HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2019 October 09.

Published in final edited form as:

J Chem Theory Comput. 2018 October 09; 14(10): 5290–5302. doi:10.1021/acs.jctc.8b00604.

Determination of Ionic Hydration Free Energies with Grand Canonical Monte Carlo/Molecular Dynamics Simulations in Explicit Water

Delin Sun[1], **Sirish Kaushik Lakkaraju**[2], **Sunhwan Jo**[2], and **Alexander D. MacKerell Jr.**[1,2]

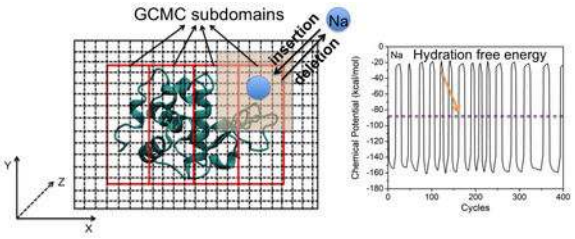
[1]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201, United States

[2]SilcsBio LLC, 8 Market Place, Suite 300, Baltimore, MD 21202, United States

Abstract

Grand Canonical Monte Carlo (GCMC) simulations of ionic solutions with explicit solvent models are known to be challenging. One challenge arises from the treatment of long-range electrostatics and finite box size in Monte Carlo simulations when periodic boundary condition and Ewald summation methods are used. Another challenge is that constant excess chemical potential GCMC simulations for charged solutes suffer from inadequate insertion and deletion acceptance ratios. In this work, we address those problems by implementing an oscillating excess chemical potential GCMC algorithm with smooth particle mesh Ewald and finite-box size corrections to treat the long-range electrostatics. The developed GCMC simulation program was combined with GROMACS to perform GCMC/MD simulations of ionic solutions individually containing Li^+, Na^+, K^+, Rb^+, Cs^+, F^-, Cl^-, Br^-, I^-, Ca^{2+} and Mg^{2+}, respectively. Our simulation results show that the combined GCMC/MD approach can approximate the ionic hydration free energies with proper treatment of long-range electrostatics. Our developed simulation approach can open up new avenues for simulating complex chemical and biomolecular systems and for drug discovery.

Graphical Abstract



Supporting Information

The Supporting Information includes a description of the method to determine the ionic excess chemical potentials and the validation of parallelization scheme of the GCMC simulation code. The Supporting Information is available free of charge on the ACS Publications website

Conflict of Interest

ADM Jr. is co-founder and CSO of SilcsBio LLC.

Keywords

Grand Canonical Monte Carlo; Ionic hydration free energy; Particle Mesh Ewald; Finite box correction; Free energy of aqueous solvation; Enhanced Sampling

1. Introduction

Understanding the molecular basis for ionic hydration is fundamentally important to a range of chemical, physical and biological sciences. Accurate theoretical models to predict ionic hydration thermodynamics are difficult to derive due to the inherent complexity of the ionic hydration process, which involves reorientation and polarization of water molecules in the vicinity of ions. Moreover, studying ionic hydration in complex heterogeneous environments, like water-octanol mixture is even more challenging. One theoretical model was proposed by Born in 1920. The Born model is based on classical electrostatics and it treats solvents implicitly as a dielectric continuum. For a single spherical ion with charge Q, the Born expression for the free energy of creating a charge (or *charging free energy*) can be written as

$$\Delta G_{\text{elec}} = \frac{Q^2}{2\sigma}\left(\frac{1}{\varepsilon} - 1\right), \quad (1)$$

where ε is the dielectric constant of the continuum solvent and σ is the so-called effective radius of the ion, which is the sum of crystal ionic radius and the solvent radius in the first ionic hydration shell. The effective radius of a new ionic compound is usually unknown *a priori*, thus limiting the wide application of the Born model. Nevertheless, the Born model successfully predicts a quadratic dependence of charging free energy on the ionic charge and a linear response of charging free energy to the ion-solvent electrostatic interaction energy. As shown by Roux et al., the Born equation can be derived by evaluating the integral which defines the excess chemical potential (μ_{ex}) for creating a charge in the solvent, i.e.,

$$\mu_{\text{ex}} = \int_0^Q dQ' \int d^3r \left(\Sigma_\theta\, h_\theta(r; Q')\rho q_\theta\right)\frac{1}{r}, \quad (1.2)$$

where $h_\theta(r; Q')$ is the equilibrium distribution function of the solvent with charge q_θ surrounding a charged solute with charge Q'. Evaluation of the integral yields the Born equation, i.e.,

$$\mu_{\text{ex}} = \int_0^Q dQ' \left[\frac{Q'}{\sigma(Q'; T, \rho)}\left(\frac{1}{\epsilon} - 1\right)\right] = \frac{Q^2}{2\sigma}\left(\frac{1}{\epsilon} - 1\right). \quad (1.3)$$

The electrostatic interaction energy between an ion and the solvent can be derived in a similar way, i.e.,

$$\left\langle U_{\text{elec}}(Q)\right\rangle = \int d^3r(\ \Sigma_\theta\ h_\theta(r;Q)\rho q_\theta)\frac{Q}{r} = Q\left[\frac{Q}{\sigma(Q;T,\rho)}\left(\frac{1}{\epsilon}-1\right)\right] = \frac{Q^2}{\sigma}\left(\frac{1}{\epsilon}-1\right). \quad (1.4)$$

Equations (1.3) and (1.4) demonstrate that the charging free energy is equal to half the ensemble average of the ion-solvent electrostatic interaction energy. The same relationship was also derived by Åqvist et al. by expanding the free energy difference between two thermodynamic states, i.e., $\Delta G = -\beta^{-1} ln\langle e^{-\beta\Delta U}\rangle$ as a Taylor series. These suggest a so-called linear response theory (LRT) method to calculating ionic charging free energy, i.e. ΔG_{elec} can be approximated from molecular dynamics (MD) or Monte Carlo (MC) simulations by extracting $<U_{\text{elec}}(Q)>$ for one ion interacting with its aqueous environment. The LRT method was verified by Åqvist et al., who showed that the ratio (R_{elec}) of ΔG_{elec} obtained from free energy perturbation (FEP) calculations to $<U_{\text{elec}}(Q)>$ varies from 0.43 to 0.55 for a series of different ionic molecules, including both atomic and molecular ions.

A more common approach to calculating the ionic hydration free energy (HFE) is performing atomistic MD or MC simulations with explicit solvents, with the HFE extracted via free energy perturbation (FEP) or thermodynamic integration (TI) methods. To achieve adequate convergence, such calculations need to pre-define a number of thermodynamic intermediate states and sufficiently long simulations are required. For simple ionic solution systems, the convergence can be satisfied within a few hundred picosecond MD simulation.[-] However, extensive applications of FEP and TI methods to more complex problems, such as estimating ligand-protein binding affinity, are still hampered due to the inherent convergence problems. For this reason, efforts continue towards the development of alternative HFE computational approaches.

Grand Canonical Monte Carlo (GCMC) simulations, which are carried out at constant chemical potential, volume and temperature, thereby allowing for the insertion and deletion of particles into and out of the simulation system, offer a new solution to the prediction of hydration thermodynamics.[-] The solute μ_{ex} provides the driving force for its dissolution, and theoretically, μ_{ex} is equivalent to the Gibbs HFE in the constant temperature and pressure condition. Accordingly, this provides a theoretical basis for prediction of HFE from GCMC simulations.

A major challenge encountered in GCMC simulations of condensed phase systems, such as bulk water, is the very low particle insertion acceptance rate mainly due to overlap of atomic van der Waals (vdW) radii even with advanced cavity-biased and configuration-biased insertion algorithms. For instance, the acceptance rates for water insertion in aqueous solutions have been estimated to be lower than 1% at ambient temperatures. For monoatomic ions, this atomic overlap challenge is small as the intrinsic water cavity is large enough to accommodate an ion. However, there are other challenges of studying ionic systems with explicit-solvent GCMC simulations. Firstly, determination of the ionic μ_{ex} (or the ionic HFE) for the GCMC simulations is nontrivial. As indicated by equation (1.2), the difference between the ionic HFE and the ion-water interaction energy is a factor of approximately 2, which corresponds to tens of kcal/mol or more. In addition, the hydration of ions involves

significant reordering of water molecules, such that the electrostatic interactions between an ion and the bulk-like water are initially energetically unfavorable. These suggest that ions will most likely never be deleted from or inserted into the solutions if the ionic μ_{ex} are kept constant at approximately the experimental HFE values. To overcome this problem, an oscillating-μ_{ex} GCMC simulation approach, as introduced in our previous work, offers a solution. Secondly, the electrostatic interactions in ionic solutions are long-ranged. We note that in our previously developed GCMC method, while applied to a range of small organic solutes and to the distribution of Mg^{2+} around RNA,[9] electrostatic interactions were treated using a cutoff method.

In this work, we report the development of an improved GCMC method in which the smooth particle-mesh Ewald (PME) method is implemented to treat the short-range and long-range electrostatics. Thirdly, when periodic boundary condition (PBC) and PME are used, movements of charged solutes in Monte Carlo simulations may be affected by the so-called "finite-box" problem. This arises from the artificial ion-image ion interactions when PBC and PME are used for charged solutes in low-dielectric solvents. In this work, we also propose a method to correct for the finite-box size effect in GCMC simulations of ionic solutions. The developed GCMC method was combined with MD simulations to study eleven ionic solutions individually containing Li^+, Na^+, K^+, Rb^+, Cs^+, F^-, Cl^-, Br^-, I^-, Ca^{2+} and Mg^{2+}. To validate the GCMC/MD method, we also calculated the ionic HFEs using FEP in conjunction with the Bennett acceptance ratio (BAR) method, with proper finite-box corrections. It is found that the presented GCMC/MD algorithm can predict the ionic HFEs with reasonable accuracy.

2. Method development

2.1 GCMC movement: theory and implementation

In the GCMC simulation method, four types of moves, *i.e.*, insertion, deletion, translation and rotation (only water molecules are rotated in the present study) are implemented. These movements are attempted and accepted governed by the Boltzmann distribution in the grand canonical ensemble. The grand canonical ensemble partition function $\Xi(\mu, V, T)$ is related to the canonical partition function $Q(N,V,T)$ through the equation

$$\Xi(\mu, V, T) = e^{\beta\mu N} Q(N, V, T), \quad (2.1)$$

and

$$Q(N, V, T) = \frac{1}{h^{3N} N!} \int\int d\boldsymbol{r}^N d\boldsymbol{p}^N exp[-\beta H(\boldsymbol{r}^N, \boldsymbol{p}^N)], \quad (2.2)$$

$Q(N,V,T)$ can be simplified by splitting the Hamiltonian into the kinetic and potential energy contributions, *i.e.* $H(\boldsymbol{r}^N, \boldsymbol{p}^N) = K(\boldsymbol{p}^N) + U(\boldsymbol{r}^N)$, which gives

$$Q(N,V,T)=\frac{1}{h^{3N}N!}\{\int d\boldsymbol{p}^{N}exp[-\beta K(\boldsymbol{p}^{N})]\}\{\int d\boldsymbol{r}^{N}exp[-\beta U(\boldsymbol{r}^{N})]\},\quad(2.3)$$

where the kinetic part can be expressed as

$$\int d\boldsymbol{p}^{N}exp[-\beta K(\boldsymbol{p}^{N})]=\int d\boldsymbol{p}^{N}e^{-\frac{\beta}{2m}\Sigma_{i}^{N}(p_{i,x}^{2}+p_{i,y}^{2}+p_{i,z}^{2})}=\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}.\quad(2.4)$$

From equations (2.1), (2.3) and (2.4), the grand canonical partition function can be derived

$$\Xi(\mu,V,T)=\frac{e^{\beta\mu N}}{\Lambda^{3N}N!}\int d\boldsymbol{r}^{N}exp[-\beta U(\boldsymbol{r}^{N})],\quad(2.5)$$

where $\Lambda=\left(\frac{h^{2}}{2\pi mk_{B}T}\right)^{1/2}$ is the average de Broglie wavelength for particles with temperature T. β equals $1/k_{B}T$, k_{B} is Boltzmann's constant, U is the potential energy for one microstate. The grand canonical ensemble probability density for a microstate can be estimated by

$$P_{N}(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N}})\propto\frac{e^{\beta\mu N}e^{-\beta U(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N}})}}{\Lambda^{3N}N!}.\quad(2.6)$$

The GCMC simulation is performed by generating a Markov chain of configurations where the successive microstates are determined from the preceding ones by randomly attempting one of the four types of movements. For each of the four types of moves, the movement acceptance probability can be generalized as

$$A_{\mathrm{trans/rot}}=\min\{1,e^{-\beta(U_{\mathrm{new}}-U_{\mathrm{old}})}\}\quad(2.7)$$

and the insertion acceptance probability can be written as

$$A_{\mathrm{insert}}=\frac{V\Lambda^{-3}}{N+1}e^{\beta\mu}e^{-\beta(U(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N+1}})-U(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N}}))}.\quad(2.8)$$

Analogously, for deletions, the deletion acceptance probability can be written as

$$A_{\mathrm{delete}}=\frac{N}{V\Lambda^{-3}}e^{-\beta\mu}e^{-\beta(U(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N-1}})-U(\boldsymbol{r_{1}},\boldsymbol{r_{2}},\ldots\boldsymbol{r_{N}}))}.\quad(2.9)$$

It is noticed that the chemical potential, μ, is a sum of the ideal gas chemical potential μ^{id} and the excess chemical potential μ^{ex}, *i.e.* $\mu = \mu^{id} + \mu^{ex}$, where the ideal gas chemical potential is related to the kinetic factor Λ,

$$\mu^{id} = k_B T \ln(\Lambda^3 N/V). \quad (2.10)$$

From equations (2.8), (2.9) and (2.10), Adams's formulation for the grand canonical ensemble insertion and deletion acceptance probabilities can be derived, *i.e.*,

$$A_{\text{insert}} = \frac{1}{N+1} e^{B} e^{-\beta(U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_{N+1}}) - U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_N}))}, \quad (2.11)$$

$$A_{\text{delete}} = N e^{-B} e^{-\beta(U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_{N-1}}) - U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_N}))}, \quad (2.12)$$

where the dimensionless Adams parameter $B = \beta\mu_{ex} + \ln N$. In our algorithm, the GCMC simulations are performed in the constant volume and temperature ensemble at each cycle. Equations (2.11) and (2.12) correspond to the random insertion algorithm, which is not efficient for GCMC simulations. In Mezei's cavity-biased insertion algorithm, equations (2.8) and (2.9) are modified by multiplying V (the volume of the GCMC region) by a dimensionless factor P_{cavity}, which is the probability of finding a void cavity in the GCMC region. Equations (2.11) and (2.12) then become

$$A_{\text{insert}} = \frac{P_{\text{cavity}}^{N}}{N+1} e^{B} e^{-\beta(U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_{N+1}}) - U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_N}))}, \quad (2.13)$$

$$A_{\text{delete}} = \frac{N}{P_{\text{cavity}}^{N-1}} e^{-B} e^{-\beta(U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_{N-1}}) - U(\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r_N}))}. \quad (2.14)$$

A cubic cavity-biased algorithm is implemented in our GCMC simulation program. As is illustrated in Fig. 1, the entire simulation box is compartmentalized into a large number of cubes. A cube is considered occupied if the cube contains one or more atoms. An array of all void cubes in the GCMC region is then constructed. During the GCMC simulations, insertions are attempted by placing the ion to a randomly selected void cube. The insertion attempts will be frequently rejected in the condensed phase system due to the high probability of atomic overlaps and the exponentially large vdW repulsive energy. To reduce the computational cost, an atom overlap check is performed, *i.e.* if the distance for an ion-water atom pair is found to be smaller than 0.5 Å, the insertion attempt is immediately rejected. Otherwise, the interaction energy of the inserted ion with its environment is

calculated using methods introduced below and the insertion acceptance probability is calculated using equation (2.13). For deletions, an ion within a GCMC region is randomly selected and the interaction energy of the ion with its environment is calculated. The probability of successful deletion is calculated using equation (2.14). The array of void cubes is updated every time an attempted movement (translation, rotation, insertion or deletion) is accepted.

2.2 Grid-searching method to calculate the short-range interaction energies

The interaction energy of an ion with its aqueous environment consists of the short-range Lennard-Jones (LJ) and electrostatic interaction energies and the long-range electrostatic energy. The long-range LJ interaction energy of a single ion with solvent is extremely weak relative to the electrostatic term, and hence this contribution is not considered in the current implementation. To further improve computational efficiency, a grid-searching algorithm is introduced to calculate the short-range interaction energies efficiently. The number of cubes in one dimension is

$$N_\alpha(\alpha = X, Y, Z) = \begin{cases} \frac{L_\alpha}{l} + 1, \text{if the remainder of } L_\alpha / l \neq 0 \\ \frac{L_\alpha}{l}, \text{if the remainder of } L_\alpha / l = 0 \end{cases}, \quad (2.15)$$

where L_α is the simulation box length and l is the size of the cube. An atom is deemed to occupy the nth cube in one dimension if the atomic coordinate (r_α) in that dimension satisfies $n_\alpha \leq (r_\alpha/1) < n_\alpha + 1$. The occupied cube's index number is thus obtained as

$$\text{index} = n_Z + n_Y \times N_Z + n_x \times N_Y \times N_Z. \quad (2.16)$$

In this way, all atomic coordinates in the simulation box are coupled to the cube index numbers by constructing a two-dimensional cube-atom array, where the first and second dimensions of the array store the cube index number and the atomic coordinate, respectively. This coupling array allows for quick identification of an atom's neighboring atoms within a cutoff distance in the GCMC simulation energy calculations. For instance, if one attempts to insert an ion into a position in the simulation box, as is illustrated in Fig. 1, to search for the neighboring atoms surrounding the ion and to calculate the ΔU, the following steps are performed:

(1) Determine n_x, n_y, n_z for the cube occupied by the ion, with this grid denoted G_{ion}.

(2) Search the neighboring cubes of G_{ion} within a cutoff distance and obtain those neighboring cube index numbers using equation (2.16). In this step, the periodic boundary condition is considered.

(3) Obtain the atomic coordinates in each of the neighboring cubes from the two-dimensional cube-atom array.

(4) Calculate the inter-particle energy if the inter-particle distance is within the cutoff distance. In the present GCMC simulations, the cutoff distance is set to 1 nm for the short-ranged LJ and electrostatic interactions.

2.3 Implementation of smooth particle mesh Ewald

The smooth PME method introduced by Essmann *et al.* is implemented in the GCMC program to account for the long-range electrostatic energy. The total electrostatic energy of a *N*-particle system is composed of four terms,

$$U_{\text{elec}} = U_{\text{short}} + U_{\text{long}} + U_{\text{corr}} - U_{\text{self}}, \quad (2.17)$$

U_{short} represents short-ranged electrostatic energy calculated in the real space,

$$U_{\text{short}} = \frac{1}{2} \Sigma_{\boldsymbol{n}} \ \Sigma_{i,j=1}^{N} \frac{q_i q_j \text{erfc}(\beta \ | r_j - r_i + \boldsymbol{n} \ |)}{| r_j - r_i + \boldsymbol{n} \ |}, \quad (2.18)$$

where q_i, q_j are charges of two interacting atoms; *erfc* is the complementary error function; β is the Ewald splitting parameter which determines the relative convergence rates and the values of U_{short} and U_{long}. An error tolerance value of 0.000001 and an electrostatic real-space cutoff distance R_{elec} of 1 nm are used to determine the value of β, using the relationship $\text{erfc}(\beta \times R_{\text{elec}})/R_{\text{elec}} \leq 0.000001$. The vector $\boldsymbol{n}=n_X L_X+n_Y L_Y+n_Z L_Z$, where L_X, L_Y, L_Z are the lengths of the three simulation box edges; n_X, n_Y and n_Z are integers indicating the use of periodic boundary condition to calculate the U_{short}. In this work, U_{short} is calculated using a grid-searching method, as introduced above. U_{long} represents the long-range electrostatic energy,

$$U_{\text{long}} = \frac{K_X K_Y K_Z}{2} \Sigma_{m_X=0}^{K_X-1} \ \Sigma_{m_Y=0}^{K_Y-1} \ \Sigma_{m_Z=0}^{K_Z-1} Q(m_X, m_Y, m_Z) \times F^{-1}[BC(m_X, m_Y, m_Z) \times F(Q)(m_X, m_Y, m_Z)], \quad (2.19)$$

where K_X, K_Y, K_Z are the number of cubic PME grid points in three dimensions and the distance between two neighboring grid points is set to 0.12 nm; m_X, m_Y, m_Z are PME grid point indexes. Equation (2.19) suggests that the PME grid point charge array Q is firstly constructed in real space and then transformed from real space to Fourier space using the fast Fourier transform algorithm, as indicated by $F(Q)$. The charge of a grid point with index of (m_X, m_Y, m_Z) is given by

$$Q(m_X, m_Y, m_Z) = \ \Sigma_{i=1}^{N} q_i M_n(\Delta X) \times M_n(\Delta Y) M_n(\Delta Z), \quad (2.20)$$

where ΔX, ΔY, and ΔZ are the distances between the atom and the selected cubic grid point in X, Y, Z dimensions; n is the cardinal B-spline order parameter and its value is set to 4 in our program. This value suggests that an atomic charge is only assigned to 4 neighboring grid points in one dimension (a total of 64 grid points in three dimensions), with the grid index, G_{index}, and the integer part of the atomic *scaled* coordinates S_α (α=X, Y, Z) satisfying G_{index}= S_α-n+k, with k=1, 2, 3, 4. Periodic boundary condition is considered when assigning the atom point charges to the grid points. Mn is calculated by,

$$M_n(u) = \frac{1}{(n-1)!}\,\Sigma_{i=0}^{n}\,(-1)^i \frac{n!}{i!(n-i)!}(u-i)_+^{n-1}. \quad (2.21)$$

Equation (2.19) also suggests that, in the Fourier space, a BC array is multiplied by $F(Q)$ and the obtained result is then reverse transformed from Fourier space to real space to do the remaining calculations. The BC array is the product of the independent B and C arrays. The elements in the B array are determined by

$$B(m_X, m_Y, m_Z) = \,|\, b_X(m_X) \,|^2 \cdot \,|\, b_Y(m_Y) \,|^2 \cdot \,|\, b_Z(m_Z) \,|^2, \quad (2.22)$$

$b_\alpha(m_\alpha)$ are complex numbers and their values are calculated using the following equation

$$b_\alpha(m_\alpha) = exp(2\pi i(n-1)m_\alpha \,/\, K_\alpha) \times [\,\Sigma_{k=0}^{n-2}\, M_n(k+1) exp(2\pi i m_\alpha k \,/\, K_\alpha)]^{-1}, \quad (2.23)$$

where n is the B-spline order parameter. The C array is defined by

$$C(m_X, m_Y, m_Z) = \begin{cases} \dfrac{1}{\pi V}\dfrac{exp(-\pi^2 \boldsymbol{m}^2 \,/\, \beta^2)}{\boldsymbol{m}^2}, & \text{if } \boldsymbol{m} \neq 0 \\ 0, & \text{if } \boldsymbol{m} = 0 \end{cases} \quad (2.24)$$

V is the volume of the simulation box**,** $\boldsymbol{m}$ is a vector and its value is determined by

$$\boldsymbol{m} = \frac{m'_X}{L_X} + \frac{m'_Y}{L_Y} + \frac{m'_Z}{L_Z}, \quad (2.25)$$

$m'_\alpha = m_\alpha$ if $0 \le m_\alpha \le K_\alpha/2$, otherwise $m'_\alpha = m_\alpha - K_\alpha$. The long-range electrostatic energy of an ion with the environment is calculated by

$$U_{\text{long}} = U_{\text{long, environment + ion}} - U_{\text{long,environment}} - U_{\text{long,ion}}, \quad (2.26)$$

where $U_{\text{long, environment+ion}}$ is the long-range electrostatic energy for the system, $U_{\text{long, environment}}$ is the long-range electrostatic energy for the environment and $U_{\text{long,ion}}$ is

the long-range electrostatic energy for a single ion in a periodic simulation box with size equal to the that of the whole simulation system. In the PME summation method, the charge density at position $\boldsymbol{r}$ in the simulation box is expressed as

$$\rho_{\mathrm{SPME}}(\boldsymbol{r}) = \rho(\boldsymbol{r}) - \sum_{i=0}^{N} \frac{q_i}{V_{\mathrm{box}}} \quad (2.27)$$

where $\rho(\boldsymbol{r})$ is the charge density computed from Coulomb's law, q_i is the atomic charge and V_{box} is the volume of the simulation box. Equation (2.27) suggests that an implicit and uniform background neutralizing charge will be created if the system has a net charge. To account for the interactions between the point charge and the background charge in real space, a net charge energy correction, i.e., U_{corr} is included. If the system has a net charge of Δq,

$$U_{\mathrm{corr}} = -\frac{\pi \times \Delta q^2}{2\beta^2 \times V_{\mathrm{box}}}. \quad (2.28)$$

Finally, the self energy $U_{\mathrm{self}} = \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2$ is not considered in the GCMC program because only the inter-molecular interaction energy needs to be calculated in the GCMC simulation.

2.4 Finite-box correction for long-range electrostatics

It has been known that application of PBC and PME leads to the finite-box size artifact in the prediction of ionic HFE. Hummer et al. and Figueirido et al. have proposed to add the Wigner self-interaction energy term to correct for this artifact. Hence, the ionic HFE values predicted from an infinitely large simulation box ($\Delta G(\infty)$) can be approximated by

$$\Delta G(\infty) \approx \Delta G(L) - \frac{1}{4\pi\varepsilon_0} \times \frac{q^2 \zeta}{2\varepsilon L}, \quad (2.29)$$

where the dimensionless constant $\zeta = 2.837297$, ε is the dielectric constant of the solvent, q is the ion charge with unit of Coulomb and L is the simulation box size. It can be seen that for solvents with low dielectric constant, the finite-box correction to the ionic HFE is not negligible.

In our GCMC simulation program with implementations of PBC and PME, the short-range ion-solvent (LJ and electrostatic) interaction energies are calculated by using the grid-searching method, as introduced above. The long-range electrostatic interaction energy can be derived from equation (2.26). Of note, the long-range electrostatic energy of an ion in a periodic simulation box is not zero, but instead this long-range electrostatic energy calculated from PME approach increases as the simulation box size is increased. It is expected that, at an infinitely large simulation box, the sum of the short-range and long-range electrostatic energy is zero, i.e., $U_{\mathrm{long}}(\infty) = -U_{\mathrm{short}}$. In GCMC simulations, the long-

range ion-solvent electrostatic interaction energy as well as the total interaction energy from an infinitely large simulation box can be estimated using an equation analogous in functional form to equation (2.29)

$$\left\langle U(\infty)\right\rangle = \left\langle U(L)\right\rangle - \frac{1}{4\pi\varepsilon_0}\times\frac{q^2\zeta}{\varepsilon L}, \quad (2.30)$$

The correction term is derived from linear response theory, i.e. $<U_{\text{elec}}(Q)>/2=\Delta G_{\text{elec}}$. To validate equation (2.30), we have run 1 ns MD simulations of one single ion solvated by water in a cubic simulation box. 1,000 configurations are extracted from the MD simulations. Eight different box sizes, i.e. L=4.8 nm, 7.2 nm, 9.6 nm, 12.0 nm, 14.4 nm, 16.8 nm and 19.2 nm are used to compute the long-range electrostatic and total interaction energy of an ion with its aqueous environment. Both GROMACS and the GCMC simulation programs are used for this calculation.

2.5 Parallelization of the GCMC simulation program

As illustrated in Fig. 1, all small molecules in the GCMC region are assigned into a few parallel GCMC subdomains. A molecule is assigned to a subdomain if the center-of-mass position of the molecule r_{com} and the lower and upper boundary positions r_{lower}, r_{upper} of the sub-GCMC domains satisfy $r_{\text{lower}} \leq r_{\text{com}} < r_{\text{upper}}$. After the cube-atom coupling array, the PME *BC* array, and the PME *Q* array and are constructed, each of the GCMC subdomains receives a copy of these arrays and GCMC simulations are run in parallel in each of the GCMC subdomains. During the GCMC simulations, the *BC* array remains fixed while the *Q* array and the cube-atom coupling array are dynamically updated once a GCMC movement is accepted. Of note, our GCMC parallelization scheme indicates that a molecule within a GCMC subdomain cannot sense the change in coordinates of particles in other subdomains. As such, it is necessary to re-construct the PME *Q* array and the cube-atom coupling array periodically, typically every 1000 GCMC steps. OpenMP is used for the program parallelization. The validations of the parallelization scheme and the robustness of interaction energy calculation of the GCMC simulation program are presented in Supporting Information.

2.6 Ionic hydration free energy calculations

To evaluate the developed GCMC/MD method, the Bennett acceptance ratio (BAR) approach is used to compute the ionic HFEs for five alkali ions (Li^+, Na^+, K^+, Rb^+, Cs^+), four halide ions (F^-, Cl^-, Br^-, I^-) and two divalent cations (Mg^{2+} and Ca^{2+}) for comparison.

2.6.1 Bennett Acceptance Ratio (BAR)—In this approach, the free energy difference between two close thermodynamic states is derived as

$$A_1 - A_2 = -k_B T \frac{\int e^{-\beta U_2(r^N)} dr^N}{\int e^{-\beta U_1(r^N)} dr^N}$$
$$= -k_B T \ln\left(\frac{\int e^{-\beta U_2(r^N)} dr^N}{\int \omega(r^N) e^{-\beta U_1(r^N) - \beta U_2(r^N)} dr^N} \times \frac{\int \omega(r^N) e^{-\beta U_1(r^N) - \beta U_2(r^N)} dr^N}{\int e^{-\beta U_1(r^N)} dr^N} \right)$$
$$= -k_B T \ln \frac{\left\langle \omega e^{-\beta U_2} \right\rangle_1}{\left\langle \omega e^{-\beta U_1} \right\rangle_2}, \quad (2.31)$$

where k_B is the Boltzmann constant, T is the temperature, N is the number of atoms, and $\omega(r^N)$ is an arbitrary weighting function. The strategy in the BAR method is to find an optimal ω value that minimizes the expected statistical error in the calculated free energy difference. The optimal weighting function has the form

$$\omega(r^N) \propto (n_1^{-1} e^{-\beta A_1 - \beta U_2(r^N)} + n_2^{-1} e^{-\beta A_2 - \beta U_1(r^N)})^{-1}. \quad (2.32)$$

The variables n_1 and n_2 are the number of trajectory configurations used in the free energy average for states1 and 2. If we assume that these are the same, we have

$$\omega(r^N) \propto (e^{-\beta A_1 - \beta U_2(r^N)} + e^{-\beta A_2 - \beta U_1(r^N)})^{-1}. \quad (2.33)$$

Finally, the following equation is derived

$$\Delta A_{1\to 2} = k_B T \ln \left[\left\langle \frac{1}{1 + e^{-\beta \Delta U + \Delta A_{1\to 2}}} \right\rangle_1 / \left\langle \frac{1}{1 + e^{-\beta \Delta U + \Delta A_{1\to 2}}} \right\rangle_2 \right] \quad (2.34)$$

The above equation can be solved by iterating until the free energy difference converges. However, the free energy value cannot converge if the two states are significantly separated in the configurational space. In practice, multiple intermediates states represented by λ are needed to bridge the two-end states. And hence, the free energy difference between state A and B can be expressed as

$$\Delta A_{1\to 2} = \Sigma \Delta A_{\lambda i \to \lambda(i+1)}. \quad (2.35)$$

For the present study, the simulation system contains one ion and 450 TIP3P water molecules, with a simulation box size of 2.4 nm. To investigate the effect of box size, we also perform BAR calculations using a larger box size of 4 nm, containing one ion and 2138 water molecules. The BAR calculations are carried out at a constant isotropic pressure of 1 atm using Parrinello-Rahman barostat and a temperature of 300 K using Nosé-Hoover thermostat,[-] as implemented in the GROMACS 5.1.0 software package. Integration of equation of motion is performed using leapfrog stochastic dynamics. The two-step procedure for decoupling LJ (cavity growth) and electrostatic (atom charging) interactions independently is employed. A series of λ=0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 are used for the cavity growth simulations and λ=0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 are used for the atom charging simulations. The cutoff distance for both short-range LJ and electrostatic interactions is 1 nm, and the long-range electrostatic interactions are treated using PME. Each window is simulated for 350 ps and the first 100 ps simulation is discarded for equilibration. It has been previously reported that the convergence can be reached within a few hundred picoseconds for MD simulations of ionic solutions using the additive CHARMM and Drude polarizable force fields.[-] A soft core potential with α=1, σ=0.3nm and λ-power of 1 is used in the cavity growth simulations in order to avoid singularities.

2.6.2 Grand Canonical Monte Carlo/Molecular Dynamics (GCMC/MD) simulations—

Eleven simulation systems are constructed and equilibrated in the NPT ensemble for 1 ns for subsequent GCMC/MD simulations. The cubic simulation box of edge-length 4.6 nm contains a single type of ion, with 11 ions and 3320 water molecules yielding a starting ion concentration of 0.15 M. To investigate whether the number of ions has an effect on the calculated free energy, we also carried out GCMC/MD simulations of Na^+ solutions using different initial number of Na^+ ions. Accordingly, the simulation systems have net charges with the net charge implicitly neutralized by the PME neutralizing background charge in the simulation systems. The iterative GCMC/MD cycles are performed in following steps: (1) 200,000 steps of GCMC simulation (insertion, deletion, translation and rotation) are run for water and ions. The initial μ_{ex} for water and ions are set to their experimental HFE values, i.e., −5.6 kcal/mol for water and the experimental HFE values for ions are listed in Table 1. We have performed additional simulations for Cl^- and Na^+ using different initial μ_{ex} values to show that the simulation results are not influenced by the chosen initial values. (2) After completion of the 200,000 steps of GCMC simulations, 1000 steps of steepest-descent energy minimization is performed followed by a 350 ps MD simulation are run in the NPT ensemble. (3) Vary the values of μ_{ex} for ions by $d\mu_{ex}$, and use this new μ_{ex} value for next iteration of GCMC simulations. The magnitude of $d\mu_{ex}$ is determined using the method introduced in the Supporting Information. (4) Perform 400 cycles of the GCMC/MD simulations from which average values of μ_{ex} are determined for ions. Throughout the GCMC simulations, the μ_{ex} value for water is kept constant at −5.6 kcal/mol.

2.7 Force fields and MD simulation details

The GROMACS 5.1.0 software package is used for the energy minimization and MD simulations. The LJ parameters for the ions are listed in Table 1. The CHARMM TIP3P model is used for water. During MD simulations, the Nosé-Hoover thermostat[-] is used to maintain the system temperature at 300 K and the Parrinello-Rahman barostat is used to maintain the system pressure at 1 atm. The LINCS algorithm is used to constrain the water geometry. LJ interactions are switched off smoothly in the range of 1-1.2 nm and the PME method is used to treat long-range electrostatics with a real space cutoff distance of 1.2 nm, with the order of B-spline interpolation set to 4 and the maximum grid spacing set to be 0.12 nm. The GROMACS long-range dispersion correction to the energy and pressure is applied.

3. Results

The primary aim of this work is developing a GCMC/MD approach in which the long-range electrostatic interactions are treated using the PME method, and then evaluating this approach for predicting ionic HFE. In the following paragraphs, we first present the ionic HFE results predicted with the BAR approach. The results obtained from the BAR approach are used to validate the developed GCMC/MD approach.

3.1 Ionic HFE predicted with the BAR approach

The absolute ionic HFEs are calculated using the BAR approach implemented in GROMACS. The thermodynamic cycle used for deriving the absolute ionic HFE, ΔG1, is illustrated in Fig. 2. ΔG1 is equal to the sum of the free energy for annihilating the ion in vacuum, ΔG2, and the free energy for growing a dummy particle into a fully charged ion in water, ΔG4. ΔG1 is also equivalent to the free energy for transferring an ion from vacuum to water if the vacuum/water interfacial electric potential is not considered. In this work, the contribution of the vacuum/water interface to ΔG1 is not considered and thus the ΔG1 value can be called the intrinsic absolute ionic HFE. This intrinsic absolute HFE can be directly compared with the experimental data from Markus, who used the tetraphenylarsonium tetraphenylborate (TATB) extrathermodynamic assumption to determine the ionic HFE. The TATB approach assumes that the TA cation and the TB anion have identical thermodynamics in water and the measured HFE for the TATB ion pair can be equally divided between the two ions. In this approach for estimation of the experimental HFE, the vacuum/water interface has no effect on the gaseous-aqueous partitioning of the neutral ion pair.

Accurate prediction of ionic HFE using the thermodynamic cycle illustrated in Fig. 2 involves *ad hoc* corrections, and explicitly taking into account these corrections is important for developing robust force field parameters for ions. Here, we only concentrate on the finite-box correction, as the energetic contributions of other corrections, including the long-range dispersion correction and the thermodynamic volume correction are relatively small. The absolute ionic HFEs using the BAR approach are calculated using two different box sizes, L=2.4 nm and L=4 nm, and the simulation results are listed in Table 2. It can be seen that the box size has a dramatic effect on the ΔG2 values whereas it has negligible effect on the ΔG4 values.

This box size effect associated with ΔG2 for ions in vacuum is attributed to the use of periodic boundary conditions and PME to treat electrostatics, i.e., an isolated ion in the periodic simulation box (vacuum) can interact with its periodic images via short-range and long-range electrostatic interactions. We used GROMACS to calculate the short-range and long-range electrostatic energies for a single ion in a periodic simulation box. It is found that the derived short-range electrostatic energy (U_{short}) is insensitive to the box size change. For the two box sizes of L=2.4 nm and L=4 nm, U_{short}= −58.52 kcal/mol for all monovalent ions and U_{short}= −234.08 kcal/mol for all divalent ions. However, the long-range electrostatic energy (U_{long}) is found to increase with the simulation box size. For monovalent ions, U_{long} increases from 38.89 kcal/mol to 46.74 kcal/mol as L increases from 2.4 nm to 4 nm. For divalent ions, U_{long} increases from 155.55 kcal/mol to 186.96 kcal/mol.

Using equation (2.30), the Wigner self-interaction corrections for ΔG2 and ΔG4 can be estimated. In water ($\varepsilon \approx 80$), the Wigner self-interaction correction energy (U_{corr}) is around 0.24 kcal/mol for monovalent ions and 0.96 kcal/mol for divalent ions when L=2.4 nm. The U_{corr} values decrease slightly to 0.15 kcal/mol for monovalent ions and 0.60 kcal/mol for divalent ions when L increases to 4 nm. Thus, U_{corr} can almost be neglected in water due to the high dielectric constant of water. For ions in vacuum ($\varepsilon \approx 1$), U_{corr}=19.2 kcal/mol and U_{corr}=76.8 kcal/mol for monovalent and divalent ions, respectively when L=2.4 nm. And these values decrease to 12.0 kcal/mol and 48.0 kcal/mol when L=4 nm (Table 2). This suggests that the ion-ion image interactions must be considered when the PME method is used to treat charged compounds in solvents with low dielectric constant, such as cyclohexane.

Here, we present an analytical method to make corrections to the finite box size artifact. When the PME method is used to treat electrostatics, the sum of U_{short} and U_{long} is expected to be zero in an infinitely large simulation box, i.e. $U_{\text{short}} + U_{\text{long}}(\infty)$ As is shown above, placing an ion in a finite-sized simulation box, the energy of the ion is found to be negative due to the underestimated long-range electrostatic energy. This suggests a linear method to make the finite box size correction for ionic free energy in vacuum, i.e.,

$$\Delta G(\infty) = \Delta G(L) - \left(U_{\text{long}}(\infty) - U_{\text{long}}(L)\right) = \Delta G(L) - \left(-U_{\text{short}} - U_{\text{long}}(L)\right). \quad (3.1)$$

The linear correction energies are listed in Table 2 and it can be seen that the linear correction energies are almost identical to ΔG2 and the Wigner self-interaction energies. This suggests a relationship

$$U_{\text{long}}(\infty) - U_{\text{long}}(L) \approx \frac{1}{4\pi\varepsilon_0} \times \frac{q^2\zeta}{2L}. \quad (3.2)$$

3.2 Effect of inclusion of long-range electrostatics on the ionic HFE predicted with the GCMC/MD approach

The developed GCMC/MD simulation approach works by continuously varying the ionic μ_{ex} value over 400 cycles of GCMC/MD simulations. The GCMC simulation is firstly performed with the ionic μ_{ex} set to at an assigned value followed by MD simulations allowing for relaxation of the entire simulation system. In the next cycle of GCMC simulations, the new ionic μ_{ex} is determined based on the difference between the actual and the target number of ions in the simulation system (see methods in Supporting Information). The simulation protocol is similar to our previous work. A major improvement in the new GCMC simulation program is the implementation of the grid searching and the PME algorithms to calculate the ionic solute-water interaction energies more accurately. Herein, we compare the ionic HFEs predicted with the new PME GCMC/MD approach and our previous GCMC/MD approach. The finite-box correction is not considered in this comparison in order to demonstrate the pure effect of inclusion of long-range electrostatics on the simulation results.

Fig. 3 plots the evolution of the ionic μ_{ex} during the 400 cycles of PME GCMC/MD simulations. For all the studied ions, the μ_{ex} fluctuates between a minimum value (μ_{ex}^{min}) and a maximum value (μ_{ex}^{max}). The μ_{ex}^{min} and μ_{ex}^{max} values correspond to the excess chemical potentials needed to remove all ions in our simulation box from the "organized" (water rearranges surrounding the ion) solutions and to insert ~0.15 M concentration of ions into bulk-like water solutions during the GCMC simulations. The variations of the number of ions in the solutions are shown in Fig. S2. It can be seen from Fig. 3 that the Markus experimental ionic HFEs (shown in Table 1) can be approximated using a linear relationship, i.e.,

$$\Delta G = (\overline{\mu_{ex}^{min}} + \overline{\mu_{ex}^{max}}) \Big/ 2, \quad (3.3)$$

where $\overline{\mu_{ex}^{min}}$ and $\overline{\mu_{ex}^{max}}$ are the averaged μ_{ex}^{min}, μ_{ex}^{max} values over the 400 cycles of GCMC/MD simulations. The obtained $\overline{\mu_{ex}^{min}}$, $\overline{\mu_{ex}^{max}}$ and the associated ionic HFEs values are listed in Table 3. Of note, the functional form of equation (3.3) resembles the linear response equation derived by Raineri et al, i.e.,

$$\Delta G = (U_{uv}^{0} + U_{uv}) \Big/ 2, \quad (3.4)$$

where $\overline{\mu_{ex}^{min}}$ and U_{uv} are the averaged solute-solvent interaction energies in the bulk-like and ion-organized solutions.

To investigate the effect of inclusion of long-range electrostatics on the GCMC/MD simulation results, we plot in Fig. 4 the evolution of the ionic μ_{ex} determined from our previously developed GCMC/MD approach (denoted cut-off GCMC/MD). The obtained

$\overline{\mu_{ex}^{max}}$, $\overline{\mu_{ex}^{max}}$ and HFE values are listed in Table 3. Compared with the PME GCMC/MD simulation results, the cut-off GCMC/MD predicted $\overline{\mu_{ex}^{min}}$ and $\overline{\mu_{ex}^{max}}$ values for the monovalent cations increase and decrease by ~10 kcal/mol, respectively. Thus, the ionic HFEs for the monovalent cations are not significantly affected when the long-range electrostatics are ignored. However, for the monovalent anions, the cut-off GCMC/MD predicted $\overline{\mu_{ex}^{min}}$ values increase by ~40 kcal/mol and the $\overline{\mu_{ex}^{max}}$ values increase by ~20 kcal/mol. As a result, the cut-off GCMC/MD predicted ionic HFEs for monovalent anions are ~30 kcal/mol higher than the corresponding results predicted with the PME GCMC/MD approach. For the divalent Ca^{2+} and Mg^{2+}, the cut-off GCMC/MD predicted ionic HFE values are also increased compared with the PME GCMC/MD results. These results suggest that long-range electrostatics have more profound effect on the hydration thermodynamics of monovalent anions and divalent cations. To investigate whether the long-range electrostatics has such marked effect on the predicted ionic HFE using the FEP/BAR approach, we calculated the HFE values for Na^+ and Cl^- using the BAR approach without PME. The results are −85.09 kcal/mol for Na^+ and −88.59 kcal/mol for Cl^-, which are only 4.5 and 4.1 kcal/mol, respectively, less favorable than the PME BAR results listed in Table 2. Thus, in the case of anions, the change in the HFE associated with the inclusion of PME in the GCMC/MD calculations is impacted in part by changes in the sampling of the ensemble of conformations.

3.3 Effect of finite box correction on GCMC simulations

Given the impact of long-range electrostatic interactions associated with the introduction of PME on the HFE values from the GCMC/MD calculations it is necessary to consider the impact of a finite box correction. Given that the acceptance rates is impacted by the electrostatic term, such a correction is of particular importance. Fig. 5 presents the interaction energy of a single Cl^-, Na^+ or Mg^{2+} with water as a function of simulation box size. Our PME GCMC program predicts consistent results with GROMACS, and it shows that the ion-water interaction energy becomes more favorable as the simulation box size is increased.

From equation (3.2) and the linear response theory, a relationship can be derived to couple the ion-water electrostatic interaction energies in a large simulation box (U_{elec}(large)) and in a small simulation box (U_{elec}(small))

$$\left\langle U_{elec}(\text{large})\right\rangle = \left\langle U_{elec}(\text{small})\right\rangle - 2\times(U_{long}(\text{large}) - U_{long}(\text{small})), \quad (3.5)$$

where U_{long} is the long-range electrostatic energy for a single ion in a cubic simulation box. To verify equation (3.5), we used the simulated U_{elec} values from a small simulation box with size of 4.8 nm to predict the U_{elec} values for simulation box with size of 19.2 nm. The results are listed in Table 4. It can be seen from Table 4 that the predicted U_{elec} values and the simulated U_{elec} values for simulation box with size of 19.2 nm are in good agreement.

Using equations (3.3) and (3.5), we come to equation (2.31) to predict the ion-environment interaction energy in an infinitely large simulation box. The obtained μ_{ex}^{min}, μ_{ex}^{max} and HFE values using the finite box corrections in the GCMC/MD simulations with PME are listed in Table 5. It can be seen that the finite box corrections leads to the HFEs becoming systematically more favorable, as expected, with the finite box correction ranging from approximately −15 kcal/mol for monovalent ions to approximately −60 kcal/mol for divalent ions.

3.4 Comparison of experimental, BAR and GCMC hydration free energies

In Fig. 6 we compare the Markus experimental HFEs with the simulated ionic HFEs using the FEP/BAR and GCMC/MD approaches (with PME and finite box corrections). For monovalent cations, the BAR approach predicts consistent results with experiment, while for monovalent anions, the BAR approach overestimates the ionic HFEs. For all the studied ions, the GCMC/MD simulations predict more favorable HFEs compared with the BAR approach. This is mainly attributed to the finite-box corrections in the GCMC simulations, which is important for GCMC simulations of divalent Mg^{2+}, as is shown in Fig. 6(c). In addition, the FEP/BAR results are obtained at infinite dilution, while the GCMC/MD results are for a concentration of 0.15 M. Finally, we note that the level of agreement between the calculated and experimental HFE values is inherently limited by the force field parameters themselves. The emphasis in the present study is the ability of the PME GCMC/MD method to satisfactorily reproduce the trends obtained with the FEP/BAR calculated values.

4. Discussion and Conclusions

Combined GCMC/MD simulation methods have proven to be a powerful technique for solvent-free and porous adsorbent systems.[-] However, in condensed phase system with explicit solvent, obtaining adequate insertion/deletion acceptance rates poses a critical challenge for GCMC simulations. On one hand, the sizes of the intrinsic cavities in bulk water and biomolecules are usually too small to efficiently insert solutes into such system. In addition, for charged solutes, the insertion attempt is frequently rejected due to the strong electrostatic repulsion between the charged solute and the bulk-like water molecules. On the other hand, the difference between the interaction energy of solute with solvent and the solute HFE (or equivalently the solute excess chemical potential) is often too large to efficiently delete the solute from the system. To overcome this, we introduced the oscillating excess chemical potential GCMC simulation approach, in which the solute excess chemical potential is determined by the difference between the actual and the target number of solute in the simulation system. In this work, we extend that approach to more accurately treat the electrostatics by implementing a grid-searching method for the rapid evaluation of short-range interactions and PME and finite box corrections for the treatment of long-range electrostatic interactions. Other studies using GCMC simulations to improve the sampling of water in condensed phase simulations include the work of Mezei and coworkers that used an increased chemical potential to facilitate water sampling around DNA and the replica-exchange GCMC method proposed by Essex and co-workers. Also of interest is the recent use of non-equilibrium Monte Carlo to achieve the sampling of Na^+ and Cl^- in aqueous systems.

In the simulations of ionic solutions in this work, the target value driving ion-insertion and ion-deletion processes are set to zero and eleven, corresponding to the ionic concentration of ~0.15 M. By varying the excess chemical potentials for the ions, minimum and maximum ionic μ_{ex} values may be obtained. It is found that the experimental ionic HFEs can be approximated using a linear relationship coupling the maximum μ_{ex} for ion insertion and the minimum μ_{ex} for ion deletion. PME combined with the finite box size corrections for long-range electrostatics yields more favorable ionic HFEs for monovalent ions than the FEP/BAR results. While for divalent ions, the finite box corrections is important for deriving improved ionic HFE results in better agreement with BAR. Differences in the targeting of infinite dilute vs. 0.15 M concentrations in the FEP/BAR and GCMC/MD methods, respectively, and the inherent limitations in the force field parameters are noted.

This work shows that, with implementation of the grid searching and the PME methods for short-range and long-range interaction energy calculations, the presented GCMC/MD approach can be used to estimate the HFEs of ions in aqueous solution, offering a significant improvement in the accuracies of the ionic HFE prediction compared with our previously developed approach. In future work, the developed GCMC/MD approach will be applied to study more complex ionic systems containing biomolecules such as ion channels, DNA, RNA and calcium binding proteins, among others, as well as in the site identification by ligand competitive saturation (SILCS) functional group mapping approach.[-]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Support from the NIH (GM109635 and GM051501) is acknowledged.

References

1. Sun DL; Forsman J; Woodward CE, Evaluating Force Fields for the Computational Prediction of Ionized Arginine and Lysine Side-Chains Partitioning into Lipid Bilayers and Octanol. J. Chem. Theory Comput 2015, 11, 1775–1791. [PubMed: 26574387]
2. Born MZ, Volumen Und Hydratationswärme Der Ionen. Physik 1920, 1:45.
3. Roux B; Yu HA; Karplus M, Molecular Basis for the Born Model of Ion Solvation. J. Phys. Chem 1990, 94, 4683–4688.
4. Åqvist J; Hansson T, On the Validity of Electrostatic Linear Response in Polar Solvents. J. Phys. Chem 1996, 100, 9512–9521.
5. Grossfield A; Ren PY; Ponder JW, Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. J. Am. Chem. Soc 2003, 125, 15671–15682. [PubMed: 14664617]
6. 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. [PubMed: 16494345]
7. Åqvist J; Medina C; Samuelsson JE, A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. Protein Eng. 1994, 7, 385–391. [PubMed: 8177887]
8. Lakkaraju SK; Raman EP; Yu WB; MacKerell AD, Jr., Sampling of Organic Solutes in Aqueous and Heterogeneous Environments Using Oscillating Excess Chemical Potentials in Grand Canonical-Like Monte Carlo-Molecular Dynamics Simulations. J. Chem. Theory Comput 2014, 10, 2281–2290. [PubMed: 24932136]

9. Ross GA; Bodnarchuk MS; Essex JW, Water Sites, Networks, and Free Energies with Grand Canonical Monte Carlo. J. Am. Chem. Soc 2015, 137, 14930–14943. [PubMed: 26509924]
10. Ross GA; Macdonald HEB; Cave-Ayland C; Martinez AIC; Essex JW, Replica-Exchange and Standard State Binding Free Energies with Grand Canonical Monte Carlo. J. Chem. Theory Comput 2017, 13, 6373–6381. [PubMed: 29091438]
11. Ben-Amotz D; Underwood R, Unraveling Water's Entropic Mysteries: A Unified View of Nonpolar, Polar, and Ionic Hydration. Acc. Chem. Res 2008, 41, 957–967. [PubMed: 18710198]
12. Woo HJ; Dinner AR; Roux B, Grand Canonical Monte Carlo Simulations of Water in Protein Environments. J. Chem. Phys 2004, 121, 6392–6400. [PubMed: 15446937]
13. Bai P; Siepmann JI, Assessment and Optimization of Configurational-Bias Monte Carlo Particle Swap Strategies for Simulations of Water in the Gibbs Ensemble. J. Chem. Theory Comput 2017, 13, 431–440. [PubMed: 28001386]
14. Graziano G, Water: Cavity Size Distribution and Hydrogen Bonds. Chem. Phys. Lett 2004, 396, 226–231.
15. Lemkul JA; Lakkaraju SK; MacKerell AD, Jr., Characterization of Mg2+ Distributions around Rna in Solution. ACS Omega 2016, 1, 680–688. [PubMed: 27819065]
16. Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG, A Smooth Particle Mesh Ewald Method. J. Chem. Phys 1995, 103, 8577–8593.
17. Lin YL; 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. [PubMed: 26586504]
18. Bennett CH, Efficient Estimation of Free Energy Differences from Monte Carlo Data. J. Comput. Phys 1976, 22, 245–268.
19. Attard P, A Grand Canonical Simulation Technique for Dense and Confined Fluids with Application to a Lennard-Jones Fluid. J. Chem. Phys 1997, 107, 3230–3238.
20. Adams DJ, Chemical Potential of Hard-Sphere Fluids by Monte Carlo Methods. Mol. Phys 1974, 28, 1241–1252.
21. Mezei M, A Cavity-Biased (T, V, M) Monte Carlo Method for the Computer Simulation of Fluids. Mol. Phys 1980, 40, 901–906.
22. Hummer G; Pratt LR; García AE, Free Energy of Ionic Hydration. J. Phys. Chem. B 1996, 100, 1206–1215.
23. Figueirido F; Del Buono GS; Levy RM, On Finite-Size Corrections to the Free Energy of Ionic Hydration. J. Phys. Chem. B 1997, 101, 5622–5623.
24. Parrinello M; Rahman A, Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. J. Appl. Phys 1981, 52, 7182–7190.
25. Hoover WG, Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys. Rev. A 1985, 31, 1695–1697.
26. Nosé S, A Molecular Dynamics Method for Simulations in the Canonical Ensemble. Mol. Phys 1984, 52, 255–268.
27. van der Spoel D; Lindahl E; Hess B; Groenhof G; Mark AE; Berendsen HJC, Gromacs: Fast, Flexible, and Free. J. Comput. Chem 2005, 26, 1701–1718. [PubMed: 16211538]
28. Hub JS; de Groot BL; Grubmuller H; Groenhof G, Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge. J. Chem. Theory Comput 2014, 10, 381–390. [PubMed: 26579917]
29. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys 1983, 79, 926–935.
30. Hess B; Bekker H; Berendsen HJC; Fraaije JGEM, Lincs: A Linear Constraint Solver for Molecular Simulations. J. Comput. Chem 1997, 18, 1463–1472.
31. MacKerell AD, Jr., et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B 1998, 102, 3586–3616. [PubMed: 24889800]
32. Joung IS; Cheatham TE, III, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. J. Phys. Chem. B 2008, 112, 9020–9041. [PubMed: 18593145]

33. Marcus Y, Thermodynamics of Solvation of Ions. Part 5.—Gibbs Free Energy of Hydration at 298.15 K. J. Chem. Soc. Faraday Trans 1991, 87, 2995–2999.

34. Marcus Y, The Thermodynamics of Solvation of Ions. Part 4.—Application of the Tetraphenylarsonium Tetraphenylborate (Tatb) Extrathermodynamic Assumption to the Hydration of Ions and to Properties of Hydrated Ions. J. Chem. Soc. Faraday Trans 1987, 87, 2995–2999.

35. Warren CL; Patel S, Hydration Free Energies of Monovalent Ions in Transferable Intermolecular Potential Four Point Fluctuating Charge Water: An Assessment of Simulation Methodology and Force Field Performance and Transferability. J. Chem. Phys 2007, 127, 064509. [PubMed: 17705614]

36. Raineri FO; Stell G; Ben-Amotz D, New Mean-Energy Formulae for Free Energy Differences. Mol. Phys 2005, 103, 3209–3221.

37. Sun DL; Zhou J, Effect of Water Content on Microstructures and Oxygen Permeation in Psima–Ipn–Pmpc Hydrogel: A Molecular Simulation Study. Chem. Eng. Sci 2012, 78, 236–245.

38. Sun DL; Zhou J, Molecular Simulation of Oxygen Sorption and Diffusion in the Poly (Lactic Acid). Chin. J. Chem. Eng 2013, 21, 301–309.

39. Guarnieri F; Mezei M, Simulated Annealing of Chemical Potential: A General Procedure for Locating Bound Waters. Application to the Study of the Differential Hydration Propensities of the Major and Minor Grooves of DNA. J. Am. Chem. Soc 1996, 118, 8493–8494.

40. Ross GA; Rustenburg AS; Grinaway PB; Fass J; Chodera JD, Biomolecular Simulations under Realistic Macroscopic Salt Conditions. J. Phys. Chem. B 2018, 122, 5466–5486. [PubMed: 29649876]

41. Guvench O; MacKerell AD, Jr., Computational Fragment-Based Bnding Site Identification by Ligand Competitive Saturation. PLoS Comput. Biol 2009, 5, e1000435. [PubMed: 19593374]

42. Raman EP; Yu WB; Lakkaraju SK; MacKerell AD, Jr., Inclusion of Multiple Fragment Types in the Site Identification by Ligand Competitive Saturation (Silcs) Approach. J. Chem. Inf. Model 2013, 53, 3384–3398. [PubMed: 24245913]

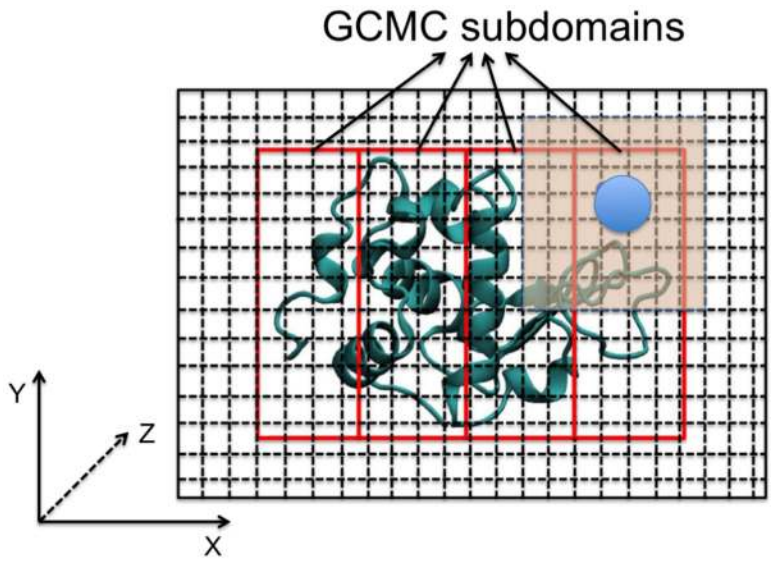


Figure 1.

In the GCMC simulation program, the three-dimensional simulation box is compartmentalized into a large number of small cubes with a length of 1 Å. All atoms in the simulation box are assigned into the cubes based on the atomic coordinates and the cube positions. By constructing a two-dimensional cube-atom array, all neighboring atoms surrounding an ion within a cutoff distance can be quickly found. In addition, a GCMC region can be defined if the simulation system contains biomolecules like protein or DNA. For parallel GCMC simulations, the GCMC region is divided into a number of GCMC subdomains.

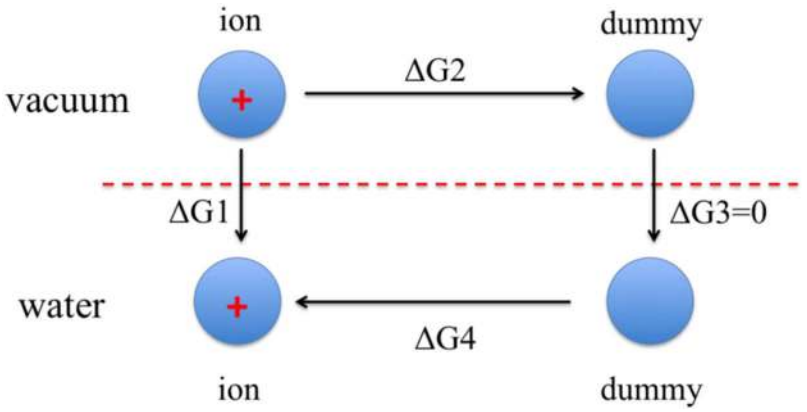


Figure 2.

Thermodynamic cycle for the partitioning of an ion from vacuum to water. The absolute ionic hydration free energy is equivalent to the free energy for transferring an ion from vacuum to water, ΔG1. Since PME and periodicity are used for the calculation, an ion in vacuum can interact with its periodic image, and thus ΔG2 is not zero. ΔG4 includes the contributions from cavity formation (dominated by LJ interactions) and charging (electrostatics). The dummy particle has no interaction with water molecules such that ΔG3=0.

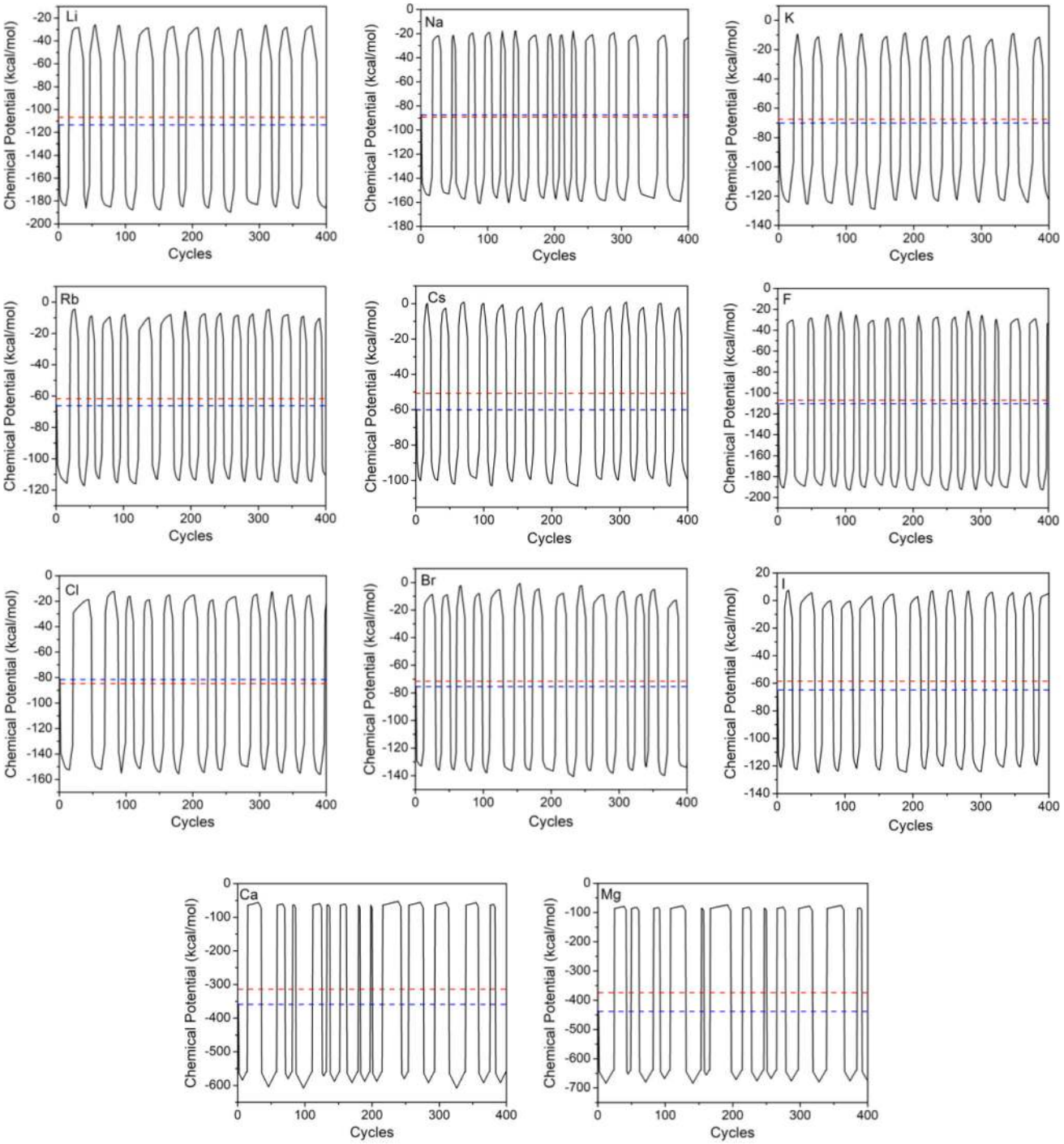


Figure 3.

Fluctuations of the excess chemical potential for the eleven ions during the 400 cycles of GCMC/MD simulations. The blue dashed line denotes the Markus experimental data while the red dashed line represents the predicted ionic HFE from the GCMC/MD simulations. In the GCMC simulation program, PME method is used for electrostatic interaction

calculations and the correction for electrostatic interaction energy due to the finite simulation box size is not considered.

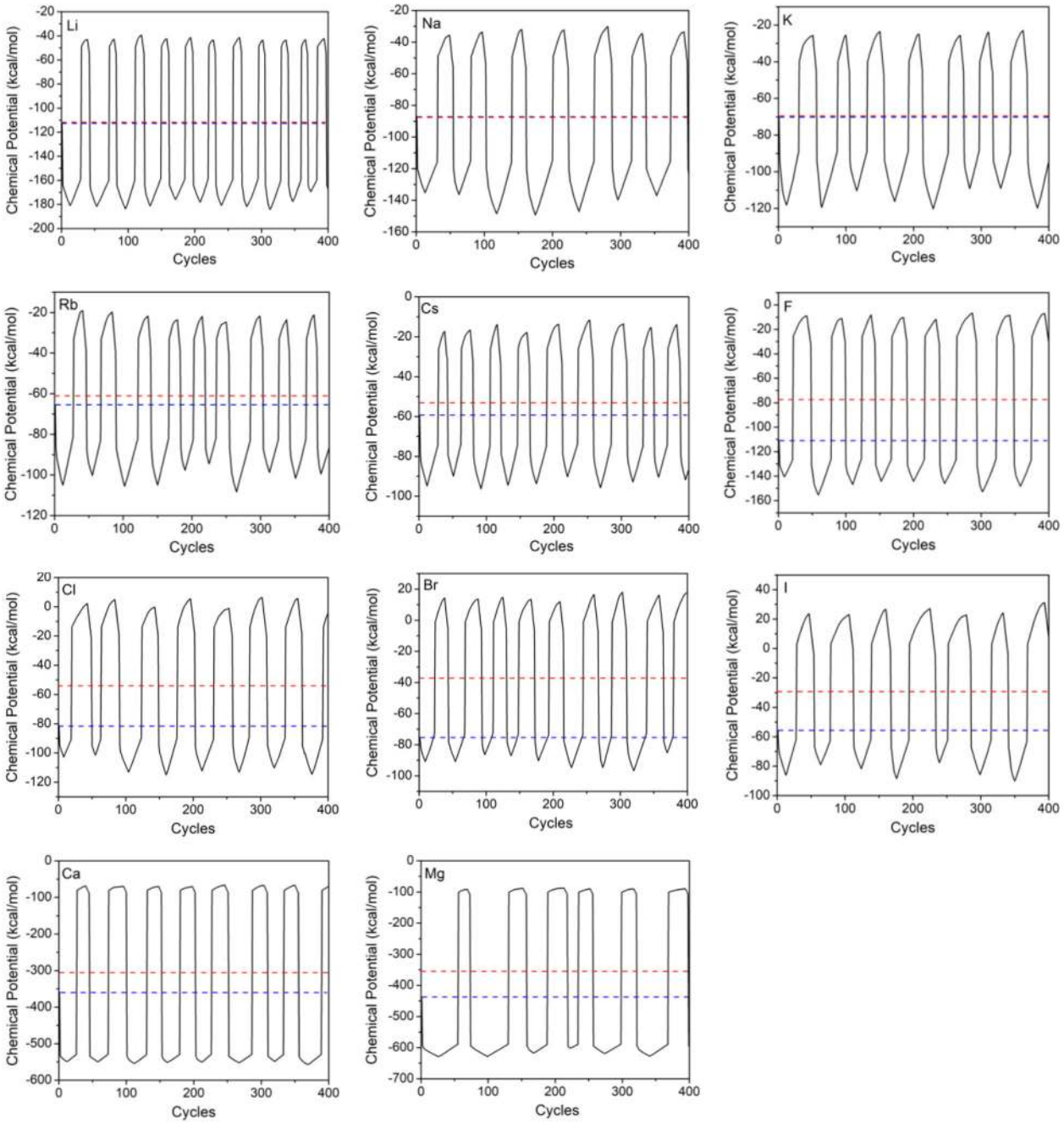


Figure 4.

Fluctuations of the excess chemical potential for the eleven ions. The results are obtained using our previously developed GCMC simulation program, in which a cut-off method is used for electrostatic interaction energy calculations.

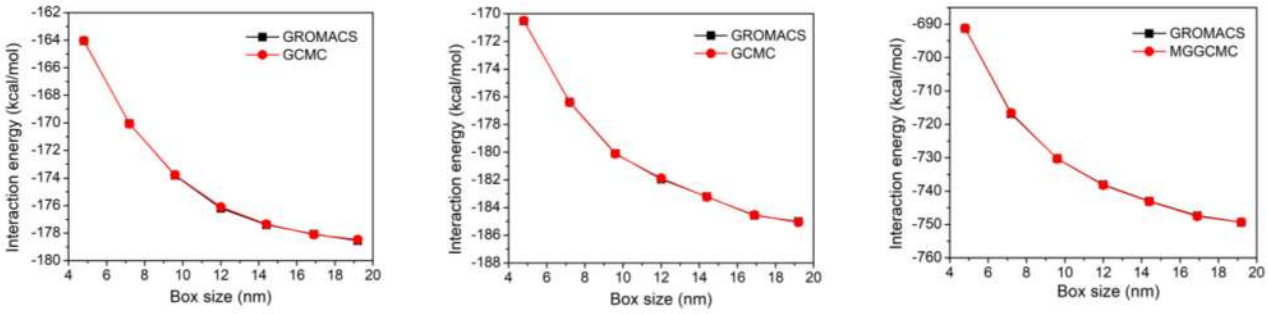


Figure 5.

Ion-water interaction energy for a single (a) Cl^- (b) Na^+ and (c) Mg^{2+}. Periodic boundary conditions and PME are applied.

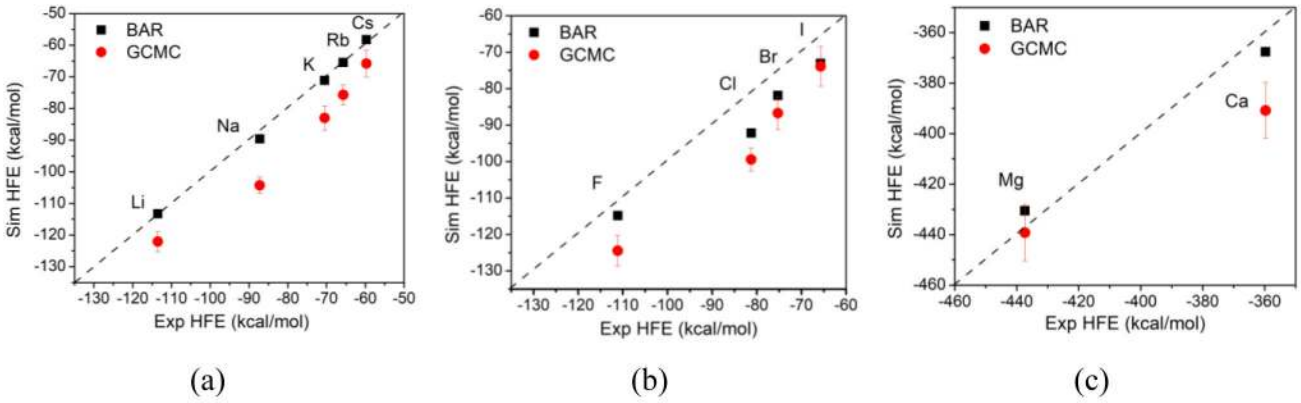


Figure 6.

Comparison of the experimental HFEs with the simulated HFEs from BAR and GCMC/MD approaches.

Table 1.

Force field parameters and the experimental hydration free energy values for the eleven ions.

Ions	LJ parameters		Exp HFE[c]
	$R_{min}/2$ (Å)	ε (kcal/mol)	(kcal/mol)
Li^{+}[a]	1.29591	0.04079	−113.52
Na^{+}[a]	1.40811	0.82103	−87.23
K^{+}[a]	1.76154	1.52302	−70.51
Rb^{+}[b]	1.81203	5.73877	−65.72
Cs^{+}[a]	2.09814	3.32611	−59.75
F^{-}[b]	2.30010	0.05887	−111.13
Cl^{-}[a]	2.26644	2.62588	−81.26
Br^{-}[a]	2.60865	1.02679	−75.28
I^{-}[b]	2.85549	0.93972	−65.72
Ca^{2+}[a]	1.36323	2.10070	−359.70
Mg^{2+}[a]	1.18372	0.26259	−437.38

[a] Parameters are from the CHARMM force field as reported in the 2017 version of the CHARMM36 GROMACS toppar files following conversion to $R_{min}/2$ in Å and epsilon in kcal/mol.

[b] Parameters are from ref.

[c] Experimental HFE values are from ref.

Table 2.

Summary of results from the FEP/BAR HFE calculations. Presented is the calculated free energy for annihilating an ion in vacuum, ΔG2; the free energy for growing a dummy particle into a fully charged ion in water, ΔG4; the Wigner energy correction for finite box size; the linear energy correction for finite box size (ΔG_{corr}); and the ionic HFE, $\Delta G1 = \Delta G2 + \Delta G4 - \Delta G_{corr}$. The unit is kcal/mol.

	ΔG2		ΔG4		Wigner correction		ΔG_{corr}		BAR HFE (ΔG1)	
Ions	**L=2.4**	**L=4.0**	**L=2.4**	**L=4.0**	**L=2.4**	**L=4.0**	**L=2.4**	**L=4.0**	**L=2.4**	**L=4.0**
Li^+	19.66	11.80	−113.24	−113.29	19.20	12.00	19.63	11.78	−113.21	−113.27
Na^+	19.66	11.80	−89.62	−89.62	19.20	12.00	19.63	11.78	−89.59	−89.60
K^+	19.66	11.80	−71.32	−71.16	19.20	12.00	19.63	11.78	−71.29	−71.14
Rb^+	19.66	11.80	−65.97	−65.49	19.20	12.00	19.63	11.78	−65.94	−65.47
Cs^+	19.66	11.80	−58.40	−58.26	19.20	12.00	19.63	11.78	−58.37	−58.24
F^-	19.66	11.80	−114.93	−114.83	19.20	12.00	19.63	11.78	−114.90	−114.81
Cl^-	19.66	11.80	−92.72	−92.24	19.20	12.00	19.63	11.78	−92.69	−92.22
Br^-	19.66	11.80	−82.15	−81.89	19.20	12.00	19.63	11.78	−82.12	−81.87
I^-	19.66	11.80	−73.47	−73.08	19.20	12.00	19.63	11.78	−73.44	−73.06
Ca^{2+}	78.63	47.21	−369.32	−367.69	76.80	48.00	78.53	47.12	−369.22	−367.60
Mg^{2+}	78.63	47.21	−432.21	−430.64	76.80	48.00	78.53	47.12	−432.11	−430.55

J Chem Theory Comput. Author manuscript; available in PMC 2019 October 09.

Table 3.

The average μ_{ex}^{min}, μ_{ex}^{max} and HFE values predicted using PME GCMC/MD and cut-off GCMC/MD simulations. The unit is in kcal/mol. The uncertainties are standard deviations.

Ion	PME			Cutoff		
	μ_{ex}^{min}	μ_{ex}^{max}	HFE	μ_{ex}^{min}	μ_{ex}^{max}	HFE
Li^+	−185.91±1.75	−27.61±1.08	−106.76±2.83	−179.58±3.98	−42.44±1.20	−111.01±5.18
Na^+	−157.64±2.06	−21.72±1.43	−88.86±3.49	−141.92±5.70	−33.12±1.67	−87.52±7.37
K^+	−124.33±1.84	−10.31±1.26	−67.32±3.10	−114.64±4.61	−24.57±1.05	−69.61±5.66
Rb^+	−114.58±1.25	−7.81±1.67	−61.20±2.92	−101.86±4.20	−21.91±1.73	−61.89±5.93
Cs^+	−100.72±1.55	−0.93±1.23	−50.83±2.78	−93.16±2.29	−14.82±1.94	−53.99±4.23
F^-	−190.53±1.63	−27.15±2.60	−108.84±4.23	−147.31±4.48	−9.03±1.74	−78.17±5.65
Cl^-	−153.32±1.64	−15.78±2.15	−84.55±3.79	−109.74±4.98	3.32±2.79	−53.21±7.77
Br^-	−136.67±1.97	−6.44±3.18	−71.55±5.15	−90.71±3.78	15.07±1.94	−37.81±5.72
I^-	−122.29±1.62	4.74±2.47	−58.77±4.09	−84.09±4.41	25.48±2.83	−29.30±7.24
Ca^{2+}	−590.46±10.47	−59.35±3.60	−324.90±14.07	−551.35±2.96	−68.68±2.52	−310.01±5.48
Mg^{2+}	−672.67±9.78	−80.46±3.71	−376.56±13.49	−611.12±3.01	−94.04±2.11	−352.58±5.12

Table 4.

The simulated and the predicted electrostatic energies of ions interacting with water. Unit is in kcal/mol.

Ions	L=4.8 nm	L=19.2 nm	
	Sim U_{elec}	Sim U_{elec}	Pre U_{elec}
Li^+	−210.93	−229.12	−229.12
Na^+	−180.03	−196.27	−194.74
K^+	−145.14	−158.45	−159.86
Rb^+	−128.85	−141.99	−143.56
Cs^+	−115.22	−130.90	−129.94
F^-	−218.38	−233.12	−233.10
Cl^-	−172.63	−184.38	−187.34
Br^-	−154.40	−170.35	−169.12
I^-	−139.13	−153.33	−153.84
Ca^{2+}	−642.86	−699.28	−701.80
Mg^{2+}	−724.54	−776.62	−783.48

Table 5

The averaged μ_{ex}^{min}, μ_{ex}^{max} and HFE values predicted using PME GCMC simulations with finite box corrections. The unit is in kcal/mol. The finite box correction contribution is the difference between the HFEs calculated from PME GCMC simulations with and without finite box correction.

Ions	Corrected PME			Finite box
	μ_{ex}^{min}	μ_{ex}^{max}	HFE	correction
Li^+	−200.48±1.69	−43.61±1.53	−122.05±3.22	−15.29
Na^+	−173.76±1.69	−34.75±0.93	−104.25±2.62	−15.39
K^+	−139.72±1.77	−26.24±1.98	−82.98±3.75	−15.66
Rb^+	−129.05±1.65	−22.33±1.46	−75.69±3.11	−14.49
Cs^+	−115.60±2.02	−16.03±2.15	−65.81±4.17	−14.98
F^-	−205.15±2.54	−43.78±1.64	−124.47±4.18	−15.63
Cl^-	−167.77±1.38	−31.20±1.77	−99.49±3.15	−14.94
Br^-	−151.53±1.47	−21.95±2.93	−86.74±4.40	−15.19
I^-	−137.48±2.34	−10.31±3.09	−73.90±5.43	−15.13
Ca^{2+}	−656.00±7.30	−125.67±3.74	−390.83±11.04	−65.93
Mg^{2+}	−735.14±8.74	−143.28±2.45	−439.21±11.19	−62.65