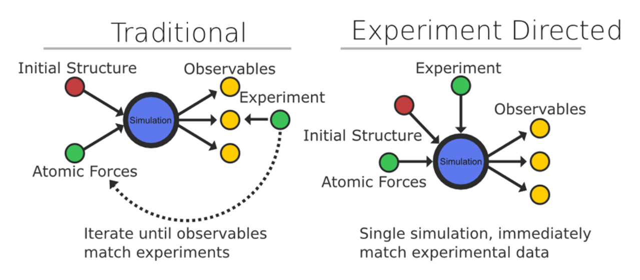
JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data

Andrew D. White and Gregory A. Voth*

Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, The University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States

ABSTRACT: A primary goal in molecular simulations is to modify the potential energy of a system so that properties of the simulation match experimental data. This is traditionally done through iterative cycles of simulation and reparameterization. An alternative approach is to bias the potential energy so that the system matches experimental data. This can be done while minimally changing the underlying free energy of the molecular simulation. Current minimal biasing methods require replicas, which can lead to unphysical dynamics and introduces new complexity: the choice of replica number and their properties. Here, we describe a new method, called experiment directed simulation that does not require replicas, converges rapidly, can match many data simultaneously, and minimally modifies the potential. The experiment directed simulation method is demonstrated on model systems and a three-component electrolyte simulation. The theory used to derive the method also provides insight into how changing a molecular force-field impacts the expected value of observables in simulation.



■ INTRODUCTION

The construction of molecular force-fields that are consistent with all available experimental data is a significant challenge. Current state-of-the-art methods rely on sophisticated iterative methods to build a molecular force-field consistent with experimental data.[1−3] These methods must be repeated for each new piece of experimental data if additional data is found. An alternative approach is to bias a molecular simulation to match experimental data.[4−8] This requires no time-consuming parameter searching and offers a mathematical framework that shows the force-field has changed as little as possible to match the experimental data.[7,9−11] The biasing approach also addresses another related problem: when a molecular model can never match experimental data because no valid parameters in the range of possible force-fields exist. For example, no radially symmetric single-site coarse-grained water model to date can reproduce the tetrahedral structure of water. When a bias on the tetrahedral order of water is added, there becomes a potential energy term that is not radially symmetric. Despite the promising attributes of these biasing methods, existing biasing approaches rely on simulating multiple replicas, with each replica subject to a different strength of bias. Theoretically, there should be many replicas per biased quantity, though in practice 2−20 replicas are sufficient.[10,12] Here, we describe a single-replica approach that can simultaneously match multiple experimental values and then apply the method to four different molecular dynamics simulations. The method is called experiment directed simulation (EDS).

Ideas for incorporating experimental data into molecular models have been studied for many decades in statistical mechanics. Some of the first experimental data came from scattering and was used to create interaction potentials with Ornstein−Zernike equation inversion.[13,14] Such approaches see modern use as well, in colloid−polymer mixtures and with the RISM method.[15−19] More recent successes in constructing force-fields consistent with experimental data have used statistical model-fitting techniques, for example in the TIP4P-2005 model.[20] Current research is pursuing semisupervised construction of force-fields given large sets of experimental data.[1] These newer methods can lead to experimentally consistent force-fields but require iterative cycles of simulations followed by parameter changes, nonlinear optimization techniques, and careful tuning through initial parameter choices and regularization. Another challenge is that there may be many combinations of parameters that lead to a better match with experimental data. Yet, it is unclear which combination is best. These iterative methods have been successful in single-component condensed phase simulations.[21] In many component systems such as protein simulations, hand tuning of parameters is still the most successful method. Thus, modifying a protein force-field to match experimental data for a specific protein can be time-consuming task.

Instead of modifying the force-field itself, other research is pursuing biasing approaches when there already exists an approximately correct force-field.[4,7,11] These methods bias a system to match some experimental observable. This method does not allow construction of novel force-fields given experimental data, but instead, it is meant for the case when an approximate force-field is known. The most obvious example

Received: April 14, 2014
Published: June 19, 2014

of this approach is the use of harmonic constraints, which has seen success in NMR structure refinement.[22−24] These constraints may alter the dynamics of the portion of the system biased.[7] Islam and co-workers[6] and later Roux and co-workers have introduced a method that reduces the undesired changes to system dynamics when a harmonic constraint is introduced using a method called restrained-ensemble or replica-averaged structural restraint ensembles.[6,9] This method has been used successfully in protein simulations to immediately improve a simulation given experimental NMR data.[11] Compared with the previous methods, which require iterative changes to the underlying force-field, this approach uses a number of replicas and thus reduces the wall-clock time of a simulation.

Pitera and Chodera used a maximum entropy argument to derive a form of bias to match an experimental value that changes the statistical ensemble the least.[7] Whereas harmonic biases and restrained-ensemble methods add terms that are quadratic in the quantity being biased, Pitera and Chodera showed that a term that is linear in the biased quantity creates a biased model that is closest to the original using the distance measure of relative entropy.[9] This linear bias only requires a single-replica and has the same properties as an infinite-replica restrained-ensemble.[7]

Harmonic biases, restrained-ensembles, and linear biases all require some choice of a coupling constant that sets the strength of the bias. From a practical standpoint, choosing a harmonic bias or restrained-ensemble coupling constant is a forgiving process because the system will move toward the desired biased value with any coupling constant choice. The linear biases introduced in Pitera and Chodera have coupling constants that determine both the value of quantity biased and how strong the coupling is, so it is critical to choose the correct value. The benefit is that there always is a correct value, whereas for harmonic methods it is not always true that there is a harmonic constant leading to exact match between the simulation and experimental data.

In this work, we describe EDS that has the benefit of both the linear bias method for being minimal and the benefit of harmonic methods that choosing adjustable parameters cannot take a system farther away from the desired biased value. The method requires only a single simulation and can bias multiple correlated or uncorrelated collective variables simultaneously. Once the EDS has converged, the simulation will be driven by the same potential shown by Pitera and Chodera to be the minimal bias.[7] An implementation of the algorithm is available by request from the authors. In the systems studied, the method converges rapidly and requires a small amount of additional simulation time for the convergence. In future work, we will test this approach for increasingly complex systems.

The paper is outlined as follows: in the methods section, we describe the simulation details of the example systems and the algorithm implementation. In Theoretical Results, we show a derivation of the method. In Simulation Results, we demonstrate our algorithm on three model systems and on a more challenging electrolyte simulation where previous force-matching methods have had limited success matching results from polarizable and ab initio methods.

■ THEORETICAL RESULTS

Assume there exists an ensemble of particles described by the probability distribution $P(r) \propto e^{-\beta U(r)}$, where $\beta = 1/k_BT$ and $U(r)$ is the potential energy of the N-particle positions, written as r. One-dimensional notation is used here without loss of generality. The expected value of a function of only the positions, $f(r)$, can be calculated as $\langle f(r)\rangle = \int \mathrm{d}r\, f(r)\, e^{-\beta U(r)}/Z$, where Z is the normalization constant of $P(r)$. The functional derivative of the expected value of $f(r)$ with respect to the potential energy can be written as

$$\begin{aligned}
\frac{\delta}{\delta U(r')} f(r) &= \frac{\delta}{\delta U(r')} \frac{\int \mathrm{d}r f(r) \mathrm{e}^{-\beta U(r)}}{Z} \\
\frac{\delta\langle f(r)\rangle}{\delta U(r')} &= \frac{1}{Z}\int \mathrm{d}r - \beta \frac{\delta U(r)}{\delta U(r')} f(r) \mathrm{e}^{-\beta U(r)} \\
&\quad - \frac{1}{Z^2}\left(\int \mathrm{d}r - \beta \frac{\delta U(r)}{\delta U(r')} \mathrm{e}^{-\beta U(r)}\right) \\
&\quad \left(\int \mathrm{d}r f(r) \mathrm{e}^{-\beta U(r)}\right) \\
\frac{\delta\langle f(r)\rangle}{\delta U(r')} &= -\beta \left\langle \frac{\delta U(r)}{\delta U(r')} f(r) \right\rangle + \beta \left\langle \frac{\delta U(r)}{\delta U(r')} \right\rangle \langle f(r)\rangle \\
\frac{\delta\langle f(r)\rangle}{\delta U(r')} &= -\beta\, \mathrm{Cov}\left(\frac{\delta U(r)}{\delta U(r')}, f(r)\right)
\end{aligned} \tag{1}$$

Eq 1 provides a description of how $\langle f(r)\rangle$ changes as a single point on the potential energy function changes. If $f(r)$ is an observable, one may have a value $\bar{f}$ either from experiments or another simulation. It is possible that $\langle f(r)\rangle \neq \bar{f}$ because the potential energy function is approximate or there are missing details from the model. The goal of EDS is to modify the potential energy function so that $\langle f(r)\rangle = \bar{f}$. Changing the potential energy function point-by-point can lead to degenerate solutions or there may exist no pairwise representable potential energy function where the observable matches the experimental value. Instead, we will assume a particular form of the potential energy and change only one variable. As shown in Pitera and Chodera;[7] Roux and Weare;[16] and Sadowsky,[25] the change to a potential, ΔU, which constrains $\langle f(r)\rangle = \bar{f}$ and has the property that the probability distribution generated $U(r) + \Delta U(r)$ is as close to the original probability distribution with respect to relative entropy is $\Delta U(r) = (\alpha/\bar{f})\cdot f(r)$. α is a number in energy units and will be called a coupling constant. The fact that this form of $\Delta U(r)$ moves the probability distribution as little as possible from $U(r)$ has led to it being called a minimal biasing potential.[7] Indeed, given that there are many possible ways to modify the potential energy, the way that changes the probability distribution least seems most satisfactory. This extra term may also be thought of as a modification of the potential energy that adds the smallest amount of new information while still matching experimental data.

Assuming it exists, the coupling constant α is defined such that $f(r) = \bar{f}$ under the potential $U'(r, \alpha) = U(r) + (\alpha/\bar{f})\cdot f(r)$. This form of the potential energy has a single parameter, the coupling constant, which can be set to match any desired $\bar{f}$. This changes the problem from changing the potential energy point-by-point to a single variable optimization procedure. We may combine the definition of the coupling constant with eq 1:

$$\frac{\partial \langle f(r) \rangle}{\partial \alpha} = -\beta \ \mathrm{Cov}\left(\frac{\partial U'(r, \alpha)}{\partial \alpha}, f(r) \right) = -\frac{\beta}{\bar{f}} \ \mathrm{Var}(f(r))$$

$$\frac{\partial (\langle f(r) \rangle - \bar{f}\ )^2}{\partial \alpha} = -\frac{2\beta}{\bar{f}} \langle f(r) - \bar{f}\ \rangle \mathrm{Var}(f(r) \qquad (2)$$

Eq 2 above provides a gradient with which to find α and a multidimensional optimization procedure to find α for an arbitrary number of collective variables. This idea of finding α via a gradient-based optimization procedure was also hypothesized in Pitera and Chodera.[7] The theory generalizes to multiple dimensions of observables as

$$U' = U(r) + \sum_i \alpha_i \frac{f_i(r)}{\bar{f}_i} \qquad (3)$$

Pitera and Chodera showed there exists solutions for coupling constants in the multidimensional case.[7]

METHOD IMPLEMENTATION

The EDS method calculates the coupling constants and biases in a simulation and is implemented in the Colvars package.[26] The source-code may obtained from the authors along with installation instructions for LAMMPS[27] and NAMD.[28] The implementation biases a simulation linearly in the chosen collective variables. The force at each time step is calculated according to

$$F_i = \frac{\alpha}{\bar{f}} \frac{\partial f(r)}{\partial r_i} \qquad (4)$$

where $\partial f(r)/\partial r_i$ is the spatial derivative of particle i of the collective variable $f(r)$, $\bar{f}$ is the set point of the collective variable $f(r)$, and α is the coupling constant, which is unique for each collective variable. The coupling constant is updated according to

$$\alpha_{\tau+1} = \alpha_\tau - \eta_\tau g_\tau$$

$$g_\tau = -2\beta \left(\frac{\langle f(r) \rangle_\tau}{\bar{f}} - 1 \right) (\langle f(r)^2 \rangle_\tau - \langle f(r) \rangle_\tau^2) \qquad (5)$$

where τ is the iteration index, g_τ is the gradient at the τth iteration, T is temperature, and $\langle \cdot \rangle_\tau$ denotes an average value over the τth iteration. The τ subscript is important to note because each iteration samples from a different statistical ensemble, since the potential at τ is different than the potential at $\tau + 1$. η_τ is the learning rate and is calculated according to

$$\eta_\tau = \frac{A}{\sqrt{\sum_1^\tau g_i^2}} \qquad (6)$$

For the multidimensional case, each collective variable is treated independently and is updated at iteration τ with probability $1/(M - 1)$, where M is the number of collective variables. This update procedure as described is called per-coordinate adaptive online stochastic gradient descent with an infinite horizon, and analysis of it may be found in McMahan and Streeter.[29,30] Briefly, it converges in the same order as online gradient descent in the single dimensional case, assuming a few conditions. The first is that $(\langle f_i(r) \rangle - \bar{f}_i)^2$ is convex in α_j for all i,j in the M collective variables. Pitera and Chodera showed that this is true provided there is no correlation between the collective variables.[7] Practically, this has not been seen to be an issue and the systems studied in this work have highly correlated collective variables. The second condition is that $(\langle f_i(r) \rangle - \bar{f}_i)^2$ is Lipschitz-continuous with respect to α_j for all i,j. The Lipshitz-continuity for $(\langle f_i(r) \rangle - \hat{f}_i)^2$ with respect to α_j is trivial since a bounded change in α_j leads to a bounded change in the probability distribution and thus a bounded change in $\langle f_i(r) \rangle$.

For numerical stability, the collective variable variances and means are calculated using the Welford, West, and Hanson one-pass method, as recommended in Chan et al.[31] The iteration period is split equally between an equilibration period that is half the update period, during which the coupling constant is linearly ramped to its new value, and a production period during which statistics are calculated for the next update. Increasing or decreasing a coupling constant is equivalent to adding or removing energy from the simulations. Constant NVE integrated ensembles are not compatible with the method due the addition and removal of energy.

Two parameters must be chosen to use this method. The first is A, the range of the coupling constant in units of energy. This is largest magnitude of the coupling constant which one expects. If this parameter is too low, the simulation will not converge. If it is too high the simulation will waste time exploring values of α that are too large. A value of 3 k_BT has worked well in the systems presented as a first choice. If the method does not converge and is reaching coupling constants where $|\alpha| > A$, it should be increased. The second parameter is the duration of an iteration, N, in number of simulation steps. This determines how long the system requires to equilibrate after a change in coupling constant ($N/2$), how long statistics are collected for an iteration ($N/2$), and how quickly energy is added to the system (at most, $A/2N$). Until the coupling constant has converged, the method as described is an optimization procedure and not an integration of a particular statistical ensemble. It is important that each step should be uncorrelated from the last so that iterations are independent. Therefore, N should be at least twice the autocorrelation time of the collective variable. The system should also be able to dissipate energy as fast as $N/2$, which can be done by adjusting thermostat parameters. Practically we have found that N can be significantly shorter than the autocorrelation time of the collective variables being biased and still converge correctly.

After the coupling constant has converged, the simulation may be stopped and the median value should be used for a production simulation (without updating of the coupling constant). Alternatively, the simulation may be continued if the coupling constant is not changing quickly enough to introduce artifacts. The frames before the coupling constant converges should not be used for analysis unless reweighting of the trajectory is done.

COORDINATION NUMBER DEFINITION

The coordination number collective variable is defined to be

$$\langle N(r_0) \rangle = \rho \int_0^R \mathrm{d}r[1 - u(r - r_0)] 4\pi r^2 g(r) \qquad (7)$$

where ρ is the number density, $g(r)$ is the radial distribution function, $u(r)$ is the unit step function, R is a cutoff and r_0 is distance of the coordinating shell. The EDS method requires spatial derivatives and the unit step function has ill-defined derivatives. It may be replaced with the following approximation:[32]

$$1 - u(r - r_0) \approx \begin{cases} \dfrac{1 - \left(\frac{r - r_0}{w}\right)^6}{1 - \left(\frac{r - r_0}{w}\right)^{12}}, & r > r_0 \\ 1 \end{cases} \tag{8}$$

where w controls how wide the region with significant derivatives is. It is important to note that this definition creates a dependence on the simulation cutoff distance, R, in the coordination number. The ith moment of the radial distribution function can be calculated as

$$\langle r^i N(r_0)\rangle = \rho \int_0^R \mathrm{d}r[1 - u(r - r_0)]4\pi r^{2+i} g(r) \tag{9}$$

Where again the mollified unit step function may be inserted if derivatives are required.

SIMULATION METHODS

All simulations were conducted using the LAMMPS simulation package with the Colvars library.[26,27] The 1D harmonic oscillator system was in reduced units and simulated at a temperature of 1 with a Langevin thermostat with a 0.5 time-constant and a time step of 0.005.[33] The coupling constant range was set to be 10 and the iteration duration to 500 steps.

The bead−spring polymer system consisted of 16 Lennard-Jones beads joined by harmonic bonds again in reduced units. The Lennard-Jones potential used a cutoff of 3σ and $\sigma = \varepsilon = 1$. The harmonic bonds had a spring constant of $5\sigma^*$, a bond length of $2\sigma^*$, and a mass of 1 m. The time step was 0.005 $(\epsilon^*/m\sigma^{*2})^{1/2}$ and the center-of-mass motion was removed every 25 steps. The system was thermostated to a temperature of 1.5 ϵ^* with a Langevin thermostat and a time-constant of $0.25(\epsilon^*/m\sigma^{*2})^{1/2}$.[33] The angular potential was harmonic with a force constant of 2.5 ϵ^*/σ^* and an equilibrium angle of 155 degrees. The coupling constant range was set to $3T$ and the iteration duration to 250 steps. The simulation was run for 25 million steps.

The Lennard-Jones simulation contained 864 particles in a unit-size cube with $\sigma = 0.9$, $\epsilon = 1.2$ in the reference simulation and $\sigma = 0.9$, $\epsilon = 0.4$ in the biased simulations. The system was simulated with a Langevin thermostat at a reduced temperature of 1.5 ϵ^* and a time constant of 0.1 $(\epsilon^*/m\sigma^{*2})^{1/2}$. The potential was shifted/truncated at $2.5\sigma^*$ and the time step was $0.001(\epsilon^*/m\sigma^{*2})^{1/2}$. The system was run for 250 000 steps with an iteration duration of 50 steps for the EDS and a coupling constant range of 50 ϵ^*.

The electrolyte solution simulation is described in detail in Jorn et al.[34] Briefly, 201 ethylene carbonate, 15 lithium ions, and 15 hexafluorophosphate ions were simulated in a 3 nm cube under a Nosé−Hoover thermostat at 298 K with a time constant of 50 fs.[35] The ethylene-carbonate force-field came from Masia et al.[36] and the remaining terms were developed in Jorn et al.[34] A cutoff of 0.9 and 1.25 nm was used for Lennard-Jones and Coulomb pair-potentials, respectively. The system was simulated for 20 ps before EDS was initiated. The time step was 1 fs. The bias was equilibrated for 500 ps. After equilibration, statistics were gathered for 1 ns.

SIMULATION RESULTS

The first system biased using an experiment directed simulation (EDS) is a one-dimensional harmonic oscillator. The equilibrium position of the harmonic oscillator is 1 without bias. The method is applied to move the equilibrium position to 4. The resulting position distribution is shown in Figure 1a.

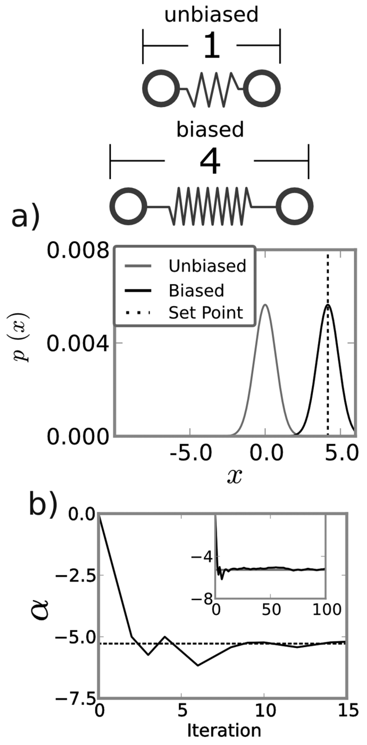


Figure 1. Experiment directed simulation of a one-dimensional harmonic oscillator. Panel a shows the probability distribution of position for the harmonic oscillator in the biased and unbiased simulations. The set point is 2.5 in the biased simulation. The biased distribution maintains the variance of the unbiased distribution and matches the set point. The convergence of the coupling constant is shown in Panel b. The dashed line is the correct coupling constant and the algorithm finds it within a single iteration. The inset shows the stability over 100 iterations of the method.

The position distribution maintains the same width and only equilibrium position is moved, as is expected for such minimally biasing potentials.[7] The speed at which the coupling constant is found is shown in Figure 1b. After only a few iterations, the algorithm has converged within a few percent of the correct coupling constant. The inset shows the stability of the algorithm over long periods of time. The algorithm converges quickly and correctly for a harmonic oscillator. Furthermore, the method provides a length distribution as similar as possible to the original distribution.

The second system is a 16-bead polymer with harmonic bonds and Lennard-Jones interactions between beads. The radius of gyration is biased in this example. This example is a bit similar to a protein simulation; it has long a time-scale and a frustrated topology. The mean radius of gyration is $3.101\sigma^*$ in the unbiased system, where σ^* are reduced length units. A “reference” system was constructed which has an additional angular potential constraining the angle between beads to be 155° (see Figure 2a). This reference system has a mean radius of gyration of $6.001\sigma^*$. Providing only the reference mean radius of gyration, the polymer simulation was biased to match the new value. The resulting distribution is shown in Figure 2b. The “biased” line correctly matches the new radius of gyration value. The biased line also matches the shape of the unbiased distribution, but it is scaled upward in order to match the new

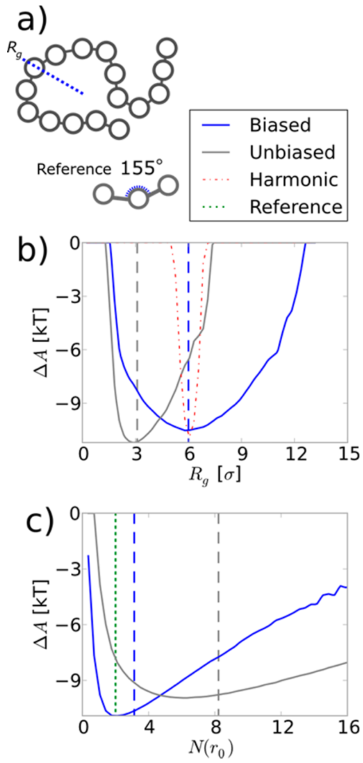


Figure 2. Radius of gyration potential mean force (PMF) for a 16-bead Lennard-Jones bead−spring polymer. The blue line in panel b shows a PMF where the radius of gyration was biased using an experiment directed simulation to match a set point of 6.0σ. The blue line maintains much of same curvature and symmetry as the unbiased PMF. The red line shows a PMF where the radius of gyration was restrained using a harmonic bias to a set point of 6.0σ. It is distorted and symmetric. Panel C shows the PMF for coordination number, which was not biased. The green line is the reference mean. The biased PMF moves closer to the reference mean while maintain the shape of the unbiased distribution (gray).

value. For comparison, a system with a harmonic constraint with a force constant of 50 is shown in red. This distribution indeed now matches the new radius of gyration, but its fluctuations are significantly damped and it lacks the asymmetry seen in the unbiased system. Furthermore, the biased polymer system matches to four significant Figures (6.001σ*) and the harmonically constrained system matches to only one (5.928σ*).

After biasing the radius of gyration to more closely match the reference polymer system, other properties may be compared between the reference and biased polymer systems. The coordination number between beads, with a cutoff of 1.41σ*, is 8.2 in the unbiased system and 2.0 in the reference system as seen in Figure 2c. After biasing, the coordination number becomes 3.1, which is much closer to the reference value. The method works well in the polymer system, which is a significantly more challenging system than the harmonic oscillator.

Often in simulations, one wishes to match the radial distribution function, $g(r)$, which can come from more accurate simulation methods or scattering experiments. The radial distribution function is a transformation of a probability distribution and can only be matched using EDS by finding a coupling constant at each pairwise distance, which is not possible in practice. However, we may use the integral transformation in eq 6 to go from the radial distribution function to the coordination number, which can be calculated instantaneously in simulations and is an average value, not a probability distribution. This reduces the number of coupling constants to one at each pairwise distance to one for each moment, where the moments are similar to a series expansion approximation of the radial distribution function. Matching the coordination number only corresponds to matching the area under the first peak of the radial distribution and is akin to a zeroth moment matching. However, there is no guarantee that the shape of the first peak of the biased simulation must match a reference radial distribution function if they have the same areas. Thus, one may also match higher moments of the radial distribution function to ensure that the peak means, widths, and skews match. Calculating the higher moments in simulations is computationally done by multiplying the coordination number between two particles, which is between 0 and 1, by their pairwise distance raised to the power of the moment being calculated. This is the left-hand side of eq 9.

This method of matching multiple moments of a radial distribution function is shown in Figure 3 for a Lennard-Jones simulation. A “reference” radial distribution function was calculated by simulating 864 Lennard-Jones particles with Langevin dynamics at a density of 0.864σ^{*-3} and a temperature of 1.5ϵ* with parameters σ = 0.9, ϵ = 1.2. That radial distribution function is shown as the dashed blue line in Figure 3a. A second simulation with parameters σ = 0.9, ϵ = 0.4 was conducted. That is the “unbiased” radial distribution function, shown in red. Next, the zeroth (area), first, second, and third moments were matched using eq 9 and EDS. The parameters for eq 8, the unit-step function approximation in eq 9, are w = 0.3σ*, r_0 = 0.75σ*, R = 1.05σ*. As seen in Figure 3a, increasing the number of matched moments increases the fit until the simulation conducted with a different parameter set closely matches the reference radial distribution function. The convergence of these properties is seen in Figure 3b for the case of matching up to the third moment, which is the black line. The red lines are the unbiased moments and the blue lines are the reference values. Both the unbiased and EDS lines are smoothed by taking a 2.5 $(\epsilon^*/m\sigma^{*2})^{1/2}$ time running average. The unsmoothed plot for the zeroth moment is shown in Figure 3c. The convergence is slower in Figure 3b than the other examples shown because of the significant fluctuations in the moments. This example also demonstrates that highly coupled biases, since the coupling constant of the moments changes as moments are added, are successfully treated in the multidimensional case. Thus, EDS may be used if complete radial distribution functions are available from experiments.

Lithium ion batteries are one of the most prevalent batteries in commercial applications due to their outstanding power density. Constantly improving lithium ion batteries requires a strong understanding of the charge cycling processes, especially the solid electrolyte interphase (SEI) where lithium ions are trapped in a surface film.[37,38] The structure and composition of the SEI is an open research problem and has been the topic of past modeling work.[39−42] Jorn et al.[34] conducted the first SEI-electrolyte simulation, whereas past models have only considered the SEI alone. The model Jorn et al. used was parametrized from ab initio calculations and built from other published models.[36,43] The model is a classical molecular dynamics potential, which makes it efficient compared with polarizable and ab initio molecular dynamics methods.

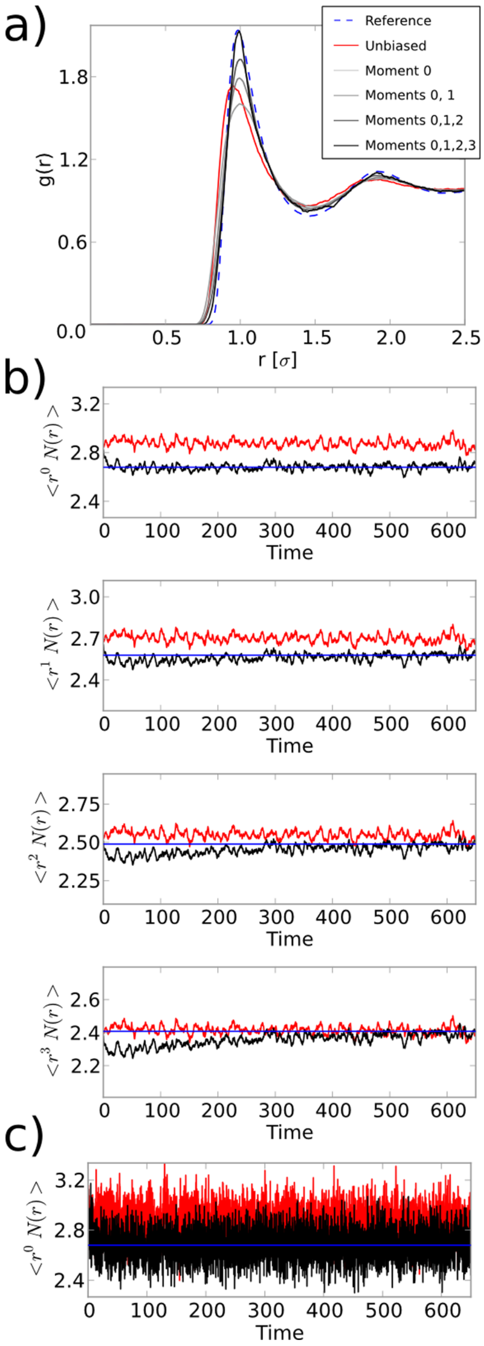


Figure 3. Radial distribution function for a Lennard-Jones simulation under different amounts of bias. The reference $g(r)$ was simulated with $\sigma = 0.9$, $\epsilon = 1.2$ for its Lennard-Jones parameters and the unbiased $g(r)$ was simulated with $\sigma = 0.9$, $\epsilon = 0.4$. The other lines were simulated with $\sigma = 0.9$, $\epsilon = 0.4$ and experiment directed simulation on the moments of $g(r)$ as indicated in the legend. Matching up to the 3rd moment provides good agreement with the reference $g(r)$. Panel b shows the convergence of the biased moments over time with a 2.5 $(\epsilon^*/m\sigma^{*2})^{1/2}$ time running average. Panel c shows the unsmoothed zeroth moment and bias over time.

One outstanding challenge of the model described in Jorn et al. is that it does not match certain structural data from Borodin and Smith,[43] whose more sophisticated polarizable model match experiments well. One approach to improving the Jorn et al. model is to perform a traditional cycle of changing parameters, running simulations, and comparing the results. Though this may lead to transferable parameters, practically this is very difficult with the number of parameters and amount of experimental data. Instead, one may use the method described here to bias a simulation to match experimental data.

Following Jorn et al., an electrolyte model system was constructed consisting of ethylene carbonate, lithium, and hexafluorophosphate. The ions were at a concentration of 1 M. The properties chosen to improve were the lithium ethylene carbonate coordination number and the lithium fluorine coordination number. The lithium ethylene coordination number is between the lithium and the carbonyl oxygen (see Figure 4d). The Jorn et al. model has a lithium ethylene carbonate coordination number of 3.52 and a lithium fluorine coordination number of 0.68. The Borodin and Smith model has a lithium ethylene carbonate coordination number of 3.1 (first peak minimum: 2.725 Å) and a lithium fluorine coordination number of 0.85 (first peak minimum: 2.5 Å). To match the Borodin and Smith model, EDS was applied using coordination number with the unit step approximation in eq 8. In order to better match the $g(r)$ width, which was also incorrect for the lithium ethylene carbonate system, the second moment of the radial distribution function was also biased for the lithium ethylene carbonate coordination number.

The parameter r_0 was set to 0.985 and 0.799 Å for the lithium ethylene carbonate and lithium fluorine coordination numbers, respectively. The width was chosen to be 1 Å for both cases and the cutoff, R in eq 8, was set to 8 Å. The iteration duration was 50 steps. The coupling range, A, was set to 100 kcal for both the lithium ethylene carbonate coordination number and second moment. The coupling range was set to 50 kcal for the lithium fluorine coordination number. These large coupling ranges are necessary because coordination number affects many particles and thus its energy scale is large.

The EDS converged quickly as shown in Figure 4b. The top panel in Figure 4b is lithium ethylene carbonate and converged to the correct value within 1 ps. The middle panel is the second moment, which converged also within 1 ps. The lithium fluorine coordination number converged slightly slower, at approximately 5 ps. The insets show the values are stable over the duration of the simulation. The unbiased coordination numbers are shown in gray and are different from what was described in the text due to the application of eq 9, which is an approximation to eq 7. Figure 4a shows the radial distribution functions corresponding to the two biased coordination numbers and the lithium phosphorus radial distribution function, which was not biased. The lithium ethylene carbonate radial distribution function is significantly improved and the first peak width is indeed closer to the Borodin and Smith model. The lithium fluorine is not as improved, due to the lack of a higher moment biasing, but is nevertheless better. Finally, the lithium phosphorus radial distribution fucntion, which was not biased, also improves due the improvement of the lithium fluorine coordination number. The coordination number of lithium phosphorus changed from 0.256 to 0.587. The Borodin and Smith model is 0.599. Also, while the present paper is devoted to the issue of statistical properties and not focused on the issue of dynamics, it is interesting to note that the self-diffusion coefficient of lithium also improved with the changes to coordination number as seen in Figure 4c. As was seen in the polymer system, unbiased properties also seem to improve as the biases improve the target properties.

Biasing of three properties in the Jorn et al. model brought the system properties much closer to the Borodin and Smith

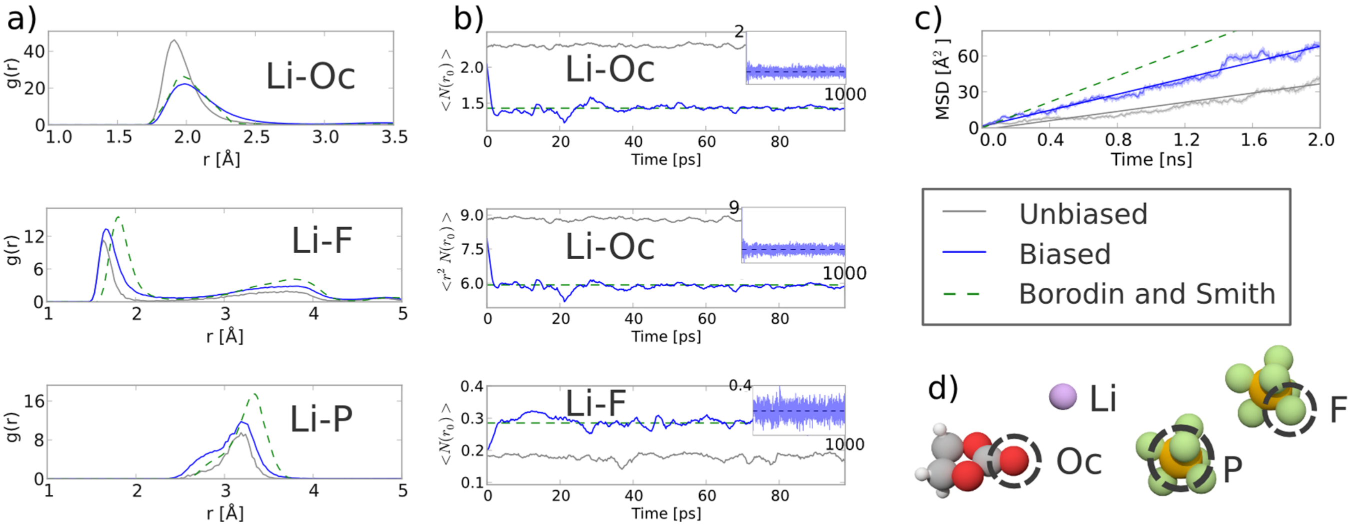


Figure 4. Results of experiment directed simulation of an electrolyte simulation. Panel a contains the radial distribution functions. The inset text indicates which atoms make up the radial distribution function and corresponds to the circles in panel d. Panel b shows the convergence of the coordination numbers that were biased. The insets are the coordination numbers over 1 ns. Panel c shows the mean-squared displacement of lithium ions over the simulation. The solid lines are least-squares fit, and their slope is proportional to the self-diffusion coefficient of lithium. Panel d shows the component atoms of the radial distribution functions in panels a and b. Red colored spheres are oxygen atoms, gray are carbon, white are hydrogen, purple are lithium, green are fluorine, and orange are phosphorus.

model without the use of polarizability or a reparameterization. The amount of extra simulation time necessary to converge EDS was less than 5 ps, which is insignificant relative to the simulation time of the system. The biased properties all match their set points over the simulation. The biasing of the three properties also improved the radial distribution functions for the different species in the simulation as well as the self-diffusion coefficient of lithium. The improvement in unbiased properties is not as significant in the biased properties, which is expected since EDS finds the smallest changes to a model to match new data.

■ CONCLUSIONS

A new biasing method was introduced that allows efficient and minimal biasing of simulations to match experimental data. The method, experiment directed simulation (EDS), was derived by combining previously reported minimal biasing potential energy terms with a functional derivative of the expected value of a collective variable and an online stochastic gradient descent optimization algorithm. Past work in minimally biased simulations required the use of replica-exchange, whereas this method requires only a single replica even to bias multiple collective variables simultaneously. EDS was applied to a one-dimensional harmonic oscillator, a bead−spring polymer, a Lennard-Jones simulation and a more challenging three-component electrolyte molecular dynamics simulation. Biasing three separate collective variables in the electrolyte simulation required less than 5 ps of additional equilibration simulation time. Biasing the structural properties in the electrolyte simulation improved the self-diffusion coefficient of lithium, showing that biasing structural properties can improve dynamical properties. The method can also be used to modify radial distribution functions to match experimental data. EDS has been implemented in the Colvars package, which is integrated into the LAMMPS and NAMD simulation engines. Exploring the application of EDS to bias explicitly dynamical properties is the next logical step in the research, albeit a challenging one, and will be the subject of future work.

■ AUTHOR INFORMATION

Corresponding Author

*Email: gavoth@uchicago.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was supported by the Office of Naval Research (ONR Award N00014-13-1-0058). A.D.W. was also supported in part by a Yen Fellowship from the University of Chicago Institute for Biophysical Dynamics.

■ REFERENCES

(1) Wang, L. P.; Chen, J. H.; Van Voorhis, T. Systematic parametrization of polarizable force fields from quantum chemistry data. *J. Chem. Theory Comput.* **2013**, *9* (1), 452−460.

(2) Pastor, R. W.; MacKerell, A. D. Development of the CHARMM force field for lipids. *J. Phys. Chem. lett.* **2011**, *2* (13), 1526−1532.

(3) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712−725.

(4) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005**, *433* (7022), 128−132.

(5) Best, R. B.; Vendruscolo, M. Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc.* **2004**, *126* (26), 8090−8091.

(6) Islam, S. M.; Stein, R. A.; Mchaourab, H. S.; Roux, B. Structural refinement from restrained-ensemble simulations based on EPR/DEER data: Application to T4 lysozyme. *J. Phys. Chem. B* **2013**, *117* (17), 4740−4754.

(7) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8* (10), 3445−3451.

(8) Rozycki, B.; Kim, Y. C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19* (1), 109−116.

(9) Roux, B.; Weare, J., On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138* (8).

(10) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10* (2), e1003406.

(11) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, 094112.

(12) De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M. Characterization of the lnterdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. *Biochemistry* **2013**, *52* (37), 6480–6486.

(13) Henderson, D.; Grundke, E. W. Direct correlation-function—Hard-sphere fluid. *J. Chem. Phys.* **1975**, *63* (2), 601–607.

(14) Johnson, M. D.; March, N. H. Ion–ion oscillatory potentials in liquid metals. *Proc. R. Soc. London, A* **1964**, *282* (1388), 283–302.

(15) Bolhuis, P. G.; Louis, A. A. How to derive and parameterize effective potentials in colloid-polymer mixtures. *Macromolecules* **2002**, *35* (5), 1860–1869.

(16) Rajagopalan, R.; Rao, K. S. Interaction forces in charged colloids: Inversion of static structure factors. *Phys. Rev. E* **1997**, *55* (4), 4423–4432.

(17) Wang, Q. F.; Keffer, D. J.; Nicholson, D. M.; Thomas, J. B. Use of the Ornstein–Zernike Percus–Yevick equation to extract interaction potentials from pair correlation functions. *Phys. Rev. E* **2010**, *81* (6), 061204.

(18) Beglov, D.; Roux, B. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem. B* **1997**, *101* (39), 7821–7826.

(19) Chandler, D.; Andersen, H. C. Optimized cluster expansions for classical fluids 0.2. Theory of molecular liquids. *J. Chem. Phys.* **1972**, *57* (5), 1930–1937.

(20) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505.

(21) Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117* (34), 9956–9972.

(22) Brunger, A. T.; Nilges, M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Q. Rev. Biophys.* **1993**, *26* (1), 49–125.

(23) Dolenc, J.; Missimer, J. H.; Steinmetz, M. O.; van Gunsteren, W. F. Methods of NMR structure refinement: Molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. *J. Biomol. NMR* **2010**, *47* (3), 221–235.

(24) Iwahara, J.; Schwieters, C. D.; Clore, G. M. Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **2004**, *126* (18), 5879–5896.

(25) Sadowsky, J. S. On the optimality and stability of exponential twisting in Monte-Carlo estimation. *IEEE Trans. Inf. Theory* **1993**, *39* (1), 119–128.

(26) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111* (22–23), 3345–3362.

(27) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.

(28) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(29) Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.

(30) McMahan, H. B.; Streeter, M. Adaptive bound optimization for online convex optimization. In *Proceedings of the 23rd Annual Conference on Learning Theory, Haifa, Isreal, June 27–29, 2010*; Tauman, K., Mohri, M.; Omnipress: Madison, WI, 2010.

(31) Chan, T. F.; Golub, G. H.; Leveque, R. J. Algorithms for computing the sample variance—Analysis and recommendations. *Am. Stat.* **1983**, *37* (3), 242–247.

(32) Iannuzzi, M.; Laio, A.; Parrinello, M., Efficient exploration of reactive potential energy surfaces using Car–Parrinello molecular dynamics. *Phys. Rev. Lett.* **2003**, *90* (23).

(33) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **1978**, *17* (3), 1302–1322.

(34) Jorn, R.; Kumar, R.; Abraham, D. P.; Voth, G. A. Atomistic modeling of the electrode–electrolyte interface in Li-ion energy storage systems: Electrolyte structuring. *J. Phys. Chem. C* **2013**, *117* (8), 3747–3761.

(35) Hoover, W. G. Canonical dynamics—Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.

(36) Masia, M.; Probst, M.; Rey, R. Ethylene carbonate, A theoretical study of structural and vibrational properties in gas and liquid phases. *J. Chem. Phys. B* **2004**, *108* (6), 2016–2027.

(37) Verma, P.; Maire, P.; Novak, P. A review of the features and analyses of the solid electrolyte interphase in Li-ion batteries. *Electrochim. Acta* **2010**, *55* (22), 6332–6341.

(38) Xu, K.; von Cresce, A. Interfacing electrolytes with electrodes in Li ion batteries. *J. Mater. Chem.* **2011**, *21* (27), 9849–9864.

(39) Winter, M. The solid electrolyte interphase—The most important and the least understood solid electrolyte in rechargeable Li batteries. *Z. Phys. Chem.* **2009**, *223* (10–11), 1395–1406.

(40) Bedrov, D.; Smith, G. D.; van Duin, A. C. T. Reactions of singly-reduced ethylene carbonate in lithium battery electrolytes: A molecular dynamics simulation study using the ReaxFF. *J. Phys. Chem. A* **2012**, *116* (11), 2978–2985.

(41) Wang, Y. X.; Nakamura, S.; Ue, M.; Balbuena, P. B. Theoretical studies to understand surface chemistry on carbon anodes for lithium ion batteries: Reduction mechanisms of ethylene carbonate. *J. Am. Chem. Soc.* **2001**, *123* (47), 11708–11718.

(42) Tasaki, K. Solvent decompositions and physical properties of decomposition compounds in Li-ion battery electrolytes studied by DFT calculations and molecular dynamics simulations. *J. Phys. Chem. B* **2005**, *109* (7), 2920–2933.

(43) Borodin, O.; Smith, G. D. Quantum chemistry and molecular dynamics simulation study of dimethyl carbonate: Ethylene carbonate electrolytes doped with $LiPF_6$. *J. Phys. Chem. B* **2009**, *113* (6), 1763–1776.