JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

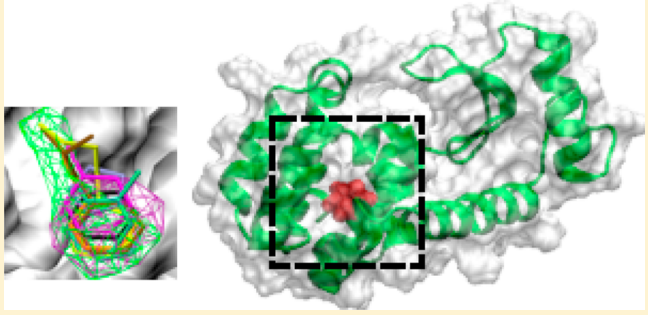
Sampling of Organic Solutes in Aqueous and Heterogeneous Environments Using Oscillating Excess Chemical Potentials in Grand Canonical-like Monte Carlo-Molecular Dynamics Simulations

Sirish Kaushik Lakkaraju, E. Prabhu Raman, Wenbo Yu, and Alexander D. MacKerell, Jr.*

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

Supporting Information

ABSTRACT: Solute sampling of explicit bulk-phase aqueous environments in grand canonical (GC) ensemble simulations suffer from poor convergence due to low insertion probabilities of the solutes. To address this, we developed an iterative procedure involving Grand Canonical-like Monte Carlo (GCMC) and molecular dynamics (MD) simulations. Each iteration involves GCMC of both the solutes and water followed by MD, with the excess chemical potential (μ_{ex}) of both the solute and the water oscillated to attain their target concentrations in the simulation system. By periodically varying the μ_{ex} of the water and solutes over the GCMC-MD iterations, solute exchange probabilities and the spatial distributions of the solutes improved. The utility of the oscillating-μ_{ex} GCMC-MD method is indicated by its ability to approximate the hydration free energy (HFE) of the individual solutes in aqueous solution as well as in dilute aqueous mixtures of multiple solutes. For seven organic solutes: benzene, propane, acetaldehyde, methanol, formamide, acetate, and methylammonium, the average μ_{ex} of the solutes and the water converged close to their respective HFEs in both 1 M standard state and dilute aqueous mixture systems. The oscillating-μ_{ex} GCMC methodology is also able to drive solute sampling in proteins in aqueous environments as shown using the occluded binding pocket of the T4 lysozyme L99A mutant as a model system. The approach was shown to satisfactorily reproduce the free energy of binding of benzene as well as sample the functional group requirements of the occluded pocket consistent with the crystal structures of known ligands bound to the L99A mutant as well as their relative binding affinities.



INTRODUCTION

Chemical potential (μ) describes the equilibrium movement of particles between two phases or states. The driving force behind this movement comes from (1) a concentration gradient—particles tend to move from high to low concentration regions to gain mixing entropy; (2) chemical affinity—particles are attracted to regions of high chemical affinity.[1] Excess chemical potential (μ_{ex}) is the quasistatic work to bring a particle (e.g., solute molecule) from the gas phase to the solvent; $\mu_{ex} = \mu - \mu_{id}$, where μ and μ_{id} are the chemical potential and the ideal gas chemical potential of the solute, respectively.[2] In the context of statistical mechanics, chemical potential allows for the thermodynamic state of a system to be defined in terms of a grand canonical (GC) ensemble (μVT) that allows for variation in the species concentrations across phases/states. Simulation procedures have long evolved toward efficiently determining Gibbs' free energy of hydration (HFE), chemical affinity, and other thermodynamically relevant properties of waters and other small solute molecules from GC ensembles[3−10] instead of the more conventional isothermal, isobaric (*NPT*), canonical (*NVT*), or microcanonical (*NVE*) ensembles where the concentration of the species is fixed. To date, many of the GC ensemble strategies have employed Monte Carlo (MC) simulations toward either driving the sampling of waters or individual small molecules around proteins[3,9−11] or crystal environments,[4] or alternately improving the accuracy of the relative HFE calculations in free energy perturbation (FEP) calculations.[5] However, since GCMC simulations of systems containing explicit solvent to represent the bulk-phase suffer from convergence problems due to low acceptance rates encountered for the solute insertions[12,13] simulations in the past were restricted toward investigation of the chemical affinities of only the solvent,[3,10,14] drive individual solute sampling in the absence of explicit solvent,[9,11] determine thermodynamic properties in crystal conditions,[4] or use expanded ensemble strategies.[6]

In the context of protein and macromolecular environments, chemical fragment sampling simulation techniques have been employed in the past toward discovery or rational design of molecules that can bind to macromolecular targets with high affinity to achieve a desired biological outcome.[15−20] The Site

Received: December 30, 2013
Published: May 6, 2014

Identification by Ligand Competitive Saturation (SILCS) method is one such technique that identifies the location and approximate affinities of different functional groups on a target macromolecular surface by performing Molecular Dynamics (MD) simulations of the target in an aqueous solution of solute molecules representative of different chemical fragments.[17,18,21] However, these MD isothermal, isobaric (*NPT*) ensembles suffer from the long diffusion time scales of the solutes through explicit solvent and macromolecule environments, especially when the macromolecular binding sites are deeply buried and inaccessible to the solvent (i.e., occluded). Sampling of the distribution of chemically diverse solutes in proteins in GC ensembles with MC simulations offers the potential to overcome these limitations, yielding more accurate solute spatial distributions in aqueous macromolecular environments.

In this paper, we present a methodology for solute sampling in explicit solvent aqueous systems and solvated protein environments using a variation of GCMC in which the μ_{ex} values are adjusted to attain a target concentration along with MD simulations in an iterative fashion (oscillating-μ_{ex} GCMC-MD). In this approach GCMC sampling is performed on both the solutes and the waters. A short time-scale MD after the GCMC allows for both conformational sampling of the solutes and configurational sampling of the aqueous system and the macromolecule. To achieve satisfactory convergence, the process of GCMC-MD is repeated through multiple iterations, with the μ_{ex} of the species being studied systematically oscillated over the iterations to drive the solute and water exchanges. Two types of aqueous systems were considered: (1) a system containing 1 M of only one type of solute in water, thereby replicating the standard state of the solute, and (2) a dilute aqueous mixture containing 0.25 M of many types of solutes. Both systems are implemented in aqueous solution alone and in the presence of the L99A mutant of T4-lysozyme (T4-L99A). GCMC was run on the water and all the different solutes in the system with the variation of their respective μ_{ex} values through the GCMC-MD iterations used to improve the solute exchange probabilities while allowing for determination of the respective μ_{ex} values required to maintain a defined concentration of the solutes and waters in the aqueous systems. As the oscillating-μ_{ex} GCMC-MD approach involves variation of μ_{ex} such that it is not formally the GC ensemble, we show that it produces a representative ensemble of conformations as the average μ_{ex} values approximate the HFE of solutes at a specified target concentration in aqueous systems. In the presence of the protein, this oscillating-μ_{ex} GCMC-MD strategy could be used for efficiently sampling the spatial distribution of the solutes in and around the protein. Multiple solute sampling in the context of SILCS allows for determination of their affinity patterns to the protein that can be used toward rational drug design.

METHOD AND SIMULATIONS DETAILS

The theory of the GCMC has been well described in several references.[9,12,22] There are four possible GCMC moves on a molecule, M (i.e., solute or water): insertion—brings M into the system A from the reservoir; deletion—removes M from the system A and moves it back into the reservoir; translation and rotation—M is translated/rotated within a subvolume surrounding the original location of M in system A. The probabilities of these moves as governed by the Metropolis criteria are

$$P_{\text{insert}} = \min\left\{1, \frac{f_n}{n+1} e^{B-\beta\Delta E}\right\}$$
$$P_{\text{delete}} = \min\left\{1, \frac{n}{f_{n-1}} e^{-B-\beta\Delta E}\right\}$$
$$P_{\text{trans/rot}} = \min\{1, e^{-\beta\Delta E}\} \quad (1)$$

where $B = \beta\mu_{ex} + \ln \overline{n}$, $\overline{n} = \overline{\rho v}$, μ_{ex} is the excess chemical potential, $\overline{n}$ is the expected number of molecules, $\overline{\rho}$ is the density, $\overline{v}$ is the volume of system A, f_n is the fractional volume of the subspace where the insertion attempts are made, ΔE is the change in energy due to a move, β is $1/k_BT$, k_B is the Boltzmann constant, and T is temperature (300 K in the present study). Through the GCMC simulation, the volume of the simulation system A, the total energy, and the total number of particles between the system A and its reservoir are fixed.[10]

As seen in eq 1, the target concentration of each solute ($\overline{n}$), the interactions of each solute with all the other molecules in the simulation system A (ΔE), and the supplied μ_{ex} determines the solute populations through GCMC simulations. Since, the move probabilities of the individual solutes or waters are driven by their $\overline{n}$ and μ_{ex} values, when the μ_{ex} of a solute is less than the work needed to move a molecule from the gas-phase reservoir to the system A, a decrease in concentration of the solute from the system A would occur. Likewise, when the μ_{ex} supplied is more than the needed work, an increase in solute concentration will occur in system A. Consequently, the value of μ_{ex} supplied to the GCMC simulation may be oscillated based on the concentration in the simulation system and the target $\overline{n}$.

With $\overline{n}$ used as a target for the oscillating-μ_{ex} GCMC calculation, the simulation system A is defined to contain water at a bulk-phase concentration of 55 M. Through the iterative oscillating-μ_{ex} GCMC-MD simulations, the concentration of the solutes and the water in system A vs their target concentration ($\overline{n}$) is used as a guide to vary their respective μ_{ex} through each subsequent GCMC iteration. The simulations start with the μ_{ex} of the solutes and the water set to 0 in the first iteration. Over every subsequent iteration or subset of iterations, μ_{ex} of the respective solutes and water is varied by a magnitude that is governed by the deviation of the solutes and water in the system A from their respective target $\overline{n}$. As the concentrations of the solutes and the waters reach their target, the width of the variation of each μ_{ex} is decreased and this defines the onset of convergence. Thus, in standard state simulations with only one type of solute at 1 M in water, the approximate HFE associated with those conditions can be calculated via the average μ_{ex}. Upon convergence, the oscillation of μ_{ex} values for waters and the solutes is continued to obtain the average μ_{ex} and to facilitate the sampling under equilibrium conditions.

The standard state simulations with only one type of solute at 1 M in system A will be referred to as Scheme I and the multiple solute aqueous mixture simulations as Scheme II. System A is a spherical region of radius, r_A, into which GCMC moves are performed. As shown in Figure 1, separate reservoirs for the solutes and the waters are coupled to system A. System A is immersed in a larger system, B, which includes additional water. For the bulk aqueous systems, system B is a larger sphere of radius $r_B = r_A + dr$ (in the present study, dr = 5 Å). The larger system B limits edge effects, such as hydrophobic solutes

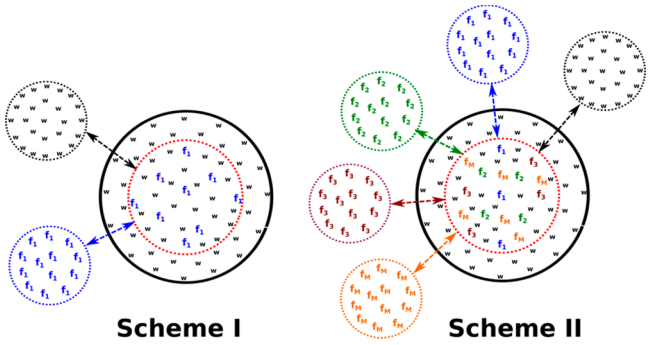


Figure 1. Setup for the standard state (Scheme I) and the aqueous solute mixture (Scheme II) oscillating-μ_{ex} GCMC-MD simulations. Water and the solute molecules are exchanged between their respective reservoirs and the spherical simulation systems, A, indicated by the dashed boundary, defined by radius r_A. System A is immersed in a larger system, B defined by the solid boundary, that included additional waters of radius r_B set to r_A+ 5 Å in the present study to prevent hydrophobic solutes from occupying the edge of the system A (see Figure S1, SI). Alternatively, the larger system in which system A is immersed may be treated using periodic boundary conditions and include other chemical entities in addition to water (see Figure S5, SI).

from occupying the edge of the system A (Supporting Information (SI), Figure S1). Alternatively, system B may be treated using periodic boundary conditions and/or include other chemical entities in addition to water. An example of a periodic system B in which a spherical system A is immersed, with the combined system containing the T4-L99A is shown in Figure S6 of the SI.

The iterative oscillating-μ_{ex} GCMC-MD procedure is described as follows:

(a) Run i steps of GCMC to exchange solutes and waters between their respective reservoirs and the simulation system A. The i steps are divided between each of the M solutes and water. In the present study $i/2$ and $(i)/(M + 1)$ GCMC moves (insertion/deletion/translation/rotation) are attempted for the solutes (F_1, F_1, F_2, ..., F_M) and the waters, in Schemes I and II, respectively, though the number of moves for the different solutes and water are not required to be equal. The order in which the four possible GCMC moves are attempted, and the molecule (solutes or water) on which the move is performed is randomized. In the first iteration, μ_{ex} of the solutes and the water is set to 0. The radii of the water and the solute sphere(s) subjected to GCMC moves are set to r_A through the GCMC process, though the energetic interactions associated with the moves also includes contributions from any waters, or other chemical entities, outside of the GCMC sphere that is defined as system B.

(b) After the GCMC, j steps of MD are run on combined system A and B. For the finite spherical systems, the solutes are retained within the spherical dimensions of system A, r_A through harmonic flat-bottom spherical restraints,[23] while the water and any other molecules in system B encompassing system A are subjected to (i) harmonic flat-bottom spherical restraints with $r_B = r_A + dr$ when system B is spherical or (ii) periodic boundary conditions.

(c) Vary the value of μ_{ex} of solute and water, μ_P (P = F_1, F_2, ..., F_M solutes or W, water) by $d\mu_P$. Use this new value of μ_P in the next iteration of GCMC. The magnitude of $d\mu_P$ is determined by the deviation of the current concentration, N_P (P = F_1, F_2, ..., F_M, or W) in system A from the target concentration (N_P^{target}) .

$$\mu_P = \mu_P + d\mu_P$$

where

$$d\mu_P = 5d\mu_P, \quad \text{when } N_P = 0;$$

$$\text{for cycle 1, } |d\mu_P| = 0.5$$

else

$$d\mu_P = \begin{cases} d\mu_P \dfrac{N_P^{target}}{N_P} & N_P < 0.7N_P^{target} \\ \text{rand}((d\mu_P - 0.5), (d\mu_P + 0.5)) & 0.7N_P^{target} < N_P < 2N_P^{target} \\ -d\mu_P \dfrac{N_P}{N_P^{target}} & N_P > 2N_P^{target} \end{cases} \quad (2)$$

With the variation of $d\mu_P$ system-dependent, various schemes (standard state, aqueous mixture, heterogeneous systems) can be generated using eq 2. We note that different schemes may be applied to vary $d\mu_P$.

(d) Perform k iterations of the oscillating-μ_{ex} GCMC-MD (steps a−c) with new values of μ_{ex} from c).

The GCMC portion of the simulations were run using an in-house C++ code that implemented the grid-based GCMC scheme[10] with the cavity-bias algorithm[13,24] to drive solute and water exchanges between their reservoirs and the aqueous systems, A. The system A is divided into a 1 Å grid lattice. When system A contains macromolecules, all the atoms of the macromolecules are considered to be of the same size, and each atom is assigned to a lattice site. Fragments within the system A are not assigned lattice sites. After deleting the lattice sites assigned to the macromolecule, a list of available lattice sites is maintained. The term f_n in eq 1 is calculated as the ratio between the total number of lattice sites, to the total number of available lattice sites. During a GC move, for example, an insertion of a solute, one of the available lattice sites is randomly chosen. The solute molecule is moved from its gas-phase reservoir such that the center of mass of the solute occupies the randomly selected lattice site. Interaction energy of the solute is then calculated with all of the macromolecule, solutes and water atoms in the system. The updated interaction energy is applied to eq 1 to determine if the move is accepted. If the insertion is accepted, the solute is now a part of the GC grid. When the insertion is rejected, the solute is taken back to the gas-phase reservoir. In general, for every GC move, the interaction energy of the selected molecule with every other solute and solvent molecules and the macromolecule is calculated. ΔE, due to the attempted move is then applied to eq 1 to determine if the move is accepted or rejected.

Solute empirical force-field parameters were obtained from the CGenFF[25] and the TIP3P water model was used.[26] The studied solute molecules were chosen to represent different chemical functionalities including apolar benzene and propane, neutral polar molecules such as acetaldehyde, methanol and formamide, and the negative and positively charged acetate and methylammonium, respectively. These solutes were those used in our recent work on an extension of the SILCS method.[21]

As detailed previously,[17,18] to prevent aggregates of hydrophobic and charged solutes, thereby promoting faster convergence, a repulsive energy term was introduced only

between benzene:benzene, benzene:propane, propane:propane, acetate:acetate, acetate:methylammonium, and methylammonium:methylammonium molecular pairs. This was achieved by adding a massless particle to the center of mass of benzene and the central carbon of propane, acetate, and methylammonium. Each such particle does not interact with any other atoms in the system but with other particles on the hydrophobic or charged molecules through the Lennard-Jones (LJ) force field term[27] with parameters (ε = 0.01 kcal/mol; R_{min} =12.0 Å). All LJ and Coulomb interactions were calculated during GCMC (i.e., no truncation of nonbonded interactions), including interactions with system B. Simulations are initiated with an empty system A. The randomized GCMC process with the solutes and the waters is initially run in multiples of 50 000 moves until the waters in the system reach the bulk 55 M concentration. At each cycle in this repetition of the 50 000 GCMC moves, the μ_{ex} of the solutes and the water is increased by 1 kcal/mol to accelerate the water and solute accumulation in system A. During this process, the concentration of the solutes may also increase beyond their target values. Once the bulk water concentration is attained, this is used as the starting configuration for the oscillating-μ_{ex} GCMC-MD procedure described in steps a−d. After about 50 iterations of the oscillating-μ_{ex} GCMC-MD in which the μ_{ex} values are varied, the concentration of the solutes approach their target values. In each iteration, 50 000 and 100 000 GCMC moves were run for Schemes I and II, respectively.

The GROMACS[28] package (version 5.0) was used for all MD simulations. For the aqueous systems, during each iteration the combined system A and B, including all the solutes and the waters, were simulated with the leapfrog integrator (GROMACS integrator "md"), with a 2 fs time-step, at 300 K through a Nose-Hoover thermostat.[29,30] The last conformation of the GCMC run was used as the starting conformation for the MD. The system was initially equilibrated over a period of 100 ps with reassignment of velocities. This was followed by a 500 ps production run. The data during the equilibration phase is not considered for analysis. The last conformation from the production MD is used as the starting conformation of the next MC run.

The LINCS algorithm[31] was used to constrain water geometries and all covalent bonds involving a hydrogen atom. van der Waals (vdW) and electrostatic interactions were switched off smoothly over the range of 8−10 Å. Solutes and waters were held within the spherical dimensions of system A or B by applying harmonic flat-bottom restraints[23] with a force constant of 1.2 kcal/mol·Å^2 on the following solute or water atoms: the massless particle at the geometric center of benzene, propane, acetate, and methylammonium; the carbon atoms of the acetaldehyde, methanol, and formamide; and the oxygen atom of water. For the bulk aqueous systems, the radius, r_B of the harmonic flat-bottom restraints used to define system B applied to only the water was 5 Å larger than the radius r_A = 20 Å defining the restraint applied to the solutes in system A.

For T4-L99A, PDB coordinates 181L with the benzene ligand was used following the deletion of the ligand. The T4-L99A structure was inserted in boxes replicating Schemes I and II systems with 1 M of benzene and 0.25 M each of the different solutes (benzene, propane, acetaldehyde, methanol, formamide, acetate, and methylammonium), respectively. Akin to the aqueous systems, 10 such boxes were built for Scheme I with 1 M benzene and Scheme II with the multiple solutes, with the solutes randomly inserted in each of the boxes. These systems were minimized for 1000 steps with the steepest descent algorithm[32] in the presence of periodic boundary conditions (PBC).[27] They were then equilibrated for 250 ps by periodic reassignment of velocities. The leapfrog version of the Verlet integrator[27] with a time step of 2 fs was used for heating and equilibration. Long range electrostatic interactions were handled with the particle-mesh Ewald method[33] with a real space cutoff of 12 Å, a switching function[34] was applied to the Lennard-Jones interactions at 12 Å, and a long-range isotropic correction[27] was applied to the pressure for Lennard-Jones interactions beyond the 12 Å cutoff length. During the minimization and equilibration harmonic positional restraints with a force constant of 2.4 kcal/mol·Å^2 were applied to protein non-hydrogen atoms. For the MD in the iterative protocol, the position restraints were removed and replaced by weak restraints only on the backbone C−α carbon atoms with a force constant (k in 1/2 $k\delta x^2$) of 0.12 kcal/mol·Å^2. This was done to prevent the rotation of the protein in the simulation box and potential denaturation due to the presence of a highly concentrated fragment solution.[17]

The oscillating-μ_{ex} GCMC-MD iterations are repeated until convergence, typically 200 iterations for the aqueous systems, yielding 100 ns of MD and 10 and 20 million MC steps for Schemes I and II, respectively. Furthermore, to ensure sufficient sampling and convergence, 10 separate oscillating-μ_{ex} GCMC-MD simulations are run for Scheme II and for each solute in Scheme I resulting in a cumulative 1 μs (10 × 100 ns) of MD for Scheme II and each solute in Scheme I. Runs differ through a randomly generated seed for both GCMC and the leapfrog MD integrator operations in each iteration. Oscillating-μ_{ex} GCMC-MD of the protein systems are repeated over 100 iterations, yielding a cumulative 500 ns of MD over the 10 separate simulations from both Scheme I and Scheme II.

RESULTS AND DISCUSSION

The goal of the present work was the implementation of an algorithm to perform equilibrium studies of solutes in aqueous solution, including mixtures of solutes and in the presence of a macromolecule, with T4-L99A used as the macromolecular model system in the present study. In the context of Scheme I, if the oscillating-μ_{ex} GCMC-MD scheme is producing an approximately correct ensemble, the average μ_{ex} should approximate the HFE corresponding to a 1 M standard state aqueous system and the binding affinity of solutes to the protein. With the solute mixture in Scheme II the approach would allow for determining the μ_{ex} required to drive sampling of the distribution of solute molecules in a heterogeneous aqueous system and determine solute affinity patterns around the protein site. There are other methods available that determine μ_{ex} required to maintain a target concentration of the water molecules using the GCMC methodology,[35] including one that relies on more complex proportional−integral−derivative (PID) controller scheme, and continuously fluctuate the μ_{ex} at every step of the GCMC. In comparison, the modulation of the excess chemical potential in the present work is driven solely by the concentrations of the solutes and the water and their target concentrations or chemical potentials.

In the initial calculations, when the solute and solvent in the finite spherical systems shared the same spherical boundary, the nonpolar molecules sampled the surface of the spherical system (SI, Figure S1). This is due to the favorable interactions of the nonpolar solutes with the nonpolar, vacuum environment outside of the spherical simulation system. To overcome this,

Table 1. Average μ_{ex} of the Solute and the Water through the Scheme I and II GCMC-MD Simulations with a Spherical System of Radius 25 Å, Obtained from Cycles 150−200

			Scheme I		Scheme II	
fragment	HFE^{exp} (kcal/mol)[a]	HFE^{fep} (kcal/mol)[b]	μ_{ex} (kcal/mol)	conc (M)	μ_{ex} kcal/mol)	conc (M)
benzene	−0.83	−0.71	−0.77 ± 0.28	1.40 ± 0.11	−0.92 ± 0.13	0.32 ± 0.19
propane	1.96	1.60	1.75 ± 0.11	1.37 ± 0.26	1.64 ± 0.45	0.36 ± 0.11
acetaldehyde	−3.5	−4.43	−3.35 ± 0.17	1.01 ± 0.11	−2.86 ± 0.1	0.24 ± 0.12
methanol	−5.1	−6.16	−5.77 ± 0.27	1.3 ± 0.49	−4.92 ± 0.11	0.22 ± 0.08
formamide	−14	−10.71	−13.7 ± 0.48	1.11 ± 0.14	−11.15 ± 0.44	0.20 ± 0.09
acetate	−79.1	−96.5	−45.4 ± 0.84	0.82 ± 0.21	−52.1 ± 0.85	0.24 ± 0.03
meth. amm.	−71.3	−52.0	−54.7 ± 0.89	0.71 ± 0.23	−52.2 ± 0.92	0.22 ± 0.13
water	−5.6		−5.2 + 0.09	53.7 ± 1.3	−4.9 ± 0.14	55.1 ± 0.32

[a]HFE^{exp} from refs 10, 41, and 42. [b]HFE^{fep} calculated using FEP calculations of 1 M solutions using a previously described methodology.[57]

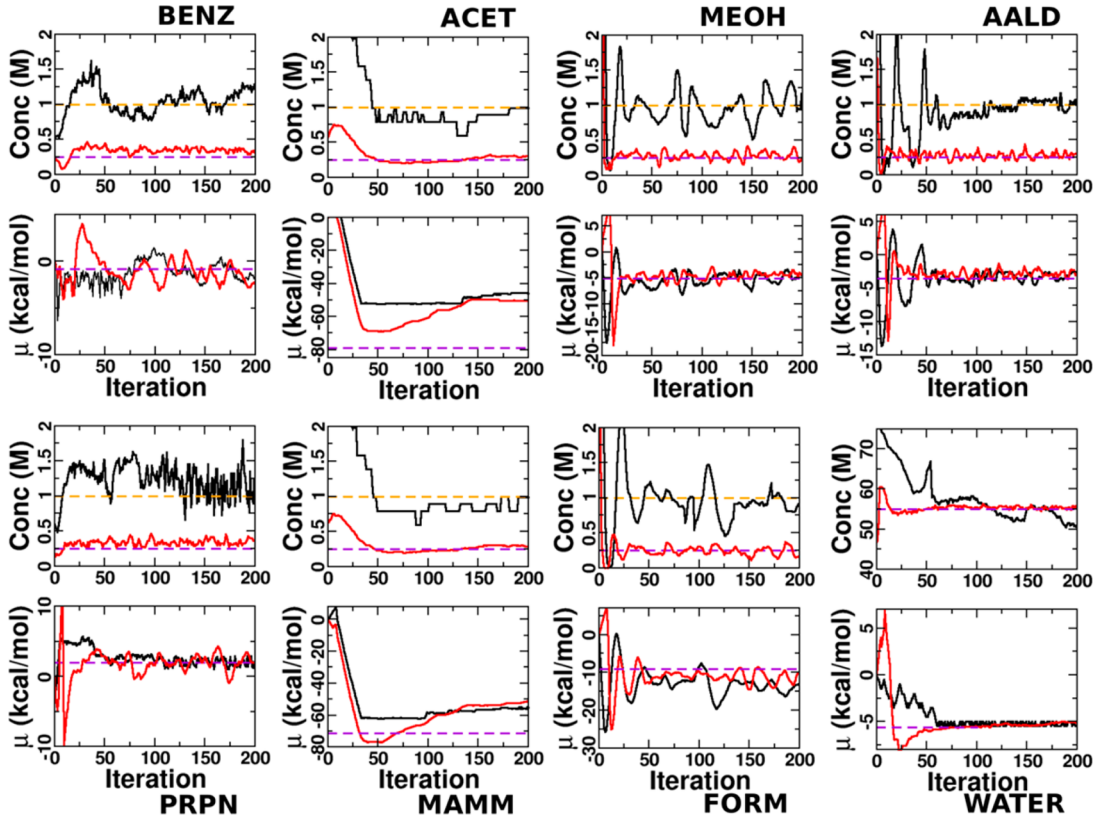


Figure 2. Plots of μ_{ex} and concentrations of the solutes and the water as a function of the oscillating-μ_{ex} GCMC-MD iterations. The μ_{ex} value for the solutes and the water was varied every iteration based on their respective concentrations in the simulation system A. The target concentration of the solutes was 1 and 0.25 M in Schemes I (black) and II (red), respectively, while water was maintained at bulk concentration of 55 M in both the systems. Solutes are benzene (BENZ), propane (PRPN), acetate (ACET), methylammonium (MAMM), methanol (MEOH), formamide (FORM), and acetaldehyde (AALD).

the radius of the simulation system for water was increased by 5 Å, such that the "pancaking" of nonpolar molecules is avoided as they stay fully hydrated. The use of such an extended system will also alleviate other potential edge effects that the other solutes may encounter. We note that the larger system B can extend to any distance beyond the solute restraints defining system A, including being modeled using PBC, as was performed in our calculations that included T4-L99A.

In the remainder of this section results using both Scheme I and II will be presented for aqueous systems alone, which will be followed by Scheme I and II calculations on a system that contains T4-L99A. This lysozyme mutant was selected as a model system as it has been widely used in experimental and computational studies of ligand binding as well as for studies of the impact of mutations on protein structure and stability.[36−39] The L99A mutation in the C-terminal domain creates a completely buried, hydrophobic cavity of ∼150 $Å^3$ that, although inaccessible in static structures, binds small hydrophobic ligands in a rapid and reversible manner.[40−42] Such an occluded pocket offers a rigorous test of the sampling effectiveness of the presented oscillating-μ_{ex} GCMC-MD methodology, including a quantitative evaluation of the approach.

Aqueous Solution Systems. For both Scheme I and aqueous mixture Scheme II the concentration is the target for the variations of the μ_{ex} through the oscillating-μ_{ex} GCMC-MD iterations, with the μ_{ex} of both the solute and the water initially set to 0. The concentration is converted to a number N_P^{target}

dx.doi.org/10.1021/ct500201y | J. Chem. Theory Comput. 2014, 10, 2281−2290

associated with the volume of system A for each solute type or water and applied in eq 2. In both cases, the average value of μ_{ex} of both the solutes and the water converged close to their HFE and the sampling of μ_{ex} values approximate Gaussian distributions (SI, Figure S2). Table 1 lists both the calculated average μ_{ex} values and the experimental HFE (HFE^{exp}) of the solutes.[43,44] The μ_{ex} values were obtained from block averages over the final five 10 cycle blocks obtained between cycles 150 and 200, with the final μ_{ex} calculated as the average of these block averages along with the associated standard errors. HFEs were also calculated using FEP[8] method by inserting a solute molecule in a 27 Å cubic box with bulk water and 1 M of the solute (ie. Nine additional solute molecules in addition to the solute being perturbed), following the simulation methodology as detailed in our previous works,[45] to account for possible limitations in the force field that would yield a HFE in disagreement with the experimental data.

Figure 2 traces the progression of both the concentration and μ_{ex} through the GCMC-MD iterations, as the μ_{ex} was being varied based on N_P^{target} in system A for both Schemes I and II. Both the concentration and the μ_{ex} at each iteration in Figure 2 are presented as the average over the 10 independent simulations. In both Schemes I and II, the solutes and the solvent of system A attain their target N_P^{target} values, corresponding to 1 and 0.25 M of solute in Schemes I and II, respectively, and 55 M of water. For most of the systems, convergence occurred within 50 iterations. However, for the charged fragments the acceptance rates for particle insertions were low, consistent with previous studies,[12,13] due to the unfavorable electrostatic interactions. The convergence took longer for these cases and simulations were hence run for 200 iterations.

Table 1 shows the average μ_{ex} and the concentration over the 10 simulations, from the final 50 iterations. The average μ_{ex} and the concentration were largely similar to those measured across iterations 50−200 (SI, Table S2), consistent with the onset of convergence from iteration 50. For the hydrophobic and polar molecules, the average μ_{ex} values compare well with the HFE^{fep}. The largest deviation occurs with formamide with the values falling between the HFE^{exp} and HFE^{fep} values. The deviation of μ_{ex} and HFE^{fep} from the HFE^{exp} for the charged fragments, acetate and methylammonium, is due to the vacuum-to-solvent interface potential not being accounted for in the present calculations.[46] For monovalent anions/cations, this contribution was calculated to be about ±12.5 kcal/mol, respectively, with the TIP3P water model.[47] Further, as GCMC methods with charged solutes are limited by low acceptance rates for particle insertions, the agreement of the μ_{ex} for the charged systems was expected to be limited, which is most notable with acetate. Overall, these results establish that the presented oscillating-μ_{ex} GCMC-MD methodology approximates the HFE of organic solutes via their μ_{ex} in aqueous systems. In addition the approach is indicated to be suitable for more complex aqueous mixtures as evidenced by the μ_{ex} values obtained from Scheme II being in satisfactory agreement with the experimental and HFE^{fep} values.

While the oscillating-μ_{ex} GCMC-MD simulation protocol achieved the correct concentration and μ_{ex}, the ability of the method to obtain the correct spatial sampling of the solutes in the finite spherical systems is important for studies of heterogeneous systems. Spatial sampling was investigated via the analysis of radial distribution functions (RDF). RDFs of selected solute atoms and the water oxygens were calculated from the cumulative MD sampling of both the Scheme I and II oscillating-μ_{ex} GCMC-MD simulations. These were compared to the RDFs obtained from explicit-water 15 ns PBC MD simulations[27] that maintain Scheme I and II concentrations of the solutes and water in a cubic box with a side of 50 Å and includes the explicit treatment of long-range nonbond interactions via the particle mesh Ewald[33] and isotropic LJ correction methods (See SI, section S1 for simulations details).[27] The RDFs from the finite spherical systems oscillating-μ_{ex} GCMC-MD and the explicit-water PBC MD match very well (SI, Figure S3). This indicates the efficiency of the use of oscillating-μ_{ex} values to drive GCMC sampling and that the treatment of the long-range nonbond interactions in the presented oscillating-μ_{ex} GCMC-MD protocol does not significantly impact the spatial sampling of the aqueous systems. Some fluctuations are seen in the RDF from Scheme II PBC, which are due to sampling issues: the MD only simulations were run for 15 ns versus a total of 100 ns MD in the oscillating-μ_{ex} GCMC-MD protocol. Thus, the present methodology attains spatial sampling consistent with that observed in unbiased MD PBC simulations. We note that the μ_{ex} settled close to the HFE in a Scheme I simulation of acetaldehyde and methanol without MD at the end of every iteration (SI, Table S3). However, since the molecules are rigid during each cycle of GCMC, MD is likely needed to preserve the correct conformational and spatial sampling of the solutes in these bulk-phase environments (see below).

Subsequent calculations focused on determining if the GCMC-MD approach could obtain equilibrium solute sampling with known, fixed μ_{ex} values. With that objective, we ran a set of GCMC-MD simulations, where instead of starting with setting μ_{ex} to 0, it was held fixed at the HFE throughout the iterations. Shown in Figure 3 are the concentration and μ_{ex} for the acetaldehyde system as a function of the oscillating-μ_{ex} GCMC-MD and fixed-μ_{ex} GCMC-MD iterations. In the fixed-μ_{ex} simulations, we found that as the number of iterations increased and waters attained bulk

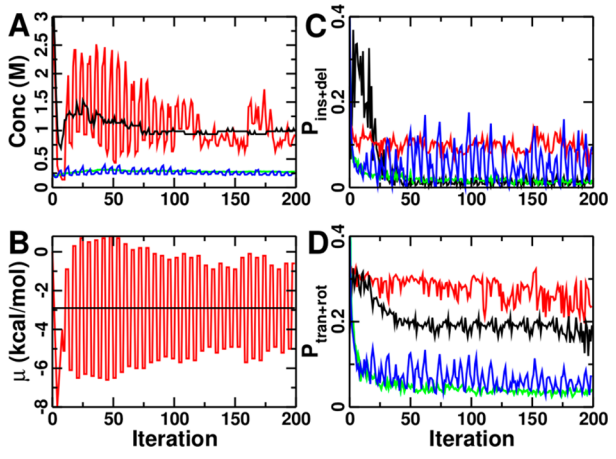


Figure 3. (A) Concentration and (B) μ_{ex} of acetaldehyde and the average probabilities of (C) insertion + deletion ($P_{ins+del}$) and (D) translation + rotation ($P_{trans+rot}$) as a function of the number of oscillating-μ_{ex} GCMC-MD iterations from the Scheme I and II oscillating-μ_{ex} GCMC-MD of acetaldehyde with μ_{ex} fixed at the HFE (black, green) or fluctuated by dμ_{ex} (red, blue), respectively. Note that the number of solute exchanges with the gas-phase reservoir are greater with the fluctuating μ_{ex} while the average concentration and μ_{ex} are approximately equivalent.

concentrations, the number of solute exchanges decreased considerably. Similar trends were seen for the other solutes (Figure S4 and S5 of the SI). This is an outcome of the cavity bias search used during GCMC moves. When the moves are performed for one molecule, the other molecules (both solutes and waters) are stationary and participate only in nonbond interactions with the current molecule being exchanged, thereby preventing overlapping moves of molecules into locations already occupied. Thus, at the 55 M bulk-phase concentration of water, it is easier for the smaller waters to fill up cavities in system A than the larger solutes, leading to a drastic decrease in the GCMC exchange probabilities for the solutes (Figure 3C and D). As continuous GCMC exchanges of fragments through insertions and deletions, and local relaxation through translations and rotations are important to maintain chemical equilibrium between system A and the coupled gas-phase fragment reservoirs this outcome was deemed problematic.

To overcome this, μ_{ex} of both the solutes and the water were cyclically fluctuated around their respective HFEs. $d\mu_P = N_P/N_P^{target}$ was alternately added and subtracted to μ_P (P = F_1, F_2, ..., F_M, W) following every three iterations of GCMC. Thus, the frequency of oscillation is six cycles, and by maintaining this oscillation through the length of the simulation, the average μ_{ex} remains constant at the values in Table 1. We note that by periodically fluctuating μ_{ex} the system is likely not a formal GC ensemble; however, by maintaining the average μ_{ex} constant over the length of the simulation the extent of deviation is minimal, which is supported by the agreement between the oscillating-μ_{ex} GCMC-MD calculated average μ_{ex} and FEP hydration free energies (Table 1).

Such variations in μ_{ex} lead to improvements in sampling over the course of the GCMC-MD iterations, as evidenced by both the change in the number of the solute molecules and the increase in the GCMC move probabilities ($P_{ins+del}$ and $P_{trans+rot}$ for the insertions and deletions and translations and rotations, respectively), as shown in Figure 3. Similar trends were observed for the apolar and the other polar solutes (SI, Figure S4, S5). Thus, together with GCMC of both the solvent and the solute, it is important to continue to fluctuate the μ_{ex} supplied to these molecules once the target concentration, as described in eq 2, is attained to maintain efficient sampling of the solutes in the simulation system A.

T4 Lysozyme Mutant L99A. Computational chemical fragment mapping studies to date[15,16,20,48] have applied MD simulations such that the proteins that could be studied had to have binding sites accessible to the bulk solvent, allowing the solutes and water to diffuse in and out of the sites on the time scale of the simulations. However, a larger number of biologically important proteins, such as the G-protein coupled receptors (GPCRs)[49] and nuclear receptors[50] as well as other macromolecules, have very deep or occluded pockets that would require simulations of extensive duration. To overcome this accessibility limitation, the present oscillating-μ_{ex} GCMC-MD method offers great potential and we selected the well-studied T4-L99A, which contains an engineered occluded binding site for benzene[39] as a model system. The oscillating-μ_{ex} GCMC-MD sampling method was identical to the aqueous systems above with the exception that system B was treated as periodic with the solute being studied included in system B at the target concentration used to drive the GCMC sampling. System A was a 20 Å sphere centered on the T4-L99A binding site defined by residues Ala 99 and Met 102 (SI, Figure S6).

Scheme I calculations on T4-L99A involved only benzene as a solute at 1 M along with 55 M water. This allowed for validation of the method to drive sampling of benzene in the occluded binding pocket of the protein as well as yield a quantitative estimate of ligand binding. Across a 10 × 37.5 ns oscillating-μ_{ex} GCMC-MD simulation with conformations saved every 10 ps, benzene was bound to the T4-L99A binding site for a total of 372 ns with the pocket being empty for only 3 ns. The average benzene concentration in the simulation system A was 1.5 ± 0.2 M. These results indicate that the oscillating-μ_{ex} GCMC-MD method can sample the occluded pocket as well as attain the defined concentration in the entire system that drives the GCMC sampling.

The pocket sampling also allows for estimation of the binding affinity of a ligand ΔG°, using

$$\Delta G^\circ = -RT \ln \frac{[PL]}{[P]} + RT \ln[L]V_{ref} \quad (3)$$

where, R is the gas constant (kcal/mol·K), T (K) is the temperature, [PL] is the concentration of the bound ligand, [L] is the total concentration of the ligand, [P] is the concentration of the protein, and V_{ref} is the reference volume in concentration units (~1660 Å^3 per one ligand molecule for 1 M of ligand).[51] Since the simulations are maintained in equilibrium, the bound vs unbound ligand concentrations can be correlated to the time fraction of ligand bound vs unbound in the binding site through the simulation. A previous technique to determine binding affinity of small molecules to the protein site using the GCMC methodology was limited to gas-phase simulations, using Generalized Born/Solvent Accessibility (GB/SA) solvation calculations to convert the gas-phase energies to aqueous energies.[52] By running oscillating-μ_{ex} GCMC-MD of both the solutes and the waters, binding affinity calculation explicitly incorporates desolvation contribution of both the binding pocket and of the ligand.

As the T4-L99A pocket is completely occluded, the presence or the absence of a solute atom at the active site is driven only by the GCMC insertion/deletion moves. Over the 10 × 37.5 ns oscillating-μ_{ex} GCMC-MD simulation that involves a total of 18.75 million GCMC insertion/deletion attempts for the benzene, with [L] ~ 1.5 M and a [PL]/[P] ratio of 99.4/0.6, ΔG° is about −3.25 kcal/mol from eq 2. Although there is some difference from the experimental binding affinity of −5.19 ± 0.16 kcal/mol,[37] it should be noted that an increase of 10^4 steps out of the 1.9×10^7 total steps with benzene in the pocket translates to nearly 1 kcal/mol difference in the ΔG° calculated. This emphasizes the difficulty of converging the calculation of a binding constant to a single site in a protein, although force field effects could also impact the obtained ΔG°.

Scheme II calculations involved the seven solutes along with water. As with the aqueous system, the target concentration for the solutions was 0.25 M. This simulation was run for a total of 10 × 50 ns. To facilitate analysis of the results, affinity patterns of the selected atoms from the different solutes in the occluded binding pocket, called "Grid Free Energy (GFE) FragMaps",[18] were calculated. GFE FragMaps are Boltzmann transformed probability distributions for the solute atoms that are normalized with the distributions of these molecules in aqueous solution in the absence of the macromolecule. This normalization accounts for the free energy penalty of solute desolvation when calculating the GFEs (see SI, section S2). These maps may then be visualized to qualitatively evaluate the ability of the oscillating-μ_{ex} GCMC-MD sampling method to

reproduce the positions of different ligands known to bind T4-L99A that have been subjected to experimental analysis. Nine ligands were considered: (1) benzene, (2) *o*-xylene, (3) *p*-xylene, (4) ethylbenzene, (5) benzofuran, (6) indene, (7) indole, (8) isobutylbenzene, and (9) *n*-butylbenzene.

Figure 4 shows the aromatic (BENC), aliphatic (PRPC), and polar hydrogen bond donor (MEOH, FORH) and acceptor

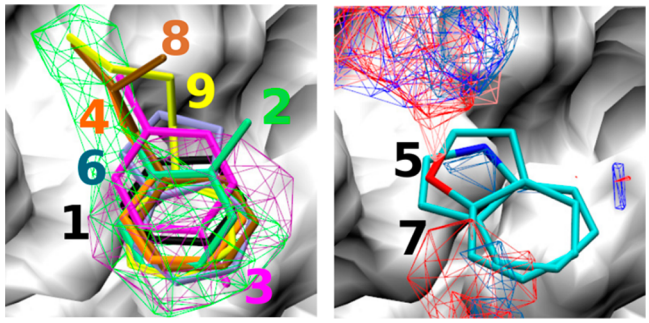


Figure 4. Selected GFE FragMaps at the ligand binding site of the T4-lysozyme L99A from a 10 × 50 ns oscillating-μ_{ex} GCMC-MD simulation and the minimized crystal conformations of the 9 ligands (see text for the list); protein atoms occluding the view of the binding pocket were removed for clear visualization. FragMaps are displayed at a cutoff of −1.2 kcal/mol for the BENC (purple), PRPC (green) and −0.5 kcal/mol for AALO (pink), MEOO (dark red), FORO (red), MEOH (blue), FORH (light blue).

(MEOO, FORO, AALO) maps along with the crystallographic orientations of the ligands. Benzene occupies the aromatic FragMap (purple) while the aliphatic moieties of the other ligands occupy the aliphatic FragMap region (green) that protrudes away from the benzene molecule. In addition, neutral H-bond donor and acceptor maps are in the binding pocket and are in the vicinity of the corresponding functional groups on benzofuran and indole. Importantly, although we use the protein structure from the T4-L99A−benzene complex (PDB 181L) as the starting conformation, the oscillating-μ_{ex} GCMC-MD simulations correctly identify the ability of the pocket to alter its conformation to allow favorable interaction with the aliphatic moieties as well as its ability to accommodate polar functionality. This is an outcome of the inclusion of MD, in which protein flexibility is included, in the presented methodology.[53,54] When a second Scheme II simulation was run without the MD, the pocket in the neighborhood of Ser117 and Leu118 was not available for sampling by the apolar fragments (SI, Figure S7). It is possible that MD sampling could be redundant when system A does not contain any macromolecule, since the μ_{ex} calculated both with and without the MD are similar (SI, Table S3). However, to maintain consistency between the macromolecular and the aqueous environment only sampling, MD is retained during the latter.

The use of GFE FragMaps also has the advantage that it allows for quantitative evaluation of the relative affinity of the ligands, based on Ligand GFE (LGFE) scores, as previously described.[21] LGFE quantifies the overlap of the atoms in the ligand with the corresponding GFE FragMaps. LGFE scores were calculated as Boltzmann weighted averages from ensembles of ligand−protein orientations generated using (i) MD sampling of the ligands bound to the protein and (ii) MC sampling of the ligands in the field of the FragMaps (see SI, section S2). As shown in Figure 5, both $LGFE^{MD}$ and $LGFE^{MC}$ correlate very well with the experimental binding free energies

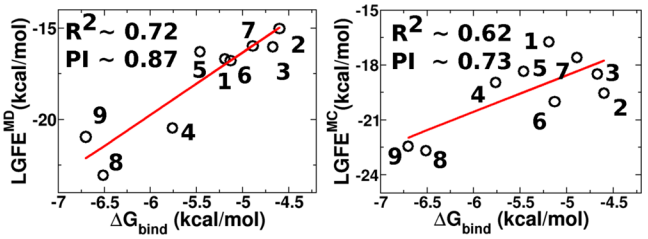


Figure 5. Correlation of the experimental binding affinity $\Delta G°$ (from ref 37) with the LGFE scores for the nine ligands considered (see text for the list). The LGFE scores are obtained from MD and MC conformational ensembles of the ligands ($LGFE^{MD}$, $LGFE^{MC}$) and the GFE FragMaps (see main text and SI, section S2). Overall maps have a very good correlation (high R^2 and predictive index, PI).

(high R^2 and predictive index (PI)[55]). Importantly, the LGFEs can distinguish between the binding activity of both congeneric series and diverse classes of ligands. We note that the range of experimental binding affinities is −2.1 kcal/mol while the LGFE scores from the protein + ligand MD ensemble and the MC sampling are spread over wider ranges of −4.4 and −6.1 kcal/mol, respectively. This is not unexpected as the LGFE scores are not true free energies of binding as numerous terms that contribute binding are omitted (e.g., the configurational entropy of the ligands).[56]

Our novel oscillating-μ_{ex} GCMC-MD methodology therefore allows for investigations of the μ_{ex} of solutes in aqueous solution, including solutions containing multiple solutes. Central to the approach is the use of variable μ_{ex} throughout the GCMC cycles of the simulations. This leads to convergence of μ_{ex} for given solute(s) and environment based on the target concentration and the maintenance of solute sampling once the system has converged with respect to the μ_{ex} or target concentration. Along with probing the μ_{ex} required to maintain the solutes at target concentrations in aqueous environments, the oscillating-μ_{ex} GCMC-MD approach is also useful toward efficient solute sampling. Notably, the method may be used to effectively sample the configurational space of an occluded pocket in a macromolecule; the engineered pocket in T4-L99A in the present study. This sampling may be performed with a complex mixture of solutes to map the functional preference of the binding pocket. When this is analyzed in the context of the SILCS methodology, the approach is shown to qualitatively reproduce the binding orientation of known ligands as well as quantitatively rank order the binding affinity of the ligands. It is anticipated that this capability will be of utility for the application of computer aided drug design methods to macromolecules with deep or occluded binding pockets such as those in protein nuclear factors[50] and GPCRs.[49]

■ ASSOCIATED CONTENT

Supporting Information

Simulation details for the aqueous systems PBC MD; LGFE calculation details; μ_{ex} distributions, RDF plots, exchange probabilities through GCMC for the various solutes in the aqueous systems simulations; average μ_{ex} and concentration of solutes across Scheme I and II oscillating-μ_{ex} GCMC-MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: alex@outerbanks.umaryland.edu.

Notes

The authors declare the following competing financial interest(s): A.D.M. Jr. is co-founder and Chief Scientific Officer of SilcsBio LLC.

ACKNOWLEDGMENTS

We thank all members of the MacKerell group for helpful discussions. This work was supported by NIH grant CA107331 and Maryland Industrial Partnerships Award 5212. The authors acknowledge computer time and resources from the Computer Aided Drug Design (CADD) Center at the University of Maryland, Baltimore.

REFERENCES

(1) Dill, K. A.; Bromberg, S. *Molecular driving forces: statistical thermodynamics in chemistry and biology*; Taylor & Francis, 2003.

(2) Pohorille, A.; Wilson, M. A. Excess chemical potential of small solutes across water−membrane and water−hexane interfaces. *J. Chem. Phys.* **1996**, *104*, 3760.

(3) 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.

(4) Resat, H.; Mezei, M. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. *J. Am. Chem. Soc.* **1994**, *116*, 7451−7452.

(5) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050.

(6) Chang, J. The calculation of chemical potential of organic solutes in dense liquid phases by using expanded ensemble Monte Carlo simulations. *J. Chem. Phys.* **2009**, *131*, 074103.

(7) Raman, E. P.; MacKerell, A. D., Jr. Rapid estimation of hydration thermodynamics of macromolecular regions. *J. Chem. Phys.* **2013**, *139*, 055105.

(8) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(9) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand canonical Monte Carlo simulation of ligand-protein binding. *J. Chem. Info Model* **2006**, *46*, 231−242.

(10) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **2004**, *121*, 6392.

(11) Kulp, J. L., III; Kulp, J. L., Jr.; Pompliano, D. L.; Guarnieri, F. Diverse fragment clustering and water exclusion identify protein hot spots. *J. Am. Chem. Soc.* **2011**, *133*, 10740−10743.

(12) Jayaram, B.; Beveridge, D. Grand canonical Monte Carlo simulations on aqueous solutions of sodium chloride and sodium DNA: excess chemical potentials and sources of nonideality in electrolyte and polyelectrolyte solutions. *J. Phys. Chem.* **1991**, *95*, 2506−2516.

(13) Mezei, M. A cavity-biased (T, V, μ) Monte Carlo method for the computer simulation of fluids. *Mol. Phys.* **1980**, *40*, 901−906.

(14) Small, M. C.; Lopes, P.; Andrade, R. B.; MacKerell, A. D., Jr. Impact of Ribosomal Modification on the Binding of the Antibiotic Telithromycin Using a Combined Grand Canonical Monte Carlo/Molecular Dynamics Simulation Approach. *PLOS Comput. Biol.* **2013**, *9*, e1003113.

(15) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J. Chem. Theory Comput* **2012**, *8*, 2435−2447.

(16) Lexa, K. W.; Carlson, H. A. Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* **2010**, *133*, 200−202.

(17) Guvench, O.; MacKerell, A. D., Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLOS Comput. Biol.* **2009**, *5*, e1000435.

(18) Raman, E. P.; Yu, W.; Guvench, O.; MacKerell, A. D., Jr. Reproducing crystal binding modes of ligand functional groups using Site-Identification by Ligand Competitive Saturation (SILCS) simulations. *J. Chem. Info Model* **2011**, *51*, 877−896.

(19) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(20) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method.Proteins. *Struct, Func, Bioinfo* **1991**, *11*, 29−34.

(21) Raman, E. P.; Yu, W.; Lakkaraju, S. K.; MacKerell, A. D., Jr. Inclusion of multiple fragment types in the Site Identification by Ligand Competitive Saturation (SILCS) approach. *J. Chem. Info Model* **2013**, *53*, 3384−3398.

(22) Torrie, G.; Valleau, J. Electrical double layers. I. Monte Carlo study of a uniformly charged surface. *J. Chem. Phys.* **1980**, *73*, 5807.

(23) Caleman, C.; Hub, J. S.; van Maaren, P. J.; van der Spoel, D. Atomistic simulation of ion solvation in water explains surface preference of halides. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6838−6842.

(24) Mezei, M. Grand-canonical ensemble Monte Carlo study of dense liquid: Lennard-Jones, soft spheres and water. *Mol. Phys.* **1987**, *61*, 565−582.

(25) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

(27) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford university press, 1989.

(28) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(29) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(30) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.

(31) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(32) Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46*, 269−279.

(33) 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.

(34) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **2004**, *15*, 667−683.

(35) Speidel, J. A.; Banfelder, J. R.; Mezei, M. Automatic control of solvent density in grand canonical ensemble Monte Carlo simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1429−1434.

(36) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.

(37) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **1995**, *34*, 8564−8575.

(38) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting ligand binding affinity with alchemical

free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747−763.
(39) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576−8588.
(40) Mulder, F. A.; Hon, B.; Mittermaier, A.; Dahlquist, F. W.; Kay, L. E. Slow internal dynamics in proteins: application of NMR relaxation dispersion spectroscopy to methyl groups in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **2002**, *124*, 1443−1451.
(41) Palmer, A. G., III Probing molecular motion by NMR. *Curr. Opin Struct Biol.* **1997**, *7*, 732−737.
(42) Eriksson, A.; Baase, W. A.; Zhang, X.; Heinz, D.; Blaber, M.; Baldwin, E. P.; Matthews, B. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **1992**, *255*, 178−183.
(43) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(44) Chalmet, S.; Ruiz-López, M. Molecular dynamics simulation of formamide in water using density functional theory and classical potentials. *J. Chem. Phys.* **1999**, *111*, 1117.
(45) Baker, C. M.; Lopes, P. E.; Zhu, X.; Roux, B.; MacKerell, A. D., Jr. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.
(46) Harder, E.; Roux, B. On the origin of the electrostatic potential difference at a liquid-vacuum interface. *J. Chem. Phys.* **2008**, *129*, 234706−234706−234709.
(47) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(48) Ben-Shimon, A.; Eisenstein, M. Computational mapping of anchoring spots on protein surfaces. *J. Mol. Biol.* **2010**, *402*, 259−277.
(49) Moepps, B.; Fagni, L. Mont Sainte-Odile: a sanctuary for GPCRs. *EMBO Rep.* **2003**, *4*, 237.
(50) Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schütz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P. The nuclear receptor superfamily: the second decade. *Cell* **1995**, *83*, 835−839.
(51) Sharp, K. A. Statistical Thermodynamics of Binding and Molecular Recognition Models. *Prot.-Lig. Interact.* **2012**, 3.
(52) Clark, M.; Meshkat, S.; Wiseman, J. S. Grand Canonical Free-Energy Calculations of Protein− Ligand Binding. *J. Chem. Inf. Model.* **2009**, *49*, 934−943.
(53) Foster, T. J.; MacKerell, A. D., Jr.; Guvench, O. Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *J. Comput. Chem.* **2012**, *33*, 1880−1891.
(54) Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007**, *50*, 3457−3464.
(55) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(56) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047−1069.
(57) Deng, Y.; Roux, B. Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.