JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

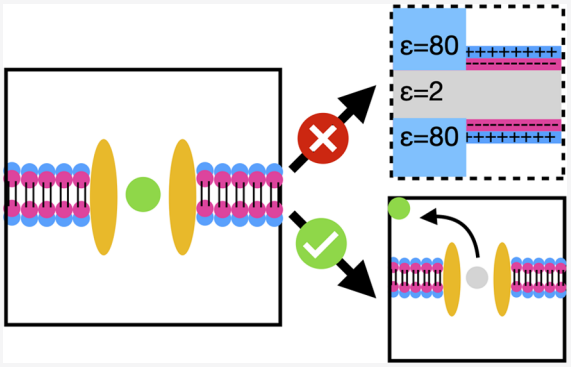
Correction Schemes for Absolute Binding Free Energies Involving Lipid Bilayers

Zhiyi Wu* and Philip C. Biggin*

Cite This: J. Chem. Theory Comput. 2022, 18, 2657−2672 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Absolute binding free-energy (ABFE) calculations are playing an increasing role in drug design, especially as they can be performed on a range of disparate compounds and direct comparisons between them can be made. It is, however, especially important to ensure that they are as accurate as possible, as unlike relative binding free-energy (RBFE) calculations, one does not benefit as much from a cancellation of errors during the calculations. In most modern implementations of ABFE calculations, a particle mesh Ewald scheme is typically used to treat the electrostatic contribution to the free energy. A central requirement of such schemes is that the box preserves neutrality throughout the calculation. There are many ways to deal with this problem that have been discussed over the years ranging from a neutralizing plasma with a post hoc correction term through to a simple co-alchemical ion within the same box. The post hoc correction approach is the most widespread. However, the vast majority of these studies have been applied to a soluble protein in a homogeneous solvent (water or salt solution). In this work, we explore which of the more common approaches would be the most suitable for a simulation box with a lipid bilayer within it. We further develop the idea of the so-called Rocklin correction for lipid-bilayer systems and show how such a correction could work. However, we also show that it will be difficult to make this generalizable in a practical way and thus we conclude that the use of a "co-alchemical ion" is the most useful approach for simulations involving lipid membrane systems.



INTRODUCTION

The accurate calculation of the free energy (FE) of many processes such as ligand binding,[1] change of protonation state,[2] or the influence of mutations[3] is a major focus of modern computational biochemistry.[4] The current state-of-art approach is to construct a periodic computational box where the protein is solvated in explicit water molecules described using a molecular mechanics (MM) force field and to perform alchemical transformation or path sampling to obtain the desired properties. The long-range electrostatics are usually calculated with lattice-sum methods like particle mesh Ewald (PME).[5] However, PME demands the simulation box to be neutral. Thus, for the annihilation/decoupling of charged ligands during binding free-energy calculations, protonation free-energy calculations, or protein mutation calculations where the net charge of the simulation box is perturbed, this has to be dealt with in some way and should not be ignored.[6]

Various solutions to this problem have been suggested. The most extensively employed strategy is where a neutralizing plasma is evenly distributed throughout the simulation box to ensure the overall neutrality is maintained. Though such a plasma does indeed maintain the neutrality of the simulation box, it generates a size-dependent artifact. This size-dependent artifact exists in free-energy estimates of ligand binding free-energy calculations involving changes of charge,[7] charge-perturbing protein mutations,[8] and protonation free-energy calculations.[9]

To tackle this finite-size effect, many approaches can be taken, ranging from ignoring charge-changing mutations,[10] calculating an explicit correction[11] through to incorporating a co-alchemical ion to counteract directly the charge effect.[12]

PME is used widely to compute long-range electrostatics, and the finite-size effect has been well characterized in such calculation systems (see Rocklin et al.[11] and references therein). Thus, one solution would be to avoid using PME. A possible alternative in this regard is the reaction field (RF) method[13] or modified RF methods, where the cutoff is based on charge groups instead of atoms[14] or the anisotropic RF method.[15] All of these methods have thus far been shown to provide comparable performance to PME without the finite-size effect. Though RF methods might work well for homogeneous systems, such as a protein in a pure water solvent, care must be taken for

Received: December 13, 2021
Published: March 22, 2022

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

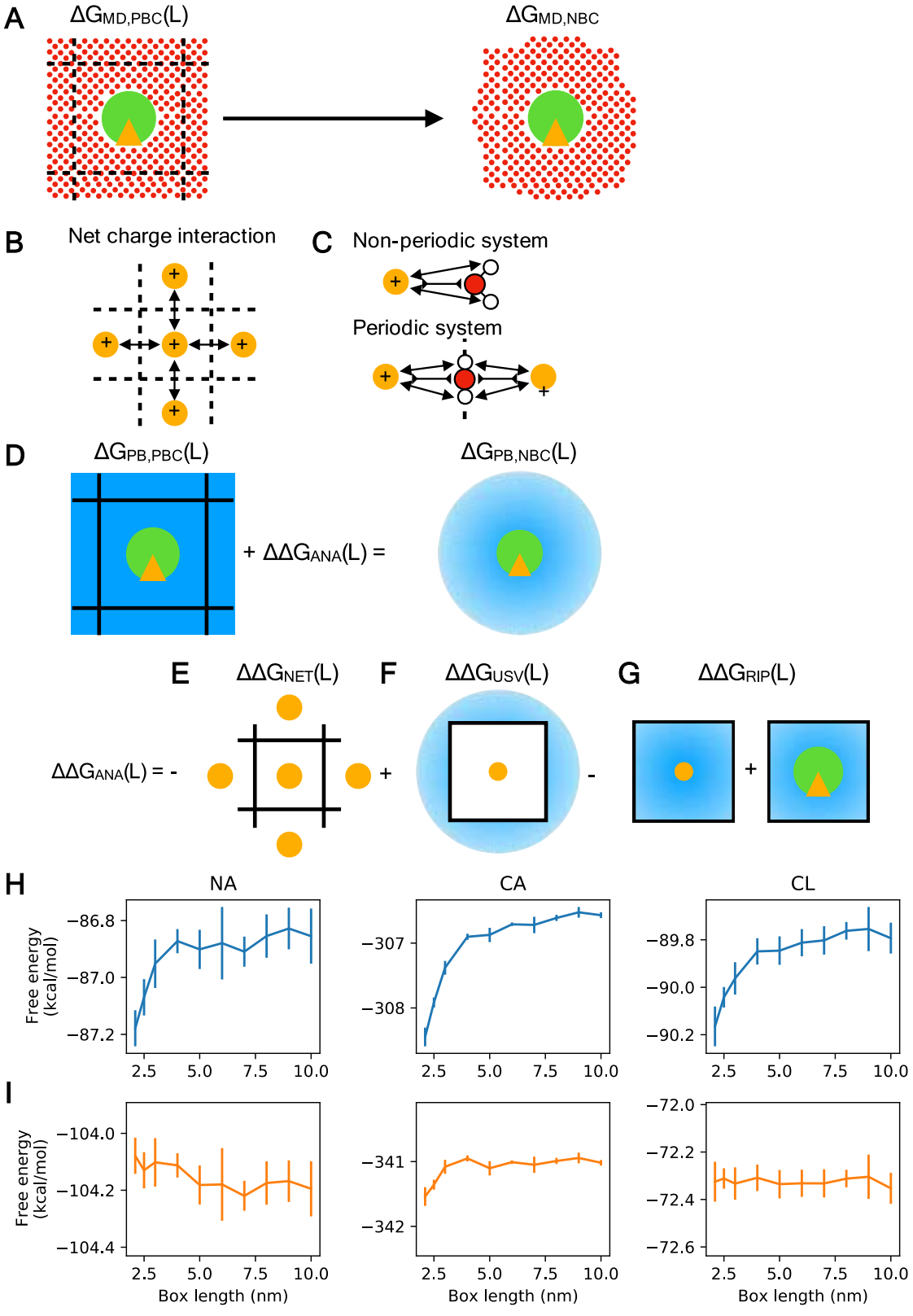


Figure 1. Rocklin correction for a single ion in water. (A) The Rocklin correction defines a semianalytic correction that converts a periodic boundary condition system to a nonperiodic condition system with infinite volume. (B) Under periodic boundary conditions, the solute will interact electrostatically with its periodic neighbor. (C) The presence of the periodic neighbor will affect how the solvent is orientated. (D) After a discrete solvent correction (DSC), a semianalytic correction is used to transform the periodic Poisson−Boltzmann system to a nonperiodic Poisson−Boltzmann system. (E) Periodicity-induced net charge interaction (NET) correction. (F) Periodicity-induced net charge undersolvation (USV) correction. (G) Residual integrated potential (RIP) correction. (H) The uncorrected charge annihilation free energy of the ions Na^+, Cl^-, and Ca^{2+} displays size dependency. (I) The same size dependency is eliminated when the Rocklin correction is applied.

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

inhomogeneous conditions such as a lipid−water system where the lipid hydrophobic core has a very different dielectric constant compared with bulk water. In such instances, PME may well be a better choice.[16]

Another alternative approach is to use force-switching electrostatics, which has been shown to perform better than PME in terms of dealing with a charged simulation box.[17] For some special cases, such as computing the binding energy of ligand, pure quantum mechanics (QM)[18] or QM/MM[19] Hamiltonians are also considered as ways to avoid the size-dependent artifact. On the other hand, electrostatic interactions computed with Poisson−Boltzmann (PB) or generalized-Born (GB), where nonperiodic bound conditions are assumed, can also give results free from finite-size effects.[20]

For ligand binding free-energy calculations, one can obtain the free energy of moving the ligand from the protein to the solvent via double-decoupling/annihilation methods, where the free energy of decoupling/annihilating the ligand from the protein and water is computed separately and the difference is the binding free energy. Thus, when the ligand is charged, the decoupling/annihilation of the ligand will remove charge from the simulation system and give rise to a finite-size effect. One solution for this specific problem would be physically moving the ligand out of the protein into the water phase. Since the ligand is always in the same box during this transition, the charge of the box stayed neutral. The free energy of binding can then be recovered by path-sampling techniques[21] such as the attach−pull−release scheme.[22]

When path-sampling techniques are used to compute the ligand binding free energy, the starting point of the simulation is the ligand bound to the protein and the end point is the ligand in the bulk solvent, quite remote from the protein. To compute the free energy, the whole physical path needs to be sampled. Thus, the calculations are very expensive, especially if the path is long. One solution is to preserve the start and end points and bridge them with alchemical transformations. For ligand binding free-energy calculations, the ligand is decoupled/annihilated in the protein and is coupled/created in the solvent phase at the same time to keep the box neutral.[23]

This approach has been generalized to other types of calculations and is sometimes also referred to as the double-system/single-box approach,[24,25] where the whole cycle of the alchemical transformations is done in the same box so that the charge is always conserved. The approach has been used to understand the effect of protein mutation on the stability of the protein, where the original residue in the protein is mutated to the target residue and the target residue in a tripeptide (to mimic the unfolded protein) is mutated back to the original residue.[3] A similar approach has been used to investigate the effect of protein mutation on the stability of a protein dimer, where the original residue in the dimer is mutated to the target residue and the target residue in the monomer protein is mutated back to the original residue.[26] Although this approach is effective, it requires a much larger simulation box, which is not computationally efficient.

Perhaps, the simplest approach designed to account for the change of net charge during an alchemical transformation is to employ an additional "co-alchemical ion", which changes its charge at the same time as the main perturbation such as to keep the box neutral overall. The simplicity of the implementation of the co-alchemical ion makes it the ideal solution to automatic workflows, which are being increasingly developed and employed.[27,28]

Overall, the strategy that has been most commonly adopted has been the neutralizing plasma approach that employs a postsimulation correction scheme. The simplest scheme is a correction derived for a naked point charge in a continuum medium.[29] However, given that for modern simulations, solvents are usually represented as discrete molecules and the protein systems are too complex to be represented as a continuum, Rocklin expanded the scheme and used adaptive Poisson−Boltzmann solver (APBS) calculations to account for the difference between the protein and a continuum medium.[11] This extended scheme, commonly referred to as the "Rocklin correction", has been used in many studies,[24,30] and its accuracy has been verified by other groups.[31] For soluble proteins, the co-alchemical ion approach and the Rocklin correction give similar results.[32] The finite-size effect has also been seen in simulations using polarizable force fields (AMOEBA),[33] and the Rocklin correction has been shown to be able to correct for that.[34] Similarly, electrostatic interactions computed using multipole methods exhibit finite-size effects and these can also be corrected with the Rocklin correction.[35] Given the complexity of the correction, a simplified version is sometimes used in automatic workflows.[36]

However, nearly all calculations to date have been performed with soluble proteins and it remains unclear as to how well these corrections can be applied to membrane protein systems. It has already been shown that the neutralizing plasma can affect the interpretation of membrane systems.[37] Though the Rocklin correction has been shown to improve the result of membrane systems,[38] the original derivation does not consider nonwater solvent and it is unclear as to how to incorporate components like lipid membranes. In other simulations where the bulk solvent is not water, path sampling has been used to avoid finite-size effects.[39]

In this study, we explore which of the various methods outlined above would offer the best performance for membrane simulation boxes. As part of this process, we derived a new postsimulation correction scheme, similar to the original Rocklin scheme. However, our results suggest that the co-alchemical ion approach may be the preferred route forward.

RESULTS

Rocklin Correction in the Case of a Single Ion in Water.

To solve the finite-size artifact during charge-changing free-energy calculations, Rocklin derived a semianalytic correction scheme. The correction converts a periodic boundary condition (PBC) system with box length L (PBC(L)) into a nonperiodic system of infinite size (e.g., a macroscopic droplet) (Figure 1A) so as to remove the size dependency of the system. The correction can be decomposed into two components. The first is the interactions between the system of interest and its periodic neighbors. In periodic systems, when there is a net charge associated with the simulation box, the net charge will be exerting significant electrostatic interactions to the solute in neighboring boxes (Figure 1B) that are not present in the nonperiodic system. The second component is that the solvent will interact with the net charge differently in the nonperiodic system compared with that in the periodic system. In a nonperiodic system, the electrostatic potential (ESP) generated by a point charge would only vanish at an infinite distance. Thus, all of the solvent molecules will reorient in response to this electrostatic potential (Figure 1C). In periodic systems, on the other hand, the electrostatic potential would be set to zero at the

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

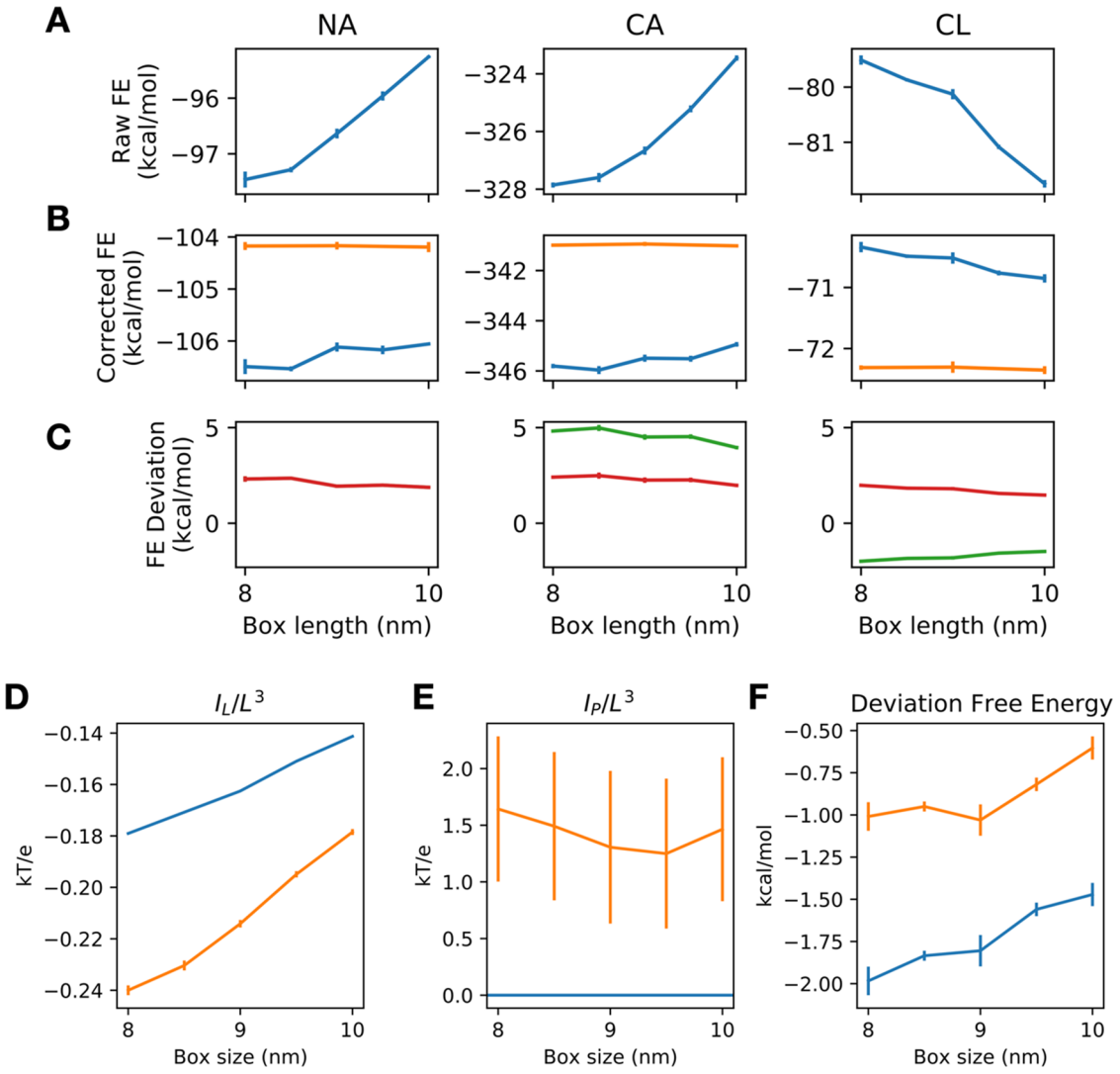


Figure 2. Effect of the Rocklin correction on the annihilation free energy of a single ion in the membrane water system. (A) The uncorrected annihilation free energy (FE) of chloride, sodium, and calcium ion has strong size dependency. (B) The annihilation free energy of a single ion in a lipid membrane with the Rocklin correction applied (blue line) converges toward the reference annihilation free energy of a single ion in pure solvent (orange line). (C) The deviation of the corrected annihilation free energy from the reference free energy (green line), when divided by the charge of the ion, gives a similar profile (red line). (D) The average integrated potential of the ligand in the absence (blue line) and the presence (orange line) of an explicit representation of the lipid. (E) The average integrated potential of the protein in the presence (orange line) and absence (blue line) of the lipid. (F) The deviation from the reference free energy (blue line) shrinks when the lipid is represented explicitly.

boundary due to the presence of the periodic neighbor and the solvent will have different orientations.

To correct the free energy derived from periodic conditions, $\Delta G_{\mathrm{MD,PBC}}(L)$, Rocklin proposed two different correction terms as shown in eq 1: the discrete solvent correction (DSC) to correct for the solvent effect $\Delta\Delta G_{\mathrm{DSC}}(L)$ and an analytical (ANA) correction $\Delta\Delta G_{\mathrm{ANA}}(L)$ to correct for net charge solute interactions. Together, these result in the free energy under non-PBC conditions $\Delta G_{\mathrm{MD,NBC}}$

$$\Delta G_{\mathrm{MD,PBC}}(L) + \Delta\Delta G_{\mathrm{DSC}}(L) + \Delta\Delta G_{\mathrm{ANA}}(L) = \Delta G_{\mathrm{MD,NBC}} \quad (1)$$

During a simulation, the instantaneous solvent rearrangements can give rise to large energy fluctuations. Thus, to obtain the ensemble energy of the system, a large number of frames need to be taken into account. The discrete solvent correction transforms the system from having explicit water molecules into a continuum-electrostatics Poisson−Boltzmann (PB) model, which avoids the necessity of applying the correction to multiple frames, as summarized in eq 2

$$\Delta G_{\mathrm{PB,PBC}}(L) + \Delta\Delta G_{\mathrm{ANA}}(L) = \Delta G_{\mathrm{PB,NBC}} \quad (2)$$

The analytical correction $\Delta\Delta G_{\mathrm{ANA}}(L)$ then transforms the periodic PB model $\Delta G_{\mathrm{PB,PBC}}(L)$ into a PB model in a nonperiodic box $\Delta G_{\mathrm{PB,NBC}}$ (Figure 1D). To achieve this transformation, two steps are taken (eq 3): the charge interactions between periodic neighbors are removed by $\Delta\Delta G_{\mathrm{NET}}(L)$ and the polarization effect of the net charge on the medium outside the simulation box can be added by $\Delta\Delta G_{\mathrm{USV}}(L)$

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

$$\Delta G_{\mathrm{PB,PBC}}(L) + \Delta\Delta G_{\mathrm{NET}}(L) + \Delta\Delta G_{\mathrm{USV}}(L) + \Delta\Delta G_{\mathrm{RIP}}(L) + \Delta\Delta G_{\mathrm{EMP}}(L) = \Delta G_{\mathrm{PB,NBC}} \quad (3)$$

The $\Delta\Delta G_{\mathrm{NET}}(L)$ is the periodicity-induced net charge interaction (NET) correction, the $\Delta\Delta G_{\mathrm{USV}}(L)$ is the periodicity-induced net charge undersolvation, the $\Delta\Delta G_{\mathrm{RIP}}(L)$ is the residual integrated potential correction, and the $\Delta\Delta_{\mathrm{GEMP}}(L)$ is an empirical correction.

The leading term of the charge interactions between the periodic neighbors is the net charge interaction, which is corrected with a periodicity-induced net charge interaction (NET) correction $\Delta\Delta G_{\mathrm{NET}}(L)$ (Figure 1E). In $\Delta\Delta G_{\mathrm{NET}}(L)$, the net charge of the system is approximated with a point charge at the center of the simulation box and $\Delta\Delta G_{\mathrm{NET}}(L)$ corrects for the self-interactions between charged species across the periodic boundaries in vacuum.

The polarization effect of the net charge on an infinite medium is corrected for by the periodicity-induced net charge undersolvation term $\Delta\Delta G_{\mathrm{USV}}(L)$ (Figure 1F), which corrects for the solvation of the charged species disrupted by the periodic boundary. Both $\Delta\Delta G_{\mathrm{NET}}(L)$ and $\Delta\Delta G_{\mathrm{USV}}(L)$ are calculated analytically, assuming that the charged species is a naked point charged (naked means no excluded volume) centered in a box of water. The difference between the real system and a point charge is corrected with the residual integrated potential correction ($\Delta\Delta G_{\mathrm{RIP}}(L)$), which performs a PB calculation to derive the average electrostatic potential difference between the solute and the point charge, where the average electrostatic potential of the point charge is computed analytically (Figure 1G).

To illustrate the effect of the Rocklin correction, we computed the charge annihilation free energy of a single ion in a box of water. The uncorrected charge annihilation free energy exhibits size dependency (Figure 1H), while the corrected free energy does not (Figure 1I).

Systems with a Lipid Bilayer. As shown above, the Rocklin scheme works very nicely for a box of homogeneous solvent. However, it is currently unclear as to how one should proceed in a nonhomogeneous environment such as a simulation box with a lipid bilayer present. Though many papers have highlighted the specific interactions between the lipid head group and the protein,[40] lipids are usually included in the simulation to provide the necessary hydrophobic environment to accommodate transmembrane proteins. One could argue that the hydrophobic region would have very little effect on the correction terms that purely deal with electrostatic interactions. Thus, in the first instance, we investigated whether simply ignoring the lipid bilayer would give correct results by investigating a test ion in a lipid−water system to check the accuracy of the correction when the membrane is not taken into account.

Similar to the charge annihilation free energy of a single ion in a box of water, the charge annihilation free energy of a single ion in a lipid−water system exhibits a strong size-dependent effect (Figure 2A). When the Rocklin correction is applied assuming the lipid has no excluded volume and no partial charge, the size-dependent effect is mostly corrected (Figure 2B). However, the corrected free energy still exhibits a small size dependency and is some distance away from the reference charge annihilation free energy (calculated from the charge annihilation free energy of a single ion in water) (Figure 2B). The deviation from the reference free energy is proportional to the net charge of the ion and converges toward 0 with a bigger box size (Figure 2C). Thus, this route, although simple, does not provide a proper route to correcting the electrostatics.

Including the Lipid Bilayer in the Residual Integrated Potential Calculations. Having shown that ignoring the lipid bilayer would give a sizable deviation from the reference value, we tried to incorporate the role of the lipid bilayer into the Rocklin correction. The Rocklin correction corrects the periodicity-induced artifact by deriving an analytical solution ($\Delta\Delta G_{\mathrm{NET}}(L)$, $\Delta\Delta G_{\mathrm{USV}}(L)$) to correct for the ideal case where the system is represented as a naked charge centered in the box representing the net charge of the system. An additional residual integrated potential $\Delta\Delta G_{\mathrm{RIP}}(L)$ is then used to compute the difference between the real system and the naked charge.

The rationale behind $\Delta\Delta G_{\mathrm{RIP}}(L)$ is that the major difference between a periodic system computed using a lattice-sum method and nonperiodic system computed using Coulombic equations is that in a periodic system, the average electrostatic potential of the simulation box is constrained to zero, whereas such a constraint is not present in nonperiodic systems. Thus, to transform the system to a nonperiodic condition, we need to obtain the energy of charging the system in nonperiodic systems, which is the product of the average electrostatic potential, computed as the integrated electrostatic potential of the simulation box over the box volume (I/L^3) and the net charge of the system (Q), as shown in eq 4

$$\Delta\Delta G_{\mathrm{RIP}} = [(I_{\mathrm{P}} + I_{\mathrm{L}})(Q_{\mathrm{P}} + Q_{\mathrm{L}}) - I_{\mathrm{P}}Q_{\mathrm{P}}]/L^3 \quad (4)$$

For the specific case of ligand binding free-energy calculations, the correction is computed as the difference of the charging energy between the apo protein and the protein−ligand complex. The charging energy of the protein−ligand complex is the product of the net charge of the complex ($Q_{\mathrm{P}} + Q_{\mathrm{L}}$) and the average electrostatic potential of the complex ($(I_{\mathrm{PL}})/L^3$). Rocklin applied an approximation that $I_{\mathrm{PL}} = I_{\mathrm{P}} + I_{\mathrm{L}}$, where I_{P} and I_{L} are the integrated potential computed using adaptive Poisson−Boltzmann solver (APBS) calculations with both ligand and protein as excluded volumes. The I_{P} is computed with protein having partial charges, while I_{L} is computed with ligand having partial charges. Thus, the sum of I_{P} and I_{L} would be I_{PL} and I_{P} can be reused during the apo protein calculation, which is the product of the average electrostatic potential of the protein (I_{P}/L_3) and the net charge of the protein (Q_{P}).

The simplest approach of incorporating the effect of the lipid would be to consider the lipid as part of the protein. Thus, the lipid would form part of the excluded volume in the I_{L} and I_{P} calculations and contribute electrostatically to the I_{P} calculation. If we consider the example of the charging free energy of a single ion in a membrane water system, when the lipid is absent, the I_{P} term will be 0. If the lipid is included in the APBS calculation in the same manner as the protein, the lipid contributes to the APBS calculations as individual atoms with a partial charge. These contributions for I_{L} and I_{P} deviate from the case where lipid is not taken into account (Figure 2D,E). The deviation in I_{L}/L_3 is very small (~0.05 kT/e), showing that treating the lipid bilayer as a hydrophobic slab has little impact on the correction. On the other hand, the deviation in I_{P}/L^3 is sizable, showing that the lipid bilayer drastically changes the electrostatic environment through charge interactions. It is also worth noting that due to the frame-to-frame fluctuation of the lipid, a standard deviation of ~0.2 kT/e is observed for I_{P}/L^3, showing that multiple frames are required to obtain an accurate estimate of I_{P}/L^3. By using the new I_{P}/L^3 and I_{L}/L^3 in the $\Delta\Delta G_{\mathrm{RIP}}(L)$

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

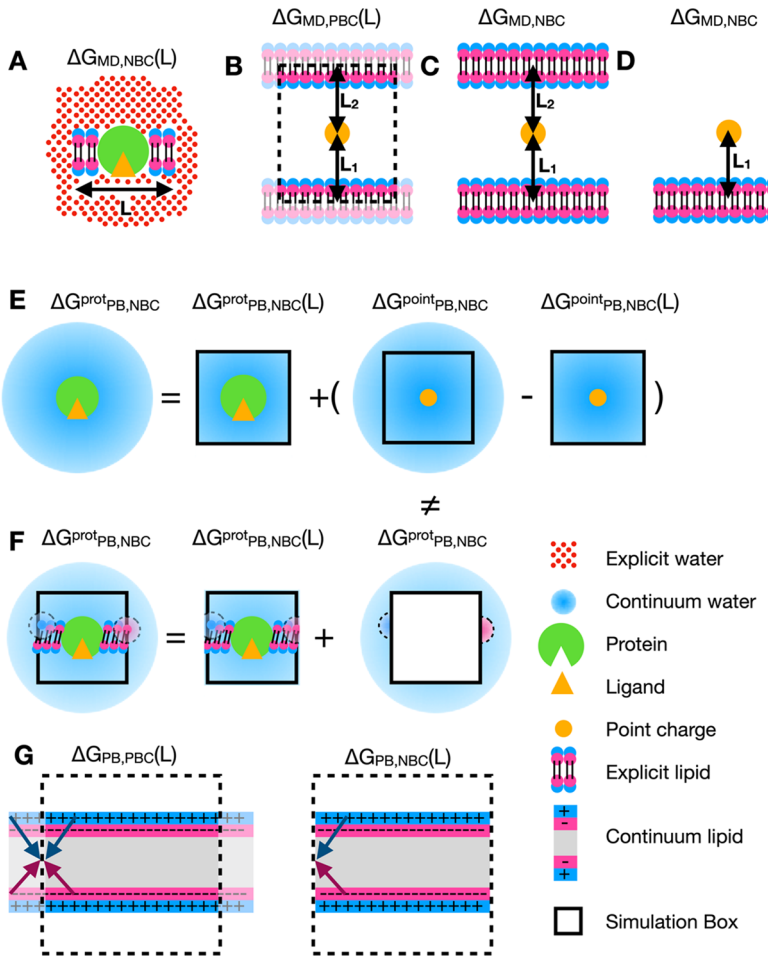


Figure 3. Problems with defining the nonperiodic boundary condition of a lipid membrane system. (A) Defining the nonperiodic boundary condition for a membrane protein system as a membrane patch floating in a macroscopic droplet might not be a good representation. (B) In periodic boundary condition systems, a single membrane could sandwich the ion (orange circle). The nonperiodic boundary condition could be defined as an ion sandwiched by two infinite membranes (C) or the case where only the membrane closest to the ion is preserved (D). (E) The nonperiodic boundary condition for a soluble protein in a continuum solvent can be defined as the sum of the simulation box calculated explicitly with APBS calculations and the space beyond the simulation box calculated analytically, assuming the protein is a naked point charge. (F) The same procedure cannot be performed for membrane proteins as the effect of the lipid extends beyond the simulation box and cannot be approximated as a naked point charge. (G) The lipids outside the box will also affect the lipids in the simulation box, which makes the Poisson−Boltzmann calculation of the simulation box different from the case under periodical boundary conditions.

calculations, the deviation from the reference value is reduced (Figure 2F), but there is still considerable residual deviation. However, it is worth noting that this may be an overestimate of bias in the uncorrected result because even ABFE, via the double-decoupling nature, does have some degree of error cancellation.

Defining the Nonperiodic Condition for Membrane Systems. Though the inclusion of an explicit lipid bilayer lowers the deviation, a size-dependent artifact is still present, indicating scope for improvement.

In the work of Rocklin et al., they solved the finite-size artifact problem by applying a correction that transforms the system from a periodic system with box length L into a nonperiodic system of infinite size. This transformation is relatively straightforward for a soluble protein as there is very little ambiguity in defining the nonperiodic system (Figure 1A). However, defining this size-independent nonperiodic system can be nontrivial for a membrane protein (Figure 3A), as it is unclear as to how one should represent the membrane. Given the membrane in a periodic system is, by definition, infinite, there should also be an infinite membrane in the nonperiodic system as well.

However, this creates a problem of how to reproduce the position of the net charge with respect to the membrane in the nonperiodic system. In the periodic system, unless the solute is centered exactly at the core of the membrane, the solute is always sandwiched between two membranes, characterized by two distances from both membranes. An extreme case would be the computation of charge annihilation free energy of a single ion in a lipid−water box (Figure 3B). One could argue that two lipid membranes need to be present to reproduce this sandwich effect (Figure 3C). On the other hand, in the periodic system, only one membrane is present in the simulation box, so it might be difficult to map the single-membrane periodic system to the double-membrane nonperiodic system and one should stick with a single membrane in a nonperiodic system (Figure 3D).

The difficulty in defining the nonperiodic condition for lipid bilayer systems poses a challenge during the $\Delta\Delta G_{RIP}(L)$ calculations. The $\Delta\Delta G_{RIP}(L)$ corrects for the nonzero average electrostatic potential of the simulation box in nonperiodic

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

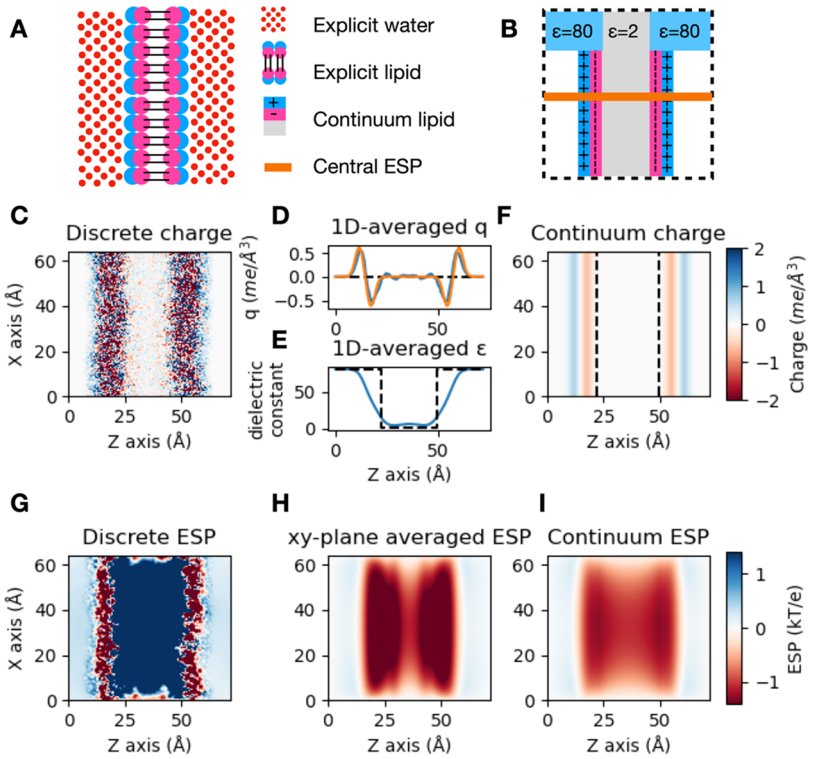


Figure 4. Simple continuum model to represent an explicit lipid. (A) The nonperiodic lipid−water system represented with explicit atoms. (B) The continuum model of the lipid−water system, where the lipid head group is represented by two oppositely charged Gaussian densities and the space is separated into a high dielectric constant zone (water and lipid head groups) and a low dielectric constant zone (hydrophobic core of the lipid). The central ESP is the ESP profile at the center of the lipid along the *z*-axis (orange). (C) The average charge density from APBS calculations with explicit lipid. (D) The average charge density reduced to the *z*-axis (blue) and the fitted charge density with two oppositely charged Gaussian densities (orange). (E) The average dielectric constant reduced to the *z*-axis (blue) and the two-step binary model of high and low dielectric constant regions (dashed black). (F) The separation between high and low dielectric constant regions (dashed black) superimposed on the oppositely charged Gaussian density charge model. (G) The average ESP derived from the APBS calculation with explicit lipid, where a boundary effect can be seen close to the edge. (H) The ESP computed with the average dielectric constant and average charge density from the APBS calculation with explicit lipid. (I) The ESP computed with the continuum lipid model.

conditions. However, no Poisson−Boltzmann solver can compute the electrostatic potential of a macroscopic nonperiodic system. Thus, a key assumption has to be made that the electrostatic potential exerted by a protein would be the same as the electrostatic potential exerted by a naked point charge of the same net charge beyond a certain distance. Thus, the electrostatic potential of a protein system in the nonperiodic condition would be the sum of the electrostatic potential of the protein in a finite-size box and the electrostatic potential of a naked point charge outside this finite-size box (Figure 3E). This assumption holds true for a soluble protein when the protein is at least 1 nm away from the box edge.

It is, however, difficult to argue that this assumption for a soluble protein still holds for a membrane protein. If a single infinite lipid membrane is assumed, lipids will be touching the box edge and the lipid outside the box would exert an electrostatic potential back into the simulation box. This creates a problem both outside and inside the box.

Given that the lipid membrane outside the box is not a homogeneous continuous medium, the original assumption that $\Delta\Delta G_{RIP}(L) + \Delta\Delta G_{USV}(L)$ expands the system to a size-independent nonperiodic system no longer holds. Furthermore, lipids touching the boundary will exert an electrostatic potential outside the box that cannot be captured by an analytical solution (Figure 3F).

In both periodic and nonperiodic systems, the lipid on the box edge will experience electrostatic potential effects from lipids outside the box. However, it is difficult to take into account the atoms outside of the simulation box (Figure 3G). For the simplest case of a single membrane in a water box, this boundary effect can be observed in the hydrophobic regions close to the box edge (Figure 4G). Thus, the average electrostatic potential (ESP) computed using a finite simulation box cannot be approximated as the box of the same size sculpted from an infinite membrane system.

One way of mimicking the infinite membrane in the nonperiodic boundary condition is to perform a Poisson−Boltzmann calculation in a box that is much larger than the simulation box and sculpt the simulation box from the result. However, though in theory this would give the average electrostatic potential of the simulation box in nonperiodic conditions, it raises many problems. As is seen in Figure 2E, lipids exhibit large fluctuation between frames (with a standard deviation of 0.7 kT/e) and, thus, a large number of frames need to be considered to obtain a converged value. Furthermore, the large number of lipid atoms that are required to mimic an infinite membrane make the calculation distinctly unattractive for postsimulation treatment.

Use of a Continuum Model for the Lipid. Given the issues above, we next considered using a continuum model (Figure 4A,B) to replace the explicit lipid as this would make the

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

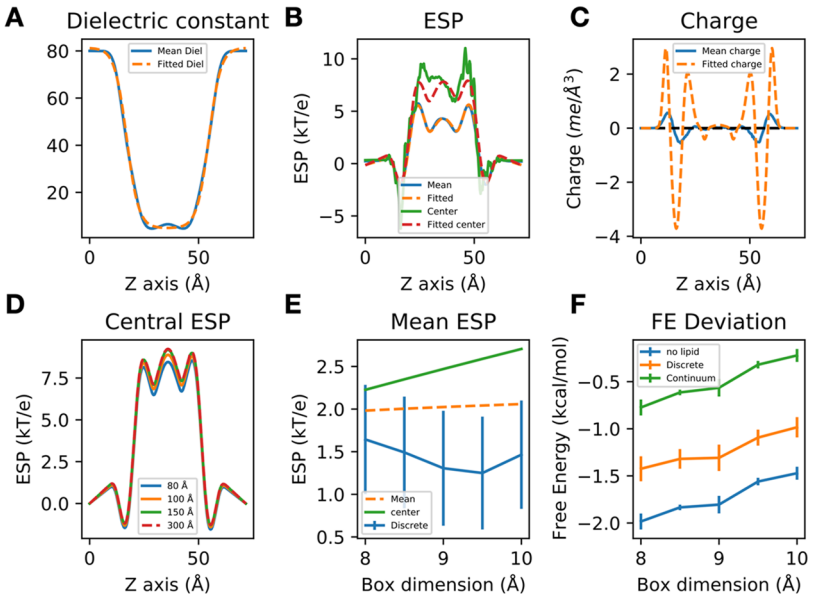


Figure 5. Continuum model reproduces the average ESP and lowers the deviation to the reference free energy. (A) The average dielectric constant profile is fitted with a logistic function. (B) The charge density profile of five Gaussians is fitted (dashed orange line) to reproduce the average ESP computed from APBS calculations with explicit lipid (blue line). The central ESP computed with the continuum model (dashed red) closely matches the result from explicit lipid (green). (C) The fitted charge density bears a similar shape to the average charge density but each Gaussian has a different magnitude. (D) The central ESP is converged when the box dimension reaches 150 Å (green), which is the same as a box dimension of 300 Å (dashed red). (E) The mean ESP computed with the continuum model (orange dashed) is similar to that computed with the explicit lipid (blue). The central ESP with a box dimension of 150 Å gives a different profile (green). (F) The Rocklin correction performed with explicit lipid (orange) would lower the deviation from the reference value compared with not taking the lipid into account (blue), and using the central ESP from the continuum model further lowers the deviation (green).

calculation of a very large lipid feasible and removes the frame-to-frame fluctuation problems. For an infinite lipid membrane, the electrostatic potential would be the same in the plane parallel to the membrane and is a function of the distance to the center of the membrane. This electrostatic potential profile of an infinite lipid membrane could be approximate as the central ESP of a sufficiently large membrane patch along the axis orthogonal to the membrane (Figure 4B).

APBSmem[41] offers a method of modeling the lipid as two regions of different dielectric constants, where the head group has a dielectric constant of 80 (for the lipid head group and the solvent) and the hydrophobic core has a dielectric constant of 2, as summarized in eq 5 and Figure 4B

$$\mathrm{diel}(z) = \begin{cases} 80 & |z - z_{\mathrm{center}}| > l_{\mathrm{hydrophobic}} \\ 2 & |z - z_{\mathrm{center}}| \le l_{\mathrm{hydrophobic}} \end{cases} \tag{5}$$

where $\mathrm{diel}(z)$ is the dielectric constant of the system as a function of its location on the z-axis, z_{center} is the center of the lipid membrane, and $l_{\mathrm{hydrophobic}}$ is the length of the hydrophobic core defined as the distance from the first carbon after glycerol to the end of the aryl tail. Given that the APBS calculation is performed to correct for artifacts arising from molecular mechanics simulations, the dielectric constant of the solvent should match the dielectric constant of the water model (TIP3P) used during the calculation. However, this would mean that the space is segregated into three zones of different dielectric constants, solvent (97 for TIP3P water[42]), lipid head group (80), and lipid hydrophobic core (2). Given that the dielectric constants of the TIP3P water and the lipid head group are quite close, for the simplicity of the calculation, the dielectric constant of the solvent was set to 80 in the following calculations.

Representing the lipid as a hydrophobic slab is insufficient, as is illustrated by large I_{P}/L^3 (~1 kT/e) where the partial charge of the lipid plays a significant role during the APBS calculation as well. Previous work[43] has been done to incorporate the charge effect of the lipid, where the negatively charged phosphate group and the positively charged head group (e.g., choline for phosphatidylcholines) were represented as a pair of ± charge sheets.

A similar continuum lipid model has been constructed to see if a more accurate $\Delta\Delta G_{\mathrm{RIP}}(L)$, which is derived from the average ESP of an infinite lipid membrane, could lower the deviation to the reference charge annihilation free energy. The dielectric constant profile is constructed in the same way using a binary step model (Figure 4E). The charge density of phosphate and choline groups is modeled as Gaussian-shaped charges, as a single sheet of charge might make the calculation sensitive to the box size and grid spacing, as summarized in eqs 6−8

$$g(x, A, \mu, \sigma) = A\, \mathrm{e}^{-1/2(x-\mu)^2/\sigma^2} \tag{6}$$

$$q(z) = g(|z - z_{\mathrm{center}}|, A_{\mathrm{Cho}}, \mu_{\mathrm{Cho}}, \sigma_{\mathrm{Cho}}) - g(|z - z_{\mathrm{center}}|, A_{\mathrm{PO_4}}, \mu_{\mathrm{PO_4}}, \sigma_{\mathrm{PO_4}}) \tag{7}$$

$q(z)$ is the charge density as a function of the position on the z-axis. The $A_{\mathrm{Cho}}, \mu_{\mathrm{Cho}}, \sigma_{\mathrm{Cho}}$ and $A_{\mathrm{PO_4}}, \mu_{\mathrm{PO_4}}, \sigma_{\mathrm{PO_4}}$ are the magnitude, the center, and the spread of the charge density of the choline group and the phosphate group, respectively. The following constrain on the magnitude and spread has been used to ensure that the sum of the charge density would be zero

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

$$A_{\mathrm{Cho}}\sigma_{\mathrm{Cho}} = A_{\mathrm{PO_4}}\sigma_{\mathrm{PO_4}} \tag{8}$$

The parameters of A_{Cho}, μ_{Cho}, σ_{Cho}, $A_{\mathrm{PO_4}}$, $\mu_{\mathrm{PO_4}}$, and $\sigma_{\mathrm{PO_4}}$ were fitted to a small lipid−water model system (64.26 * 67.47 * 71.67 Å) (Figure 4C). Though there was charge density arising from the small dipole of the aryl chain, they were ignored (Figure 4D). The charge density modeled with two Gaussians and the dielectric constant modeled with a binary step model (Figure 4F) were then used to calculate the ESP (Figure 4I). The ESP computed using the continuum model (Figure 4I) deviated a lot from the average ESP computed using explicit atoms (Figure 4G).

This deviation is not unexpected as the lack of explicit screening will affect the computed electrostatic potential. To compensate for the absence of explicit screening, the use of a coarse-grained Martini model preserves the total charge of the residue but reduces the dielectric constant to 15.[44] On the other hand, another strategy, Choe et al.,[43] is to preserve the dielectric constant but scale the charge sheet until the internal potential at the center of the bilayer reaches +300 mV. To confirm that the deviation does indeed come from the lack of explicit screening, the average charge density (Figure 4C) and dielectric constant (Figure 4E) profile from the explicit APBS calculations were used to reconstruct the ESP (Figure 4H), and as expected, it was similar to the result from the continuum model (Figure 4I).

Attempting to reproduce the ESP obtained without explicit screening, the dielectric constant profile was fitted with a logistic function to reproduce the smooth transition from the high dielectric constant to low dielectric constant region, as described by eq 9

$$\sigma(z) = \frac{k_2}{1 + \mathrm{e}^{-k_1(|z-z_{\mathrm{center}}|+b_1)}} + b_2 \tag{9}$$

where k_1, k_2, b_1, and b_2 are fitted to reproduce the dielectric constant profile (Figure 5A), and the fitted parameters are in Table S1.

Five Gaussians were required to reproduce the ESP profile averaged across the x−y-plane (Figure 5B), and the fitted parameters are given in Table S2. The central ESP profile of the discrete model, though not the direct target of the fitting procedure, matches with that of the continuum model (Figure 5B).

The fitted charge density is written as

$$\begin{aligned} q(z) = {} & g(|z - z_{\mathrm{center}}|, A_{\mathrm{Cho}}, \mu_{\mathrm{Cho}}, \sigma_{\mathrm{Cho}}) \\ & - g(|z - z_{\mathrm{center}}|, A_{\mathrm{PO_4}}, \mu_{\mathrm{PO_4}}, \sigma_{\mathrm{PO_4}}) \\ & + g(|z - z_{\mathrm{center}}|, A_{\mathrm{GL}}, \mu_{\mathrm{GL}}, \sigma_{\mathrm{GL}}) \\ & - g(|z - z_{\mathrm{center}}|, A_{\mathrm{C-}}, \mu_{\mathrm{C-}}, \sigma_{\mathrm{C-}}) \\ & + g(|z - z_{\mathrm{center}}|, A_{\mathrm{C+}}, \mu_{\mathrm{C+}}, \sigma_{\mathrm{C+}}) \end{aligned} \tag{10}$$

where the magnitude A, center μ, and spread σ of the choline Cho, phosphate PO_4, ester GL (possibly), and the positive and negative dipoles of the aryl chain C+/C− were allowed to fit freely. The center of the positive dipole of the aryl chain $\mu_{\mathrm{C+}}$ is fixed to 0. The new charge density has a similar shape to the original charge density (Figure 5C), where the magnitude is increased to compensate for the lack of explicit screening.

To obtain the ESP profile of an infinite lipid membrane, the size of the membrane patch was expanded in the xy-plane until the central ESP was converged (Figure 5D), where a membrane patch with dimensions 150 * 150 Å in the xy-plane gives the same central ESP as a membrane patch of 300 * 300 Å.

In terms of calculating $\Delta\Delta G_{\mathrm{RIP}}(L)$, the integrated potential I is computed as

$$I = B_{\mathrm{HET}}[X, L_{\mathrm{ref}}] - B_{\mathrm{HET}}[Q_X, L_{\mathrm{ref}}] \tag{11}$$

where $B_{\mathrm{HET}}[X,L_{\mathrm{ref}}]$ is the integrated potential of the simulation box and is defined as

$$B_{\mathrm{HET}}[X, L_{\mathrm{ref}}] = \int_{L_{\mathrm{ref}}} d^3r\phi_{\mathrm{HET},X}(r) \tag{12}$$

where $\phi_{\mathrm{HET},X}(r)$ is the ESP. $B_{\mathrm{HET}}[Q_X,L_{\mathrm{ref}}]$ is the reference integrated potential computed with a naked point charge in a box with the dimension of L_{ref} and is defined as

$$B_{\mathrm{HET}}[Q_X, L_{\mathrm{ref}}] = \int_{L_{\mathrm{ref}}} d^3r\phi_{\mathrm{HET},Q_X}(r) = \int_{L_{\mathrm{ref}}} d^3r\frac{Q_x}{4\pi\epsilon_0\epsilon_S r} \tag{13}$$

Since the lipid has a total charge of 0, $B_{\mathrm{HET}}[Q_X,L_{\mathrm{ref}}]$ evaluates to 0 in this case.

We transformed the $B_{\mathrm{HET}}[X,L_{\mathrm{ref}}]$ such that it is the ESP at the equivalent position in the z-axis from the center of the membrane $\phi_{\mathrm{HET},X}(z)$. Thus, $B_{\mathrm{HET}}[X,L_{\mathrm{ref}}]$ is defined as

$$B_{\mathrm{HET}}[X, L_{\mathrm{ref}}] = \int_{l_{\mathrm{ref}}} d^3r\phi_{\mathrm{HET},X}(z) \tag{14}$$

The $\Delta\Delta G_{\mathrm{RIP}}(\mathrm{L})$ computed with the center ESP of a large continuum lipid membrane gives very different I_{P}/L^3 compared with the average ESP of the cubic simulation box with the same Z dimension (Figure 5E), and the result from the center ESP lowers the deviation from the reference value (Figure 5F). Note that in terms of computing I_{P}/L^3 from the simulation box, the continuum lipid model gives a similar value compared with the result computed with explicit atoms for lipids (Figure 5E). Also note the contribution from the $\Delta\Delta G_{\mathrm{RIP}}(L)$ is at a scale of ∼1 kcal/mol. During the lipid APBS calculations, the dielectric constant of the water, for simplicity, was set to 80, instead of 97 for TIP3P water. If the dielectric constant was set to 97 instead of 80, we would expect the $\Delta\Delta G_{\mathrm{RIP}}(L)$ to be 1 * 80/97 ≈ 0.82 kcal/mol. Since the difference of ∼0.18 kcal/mol is much smaller than the force field error, which is usually estimated to be ∼1 kcal/mol, we conclude that setting the dielectric constant of the water to 80 is a reasonable approximation.

Treatment of Residual Errors. The previous section described our attempts to solve the problem that a Poisson−Boltzmann calculation done with a finite-size simulation box cannot give the same average ESP as the same box in a nonperiodic condition due to the lack of electrostatic interactions from the lipid outside the box (Figure 3G). However, the undersolvation of the ion $\Delta\Delta G_{\mathrm{USV}}(L)$ by the lipid outside the simulation box is still not properly accounted for (see Figure 3F). This can be addressed by deriving a further correction that builds on the Rocklin correction such that it can be applied to the periodic lipid−water system.

The sum of $\Delta\Delta G_{\mathrm{NET}}(L)$ and $\Delta\Delta G_{\mathrm{USV}}(L)$ describes the periodic self-interaction of a naked point charge in a homogeneous medium described by a dielectric constant ϵ_S and is defined as

$$\begin{aligned} & \Delta\Delta G_{\mathrm{NET}}(L) + \Delta\Delta G_{\mathrm{USV}}(L) \\ & = -\frac{\xi_{\mathrm{LS-3D}}}{8\pi\epsilon_0\epsilon_S}[(Q_{\mathrm{P}} + Q_{\mathrm{L}})^2 - Q_{\mathrm{P}}^2]\frac{1}{L} \end{aligned} \tag{15}$$

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

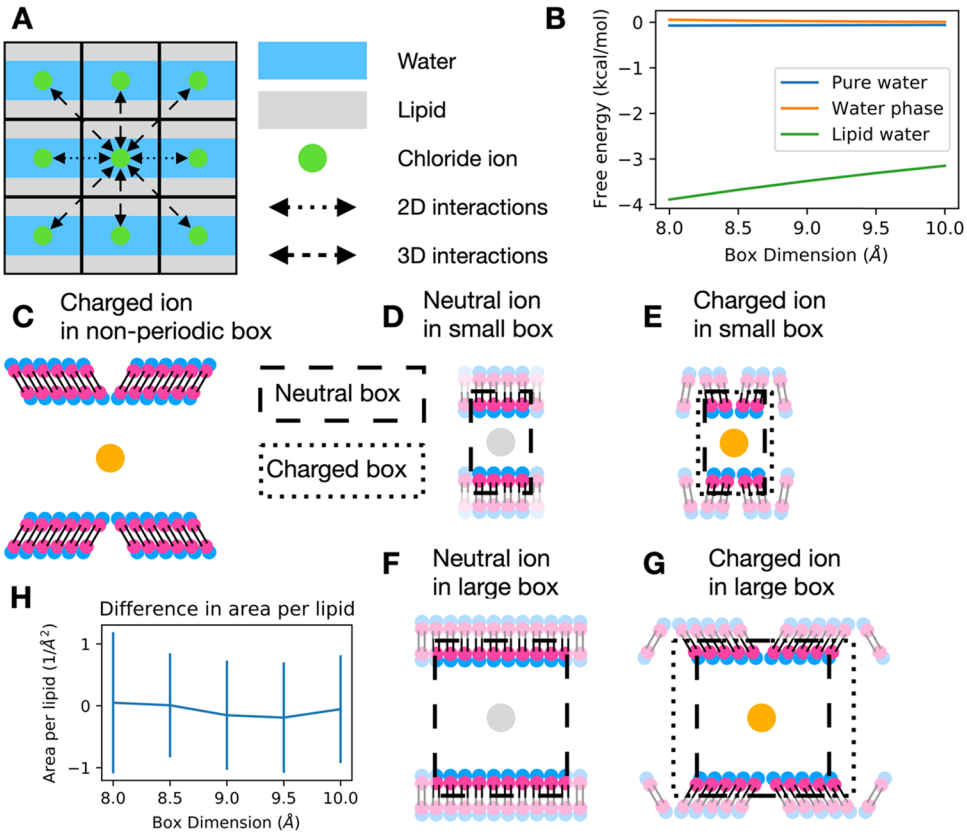


Figure 6. Self-interaction term in a lipid−water system and the discrete solvent effect of the lipid. (A) The plane self-interaction of the ion is through the water phase, whereas the off-plane self-interactions are through the lipid−water mixture. (B) The self-interaction is very small and exhibits very little size dependency when the ion is in a pure water medium (blue) or in the water phase of the lipid−water system (orange). The self-interaction term could be significant for the case where the charged particle is in the lipid phase within a lipid−water system (green). (C) In nonperiodic conditions, the lipid can reorientate in response to the charged ion. In periodic conditions, the lipid will not reorientate when the ion does not harbor charge. (D, F) When the ion is charged, the lipid can reorientate but the extent of the reorientation depends on the size of the box. (E, G) A larger box would permit a larger reorientation and a larger change in the box dimension. (H) The area per lipid as a function of box dimension.

where $\xi_{\text{LS-3D}}$ is the cubic LS (Wigner) integration constant $\xi_{\text{LS-3D}} \approx -2.837297$;[45] ϵ_0 is the permittivity of vacuum, where $\frac{1}{4\pi\epsilon_0}$ evaluates to 138.93545585 kJ nm e^{-2} mol^{-1}; ϵ_S is the dielectric constant of the solvent, which in this case is TIP3P water $\epsilon = 97$; Q_P and Q_L are the total charge of the protein and the ligand, respectively; and L is the box dimension.

For a test case of computing the charge annihilation free energy of a chloride ion, the self-interaction in the pure water box can be computed analytically by approximating the water to a homogeneous continuum medium

$$\Delta\Delta G_{\text{coul}}^{\text{water}}(L) = -\frac{\xi_{\text{LS-3D}}}{8\pi\epsilon_0\epsilon_{\text{water}}}\frac{q^2}{L} \tag{16}$$

where q is the total charge of the ion and ϵ_{water} is the dielectric constant of the water.

For a single ion in the lipid−water system (Figure 6A), the situation is different as the horizontal interactions parallel to the membrane are through the water medium (ϵ_{water}), while all of the off-plane interactions are through a mixture of lipid and water (ϵ_{mix}). Thus, the self-interaction term $\Delta\Delta G_{\text{coul}}^{\text{lipid-water}}(L)$ is defined as

$$\Delta\Delta G_{\text{coul}}^{\text{lipid-water}}(L) = \Delta\Delta G_{\text{coul}}^{\text{mix-3D}}(L) - \Delta\Delta G_{\text{coul}}^{\text{mix-2D}}(L) + \Delta\Delta G_{\text{coul}}^{\text{water-2D}}(L) \tag{17}$$

where $\Delta\Delta G_{\text{coul}}^{\text{mix-3D}}(L)$ is the self-interaction in the 3D space with a homogeneous medium of ϵ_{mix}, $\Delta\Delta G_{\text{coul}}^{\text{mix-2D}}(L)$ is the self-interaction in the 2D plane, and $\Delta\Delta G_{\text{coul}}^{\text{water-2D}}(L)$ is the self-interaction in the 2D plane with another homogeneous medium such as ϵ_{water}. The terms are defined as

$$\Delta\Delta G_{\text{coul}}^{\text{mix-3D}}(L) = -\frac{\xi_{\text{LS-3D}}}{8\pi\epsilon_0\epsilon_{\text{mix}}}\frac{q^2}{L} \tag{18}$$

$$\Delta\Delta G_{\text{coul}}^{\text{mix-2D}}(L) = -\frac{\xi_{\text{LS-2D}}}{8\pi\epsilon_0\epsilon_{\text{mix}}}\frac{q^2}{L} \tag{19}$$

$$\Delta\Delta G_{\text{coul}}^{\text{water-2D}}(L) = -\frac{\xi_{\text{LS-2D}}}{8\pi\epsilon_0\epsilon_{\text{water}}}\frac{q^2}{L} \tag{20}$$

where $\xi_{\text{LS-2D}}$ is the squared LS (Wigner) integration constant $\xi_{\text{LS-2D}} \approx -3.90025$[45] and ϵ_{mix} is the dielectric constant of the water−lipid system

$$\frac{1}{\epsilon_{\text{mix}}} = \frac{1}{\epsilon_{\text{lipid}}}^{2l_{\text{hydrophobic}}/L} \times \frac{1}{\epsilon_{\text{water}}}^{L-2l_{\text{hydrophobic}}/L} \tag{21}$$

where $l_{\text{hydrophobic}}$ is the length of the aryl chain thickness of a single lipid ($l_{\text{hydrophobic}} \approx 18.87$ Å, computed based on the distance between the first atom in the aryl chain to the end of the lipid on 100 ns of equilibrium simulation of a 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) bilayer in an NPT

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

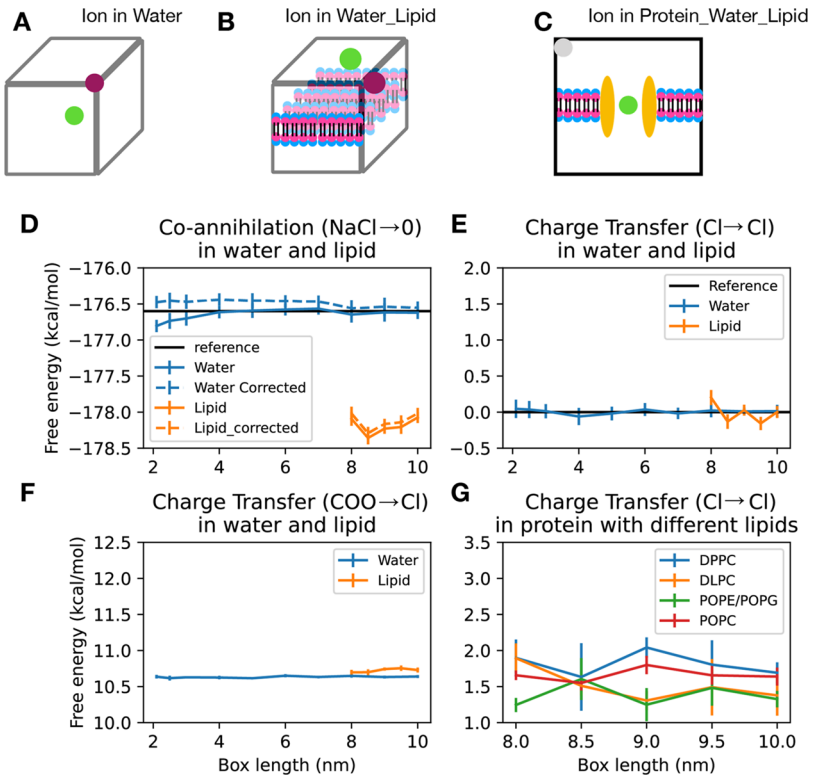


Figure 7. Use of a co-alchemical ion to maintain charge neutrality. (A) The test system for the coannihilation of NaCl or charge transfer from the chloride ion or formic acid to chloride ion, where one molecule is put at the center of the box and the other at the edge of the box. (B) The test case for the lipid−water system, where one molecule is put at the center of the box and the other molecule at the same *xy*-plane, is at the edge. (C) The test case where protein is embedded in the lipid and the chloride ion is restrained in the center of the protein, where the charge is transferred from the chloride (green) in the protein (brown) to another chloride at the corner of the box (gray). (D) The coannihilation free energies of NaCl in water (blue) and the lipid−water system (orange) are different. (E) The free energy of charge transfer from chloride to chloride in both pure water and lipid membrane systems. (F) The charge transfer from formic acid to chloride ion. (G) The free energy of charge transfer of the chloride ion in a membrane protein to a chloride ion in solvent in different lipids.

ensemble); a factor of 2 is used to account for the bilayer. ϵ_{lipid} is the dielectric constant of the lipid aryl chain ($\epsilon_{\text{lipid}} = 2$).[41]

ϵ_{water} is the dielectric constant describing interactions in the *xy*-plane. Since the charge species is an ion in the water phase, ϵ_{water} is the dielectric constant of water and if the charge species is at the center of the membrane (e.g., charged ligand in the membrane protein), ϵ_{water} would be the dielectric constant of the hydrophobic core of the membrane.

To show the effect of this treatment and its influence on the finite-size artifact, the self-interaction terms have been computed using three test cases: monovalent ion in pure water, monovalent ion in the water phase of a lipid−water system, and a hypothetical monovalent ion in the hydrophobic core of the lipid−water system (Figure 6B). Note that for the ion in the lipid−water system, where the 2D plane (i.e., *xy*) self-interaction is through the continuum medium of water (ϵ_{water}), the $\Delta\Delta G_{\text{coul}}^{\text{lipid-water}}(L)$ is very small (<0.1 kcal/mol) but can be quite significant (~4 kcal/mol) when the net charge is in the hydrophobic core of the lipid bilayer (e.g., a charged ligand in the membrane).

Discrete Solvent Correction for Lipid. Though the continuum lipid model can reproduce the ESP generated by the explicit lipid in a neutral state, it cannot reproduce the configurational change of the lipid in response to a charged ion. For water, the effect of the charged species on the orientation of the solvent (Figure 1C) is accounted for by $\Delta\Delta G_{\text{DSC}}(L)$. In theory, these effects should still exist for lipids, where the lipid would reorient in response to the charged species in the nonperiodic conditions (Figure 6C). Periodic boundary conditions would be expected to disrupt this lipid reorientation.

This disruption would also be size-dependent. For large boxes, the lipid is not reoriented when the ligand is unchanged (Figure 6F) but would reorient when the ligand is charged (Figure 6G). The reorientation would result in an expansion in the *xy*-plane and a larger area per lipid. For smaller boxes, when the ligand is unchanged, the lipid is still not reoriented (Figure 6D). However, when the ligand is charged, the lipid should reorient but is disrupted by the periodic boundary (Figure 6E). This would result in a smaller increase in the *xy*-plane and a smaller increase in area per lipid. To check if this effect is significant and if it is size-dependent, the difference in area per lipid between the charged and uncharged states has been computed for different box sizes. As shown in Figure 6H, the change of the area per lipid when the ion is charged is very small, suggesting that there is very little reorientation. Furthermore, the effect of the box size on this difference is very small, suggesting that $\Delta\Delta G_{\text{DSC}}(L)$ from lipid reorientation would be negligible.

Additional Co-Alchemical Ion as Alternative Solution. In the previous sections, we have shown that Rocklin correction could be used to remove the size-dependent artifact during charge annihilation free-energy calculations and, using a continuum model for the lipid bilayer, the deviation from the reference value could be lowered from ~2 kcal/mol to ~0.5

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

kcal/mol for a simple lipid−water system. However, it is unclear as to how such a continuum model for a lipid bilayer would be applied to a protein−lipid system. Another way of solving the size-dependent artifact is to use a co-alchemical ion to maintain the neutrality of the box.

There could be two ways of applying such an alchemical ion. For example, for the charge annihilation of a ligand with a charge of −1, a sodium ion of +1 charge could be annihilated simultaneously or a chloride ion with zero charge could be recharged to a normal chloride ion. For a water system (Figure 7A), both approaches yield the correct answer, where the simultaneous annihilation of chloride and sodium ion yields the sum of the annihilation free energy of a chloride and a sodium ion (Figure 7D). Some deviation can be seen where the coannihilation free energy has some very small size dependency. This deviation could be removed by accounting for the interaction between the sodium and chloride ions, assuming that the water is a continuum medium (Figure 7D).

Another option is to charge another ion to keep the total charge neutral. To account for the charge annihilation of a chloride ion, another neutral chloride ion will be charged. The charge transfer from one chloride ion to another chloride ion yields a zero free-energy difference, regardless of the presence of lipid (Figure 7E). Since the charge transfer between chloride ions will always give zero, formic acid is used as a test case, where the charge-transfer free energy is constant across different box sizes (Figure 7F).

The situation is, however, different for lipid−water systems (Figure 7B), where the simultaneous annihilation of chloride and sodium ions yields a free energy of ∼1.5 kcal/mol lower than the case for a pure water system (Figure 7D). Many factors could give rise to this deviation. The ions could have interactions with the lipid bilayer, which are absent in a pure water system. The lipid bilayer could disrupt the solvation shell around the ion in a different sense compared with the ion in a pure water system. Finally, the self-interaction free energy of the ions across the periodic boundary condition could be different due to the low dielectric constant of the lipid acryl tail. Among these three factors, the final factor is the only one that we would consider here. Given that an analytical solution cannot be derived to describe the interactions between the sodium and chloride ions across the periodic boundary, we attempted to remove this artifact numerically. The lattice-sum energy of the system-{Rocklin, 2013 #7126} can be described as

$$U_{LS} = \frac{1}{4\pi\epsilon_0}\left[\sum_{i=1}^{N}\sum_{j>i}^{N} q_i q_j \psi_{LS}(r_{ij}) + \frac{1}{2}\psi^0_{LS}\sum_{i=1}^{N} q_i^2\right] \quad (22)$$

where the system has N charged particles, where each particle has charge q_i at location r_i ($r_{ij} = r_j - r_i$). ψ_{LS} is the lattice-sum influence function, which is the electric potential generated by a unit point charge at the origin multiplied by $4\pi\epsilon_0$, and ψ^0_{LS} is the Wigner self-term constant (difference between ψ_{LS} and r^{-1} in the limit of infinitesimal distances). For a cubic box of edge L in a homogeneous medium with a dielectric constant of ϵ_S, the ψ_{LS} and ψ^0_{LS} are defined as

$$\nabla^2\psi_{LS} = -\frac{4\pi}{\epsilon_S}\sum_{n\in\mathbb{Z}^3}[\delta(r+Ln) - L^{-3}] \quad (23)$$

$$\langle\nabla\psi_{LS}\rangle = 0$$

$$\langle\psi_{LS}\rangle = 0$$

$$\psi^0_{LS} = L^{-1}\xi_{LS}$$

where δ is the three-dimensional Dirac delta function. For the lipid−water system, where the dielectric constant is different for plane interactions and off-plane interaction (Figure 6A), ψ_{LS} is defined as

$$\nabla^2\psi_{LS} = \begin{cases} \dfrac{4\pi}{\epsilon_{water}} - \displaystyle\sum_{n\in\mathbb{Z}^3}[\delta(r+Ln) - L^{-3}] & z = 0 \\ -\dfrac{4\pi}{\epsilon_{mix}}\displaystyle\sum_{n\in\mathbb{Z}^3}[\delta(r+Ln) - L^{-3}] & z \neq 0 \end{cases} \quad (24)$$

where ϵ_{water} is the dielectric constant of water and ϵ_{mix} is the dielectric constant of the lipid−water system. z is the distance to the center of the water phase and describes the interactions going purely through the water phase in the lipid−water system.

To obtain this lattice-sum energy numerically, three PME calculations were performed

$$U_{LS} = U^{\epsilon_{mix}}_{PME}(L, L, L) - U^{\epsilon_{mix}}_{PME}(L, L, \infty) + U^{\epsilon_{water}}_{PME}(L, L, \infty) \quad (25)$$

where $U^{\epsilon_{mix}}_{PME}(L, L, L)$ is the lattice-sum energy of the system in a continuous medium of ϵ_{mix}, $U^{\epsilon_{mix}}_{PME}(L, L, \infty)$ is the lattice-sum energy of the system in a continuous medium of ϵ_{mix} with an infinite z-axis dimension (10,000 nm as employed in the calculation), and $U^{\epsilon_{water}}_{PME}(L, L, \infty)$ is the lattice-sum energy of the system in a continuous medium of ϵ_{water} with an infinite z-axis dimension. The lattice-sum energy removes a very small amount of the deviation but the majority persists. Thus, we conclude that the deviation could not be corrected easily and further considerations are required if the simultaneous annihilation method were to be used in lipid membrane systems.

The charge-transfer method, on the other hand, still gives a very small amount of deviation in the membrane system compared with the pure water system. This deviation, however, is negligible (∼0.1 kcal/mol) and relatively size-independent.

Co-Alchemical Ion in the Case of a Membrane Protein. Although we have shown above that the charge-transfer method using a co-alchemical ion can derive the free-energy difference in a size-independent manner in an idealized system, it remains to be seen as to how it would behave in the real world, where the ligand is bound to a protein embedded in a membrane. We therefore constructed a system with the 14-stranded outer-membrane porin OmpG[46] embedded in a POPC membrane. The ion to be annihilated is situated at the core of the protein at the same level as the core of the hydrophobic region of the lipid (Figure 7C). Given that the chloride ion in the protein is in an environment that is different from the solvent, the charge transfer from one chloride ion to another chloride ion will not be the same as the case when both of them are in solvent (Figure 7E). As shown in Figure 6B, the self-interactions across the hydrophobic region of the lipid would give a significant self-interaction energy that would be expected to be size-dependent; this problem should be greatly reduced when the ligand is in the protein. Our calculations show that for the model protein system, this effect of the box dimension on the self-interaction term is smaller than the error and can be safely ignored (Figure

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

7G). Furthermore, the result is invariant with respect to the lipid saturation level (saturated DPPC, half-saturated POPC) or thickness (short DLPC/long DPPC) or when the lipid is charged (POPE POPG mixture) showing that the method is robust.

DISCUSSION

In this work, we have shown that the Rocklin correction can correct for the finite-size artifact for the soluble system but will exhibit a large deviation in the lipid−water system. By incorporating a continuum lipid model via the Poisson−Boltzmann calculation, we were able to greatly reduce the error. Although this gives a reasonable improvement, this approach remains rather limited as the continuum lipid model could only be derived for a pure lipid bilayer and it is unclear as to how to generalize this to all lipid−protein systems. Instead, we propose to avoid this problem altogether and advocate the use of the co-alchemical ion approach, where the charge of the charged ligand is transferred to the co-alchemical ion. However, it is also the case that this approach gives rise to a small systematic error when the membrane is present. Note that in the case where a large enough bulk solvent volume can be included, then one might use the same volume to couple another copy of the ligand, thus reverting to the double-system single-box approach {Macchiagodena, 2021 #10371}, but there are practical implications that can make this approach less trivial than a simple co-alchemical ion treatment.

Our results here shed some light on the apparent inconsistencies in the literature, where many people have strongly advocated the Rocklin correction and the necessity of correcting for the finite-size artifact. Others have shown that ignoring the size-dependent artifact can still give results that give a good correlation with the experimental data.[47] Similarly, there are studies that have shown that the Rocklin correction has very little effect on the RBFE calculations.[48] This is likely best explained by the fact that the Rocklin correction mainly corrects for the long-range effects of the electrostatic interactions. The $\Delta\Delta G_{DSC}(L)$, which is usually the largest component in Rocklin correction, depends on the proportion of the box that is filled with water. For RBFE calculations, this proportion is unlikely to change, which means the $\Delta\Delta G_{DSC}(L)$ is very similar for different ligands tested in RBFE. Though the specific ligand−protein interaction would give rise to a different $\Delta\Delta G_{RIP}(L)$, different ligands would give a very similar average electrostatic potential and $\Delta\Delta G_{RIP}(L)$ is usually very small unless lipids are involved.

One way of avoiding the finite-size artifact would be to ensure the neutrality of the simulation box. For the computation of ligand binding free energy, an alternative would be path sampling, where the free energy of pulling the ligand out of the binding pocket is taken as the binding free energy. However, it has been observed that the binding free energy computed with path sampling did not converge to the same end result as those done alchemically and corrected with the Rocklin correction.[49] However, one study has also shown that they can give similar results.[50] One explanation for this discrepancy would be that the self-interaction term would be size-dependent when the species of opposite charge cannot sample the same configurational space. For example, in Figure 6B, we have shown that the self-interaction energy of an ion in the water phase of the lipid−water system can be quite different compared with the same ion in the hydrophobic core. Thus, in the case where the charge species cannot sample the whole configurational space, such as when the ligand is restrained to the protein during path sampling, the self-interaction energy of the charged ligand cannot be compensated by an ion of opposite charge in the water phase. Thus, the self-interaction energy of the ligand will be size-dependent and could give different results when the simulation box has different sizes. This difference in the self-interaction energy could also explain the conformational dependence on the box size,[51] where smaller boxes would favor the conformation that minimizes the self-interaction energy.

CONCLUSIONS

To solve the finite-size effect, we have shown that the Rocklin correction, the co-alchemical ion with charge transfer, and coannihilation of the charge would all work for a soluble protein system. For systems with lipid bilayers, however, only the co-alchemical ion using charge transfer will avoid the finite-size effects, and thus for these kinds of calculations, this is the recommended approach.

METHODS

Construction of Water Box. To investigate the charge annihilation free energy, empty boxes were constructed, where the box dimension, L, was 2.1, 2.5, 3, 4, 5, 6, 7, 8, 9, and 10 nm. For the single-ion charge annihilation free-energy calculations, the ion was placed at the center of the box and the box was solvated with the gmx solvate tools.[52] For the charge transfer or coannihilation calculations, the chloride ion was position-restrained at the center of the box with a restraint strength of 1000 kJ/(mol nm^2) at the x-, y-, and z-axes. The sodium ion, chloride ion, or formic acid was position-restrained to the edge of the box (0, 0, 0) at the same strength.

Construction of the Lipid−Water Box. The lipid−water box was constructed using a pre-equilibrated POPC system obtained from the slipid website (http://www.fos.su.se/~sasha/SLipids/Downloads.html, accessed 09/12/2021).[53] The lipid was replicated in the x- and y-dimensions with MDAnalysis[54] and was trimmed to the desired dimension, L, as required (i.e., 8, 8.5, 9, 9.5, 10 nm). The box was solvated with water and equilibrated for 100 ns under isobaric and isothermal ensemble (NPT) conditions. For the charge annihilation of a single ion, the ion was placed at the center of the box and was position-restrained on the z-axis with a restraint of 1000 kJ/(mol nm^2). For the charge transfer or coannihilation calculations, the chloride ion was placed at the center of the box ($L/2$, $L/2$, $L/2$) and the sodium ion, chloride ion, or formic acid was placed at the edge at the same z-axis (0, 0, $L/2$) and position-restrained in all (x, y, and z) axes. To ensure the ion remains in the middle of the water phase (in the middle of the box), the equilibrated box was centered with respect to the water phase.

Construction of the Protein−Lipid−Water Box. The open form of OmpG (PDB: 2IWV)[46] was embedded in a POPC bilayer using a self-assembly protocol, as described previously by us.[55] Both protein and lipid were described using the Martini 3 force field.[44] The lipids were trimmed to the desired box size and equilibrated for 20 ns with the "New-RF" parameters.[56] The system was then converted to an atomistic representation with cg2at[57] and then further equilibrated for 100 ns. The ion to be annihilated was position-restrained to the center of the protein (i.e within the barrel), where the reference z-axis coordinate is the center of the membrane defined as the mean of the phosphate groups of the lipids. The coordinate in the x−y-plane was chosen such that it is the furthest from any atoms in the protein. To ensure that the same result is obtained in different

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

box sizes, the protein was position-restrained to the center of the box using the same reference coordinate from the crystal structure. The co-alchemical ion where the charge is created was restrained to the corner of the box [0, 0, 0]. All positional restraints were set to 1000 kJ/(mol nm^2).

Simulation Setup. Simulations were run with the GROMACS 2020 MD engine.[52] The force field for lipids was slipid,[53] and the TIP3P water model was used for water. Formic acid was parameterized with GAFF2.[58] The electrostatic interactions and Lennard-Jones interaction were computed with particle mesh Ewald (PME).[5] The pme-order was set to 6, the Fourier spacing was set to 0.1, and the ewald-rtol was set to 1×10^{-6} for electrostatic interactions and 1×10^{-3} for Lennard-Jones interactions. The direct space cutoff was set to 1 nm. The H-bond length was constrained by LINCS[59] with lincs-iter and lincs-order set to 2 and 6, respectively. The time step of integration was 2 fs. The alchemical transformation was done via 11 steps with a delta of 0.1. The charge of the chloride, sodium, or formic acid was scaled linearly from the full charge to zero or vice versa.

The system was energy-minimized before a 100 ps NPT equilibration with Langevin dynamics at 298.15 K,[60] and the pressure was restrained to 1 bar with the Berendsen barostat.[61] Production runs employed replica exchange, and exchanges were performed every 1000 steps and used the Parrinello–Rahman barostat.[62] For the calculations involving only sodium and chloride ions, production runs were 1 ns and were repeated 30 times. For the calculations involving formic acid, five repeats of 30 ns production runs were performed. The free-energy estimate was done with MBAR[63] and alchemlyb.[64]

Poisson–Boltzmann Calculations. APBS 1.5[65] was used to numerically solve Poisson–Boltzmann equations. A POPC water box with a dimension of (64.26 * 67.47 * 71.67 Å) was simulated in the NVT ensemble for 100 ns, where 100 snapshots were taken to compute the electrostatic potential profile. The APBS input files were modeled on the Rocklin correction.[11]

For the computation of the continuum model, the input files (charge density, dielectric constant mesh) were prepared with rocklinc (https://github.com/bigginlab/rocklinc, accessed 11/11/2021) and all analysis codes are included in the same depository.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01251.

Fitted parameters used for the dielectric constant in eq 9 (Table S1) and fitted parameters for the charge density used in eq 10 (Table S2) (PDF)

(PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Zhiyi Wu − *Department of Biochemistry, Oxford OX1 3QU, U.K.;* orcid.org/0000-0002-7615-7851; Email: Zhiyi.wu@bioch.ox.ac.uk

Philip C. Biggin − *Department of Biochemistry, Oxford OX1 3QU, U.K.;* orcid.org/0000-0001-5100-8836; Email: Philip.biggin@bioch.ox.ac.uk

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c01251

Author Contributions

Z.W. designed the research and performed and analyzed simulations. Z.W. and P.C.B. wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Irfan Alibay for useful discussions, the Advanced Research Computing facility, JADE2 (EP/T022205/1), and the ARCHER U.K. National Supercomputing Services for computer time granted via the High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://www.hecbiosim.ac.uk), supported by EPSRC (EP/R029407/1).

■ REFERENCES

(1) (a) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* **2017**, *139*, 946−957. (b) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016**, *7*, 207−218.

(2) de Oliveira, C.; Yu, H. S.; Chen, W.; Abel, R.; Wang, L. Rigorous free energy perturbation approach to estimating relative binding affinities between ligands with multiple protonation and tautomeric states. *J. Chem. Theory Comput.* **2019**, *15*, 424−435.

(3) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate estimation of ligand binding affinity changes upon protein mutation. *ACS Cent. Sci.* **2018**, *4*, 1708−1718.

(4) Mobley, D. L.; Gilson, M. K. Predicting binding free energies: Frontiers and benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531−558.

(5) 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.

(6) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. An overview of electrostatic free energy computations for solutions and proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690−2709.

(7) Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564−1582.

(8) Friedman, R. Computational studies of protein−drug binding affinity changes upon mutations in the drug target. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, No. e1563.

(9) Farafonov, V. S.; Lebed, A. V.; McHedlov-Petrossyan, N. O. Computing pKa Shifts Using Traditional Molecular Dynamics: Example of Acid−Base Indicator Dyes in Organized Solutions. *J. Chem. Theory Comput.* **2020**, *16*, 5852−5865.

(10) Vilseck, J. Z.; Ding, X.; Hayes, R. L.; Brooks, C. L. Generalizing the discrete gibbs sampler-based λ-dynamics approach for multisite sampling of many ligands. *J. Chem. Theory Comput.* **2021**, *17*, 3895−3907.

(11) 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*, No. 184103.

(12) Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A. Relative binding affinity prediction of charge-changing sequence mutations with FEP in protein−protein interfaces. *J. Mol. Biol.* **2019**, *431*, 1481−1493.

(13) Ge, Y.; Hahn, D. F.; Mobley, D. L. A benchmark of electrostatic method performance in relative binding free energy calculations. *J. Chem. Inf. Model.* **2021**, *61*, 1048−1052.

(14) Kubincová, A.; Riniker, S.; Hünenberger, P. H. Reaction-field electrostatics in molecular dynamics simulations: development of a conservative scheme compatible with an atomic cutoff. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26419−26437.

(15) Sidler, D.; Frasch, S.; Cristòfol-Clough, M.; Riniker, S. Anisotropic reaction field correction for long-range electrostatic interactions in molecular dynamics simulations. *J. Chem. Phys.* **2018**, *148*, No. 234105.
(16) Karttunen, M.; Rottler, J.; Vattulainen, I.; Sagui, C. Electrostatics in Biomolecular Simulations: Where Are We Now and Where Are We Heading?. In *Current Topics in Membranes*; Feller, S. E., Ed.; Academic Press, 2008; Chapter 2, Vol. *60*, pp 49−89.
(17) Hayes, R. L.; Vilseck, J. Z.; Brooks Iii, C. L. Approaching protein design with multisite λ dynamics: Accurate and scalable mutational folding free energies in T4 lysozyme. *Prot. Sci.* **2018**, *27*, 1910−1922.
(18) Wang, Y.; Murlidaran, S.; Pearlman, D. A. Quantum simulations of SARS-CoV-2 main protease Mpro enable high-quality scoring of diverse ligands. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 963−971.
(19) Holden, Z. C.; Rana, B.; Herbert, J. M. Analytic gradient for the QM/MM-Ewald method using charges derived from the electrostatic potential: Theory, implementation, and application to ab initio molecular dynamics simulation of the aqueous electron. *J. Chem. Phys.* **2019**, *150*, No. 144115.
(20) Cruzeiro, V. W. D.; Amaral, M. S.; Roitberg, A. E. Redox potential replica exchange molecular dynamics at constant pH in AMBER: Implementation and validation. *J. Chem. Phys.* **2018**, *149*, No. 072338.
(21) Song, L. F.; Bansal, N.; Zheng, Z.; Merz, K. M. Detailed potential of mean force studies on host−guest systems from the SAMPL6 challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1013−1026.
(22) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-pull-release calculations of ligand binding and conformational changes on the first BRD4 bromodomain. *J. Chem. Theory Comput.* **2017**, *13*, 3260−3275.
(23) (a) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical transfer approach to absolute binding free energy estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309−3319. (b) Procacci, P. A remark on the efficiency of the double-system/single-box nonequilibrium approach in the SAMPL6 SAMPLing challenge. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 635−639. (c) Ekimoto, T.; Yamane, T.; Ikeguchi, M. Elimination of finite-size effects on binding free energies via the warp-drive method. *J. Chem. Theory Comput.* **2018**, *14*, 6544−6559.
(24) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 601−633.
(25) (a) Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H. Calculation of Binding Free Energies. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Springer: New York, 2015; Vol. *1215*, pp 173−209. (b) Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual Double-System Single-Box: A nonequilibrium alchemical technique for absolute binding free energy calculations: Application to ligands of the SARS-CoV-2 main protease. *J. Chem. Theory Comput.* **2020**, *16*, 7160−7172. (c) Macchiagodena, M.; Karrenbrock, M.; Pagliai, M.; Procacci, P. Virtual double-system single-box for absolute dissociation free energy calculations in GROMACS. *J. Chem. Inf. Model.* **2021**, *61*, 5320−5326.
(26) Patel, D.; Patel, J. S.; Ytreberg, F. M. Implementing and Assessing an Alchemical Method for Calculating Protein−Protein Binding Free Energy. *J. Chem. Theory Comput.* **2021**, *17*, 2457−2464.
(27) Suruzhon, M.; Senapathi, T.; Bodnarchuk, M. S.; Viner, R.; Wall, I. D.; Barnett, C. B.; Naidoo, K. J.; Essex, J. W. ProtoCaller: Robust automation of binding free energy calculations. *J. Chem. Inf. Model.* **2020**, *60*, 1917−1921.
(28) Carvalho Martins, L.; Cino, E. A.; Ferreira, R. S. PyAutoFEP: An automated free energy perturbation workflow for GROMACS integrating enhanced sampling methods. *J. Chem. Theory Comput.* **2021**, *17*, 4262−4273.
(29) Simonson, T.; Roux, B. Concepts and protocols for electrostatic free energies. *Mol. Simul.* **2016**, *42*, 1090−1101.
(30) (a) Huai, Z.; Yang, H.; Li, X.; Sun, Z. SAMPL7 TrimerTrip host−guest binding affinities from extensive alchemical and end-point free energy calculations. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 117−129. (b) Ruiz-Blanco, Y. B.; Sanchez-Garcia, E. CL-FEP: An end-state free energy perturbation approach. *J. Chem. Theory Comput.* **2020**, *16*, 1396−1410. (c) Cruz, J.; Wickstrom, L.; Yang, D.; Gallicchio, E.; Deng, N. Combining alchemical transformation with a physical pathway to accelerate absolute binding free energy calculations of charged ligands to enclosed binding sites. *J. Chem. Theory Comput.* **2020**, *16*, 2803−2813. (d) Ben-Shalom, I. Y.; Lin, Z.; Radak, B. K.; Lin, C.; Sherman, W.; Gilson, M. K. Accounting for the central role of interfacial water in protein−ligand binding free energy calculations. *J. Chem. Theory Comput.* **2020**, *16*, 7883−7894. (e) Bhati, A. P.; Wan, S.; Hu, Y.; Sherborne, B.; Coveney, P. V. Uncertainty quantification in alchemical free energy methods. *J. Chem. Theory Comput.* **2018**, *14*, 2867−2880. (f) Bellucci, M. A.; Walsh, M. R.; Trout, B. L. Molecular dynamics analysis of anti-agglomerant surface adsorption in natural gas hydrates. *J. Phys. Chem. C* **2018**, *122*, 2673−2683.
(31) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(32) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate calculation of relative binding free energies between ligands with different net charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346−6358.
(33) Shi, Y.; Laury, M. L.; Wang, Z.; Ponder, J. W. AMOEBA binding free energies for the SAMPL7 TrimerTrip host−guest challenge. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 79−93.
(34) Qi, R.; Walker, B.; Jing, Z.; Yu, M.; Stancu, G.; Edupuganti, R.; Dalby, K. N.; Ren, P. Computational and experimental studies of inhibitor design for aldolase A. *J. Phys. Chem. B* **2019**, *123*, 6034−6041.
(35) Urano, R.; Shinoda, W.; Yoshii, N.; Okazaki, S. Exact long-range Coulombic energy calculation for net charged systems neutralized by uniformly distributed background charge using fast multipole method and its application to efficient free energy calculation. *J. Chem. Phys.* **2020**, *152*, No. 244115.
(36) Macchiagodena, M.; Karrenbrock, M.; Pagliai, M.; Procacci, P. Virtual Double-System Single-Box for Absolute Dissociation Free Energy Calculations in GROMACS. *J. Chem. Inf. Model.* **2021**, *61*, 5320−5326.
(37) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **2014**, *10*, 381−390.
(38) Kyrychenko, A.; Lim, N. M.; Vasquez-Montes, V.; Rodnin, M. V.; Freites, J. A.; Nguyen, L. P.; Tobias, D. J.; Mobley, D. L.; Ladokhin, A. S. Refining protein penetration into the lipid bilayer using fluorescence quenching and molecular dynamics simulations: The case of diphtheria toxin translocation domain. *J. Membr. Biol.* **2018**, *251*, 379−391.
(39) Esposito, C.; Vitalis, A. Precise estimation of transfer free energies for ionic species between similar media. *Phys. Chem. Chem. Phys.* **2018**, *20*, 27003−27010.
(40) Duncan, A. L.; Corey, R. A.; Sansom, M. S. P. Defining how multiple lipid species interact with inward rectifier potassium (Kir2) channels. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 7803−7813.
(41) Marcoline, F. V.; Bethel, N.; Guerriero, C. J.; Brodsky, J. L.; Grabe, M. Membrane protein properties revealed through data-rich electrostatics calculations. *Structure* **2015**, *23*, 1526−1537.
(42) 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.
(43) Choe, S.; Hecht, K. A.; Grabe, M. A Continuum Method for Determining Membrane Protein Insertion Energies and the Problem of Charged Residues. *J. Gen. Physiol.* **2008**, *131*, 563−573.
(44) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.;

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672

Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382−388.
(45) 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.
(46) Yildiz, z.; Vinothkumar, K. R.; Goswami, P.; Kühlbrandt, W. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J.* **2006**, *25*, 3702−3713.
(47) Krepl, M.; Damberger, F. F.; von Schroetter, C.; Theler, D.; Pokorná, P.; Allain, F. H. T.; Šponer, J. Recognition of N6-methyladenosine by the YTHDC1 YTH domain studied by molecular dynamics and nmr spectroscopy: The role of hydration. *J. Chem. Phys. B* **2021**, *125*, 7691−7705.
(48) Lim, V. T.; Geragotelis, A. D.; Lim, N. M.; Freites, J. A.; Tombola, F.; Mobley, D. L.; Tobias, D. J. Insights on small molecule binding to the Hv1 proton channel from free energy calculations with molecular dynamics simulations. *Sci. Rep.* **2020**, *10*, No. 13587.
(49) Öhlknecht, C.; Perthold, J. W.; Lier, B.; Oostenbrink, C. Charge-changing perturbations and path sampling via classical molecular dynamic simulations of simple guest−host systems. *J. Chem. Theory Comput.* **2020**, *16*, 7721−7734.
(50) Deng, N.; Cui, D.; Zhang, B. W.; Xia, J.; Cruz, J.; Levy, R. Comparing alchemical and physical pathway methods for computing the absolute binding free energy of charged ligands. *Phys. Chem. Chem. Phys.* **2018**, *20*, 17081−17092.
(51) Kasahara, K.; Sakuraba, S.; Fukuda, I. Enhanced Sampling of Molecular Dynamics Simulations of a Polyalanine Octapeptide: Effects of the Periodic Boundary Conditions on Peptide Conformation. *J. Chem. Phys. B* **2018**, *122*, 2495−2503.
(52) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(53) Grote, F.; Lyubartsev, A. P. Optimization of Slipids force field parameters describing headgroups of phospholipids. *J. Phys. Chem. B* **2020**, *124*, 8784−8793.
(54) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Dotson, D. L.; Domanski, J.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Python in Science*; Benthall, S.; Rostrup, S., Eds.; Los Alamos National Lab: Austin, TX, 2016.
(55) Wu, Z.; Alibay, I.; Newstead, S.; Biggin, P. C. Proton control of transitions in an amino acid transporter. *Biophys. J.* **2019**, *117*, 1342−1351.
(56) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.
(57) Vickery, O. N.; Stansfeld, P. J. CG2AT2: an Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 6472−6482.
(58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(59) Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(60) Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient algorithms for langevin and DPD dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637−3649.
(61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(62) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(63) Chodera, J. D. A simple method for automated equilibration detection in molecular simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799−1805.
(64) alchemistry/alchemlyb, 2021. https://doi.org/10.5281/zenodo.4973744.
(65) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

https://doi.org/10.1021/acs.jctc.1c01251
J. Chem. Theory Comput. 2022, 18, 2657−2672