Supporting Information for

Determination of Ionic Hydration Free Energy 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

Method to determine the dμ_{ex} values for ions

The GCMC/MD simulations are initiated by assigning the excess chemical potential (μ_{ex}) of the ion to its experimentally measured hydration free energy (HFE), though the final results are insensitive to the initial value of the HFE (see below, Figure S3).[1] The difference between the ionic HFE and the averaged ion-water interaction energy is tens of kcal/mol or more, consistent with linear response theorem.[2-3] Thus, the probability of removing or inserting the ions in a condensed-phase system is extremely low if the μ_{ex} of ions are kept constant at the experimental HFE. This motivated the development of the oscillating μ_{ex} GCMC approach in which μ_{ex} is varied by an amount dμ_{ex} according to the number of solutes or ions in the system relative to a target value.[4] Accordingly, to remove ions from the system, the μ_{ex} is stepwise lowered by an amount of dμ_{ex} (e.g., making the μ_{ex} more favorable or more negative). In this so-called "ion deletion" process, the dμ_{ex} value in the present study is determined by the following equation

$$\mathrm{d}\mu_{ex} = \begin{cases} -4\ kcal/mol, \left(if\ N_i \geq\ N_{target}\right) \\ \frac{-4N_i}{\left(2\times N_{target}\right)} kcal/mol, \left(if\ N_i < N_{target}\right) \end{cases}, \quad \text{(S1)}$$

where N_i is the number of ions in the simulation box, N_{target} is the targeted number of ions, which is set at the start of the GCMC/MD simulation. In this work, N_{target}=11 corresponds to an ionic concentration of ~0.15 M. Based on equation (S1), if N_i is larger than or equal to N_{target}, a dμ_{ex} value of -4 kcal/mol is used. If N_i is smaller than N_{target} , a small dμ_{ex} value is used to gradually and smoothly change μ_{ex} while continuing to remove ions from the system. The ion-deletion process continues until the number of ions in the system becomes zero following which the ion-insertion process is initiated.

Inserting ions into the system is performed by increasing the ionic μ_{ex} value (e.g., making the μ_{ex} less favorable or more positive). Once the ion-insertion process has been initiated, the dμ_{ex} value is determined by

$$\mathrm{d}\mu_{ex} = \begin{cases} 4\ kcal/mol, (if\ dN_i = 0) \\ \frac{4(N_{target} - N_i)}{\left(5\times N_{target}\right)} kcal/mol, \left(if\ 0 < N_i < N_{target}\right) \end{cases} \quad \text{(S2)}$$

where if no ion was inserted into the system in the previous cycle of GCMC simulation ($dN_i = 0$), a dμ_{ex} value of 4 kcal/mol is used. If an ion was inserted in the previous GCMC cycle, a smaller dμ_{ex} value is used to more gradually increase dμ_{ex} while continuing to insert ions into the system. Ions will continue to be inserted into the system until the number of ion becomes larger than or equal to N_{target}. At this stage, the ion-deletion process is initiated again.

However, when the above protocol alone is applied, the variation between the target number of ions and zero ions requires a large number of GCMC/MD cycles to start inserting or deleting ions, making the method inefficient. Fig. S1 shows the evolution of μ_{ex} of Na^+ in aqueous solution using the above protocol. It can be seen from Fig. S1 that a significant number of GCMC/MD simulation cycles occur in the range of -130 kcal/mol and -50 kcal/mol for μ_{ex}. While in this range, the probability of Na^+ ions being either inserted into or deleted from the solutions during the GCMC simulations is extremely small consistent with the Metropolis-Hasting criteria. To avoid spending a large number of cycles in which minimal insertions or deletions occur, we simply bypass the -130 kcal/mol to -50 kcal/mol μ_{ex} region (as indicated by the two solid lines for the two boundary μ_{ex} values). This is performed by setting μ_{ex} to -130 or -50 kcal/mol when μ_{ex} reaches the opposing values based on the $d\mu_{ex}$ values used according to equations S1 and S2. The "jump" of μ_{ex} is applied in a single cycle following which the variation of $d\mu_{ex}$ between cycles reverts to that in equations S1 or S2 for the ion-deletion and ion-insertion phases, respectively. Application of this method significantly decreases the number of GCMC/MD cycles required to go from zero to N_{target} and vice versa. Different boundary μ_{ex} values are set for the different ions, with the boundary μ_{ex} values determined empirically and listed in Table S1.

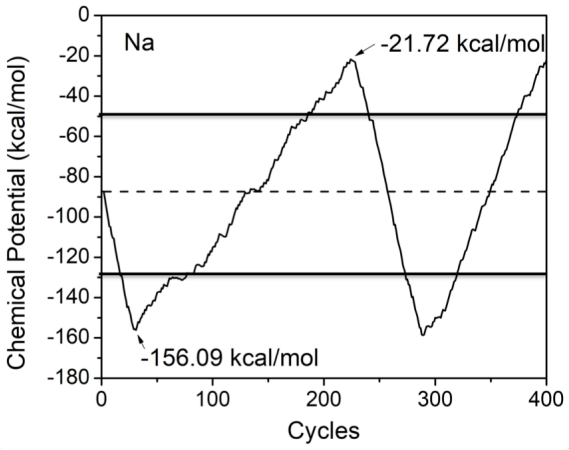


Figure S1. The fluctuation of the excess chemical potential of Na^+.

Table S1 The lower (L- μ_{ex}) and upper (U- μ_{ex}) boundary μ_{ex} values used for PME GCMC simulations. Units are in kcal/mol.

Ions	L- μ_{ex}	U- μ_{ex}
Li^+	-170	-55
Na^+	-140	-30
K^+	-100	-30
Rb^+	-100	-22
Cs^+	-85	-17
F^-	-175	-38
Cl^-	-135	-33
Br^-	-125	-23
I^-	-108	-10
Ca^{2+}	-560	-68
Mg^{2+}	-640	-80

Subsequent GCMC/MD simulations were performed using the μ_{ex} offset approach combined with application of dμ_{ex} based on equations S1 and S2. Analysis of the number of ions in the simulation boxes for the different ions is shown in Fig. S2. As may be seen, the value of N_{target} is over shot during many of the ion insertion stages of the GCMC/MD simulations. Similarly, once N_i reaches zero, its value remains at zero for multiple cycles. This is an outcome of the gradual variation of μ_{ex} between GCMC/MD cycles as prescribed by equations S1 and S2, such that multiple GCMC/MD cycles are required until the μ_{ex} boundary values listed in Table S1 are reached and the large jump of μ_{ex} undertaken.

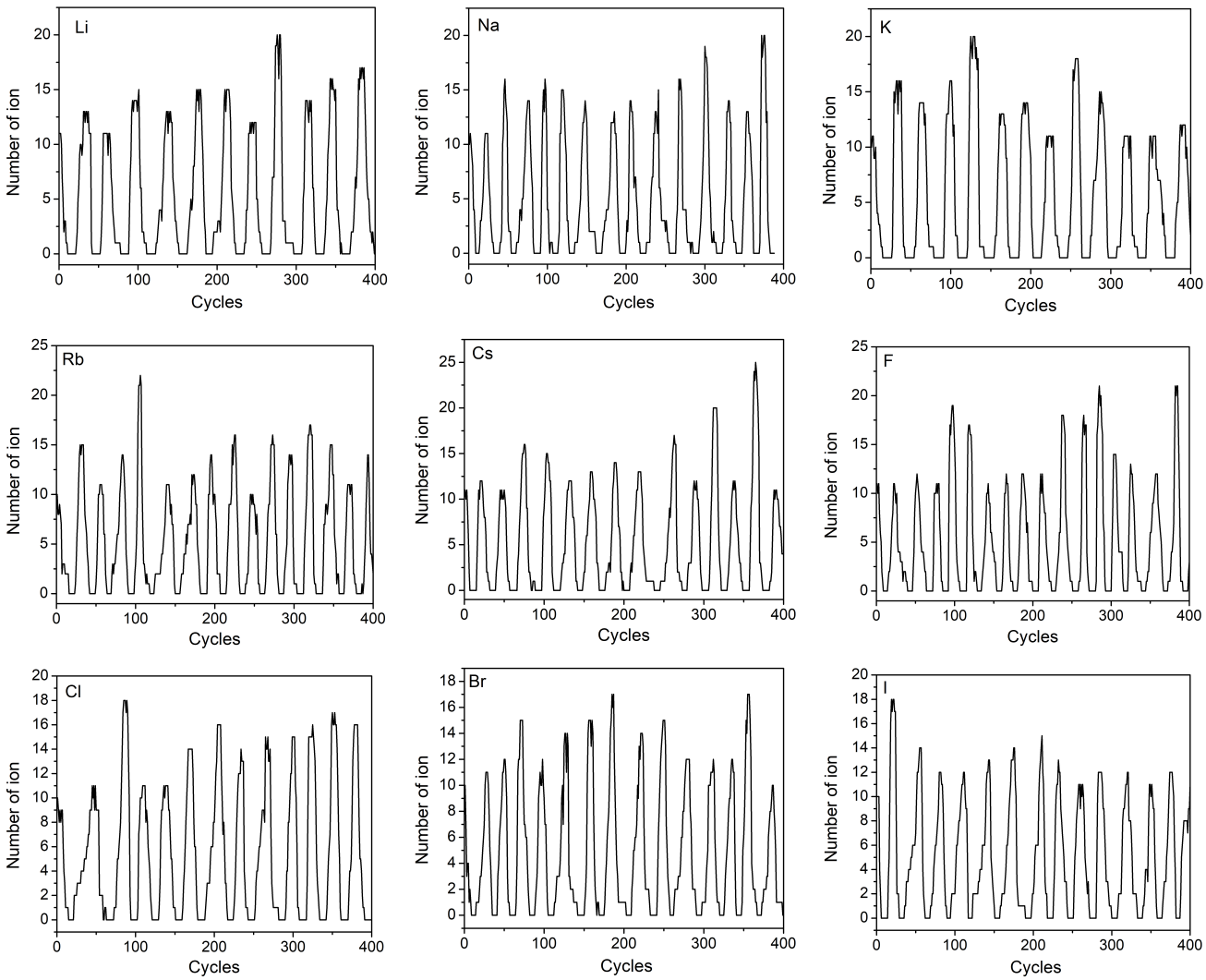


Figure S2: Fluctuations of the number of ions in the simulation box during the 400 cycles of GCMC/MD simulations.

Additional calculations were undertaken to determine the impact of the initial assignment of the value of the HFE and on the number of ions in the system on the final calculated HFE values. Shown in Figure S3 are results for Cl^- and Na^+ with different initial HFE values, with the average HFE values from the GCMC/MD simulations included as lines in the individual plots. As may be seen significant changes in the initial HFE values yield nearly identical average HFE values from the GCMC/MD simulations. Similar results were obtained when the number of target ions in the system

is varied as shown in Figure S4. With N_{target} set to 1, 3 and 7 near identical final HFE values are obtained, which are similar to those obtained with 11 ions as shown in Figure S3 and in the main manuscript.

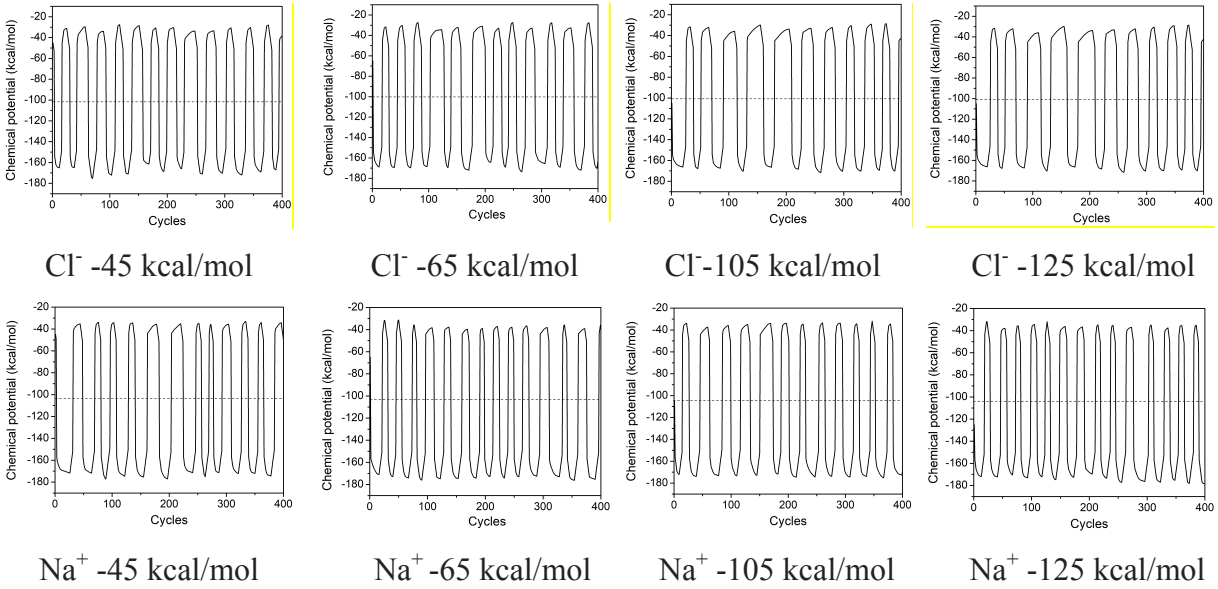


Figure S3: Effect of initial value assigned to the chemical potential on the GCMC/MD simulation hydration free energy results for Cl^- and Na^+. The average HFE value calculated over the simulations are indicated by the horizontal line in the plots.

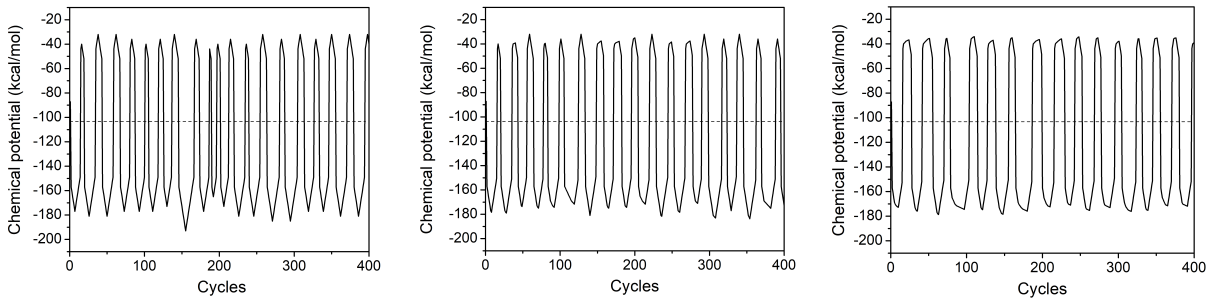
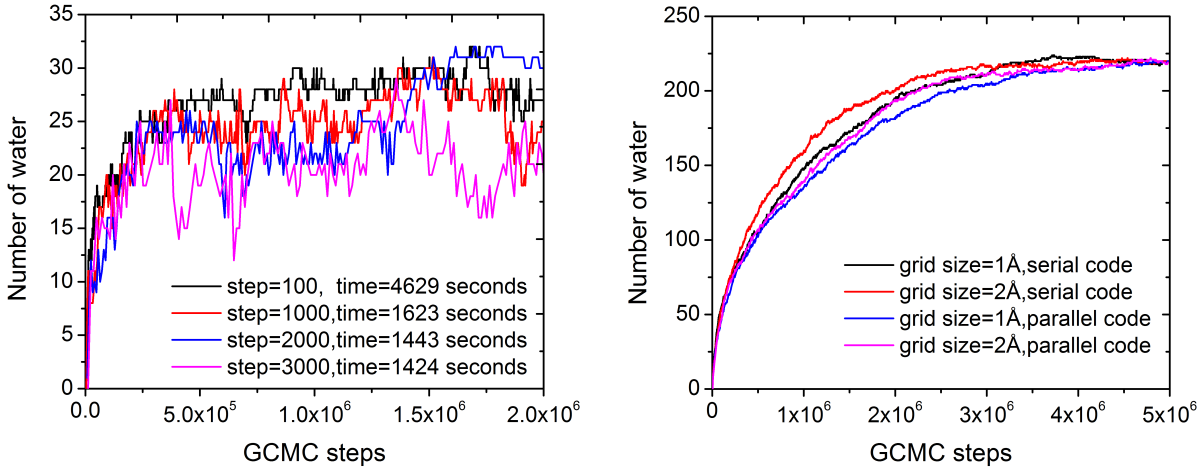


Figure S4. Effect of number of ions on the GCMC/MD simulation hydration free energy results. The initial chemical potential assigned in the simulations was -87 kcal/mol. The average HFE value calculated over the simulations are indicated by the horizontal line in the plots

Validation of the parallelization scheme of the GCMC simulation program

In the parallel GCMC simulation method, the GCMC subdomains run the simulations independently, such that the ions or molecules inside one subdomain cannot sense the movement of those in neighboring subdomains. This parallelization scheme is valid only when the neighboring subdomains do not cause significant perturbations of the local subdomain liquid structure. To avoid problematic perturbations of the liquid structure, the simulation system (*i.e.*, the PME Q array and the cube-atom coupling array) has to be updated every N GCMC steps. To determine an optimum value for N, a void cubic region with length 10 Å in the center of a water box was created, and then parallel GCMC simulations were run with μ_{ex} for water set to -5.6 kcal/mol. Fig. S5 (a) plots the number of water molecules inserted into the cubic region when N was set to 100, 1000, 2000 and 3000. It can be seen that when N was set to 100, a converged number of ~28 water molecules can be obtained requiring 4,629 seconds to run the 2,000,000 GCMC steps. When N was set to 1000, a satisfactorily converged number of ~26 was obtained with the computational cost is reduced by ~3 fold (1,623 seconds) compared to N=100. When N was further increased to 2000 and 3000, the computational costs were not significantly reduced compared to N=1000, although the average number of water molecules was far from converged within the 2,000,000 total simulation steps. Based on this analysis, N was set to 1000 in the GCMC simulations unless stated otherwise.

To further test the effects of the parallelization scheme and the cubic grid size used for the short-range interaction grid search method on the simulation result, we created a larger void cubic region with length of 20 Å in the water box and the GCMC simulations were run in serial and in parallel, respectively. Fig.S5 (b) shows that, for both tested grid sizes of 1 Å and 2 Å, the serial simulations give slightly faster converged results compared to the parallel simulations. Nevertheless, both serial and parallel simulations predict almost identical numbers of inserted water molecules at the end of 5,000,000 GCMC steps irrespective of the grid size. These benchmark results indicate that the proposed GCMC simulation parallelization scheme can significantly speed up the calculations without markedly affecting the simulation results.



(a) (b)

Figure S5: (a) Effects of *N*, the number of GCMC steps between the PME Q array and the cube-atom coupling array updates on the number of water molecules inside a cubic region of length 10 Å and the total simulation time. (b) Effect of cubic grid size on the number of water molecules inside a cubic region of length 20 Å from serial and parallel simulations. AMD FX-6350 Six-Core processor was used for simulation parallelization.

1. 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.

2. Åqvist, J.; Hansson, T., On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **1996**, *100*, 9512-9521.

3. Roux, B.; Yu, H. A.; Karplus, M., Molecular Basis for the Born Model of Ion Solvation. *J. Phys. Chem.* **1990**, *94*, 4683-4688.

4. Lakkaraju, S. K.; Raman, E. P.; Yu, W. B.; MacKerell, A. D., 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.

5. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., Gromacs: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701-1718.