noise, which is around 0.5 kcal/mol.

Third, there are large, systematic PME/SSBP differences for the charging free energies, around 7−8 kcal/mol (Table 2). These can be viewed as an effect of charge sorting.[77] Indeed, we saw above that summing the potential with Ewald's method gives the same result as summation with a particle-based cutoff R_c, which then goes to infinity. For pure water with a small uncharged solute, the corresponding potential is close to zero.[20,64] With tinfoil boundary conditions, there is a small additional shift of the potential so that its box average is exactly zero, which does not affect the basic picture. In contrast, with SSBP, we sum over entire molecules up to the dielectric boundary, with an additional dielectric reaction potential. In theory, when the solute is uncharged (λ = 0), water structure should be bulk-like right up to the boundary; in practice, the dielectric model is imperfect and the solvent structure is somewhat different and less isotropic;[26,77] nevertheless, the potential due to the inner waters should be close to the estimate for a planar liquid−gas interface in eq 6. This is also the result expected from M-summation with a cutoff which then goes to infinity. Beyond the dielectric boundary, with λ = 0, the dielectric is not polarized and makes a very small contribution. Overall, we are left with a potential that resembles the M-sum result for pure water. This picture changes somewhat as we charge the ion, and so the overall PME/SSBP difference, 7−8 kcal/mol, is somewhat less than would be expected from a simple application of eq 6 and eq 10.

Fourth, the PME/SSBP differences are indeed rather different for different λ values (Table 1, Figure 3). For Asp, the differences (in kcal/mol/e) are +4 ± 3 when λ = 0 (protonated state), −9 ± 1 when λ = 0.5, and −13 ± 1 when λ = 1 (ionized state). For sodium, the PME/SSBP differences are

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Cancellation of Interface Artifacts for Neutral Pairs of Ions (kcal/mol)[a]

ion(s)	method	model size(s) (Å)	−ΔG (kcal/mol)	selected setup details
Na	PME	20/30/40	94.4/94.2/94.0	with a Cl^- spectator counterion
Na	PME	40/51	93.9/94.6 ± 0.5	100 ns total run time
Na	SSBP	26/32	102.2/101.1	20 ns total run time
Asp	PME	49	59.9 ± 0.3	50 ns total run time
Asp	SSBP	38/40/44/56	51.7/52.5/52.5/53.8	70 ns total run time
Asp + Na	PME	40 × 40 × 80	**154.0** ± 1.0	in a single box; 50 ns run time
Asp + Na	PME	49/51	**154.4**	in two separate boxes; see above
Asp + Na	SSBP	see above	**154.2**	in two separate spheres
Hsd	PME	45	14.0 ± 0.3	His, Nδ singly protonated form
Hsd	SSBP	48	21.3 ± 0.1	40 ns total run time
Hsd + Asp	PME	45/49	**73.9**	from the runs above
Hsd + Asp	SSBP	48/38−56	**73.9**	from the runs above
Hse	PME	45	11.7 ± 0.3	His, Nε singly protonated form
Hse	SSBP	48	19.6 ± 0.1	40 ns total run time
Hse + Asp	PME	45/49	**71.6**	from the runs above
Hse + Asp	SSBP	48/38−56	**72.2**	from the runs above

[a]Charging free energies for ions and ion pairs in water. Values in bold demonstrate the scaling of model artifacts with the ionic charge and their near-perfect cancellation when we charge a neutral pair with either PME or SSBP; see main text. PME indicates PBC with PME, tinfoil boundary conditions, and a cubic box unless otherwise mentioned; SSBP indicates a spherical cluster surrounded by a dielectric continuum with a dielectric of 80. Model sizes are edge lengths of the PME box or the spherical cluster diameter. Representative total run times are shown; uncertainties correspond to the deviation between runs.

+9.3 ± 0.5 and +4.6 kcal/mol/e for the uncharged and charged end points, respectively. The potential values are noisier than the total ΔG, as shown in Figure 3. Charge sorting contributions will depend on the λ value, but not on the PME box size when it is large (40−50 Å). In general, contributions from the SSBP boundary will depend on the λ value and the SSBP sphere size, at least for the smaller spheres.

Fifth, despite all these variations, for the overall charging free energies, the PME/SSBP differences cancel accurately when we consider the Asp/sodium, Asp/Hsd, and Asp/Hse pairs. The differences are −7.7 kcal/mol for Asp, +7.6 kcal/mol for sodium, +7.3 kcal/mol for Hsd, and +7.9 kcal/mol for Hse. The total charging free energy for the sodium/Asp pair is −153.9 kcal/mol with PME and −154.2 kcal/mol with SSBP. If sodium and Asp are simulated with PME together in a single, larger box ($40 \times 40 \times 80$ Å^3), the result is almost identical, −154.0 kcal/mol. For the Hsd/Asp pair, the total charging free energy is −74.0 kcal/mol with PME and −73.9 kcal/mol with SSBP. For the Hse/Asp pair, the PME/SSBP values are −71.6/−72.2 kcal, for a difference of 0.6 kcal/mol. Thus, for all four of these charge conserving processes, potential shifts due to PME and any charge sorting or boundary effects on the free energy accurately cancel out, as expected from the general argument above, despite the very different structures and charge distributions of the four solutes.

Similar cancellation was seen by Lu and Cui for a fourth, sodium/chloride pair,[77] when comparing PME and a method very similar to SSBP ("GSBP" method in CHARMM[78]), with PME/SSBP differences of +9.8 kcal/mol for sodium and −8.6 kcal/mol for chloride. Those calculations used a linear response approximation (LRA) to interpolate the end point free energy derivatives, which can account for the somewhat different value obtained for sodium (+9.8 kcal/mol, vs +7.6 kcal/mol in this work). Indeed, with the linear response approximation, our sodium charging free energy is about 3 kcal/mol larger than the TI value; for Asp, the differences are also a few kilocalories per mole. With PME, our sodium LRA estimate of 97.8 kcal/mol exactly matches that of Lu and Cui.[77] In a related earlier example, group-based truncation was shown to give good agreement with Ewald summation for uncharging a (neutral) water molecule.[79]

3.2. Polarization Artifacts and Potential Shifting with PME. *3.2.1. Simulation Study of Toy Models.* To illustrate the issue of potential shifting with PME, we now describe calculations for a series of "toy" models that involve a K^+-like cation and a Cl^--like anion. The parameters of the two ions were adjusted so that they yield identical charging free energies in an isolated sphere of water in a vacuum. In a first series of models, the ion is held at the center of a small sphere of about 10 Å radius comprising 125 TIP3 water molecules. The sphere is then inserted in a cubic lattice whose edge length L varied from 20 Å up to 60 Å, thereby changing the volume fraction occupied by the vacuum in the system. This series of toy systems was simulated with PME and tinfoil boundary conditions to compute the charging free energy of the ion (cation or anion). The charging free energies were also calculated with the small sphere in a vacuum, without any PBC ($L = \infty$), corresponding to a vacuum volume fraction of 1.0 exactly. In a second series of models, the ion is held at the center of a 15.5-Å-thick planar slab of 125 TIP3 water molecules. The slab is realized in the xy plane via orthorhombic PBC, with $L_x = L_y = 15.5$ Å while L_z is varied up to 90 Å. For comparison, the charging free energy of the cation and anion were also calculated for a system with cubic PBC/PME and L = 15.5 Å (vacuum volume fraction of 0.0 exactly), as well as for a system with SSBP, both meant to represent ion solvation in the bulk.

In these series of toy models, the dominant contribution to the charging free energy of the ion is expected to be well approximated by

$$\Delta G_{\mathrm{elec}}(q) = \langle \phi_{\mathrm{w}} \rangle q + Bq^2 \tag{18}$$

where $\langle \phi_{\mathrm{w}} \rangle$ is the mean potential of the water phase when $q = 0$ (in the absence of the ion's charge) and B reflects the quadratic reaction field response of the environment. These quantities depend on the lattice parameters and geometry of the problem (sphere or slab). However, if the cation and anion parameters have been adjusted to match one another for one basic lattice-independent system, such as a water sphere isolated in a vacuum without PBC/PME, then any deviations between the cation and the anion as the lattice changes is a clear indication of a change in the offset potential affecting the charging free energy. The linear term can then be separated from the quadratic reaction field term by taking the difference of the cation and anion's free energy, $[\Delta G_{\mathrm{elec}}(q = +1) - \Delta G_{\mathrm{elec}}(q =$

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$-1)]/2q = \langle\phi_w\rangle$. How should $\langle\phi_w\rangle$ be expected to vary in the different systems? In the water region, the mean potential is $\langle\phi_w\rangle$, and in the vacuum region, it is $\langle\phi_v\rangle$. If we assume that the water region is sufficiently large, then we can expect that the water−vacuum interfacial potential will be established between the liquid and vacuum regions:

$$\Phi_{wv} = \langle\phi_w\rangle - \langle\phi_v\rangle \quad (19)$$

Furthermore, the mean potential averaged over the simulation box, $\overline{\langle\phi_{box}\rangle}$, is constrained to be zero by virtue of the tinfoil PME treatment, i.e.,

$$\overline{\langle\phi_{box}\rangle} = (1 - f_v)\langle\phi_w\rangle + f_v(\langle\phi_w\rangle - \Phi_{wv}) = 0 \quad (20)$$

where f_v represents the volume fraction of the vacuum in the cell. This implies that the mean potential in the water phase is

$$\langle\phi_w\rangle = f_v \Phi_{wv} \quad (21)$$

In these toy models depicted in Figure 4, the mean potential in the water phase is a fraction of the full water−vacuum interfacial potential Φ_{wv}, which is about −520 mV for TIP3, or −12 kcal/mol for a monovalent cation. The analysis is shown in Figure 5, where the free energies are plotted as a function of the volume fraction f_v of the vacuum in the various systems. Both for the system with a spherical water region and the planar water slab, it is clear that $[\Delta G_{elec}(q = +1) - \Delta G_{elec}(q = -1)]/2q = \langle\phi_w\rangle$ (blue line) is directly comparable to $f_v\Phi_{wv}$ (cyan line). The agreement is excellent, indicating that this analysis is valid.

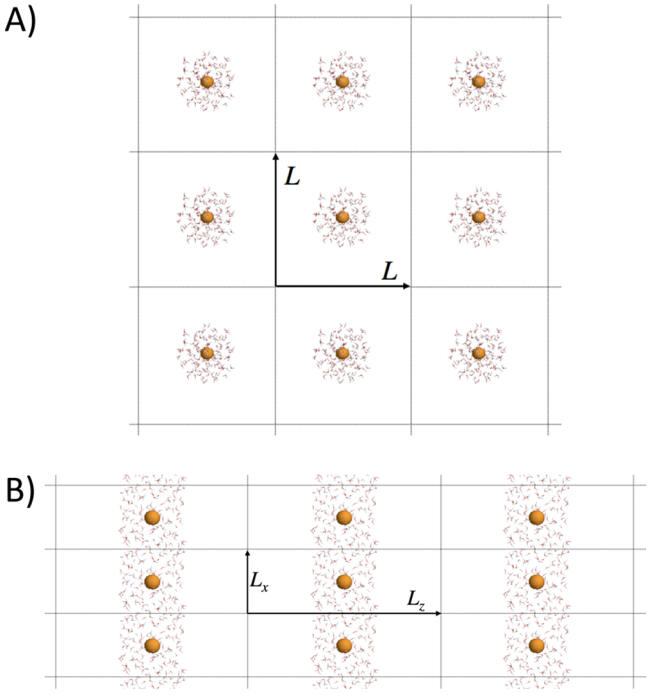


Figure 4. Charging free energies in inhomogeneous toy systems. Charging free energy of a K^+-like cation or Cl^--like anion in a series of toy model models as a function of f_v, the volume fraction of vacuum within the simulation cell. A sphere of 125 water molecules (radius of about 9.3 Å) in a cubic lattice (A) and a slab of 125 water molecules (thickness of 15.5 Å) in periodic lattice (B).

While the concepts were illustrated with a series of simple model systems, the same issues are expected to play a role in more complex inhomogeneous systems, such as liquid−liquid junctions, phospholipid membranes, micelles, and solvated proteins. For instance, the free energy of a ligand carrying a net charge in the binding site of a solvated protein will normally experience a reference offset solvent potential that differs from reference offset potential of the same charged ligand charged in a box of pure solvent. These considerations could be important in the computations of absolute binding free energy of charged ligands using alchemical FEP/MD simulations. As realistic examples, in the next sections we will estimate the magnitude of

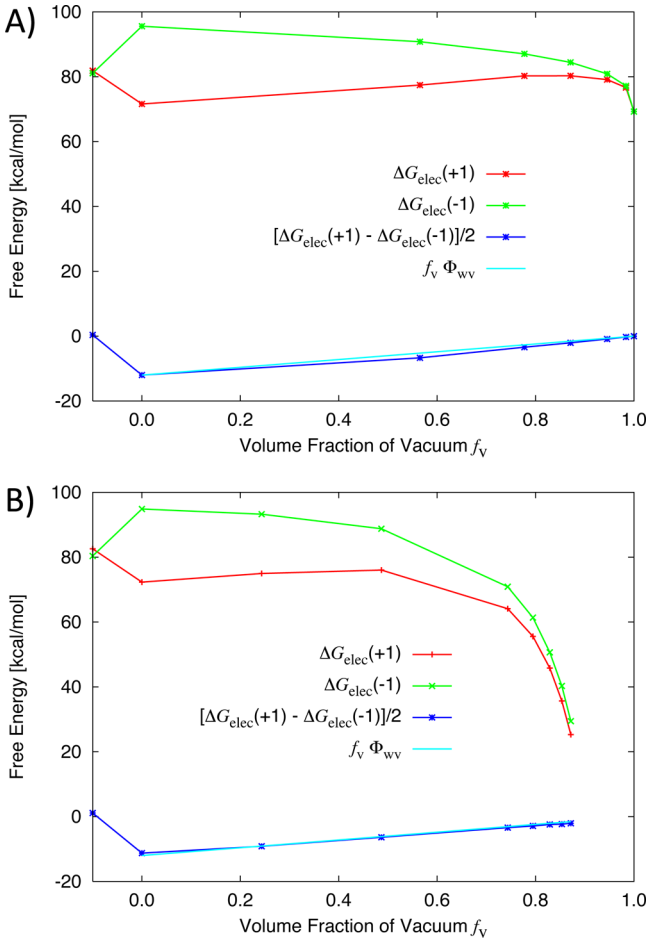


Figure 5. Charging free energies in inhomogeneous toy systems. Charging free energy of a K^+-like cation or Cl^--like anion in a series of toy model models as a function of f_v, the volume fraction of vacuum within the simulation cell. In A, results are shown for a sphere of 125 water molecules (radius of about 9.3 Å). Cubic systems were simulated with L = 15.5, 20, 25, 30, 40, 60, and ∞ Å, yielding a vacuum volume fraction f_v of 0.0, 0.57, 0.78, 0.87, 0.95, 0.98, and 1.0, respectively. In B, results are shown for a planar slab of 125 water molecules (thickness of 15.5 Å). Orthorhombic systems were simulated with L_z = 15.5, 20.5, 31, 46.7, 62.2, 77.7, 93.3, 108.8, and 124.4 Å, yielding a vacuum volume fraction f_v of 0.0, 0.24, 0.49, 0.74, 0.79, 0.83, 0.85, and 0.87, respectively. The value at f_v = 0 corresponds to the charging free energy of the ions in a PBC/PME water box with 125 TIP3 water molecules. The cyan line corresponds to $f_v \times 12$ kcal/mol. Included for comparison (at $f_v = -0.1$) is the charging free energy of the same ions in a sphere of 125 TIP3 water molecules with SSBP.[27] All charging free energy calculations were carried out using the PERT module of CHARMM[91] with at least 1 ns of sampling. Flat-bottom half-harmonic restraining potentials were applied to the solvent molecules via the GEO module of CHARMM[91] to retain the planar slab or sphere shape.

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

these effects for the deprotonation of an aspartic acid side chain in thioredoxin and for the binding of the anticancer drug Gleevec to the Abl tyrosine kinase.

3.2.2. Deprotonation of Thioredoxin Asp Side Chain. We simulated Asp26 deprotonation in thioredoxin and a model compound in solution, using PME with an octahedral simulation box. Earlier simulations of the same system used a cubic box with an edge length of about 59 Å, containing thioredoxin, three sodium ions, and 6379 waters, close to the 6445 waters used here. However, the present box has a different, more symmetrical shape, so that polarization artifacts due to periodicity should be significantly altered. Averaging over 16 runs (40 ns total) for each system, the computed free energy to deprotonate Asp26 in the protein is −47.9 (±0.6) kcal/mol; for the model compound in solution, it is −59.4 (±0.9). Above, for Asp with a cubic box, we obtained −59.9 kcal/mol. The protein−solvent difference is +11.5 (±1.0) kcal/mol, strongly favoring the protonated form of Asp26. While the result is significantly larger than the experimental estimate (about +3.8 kcal/mol), it is in very good agreement with the earlier calculations using the cubic simulation box, which gave −49.3 ± 1.6 and −60.5 ± 0.6 kcal/mol for the protein and model compound, respectively, and +11.2 (±1.7) for their difference.[80] The deviation from experiment can be attributed to the use of a fixed charge force field.[81]

We also simulated deprotonation of Asp20, using a cubic box; the charging free energy was −58.9 kcal/mol. For the Asp model compound in solution, above, we obtained −59.9 kcal/mol. The difference corresponds to a pK_a shift of 0.7 units, close to the experimental value of zero,[80] and giving a free energy deviation of about 1 kcal/mol. When we charge Asp20 and a sodium counterion in the same box (cubic, L = 68 Å), we obtain −154.0 kcal/mol. Summing the results for thioredoxin and sodium in two separate boxes gives −152.9 kcal/mol, within 1.1 kcal/mol of the single-box result, even though the box sizes are different, polarization artifacts and potential shifting effects are different, and the single-box simulation has no gellium (the system is neutral at all stages of the free energy perturbation). Given the long simulations and modest noise level, we conclude that systematic error associated with these effects, and possibly with an imperfect cancellation of charge sorting effects due to finite box size, is not much larger than 1 kcal/mol.

3.2.3. Binding of Gleevec to Abl Kinase. To illustrate that these issues arise in the treatment of protein−ligand association, we compute the electrostatic component to the absolute binding free energy of the kinase inhibitor Gleevec (or imatinib) to Abl kinase.[82,83] The electrostatic component appears in the staged protocol for calculating the absolute binding free energy, and it corresponds to the difference in charging free energy of the ligand in bulk solution and in complex with the protein in the presence of a set of translational, orientational, and conformational restraints.[82−84] While Gleevec can be either neutral or positively charged under physiological conditions in solution, there are strong indications that it carries a net positive charge in the bound complex with tyrosine kinases.[82,83] In the following, we consider that Gleevec always carries a net charge of +1.[85] For comparison, the FEP simulations were carried out with PBC, PME, and tinfoil boundary conditions, using various box sizes. The unbound ligand was simulated either in the same box as the protein, or in its own, separate box. The simulation systems are shown in Figure 6. The results are summarized in Table 3. We obtain consistent agreement between all runs and protocols, including a wide range of box sizes.

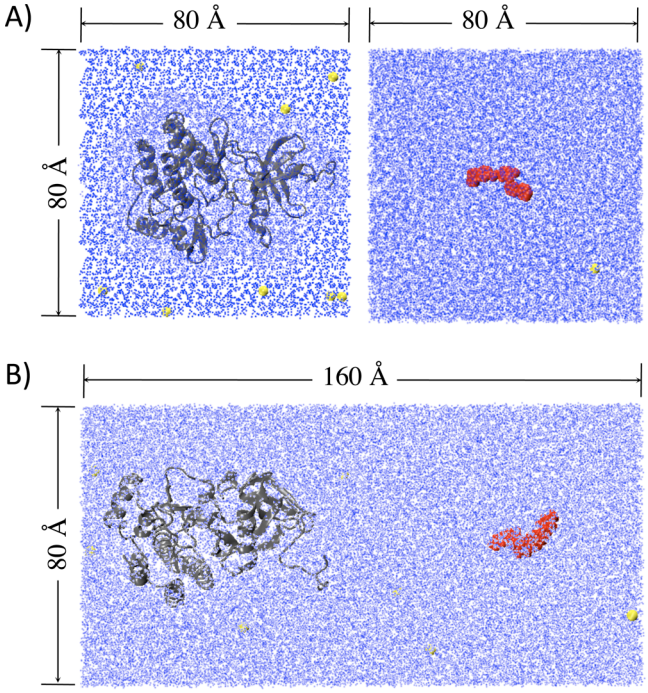


Figure 6. Atomic models for the calculations of the charging contribution to the binding free energy of Gleevec to Abl kinase. (A) The calculation was carried out with separate boxes for the ligand in complex with the protein (left) and the ligand in solution (right). (B) The calculation was carried out within a unique boxes, and the ligand in complex with the protein (left) is rematerialized in the bulk solution (right).

The charging free energy of Gleevec in solution is −67.0 ± 0.1 for cubic boxes with edges of L = 45 or 80 Å, with or without a spectator chloride counterion (n.b., the charge of the spectator does not change during the perturbation). Evidently, polarization artifacts due to PBC and potential shifting effects have already converged to the large box limit when L = 45 Å. The charging free energy of the ligand in complex with Abl kinase is slightly more negative, −67.7 kcal/mol, so that there is a small favorable contribution of −0.7 kcal/mol to the binding free energy. We refer to this as the two-box result. A previous study showed that the electrostatic contribution is small and that the binding affinity is dominated by van der Waals dispersion interactions.[82,83] As a comparison, we considered the charging free energy process with Abl and Gleevec kept in the same box (80 × 80 × 160 Å^3). We refer to this as the one-box result. The ligand is first uncharged in the bulk, for a charging free energy of −68.1 kcal/mol, then recharged in the protein binding site, for a charging free energy of −69.5 kcal/mol. The electrostatic contribution to the binding free energy is thus −1.4 kcal/mol, within 0.7 kcal/mol of the two-box result. This comparison of the one-box and two-box results suggests that the various artifacts are small, or that that they cancel out. For this reason, it is informative to evaluate their magnitude in detail.

In the one-box calculation (Figure 6B), the mean potential of the solvent phase is the same, whether the ligand is being charged in the binding site of the protein or in the bulk region. Under these conditions, the origin of the potential is

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Charging Free Energies for Gleevec and Abl (kcal/mol)

	PBC box (in Å)	Gleevec	Gleevec:Abl	difference
protein/ligand in separate (cubic) boxes	80	−67.0 ± 0.1	−67.7 ± 0.5	−0.7 ± 0.5
ligand in cubic box[a]	45	−67.1 ± 0.1		
PB correction for periodicity	80			0.19
solvent phase potential correction from PMEPot	80			0.74 (32 mV)
corrected final value				0.23 ± 0.5
protein/ligand in the same (orthorhombic) box	80 × 80 × 160	−68.1	−69.5	−1.4 ± 0.8
PB correction for periodicity				1.09
PB correction for interactions at distance r				−0.06
solvent phase potential correction				0.00
corrected final value				−0.37 ± 0.8

[a]With a Cl^- spectator counterion.

maintained across the calculation and no potential shift correction is needed. However, in the two-box calculation (Figure 6B), the mean electrostatic potential of the solvent phase is not necessarily the same in the box modeling the bound complex in solution and in the box modeling the ligand in solution. The solvent reference imposed by the PME with tinfoil conditions is constrained in these two systems. To clarify this issue, we calculated the average electrostatic potential in the cubic box with uncharged Gleevec bound to Abl kinase in the cubic box (L = 80 Å). The potential was calculated from a 50 ns MD simulation using the PMEPot plugin[86] in VMD.[87] The result is shown in Figure 7A. As expected, the potential is uniform in most of the bulk solvent region. Inside the protein, the potential becomes more positive than the solvent region. As observed from Figure 7B, the potential in the solvent region is sharply peaked around a single value, while the potential in the protein spreads over a wide interval. In the cubic box with L = 80 Å for the solvated Abl kinase, the average potential in the protein region is 206 mV and the average potential in the solvent region is −32 mV. The potential averaged over the entire box is zero, as constrained from tinfoil PME. The volume fraction of the solvent and the protein is 0.87 and 0.13, respectively. This analysis yields a mean potential correction that is +32 mV to counter the downshift of the solvent phase, corresponding to an energy shift of +0.74 kcal/mol for a unit charge.

The potential of the protein region is 238 mV more positive than the solvent. This difference is considerably smaller in magnitude than the water−vacuum interfacial Galvani potential $-\Phi_{wv}$ (−520 mV from the TIP3 model). It may be anticipated that, if we turned off all of the protein charges and transformed the macromolecule into a nonpolar obstacle for the solvent, the interfacial water−protein potential would start to approach the value of the water−vacuum interfacial Galvani potential. Alternatively, if the interior of the protein was transformed into a water-like medium, the interfacial potential would become zero. The solvent potential in the PBC box with the solvated protein is not sensitive to small alterations, such as turning off the charge of the ligand or a few residues. For instance, the mean potential of the protein region relative to the solvent region is 238 mV or 243 mV, with Gleevec charged or uncharged, respectively, while the potential of the solvent shifted from −32 mV to −27 mV. Fundamentally, the mean potential of the protein region relative to the solvent phase arises from an interfacial phenomenon, thus, it should not be affected if the protein region becomes larger. Once a globular protein is large enough to have a representative amino acid composition, such as is the case for Abl (274 residues), the

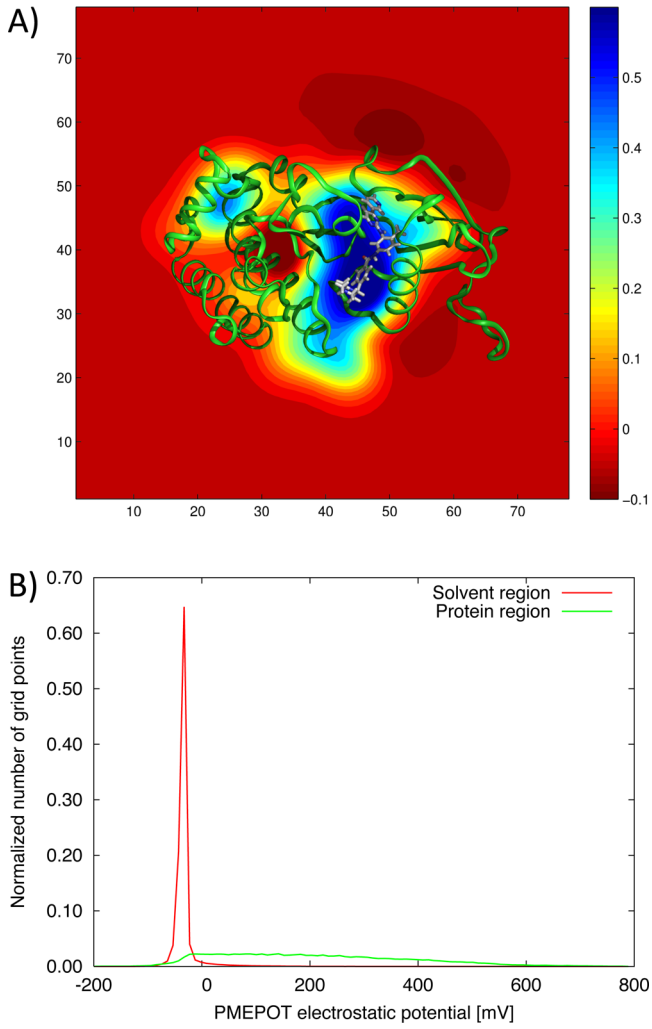


Figure 7. Average potential for Abl kinase with uncharged bound Gleevec in a cubic box (L = 80 Å) calculated from 50 ns MD simulation using the PMEPot plugin[86] of VMD.[87] (A) Contour plot showing a cut of the potential through the center of the cubic box. (B) Distribution of the values of the potential in the protein and solvent region. The average potential in the protein region is 206 mV, and the average potential in the solvent region is −32 mV. The solvent region is defined as any grid point farther than 5 Å from a protein atom (based on this definition, the volume fraction of the solvent and the protein is 0.87 and 0.13, respectively). The potential averaged over the entire box is zero, as constrained from tinfoil PME.

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

interfacial potential should be statistically converged. This points to the concept that the protein−solvent interfacial potential is likely to be a universal quantity on the order of 240 mV (protein region positive). Interestingly, a previous PB study of 305 proteins concluded that the average potential in protein interiors was 109 mV.[88] While this estimate based on continuum dielectric does not include the Galvani potential of the protein−solvent interface, the qualitative agreement is generally supportive of a universal protein−solvent interfacial potential. A direct evaluation of the shift in the mean solvent potential through an analysis such as carried out here is preferable, although it is possible to obtain a very rough estimate as

$$\langle \phi_w \rangle \approx -f_p \times 238\text{mV} \tag{22}$$

where f_p is the volume fraction occupied by the protein in the box. f_p can be roughly approximated as $(1 - Nv/V)$, where v is the standard molecular volume of a water molecule, 29.9 Å^3, V is the volume of the box, and N is the number of water molecules in the box. This rough analysis for the Abl system ($f_p \approx 10\%$) suggests a shift in the mean solvent potential on the order of 24 mV (n.b., the estimate is expected to be a lower bound because the volume fraction occupied by the solvent in the box is slightly overestimated by this treatment). A more rigorous approach is to carry out an analysis similar to that of Figure 7.

To obtain more information on polarization artifacts, we carried out PB continuum dielectric calculations with and without PBC using the PBEQ module[89] of CHARMM.[90,91] In the following, the superscript "Abl-Gleevec" represents the protein−ligand complex, "Abl" represents the isolated protein, "Gleevec" represents the isolated ligand, and "Abl-(r)-Gleevec" represents the protein and ligand separated by a distance r (80 Å in the case of the 1-box FEP calculation). For each of those situations, the continuum electrostatic free energy was computed from eq 26 with no salt concentration ($\kappa = 0$). The subscript "PBC" implies that the PB equation was solved by imposing the same periodic boundary conditions as used in the FEP/MD simulation, while the subscript "noPBC" implies that the PB equation was solved by setting boundary conditions on the edge of a very large box with a coarse grid followed by a focusing procedure with a finer grid. Based on these definitions, the correction for periodicity artifacts in the two-box calculation is

$$\begin{aligned}\Delta G_{\text{two-box}} = &[G_{\text{noPBC}}^{\text{Abl-Gleevec}} - G_{\text{noPBC}}^{\text{Abl}} - G_{\text{noPBC}}^{\text{Gleevec}}] \\ &- [G_{\text{PBC}}^{\text{Abl-Gleevec}} - G_{\text{PBC}}^{\text{Abl}} - G_{\text{PBC}}^{\text{Gleevec}}]\end{aligned} \tag{23}$$

the correction for the periodicity artifact in the one-box calculation is

$$\begin{aligned}\Delta G_{\text{one-box}} = &[G_{\text{noPBC}}^{\text{Abl-Gleevec}} - G_{\text{noPBC}}^{\text{Abl-(r)-Gleevec}}] \\ &- [G_{\text{PBC}}^{\text{Abl-Gleevec}} - G_{\text{PBC}}^{\text{Abl-(r)-Gleevec}}]\end{aligned} \tag{24}$$

and the correction for the interaction between Abl and Gleevec separated by a finite distance r is

$$\Delta G_{(r)} = G_{\text{noPBC}}^{\text{Abl-(r)-Gleevec}} - G_{\text{noPBC}}^{\text{Abl}} - G_{\text{noPBC}}^{\text{Gleevec}} \tag{25}$$

In all cases, the reference calculation with no PBC was obtained via a focusing calculation that first set the boundary potential at the edge of a large box of 160 Å on the side (a large orthorhombic box 160 Å × 160 Å × 320 Å was used to calculate "Abl-(r)-Gleevec" the interaction between Abl and Gleevec separated by a finite distance r). The results are given in Table 3. Comparing now the ligand charging steps with and without PBC, the correction is +0.19 kcal/mol (cubic box with L = 80 Å). Comparing now the ligand charging steps with and without PBC, the correction for the one-box result is +1.09 kcal/mol (periodic orthorhombic 80 × 80 × 160 Å^3 box). The correction is larger with the larger box, somewhat surprisingly. The difference is due to partial ligand−protein interactions when they remain in the same box.

The total correction for the two-box system, which includes the shift in solvent phase potential as well as the PB correction for periodicity, is 0.93 kcal/mol. The total correction for the one-box system, which includes the PB correction for periodicity and for the interaction between Abl and Gleevec separated by 80 Å, is 1.03 kcal/mol. The analysis makes it clear that the magnitude of the corrections is fairly similar, whether one is choosing to pursue a one-box or a two-box computational strategy. Furthermore, it is worth noting that the one-box system could equivalently support a binding calculation based on a PMF approach.[92] As the similar sizes of the various size corrections make it clear, there would be no clear disadvantage to using a PMF-based approach for this system, contrary to the conclusion of Rocklin et al.[65]

While the corrections are small, they are not negligible if one wishes to reach a consistency of less than 1 kcal/mol for this type of calculation. The final two-box estimate of the electrostatic component is 0.23 kcal/mol (two cubic boxes with L = 80 Å), and the final one-box estimate is −0.37 kcal/mol (single orthorhombic box with 80 Å × 80 Å × 160 Å). The discrepancy is $\sim k_B T$ between the two sets of FEP calculations, which is essentially on the order of the statistical uncertainty for the present systems (Table 3). Obtaining definitive computational results of high accuracy that can be reliably compared with experiments remains challenging because a number of factors can affect the measurements. For example, PB calculations indicate that the solvation free energy of Gleevec would change by −0.3 kcal/mol and the binding free energy to Abl would change by 1.0 kcal/mol by simply going from 0 to 150 mM salt. Thus, the order of magnitude of the discrepancy here is less than would be expected by from changing salt concentration.

3.2.4. Electrostatic Potentials in the Crk SH3 Domain. To further illustrate the sensitivity of electrostatic potentials to the size of the simulation box, we computed potentials on selected positions in Crk, using PME with a 60 or 74 Å cubic box. The two box volumes differ by a factor of 1.88. We considered the 57 C_α's and one atom from each of the side chains. Results are summarized in Figure 8. Comparing the same atoms in the two simulations, the mean unsigned difference is 1.1 kcal/mol/e for the C_α's and 1.0 kcal/mol/e for the side chain atoms. The standard deviation over each trajectory of the potential itself, averaged over the atomic positions, is about 7 kcal/mol for both simulations and both sets of atoms (backbone, side chain). The Pearson correlation coefficient between the potential values from the two simulations is 0.9996 for the C_α's and 0.9999 for the side chain atoms. These differences are small, not much larger than the expected noise level, confirming that systematic artifacts due to periodicity and PME are modest with these box sizes.

Nevertheless, the systematic differences are not completely negligible. Averaging over all C_α positions, the mean electrostatic potential is 323 mV with the 60 Å box and 344 mV with

dx.doi.org/10.1021/ct500195p | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

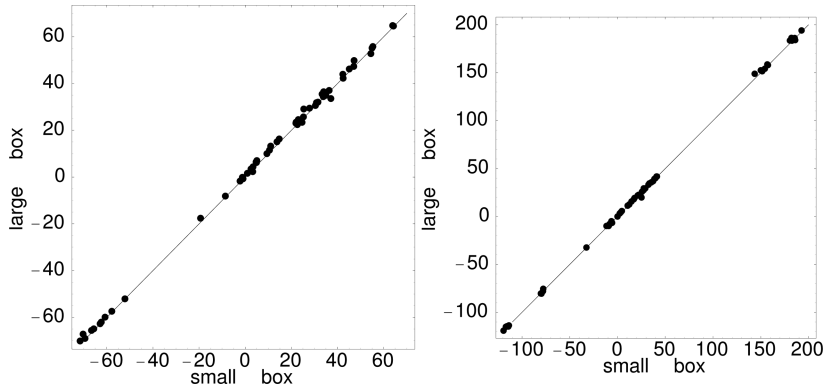


Figure 8. Electrostatic potentials in Crk (kcal/mol/e) from MD with PME and a 60 or 74 Å box. C_α's (left) and selected side chain positions (right; one per residue).

the 74 Å box. This yields a positive increase of 21 mV, which corresponds to an energy shift of +0.48 kcal/mol for a unit charge. An increase of the mean potential with box size was also seen above in the case of Abl kinase (Figure 7). A possible source for this (small) increase is the potential shifting due to the tinfoil boundary conditions. Indeed, the increase in solvent volume in the larger box modifies, in principle, the value of the potential origin imposed by the tinfoil boundary conditions. In the absence of any potential shifting (no surrounding tinfoil), and judging from pure TIP3P water, the mean potential of the solvent phase of our box is expected to be negative, at least not too close to the protein. By increasing the solvent volume, we increase this negative region, and so we increase the offset that is needed to shift the mean box potential to zero. As a result, the effective potential we measure within the protein should become more positive. To be precise, the protein region (augmented by a layer of about 5 Å, as in Figure 7) has a volume of about 28 000 to 30 000 $Å^3$, occupying 13−14% of the small box and 7−8% of the large one. According to eq 22, the mean potential of the water phase in these systems should be $\langle\phi_w\rangle = -33$ mV in the smaller box ($f_p = 0.14$) and $\langle\phi_w\rangle = -16$ mV in the larger box ($f_p = 0.07$). The difference is 17 mV, fairly close to the observed difference of 21 mV.

Overall, we conclude that with these box sizes, systematic artifacts are rather small, but potential shifting may contribute an offset as large as 0.5 kcal/mol to the potential and any charging free energies.

4. CONCLUDING DISCUSSION

Systematic artifacts in electrostatic free energy calculations have been extensively discussed and analyzed. With an infinite simulation model, the potential is undetermined relative to the normal zero of potential, i.e., empty space at infinity.[42,55] In real, finite systems, the system is bounded and there is an offset potential at the boundary; this offset potential affects charged solutes and does not vanish, even in the limit of an arbitrarily large system. The situation is perhaps somewhat confusing because it is common to assume that boundary effects become unimportant in the thermodynamic limit.

The role of conditional convergence and charge sorting was pointed out in the 1990s, but there is still some debate about the respective significance of particle- and molecule-based summation schemes. Both have been described variously as correct, incorrect, or unimportant. Clearly, any infinite simulation model is "incorrect," since it fills all space. With an infinite system, the potential is mathematically ambiguous (conditionally convergent), and there are several answers that are all equally correct mathematically. When one considers the potential summation methods used today, the ambiguity is fortunately weak, since the corresponding charging free energies differ through a term that is proportional to the charge that is introduced, independently of the solute structure.

Several authors have argued, at least implicitly, that an M-summation scheme is preferable because it breaks the indetermination in the potential of an infinite system and gives a result that is closer to the real system, which has a liquid−gas boundary.[44,45,65−67] However, this amounts to using a fictitious interface as a proxy for a real one, which is not very satisfying. Solvent structure near a fictitious interface (or its ODL analytical approximation) does not accurately represent a real physical interface, and there is no reason to choose such a crude model when a real interface is easy to model explicitly. The potential drop across a fictitious boundary is also dependent on the arbitrary choice of molecular center used to sort the solvent molecules and to define their quadrupole. Importantly, the contributions of charge sorting and any boundaries, real or fictitious, cancel out for neutral pairs of charges; therefore, the real interface need only be studied for one particular reference ion, as discussed at length above.

Another important artifact with Ewald summation and tinfoil boundary conditions is the shifting of the box potential to be zero. This imposes a different potential origin for every system, depending on its composition. The effect of the shifting potential of the solvent phase was illustrated explicitly with computations on a series of toy models. Unlike the effects of a boundary (real or fictitious), this system-dependent potential origin contributes directly to free energy differences between different molecules, like Gleevec and the Abl:Gleevec complex, unless they are simulated within the same box. As shown with illustrative computations, the magnitude of the artifact can be quantitatively estimated, and its impact can be considerably alleviated by using a sufficiently large solvent box. In all the practical applications considered here, the effect was fairly small. Nevertheless, the conditions for confirming that the phase potential of the solvent is converged to the pure solvent limit should always be checked.

Despite these artifacts and subtleties, we wish to emphasize that clear computational protocols are available to tackle a wide range of problems, as reviewed above. Indeed, good agreement with experimental results has previously been obtained for several proteins, in several cases using PME, tinfoil boundary conditions, and large boxes. These include (to mention a few)

redox potential shifts relative to a model compound,[93,94] acid/base constants,[1,80,81,95] and several protein:ligand binding free energies or free energy differences (involving charged ligands).[82,83,96,97] With the continuing growth in computer power, improving force fields and quantum methods, and sophisticated sampling algorithms, the future of electrostatic free energy calculations in complex biomolecular systems is very bright.

COMPUTATIONAL METHODS

We considered the following ions or molecules in aqueous solution: sodium (with or without a chloride counterion), aspartic acid and histidine with blocking N-acetyl and *N*-methylamide blocking groups (Asp, His),[80] the kinase inhibitor imatinib, or Gleevec, and the proteins Abl kinase (PDB code 1IEP), thioredoxin (PDB code 2TRX), and the SH3 domain of Crk (PDB code 1CKA; referred to simply as Crk). All systems except thioredoxin were solvated in a cubic water box with various edge lengths, and MD simulations were run with the CHARMM program[90,91] using periodic boundary conditions, Particle Mesh Ewald summation (PME)[34] with conducting or tinfoil boundary conditions, at room temperature and pressure.[98,99] Thioredoxin used the same setup, but an octahedral box. Asp, His, and sodium were also simulated in a water cluster surrounded by a water-like dielectric continuum,[27] using Langevin dynamics and constant pressure (SSBP method in CHARMM). For all systems, we used a modified TIP3P water model,[100,101] CHARMM27 force field parameters for the proteins, sodium, Asp, and His,[102] and Gleevec parameters optimized earlier.[85] Ions were held at appropriate positions (usually near the center of each box or sphere) by harmonic restraints on their centers. For two ions in the same box, they were restrained to positions well separated in the box.

For all systems, one or more perturbing charges were introduced by incrementing individual atomic partial charges. The charge increments were scaled from zero to one by means of a coupling parameter λ. To obtain charging free energies, we performed MD simulations at a series of λ values, usually around 5. The corresponding free energy increments were obtained either by using WHAM to analyze histograms of the energy gap, $\Delta U = U(\lambda + \delta\lambda) - U(\lambda)$,[103−105] or by thermodynamic integration (TI). Free energy derivatives, $\partial G/\partial\lambda = \langle \partial U/\partial\lambda \rangle_\lambda$, were interpolated with cubic splines then integrated.[106,107] Here, U is the energy function (including lattice or reaction field energies), and the brackets $\langle\rangle_\lambda$ represent an average over an MD simulation with a particular λ value. The energy derivatives were computed with a finite-difference approximation, as the time average of $[U(\lambda + \delta\lambda) - U(\lambda)]/\delta\lambda$. This approximation is valid, since the free energy derivative is found to vary smoothly with the atomic charges, in accord with linear response.[15,93]

For the Abl kinase, we computed the free energy to completely uncharge the Gleevec ligand, following the method used earlier,[82,83] with two different box sizes. A total of 20 runs were carried out and the statistical uncertainties correspond to the deviation over the last 10 runs. For thioredoxin, we computed the protonation free energy of the Asp26 side chain, with the partial charge model used earlier[80] but a different box shape. For Crk, we considered a perturbation due to a test charge, δq, placed successively on each C_α atom and on the following side chain atoms (one at a time): C_β's of Ala, Ser, Val; C_γ's of Asn, Asp, Leu, Met, Pro; C_δ's of Gln, Glu, Ile; C_ζ/N_ζ of Arg, Lys, Phe, Tyr; and $N_{\varepsilon 1}$ of Trp (110 positions in all). Using a 20 ns MD simulation of the unperturbed protein, with an L = 60 or 74 Å box, we estimated the derivative with respect to δq of the free energy for each test charge position.

The PB continuum dielectric calculations with and without PBC were carried out using the PBEQ module[89] of CHARMM[90,91] The optimized atomic Born radii were used.[108] The finite-difference PB calculations used a grid spacing of 0.5 Å. In all cases, the continuum electrostatic free energy was computed as

$$G = \frac{1}{2}\sum_i q_i \phi(\mathbf{r}_i) \tag{26}$$

where $\phi(\mathbf{r})$ is the solution to the finite-difference Poisson−Boltzmann equation.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: thomas.simonson@polytechnique.fr.

*E-mail: roux@uchicago.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Helpful discussions with Bernard Brooks and Othmar Steinhauser are acknowledged. The help of Huan Rui with Figure 7 was greatly appreciated. The authors are grateful for the hard work from the anonymous reviewers. This work was supported by grant MCB-0920261 from the National Science Foundation (Y.-L.L. and B.R.).

REFERENCES

(1) Warshel, A. *Computer Modelling of Chemical Reactions in Enzymes and Solutions*; John Wiley: New York, 1991.

(2) Warshel, A.; Sussman, F.; King, G. Free energy changes in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **1986**, *25*, 8368−8372.

(3) Warshel, A.; Chu, Z.; Parson, W. Dispersed polaron simulations of electron transfer in photosynthetic reaction centers. *Science* **1989**, *246*, 112−116.

(4) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395.

(5) Ulstrup, J. *Charge Transfer Processes in Condensed Media*; Springer: Berlin, 1979.

(6) Warshel, A. Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron transfer and proton transfer studies. *J. Phys. Chem.* **1982**, *86*, 2218−2224.

(7) Marcus, Y. *Ion Solvation*; Wiley: New York, 1985.

(8) Marcus, R. Electron transfer reactions in chemistry: theory and experiment. *Nobel Lectures, Chemistry 1991−1995*; Malmström, B. G., Ed.; World Scientific Publishing Co.: Singapore, 1997.

(9) Hünenberger, P.; Reif, M. *Single-Ion Solvation: Experimental and Theoretical Approaches to Elusive Thermodynamic Quantities*; Royal Society of Chemistry: London, 2011.

(10) Warshel, A.; Sharma, P.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochem. Biophys. Acta* **2006**, *1764*, 1647−1676.

(11) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: the protein−ligand recognition problem. *Acc. Chem. Res.* **2002**, *35*, 430−437.

(12) Simonson, T. Electrostatics and dynamics of proteins. *Rep. Prog. Phys.* **2003**, *66*, 737−787.

(13) Chipot, C.; Mark, A. E.; Pande, V. S.; Simonson, T. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*;

Chipot, C., Pohorille, A., Eds.; Springer Verlag, New York, 2007; Chapter 13.
(14) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(15) Aqvist, J.; Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **1996**, *100*, 9512−9521.
(16) Hummer, G.; Pratt, L.; Garcia, A. Free energy of ionic hydration. *J. Phys. Chem.* **1996**, *100*, 1206−1215.
(17) Hummer, G.; Pratt, L.; Garcia, A.; Berne, B. J.; Rick, S. W. Electrostatic potentials and free energies of solvation of polar and charged molecules. *J. Phys. Chem. B* **1997**, *101*, 3017−3020.
(18) Ashbaugh, H. S.; Wood, R. H. Effects of long-range electrostatic potential truncation on the free energy of ionic hydration. *J. Chem. Phys.* **1997**, *106*, 8135−8139.
(19) Aqvist, J.; Hansson, T. Analysis of Electrostatic potential truncation schemes in Polar Solvents. *J. Phys. Chem. B* **1998**, *102*, 3837−3840.
(20) Hummer, G.; Pratt, L.; Garcia, A.; Garde, S.; Berne, B. J.; Rick, S. W. Reply to comment on "Electrostatic potentials and free energies of solvation of polar and charged molecules". *J. Phys. Chem. B* **1998**, *102*, 3841−3843.
(21) Ashbaugh, H. S.; Sakane, S.; Wood, R. H. Reply to comment on "Electrostatic potentials and free energies of solvation of polar and charged molecules". *J. Phys. Chem. B* **1998**, *102*, 3844−3845.
(22) Darden, T.; Pearlman, D.; Pedersen, L. Ionic charging free energies: spherical versus periodic boundary conditions. *J. Chem. Phys.* **1998**, *109*, 10921−10935.
(23) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1−20.
(24) Kirkwood, J.; Westheimer, F. The electrostatic influence of substituents on the dissociation constant of organic acids. *J. Chem. Phys.* **1938**, *6*, 506−512.
(25) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−129.
(26) King, G.; Warshel, A. A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* **1989**, *91*, 3647−3661.
(27) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(28) Cramer, C.; Truhlar, D. Implicit solvent models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.
(29) Allen, M.; Tildesley, D. *Computer Simulations of Liquids*; Clarendon Press: Oxford, 1991.
(30) Born, M.; Huang, K. *Dynamical Theory of Crystal Lattices*; Clarendon Press: Oxford, 1954.
(31) De Leeuw, S.; Perram, J.; Smith, E. Simulation of electric systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. London* **1980**, *A373*, 27−56.
(32) Nijboer, B. R. A.; Ruijgrok, T. W. On the energy per particle in three- and two-dimensional Wigner lattices. *J. Stat. Phys.* **1988**, *53*, 361−382.
(33) Figueirido, F. E.; Levy, R. On finite-size effects in computer simulations using the Ewald potential. *J. Chem. Phys.* **1995**, *103*, 6133−6142.
(34) Darden, T. In *Computational Biochemistry & Biophysics*; Becker, O., Mackerell, A., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, New York, 2001; Chapter 4.
(35) Barker, J.; Watts, R. Monte Carlo studies of the dielectric proterties of water-like models. *Mol. Phys.* **1973**, *26*, 789−792.
(36) van Gunsteren, W.; Berendsen, H.; Rullmann, J. Inclusion of reaction fields in molecular dynamics: application to liquid water. *Faraday Discuss. Chem. Soc.* **1979**, *66*, 58−70.
(37) Wilson, M. A.; Pohorille, A.; Pratt, L. R. Comment on "Study on the liquid-vapor interface of water. I. Simulation results of thermodynamic properties and orientational structure. *J. Chem. Phys.* **1989**, *90*, 5211−5213.
(38) Lamoureux, G.; Roux, B. Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field. *J. Phys. Chem. B* **2006**, *110*, 3308−3322.
(39) Feller, S.; Pastor, R.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. Effect of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J. Phys. Chem.* **1996**, *100*, 17011−17020.
(40) Harder, E.; Roux, B. On the origin of the electrostatic potential difference at a liquid-vacuum interface. *J. Chem. Phys.* **2008**, *129*, 234706.
(41) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(42) Jackson, J. *Classical Electrodynamics*; Wiley: New York, 1975.
(43) Berendsen, H.; Postma, J.; van Gunsteren, W.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1981.
(44) Kastenholz, M. A.; Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids. *J. Chem. Phys.* **2006**, *124*, 224501.
(45) Kastenholz, M. A.; Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J. Chem. Phys.* **2006**, *124*, 224501.
(46) Leung, K.; Rempe, S. B.; von Lilienfeld, O. A. Ab initio molecular dynamics calculations of ion hydration free energies. *J. Chem. Phys.* **2009**, *130*, 204507.
(47) Kathmann, S. M.; Kuo, I.; Mundy, C. J.; Schenter, G. K. Understanding the Surface Potential of Water. *J. Phys. Chem. B* **2011**, *115*, 4369−4377.
(48) Ashbaugh, H. S. Convergence of molecular and macroscopic continuum descriptions of ion hydration. *J. Phys. Chem. B* **2000**, *104*, 7235−7238.
(49) Asthagiri, D.; Pratt, L.; Ashbaugh, H. S. Absolute hydration free energies of ions, ion-water clusters, and quasichemical theory. *J. Chem. Phys.* **2003**, *119*, 2702−2708.
(50) Simonson, T.; Archontis, G.; Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B* **1997**, *101*, 8349−8362.
(51) Nina, M.; Simonson, T. Molecular dynamics of the $tRNA^{Ala}$ acceptor stem: comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.
(52) Woo, H. J.; Dinner, A.; Roux, B. Grand canonical Monte Carlo simulation of water in protein environments. *J. Chem. Phys.* **2004**, *121*, 6392−6400.
(53) Born, M. Volumen und hydrationswärme der ionen. *Zeit. Phys.* **1920**, *1*, 45−48.
(54) Godement, R. *Analyse Mathématique I*; Springer: Berlin, 1998.
(55) Landau, L.; Lifschitz, E. *Electrodynamics of Continuous Media*; Pergamon Press: New York, 1980.
(56) Figueirido, F. E.; Del Buono, G. S.; Levy, R. On finite-size corrections to the free energy of hydration. *J. Phys. Chem. B* **1997**, *101*, 5622−5623.
(57) Sakane, S.; Ashbaugh, H. S.; Wood, R. H. Continuum electrostatics to the polarization and thermodynamic properties of Ewald sum simulations for ions and ion pairs at infinite dilution. *J. Phys. Chem. B* **1998**, *102*, 5673−5682.
(58) Hunenberger, P.; McCammon, J. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: A continuum electrostatics study. *Biophys. Chem.* **1999**, *78*, 69−88.
(59) Bogusz, S.; Cheatham, T. E.; Brooks, B. R. Removal of pressure and free energy artefacts in charged periodic systems via net charge corrections to the Ewald potential. *J. Chem. Phys.* **1998**, *108*, 7070−7084.

(60) Boresch, S.; Ringhofer, S.; Höchtl, P.; Steinhauser, O. Towards a better understanding of biomolecular solvation. *Biophys. Chem.* **1999**, *78*, 43−68.
(61) Allen, T.; Andersen, O.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117−122.
(62) Allen, T.; Andersen, O.; Roux, B. Molecular dynamics - potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys. Chem.* **2006**, *124*, 251−267.
(63) Allen, T.; Andersen, O.; Roux, B. Ion permeation through a narrow channel: using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields. *Biophys. J.* **2006**, *90*, 3447−3468.
(64) Hummer, G.; Pratt, L.; Garcia, A. Multistate gaussian model for electrostatic solvation free energies. *J. Am. Chem. Soc.* **1997**, *119*, 8523−8527.
(65) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.
(66) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 3705−3723.
(67) Reif, M. M.; Oostenbrink, C. Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. *J. Chem. Theory Comput.* **2013**, in press.
(68) Jorgensen, W.; Buckner, K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(69) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(70) Hummer, G.; Pratt, L.; Garcia, A. Ion sizes and finite-size corrections for ionic-solvation free energies. *J. Phys. Chem. B* **1997**, *101*, 9275−9277.
(71) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774−786.
(72) Horvath, L.; Beu, T.; Manghi, M.; Palmeri, J. The vapor-liquid interface potential of (multi)polar fluids and its influence on ion solvation. *J. Chem. Phys.* **2013**, *138*, 154702.
(73) Beck, T. The influence of water interfacial potentials on ion hydration in bulk water and near interfaces. *Chem. Phys. Lett.* **2013**, *561*, 1−13.
(74) Grossfield, A.; Ren, P.; Ponder, J. W. Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.* **2003**, *125*, 15671−15682.
(75) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pK_a values in proteins. *Proteins* **2011**, *79*, 3260−3275.
(76) Simonson, T.; Archontis, G.; Karplus, M. A Poisson-Boltzmann study of charge insertion in an enzyme active site: the effect of dielectric relaxation. *J. Phys. Chem. B* **1999**, *103*, 6142−6156.
(77) Lu, X.; Cui, Q. Charging free energy calculations using the Generalized Solvent Boundary Potential (GSBP) and periodic boundary condition: a comparative analysis using ion solvation and oxidation free energy in proteins. *J. Phys. Chem. B* **2012**, *117*, 2005−2018.
(78) Im, W.; Bernèche, S.; Roux, B. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(79) Brunsteiner, M.; Boresch, S. Influence of the treatment of electrostatic interactions on the results of free energy calculations of dipolar systems. *J. Chem. Phys.* **2000**, *112*, 6953−6955.
(80) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.
(81) Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pK_a shifts for Asp26/Asp20 in thioredoxin. *Biophys. J.* **2008**, *95*, 1080−1088.
(82) Lin, Y. L.; Meng, Y.; Jiang, W.; Roux, B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 1664−1669.
(83) Lin, Y. L.; Roux, B. Computational analysis of the binding specificity of Gleevec to Abl, c-Kit, Lck, and c-Src tyrosine kinases. *J. Am. Chem. Soc.* **2013**, *135*, 14741−14753.
(84) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(85) Aleksandrov, A.; Simonson, T. A molecular mechanics model for imatinib and imatinib:kinase binding. *J. Comput. Chem.* **2010**, *31*, 1550−1560.
(86) Aksimentiev, A.; Schulten, K. Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* **2005**, *88*, 3745−3761.
(87) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(88) Gunner, M. R.; Saleh, M.; Cross, E.; ud-Doula, A.; Wise, M. Backbone dipoles generate positive potentials in all proteins: origins and implications of the effect. *Biophys. J.* **2000**, *78*, 1126−1144.
(89) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(90) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(91) Brooks, B.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(92) Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794−802.
(93) Simonson, T. Gaussian fluctuations and linear response in an electron transfer protein. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6544−6549.
(94) van den Bosch, M.; Swart, M.; Snijders, J.; Berendsen, H.; Mark, A.; Oostenbrink, C.; van Gunsteren, W.; Canters, G. Calculation of the redox potential of the protein azurin and some mutants. *ChemBioChem.* **2005**, *6*, 738−746.
(95) Parson, W. W.; Warshel, A. In *Biophysical Techniques in Photosynthesis Research II*; Matysik, J., Aartsma, T. J., Eds.; Springer: Dordrecht, The Netherlands, 2006; p 000.
(96) Aleksandrov, A.; Thompson, D.; Simonson, T. Alchemical free energy simulations for biological complexes: powerful but temperamental. *J. Mol. Recognit.* **2010**, *23*, 117−127.
(97) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290−6295.
(98) Nose, S. A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **1984**, *81*, 511−519.
(99) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(100) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(101) Höchtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. Rationalization of the dielectric properties of common three-site

water models in terms of their force field parameters. *J. Chem. Phys.* **1998**, *109*, 4927−4937.
(102) Mackerell, A. D.; et al. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(103) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(104) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(105) Lu, N.; Woolf, T. B. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer Verlag: New York, 2007; Chapter 6.
(106) Simonson, T. In *Computational Biochemistry & Biophysics*; Becker, O., Mackerell, A., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001; Chapter 9.
(107) Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations II: Improvements for Thermodynamic Integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(108) Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.