grand: A Python Module for Grand Canonical Water Sampling in OpenMM

Marley L. Samways,[†] Hannah E. Bruce Macdonald,[‡] and Jonathan W. Essex[*,†]

†*School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK*

‡*Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA*

E-mail: j.w.essex@soton.ac.uk

Abstract

Networks of water molecules can play a critical role at the protein-ligand interface, and can directly influence drug-target interactions. Grand canonical methods aid in the sampling of these water molecules, where conventional molecular dynamics equilibration times are often long, by allowing waters to be inserted and deleted from the system, according to the chemical potential. Here, we present our open source Python module, *grand* (https://github.com/essex-lab/grand), which allows molecular dynamics simulations to be performed in conjunction with grand canonical Monte Carlo sampling, using the OpenMM simulation engine. We demonstrate the accuracy of this module by reproducing the density of bulk water observed from constant pressure simulations. Application of this code to the bovine pancreatic trypsin inhibitor protein reproduces three buried crystallographic water sites that are poorly sampled using conventional molecular dynamics.

Introduction

The importance of water molecules in structure-based drug design has been widely recognized for many years,[1] with a well-known example being the series of cyclic urea inhibitors of HIV protease, for which displacement of a tightly-bound water molecule offers a significant boost in affinity.[2] The prevalence of water-mediated interactions between proteins and small molecules was highlighted by a 2007 study, in which it was found that 85 % of 392 high resolution crystal structures analyzed contained at least one water molecule bridging the protein and ligand.[3]

Molecular dynamics (MD) is now routinely used in computer-aided drug design,[4] given the level of atomistic detail that can be obtained regarding the dynamics of protein structures. However, as the locations of protein-bound water molecules can be difficult to study using experimental methods, especially when the sites are disordered,[5–7] the initial structure for these simulations may not contain water in the correct locations within the binding site of interest. Problematically, the binding and unbinding of buried water molecules can occur on a timescale of microseconds to milliseconds,[8] which would prevent the equilibration of conventional molecular dynamics simulations. Monte Carlo methods can be used in an attempt to bypass these kinetic barriers to water diffusion into the binding site, by adding a random element to the water sampling, such as the method recently presented by Ben-Shalom et al., which employs Monte Carlo translation moves of water molecules between protein binding sites and bulk solvent.[9]

Grand canonical Monte Carlo (GCMC) allows the number of particles in a simulation to fluctuate according to a specified chemical potential,[10–13] and in recent years has been used to enhance the sampling of water molecules at protein-ligand binding interfaces.[14–20] By performing simulations at a range of chemical potentials, the binding free energies of water networks may be estimated.[16,17] GCMC has also been employed in conjunction with relative free energy calculations, to allow the water network around a ligand to respond to an alchemical perturbation, with the effect of water binding or displacement explicitly included

in the free energy calculation.[15,19] Given the sampling limitations of Monte Carlo simulations of proteins, and the more widespread usage of molecular dynamics, we have developed a Python module, named *grand*, which allows GCMC moves to be used to improve water sampling in molecular dynamics simulations using the OpenMM simulation engine.[21] It is important to note that a number of approaches have been developed to allow the particle number to vary during MD simulations.[22–27] Notably, the approach used here of alternating GCMC moves with MD sampling has been used previously by Deng and Roux to improve the sampling of water molecules in protein systems.[15] The focus of this publication is the presentation of the *grand* module, which allows this to be performed in OpenMM.

The nomenclature used here is that GCMC moves refer to particle insertion/deletion moves, the GCMC region is the volume within which these insertion and deletion moves are attempted, and GCMC/MD refers to a simulation in which both molecular dynamics sampling, and GCMC moves are employed.

Theory

To sample states with different numbers of particles, GCMC employs particle insertion and deletion moves, in which a particle is added to, or removed from, the simulation via an ideal gas reservoir. The acceptance probabilities (using the Adams formulation[10,11]) of these moves are as follows,

$$P_{insert} = \min \left[1, \ \frac{1}{N+1} e^B e^{-\beta \Delta U} \right] \tag{1}$$

$$P_{delete} = \min \left[1, \ N e^{-B} e^{-\beta \Delta U} \right] \tag{2}$$

where N is the number of particles in the initial state, B is the Adams parameter,[10,11] β is the thermodynamic beta and ΔU is the potential energy change associated with the proposed

move. Here, the Adams parameter is defined as

$$B = \beta\mu + \ln\left(\frac{V_{GCMC}}{\Lambda^3}\right) \qquad (3)$$

where μ is the chemical potential of the system, V_{GCMC} is the volume of the region in which insertion and deletion moves are attempted, and Λ is the thermodynamic wavelength of the particle. When simulating a GCMC region at equilibrium with bulk water, the corresponding Adams value (B_{equil}) is given by[17]

$$B_{equil} = \beta\mu'_{sol} + \ln\left(\frac{V_{GCMC}}{V^\circ}\right) \qquad (4)$$

where μ'_{sol} is the excess chemical potential of bulk water and V° is the standard state volume of water, with experimental values of -6.324 kcal mol^{-1} and 30.003 Å^3, respectively,[28] where the latter is calculated from the density of bulk water.[29] These values have been calculated here for TIP3P water as $\mu'_{sol} = -6.09$ kcal mol^{-1} and $V^\circ = 30.345$ Å^3, by calculating the hydration free energy of water and the average volume per water molecule, respectively, via constant pressure molecular dynamics simulations (see Supporting Information). Unless explicitly stated otherwise, all values of B_{equil} reported in this work were calculated using these simulated values.

It should be noted that this section describes a basic implementation of GCMC moves, for which the acceptance rate is typically low in condensed systems. Enhanced sampling techniques, such as the configurational[30] and cavity bias[12] methods, as well as continuous fractional component[31] and molecular exchange approaches,[32] can be used to increase the acceptance rate observed,[33] albeit at an increased computational cost and often the introduction of additional user-defined parameters. These methods to boost the efficiency will be the subject of future work, but version 1.0.0 of *grand* does not support this.

Implementation

Our previous work[16,17,19] has involved the use of GCMC simulations in ProtoMS,[34] where the configurational sampling of the system is carried out using conventional Monte Carlo (MC) sampling. However, this approach is typically limited, as configurational sampling is often somewhat restricted in MC, particularly for proteins, as the sampling of large-scale concerted motions is challenging. GCMC/MD offers a solution to these issues, given that MD offers better configurational sampling of the protein than MC, and additionally facilitates the use of enhanced sampling techniques to further improve the sampling of a protein structure.

To improve the convergence of GCMC moves, typically the insertion and deletion moves are restricted to a region of particular interest — usually defined within the protein active site. When using GCMC sampling with MC simulations, where protein sampling is limited, it is trivial to define a hard-wall constraint in Cartesian space to define this region. Insertion and deletion moves are only attempted within this defined region, and any translation moves that would move a water out of the region, or a bulk water into the region, are rejected. The use of a hard-wall GCMC region in MD simulations is less trivial, and therefore we now permit diffusion of water molecules both into, and out of the GCMC region. A region defined using Cartesian coordinates is not useful in MD, where large-scale protein movement and rigid-body rotation is possible, and therefore we have moved to defining the GCMC region based on internal coordinates of the system. When sampling at the equilibrium chemical potential, this permeable barrier will be appropriate, but this may be problematic for titration simulations, where at some chemical potentials the region will be significantly under-hydrated.

To allow GCMC insertion and deletion of water molecules to be executed during MD simulations *grand* regularly executes GCMC moves to update the particle number, just as simulations in the NPT ensemble periodically update the simulation box dimensions via volume changes implemented through a barostat. This Python module includes several 'Sampler' objects which handle the aspects of the simulation related to GCMC, including

the execution of insertion and deletion moves. This is designed to be compatible with the Python API of OpenMM, such that only a minimal amount of additional knowledge is required to run GCMC/MD simulations. The various simulation parameters, including μ'_{sol} and $V°$ (which default to the values of -6.09 kcal mol^{-1} and 30.345 Å^3 determined in this work) are defined when the Sampler object is created, along with the GCMC region, which can be set to either a sphere, centered on a group of reference atoms, or the entire simulation volume.

To test the accuracy of this implementation of GCMC/MD, and the underlying theory, a bulk water system was simulated using NPT MD and GCMC/MD to allow the mass density to fluctuate; the former does this by varying the system volume, while the latter varies the particle number. Further, to assess the sensitivity of the grand canonical approach to the values of the excess chemical potential and standard state volume of water, the same test was repeated using the experimental values of $\mu'_{sol} = -6.324$ kcal mol^{-1} and $V° = 30.003$ Å^3. Bovine pancreatic trypsin inhibitor (BPTI) was chosen as a protein test case, as the binding site containing three crystallographic waters has been used to test GCMC sampling in previous work.[16,17] This system was sampled using GCMC/MD to see if the locations of these three waters would be reproduced, with the results compared to those from conventional MD.

Methods

The AMBER ff14sb force field was used to model the protein,[35] the TIP3P force field for the water,[36] and Joung-Cheatham parameters for the ions.[37,38] The real-space cutoff for nonbonded interactions was taken as 12 Å, with a switching function applied between 10 - 12 Å for Lennard-Jones interactions (a long-range correction was not used - see Supporting Information), and particle mesh Ewald (PME) was used to calculate the effect of long-range electrostatic interactions.[39] All bonds involving hydrogen atoms were constrained using the

SHAKE algorithm for the protein,[40,41] and the SETTLE algorithm for water.[42] The dynamics were sampled at a temperature of 298 K using the BAOAB Langevin integrator[43,44] (with a timestep of 2 fs and friction coefficient of 1 ps^{-1}), as this integrator has been found to be extremely accurate for equilibrium sampling.[45] Where constant pressure simulations were performed, a Monte Carlo barostat was used to maintain a pressure of 1 bar, with volume changes attempted every 25 timesteps. All simulations used version 7.2.2 of OpenMM,[21] and version 1.0.0 of *grand*.

Bulk Water. In each case, simulations were run for 100 ns, starting from a pre-equilibrated box containing 2094 water molecules at a density of 1.004 g mL^{-1}, with densities recorded every 0.5 ns — this was run three times. During the GCMC/MD simulations, 125 GCMC moves were executed every 250 fs — these moves sampled the entire simulation volume ($B_{equil} = -2.655$), as the very large values of N and V_{GCMC} would be expected to exacerbate any underlying errors in the algorithm or associated parameters. As previously stated, a comparison was also carried out using the values of $\mu'_{sol} = -6.324$ kcal mol^{-1} and $V^{\circ} = 30.003$ Å^3 ($B_{equil} = -3.039$).

BPTI. First, all hydrogen/deuterium atoms were removed from the starting structure (PDB ID: 5PTI), and any sidechain conformational issues were resolved by selecting the more occupied conformation. The system was protonated using the *Modeller* tool found in OpenMM,[21] with charged termini. The protein was then solvated in a box of TIP3P water (with 8 Å of padding around the protein in each dimension) and chloride ions were added to neutralize the system charge, using AmberTools.[46] The GCMC region was defined as a sphere, centered on the mean coordinate of the C$_\alpha$ atoms of the Tyr10 and Asn43 residues, with a radius of 4.2 Å ($B_{equil} = -7.959$), and any waters present in this region were deleted. The equilibration process was executed in three stages: GCMC/MD to equilibrate the water distribution, NPT MD to equilibrate the system volume, and further GCMC/MD to equilibrate the waters at the new system volume. The first GCMC/MD stage involved an initial 10k GCMC moves, followed by 1 ps of GCMC/MD (1000 GCMC moves every 10 fs of

MD) — the ratio of GCMC moves to MD steps is high here, to prevent the dry pocket from collapsing. The NPT stage lasted 500 ps, as did the second GCMC/MD stage (200 GCMC moves every 1 ps of MD) — the GCMC moves are fewer and less frequent here, to allow more MD sampling. The production simulation involved 10 ns of GCMC/MD (50 GCMC moves carried out every 1 ps of MD). For comparison, an equivalent set of simulations was carried out using conventional MD, in which all of the GCMC/MD stages described were replaced with NVT MD.

Results and Discussion

Bulk Water. Figure 1 shows the density variation over time for the bulk water GCMC/MD simulations carried out, compared with that of the NPT simulations. As can be seen from Figure 1a, the GCMC/MD simulations using the newly parametrized values of the excess chemical potential and standard state volume show excellent agreement with NPT results. It was also found that the accuracy of the GCMC/MD results is good when starting the simulation from densities far from the equilibrium value (Figure S1). This would suggest that the implementation presented in the *grand* module is correct, within the limits of the simulation parameters. When using the experimental values for the excess chemical potential and standard state volume, there is a significant discrepancy between the GCMC/MD and NPT results (Figure 1b). This would imply that the results are very sensitive to these values, and also demonstrates the importance of using internally consistent parameters. However, as previously mentioned, this is a large system, where any errors will be magnified, so the sensitivity may not be as significant for typical protein-binding site applications. Nevertheless, when simulating under drastically different conditions to those presented here (such as different water model, long-range electrostatic treatment, interaction cutoff, etc.), it is recommended that these parameters be recalibrated appropriately. This sensitivity is further demonstrated in Figure S2, where shorter versions of these simulations were carried
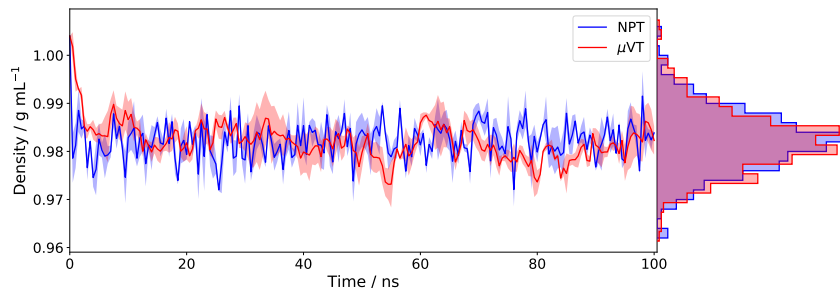
(a) GCMC/MD simulation densities of water using $\mu'_{sol} = -6.09$ kcal mol^{-1} and $V^\circ = 30.345$ Å^3.
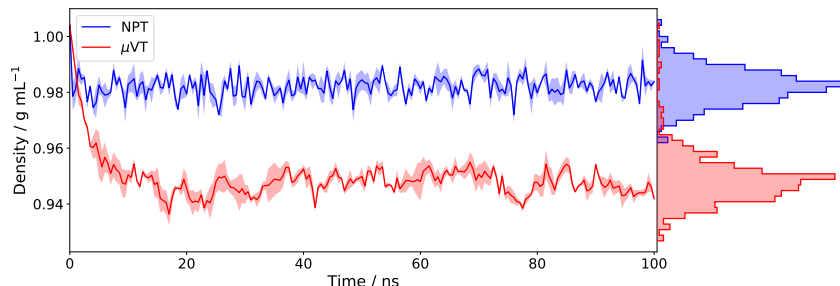
(b) GCMC/MD simulation densities of water using $\mu'_{sol} = -6.324$ kcal mol^{-1} and $V^\circ = 30.003$ Å^3.

Figure 1: Variation in the water density over time for the GCMC/MD simulations performed (red data), compared with the NPT results (blue data). The solid line represents the mean density over the three independent repeats, and the shaded region represents the associated standard error. The histograms were plotted using all density samples.

out using several values of μ'_{sol} between -6.15 and -6.25 kcal mol^{-1}.

BPTI. The observed water positions in both the GCMC/MD and NVT simulations of BPTI were clustered, using average-linkage hierarchical clustering of the oxygen atom locations, with a distance cutoff of 2.4 Å — it should be noted that the distance observed between waters in the same simulation frame was set to an arbitrarily high value, in order to prevent distinct water sites from being clustered together. The location of each cluster was taken as the closest observed oxygen location to the cluster centroid. These sites are shown alongside the three crystallographic sites in Figure 2, where the simulated data were superimposed on the experimental data by aligning the C$_\alpha$ atoms of the protein. As can be seen, the crystallographic sites are very well reproduced by the GCMC/MD simulation — all three GCMC/MD clusters show occupancies greater than 99 %, and match the experimental locations to within 0.6 Å. The experimental agreement is significantly worse for the

NVT results, where only two of the sites are reproduced within 1.0 Å, and the occupancies are much lower (51 & 57 %) — free energy analysis has shown these three sites to be very stable,[17] and they would therefore be expected to be highly occupied. It should be noted that a water site with a binding free energy of zero would be expected to have an occupancy of 50 % in GCMC/MD. Several additional sites were observed for both simulations, though these were all occupied for 12 % or less of the total simulation time, indicating that they are not stable and therefore of little interest. In terms of speed, the entire GCMC/MD process (including equilibration) was around 4.4 times more computationally expensive than the NVT equivalent; however, this increased cost is dependent on the efficiency of the implementation and the frequency and number of GCMC moves executed, and so could easily be reduced, if required. The three crystallographic water sites were all occupied within minutes of the GCMC/MD equilibration process, using a single GPU, whereas the NVT simulation never samples all three sites simultaneously in over two hours. In this context, the *grand* implementation of GCMC/MD is significantly more efficient than conventional MD.

Conclusions

An implementation of GCMC moves for the insertion and deletion of water molecules from a molecular dynamics simulation has been presented, in a freely available Python module, named *grand*. The accuracy of this implementation was verified by using GCMC/MD to reproduce the bulk water density distribution observed from constant pressure simulations. It should be noted that this accuracy test appears sensitive to the values of the excess chemical potential and standard state volume of water used by GCMC; users would be well advised to re-parametrize these values if their simulations conditions diverge from those presented here. This module was also applied to the BPTI protein, where three buried water sites were well reproduced. In contrast, standard NVT MD did not sample these sites well, implying that their equilibration time is longer than the 10 ns simulated. Therefore, whilst the GCMC/MD
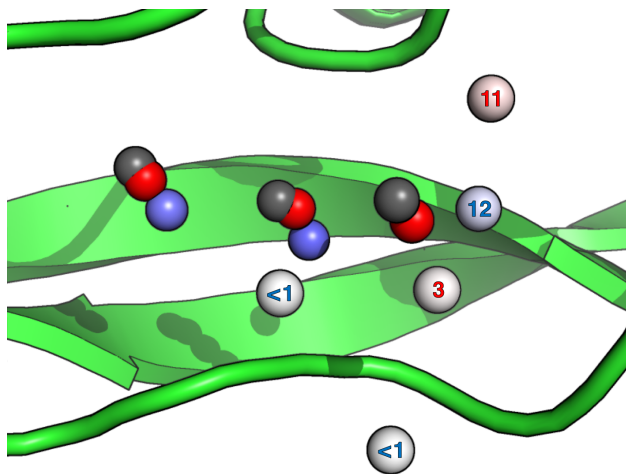
Figure 2: Clustered water positions from the GCMC/MD (red) and MD (blue) simulations of BPTI, with the crystallographic waters shown in gray. The clustered sites are colored by their occupancy, where the more highly occupied sites are shown darker. Where the color is unclear, the sites are labeled by their percentage occupancies.

protocol is more computationally expensive, the water distribution is equilibrated much more rapidly, requiring fewer samples to be discarded as unequilibrated. This cost could be reduced by using *grand* to equilibrate binding site waters, prior to MD simulation, given that the three crystallographic waters in the BPTI binding site were inserted rapidly during GCMC/MD equilibration, and sampled poorly by NVT simulation.

It also should be noted that the acceptance rates observed for the GCMC/MD simulations here are very low (on the order of 0.01 %), relative to typical MC simulations, as is commonly observed for unbiased GCMC simulation in condensed phases.[17] Future work will seek to optimize the performance of the *grand* module, including the use of enhanced sampling techniques to boost the acceptance rates observed. However, the results presented here have demonstrated the accuracy and utility of GCMC/MD for the sampling of occluded water sites. The *grand* module is released under the MIT license, and can be downloaded at https://github.com/essex-lab/grand. Documentation is available at

https://grand.readthedocs.io, and several examples are provided in the GitHub repository.

Acknowledgement

Supporting Information Available

Supporting information is included. All scripts and data required to reproduce the results presented in this Application Note can be found at https://github.com/essex-lab/grand-paper.

References

(1) Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973–980.

(2) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C.-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380–384.

(3) Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.

(4) Defelipe, L.; Arcon, J.; Modenutti, C.; Marti, M.; Turjanski, A.; Barril, X. Solvents to Fragments to Drugs: MD Applications in Drug Design. *Molecules* **2018**, *23*, 3269.

(5) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed.* **2003**, *42*, 2718–2736.

(6) Ohlendorf, D. H. Acuracy of refined protein structures. II. Comparison of four independently refined models of human interleukin 1β. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 808–812.

(7) Fields, B. A.; Bartsch, H. H.; Bartunik, H. D.; Cordes, F.; Guss, J. M.; Freeman, H. C. Accuracy and precision in protein crystal structure analysis: two independent refinements of the structure of poplar plastocyanin at 173 K. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 709–730.

(8) Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017**, *117*, 10694–10725.

(9) Ben-Shalom, I. Y.; Lin, C.; Kurtzman, T.; Walker, R. C.; Gilson, M. K. Simulating Water Exchange to Buried Binding Sites. *J. Chem. Theory Comput.* **2019**, *15*, 2684–2691.

(10) Adams, D. Chemical potential of hard-sphere fluids by Monte Carlo methods. *Mol. Phys.* **1974**, *28*, 1241–1252.

(11) Adams, D. Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Mol. Phys.* **1975**, *29*, 307–311.

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

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

(14) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **2004**, *121*, 6392–6400.

(15) Deng, Y.; Roux, B. Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **2008**, *128*, 115103.

(16) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137*, 14930–14943.

(17) Ross, G. A.; Bruce Macdonald, H. E.; Cave-Ayland, C.; Cabedo Martinez, A. I.; Essex, J. W. Replica-Exchange and Standard State Binding Free Energies with Grand Canonical Monte Carlo. *J. Chem. Theory Comput.* **2017**, *13*, 6373–6381.

(18) Wahl, J.; Smieško, M. Assessing the Predictive Power of Relative Binding Free Energy Calculations for Test Cases Involving Displacement of Binding Site Water Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 754–765.

(19) Bruce Macdonald, H. E.; Cave-Ayland, C.; Ross, G. A.; Essex, J. W. Ligand Binding Free Energies with Adaptive Water Networks: Two-Dimensional Grand Canonical Alchemical Perturbations. *J. Chem. Theory Comput.* **2018**, *14*, 6586–6597.

(20) Bodnarchuk, M. S.; Packer, M. J.; Haywood, A. Utilizing Grand Canonical Monte Carlo Methods in Drug Discovery. *ACS Med. Chem. Lett.* **2020**, *11*, 77–82.

(21) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(22) Çagin, T.; Pettitt, B. M. Grand Molecular Dynamics: A Method for Open Systems. *Mol. Simulat.* **1991**, *6*, 5–26.

(23) Papadopoulou, A.; Becker, E. D.; Lupkowski, M.; van Swol, F. Molecular dynamics and Monte Carlo simulations in the grand canonical ensemble: Local versus global control. *J. Chem. Phys.* **1993**, *98*, 4897–4908.

(24) Heffelfinger, G. S.; Swol, F. v. Diffusion in Lennard-Jones fluids using dual control volume grand canonical molecular dynamics simulation (DCV-GCMD). *J. Chem. Phys.* **1994**, *100*, 7548–7552.

(25) Im, W.; Seefeld, S.; Roux, B. A Grand Canonical Monte Carlo–Brownian Dynamics Algorithm for Simulating Ion Channels. *Biophys. J.* **2000**, *79*, 788–801.

(26) Pool, R.; Heringa, J.; Hoefling, M.; Schulz, R.; Smith, J. C.; Feenstra, K. A. Enabling grand-canonical Monte Carlo: Extending the flexibility of GROMACS through the GromPy python interface module. *J. Comput. Chem.* **2012**, *33*, 1207–1214.

(27) Jung, C. K.; Braunwarth, L.; Jacob, T. Grand Canonical ReaxFF Molecular Dynamics Simulations for Catalytic Reactions. *J. Chem. Theory Comput.* **2019**, *15*, 5810–5816.

(28) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027.

(29) Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data* **1975**, *20*, 97–105.

(30) Bai, P.; Siepmann, J. I. 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.

(31) Shi, W.; Maginn, E. J. Continuous Fractional Component Monte Carlo: An Adaptive Biasing Method for Open System Atomistic Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 1451–1463.

(32) Soroush Barhaghi, M.; Torabi, K.; Nejahi, Y.; Schwiebert, L.; Potoff, J. J. Molecular exchange Monte Carlo: A generalized method for identity exchanges in grand canonical Monte Carlo simulations. *J. Chem. Phys.* **2018**, *149*, 072318.

(33) Torres-Knoop, A.; Burtch, N. C.; Poursaeidesfahani, A.; Balaji, S. P.; Kools, R.; Smit, F. X.; Walton, K. S.; Vlugt, T. J.; Dubbeldam, D. Optimization of Particle Transfers in the Gibbs Ensemble for Systems with Strong and Directional Interactions Using CBMC, CFCMC, and CB/CFCMC. *J. Phys. Chem. C* **2016**, *120*, 9148–9159.

(34) Woods, C. J.; Michel, J.; Bodnarchuk, M. S.; Bradshaw, R. T.; Ross, G. A.; Cave-Ayland, C.; Bruce Macdonald, H. E.; Cabedo Martinez, A. I.; Samways, M. L.; Graham, J. A. ProtoMS 3.4. 2018; http://protoms.org, (Date accessed: January 17, 2020).

(35) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(36) 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*, 926–935.

(37) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(38) Joung, I. S.; Cheatham, T. E. Molecular Dynamics Simulations of the Dynamic and

Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, *113*, 13279–13290.

(39) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N .log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(40) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(41) Yoneya, M.; Berendsen, H. J. C.; Hirasawa, K. A Non-Iterative Matrix Method for Constraint Molecular Dynamics Simulations. *Mol. Simulat.* **1994**, *13*, 395–405.

(42) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(43) Leimkuhler, B.; Matthews, C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Appl. Math. Res. eXpress* **2012**, abs010, arXiv: 1203.5428.

(44) Chodera, J. D.; Rizzi, A.; Naden, L.; Beauchamp, K. A.; Grinaway, P. B.; Fass, J.; Rustenburg, A. S.; Ross, G. A.; Simmonett, A. C.; Swenson, D. W. H. openmm-tools. 2018; `https://github.com/choderalab/openmmtools`, (Date accessed: April 24, 2020).

(45) Fass, J.; Sivak, D.; Crooks, G.; Beauchamp, K.; Leimkuhler, B.; Chodera, J. Quantifying Configuration-Sampling Error in Langevin Simulations of Complex Molecular Systems. *Entropy* **2018**, *20*, 318.

(46) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.;

Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2018. 2018.

Graphical TOC Entry