JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Well-Tempered Variational Approach to Enhanced Sampling

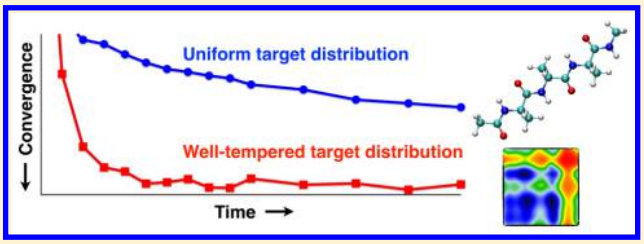
Omar Valsson*[,†,‡] and Michele Parrinello[†,‡]

[†]Department of Chemistry and Applied Biosciences, ETH Zurich, c/o USI Campus, Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland

[‡]Facoltà di Informatica, Instituto di Scienze Computationali, and National Center for Computational Design and Discovery of Novel Materials MARVEL, Università della Svizzera italiana (USI), Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland

Supporting Information

ABSTRACT: We propose a simple yet effective iterative scheme that allows us to employ the well-tempered distribution as a target distribution for the collective variables in our recently introduced variational approach to enhanced sampling and free energy calculations [Valsson and Parrinello *Phys. Rev. Lett.* **2014**, *113*, 090601]. The performance of the scheme is evaluated for the three-dimensional free energy surface of alanine tetrapeptide where the convergence can be rather poor when employing the uniform target distribution. Using the well-tempered target distribution on the other hand results in a significant improvement in convergence. The results observed in this paper indicate that the well-tempered distribution is in most cases the preferred and recommended choice for the target distribution in the variational approach.



1. INTRODUCTION

The use of atomic based molecular dynamics (MD) and Monte Carlo (MC) simulation is widespread in many fields of science. However, many systems are characterized by metastable states separated by high free energy barriers, which makes it rather difficult to obtain a proper and complete sampling of phase space. This has prompted the introduction of a wide variety of enhanced sampling methods that aim at solving this pressing issue. Many of these methods are based on the principle of umbrella sampling[1] and introduce an external bias potential that facilitates transitions between metastable states.[2−7] The bias potential normally acts on a low-dimensional space spanned by so-called collective variables (CVs) designed to give an appropriate coarse grained description of the system.

We have recently introduced a new CV based enhanced sampling approach where the bias potential is constructed by variationally minimizing a convex functional.[8] At the minimum the bias potential is such that the CVs are sampled according to a predefined target distribution, and furthermore the bias potential relates in a simple manner to the free energy as a function of the CVs. The freedom of selecting the target distribution gives a lot of flexibility to the method. The simplest option is to take the target distribution as uniform such that all values in CV space are evenly sampled. However, this choice might not always be optimal since much time will be spent on sampling regions high in free energy that are not of interest, which in some cases might lead to poor convergence. Another issue with the choice of the uniform target distribution is that this often requires restrains on the values of the CVs in order to avoid sampling irrelevant (and perhaps unphysical) regions of phase space, especially in the case of unbound CVs. Therefore, it is probably better to instead consider a target distribution that focuses sampling on the most relevant low free energy regions.

One intriguing choice for the target distribution that achieves this goal is the distribution asymptotically sampled in well-tempered metadynamics.[9,10] This well-tempered distribution can be viewed as effectively sampling the CV distribution at some higher, but finite, temperature. However, the well-tempered distribution depends on the a priori unknown free energy surface and thus cannot be used directly as a target distribution in the variational approach. To solve this we introduce here a simple yet effective iterative scheme that allows for converging the target distribution to the well-tempered distribution. The performance of the scheme is evaluated by considering as a numerical example the three-dimensional free energy surface of alanine tetrapeptide. We find that the use of the well-tempered target distribution results in a significant improvement in the convergence as compared to the uniform target distribution.

The paper is organized as follows. In Section 2 we review the variational approach and introduce the iterative scheme used to achieve the well-tempered distribution. In Section 3 we give the details of the computational setup and introduce the error metric employed to measure the convergence of the simulations. In Section 4 we show our results for alanine tetrapeptide, and finally we conclude the paper in Section 5.

Received: January 28, 2015
Published: April 20, 2015

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

2. METHODS

2.1. Variational Approach. In order to keep the paper self-contained we will start by briefly reviewing the main principles of the variational approach introduced in our previous paper.[8]

We consider a system described by microscopic coordinates $\mathbf{R}$ and a potential energy function $U(\mathbf{R})$ and introduce a set of suitably chosen CVs $\mathbf{s}(\mathbf{R}) = (s_1(\mathbf{R}),s_2(\mathbf{R}), ..., s_d(\mathbf{R}))$ that properly describe the different metastable states of the system. The free energy surface (FES) corresponding to these CVs is defined up to a constant as

$$F(\mathbf{s}) = -\frac{1}{\beta}\log\int d\mathbf{R}\,\delta(\mathbf{s}-\mathbf{s}(\mathbf{R}))\,e^{-\beta U(\mathbf{R})} \tag{1}$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature, k_B is the Boltzmann constant, and T is the temperature. The unbiased equilibrium distribution of the CVs is then defined as $P(\mathbf{s}) = e^{-\beta F(\mathbf{s})}/Z$ where $Z = \int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}$ is the partition function of the system.

We introduce a bias potential $V(\mathbf{s})$ that acts on the CVs and the functional $\Omega[V]$ that depends on $V(\mathbf{s})$

$$\Omega[V] = \frac{1}{\beta}\log\frac{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s}\, p(\mathbf{s})V(\mathbf{s}) \tag{2}$$

where $p(\mathbf{s})$ is a chosen target probability distribution that is assumed to be normalized. The $\Omega[V]$ functional is closely related to the relative entropy[11] and to the Kullback−Leibler divergence.[12] It can be shown that the bias potential that minimizes $\Omega[V]$ is related to $F(\mathbf{s})$ by

$$V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta}\log p(\mathbf{s}) - \frac{1}{\beta}\log Z_V \tag{3}$$

where $Z_V = \int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}$ is a constant that usually can be ignored. This is also the global minimum since $\Omega[V]$ is a convex functional.

The distribution $p(\mathbf{s})$ is called the target distribution as the CVs will be sampled according to $p(\mathbf{s})$ when the optimal bias potential in eq 3 acts on the system. In other words, $p(\mathbf{s})$ is the CV distribution that we want to achieve when minimizing $\Omega[V]$. We can select the target distribution as we desire, which gives a lot of flexibility to the method. The only requirement is that $p(\mathbf{s})$ is a proper probability distribution (i.e., positive and normalized) and, in principle, a priori known although this later requirement can be circumvented as we will show below. The simplest choice is to consider the uniform target distribution $p(\mathbf{s}) = 1/\Omega_{\mathbf{s}}$ where $\Omega_{\mathbf{s}} = \int d\mathbf{s}$ is the volume of CV space. This choice results in a direct relation between the bias potential and the FES, $F(\mathbf{s}) = -V(\mathbf{s})$ where we have dropped an irrelevant constant.

The variational property of $\Omega[V]$ is used by assuming a functional form for the bias potential $V(\mathbf{s};\boldsymbol{\alpha})$ that depends on a set of variational parameters $\boldsymbol{\alpha} = (\alpha_1,\alpha_2,...,\alpha_K)$. The function $\Omega(\boldsymbol{\alpha}) = \Omega[V(\boldsymbol{\alpha})]$ is then minimized with respect to α by iteratively updating the parameters using the gradient $\Omega'(\boldsymbol{\alpha})$

$$\frac{\partial\Omega(\boldsymbol{\alpha})}{\partial\alpha_i} = -\left\langle\frac{\partial V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_i}\right\rangle_{V(\alpha)} + \left\langle\frac{\partial V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_i}\right\rangle_p \tag{4}$$

and the Hessian $\Omega''(\boldsymbol{\alpha})$

$$\frac{\partial^2\Omega(\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_i} = \beta\cdot\mathrm{cov}\left[\frac{\partial V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_j},\frac{\partial V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_i}\right]_{V(\alpha)} - \left\langle\frac{\partial^2 V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_i}\right\rangle_{V(\alpha)} + \left\langle\frac{\partial^2 V(\mathbf{s};\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_i}\right\rangle_p \tag{5}$$

where the expectation values and covariance are either obtained in a biased simulation employing the potential $V(\mathbf{s};\boldsymbol{\alpha})$ or over the target distribution $p(\mathbf{s})$. The gradient and the Hessian can only be evaluated stochastically and thus they can be rather noisy so it is generally most effective to use stochastic optimization methods.[13,14]

The FES can be obtained directly from eq 3 if the assumed functional form has sufficient variational flexibility. Otherwise, $F(\mathbf{s})$ can be estimated using standard umbrella sampling reweighting, and this can even be done on the fly during the optimization process when the bias potential has reached a quasi-stationary state.

We write the bias potential as a linear expansion

$$V(\mathbf{s};\boldsymbol{\alpha}) = \sum_k \alpha_k f_k(\mathbf{s}) \tag{6}$$

where the basis functions $f_k(\mathbf{s})$ can be for example plane waves, splines, or Chebyshev polynomials. A small number of terms in the expansion normally suffices as $F(\mathbf{s})$ is generally smooth. For the linear expansion in eq 6 the gradient $\Omega'(\boldsymbol{\alpha})$ and the Hessian $\Omega''(\boldsymbol{\alpha})$ simplify

$$\frac{\partial\Omega(\boldsymbol{\alpha})}{\partial\alpha_i} = -\langle f_i(\mathbf{s})\rangle_{V(\alpha)} + \langle f_i(\mathbf{s})\rangle_p \tag{7}$$

$$\frac{\partial^2\Omega(\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_i} = \beta\cdot\mathrm{cov}[f_i(\mathbf{s}),f_j(\mathbf{s})]_{V(\alpha)} \tag{8}$$

and just involve the expectation values and the covariances of the basis functions.

The variational approach is trivially implementable in a multiple walker framework[15] where the different walkers share the same bias potential $V(\mathbf{s};\boldsymbol{\alpha})$ (i.e., same variational parameter $\boldsymbol{\alpha}$) and together contribute to sampling the averages and covariances needed for the gradient and the Hessian. The communication between walkers is minimal as the averages and covariances are sampled separately for each walker and only need to be combined when updating the variational parameters.

2.2. Stochastic Optimization Scheme. The optimization process is performed using the averaged stochastic gradient descent algorithm introduced in ref 14. In this algorithm we consider at iteration n both the instantaneous parameters $\boldsymbol{\alpha}^{(n)}$ and the averaged parameters $\bar{\boldsymbol{\alpha}}^{(n)} = n^{-1}\sum_{l=0}^{n}\boldsymbol{\alpha}^{(l)}$. The instantaneous parameters are updated according to

$$\boldsymbol{\alpha}^{(n+1)} = \boldsymbol{\alpha}^{(n)} - \mu[\Omega'(\bar{\boldsymbol{\alpha}}^{(n)}) + \Omega''(\bar{\boldsymbol{\alpha}}^{(n)})[\boldsymbol{\alpha}^{(n)} - \bar{\boldsymbol{\alpha}}^{(n)}]] \tag{9}$$

where μ is fixed step size, and the gradient and the Hessian are obtained by sampling from a simulation where the bias potential $V(\mathbf{s};\bar{\boldsymbol{\alpha}}^{(n)})$ is acting on the system. In practice we have found that it is sufficient to consider only the diagonal part of the Hessian, and this is done in all calculations here.

This algorithm has two benefits. First, the bias potential, and the corresponding estimate of $F(\mathbf{s})$ from eq 3, depend on the

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

averaged parameter $\bar{\boldsymbol{\alpha}}^{(n)}$ and thus change and converge smoothly. Second, we can generally use rather short sampling times at each iteration. For example, in the classical systems studied here we use a sampling time on the order of a picosecond.

2.3. Iterative Scheme for the Well-Tempered Target Distribution. The well-tempered distribution is the CV distribution asymptotically sampled in well-tempered metadynamics[9,10] and achieved in other enhanced sampling methods by assuming an adiabatic separation between the CVs and the other degrees of freedom.[16−18] We stress that no such assumption is needed in the variational approach and in metadynamics as well. The well-tempered distribution reads as

$$p(\mathbf{s}) = \frac{e^{-\beta' F(\mathbf{s})}}{\int d\mathbf{s}\, e^{-\beta' F(\mathbf{s})}} \propto [P(\mathbf{s})]^{1/\gamma} \tag{10}$$

where $F(\mathbf{s})$ is the target FES at temperature T as defined in eq 1 above, $\beta' = [k_B(T+\Delta T)]^{-1}$ where ΔT is a parameter with the units of temperature, and $P(\mathbf{s}) = e^{-\beta F(\mathbf{s})}/Z$ is the unbiased distribution. The parameter $\gamma = \beta/\beta' = (T+\Delta T)/T$ is the so-called bias factor, and as it is increased $p(\mathbf{s})$ becomes flatter until for $\gamma \to \infty$ the uniform distribution is obtained.

The well-tempered distribution can be viewed as a distribution where the CVs are effectively sampled at some higher, but finite, temperature $T+\Delta T$, while $F(\mathbf{s})$ still remains the FES at temperature T. It can also be viewed as sampling on an effective FES $\tilde{F}(\mathbf{s}) = (1/\gamma)F(\mathbf{s})$ which has largely the same metastable states as the original $F(\mathbf{s})$ but with free energy barriers that have been reduced by γ.

The well-tempered distribution as defined above in eq 10 cannot be used directly in the variational approach since $F(\mathbf{s})$ is not known a priori. Thus, to solve this we propose an iterative scheme where we build up an estimate of the well-tempered distribution during the optimization process. Once the target distribution has converged to the well-tempered distribution the relation between the FES and bias potential is $V(\mathbf{s}) = -(1-\gamma^{-1})F(\mathbf{s})$ which is the same relation that asymptotically holds in well-tempered metadynamics.[9,10]

In this iterative scheme we update the target distribution as

$$p^{(k+1)}(\mathbf{s}) = \frac{\exp[-\beta' F^{(k+1)}(\mathbf{s})]}{\int d\mathbf{s}\, \exp[-\beta' F^{(k+1)}(\mathbf{s})]} \tag{11}$$

where $F^{(k+1)}(\mathbf{s})$ is the current estimate of the FES obtained from eq 3

$$\begin{aligned} F^{(k+1)}(\mathbf{s}) &= -V^{(k)}(\mathbf{s}) - (1/\beta)\log p^{(k)}(\mathbf{s}) \\ &= -V^{(k)}(\mathbf{s}) + (1/\gamma)F^{(k)}(\mathbf{s}) \end{aligned} \tag{12}$$

where $V^{(k)}(\mathbf{s})$ is the bias potential obtained using the previous target distribution $p^{(k)}(\mathbf{s})$ (the constant can be ignored as the target distribution is normalized in eq 11). Usually the initial target distribution is taken to be uniform $p^{(0)}(\mathbf{s}) = C$ which amounts to taking $F^{(0)}(\mathbf{s}) = 0$. We can also start with a better initial guess, for example from a close-by temperature or a different level of theory.

The only parameter in this iterative scheme is the updating stride Δt_{WT} that specifies how many bias potential iterations should be performed between updating the target distribution, and this parameter is discussed in further detail in Section S2 in the Supporting Information (SI). Note that when we combine this iterative scheme with the stochastic optimization algorithm from the previous Section the averaging of $\bar{\alpha}^{(n)}$ is done in an continuous manner (i.e., the averaging is not restarted when updating the target distribution).

The expectation values $\langle \cdots \rangle_p$ have to be recalculated each time the target distribution is updated. Generally it is most convenient to do this by numerical integration. In cases where numerical integration might be computationally expensive or unfeasible an alternative option could be to calculate the expectation values by employing Monte Carlo sampling to generate points distributed according to $p^{(k+1)}(\mathbf{s})$.

As discussed in Section S1 in the SI, another option for updating the well-tempered target distribution would be to average over the FESs obtained from eq 11 which should lead to smoother changes in the expectation values $\langle \cdots \rangle_p$. The performance of this averaged option is evaluated in Section S2 in the SI where it is observed to be a feasible option although exhibiting a slightly slower convergence rate as compared to the direct option (eqs 11 and 12) employed here.

3. COMPUTATIONAL DETAILS

3.1. Alanine Tetrapeptide. We consider as a benchmark system alanine tetrapeptide (Ace-Ala_3-Nme), both in vacuum, like we did in our previous publication,[8] and solvated in water. Unless explicitly stated all computational details given in this Section refer both to the simulations in vacuum and water.

The configurational space of alanine tetrapeptide is described in terms of the six backbone dihedral angles indicated in Figure 1 (Φ_1,Ψ_1, Φ_2,Ψ_2, Φ_3,Ψ_3). The CVs employed to bias the system are the three Φ angles $\mathbf{s} = (\Phi_1,\Phi_2,\Phi_3)$ as they are sufficient to obtain proper sampling of phase space for the force field employed here, both in vacuum and water. In other words, the Ψ angles are fast degrees of freedom as compared to the Φ angles.

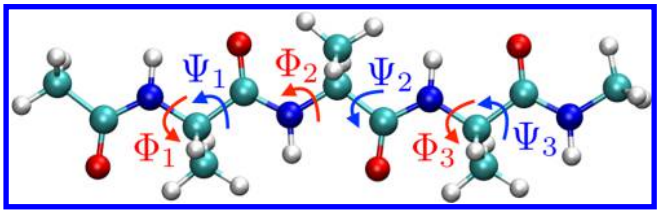


Figure 1. Alanine tetrapeptide (Ace-Ala_3-Nme) along with the backbone dihedral angles used to describe it. The Φ_1, Φ_2, and Φ_3 angles are the CVs used to bias the system.

3.2. Molecular Dynamics Parameters. All MD simulations are performed using the Gromacs 4.5.5[19] MD code, patched with a private development version of the PLUMED 2[20] enhanced sampling plug-in. The alanine tetrapeptide is described using the Amber99-SB[21] force field, and the TIP3P[22] water model is used for the solvent simulations. For all simulations we use a time-step of 0.002 ps and constrain the bonds involving hydrogens using the LINCS[23] algorithm. All simulations are performed at a constant temperature of 300 K using the stochastic velocity rescaling thermostat[24] with a relaxation time of 0.1 ps. The simulations for the vacuum system are performed without periodic boundary conditions and without cut-offs for the electrostatic and nonbonded van der Waals interactions. The simulations for the solvated system are performed in the NVT ensemble using a periodic cubic box of side length 3.327 nm that included 1200 TIP3P water molecules. The long-range electrostatic interactions are handled

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

using the Particle-Mesh Ewald[25] method with a real space cutoff of 0.95 nm. We employ a cutoff of 0.95 nm for the nonbonded van der Waals interactions.

3.3. Basis Set for Bias Potential. The bias potential is expanded in a Fourier series, most compactly written as

$$V(\mathbf{s};\boldsymbol{\alpha}) = \sum_{\mathbf{n}=-\mathbf{N}_{\max}}^{\mathbf{N}_{\max}} \alpha_{\mathbf{n}} \exp(i\mathbf{n}\mathbf{s}) \tag{13}$$

where $\mathbf{s} = (\Phi_1,\Phi_2,\Phi_3)$ are the CVs that are periodic and defined in range $[-\pi,\pi]$, $\mathbf{n} = (n_1,n_2,n_3)$ is the order of the basis function, and $\mathbf{N}_{\max} = (N_{\max},N_{\max},N_{\max})$ is the maximum order of the basis set expansion. The overall constant in the expansion is dropped, $\alpha_0 = 0$. The parameter $N_{\max}$ is used to indicate the size of the expansion, and the total number of basis functions is $(2{\cdot}N_{\max}{+}1)^3 - 1$. In practice the basis functions are taken to be cosines and sines, and all coefficients are real.

Below we show results obtained with basis set expansions of size $N_{\max} = 3$ (342 basis functions), $N_{\max} = 6$ (2196 basis functions), or $N_{\max} = 10$ (9261 basis functions). The minimal $N_{\max} = 3$ expansion does not have sufficient variational flexibility to fully describe the finer features of the $F(\Phi_1,\Phi_2,\Phi_3)$ surface but nevertheless gives a quite reasonable description of the main features (see Figure S8 in the SI and also ref 8). Full variational flexibility is obtained with the larger $N_{\max} = 6$ expansion that can completely describe the $F(\Phi_1,\Phi_2,\Phi_3)$ surface. Finally we employ the largest $N_{\max} = 10$ expansion to check the robustness of the convergence with respect to noise as it has considerably more basis functions than are really needed to describe the FES.

3.4. Parameters for the Variational Approach. For all simulations the sampling time for each iteration in the optimization is 1.0 ps (500 timesteps), and the step size is $\mu = 0.08$ kJ/mol. All calculations are started with all coefficients set to zero ($\bar{\boldsymbol{\alpha}}^{(0)} = \boldsymbol{\alpha}^{(0)} = \mathbf{0}$).

For the iterative scheme in Section 2.3, the well-tempered target distribution is updated every $\Delta t_{WT} = 500$ ps (or 500 bias potential iterations). As discussed in Section S2 of the SI, the performance of the iterative scheme appears to be insensitive to how often the target distribution is updated as employing values of Δt_{WT} from 50 to 1000 ps results in very similar convergence rates. The expectation values $\langle\cdots\rangle_p$ are calculated using numerical integration on a $100 \times 100 \times 100$ equally spaced grid. All other parameters are specified as the results are presented.

3.5. Error Metric for Free Energy Surfaces. To assess the convergence of $F(\mathbf{s})$ we employ the error metric used previously in refs 9 and 26. This metric measures the deviation of a given $F(\mathbf{s})$ from a reference $F_r(\mathbf{s})$ and is defined as

$$\varepsilon = \sqrt{\frac{\int d\mathbf{s}\,[\bar{F}_r(\mathbf{s}) - \bar{F}(\mathbf{s})]^2\theta(\nu - F_r(\mathbf{s}))}{\int d\mathbf{s}\,\theta(\nu - F_r(\mathbf{s}))}} \tag{14}$$

where the Heaviside step function $\theta(\nu - F_r(\mathbf{s}))$ limits the comparison to a reference region that is defined by all points that lie within ν of the minimum of $F_r(\mathbf{s})$. The FESs are rigidly shifted by the average value in the reference region:

$$\bar{F}_r(\mathbf{s}) = F_r(\mathbf{s}) - \frac{\int d\mathbf{s}\,F_r(\mathbf{s})\theta(\nu - F_r(\mathbf{s}))}{\int d\mathbf{s}\,\theta(\nu - F_r(\mathbf{s}))} \tag{15}$$

$$\bar{F}(\mathbf{s}) = F(\mathbf{s}) - \frac{\int d\mathbf{s}\,F(\mathbf{s})\theta(\nu - F_r(\mathbf{s}))}{\int d\mathbf{s}\,\theta(\nu - F_r(\mathbf{s}))} \tag{16}$$

The metric ε is calculated using the projection $F(\Phi_1,\Phi_2) = -(1/\beta)\log\int d\Phi_3 \exp[-\beta F(\Phi_1,\Phi_2,\Phi_3)]$ and a cutoff of $\nu = 20$ kJ/mol $\approx 8\ k_BT$. Very similar results are obtained for ε when using a lower cutoff of $\nu = 10$ kJ/mol or the other two projections $F(\Phi_1,\Phi_3)$ and $F(\Phi_2,\Phi_3)$.

Both for vacuum and in water the reference $F_r(\Phi_1,\Phi_2,\Phi_3)$ is taken from a 400 ns simulation with the variational approach using the $N_{\max} = 6$ expansion and the well-tempered target distribution with a bias factor of $\gamma = 10$. These FESs are in excellent agreement with results obtained from independent simulations and reference results obtained with other methods (see Sections S6 and S7 in the SI).

Note that it is equivalent to talk about convergence of the FES and the bias potential as $F(\mathbf{s})$ is always estimated from $V(\mathbf{s})$ according to eq 3.

4. RESULTS

At first we consider alanine tetrapeptide in vacuum and start by briefly examining the FES convergence when employing uniform sampling and show in Figure 2 results obtained with

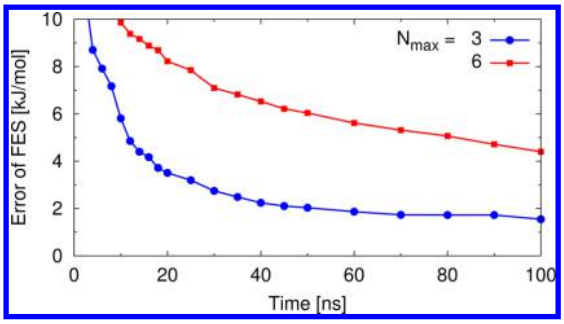


Figure 2. Convergence of the FES for alanine tetrapeptide in vacuum obtained using the uniform target distribution and basis set expansions $N_{\max} = 3$ (blue line) and $N_{\max} = 6$ (red line). See also Figure S4 in the SI for other basis set expansions not shown here. The convergence is evaluated using the ε error metric (eq 14).

the $N_{\max} = 3$ and $N_{\max} = 6$ basis set expansion. For the minimal $N_{\max} = 3$ expansion the bias potential converges after around 50 ns, and we obtain a reasonable description of the main features of the $F(\Phi_1,\Phi_2,\Phi_3)$ surface (see ref 8 for a more extended discussion of the results obtained using the uniform target distribution and the minimal $N_{\max} = 3$ expansion). For the larger $N_{\max} = 6$ expansion the convergence behavior is however considerably worse as the bias potential does not converge within the allocated time of 100 ns. As observed in Figure S4 in the SI, the convergence behavior generally degrades as the basis set size is increased and can be considered unsatisfactory for all basis set expansions higher than $N_{\max} = 3$.

We will now turn our attention to the use of the well-tempered target distribution and show in Figure 3 results obtained with $N_{\max} = 3$, $N_{\max} = 6$, and $N_{\max} = 10$ expansions. Here the bias factor is taken to be $\gamma = 10$, while other values are considered below. We can clearly observe that the convergence is greatly improved when employing the well-tempered target distribution, with a full convergence achieved within around 10 ns of simulation time. We also observe that the rate of convergence is completely basis set size independent, and the

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

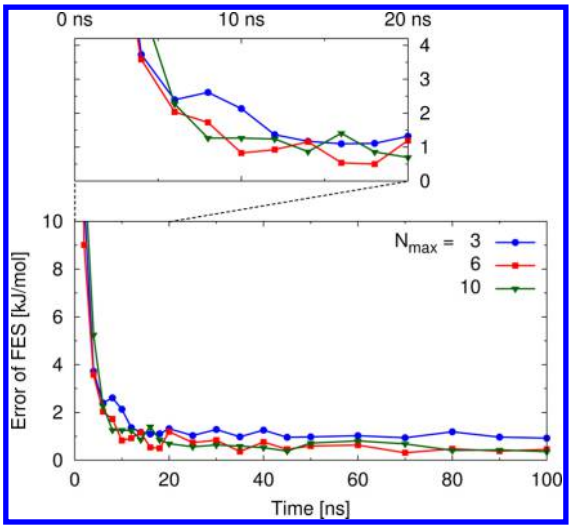


Figure 3. Convergence of the FES for alanine tetrapeptide in vacuum obtained using the well-tempered target distribution with a bias factor of $\gamma = 10$ and basis set expansions $N_{max} = 3$ (blue line), $N_{max} = 6$ (red line), and $N_{max} = 10$ (green line). See also Figure S5 in the SI for other basis set expansions not shown here. In the upper panel we show a blow-up of the first 20 ns. The convergence is evaluated using the ε error metric (eq 14).

robustness with respect to noise is clearly indicated by the more stringent test of $N_{max} = 10$ expansion that includes far more basis functions than are needed to fully describe the FES (see also Figure S5 in the SI for other basis set expansions not shown here). The converge rate is also completely independent of the initial conditions as observed in Figure S3a in the SI.

To show how the choice of the target distribution affects the sampling of CV space do we show in Figure 4 the (Φ_1,Φ_2) values sampled during the first 40 ns of the simulations in Figures 2 and 3. We can clearly see in Figure 4b that the well-tempered target distribution focuses sampling toward the free energy minima in the $F(\Phi_1,\Phi_2)$ surface and generally limits the sampling to the lowest 60 kJ/mol of the FES. On the other hand the sampling for the uniform target distribution shown in Figure 4a is much more homogeneous, as expected. The poor convergence observed for the uniform target distribution is apparent in the fact that after 40 ns we still have not started to sample the highest free energy regions of the FES.

As discussed in Section 2.3, the well-tempered distribution can be viewed as sampling on an effective FES $\tilde{F}(\mathbf{s}) = (1/\gamma)F(\mathbf{s})$ that has roughly the same metastable states as the original $F(\mathbf{s})$ but with reduced barriers. Thus, instead of trying to sample whole CV space evenly like in the uniform distribution, the well tempered distribution focuses sampling on the relevant low free energy regions, as seen in Figure 4. This is possibly the reason for the improvement in convergence that we observe here. The bias factor γ in the well-tempered distribution determines the extent of the free energy exploration and by how much the barriers are effectively reduced. In general, one would like to employ a bias factor which makes the effective barriers to be on the order of k_BT such that thermal fluctuations can easily drive the system between different metastable states. Larger bias factors will result in an more extensive free energy exploration but normally at the cost of a slower convergence. For very large bias factors the barriers are reduced by so much that the well-tempered distribution will effectively be the uniform distribution.

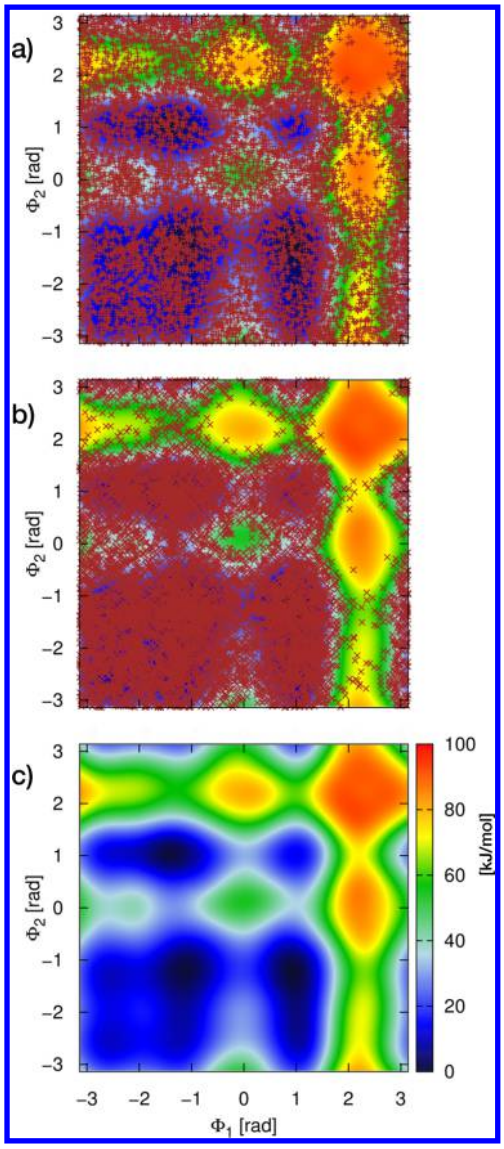


Figure 4. Brown points indicate (Φ_1,Φ_2) values sampled every 4 ps during the first 40 ns of the $N_{max} = 6$ simulations shown in Figures 2 and 3: a) results obtained employing the uniform target distribution; b) results obtained employing the well-tempered target distribution with a bias factor of $\gamma = 10$. The underlying $F(\Phi_1,\Phi_2)$ surface is shown in panel c) and also in the background of panels a) and b). The same color scale (kJ/mol) is used for the FES in all panels.

In the current case the free energy barriers are around 8−16 k_BT, so we can expect the optimal bias factor to be around 10. This is exactly what we observe in Figure 5 where we show results obtained using bias factors from $\gamma = 7$ to $\gamma = 40$. Employing bias factors $\gamma = 7$ and $\gamma = 10$ results in a similar convergence rate, while the convergence is slightly slower when employing a bias factor of $\gamma = 20$. Increasing the bias factor to $\gamma = 40$ then results in considerable slower convergence as expected.

The variational approach appears to be sensitive to choosing too low bias factors that do not reduce the barriers sufficiently. For example, employing a bias factor of 5 and the same computational setup as in Figure 5 generally resulted in an unstable optimization process as the system tended to get stuck in one of the metastable states. The most effective way to solve this seems to be to use multiple walkers as just using two

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

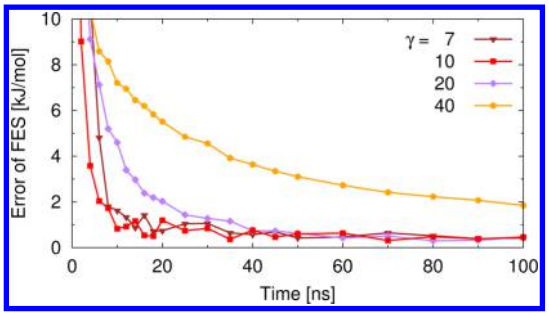


Figure 5. Convergence of the FES for alanine tetrapeptide in vacuum obtained using the well-tempered target distribution and different bias factors γ from 7 to 40. The results are obtained using a N_{max} = 6 expansion. The convergence is evaluated using the ε error metric (eq 14).

walkers resulted in a stable optimization process. Another option is to reduce the step size μ in the optimization. Furthermore, problems with a too low bias factor should be much less severe when the variational approach is employed in conjunction with replica exchange methods like parallel tempering[27−29] or bias exchange.[30]

We finally consider alanine tetrapeptide in water to explore the performance when including the additional complexity of an explicit solvent. In Figure S6 in the SI we show results obtained using the uniform target distribution, while the results obtained using the well-tempered target distribution with a bias factor of γ = 10 are shown in Figure 6. We observe that the

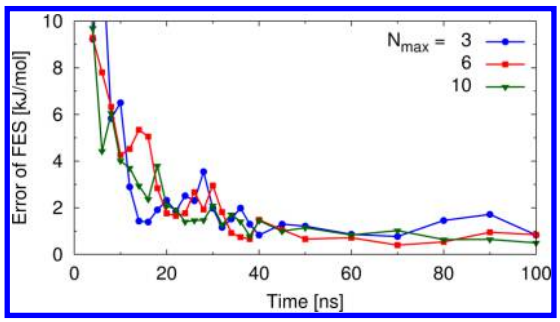


Figure 6. Convergence of the FES for alanine tetrapeptide in water obtained using the well-tempered target distribution with a bias factor of γ = 10 and basis set expansions N_{max} = 3 (blue line), N_{max} = 6 (red line), and N_{max} = 10 (green line). The convergence is evaluated using the ε error metric (eq 14).

presence of the solvent possesses no problems as the behavior is essentially the same as before. Namely, poor convergence for the uniform target distribution while the well-tempered target distribution results in a greatly improved convergence. As for the vacuum case the robustness with respect to noise is indicated by the fact that the rate of convergence is basis set size independent when employing the well-tempered target distribution. The only apparent difference is that the rate of convergence for the well-tempered target distribution is slightly slower than for the vacuum case as can be expected when we add the additional complication of a solvent environment.

5. DISCUSSION AND CONCLUSIONS

We have here introduced a simple but effective iterative scheme that allows us to use the well-tempered distribution as a target distribution in our recent variational approach[8] and evaluated its performance on the three-dimensional FES of alanine tetrapeptide.

The use of the well-tempered target distribution results in a significant improvement in convergence as compared to the uniform target distribution. This is most likely due to the fact that the well-tempered distribution focuses the sampling toward most relevant regions low in the free energy and does not waste time trying to flatten the sampling in the irrelevant regions high in free energy.

Note that convergence problems like the ones observed here are not completely universal. For example, for alanine dipeptide, the classical benchmark system for free energy methods, we did not observe any convergence problems when using the uniform target distribution in our previous publication,[8] both for simulations in vacuum and in water. Furthermore, we observed no major improvements in convergence for alanine dipeptide when employing the well-tempered target distribution. However, it is important to note that the convergence rate did not deteriorate either when using the well-tempered target distribution.

Given improvements in convergence observed in this paper we do expect that there are generally significant computational benefits in employing the well-tempered distribution as a target distribution in the variational approach. This choice will also efficiently solve the issue of having to employ restraints on the CVs to avoid sampling of irrelevant (and perhaps unphysical) regions of phase space as might be needed when employing the uniform target distribution. Therefore, in most cases the well-tempered distribution will be the recommended and the preferred choice for the target distribution in the variational approach.

That we can now use the well-tempered target distribution with the variational approach provides many possibilities. For example, the variational approach can be employed to realize the well-tempered ensemble,[31] that has proven very useful in reducing the number of replicas needed for parallel tempering simulation of solvated protein systems[32] or for obtaining the thermodynamic behavior of phase-transitions.[33] Another possibility is to use the variational approach in place of well-tempered metadynamics in the recently introduced replica exchange with collective-variable tempering method.[34]

If we are missing some slow degrees of freedom from our CV set, then the performance of variational approach and iterative scheme presented here will very likely be rather poor due to the resulting hysteresis behavior (see the Supporting Information of ref 8). One option for circumventing this issue is to combine the variational approach with parallel-tempering, as has proven very successful for metadynamics,[29] or some other variant of replica exchange.

Here we have used the iterative scheme in conjunction with the stochastic optimization method from ref 14. However, this iterative scheme can be used in combination with any other optimization scheme where comparable improvements in convergence can be expected. The iterative scheme can furthermore be used to update other target distributions that depend on some quantities that are a priori unknown but can be estimated during the simulation. For example, one of the potential options one could explore in the future is to employ the iterative scheme to achieve the diffusion optimized distribution,[35] $p(\mathbf{s}) \propto (1)/(D(\mathbf{s}))^{1/2}$ where $D(\mathbf{s})$ is the $\mathbf{s}$-dependent local diffusivity that can be estimated by employing methods available in the literature.[35−44] The efficiency of this choice will have to be assessed in future publications.

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002

■ ASSOCIATED CONTENT

Supporting Information

Discussion on the averaged update option for the well-tempered target distribution. Evaluation of computational options for updating the well-tempered target distribution. Additional convergence results. Reference free energy surfaces used for the ϵ error metric. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00076.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: omar.valsson@phys.chem.ethz.ch.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Patrick R. Shaffer for useful discussions and for reading over the manuscript. The authors acknowledge funding from the National Center for Computational Design and Discovery of Novel Materials MARVEL and the European Union Grant No. ERC-2009-AdG-247075. All calculations were performed on the Brutus HPC cluster at ETH Zurich.

■ REFERENCES

(1) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187−199.
(2) Huber, T.; Torda, A. E.; Gunsteren, W. F. *J. Comput.-Aided Mol. Des* **1994**, *8*, 695−708.
(3) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169.
(4) Wang, F.; Landau, D. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(5) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(6) Hansmann, U.; Wille, L. *Phys. Rev. Lett.* **2002**, *88*, 068105 .
(7) Maragakis, P.; van der Vaart, A.; Karplus, M. *J. Phys. Chem. B* **2009**, *113*, 4664−4673.
(8) Valsson, O.; Parrinello, M. *Phys. Rev. Lett.* **2014**, *113*, 090601.
(9) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(10) Dama, J. F.; Parrinello, M.; Voth, G. A. *Phys. Rev. Lett.* **2014**, *112*, 240602.
(11) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 094112.
(12) Bilionis, I.; Koutsourelakis, P. *J. Comput. Phys.* **2012**, *231*, 3849−3870.
(13) Kushner, H. J.; Yin, G. G. *Stochastic Approximation and Recursive Algorithms and Applications*; Springer-Verlag: New York, 2003.
(14) Bach, F.; Moulines, E. In *Advances in Neural Information Processing Systems 26*; Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, 2013; pp 773−781.
(15) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533−3539.
(16) VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 203−208.
(17) Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. E. *J. Chem. Phys.* **2002**, *116*, 4389.
(18) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168−175.
(19) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. *Bioinformatics* **2013**, *29*, 845−854.
(20) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(23) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(24) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(26) Branduardi, D.; Bussi, G.; Parrinello, M. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.
(27) Hansmann, U. H. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(28) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(29) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(30) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(31) Bonomi, M.; Parrinello, M. *Phys. Rev. Lett.* **2010**, *104*, 190601.
(32) Deighan, M.; Bonomi, M.; Pfaendtner, J. *J. Chem. Theory Comput.* **2012**, *8*, 2189−2192.
(33) Valsson, O.; Parrinello, M. *J. Chem. Theory Comput.* **2013**, *9*, 5267−5276.
(34) Gil-Ley, A.; Bussi, G. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.
(35) Trebst, S.; Huse, D.; Troyer, M. *Phys. Rev. E* **2004**, *70*, 046701.
(36) Hummer, G. *New J. Phys.* **2005**, *7*, 34−34.
(37) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *107*, 1088−1093.
(38) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13841−13846.
(39) Krivov, S. V. *J. Phys. Chem. B* **2011**, *115*, 11382−11388.
(40) Krivov, S. V. *J. Phys. Chem. B* **2011**, *115*, 12315−12324.
(41) Singh, S.; Chiu, C.-c.; de Pablo, J. J. *J. Stat. Phys.* **2011**, *145*, 932−945.
(42) Singh, S.; Chiu, C.-C.; de Pablo, J. J. *J. Chem. Theory Comput.* **2012**, *8*, 4657−4662.
(43) Jiang, P.; Yaşar, F.; Hansmann, U. H. E. *J. Chem. Theory Comput.* **2013**, *9*, 3816−3825.
(44) Tian, P.; Jónsson, S. Æ.; Ferkinghoff-Borg, J.; Krivov, S. V.; Lindorff-Larsen, K.; Irbäck, A.; Boomsma, W. *J. Chem. Theory Comput.* **2014**, *10*, 543−553.

DOI: 10.1021/acs.jctc.5b00076
J. Chem. Theory Comput. 2015, 11, 1996−2002