The Adaptive Biasing Force Method: Everything You Always Wanted To Know, But Were Afraid To Ask

Supporting Information

Jeffrey Comer,[†] James C. Gumbart,[‡] Jérôme Hénin,[¶] Tony Lelièvre,[§] Andrew Pohorille,[∥,⊥] and Christophe Chipot[*,†,#]

Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana-Champaign, Unité Mixte de Recherche CNRS n°7565, Université de Lorraine, B.P. 70239, 54506 Vandœuvre-lès-Nancy cedex, France, School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, United States, Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, 75005 Paris, France, Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique, École des Ponts ParisTech, 6 et 8, Avenue Blaise Pascal, Cité Descartes — Champs sur Marne, 77455 Marne la Vallée Cedex 2, France, and NASA Ames Research Center Exobiology Branch, Mail Stop 239-4, Moffett Field, California, 94035-1000

E-mail: chipot@ks.uiuc.edu

Justification of a stratification strategy: Diffusion timescale

In the canonical ensemble, a system of dimension d is equipped with the Boltzmann-Gibbs probability measure, i.e. the canonical measure[1]

$$\mu(\mathrm{d}\mathbf{x}) = Z^{-1}\ \exp[-\beta V(\mathbf{x})]\ \mathrm{d}\mathbf{x}, \tag{S1}$$

where $\mathbf{x} \in \mathbb{R}^d$ is the system configuration, $V : \mathbb{R}^d \to \mathbb{R}$ is the potential energy function and $Z = \int_{\mathbb{R}^d} \mathrm{d}\mathbf{x}\ \exp[-\beta V(\mathbf{x})]$ is the normalization constant or the so-called partition function. To sample this measure, one can use the overdamped Langevin dynamics,

$$\begin{cases} \mathrm{d}\mathbf{x}_t = -\nabla(V - A_t \circ \xi)(\mathbf{x}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \\ A_t'(z) = \langle F_\xi(\mathbf{x}_t) | \xi(\mathbf{x}_t) = z \rangle, \end{cases} \tag{S2}$$

where $A_t \circ \xi$ denotes the composition of A_t with ξ, so that $A_t \circ \xi(x) = A_t[\xi(x)]$, and A_t' is the estimated mean force, i.e. $-\nabla V + (A_t' \circ \xi)\nabla\xi$. The dynamics can be viewed as a standard overdamped Langevin dynamics associated with the time varying potential $V - A_t \circ \xi$. The law of the process $\mathbf{x}_t$ will be denoted by $\psi_t(\mathbf{x})\mathrm{d}\mathbf{x}$.

Let us consider a geometrical transformation carried over a reaction pathway of length $\mathscr{L}$ and assume that convergence of the free energy is achieved within t_0. Let us further consider that the reaction path can be broken down in a series of N non-overlapping windows of lengths $\mathscr{L}_1, \ldots, \mathscr{L}_N$, for which convergence is attained after $t_1', \ldots, t_N'$ — continuity of the free-energy

*To whom correspondence should be addressed

†Université de Lorraine

‡Georgia Institute of Technology

¶Institut de Biologie Physico-Chimique

§École des Ponts ParisTech

‖NASA Ames

⊥University of California San Francisco, San Francisco, California, 94143-2280

#Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Engineering, University of Illinois at Urbana-Champaign, 405 North Mathews, Urbana, Illinois 61801 and Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801

gradient in the adaptive biasing force algorithm obviates the need for overlapping windows. We want to demonstrate that $t_0 > \sum_i t'_i$. For all intents and purposes, let us drop the drift part of the adaptive biasing force equation and only retain the Brownian motion of the latter. Let us further assume that the probability distribution of configurations, $\psi(t,x)$, follows,

$$|\psi(t,x) - \psi_\infty(x)| \propto \exp[-\lambda(\mathscr{L})t] \tag{S3}$$

If we compare the free-energy profile at some value of the transition coordinate obtained from a single simulation and from a stratification scheme, the difference is equal to some tolerance, ϵ. Typically,

$$\exp[-\lambda(\mathscr{L}_1)t'_1] = \epsilon \tag{S4}$$

which can be rewritten as $\lambda(\mathscr{L}_1)t'_1 = \epsilon' = -\ln\epsilon$, or, in an equivalent form, as $t'_1 = \dfrac{\epsilon'}{\lambda(\mathscr{L}_1)}$.

More generally, $\sum_i t'_i = \epsilon' \sum_i \dfrac{1}{\lambda(\mathscr{L}_i)}$ and $t_0 = \dfrac{\epsilon}{\lambda(\sum_i \mathscr{L}_i)}$. From the above, it is rather obvious that if $\mathscr{L}$ increases, λ should decrease accordingly. For the sake of argument, let us now choose a law where λ varies with $1/\mathscr{L}^2$ and consider the Fokker-Planck equation for pure diffusion (i.e. in the absence of drift), $\partial_t\psi = \Delta\psi$.

When t tends towards infinity, ψ is expected to be constant. We can write the probability distribution of configurations at time t as $\psi(t,x) = \psi = \sum_k \psi_k(x)A_k(t)$, where $\psi_k(x)$ is a solution of $\Delta\psi_k = -\lambda_k\psi_k$. From the latter, it follows that $\sum_k \dfrac{\mathrm{d}A_k}{\mathrm{d}t} = -\sum_k A_k\lambda_k\psi_k$, which are independent vectors. From $\mathrm{d}A_k/\mathrm{d}t = -A_k\lambda_k$, one can infer the solutions $A_k(t) = A_k(0)\exp(-\lambda_k t)$. These solutions are valid over the interval $[0;\mathscr{L}]$.

As an example, let us consider $\psi_k(x) = \sin(k\pi/\mathscr{L})x$. The second derivative of $\psi_k(x)$ is $-(k\pi/\mathscr{L})^2\sin(k\pi/\mathscr{L})x$ and $\lambda_k = (k\pi/\mathscr{L})^2$. For $k=1$, the eigenvalue is $(\pi/\mathscr{L})^2$, which is in line with our law for λ in $1/\mathscr{L}^2$. Over the series of sequential windows, $(\sum_i \mathscr{L}_i)^2 = (\mathscr{L}_1 + \mathscr{L}_2 + \cdots)^2 +$ cross-terms, which is greater than $\mathscr{L}_1^2 + \mathscr{L}^2 + \cdots$. It follows that $(\sum_i \mathscr{L}_i)^2$ is always

greater than $(\sum_i \mathscr{L}_i^2)$. The time required for convergence in a single run is $t_0 = \epsilon'(\sum_i \mathscr{L}_i)^2$, and $\sum_i t_i = \epsilon' \sum_i \mathscr{L}_i^2$. As a result, $t_0 > \sum_i t_i'$.

The adaptive biasing force algorithm formally converges

In this section, we would like to study the longtime convergence of the adaptive biasing force algorithm. We restrict ourselves to the following simple setting. We consider the overdamped Langevin dynamics,

$$\mathrm{d}\mathbf{x}_t = -\nabla V(\mathbf{x}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t \tag{S5}$$

where $\mathbf{x}_t$ lives in the N-dimensional torus $\mathbb{T}^N$ (namely in $[0,1]^N$ with periodic boundary conditions) and $\xi(x_1,\ldots,x_N) = x_1$. We direct the reader to references 2,3 for extensions to more general situations of the results presented below.

Starting from equation (S5) and using the above choice of the transition coordinate ξ, the adaptive biasing force dynamics writes,

$$\begin{cases} \mathrm{d}\mathbf{x}_t = -\nabla V(\mathbf{x}_t) + A_t'(X_t^1)e_1\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t \\ A_t'(x_1) = \mathbb{E}(\partial_1 V(\mathbf{x}_t)|X_t^1 = x_1) \end{cases} \tag{S6}$$

where X_t^1 denotes the first coordinate of the vector $\mathbf{x}_t$, e_1 is the vector with coordinates $(1,0,\ldots,0)$ and $\partial_1 V$ denotes the partial derivative of $V(x_1,\ldots,x_N)$ with respect to x_1. We would like to understand under which assumptions the adaptive biasing force dynamics (S6) indeed converges faster to equilibrium than the original unbiased dynamics (S5).

One way to understand this convergence is to look at the way the law of $\mathbf{x}_t$ evolves. Let us denote $\psi(t,x)$ the density of $\mathbf{x}_t$. For the original dynamics (S5), the density ψ satisfies the Fokker-Planck equation,

$$\partial_t \psi = \mathrm{div}\,(\nabla V \psi + \beta^{-1}\nabla\psi). \tag{S7}$$

For the adaptive biasing force dynamics, the density ψ satisfies

$$\begin{cases} \partial_t \psi = \operatorname{div}\left(\nabla V \psi + \beta^{-1} \nabla \psi\right) - \partial_1 (A_t'(x_1)\psi) \\ A_t'(x_1) = \dfrac{\displaystyle\int \partial_1 V \psi \, \mathrm{d}x_2 \ldots \mathrm{d}x_N}{\displaystyle\int \psi \, \mathrm{d}x_2 \ldots \mathrm{d}x_N} \end{cases} \tag{S8}$$

It ought to be noted that equation (S8) is a nonlinear partial differential equation (PDE), which makes the study of its longtime behavior much more complicated than for the linear Fokker-Planck PDE (S7).

To study the longtime behaviors of these densities, we need a mathematical tool — Entropy techniques and logarithmic Sobolev inequalities.[4] The convergence rate is computed using the relative entropy of the law at time t with respect to the equilibrium measure. As a preliminary step, we need to introduce three definitions. Let ψ and ψ_∞ denotes two densities of probability measures. First, the *entropy of ψ with respect to ψ_∞* is defined by $H(\psi|\psi_\infty) = \displaystyle\int \ln\left(\frac{\psi}{\psi_\infty}\right)\psi$. We recall that the entropy is a non-negative quantity which can be seen as a measure of the distance between two probability measures since, for all probability densities ψ and ψ_∞, $\displaystyle\int |\psi - \psi_\infty| \leq \sqrt{2H(\psi|\psi_\infty)}$. Second, *the Fisher information of ψ with respect to ψ_∞* (we invite the reader to check reference 4 for more on this subject) is defined by $I(\psi|\psi_\infty) = \displaystyle\int \left|\nabla \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi$. Third, the measure $\psi_\infty(\mathbf{x})\, d\mathbf{x}$ is said to satisfy *the Logarithmic Sobolev inequality* with constant ρ (in short LSI(ρ)) if, for any probability density ψ, $H(\psi|\psi_\infty) \leq \frac{1}{2\rho} I(\psi|\psi_\infty)$.

The Logarithmic Sobolev inequality constant can be seen as a way to measure the multimodality of a probability measure, or the metastability of the associated overdamped Langevin dynamics, as illustrated by the following result. Let ψ satisfy the linear Fokker-Planck equation (S7). Then, the two following statements are equivalent,

- The probability density $Z^{-1}\exp[-\beta V(\mathbf{x})]$ satisfies a LSI(R).
- The probability density ψ satisfies: for any initial condition ψ_0, for all $t \geq 0$, $H(\psi(t,\cdot)|\psi_\infty) \leq H(\psi_0|\psi_\infty)\exp(-2\beta^{-1}Rt)$.

The original overdamped Langevin dynamics (S5) thus converges to equilibrium at a rate $\beta^{-1}R$, R being the logarithmic Sobolev inequality constant associated with the canonical measure $Z^{-1}\times \exp[-\beta V(\mathbf{x})]\,\mathrm{d}\mathbf{x}$.

Concerning the adaptive biasing force dynamics, one expects an exponential convergence to the equilibrium density $\psi_\infty = \tilde{Z}^{-1}\exp[-\beta(V - A\circ\xi)]$. The following can be proven.[2] Let us make the following assumptions,

(1) Bounded coupling: $\forall j \in \{2,\dots,N\}$, $\|\partial_{1,j}V\|_{L^\infty(\mathbb{T}^N)} < \infty$.

(2) LSI for the conditional measures: there exists $\rho > 0$ such that for all $x_1 \in \mathbb{T}$, the conditional probability densities $(x_2,\dots,x_N) \mapsto \dfrac{\exp[-\beta V(x_1,x_2,\dots x_N)]}{\displaystyle\int \mathrm{d}x_2\dots\mathrm{d}x_N\ \exp[-\beta V(x_1,x_2,\dots x_N)]}$ satisfy LSI(ρ).

Then, $\sqrt{H(\psi(t,\cdot)|\psi_\infty)}$ and $\|A'_t - A'\|_{L^2(\mathbb{T})}$ converge exponentially fast to zero with rate (at least) $\beta^{-1}\min(4\pi^2,\rho)$. The upper bound $4\pi^2$ to the rate of convergence corresponds to the rate of convergence of a diffusion to equilibrium on the torus $\mathbb{T}$. Typically, one expects $\rho < 4\pi^2$.

In summary, going from the standard overdamped Langevin dynamics (S5) to the adaptive biasing force dynamics (S6), the rate of convergence is changed from $\beta^{-1}R$ to $\beta^{-1}\rho$, where R denotes the LSI constant associated with the canonical measure $Z^{-1}\exp(-\beta V(\mathbf{x}))\,\mathrm{d}\mathbf{x}$ and ρ the LSI constant associated with the conditioned canonical measures

$$(x_2,\dots,x_N) \mapsto \frac{\exp[-\beta V(x_1,x_2,\dots x_N)]}{\displaystyle\int \exp[-\beta V(x_1,x_2,\dots x_N)]\mathrm{d}x_2\dots\mathrm{d}x_N}\,\mathrm{d}x_2,\dots,\mathrm{d}x_N.$$

If ξ is well chosen, one typically expects ρ to be much larger than R. This is typically the case for the simple illustrative two-dimensional potentials in Figures 1 and 2 of the article, if $\xi(x_1,x_2) = x_1$.

This gives actually a way to measure the quality of the transition coordinate: The transition coordinate, ξ, is good if the ratio ρ/R of the LSI constants of the conditioned canonical measures $\mu(\cdot|\xi(\mathbf{x})) = z$ over the LSI constant of the canonical measure μ, is large.

The proof of the longtime convergence of the adaptive biasing force dynamics mentioned above is based on two ingredients. The first one is that the law of $\xi(\mathbf{x}_t)$ follows a simple diffusion

dynamics if $\mathbf{x}_t$ is solution of the adaptive biasing force dynamics (S6). Indeed, if one considers $\psi^\xi(t,x_1) = \int \psi(t,x_1,x_2,\ldots,x_N)\,\mathrm{d}x_2\ldots\mathrm{d}x_N$ where ψ satisfies (S8), it is easy to check that,

$$\partial_t \psi^\xi = \beta^{-1}\partial_{x_1,x_1}\psi^\xi. \tag{S9}$$

This rigorously justifies the intuition that the adaptive bias flattens the potential energy landscape in the ξ direction. The second ingredient is a two-scale analysis of entropy following ideas introduced by F. Otto and co-workers, see references 5, 6 and 7.

To conclude this section, let us emphasize that the above analysis assumes that the conditional expectation appearing in the adaptive biasing force dynamics is exactly computed. This analysis is therefore well adapted to discretizations using many replicas in parallel, which indeed converge to the adaptive biasing force dynamics with the exact conditional expectation, see.[8] The analysis of the adaptive algorithms (adaptive biasing force or adaptive biasing potential) with estimates of the conditional expectations based on trajectory averages along a single path are much more complicated. See references 9 and 10 for preliminary results for the Wang-Landau algorithm.

In practice, in order to check convergence, the following technique can be used — The biasing force is fixed, and from the biased trajectory, the mean force is evaluated (see equation (7) of the article), which shows that the conditional expectation of the local mean force, indeed, yields the mean force, even on the biased potential. If the obtained mean force is similar to the fixed biasing force, this indicates with good confidence that the computed free-energy changes are converged.

Measure of the stastical error

Statistical error of the mean force According to the adaptive biasing force protocol, the mean force is calculated by accumulating the instantaneous force in bins along the transition coordinate. At each simulation step t, the system takes on a configuration $\mathbf{x}_t$, which determines the instantaneous force of the system on the transition coordinate, denoted $F_\xi(\mathbf{x}_t)$. Choosing small bins of

size $\delta\xi$ defined by the bin edges $\{Z_i\}_{i=0}^{M}$, the mean force in bin i is estimated by

$$\langle F_\xi \rangle_i = \frac{1}{n_i} \sum_{Z_i \le \xi < Z_{i+1}} F_\xi(\mathbf{x}_t), \tag{S10}$$

where the sum runs over all simulation steps in which $\xi(\mathbf{x}_t)$ lies in the ith bin and n_i is the total number of such simulation steps. Note that $\langle F_\xi \rangle_i$ and $F_\xi(\mathbf{x}_t)$ represent intrinsic system forces, and therefore do not include contributions from the biasing force. Let $\Delta F_\xi(\mathbf{x}_t)$ be the random component of the instantaneous force — namely, $\Delta F_\xi(\mathbf{x}_t) = F_\xi(\mathbf{x}_t) - \langle F_\xi \rangle_i$ with the bin index i defined by $Z_i \le \xi(\mathbf{x}_t) < Z_{i+1}$. Using this definition, the error of the mean force can be identified with the standard error of the mean,[11]

$$\mathrm{Err}\left[\langle F_\xi \rangle_i\right] = \sqrt{\frac{\langle \Delta F_\xi^2 \rangle_i}{m_i}}, \tag{S11}$$

where m_i is the number of independent samples obtained in bin i and $\langle \Delta F_\xi^2 \rangle_i$ is the variance of the random force pertaining to bin i. The number of independent samples is generally not equal to n_i since the instantaneous forces are, in practice, correlated over many time steps. An estimate of the number of independent samples is obtained by dividing the simulated time spent in the bin by the autocorrelation time of the random force, $m_i = n_i \Delta t / \tau_i$, where Δt is the simulation time step. We then obtain the following estimate for the statistical error of the mean force,

$$\mathrm{Err}\left[\langle F_\xi \rangle_i\right] = \sqrt{\frac{\tau_i}{n_i \Delta t} \langle \Delta F_\xi^2 \rangle_i}. \tag{S12}$$

Obtaining convergence of the autocorrelation function requires a large number of samples; thus, in practice we compute it over several neighboring bins by,

$$\Phi_{i,i+j}(t') = \frac{\langle \Delta F_\xi(\mathbf{x}_0)\, \Delta F_\xi(\mathbf{x}_{t'}) \rangle_{i,i+j}}{\langle \Delta F_\xi^2 \rangle_{i,i+j}}, \tag{S13}$$

where the averages are taken over subtrajectories for which $Z_i \leq \xi(\mathbf{x}_t) < Z_{i+j}$ and $i, i+j$ represents the chosen range of j bins. Another option to obtain $\mathrm{Err}\left[\langle F_\xi \rangle_i\right]$ is to apply the approach of Flyvbjerg and Petersen[12] to the time series of random forces, which yields the statistical error of the mean force without explicit computation of the correlation time.

The equation given in the main text for the error of a free-energy difference,

$$\mathrm{Err}[\Delta A_{a\to b}] = \delta\xi \left(\sum_{i=i_a}^{i_b} \frac{\tau_i}{n_i \Delta t} \left\langle \Delta F_\xi^2 \right\rangle_i \right)^{1/2} . \tag{S14}$$

has similar properties to the formula derived in the appendix of Rodriguez-Gomez et al.;[13] however, we do not make the approximation that the number of samples in each bin is uniform ($n_i = \langle n \rangle$). Although the adaptive biasing force method yields uniform sampling in the long-time limit, significant nonuniform sampling is sometimes observed in practice, as demonstrated below. However, the form given by Rodriguez-Gomez et al., $\mathrm{Err}[\Delta A_{a\to b}] \propto z_b - z_a$, is instructive in that the dependence on the size of the interval can be clearly seen. Thus, free-energy differences over small intervals in ξ are more reliable than free-energy differences over larger intervals. This effect is also apparent in equation S14 — more positive terms are included in the sum as the size of the interval is increased.

Calculating the standard deviation and autocorrelation time of the random force. Our results for reversible folding of deca-alanine as well as for other systems, including passive permeation of water through lipid bilayers, suggest that the standard deviation and the autocorrelation time of the instantaneous system force often vary little in the transition coordinate. It, therefore, may be unnecessary to carefully compute the ξ dependence of these quantities, However, it is advisable to check that a few different regions of the transition coordinate yield similar values for $\sqrt{\left\langle \Delta F_\xi^2 \right\rangle_i}$ and τ_i. Substantial ξ dependence is likely in heterogeneous systems (e.g. liquid-gas interfaces), in the vicinity of phase transitions, and for transition coordinates with intrinsic nonuniformity (e.g. ξ equal to the root-mean-square deviation of a large molecule from a reference

structure).

As a word of caution, we note that solutes in vacuum (or in implicit solvent) may show unusual behavior with respect to the correlation time of the system force that is not representative of explicitly solvated systems. For instance, using a Langevin thermostat with a damping constant of 0.2 ps^{-1}, we found appreciable oscillations in autocorrelation function persisting beyond 2 ps. These oscillations complicated the calculation of the correlation time and led to implausibly large values for the error of the system force, especially given that the Langevin damping constant has a negligible effect on the resulting mean force profile. For this reason, the graphs in figure 8 are derived from a simulation employing a damping constant of 5 ps^{-1}.

Reconciling thermodynamics and kinetics: Precision and reliability

Statistical error. Here we briefly address the question as to how to estimate the statistical error of the results of the Bayesian scheme. One method that has been used in the past[14,15] is to partition the trajectory into multiple subsets and perform the analysis on each of the subsets individually. The deviation of the results for these subsets from the results for the entire trajectory yields the error estimate.

Consistency of the diffusive model. One might ask why we use the Bayesian scheme to determine the mean system force $F(z)$, or equivalently, the free energy $A(z)$, when it is already available within the adaptive biasing framework. First, keeping $F(z)$ as a free parameter allows some validation Bayesian inference scheme. For example, disagreement between the system force derived from adaptive biasing, $F^{\mathrm{ABF}}(z)$, and that predicted by the Bayesian scheme, $F^{\mathrm{Bayes}}(z)$, can indicate that the Metropolis-Hastings algorithm used in the latter has not converged. Choosing too large a value for Δt or using an inappropriate diffusive model can also be manifested as a difference between $F^{\mathrm{ABF}}(z)$ and $F^{\mathrm{Bayes}}(z)$. Indeed, we have observed that leaving out the

$\nabla D(z_{t_\alpha})\Delta t$ term in equation 39 of the main text can result in noticeable distortions in $F^{\mathrm{Bayes}}(z)$.[14] Finally, $F^{\mathrm{ABF}}(z)$ is not exact, and it may be possible to construct a more robust diffusive model by allowing $F^{\mathrm{Bayes}}(z)$ to deviate slightly, especially in undersampled regions. Figure 13A shows that for reversible unfolding of deca-alanine, the mean force yielded by the adaptive biasing force algorithm and the Bayesian scheme are quite similar. If desired, the correspondence between $F^{\mathrm{ABF}}(z)$ and $F^{\mathrm{Bayes}}(z)$ can be enforced through the appropriate prior distribution.[14]

It is furthermore important to verify the self-consistency of the diffusive model. In overdamped Langevin dynamics, consecutive displacements are uncorrelated; thus, one should check that significant correlation is absent from the simulation trajectory on the observation time interval Δt. Values of Δt must be chosen to be both sufficiently large that consecutive displacements are uncorrelated and sufficiently small that discretization error is negligible.[15] Another major sign that Langevin dynamics is not an appropriate model is if the results of the Bayesian scheme show a significant dependence on Δt (in the absence of large discretization error), implying non-Markovian behavior.[16] Indeed, figure 13B of the main text reveals just such a strong dependence for reversible folding of deca-alanine in vacuum. Many systems with non-Markovian dynamics have been identified,[17–19] and developing more sophisticated models to study such systems is an ongoing challenge.

References

(1) Lelièvre, T.; Stoltz, G.; Rousset, M. *Free energy computations: A mathematical perspective*; Imperial College Press, 2010.

(2) Lelièvre, T.; Rousset, M.; Stoltz, G. Long-time Convergence of an Adaptive Biasing Force Method. *Nonlinearity* **2008**, *21*, 1155–1181.

(3) Lelièvre, T.; Minoukadeh, K. Long-time Convergence of an Adaptive Biasing Force Method: the Bi-channel Case. *Archive for Rational Mechanics and Analysis* **2011**, *202*, 1–34.

(4) Ané, C.; Blachère, S.; Chafaï, D.; Fougères, P.; Gentil, I.; Malrieu, F.; Roberto, C.; Schef-

fer, G. *Sur Les Inégalités De Sobolev Logarithmiques*; Société Mathématique de France, 2000; In French.

(5) Otto, F.; Reznikoff, M. A New Criterion for the Logarithmic Sobolev Inequality and Two Applications. *J. Funct. Anal.* **2007**, *243*, 121–157.

(6) Grunewald, N.; Otto, F.; Villani, C.; Westdickenberg, M. A Two-scale Approach to Logarithmic Sobolev Inequalities and the Hydrodynamic Limit. *Ann. Inst. H. PoincarÃ© Probab. Statist.* **2009**, *45*, 302–351.

(7) Lelièvre, T. A General Two-scale Criteria for Logarithmic Sobolev Inequalities. *J. Funct. Anal.* **2009**, *256*, 2211–2221.

(8) Jourdain, B.; Lelièvre, T.; Roux, R. Existence, Uniqueness and Convergence of a Particle Approximation for the Adaptive Biasing Force Process. *ESAIM-Math. Model. Num.* **2010**, *44*, 831–865.

(9) Fort, G.; Jourdain, B.; Kuhn, E.; Lelièvre, T.; Stoltz, G. Convergence of the Wang-Landau Algorithm. 2014; (in press).

(10) Fort, G.; Jourdain, B.; Kuhn, E.; Lelièvre, T.; Stoltz, G. Efficiency of the Wang-Landau Algorithm: a Simple Test Case. 2014; (in press).

(11) Taylor, J. R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books: Sausalito, California, 1997.

(12) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461–466.

(13) Rodríguez-Gómez, D.; Darve, E.; Pohorille, A. Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.* **2004**, *120*, 3563–3578.

(14) Comer, J.; Dehez, F.; Cai, W.; Chipot, C. Water conduction through a peptide nanotube. *J. Phys. Chem. C* **2013**, *117*, 26797–26803.

(15) Comer, J.; Dehez, F.; Cai, W.; Chipot, C. Water Conduction Through a Peptide Nanotube. *J. Phys. Chem. C* **2013**, *117*, 26797–26803.

(16) Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **2005**, *7*, 34.

(17) Flenner, E.; Das, J.; Rheinstädter, M. C.; Kosztin, I. Subdiffusion and Lateral Diffusion Coefficient of Lipid Atoms and Molecules in Phospholipid Bilayers. *Phys. Rev. E* **2009**, *79*, 011907.

(18) Weiss, M.; Elsner, M.; Kartberg, F.; Nilsson, T. Anomalous Subdiffusion Is a Measure for Cytoplasmic Crowding in Living Cells. *Biophys. J.* **2004**, *87*, 3518–3524.

(19) Kou, S.; Xie, X. Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule. *Phys. Rev. Lett.* **2004**, *93*, 180603.