

Biomolecular simulations at constant macroscopic counterion concentration

John D. Chodera^{1,*}

¹Computational Biology Program, Sloan Kettering Institute,
Memorial Sloan Kettering Cancer Center, New York, NY 10065

(Dated: January 4, 2016)

While biochemical and biophysical experiments are generally performed at a fixed macroscopic salt concentration (e.g. 50 mM NaCl), biomolecular simulations are almost universally performed with a fixed *microscopic* concentration of counterions that is highly unrepresentative of the actual average local concentration and fluctuations of counterion species surrounding often-charged biomolecules. Here, we describe a simple approach to faithfully simulate the correct equilibrium distribution of counterion species in an equilibrium molecular dynamics simulation by incorporating Monte Carlo moves in which counterions are inserted and deleted. To ensure high acceptance rates, we utilize nonequilibrium candidate Monte Carlo (NMC) moves for counterion insertion and deletion, and self-adjusted mixture sampling (SAMS) is used to rapidly compute the reference chemical potential for the desired macroscopic counterion concentration. We illustrate this approach on a few biomolecules of typical interest to biomolecular simulations, including a small protein (DHFR), an activated kinase (phosphorylated Abl), and a nucleic acid (DNA).

Keywords: biomolecular simulation; molecular dynamics; counterion concentration; chemical potential; semigrand canonical ensemble; self-adjusted mixture sampling

INTRODUCTION

Wolf chartreuse beard, paleo bushwick locavore tumblr selvage health goth narwhal post-ironic meggings cronut DIY etsy. Tote bag viral craft beer migas, brooklyn keffiyeh shabby chic wayfarers godard scenester affogato pabst. Humblebrag chartreuse schlitz, post-ironic wolf ethical narwhal salvia everyday carry gastropub venmo kale chips. You probably haven't heard of them cornhole tilde readymade mixtape irony. Sriracha occupy yuccie, green juice roof party fap tumblr hammock mumblecore ramps pabst. Artisan listicle truffaut kogi, shabby chic kombucha distillery etsy cronut +1 pabst mustache VHS vinyl green juice. Before they sold out brooklyn yuccie, gluten-free sriracha lumbersexual four loko kombucha semiotics letterpress biodiesel kale chips art party normcore slow-carb.

instantaneous number of salt molecules in the system. The thermodynamic parameters that define the equilibrium ensemble are the inverse temperature $\beta \equiv (k_B T)^{-1}$, the pressure p , and the number of solvent molecules N . μ_{salt} denotes the chemical potential of the neutral salt species.

We envision a system that is in equilibrium with a much larger bath of macroscopic salt concentration c_0 , but where pairs of salt molecules are allowed to exchange with solvent molecules through a permeable membrane to alter the instantaneous concentration of salt molecules within the simulation volume we are considering. For example, we may consider an experiment that takes place inside a test tube with a macroscopic concentration c_0 of 50 mM NaCl.

Simulating from the equilibrium distribution

In order to sample from the equilibrium ensemble given in Eq. , we utilize an *expanded ensemble* simulation [?] in which we alternate between sampling the configuration at fixed counterion concentration and sampling a new counterion concentration. Because we are simulating in an explicit solvent system where solvent polarization around ions means that instantaneous withdrawal of an ion would lead to high rejection rates, we utilize the nonequilibrium candidate Monte Carlo (NMC) scheme [?] to greatly enhance Monte Carlo acceptance rates when ions are inserted or deleted.

To simplify the simulation procedure and avoid the need for fully transdimensional simulations where new particle coordinates are created or destroyed, we choose to transform solvent molecules into ions and vice-versa. This approach is straightforward when the solvent is water (as most water models—such as TIP3P [?]—have a single Lennard-Jones site) and only monovalent counterions are used (such as NaCl molecules), this approach is more complex when multisite ion models are used (such as multisite divalent ion models [?] or molecular salt species). In those cases, a

METHODS

Semigrand canonical ensemble

Consider a system in the μ PT ensemble, where the number of water molecules is fixed but the number of *counterions* is allowed to change

$$\pi(x|\mu_{\text{salt}}, \beta, p, N) \equiv Z(\mu, \beta, p, N)^{-1} q(x|\mu, \beta, p, N) \quad (1)$$

$$q(x|\mu, \beta, p, N) \equiv e^{-\beta[U(x)+pV(x)+\mu_{\text{salt}}N_{\text{salt}}(x)]} \quad (2)$$

Here, x denotes all dynamical variables of the simulation, which includes the configuration of all particles in the system, the *number* of salt particles, and the instantaneous box dimensions. $U(x)$ denotes the instantaneous potential energy, $V(x)$ the instantaneous box volume, and $N_{\text{salt}}(x)$ the

* Corresponding author; random@choderalab.org

number of noninteracting dummy salt molecules may be introduced into the system at the beginning of the simulation to avoid the need to dynamically create or destroy new degrees of freedom.

The simulation proceeds as follows:

- **Propagate the simulation at fixed counterion number.**

$$x^{(n+1)} \sim \pi(x|x^{(n)}, N_{\text{salt}}^{(n)}) \quad (3)$$

$$N_{\text{salt}}^{(n+1)} \sim \pi(N_{\text{salt}}|x^{(n+1)}) \quad (4)$$

Calibrating the chemical potential

Consider an experiment that takes place inside a test tube with a macroscopic concentration c_0 of 50 mM NaCl. What does this mean in terms of a chemical potential μ appearing in Eq. ? If our simulation volume consisted of a very large solvent box that does not contain any biomolecules or other impurities, a given choice of chemical potential μ_{salt} will have a one-to-one correspondence with a macroscopic salt concentration:

$$c(\mu, \beta, p, N) \equiv \frac{\langle N_{\text{salt}} \rangle_{\mu, \beta, p, N} / N_A}{\langle V \rangle_{\mu, \beta, p, N}} \quad (5)$$

where N_A is Avogadro's number, and the resulting concentration c is a molar concentration if volume V is expressed in liters. This concentration $c(\mu, \beta, p, N)$ obviously depends on the thermodynamic parameters β, p, N in addition to the chemical potential μ , but for sufficiently large N , it will be independent of N . [JDC: Correct equation to be molarity instead of number density.]

For a given set of simulation conditions (β, p, N) and a given water and ion model, we simply need to identify the choice of μ^* that yields the desired macroscopic concentration c_0 :

$$c_0 = c(\mu^*, \beta, p, N) \quad (6)$$

To do this, we make use of a calibration simulation containing only N solvent molecules (with no biomolecule) at the same inverse temperature β and pressure p , allowing the number of salt molecules N_{salt} to fluctuate dynamically.

While there are many ways to determine μ^* , here, we utilize a *self adjusted mixture sampling* (SAMS) simulation [?]. In a SAMS simulation, we allow the number n of neutral salt

molecules to fluctuate dynamically and the relative reduced free energy $f_n \equiv -\ln Z_n$ of each is computed through a recursion strategy in which the f_n are dynamically evolved during the simulation. [JDC: More details on SAMS.]

$$Z_n \equiv \int dx e^{-\beta[U_n(x) + pV(x)]} \quad (7)$$

where $U_n(x)$ denotes there are exactly n salt pairs in the system.

Once the relative reduced free energies f_n have been determined, we can compute the chemical potential μ^* by recognizing that

$$\langle n \rangle_{\mu, \beta, p, N} = \sum_{n=0}^{\infty} n \int dx e^{-\beta[U_n(x) + pV(x) + \mu n]} \quad (8)$$

$$= \sum_{n=0}^{\infty} n e^{\beta \mu n} \int dx e^{-\beta[U_n(x) + pV(x)]} \quad (9)$$

$$= \sum_{n=0}^{\infty} n e^{\beta \mu n} e^{-f_n} \quad (10)$$

$$= \sum_{n=0}^{\infty} n e^{\beta \mu n - f_n} \quad (11)$$

and write $c(\mu, \beta, p, N)$ in terms of the $\{f_n\}$ and the associated average volumes $V_n \equiv \langle V \rangle_{\mu, \beta, p, N, n}$

$$c(\mu, \beta, p, N) = \left\langle \frac{n}{N_A} \right\rangle_{\mu, \beta, p, N} / \langle V \rangle_{\mu, \beta, p, N} \quad (12)$$

$$= \frac{\sum_{n=0}^{\infty} \frac{n}{N_A} e^{\beta \mu n - f_n}}{\sum_{n=0}^{\infty} V_n e^{\beta \mu n - f_n}} \quad (13)$$

which allows us to solve for μ^* :

$$c_0 = \frac{\left\langle \frac{n}{N_A} \right\rangle_{\mu, \beta, p, N}}{\langle V \rangle_{\mu, \beta, p, N}} \quad (14)$$

$$= \frac{\sum_{n=0}^{\infty} \frac{n}{N_A} e^{\beta \mu n - f_n}}{\sum_{n=0}^{\infty} V_n e^{\beta \mu n - f_n}} \quad (15)$$

ACKNOWLEDGMENTS

JDC acknowledges a Louis V. Gerstner Young Investigator Award, NIH core grant P30-CA008748, and the Sloan Kettering Institute for funding during the course of this work.