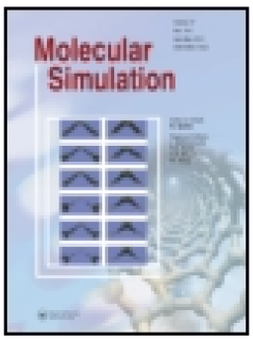

Molecular Simulation

ISSN: 0892-7022 (Print) 1029-0435 (Online) Journal homepage: https://www.tandfonline.com/loi/gmos20

Recent advances in maximum entropy biasing techniques for molecular dynamics

D. B. Amirkulova & A. D. White

To cite this article: D. B. Amirkulova & A. D. White (2019): Recent advances in maximum entropy biasing techniques for molecular dynamics, Molecular Simulation, DOI: 10.1080/08927022.2019.1608988

To link to this article: https://doi.org/10.1080/08927022.2019.1608988

Published online: 02 May 2019.

MOLECULAR SIMULATION
https://doi.org/10.1080/08927022.2019.1608988

Taylor & Francis
Taylor & Francis Group

Recent advances in maximum entropy biasing techniques for molecular dynamics

D. B. Amirkulova and A. D. White

Department of Chemical Engineering, University of Rochester, Rochester, NY, USA

ABSTRACT

This review describes recent advances by the authors and others on the topic of incorporating experimental data into molecular simulations through maximum entropy methods. Methods which incorporate experimental data improve accuracy in molecular simulation by minimally modifying the thermodynamic ensemble. This is especially important where force fields are approximate, such as when employing coarse-grain models, or where high accuracy is required, such as when attempting to mimic a multiscale self-assembly process. The authors review here the experiment directed simulation (EDS) and experiment directed metadynamics (EDM) methods that allow matching averages and distributions in simulations, respectively. Important system-specific considerations are discussed such as using enhanced sampling simultaneously, the role of pressure, treating uncertainty, and implementations of these methods. Recent examples of EDS and EDM are reviewed including applications to *ab initio* molecular dynamics of water, incorporating environmental fluctuations inside of a macromolecular protein complex, improving RNA force fields, and the combination of enhanced sampling with minimal biasing to model peptides

ARTICLE HISTORY
Received 22 January 2019
Accepted 9 April 2019

KEYWORDS
Molecular dynamics; maximum entropy

1. Introduction

A common task in molecular simulation is ensuring that observables of the simulation match experimentally measured values. For example, in simulations of protein structure [1] or liquids [2], quantitative agreement with experiments is the standard for assessing correctness of a model. When there is no quantitative agreement, changing the potential energy function or adding additional components to the simulation are possible ways to improve the fit. Making such changes can be an ambiguous and challenging process, especially if the potential energy function has multiple terms that can be modified. Minimal biasing techniques are a class of methods that modify a potential energy function to improve quantitative agreement with experimental values while minimizing the the change in the potential energy function. The definition of 'minimal' and the way the potential energy function is modified vary from method to method.

Recent reviews of minimal biasing methods can be found in [3–5]. Table 1 in [4] provides an overview of 28 minimal biasing methods, categorizing them by whether they maximize entropy, maximize parsimony, or use Bayesian inference. These three categories correspond broadly to the criteria used to ensure that the biasing function introduces a *minimal* change to the potential energy function. This review focuses on two techniques developed by the authors that are categorized as maximum entropy methods: experiment directed simulation [6] (EDS) and experiment directed metadynamics [7] (EDM). EDS is for matching ensemble average scalars and EDM is for matching free energy surfaces (probability distributions of observables).

EDS, like other minimal biasing methods, modifies a potential energy function to change ensemble averages of observables to match a specific value. These observables are typically equivalent to collective variables, but are intended to be only those that area experimentally verifiable quantities. What separates EDS from other methods is that it does not use replicas and can be used to construct a continuous NVE trajectory. For example, the Bayesian landscape tilting method of [8] relies on post-processing so that there cannot be a continuous NVE trajectory. Another example is the replica method of [9] which relies on replica-exchange of biasing forces and thus cannot result in a continuous trajectory. This does not mean EDS is 'better'; indeed these two methods appear to provide better sampling and scaling than EDS. However, the ability to compute an NVE trajectory allows dynamic observables like hydrogen-bonding lifetimes to be computed. The key results from EDS have been to improve thermodynamic observables and indirectly improve dynamic observables. For example, EDS was recently used to create a state-of-the-art DFT water model that gives near perfect agreement with X-ray scattering results, water diffusivity, and proton-hopping behavior by improving only the water oxygen–oxygen coordination number [10].

EDM is a maximum entropy method that matches ensemble probability distributions, or equivalently free energy surfaces, using prescribed functions. So far, EDM has proved most useful for matching radial distribution functions (RDFs), e.g. to then do coarse-grained modeling. EDM is less often used because it is rare to have experimental data that gives a probability distribution, other than a normal distribution (which is better treated with methods like in [11]). Other groups have since arrived at the same approach as EDM and typically it is now called 'targeted metadynamics' because it uses the method of

CONTACT A. D. White andrew.white@rochester.edu

metadynamics to arrive at a target free energy surface [12,13]. EDM is also equivalent to variationally enhanced sampling metadynamics [14], although there the target is a tool to improve sampling, and the biased simulation is not the goal. The name EDM was chosen by the authors because at the time there were new metadynamics methods that could target a collective variable domain [15]. Now the term 'targeted' seems to apply exclusively to target distributions, and EDM is probably best described as falling under the umbrella of targeted metadynamics methods.

Here, we review the maximum entropy derivations that underpin both EDS and EDM, describe some of the recent research benefiting from these methods, and discuss implementation details such as understanding the effect of EDS and EDM on the virial and accounting for uncertainty in the experimental data.

2. Theory

EDS minimally modifies an ensemble so that the ensemble average of some scalar matches a desired value, such as a value obtained from an experiment. The EDS bias is minimal because the resulting biased ensemble maximizes entropy [16,17]. Maximum entropy and minimum relative entropy are equivalent approaches to derive these minimal biasing equations [18]. We will use the minimum relative entropy approach because it has an intuitive interpretation as a distance metric. Figure 1 illustrates how the biased ensemble is one of many choices that matches our constraints, but is as close as possible to the unbiased ensemble as measured via relative entropy.

Consider an unbiased potential energy function, $U(\vec{r})$, which has a probability distribution of $P(\vec{r})$ under the NVT ensemble, following the Boltzmann distribution: $P(\vec{r}) \propto \exp(-\beta U(\vec{r}))$. $\vec{r}$ is the set of coordinate vectors of the N particles of the ensemble and $\beta = 1/kT$. k is the Boltzmann constant and T is absolute temperature. We would like to find a biased ensemble $P'(\vec{r})$ which is as similar as possible to the unbiased ensemble.

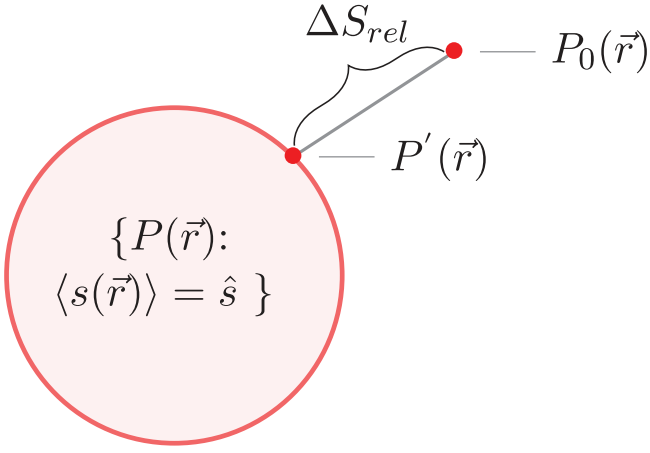

Figure 1. (Colour online) A schematic of the minimum relative entropy derivation. $P_0(\vec{r})$ is the unbiased probability distribution from the unbiased potential energy. We are finding a biased probability distribution, $P(\vec{r})$, that is consistent with Equation (2). There is a hypersurface of possible such probability distributions. With the condition that we minimize relative entropy, or 'distance' in this schematic, we find a unique point in the hypersurface $P'(\vec{r})$.

Similarity can be defined via the relative entropy:

$$\Delta S_{\text{rel}} = \int \mathrm{d}\vec{r} P'(\vec{r}) \ln \frac{P'(\vec{r})}{P(\vec{r})} \tag{1}$$

where $P'(\vec{r})$ is the biased ensemble probability distribution, and the integral is taken over all coordinates. Having a lower ΔS_{rel} means that the biased and unbiased ensembles are more similar. The biased ensemble should also have an observable average that matches a target value, which could be obtained from an experiment. This constraint is represented as

$$< s(\vec{r}) >= \int \mathrm{d}\vec{r} P'(\vec{r}) s(\vec{r}) = \hat{s} \tag{2}$$

where $s(\vec{r})$ is an instantaneous value for our observable (collective variable) which we are matching to $\hat{s}$, the desired scalar value. This is sometimes called a forward model and depends only on positions. We will relax this assumption below.

These equations represent (1) a constraint (to have our average simulation value match the target $\hat{s}$), and (2) a scalar to minimize (the relative entropy). When these conditions are present, we can find the optimal value via the method of Lagrange multipliers. The Lagrange funtion is

$$\mathcal{L}[\lambda, P(\vec{r}), P'(\vec{r})] = \Delta S_{\text{rel}} - \lambda\left(\int \mathrm{d}\vec{r} P'(\vec{r}) s(\vec{r}) - \hat{s}\right) \tag{3}$$

Equation (3) can be minimized by taking the functional derivative equal to zero: $\delta\mathcal{L}/\delta P'(\vec{r}) = 0$. Solving for $P'(\vec{r})$ gives this expression for the biased ensemble:

$$P'(\vec{r}) = \frac{1}{Z'} e^{-\beta(U(\vec{r}) + \lambda s(\vec{r}))} \tag{4}$$

where Z' is a normalization constant. This gives an expression for the biased potential energy as $U'(\vec{r}) = U(\vec{r}) + \lambda s(\vec{r})$. This result shows that if the bias is *linear* in the instantaneous observable, the ensemble is minimally biased. This derivation can be repeated for multiple dimensions [16] and for functions instead of scalars [7]. The general result is

$$U'(\vec{r}) = U(\vec{r}) + \sum_i \lambda_i s_i(\vec{r}) + \sum_j \mu_j[\nu(\vec{r})] \tag{5}$$

where i is the index of ensemble averages that are matched to a desired value and j is the index of free energy surfaces that are matched to desired functions. $\mu_j[\nu(\vec{r})]$ is a bias added that depends on $\nu(\vec{r})$, another instantaneous observable, and causes the biased ensemble to match a specific distribution in the free energy surfaces $F_t[\nu(\vec{r})]$. Specifically,

$$\int \mathrm{d}\vec{r} \delta(\nu(\vec{r}) - \nu') P'(\vec{r}) = q(\nu') \tag{6}$$

where δ is the Dirac delta function and $q(\nu')$ is a desired probability distribution for $\nu(\vec{r})$ (i.e. $q[\nu(\vec{r})] = -1/\beta \ln F_t[\nu(\vec{r})]$). $q(\nu)$ could be obtained for example from a scattering or FRET experiment [19].

The derivation above gives the form of the biased potential energy that is minimally biased, but it does not enable calculation of the Lagrange multipliers. That is the purpose of the EDS and EDM methods. EDS and EDM are time-dependent methods that change the potential energy of a simulation to

arrive at the bias in Equation (5). EDS is intended to be used in a two-step process: finding the Lagrange multiplier (adaptive) and then running a standard MD simulation using the modified force field given by Equation (5).

During the adaptive phase of EDS, the Lagrange multipliers are called coupling constants to distinguish from the time-independent Lagrange multipliers. These are indicated as α_τ where τ is a discrete step index. The Lagrange multipliers are set to be the average of α_τ. α_τ is defined as

$$\alpha_{\tau+1} = \alpha_\tau + \eta_\tau g_\tau \tag{7}$$

$$g_\tau = \frac{2\beta}{w}(< s >_\tau - \hat{s})\left(< s^2 >_\tau - < s >_\tau^2\right) \tag{8}$$

where w is an arbitrary constant used to ensure unit homogeneity, $< \cdot >_\tau$ is the ensemble average between step $\tau - 1$ and τ, and η_t is

$$\eta_\tau = \frac{A}{\sqrt{\sum_i^\tau g_i^2}} \tag{9}$$

where A is a user-defined constant that controls the size of the first step. The point of η_τ is to reduce the size of the steps over time. Note that because g_τ is in the sum, $|\alpha_0| = A$. Typically, the gradients should be clipped for stability and A is a natural upper/lower bound for clipping it. Thus, a user of this method must choose A and the time between updates. This update procedure is derived in [6] and is based on per-coordinate infinite horizon stochastic gradient descent [20]. Hocky et al. [21] recently found that Equation (8) can be modified to use covariance in multiple dimensions to improve convergence and that replacing Equation (7) with Levenberg–Marquardt optimization further improves convergence.

The EDM method finds the $\mu[v(\vec{r})]$ bias function in Equation (5). For compactness, we will now omit the dependence on $\vec{r}$. Unlike EDS, EDM consists of a single phase where $\mu(v)$ changes less over time. The update equation is

$$\mu(v)_{\tau+1} = \mu(v)_\tau + \frac{1}{q(v_\tau)}\exp\left(-\theta(v_\tau, v)_\tau\right)G(v, v_\tau) \tag{10}$$

where v_τ is the value of $v(\vec{r})$ at time τ, G is a kernel function (e.g. Gaussian), $q(v_\tau)$ is the probability at position v_τ from the target probability distribution and $\theta(v_\tau, v)_\tau$ is a function which controls convergence. Following the above update rule causes the simulation to converge to the following distributions [7,22], depending on the choice of θ:

$$\theta(v_\tau, v)_\tau = \begin{cases} 1, & \text{non-convergent} \\ \hat{\mu}_\tau, & q(v) \\ \beta\Delta T\mu(v_\tau)_\tau/T, & q(v)^{\Delta T/(\Delta T+T)}P(v)^{T/(\Delta T+T)}) \end{cases} \tag{11}$$

where ΔT is the 'tempering factor' [7], $P(v)$ is the marginal unbiased distribution for $v(\vec{r})$, and $\hat{\mu}$ is the total or average of all previous μ_τ. The first condition, $\theta = 1$, is similar to normal metadynamics; it reaches a distribution similar to $q(v)$ but may oscillate due to a lack of dampening in the update size. Condition two, called globally tempered, converges correctly to $q(v)$. The last condition, locally tempered, converges to an adjustable mixture of the unbiased and target distribution. EDM is traditionally implemented as globally tempered, so that the final distribution is indeed the target. The first condition is good for tuning parameters, since it makes progress more quickly. The final condition, locally tempered, allows a pseudo-Bayesian tuning of prior belief in the unbiased ensemble. It is pseudo-Bayesian because the ratio of influence from the prior belief to evidence is computed, not set, in Bayesian modeling.

An important consideration of both EDS and EDM is that they add potential energy during the update step which quickly becomes kinetic energy. Thus, it is important that the thermostat used to maintain constant temperature in the NVT can dissipate energy faster than it is added by the update steps. This is necessary during the adaptive phase of EDS and EDM prior to convergence.

2.1. Treating uncertainty in experimental data

One complication of minimal biasing methods is uncertainty in the experimental data. For example, the experimental data could be the radius of gyration of a polymer with a reported uncertainty in the mean of 5 nm. Should the radius of gyration be matched exactly, or only to within 5 nm? It is possible to stop the adaptive phase of EDS early, for example when the average is within the uncertainty of the experimental data. An early stop is ad-hoc and not part of the maximum entropy derivation unlike other methods which are built to address uncertainty [11,23].

Cesari et al. [24] proposed a modification to the maximum entropy derivation above to address experimental uncertainty. Instead of the constraint in Equation (2), this constraint is used

$$\int \mathrm{d}\vec{r}\, P''(\vec{r}, \boldsymbol{\epsilon})[s(\vec{r}) + \boldsymbol{\epsilon}] = \hat{s} \tag{12}$$

where ε is an auxiliary variable that allows deviations in the average of $s(\vec{r})$ and $P''(\cdot)$ is the probability distribution that will be solved after maximizing entropy. ε is a random variable from a prior distribution $P_0(\boldsymbol{\epsilon})$ that describes the uncertainty in the experimental data. For example, $P_0(\boldsymbol{\epsilon})$ could be a normal distribution or a Laplace distribution. Cesari et al. [24] show that $P''(\vec{r}, \boldsymbol{\epsilon}) = 1/Z''P'(\vec{r})P_0(\boldsymbol{\epsilon})e^{\lambda\boldsymbol{\epsilon}}$. This leads to a different update step of

$$g_\tau = \frac{2\beta}{w}(< s >_\tau + \xi(\alpha_\tau) - \hat{s}) \times \left(< s^2 >_\tau + \frac{\partial\xi(\alpha_\tau)}{\partial\alpha} - < s >_\tau^2\right) \tag{13}$$

where $\xi(\alpha_\tau)$ is the analytic posterior of ε:

$$\xi(\alpha_\tau) = \frac{\int \mathrm{d}\boldsymbol{\epsilon}\, P_0(\boldsymbol{\epsilon})e^{-\alpha_\tau\boldsymbol{\epsilon}}\boldsymbol{\epsilon}}{\int \mathrm{d}\boldsymbol{\epsilon}\, P_0(\boldsymbol{\epsilon})e^{-\alpha_\tau\boldsymbol{\epsilon}}} \tag{14}$$

Returning to the radius of gyration example above, if the uncertainty, $P_0(\boldsymbol{\epsilon})$, is assumed to be a normal distribution normal with $\sigma = 5$ nm, then $\xi(\alpha_\tau) = -25\alpha_\tau$. This new update-step can be used to rigorously include experimental uncertainty in to the EDS method.

EDM is able to tune the relative importance of the target distribution and the unbiased ensemble with the locally tempered variant in Equation (11). This could be used to match intuition.

For example, if you believe that the target distribution which comes from experimental data is twice as accurate as the molecular dynamics simulation, you could choose $\Delta T/(\Delta T + T) = 0.66$ giving about twice as much weight to the target distribution.

3. Applications of EDS and EDM

A model 1-D system is shown in Figure 2 as a probability distribution function, $P(r)$. EDS is being used to modify the average value of r to match a new set point, $\hat{r}$. EDS adds a linear bias, whose strength is indicated with the red dot, to create the biased PDF $P'(r)$ according to Equation (5). Notice how the features of $P(r)$ are mostly maintained in $P'(r)$. The bottom plot shows how each value of $\hat{r}$ corresponds to a unique biasing strength.

A more sophisticated system which demonstrates the capabilities of improving dynamic observables is the recent work on EDS *ab initio* molecular dynamics (AIMD) simulations of water [10]. DFT water with the BLYP exchange functional poorly represents water structure as seen in Figure 3 (black line). It is over-structured and has water self-diffusion coefficients that are too high (0.005–0.005 Å^2 / ps) [25] compared with the experimental value of 0.23 Å^2/ ps [26]. EDS was used to improve the coordination number of the oxygen–oxygen (O_w–O_w) water molecules and resulted in good agreement with experimental radial distribution function (Figure 3). A number of other observables improved as well, including RDFs and the water self-diffusion coefficient, which increased to 0.06 ± Å^2/ps. White et al. [10] further demonstrated that the EDS bias could be transferred to excess proton-water simulations and that when combined with DFT dispersion corrections [28], the agreement further improves. The accuracy improvement with the dispersion corrections shows that EDS is general in its applicability to DFT methods.

Cortina et al. [29] used EDS to study the KPC-2 carbapenemase enzyme, which is responsible for drug resistance in the majority of carbapenem-resistant Gram-negative bacteria [30]. They used EDS to modify protein-protein distances while doing a committor analysis [31] to identify transition states in the carbapenem-enzyme complex. Cesari et al. [24] used a method similar to EDS (modified update step) to improve agreement with experimental NMR^{3}J coupling data for RNA oligonucleotides. After biasing, they used the simulation results to improve the underlying Amber force field and thus create a transferrable model. Cesari et al. [24] also developed a novel approach to account for experimental data uncertainty by adding auxiliary variables to the EDS update step.

EDM has been used less than EDS due to the rarity of experimental data giving an exact probability distribution. One example explored in the original EDM paper was to construct a mean field bias that mimics an alanine dipeptide being in the backbone of a protein [7]. This was done by computing a potential of mean force (PMF) for ϕ,ψ dihedral angles from PDB crystallography data for the alanine-alanine sequence in protein structures. This corresponds to a probability distribution and a molecular dynamics simulation of alanine dipeptide was done with its ϕ,ψ dihedral angles biased to match the desired values. The result was a molecular dynamics simulation where the dipeptide was biased to behave as part of a longer protein structure. A similar approach was later used by [21]

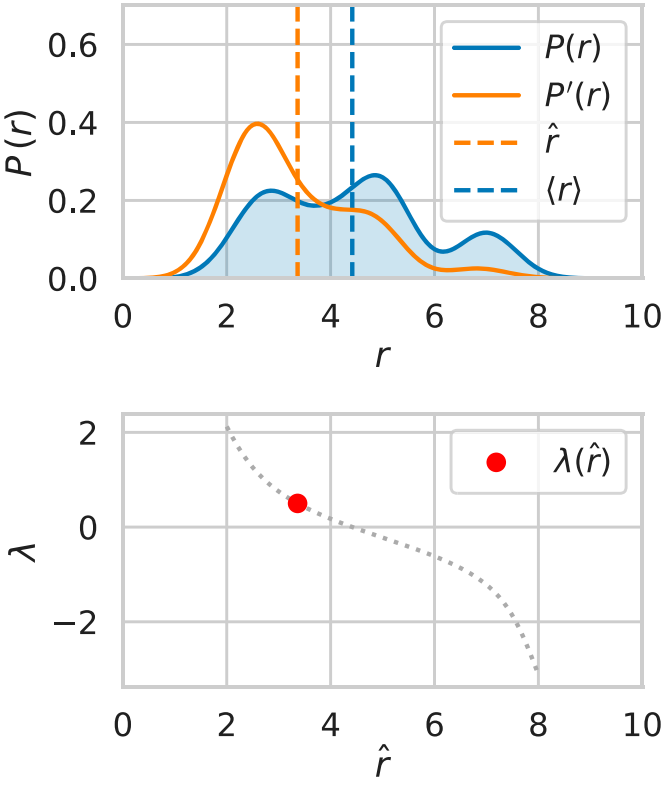

Figure 2. (Colour online) A 1-D EDS calculation where the mean of a probability distribution function (PDF) is being biased to match the dash vertical orange line. The top plot shows the unbiased and biased PDFs. The biased PDF shows as much of the shape of the unbiased PDF as possible while matching the new biased mean. The bottom shows the Lagrange multipliers that give all possible biased means. The red dot indicates the current Lagrange multiplier for the biased PDF. There is a unique λ for all possible biased means, as discussed in the Theory section.

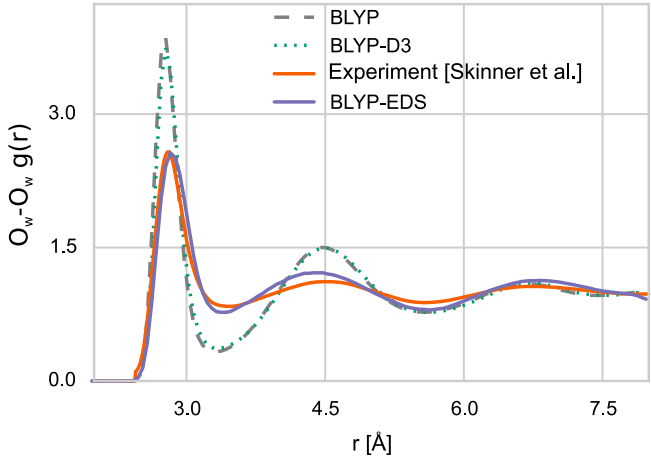

Figure 3. (Colour online) Water oxygen–oxygen radial distribution function from *ab inito* molecular dynamics at 300 K NVT and experiments from [27]. BLYP and BLYP-D3 are from DFT with and without dispersion corrections with the BLYP exchange functional. Note that BLYP-D3 is typically done at 330 K, which gives much better agreement [25]. See [10] for complete system details. The BYLP-EDS line is DFT with EDS bias added to the water oxygen–oxygen coordination number. EDS shows near quantitative agreement with experiment. Reproduced from White AD, Knight C, Hocky GM, et al. Communication: improved ab initio molecular dynamics by minimally biasing with experimental data. J Chem Phys. 2017;146(4), ISSN 00219606., with the permission of AIP Publishing.

with EDS to built a pseudo-mean field model for actin filaments.

Another example of EDM can be found in [13] who used it (called targeted metadynamics) to improve agreement of RNA oligonucleotides' dihedral angles with PDB crystallography data. They similarly built dihedral angle PMFs using the crystallography data and biased molecular dynamics simulations of RNA to adopt the dihedral angle probability distribution functions. Gil-Ley et al. [13] then took the converged EDM bias and transferred it to a different simulation of an RNA tetramer and found little to no improvement of agreement with the crystallography data. This may have been due to the underlying assumption that a PMF derived from crystallographic data is not representative of the room temperature ensemble. Nevertheless, EDM is a promising technique for incorporating experimental data as a type of auxiliary potential energy function in simulations.

3.1. Coarse-grained modeling

EDS and EDM are well-suited for coarse-grain (CG) modeling because they guarantee that a CG model matches experimental data. Dannenhoffer-Lafage et al. [32] showed that EDS can be applied before force-matching (a CG method) and the improvements of agreement with experimental data is maintained. Dannenhoffer-Lafage et al. [32] began with all atom molecular dynamics simulations of Ethylene carbonate with the force field from [33], which is known to not match center-of-mass coordination numbers from more accurate DFT calculations from [34]. Dannenhoffer-Lafage et al. [32] biased the all-atom simulations with EDS to match these known coordination numbers. They then used force-matching to create multiple CG models with one, two, and three site CG beads, using either biased all-atom or unbiased all-atom simulation data. The biased all-atom simulations showed better agreement with coordination numbers, indicating that the improvement in all-atom simulations translates to improvement in the CG model with EDS. Of course, EDS could be used directly on the CG model but that can lead to complications in how observables are calculated on CG models [35].

EDS has been applied to CG simulations of the G- and F-actin proteins as monomers and trimers [21]. Actin proteins are globular proteins that can polymerize into long semi-flexible filaments. Their hypothesis was that they could model a subsystem from within the polymerized actin structure by incorporating information about structural fluctuations from simulations of the larger system via EDS. By biasing the first and second moment of two important collective variables in actin monomer structure, they were able to observe filament-like conformations, and importantly, the fluctuations of these observables for an actin monomer, even in the system which contained only a single monomer solvated in water. Hocky et al. [21] also studied a number of questions about the EDS method in their system and found the following conclusions: (1) the linear term of EDS better matches target values and maintains system fluctuations than harmonic biases; (2) replacing the variance term in Equation (8) with a covariance matrix for all biased dimensions has faster convergence; (3) Levenberg–Marquardt [36] converges faster than stochastic gradient descent. This paper also derived a simple equation that can be used to guess the value of λ without doing a stochastic minimization (whose accuracy depends on the distance of the target observables from the unbiased observables), and hence can serve as a good initial guess for starting an EDS simulation.

3.2. Enhanced sampling

EDM may require enhanced sampling if there are slow degrees of freedom orthogonal to the biased collective variable. This is most conveniently treated via the extensive literature on enhanced sampling with metadynamics, since EDM is a type of metadynamics and typically implemented within a metadynamics code. EDS is not as simple because it requires that the $< \cdot >_\tau$ term in Equation (8) be taken over an NVT ensemble. This limits the enhanced sampling techniques to those that still give correct NVT ensemble averages. One example is parallel-tempering replica-exchange. It is possible to use metadynamics if an appropriate estimator [37] is done to compute the averages but this has not been explored in practice.

Amirkulova and White [38] demonstrated the use of enhanced sampling and EDS with the parallel-tempering well-tempered ensemble (PT-WTE) [39]. The PT-WTE method is a enhancement of enhanced parallel-tempering replica-exchange that improves exchange rates and reduces the required number of replicas [40]. PT-WTE satisfies the observable that the ensemble averages, $< \cdot >_\tau$, can be computed during the course of the simulation because PT-WTE only changes the magnitude of potential energy fluctuations, not their expectations. One apparent drawback is that the method loses the one-replica observable of EDS that allows computation of dynamic observables. However, this only applies to the adaptive phase. During the fixed-bias second phase of EDS, one replica can again be used to allow analysis of dynamic observables.

Amirkulova and White [38] studied the GYG peptide with the EDS plus enhanced sampling approach. Eight simulations were conducted, each with 1, 8, or 16 replicas. The simulations had EDS, PT-WTE, and/or parallel-tempering. EDS was used to bias proton chemical shifts to improve agreement with experimental NMR data. The simulation results showed that PT-WTE improves sampling in the EDS method and does not change agreement with experimental data. PT-WTE also converged the EDS bias with fewer replicas than the PT method. One consideration in all EDS simulations is the effect on unbiased observables. The Ramachandran plots of the simulations with and without EDS bias are compared in Figure 4. When EDS and enhanced sampling are used (Figure 4 (a)), the simulation explores a larger region of configurational space relative to the control simulation (Figure 4 (b)). Also, the global minimum changes when using EDS in Figure 4, bringing it closer to what was found in [41] for the GYG sequence.

4. The role of pressure

Thus far, we have discussed NVT and NVE ensembles. EDM and EDS add new forces to the simulation and affect the system virial and pressure. This can lead to undesirable density changes when the bias is subsequently used in an NPT

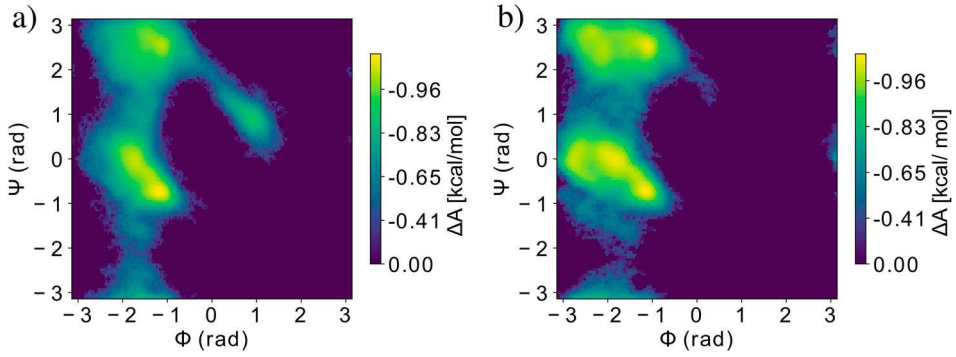

Figure 4. (Colour online) Free energy surface of GYG peptide along dihedral angles (Ramachandran plot) from [38]. (a) EDS and enhanced sampling with PT-WTE and (b) no EDS and no enhanced sampling. The difference between panels (a) and (b) shows that EDS changes the global free energy minimum, which better matches data from [41], and PT-WTE improves sampling based on explored regions. Republished with permission of World Scientific Publishing Co., Inc., from Combining enhanced sampling with experiment-directed simulation of the GYG peptide, Amirkulova DA and White AD, 17, 3 and 2018; permission conveyed through Copyright Clearance Center, Inc.

simulation. It may be possible to rectify this change in pressure, and thus density, by adding a further constraint that the contribution to the pressure from potential energy be unchanged while still maximizing entropy. As shown in the Appendix, this leads to an unsolvable equation or requires changing changing density to constrain pressure. One intuitive reason for this is that the Clausius virial theorem requires the average virial be proportional to average temperature: $< T >= -1/2 \sum_i^N < \vec{F}_i \cdot \vec{r}_i >$. Therefore, if we tried to bias an ensemble with fixed density so that its average virial is constrained to a new value, then the average temperature would change, violating our constrained average temperature.

One way around this challenge lies in the correlation between biased observables in EDS. When biasing multiple observables which are correlated, there is in fact no longer a unique set of Lagrange multipliers [16]. These extra degrees of freedom in the choice of Lagrange multipliers that maximize Equation (3) enable us to choose the Lagrange multipliers that minimally change the virial. Equation (8) can be modified so that our coupling constants minimize the additional pressure which they exert [42]:

$$g_\tau = \frac{2\beta}{w}(< s >_\tau - \hat{s})\left(< s^2 >_\tau - < s >_\tau^2\right) + 2\nu\Delta p_\tau \frac{\partial \Delta p_\tau}{\partial \alpha_\tau} \quad (15)$$

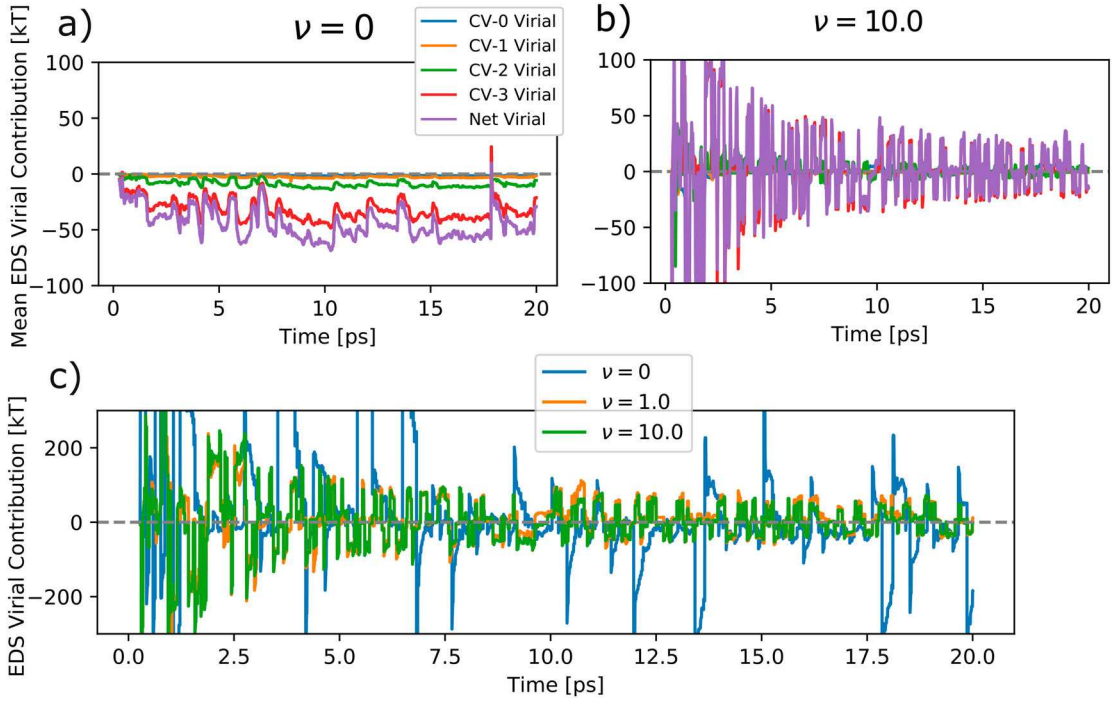

Figure 5. (Colour online) EDS applied to water coordination number moments 0 through 3 in molecular dynamics simulation of over-structured SCP/E water model. The columns show increasing strength of virial minimization. Each plot is collective variable scaled by its set-point. The vertical dashed line spearates the NVT adaptive simulation and a fixed bias NPT phase. The results show that increasing the strength of the virial minimization actually improves convergence by reducing large-magnitude changes in biasing force. Without virial minimization, EDS can produce nonphysical densities when the NVT bias is transferred to NPT. The change in density moves the CVs. For $\nu^* = 0$, this meant CV values beyond the *y*-limits of the plot.

$$\Delta p_\tau = -\frac{\alpha_\tau}{w}\frac{ds}{dV} = -\frac{\alpha_\tau}{w}\sum_{ij}\left(\frac{\partial s}{\partial v_{ij}}\frac{\partial v_{ij}}{\partial V}\right) \tag{16}$$

where ν is a parameter that controls the importance of minimizing the virial, Δp_τ is the change in the system virial pressure due to the EDS bias, $\partial s/\partial v_{ij}$ is the partial derivative of the observable with respect to one component of the triclinic box matrix, and $\partial v_{ij}/\partial V$ can be computed from the adjugate of the triclinic box matrix. When biasing multiple observables, Δp_τ should be the total over all observables. g_τ is in units of s^2 per energy, so ν must be in units of volume squared times s^2 per energy squared. Practically, we can choose a unitless ν^*, which is defined from

$$\nu = \nu^* \frac{w^2\beta^2}{\rho^2} \tag{17}$$

Equation (15) was used in a molecular dynamics simulation of 128 modified SPC/E water molecules. This modified SPC/E water has increased charges ($q_O = -0.94$) to distort its coordination number. EDS was then applied to correct the coordination number using experimentally derived coordination numbers from [27] with varying strengths of the virial correction. The coordination number moment definition may be found in [6]. Figure 5 shows that coordination number and its moments are still correctly biased with the virial term. The EDS parameters were a range (A) of 50 kj/mol, a period of 25 fs, and the Levenberg–Marquardt optimization procedure. A Nose-Hoover thermostat with a time constant of 25 fs and a timestep of 0.5 fs were used for the molecular dynamics in the LAMMPS simulation engine [43]. The NPT barostat was Parrinello-Rahman.

Figure 6 shows the effect of the virial penalty term on the EDS virial contribution. The virial contribution plotted and computed here is the mean per-particle virial energy contribution. That is $\Delta p_\tau/\rho/kT$ which removes the effect of particle number and energy scale. Panel (a) shows the cumulative mean virial contribution of each collective variable with $\nu^* = 0$ (no virial minimization). The net average virial contribution is 39.2 kT. Panel (b) shows $\nu^* = 10$ and there is a much lower contribution of 5.02 kT. Panel (c) compares the instantaneous net virial contribution of three different ν^* values. The lack of virial minimization shows large swings in the virial contribution, even after EDS has converged longer timesteps.

Figure 5 shows the impact on the convergence of the biased coordination number and its moments with the virial minimizing terms. Interestingly, adding a virial minimization term actually improves convergence. This is due to the well-known effect of regularization, which improves convergence in optimization [44]. The virial minimization term is proportional to α_τ, so it brings down the magnitude of α_τ. This prevents the large

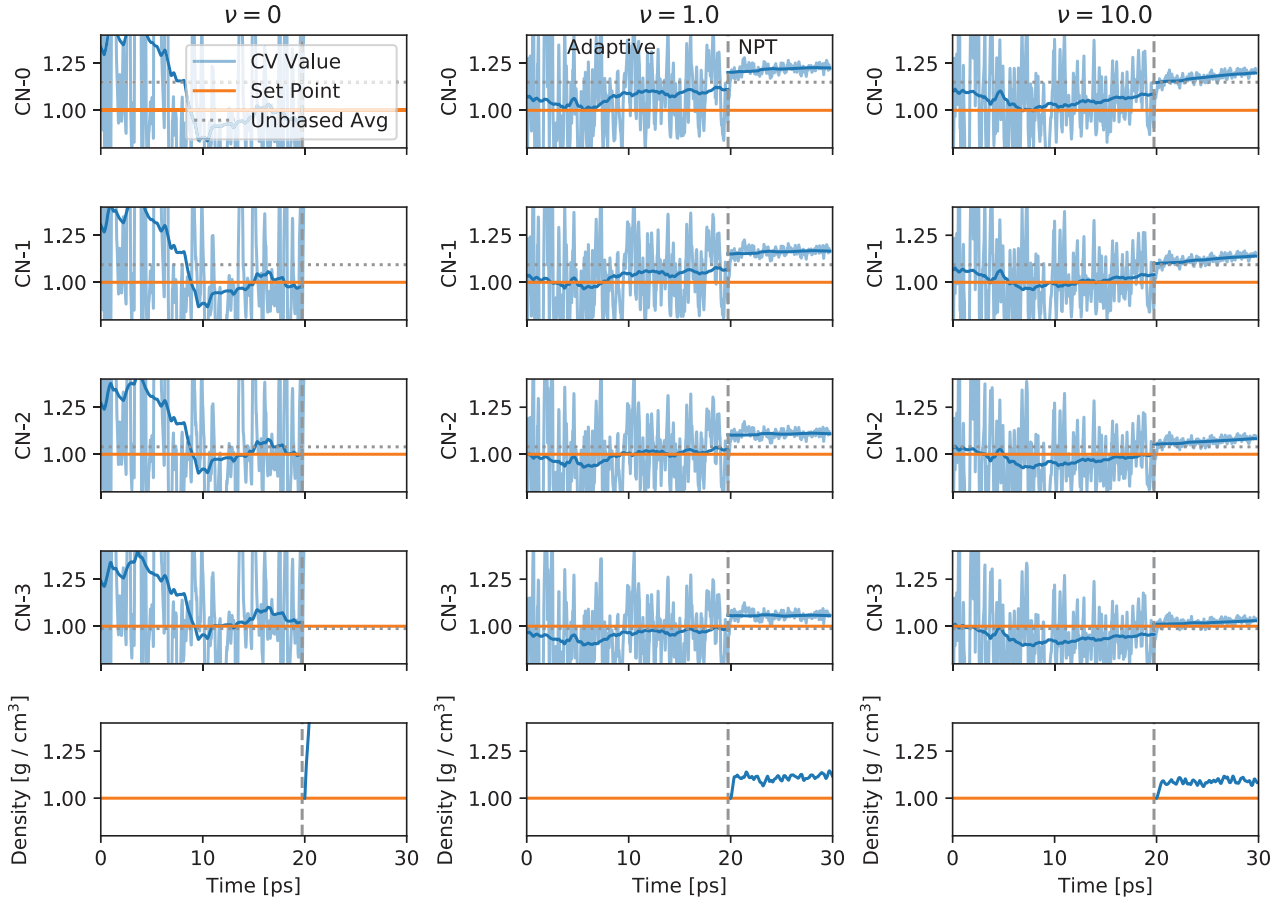

Figure 6. (Colour online) The EDS contribution to virial as EDS is applied to water coordination number moments 0 through 3 in molecular dynamics simulation of over-structured SCP/E water model. Panels (a) and (b) compare the cumulative mean virial contribution in a normal EDS and EDS with virial minimization strength of $\nu^* = 10$ broken out by individual CV. There is lower per collective variable mean virial contribution with virial minimization. Panel (c) shows the effect of different ν^* values on the instantaneous net virial contribution from EDS.

swings seen in the $\nu = 0$ system. Thus, adding the virial minimization term not only reduces the virial contribution but can improve convergence by minimizing magnitudes of the coupling constants. The true test can be seen in the NPT portion past 20ps where the bias is fixed but the box dimensions are free. EDS with $\nu^* = 0$ gives poor densities in NPT with a bias computed in NVT. The virial minimization reduces the change in density with a trade-off in match to the bias set-points. Virial minimization is in general recommended due to its enhanced convergence and better representation of virial forces.

5. Implementations

There are three implementations of EDS available: Colvars [45], Plumed1.3 [46], and Plumed2 [47]. The Colvars and Plumed1.3 implementations match the original EDS manuscript. The Plumed2 implementation is actively maintained as a plugin [47] and has features from more recent EDS articles [21]. For example, the covariance term in Equation (8) can be computed using the full sample covariance matrix or the sample variance. The update steps can also be computed using the Levenberg–Marquardt method [21,36], instead of Equation (7). The Plumed2 implementation is recommended.

EDM is implemented in Plumed2 as well, under the 'targeted metadynamics' keyword within the metadynamics biasing class. One of the common use-cases for EDM is to bias RDFs, which is not exactly a collective variable. RDFs are 'function collective variables' in the sense that there is a distribution at each timestep. Thus the bias update equation, shown in Equation (10), applies to each observed point in the RDF at each step (see [7] for equations). This can lead to thousands of updates per timestep on even small systems. To improve performance and scaling, we have created an implementation in LAMMPS that is more tightly integrated with the simulation engine than Plumed2 [7]. Note that the addition of hills is parallelized here, which is unusual and enables the scaling as a function of particle size. Hill addition is 90% of the CPU utilization. This was the implementation used in [7] for an ethylene carbonate electrolyte simulation and Lennard-Jones benchmark systems.

6. Conclusions

Minimal biasing techniques are an emerging area that improve the accuracy of molecular simulation and better utilize comparable experimental data. There are many proposed methods for minimal biasing [3–5]. EDS and EDM are two maximum entropy techniques developed by the authors that minimally bias an ensemble so that average observables or probability distributions of observables match set values. EDS and EDM both converge the potential energy of a simulation to Equation (5), allowing a subsequent NVE simulation with fixed potential energy in a single replica so that dynamic observables can be computed. EDS and EDM can bias multiple observables simultaneously and be combined with enhanced sampling methods. It is possible to treat uncertainty in experimental data using methods from [24]. Explicitly setting pressure is not possible in general for maximum entropy biasing as shown in this work. However, it is possible to minimize the change to pressure when biasing multiple collective variables in EDS. This minimization can actually improve convergence of EDS by acting as a regularization. A variety of example systems have been presented here and there are implementations available in most simulation engines for both EDS and EDM.

Acknowledgments

The authors thank Prof. Pengfei Huo for reviewing the appendix mathematics, Rainier Barrett for help preparing the manuscript and Prof. Glen Hocky for assistance in surveying the literature and preparing the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors..

Funding

This work was supported by the National Science Foundation CBET Division of Chemical, Bioengineering, Environmental, and Transport Systems [grant no. 1751471].

References

[1] Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the amber ff99SB protein force field. Proteins. 2010;78(8):1950–1958. ISSN 08873585.
[2] Shivakumar D, Williams J, Wu Y, et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. J Chem Theory Comput. 2010;6(5):1509–1519. ISSN 15499626.
[3] Sormanni P, Piovesan D, Heller GT, et al. Simultaneous quantification of protein order and disorder. Nat Chem Biol. 2017;13(4):339–342. ISSN 1552-4450.
[4] Bonomi M, THeller G, Camilloni C, et al. Principles of protein structural ensemble determination. Curr Opin Struct Biol. 2017;42:106–116. ISSN 1879033X.
[5] Olsson S, Frellsen J, Boomsma W, et al. Inference of structure ensembles of flexible biomolecules from sparse, averaged data. PLoS ONE. 2013;8(11):e79439. ISSN 1932-6203.
[6] White AD, Voth GA. Efficient and minimal method to bias molecular simulations with experimental data. J Chem Theory Comput. 2014;10(8):3023–3030.
[7] White A, Dama J, Voth GA. Designing free energy surfaces that match experimental data with metadynamics. J Chem Theory Comput. 2015112451–2460. ISSN 1549-9618.
[8] Beauchamp KA, Pande VS, Das R. Bayesian energy landscape tilting: towards concordant models of molecular ensembles. Biophys J. 2014;106(6):1381–1390. ISSN 0006-3495.
[9] Lindorff-Larsen K, Best RB, DePristo MA, et al. Simultaneous determination of protein structure and dynamics. Nature. 2005;433(7022):128–132.ISSN 0028-0836.
[10] White AD, Knight C, Hocky GM, et al. Communication: improved ab initio molecular dynamics by minimally biasing with experimental data. J Chem Phys. 2017;146(4), ISSN 00219606.
[11] Hummer G, K öfinger J. Bayesian ensemble refinement by replica simulations and reweighting. J Chem Phys. 2015;143(24):243150. ISSN 00219606.
[12] Marinelli F, Faraldo-Gómez J. Ensemble-biased metadynamics: a molecular simulation method to sample experimental distributions. Biophys J. 2015;108(12):2779–2782. ISSN 15420086.
[13] Gil-Ley A, Bottaro S, Bussi G. Empirical corrections to the amber RNA force field with target metadynamics. J Chem Theory Comput. 2016;12(6):2790–2798. ISSN 15499626.

[14] Valsson O, Parrinello M. Variational approach to enhanced sampling and free energy calculations. Phys Rev Lett. 2014;113 (9):090601.
[15] Dama JF, Hocky GM, Sun R, et al. Exploring valleys without climbing every peak: more efficient and forgiving metabasin metadynamics via robust on-the-fly bias domain restriction. J Chem Theory Comput. 2015;11(12):5638–5650. ISSN 15499626.
[16] Pitera JW, Chodera JD. On the use of experimental observations to bias simulated ensembles. J Chem Theory Comput. 2012;8 (10):3445–3451. ISSN 1549-9618.
[17] Jaynes ET. Information theory and statistical mechanics. Phys Rev. 1957;106(4):620–630.
[18] Roux B, Weare J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. J Chem Phys. 2013;138(8):084107ISSN 00219606.
[19] Boura E, Rozycki B, Herrick DZ, et al. Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. Proc Natl Acad Sci. 2011;108(23):9437–9442. ISSN 0027-8424.
[20] McMahan HB, Streeter M. Adaptive bound optimization for online convex optimization. arXiv. 2010; 1–19. ISSN 09505849.
[21] Hocky GM, Dannenhoffer-Lafage T, Voth GA. Coarse-grained directed simulation. J Chem Theory Comput. 2017;13(9):4593–4603. ISSN 15499626.
[22] Dama JF, Parrinello M, Voth GA. Well-tempered metadynamics converges asymptotically. Phys Rev Lett. 2014;112(24):240602. ISSN 0031-9007.
[23] Brookes DH, Head-Gordon T. Experimental inferential structure determination of ensembles for intrinsically disordered proteins. J Am Chem Soc. 2016;138(13):4530–4538. ISSN 15205126.
[24] Cesari A, Gil-Ley A, Bussi G. Combining simulations and solution experiments as a paradigm for RNA force field refinement. J Chem Theory Comput. 2016;12(12):6192–6200. ISSN 15499626.
[25] Steve Tse YL, Knight C, Voth GA. An analysis of hydrated proton diffusion in ab initio molecular dynamics. J Chem Phys. 2015;142 (1, ISSN 00219606.
[26] Krynicki K, Green CD, Sawyer DW. Pressure and temperature dependence of self-diffusion in water. Faraday Discuss Chem Soc. 1978;66(0):199–208.
[27] Skinner LB, Huang C, Schlesinger D, et al. Benchmark oxygen–oxygen pair-distribution function of ambient water from X-ray diffraction measurements with a wide Q-range. J Chem Phys. 2013;138(7), ISSN 00219606.
[28] Sitarz M, Wirth-Dzieciolowska E, Demant P. Loss of heterozygosity on chromosome 5 in vicinity of the telomere in γ-radiation-induced thymic lymphomas in mice. Neoplasma. 2000;47(3):148–150. ISSN 00282685.
[29] Cortina GA, Hays JM, Kasson PM. Conformational intermediate that controls KPC-2 catalysis and beta-Lactam drug resistance. ACS Catal. 2018;8(4):2741–2747. ISSN 21555435.
[30] Pemberton OA, Zhang X, Chen Y. Molecular basis of substrate recognition and product release by the klebsiella pneumoniae carbapenemase (KPC-2). J Med Chem. 2017;60(8):3525–3530.
[31] Du R, Pande VS, Grosberg AY, et al. On the transition coordinate for protein folding. J Chem Phys. 1998;108(1):334–350.
[32] Dannenhoffer-Lafage T, White AD, Voth GA. A direct method for incorporating experimental data into multiscale coarse-grained models. J Chem Theory Comput. 2016;12(5):2144–2153. ISSN 1549-9626.
[33] Masia M, Probst M, Rey R. Ethylene carbonate-Li+: a theoretical study of structural and vibrational properties in gas and liquid phases. J Phys Chem B. 2004;108(6):2016–2027. ISSN 1520-6106.
[34] Borodin O, Smith GD. Quantum chemistry and molecular dynamics simulation study of dimethyl carbonate: ethylene carbonate electrolytes doped with liPF6. J Phys Chem B. 2009;113(6):1763–1776. ISSN 1520-6106.
[35] Wagner JW, Dama JF, Durumeric AEP, et al. On the representability problem and the physical meaning of coarse-grained models. J Chem Phys. 2016;145(4, ISSN 00219606.
[36] Stinis P. A maximum likelihood algorithm for the estimation and renormalization of exponential densities. J Comput Phys. 2005;208 (2):691–703. ISSN 00219991.
[37] Tiwary P, Parrinello M. A time-independent free energy estimator for metadynamics. J Phys Chem B. 2015;119(3):736–742. ISSN 1520-6106.
[38] Amirkulova DB, White AD. Combining enhanced sampling with experiment-directed simulation of the GYG peptide. J Theor Comput Chem. 2018;17(03):1840007. ISSN 0219-6336.
[39] Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. Phys Rev Lett. 2010;104(19):1–4. ISSN 00319007.
[40] Deighan M, Bonomi M, Pfaendtner J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. J Chem Theory Comput. 2012;8(7):2189–2192. ISSN 15499618.
[41] Ting D, Wang G, Shapovalov M, et al. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model. PLOS Comput Biol. 2010;6(4). ISSN 1553734X.
[42] Louwerse MJ, Baerends EJ. Calculation of pressure in case of periodic boundary conditions. Chem Phys Lett. 2006;421(1–3):138–141. ISSN 00092614.
[43] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995;117(1):1–19. ISSN 0021-9991.
[44] Scholkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge (MA): MIT Press; 2001.
[45] Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Phys. 2013;111(22–23):3345–3362.
[46] Bonomi M, Branduardi D, Bussi G, et al. A portable plugin for free-energy calculations with molecular dynamics. Comput Phys Commun. 2009;180(10):1961–1972. ISSN 00104655.
[47] Tribello GA, Bonomi M, Branduardi D, et al. New feathers for an old bird. Comput Phys Commun. 2014;185(2):604–613. ISSN 0010-4655.

Appendix Minimal biasing pressure derivation

We would like to constrain the average contribution to virial pressure from potential energy while minimizing relative entropy in a biased ensemble. The goal is to modify the potential energy while keeping temperature, number of particles and volume the same. We define the virial contribution from potential energy to be $-(\mathrm{d}U'/\mathrm{d}V)$ where V is volume and U' is the potential energy of our biased ensemble. We can rewrite this as

$$\frac{1}{\beta}\frac{\mathrm{d}\ln P(\vec{r})}{\mathrm{d}V} = \frac{1}{\beta}\frac{\mathrm{d}}{\mathrm{d}V}\left[-\beta U'(\vec{r}) - \ln Z\right] = -\frac{\mathrm{d}U'}{\mathrm{d}V} \tag{A1}$$

where V is volume and $Z = \int \mathrm{d}\vec{r} P(\vec{r})$. The constraints for this ensemble are

$$\int \mathrm{d}\vec{r} P'(\vec{r}) = 1 \tag{A2}$$

which enforces normalization and to set the average $\frac{-\mathrm{d}U'}{\mathrm{d}V}$ to be equal to the scalar p_e:

$$\begin{aligned} < -\frac{\mathrm{d}U'}{\mathrm{d}V} > &= \frac{1}{\beta} < \frac{\mathrm{d}\ln P'(\vec{r})}{\mathrm{d}V} > = \frac{1}{3\beta V}\int \mathrm{d}\vec{r}\vec{r}\cdot\frac{\partial \ln P'(\vec{r})}{\partial \vec{r}} P'(\vec{r}) \\ &= \frac{1}{3\beta V}\int \mathrm{d}\vec{r}\vec{r}\cdot\nabla P'(\vec{r}) = p_e \end{aligned} \tag{A3}$$

where we have taken the $\partial \ln P(\vec{r})/\partial V|_{\vec{r}}$ term, which describes change in potential energy with fixed atom positions, to be zero. The quantity to be minimized is relative entropy of the biased probability distribution with respect to the unbiased probability distribution:

$$\Delta S_{\mathrm{rel}} = \int P'(\vec{r}) \ln \frac{P'(\vec{r})}{P(\vec{r})} \tag{A4}$$

To solve this system, we compute the functional derivative of the Lagrange function with respect to the biased probability distribution

$P'(\vec{r})$. This requires a functional derivative of Equation (A3), which is

$$\frac{\delta < -\frac{dU'}{dV} >}{\delta P'(\vec{r})} = -\frac{1}{3\beta V}\nabla[r_1, r_2, \ldots] = -\frac{3N}{3\beta V} \tag{A5}$$

Some intuition for Equation (A5) can be gained from the definition of the functional derivative. Equation (A5) says that adding an arbitrary test function $\phi(\vec{r})$ to the probability distribution scaled by ε, where ε is small has the following result:

$$\int d\vec{r}\,\vec{r}\cdot\nabla[P(\vec{r}) + \epsilon\phi(\vec{r})] - \int d\vec{r}\,\vec{r}\cdot\nabla P(\vec{r}) \approx -\frac{N}{\beta V}\int dr\,\epsilon\phi(\vec{r}) \tag{A6}$$

The right-hand side is independent of the probability distribution, so changing the probability distribution decreases the virial by N regardless of where the change occurs. With the functional derivative of the virial being constant, the functional derivative of the Lagrange function is

$$\frac{\delta\mathcal{L}}{\delta P'(\vec{r})} = \ln P'(\vec{r}) - \ln P(\vec{r}) + \lambda_1 - \lambda_2\frac{N}{\beta V} = 0 \tag{A7}$$

$$P'(\vec{r}) = \lambda_1 P(\vec{r}) e^{-\lambda_2\frac{N}{\beta V}} \tag{A8}$$

which has no solution if N and V are not variables because λ_1 and λ_2 are colinear. Thus, there is no way constrain the potential energy contribution to pressure. This argument does not apply to constraining pressure by changing volume or density, which is the definition of an NPT ensemble or a grand canonical ensemble.