
Contents

1	Theoretical Background	1
1.1	Statistical thermodynamics and free energy differences	1
1.2	Adaptive Biasing Methods	2
1.2.1	Well-Tempered Metadynamics (WT-MetaD)	5
1.2.2	Adaptive Biasing Force Method (ABF)	6
1.2.3	Meta-eABF	10
1.3	Standard vs geometric free energy	10
	Bibliography	IV

1 Theoretical Background

1.1 Statistical thermodynamics and free energy differences

The formulation of the problem of free energy estimation starts with the Hamiltonian

$$\mathbf{H}(\mathbf{x}, \mathbf{p}) = \sum_{i=0}^N \frac{p_i^2}{2m_i} + U(\mathbf{x}) \quad (1.1)$$

of an arbitrary chemical system, where (\mathbf{x}, \mathbf{p}) denotes a point in the phase space of the N -particle system of interest with atomic coordinates $\mathbf{x} = (x_1, \dots, x_N)$, momenta $\mathbf{p} = (p_1, \dots, p_N)$ and masses (m_1, \dots, m_N) . $U(\mathbf{x})$ is the potential energy given by any quantum chemical method or force field. If the canonical (N,V,T)-ensemble of such a system is sampled, for example by means of Langevin dynamics or Monte Carlo simulations, the probability distribution $\rho(\mathbf{x})$ in the configuration space \mathbf{x} follows the Boltzmann distribution

$$\rho(\mathbf{x}) = \frac{e^{-\beta U(\mathbf{x})}}{\int e^{-\beta U(\mathbf{x})} d\mathbf{x}} = Z^{-1} e^{-\beta U(\mathbf{x})} \quad (1.2)$$

where Z denotes the partition function and $\beta = (k_B T)^{-1}$ the inverse temperature with Boltzmann constant k_B . In this work we will focus on alchemical transformations, $A \rightarrow B$, which are represented by a small set of collective variables (CV's) $(\xi_1(\mathbf{x}) : \mathbb{R}^{3N} \rightarrow \mathbb{R}, \dots, \xi_n(\mathbf{x}))$, that are sufficient to distinguish between key states of interest. In this framework the probability distribution can be expressed by a histogram along collective variables

$$\rho(\xi) = \int \delta[\xi - \xi(\mathbf{x})] \rho(\mathbf{x}) = \langle \delta[\xi - \xi(\mathbf{x})] \rangle_\xi \quad (1.3)$$

where $\langle \rangle_\xi$ denotes the ξ -conditioned marginal distribution and the free energy difference between two states A and B is defined as

$$\Delta A_{A \rightarrow B} = -\beta^{-1} \ln \frac{\rho(\xi(\mathbf{x}_B))}{\rho(\xi(\mathbf{x}_A))} \quad (1.4)$$

However, in computer simulations the direct phase-space integrals used in equation 1.2 and 1.3 are impossible to calculate. Instead the time average $P(\xi)$ is computed by monitoring ξ during the simulation.

$$P(\xi) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \rho[\xi(t')] dt' \quad (1.5)$$

In practice, $P(\xi)$ is approximated as a histogram along the reaction coordinate with bins of fixed size. Assuming that the system at hand behaves *ergodic*, i.e. every point in phase space is visited during a infinitely long simulation, the ensemble average $\rho(\xi)$ converges to the time average $P(\xi)$ for long simulations. However, transitions of high energy barriers during chemical reactions are typically rare events, that might happen on timescales far beyond what is feasible to compute. Therefore most processes, for example in biochemical context, are *quasinonergodic*. The problem of calculating accurate free energy differences hence really consists in adequate sampling of regions with high free energy (transition states) along the reaction coordinate.

For this purpose several *enhanced sampling* methods were developed.[‡] In this work we will focus on CV-based approaches, that are based on altering the potential energy in a way, that increases the time spent in a certain region of configuration space during a simulation. One of the oldest and simplest methods to achieve the same is *Umbrella Sampling* (US). In its most common form US introduces an fixed harmonic bias potentials of the form

$$U_B(\xi) = \frac{k_i}{2} (\xi(t) - \xi_i)^2 \quad (1.6)$$

where the index i indicates a given bias window along the reaction coordinate with force constant k_i .

<reweighting and WHAM>

1.2 Adaptive Biasing Methods

Instead of dividing the reaction coordinate in several windows, with adaptive biasing methods the free energy can be estimated from one single simulation. For this purpose the systems dynamics are biased towards states corresponding to large values of the free energy along the transition coordinate via a history-dependent biasing potential. In contrast to other importance sampling strategies like umbrella sampling, this methods require no prior knowledge of the free-energy landscape at hand. Instead, the biasing potential automatically converges towards the free energy, enabling diffusive behavior along the transition coordinate.

There are multiple adaptive biasing methods available, only differing in the construction of the bias. Methods based on metadynamics (metaD) disfavor already visited states by accumulating repulsive potentials along the CV (section 1.2.1), while adaptive biasing force (ABF) methods compensate the mean force along the reaction coordinate to obtain uniform sampling (section 1.2.2). Meta-eABF finally combines both complementary approaches to further speed up convergence (section 1.2.3).

In principle adaptive biasing methods only rely on the sampling of the canonical ensemble, which will be obtained using Langevin dynamics in this work. A schematic procedure of adaptively biased Langevin dynamics is given in Algorithm 1.

Algorithm 1 Velocity Verlet integrator for adaptively biased Langevin dynamics with atomic masses \mathbf{M} , coordinates $\mathbf{x}(t)$, momenta $\mathbf{p}(t)$, potential $V(\mathbf{x}(t))$, forces $F(\mathbf{x}(t))$ and friction coefficient γ ,

while $t < t_{end}$ **do**

$\mathbf{p}(t + \frac{1}{2}\Delta t) \leftarrow \mathbf{p}(t) + \frac{1}{2}(F(\mathbf{x}(t))dt - \gamma\mathbf{M}^{-1}\mathbf{p}(t)dt + \sqrt{2\gamma\beta^{-1}}dW_t)$
 $\mathbf{x}(t + \Delta t) \leftarrow \mathbf{x}(t) + \frac{2}{2+\gamma dt}\mathbf{M}^{-1}\mathbf{p}(t + \frac{1}{2}\Delta t)dt$
 /* Langevin dynamics

$F(\mathbf{x}(t + \Delta t)) \leftarrow -\nabla U(\mathbf{x}(t + \Delta t))$
 /* get physical QM/MM forces

$\xi \leftarrow f(\mathbf{x}(t + \Delta t))$
 /* Calculate reaction coordinate from Cartesian coordinates

if $\xi_{min} \leq \xi \leq \xi_{max}$ **then**

$F_B(\xi, t + \Delta t) \leftarrow F_B(\xi, t) + \Delta F_B(\xi, t + \Delta t)$
 /* update history dependent bias force along reaction coordinate incrementally

$F(\mathbf{x}(t + \Delta t)) \leftarrow -\nabla U(\mathbf{x}(t + \Delta t)) + F_B(\xi, t + \Delta t)\nabla \xi$
 /* Add biasing force to physical force

else

$F(\mathbf{x}(t + \Delta t)) \leftarrow -\nabla U(\mathbf{x}(t + \Delta t)) + k \min(|\xi - \xi_{min}|, |\xi - \xi_{max}|)\nabla \xi$
 /* confine to range of interest with force constant k

end if

$\mathbf{p}(t + \Delta t) = \frac{2-\gamma dt}{2+\gamma dt}\mathbf{p}(t + \frac{1}{2}\Delta t) - \frac{1}{2}(F(\mathbf{x}(t + \Delta t))dt - \gamma\mathbf{M}^{-1}\mathbf{p}(t + \frac{1}{2}\Delta t)dt + \sqrt{2\gamma\beta^{-1}}dW)$
 /* Langevin dynamics

end while

/* get unbiased free energy

1.2.1 Well-Tempered Metadynamics (WT-MetaD)

MetaD biases a systems dynamic towards undersampled regions along the reaction coordinate $\xi(\mathbf{x})$, by accumulating repulsive potentials in regions that have already been visited. The bias potential is typically build by a superposition of repulsive Gaussian kernels and can be written:^[1]

$$U_{metaD}(\xi, t) = \sum_{k < \frac{t}{\tau_G}} \tau_G \omega \exp\left(-\sum_{i=1}^{N_{dim}} \frac{1}{2\sigma_i^2} (\xi_i(\mathbf{x}) - \xi_i(\mathbf{x}, t_k))^2\right) \quad (1.7)$$

with deposition rate τ_G , Gaussian height $\omega = W/\tau_G$ and variance σ^2 as free input parameters. Over the course of a simulation the bias potential fills local minima along the reaction coordinate until the systems evolution finally resembles a Brownian motion along the flattened free energy surface. The converged bias potential provides an unbiased estimate of the underlying free energy surface

$$A(\xi) = -U_{metaD}(\xi, t \rightarrow \infty) + C \quad (1.8)$$

To avoid oscillation of U_{metaD} around the correct free energy, Well-Tempered metadynamics (WTM) introduces an additional scaling factor of the Gaussian height:^[2]

$$\omega(\xi, t) = \frac{W}{\tau_G} \exp\left(-\frac{U_{WTM}(\xi, t)}{k_B \Delta T}\right) \quad (1.9)$$

ensuring an decrease of ω over time and smooth convergence of the Well-Tempered bias potential $U_{WTM}(\xi, t)$. The new bias potential does not fully compensate the free energy surface, but can be controlled by parameter ΔT . For $\Delta T \rightarrow 0$ the bias is zero and ordinary MD is recovered, whereas the limit $\Delta T \rightarrow \infty$ corresponds to normal metaD. To obtain a unbiased free energy estimate from $U_{WTM}(\xi, t)$ it has to be scaled accordingly:

$$A(\xi) = -\frac{T + \Delta T}{\Delta T} U_{WTM}(\xi, t) \quad (1.10)$$

1.2.2 Adaptive Biasing Force Method (ABF)

The intuition behind ABF is, that adding a force $A'(\xi(\mathbf{x}))\nabla\xi(\mathbf{x})$ that exactly compensates the average of the original force $-\nabla U(\mathbf{x})$ along a given coordinate would result in uniform sampling along this coordinate.^[3] Historically, this idea emerged from thermodynamic integration (TI), where the free energy derivative is computed as the ensemble average of the instantaneous force, F , acting along the collective variable:

$$\frac{dA}{d\xi} = -\langle F \rangle_\xi \quad (1.11)$$

and the free energy is calculated as the integral over this force.^[4,5] In practice, as one has no prior knowledge of the free energy derivative, ABF uses an on-the-fly estimate of the mean force acting along the reaction coordinate. For this purpose the transition coordinate ξ , connecting two end points, is divided in M equally spaced bins. The approximation of the bias force $\bar{F}(N_{Step}, k)$ in bin k is then the average of all collected force samples:^[3]

$$\bar{F}(N_{Step}, k) = \frac{R(N_{Step}^k, k)}{N_{Step}^k} \sum_{\mu=1}^{N_{Step}^k} F_\mu^k \quad (1.12)$$

$$R(N_{step}^k, k) = \begin{cases} N_{full}, & N_{step}^k < N_{full} \\ 1, & N_{step}^k \geq N_{full} \end{cases} \quad (1.13)$$

with the linear ramp function $R(N_{step}^k, k)$ preventing large fluctuations of the force estimate at the beginning of the simulation from driving the system away from equilibrium. The number of samples when the full biasing force is applied, N_{full} , and the bin size are the only free parameters that have to be chosen by the user before the simulation. For a sufficiently large number of samples N_{step}^k equation 1.12 approaches the correct average force in each bin.^[3] For 1D reaction coordinates the unbiased free energy difference ΔA can be trivially obtained from the force estimates with any numerical integration scheme (e.g. Simpson's rule).

However, for multidimensional reaction coordinates calculating ΔA from ABF-forces is more involved and has to be done in a postprocessing step. Because of statistical errors the mean force in individual bins it is not conservative. This problem can be avoided using the finite element method (FEM). For this purpose tent functions B_l are used to approximate $A(\xi)$:^[6]

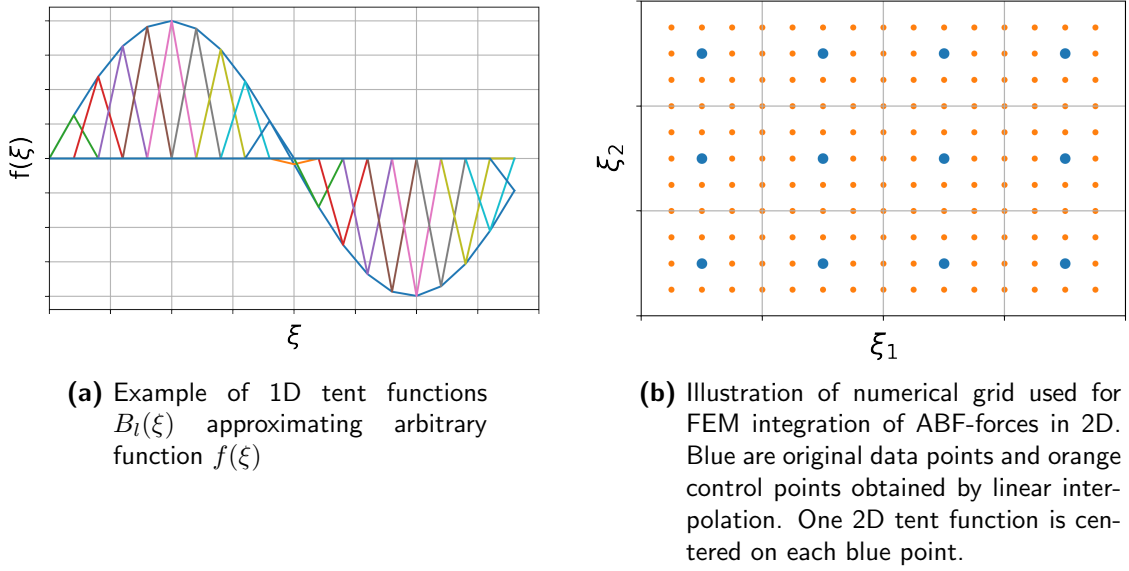
$$A(\xi) = \sum_l \alpha_l B_l(\xi) \quad (1.14)$$

Coefficients α_l are obtained by minimizing

$$RMSE(\vec{\alpha}) = \sqrt{\frac{1}{M} \sum_k \left(\left(\sum_l \alpha_l \nabla B_l(\xi_k) \right) - \langle F(\xi_k) \rangle \right)^2} \quad (1.15)$$

with some numerical minimization routine (e.g. BFGS algorithm^[7]) for all partial derivatives simultaneously. An 1D example of the approximation of an arbitrary function by tent functions is given in 1.1a. To increase the smoothness of the reconstructed free energy surface four control points per bin and dimension are chosen between data points (see figure 1.1b).

Figure 1.1: Illustration of FEM method



The last and critical ingredient for the ABF method now is an explicit expression for the instantaneous force F_ξ . Carter et al.^[8] gave a first general expression:

$$F(\xi, \mathbf{q}) = -\frac{\partial V(\xi, \mathbf{q})}{\partial \xi} + \beta^{-1} \frac{\partial \ln |J(\xi, \mathbf{q})|}{\partial \xi} \quad (1.16)$$

which depends implicitly on a vector field $\partial x_i / \partial \xi$, hereafter referred to as "inverse gradient" and on an Jacobian correction term purely geometric in origin. The inverse gradient can be thought of as direction along which an infinitesimal change in ξ is propagated in Cartesian coordinates, the complementary coordinates \mathbf{q} being kept constant. A major drawback of this formalism is the requirement of an full coordinate transform from Cartesian coordinates (\mathbf{x}) to generalized coordinates (ξ, \mathbf{q}) .

This requirement could be lifted by den Otter,^[9] who put forward the breakthrough idea

that the change in ξ can be propagated along an arbitrary vector field \mathbf{v}_i ($\mathbb{R}^{3N} \rightarrow \mathbb{R}^{3N}$), provided it satisfies some orthonormality conditions. Extended to multidimensional reaction coordinates $\xi = (\xi_i)$ and in presence of a set of constraints $\sigma_k(\mathbf{x}) = 0$ these read:^[10]

$$\mathbf{v}_i \cdot \nabla \xi_j = \delta_{ij} \quad (1.17)$$

$$\mathbf{v}_i \cdot \nabla \sigma_k = 0 \quad (1.18)$$

If all reaction coordinates ξ_i are orthogonal to one another and to all constraints, $\mathbf{v}_i = \nabla \xi_j / |\nabla \xi_j|^2$ is always a valid, but not necessarily the best, option. Otherwise conditions 1.17 and 1.18 can be fulfilled by orthogonalization:^[10]

$$v_i(\mathbf{x}) = \frac{Q^i \nabla \xi_i(\mathbf{x})}{|Q^i \nabla \xi_i(\mathbf{x})|} \quad (1.19)$$

with projector Q^i given by the orthonormal basis $\{\hat{n}_j^i(\mathbf{x})\}_{j \neq i}$ in the subspace spanned by $\{\nabla \xi_j(\mathbf{x})\}_{j \neq i} \cup \{\nabla \sigma_j(\mathbf{x})\}_{j=1, \dots, M}$:

$$Q^i = \mathbf{1} - \sum_{j \neq i} \hat{n}_j^i(\mathbf{x}) \otimes \hat{n}_j^i(\mathbf{x}) \quad (1.20)$$

Replacing the inverse gradient by vectorfield \mathbf{v}_i , expression 1.16 finally reduces to:

$$F(\xi_i, \mathbf{x}) = -\nabla V(\mathbf{x}) \cdot \mathbf{v}_i(\mathbf{x}) + \beta^{-1} \nabla \cdot \mathbf{v}_i(\mathbf{x}) \quad (1.21)$$

but still involves the calculation of second derivatives in the form of the divergence of vector fields \mathbf{v}_i .^[3] Analytic expressions for bend angles and torsion angles, used in the present work, are given in the appendix. However, especially for torsion angles orthogonalization via equation 1.19 becomes exceedingly tedious and impracticel, significantly limiting the applicability of ABF for multidimensional reaction coordinates.

To circumvent the technical requirements of ABF Lesage et al.^[11] proposed an more flexible approach named eABF. In eABF the physical system is extended by additional coordinates λ with mass m_λ , which are coupled to the reaction coordinates ξ_i with harmonic potentials. The extended system (\mathbf{x}, λ) evolves according to Langevin dynamics in the extended potential

$$V^{ext}(\mathbf{x}, \lambda_i) = V(\mathbf{x}) + \frac{k_i}{2} (\xi_i(\mathbf{x}) - \lambda_i)^2. \quad (1.22)$$

The key intuition behind eABF is, that in the tight coupling limit efficient sampling of λ will result in efficient sampling of ξ . Therefore, to obtain uniform sampling along ξ biasing

of λ is sufficient. The inverse gradient is chosen as null for all physical coordinates \mathbf{x} and 1 for λ . This way constraints 1.17 and 1.18 are always satisfied, which is especially useful for calculations involving a set of non-orthogonal reaction coordinates. Sampling the extended system gives the following unbiased Boltzmann distribution in λ :

$$\begin{aligned}\rho^k(\lambda) &\propto \int \exp\left[-\beta\left(V(\mathbf{x}) + \frac{k}{2}(\xi(\mathbf{x}) - \lambda)^2\right)\right] d\mathbf{x} \\ &= \int \exp\left[-\beta V(\mathbf{x}) - \frac{(\xi(\mathbf{x}) - \lambda)^2}{2\sigma^2}\right] d\mathbf{x}\end{aligned}\quad (1.23)$$

which depends on the force constant k or variance of the Gaussian kernel $\sigma^2 = (\beta k)^{-1}$. The bias on λ is the running average over the spring force in λ -bin k :

$$\bar{F}(\lambda_i, k) = \frac{\partial A^k(\lambda_i)}{\partial \lambda_i} = \frac{1}{N_{Step}^k} \sum_{\mu=1}^{N_{Step}^k} k(\lambda_{i,\mu}^k - \xi_{i,\mu}^k) \quad (1.24)$$

For small values of N_{Step}^k again the linear ramp function $R(N_{Step}, k)$ given by equation 1.13 is used. This generates the following biased Boltzmann distribution:

$$\tilde{\rho}(\lambda) \propto \int \exp\left[-\beta V(\mathbf{x}) - \frac{(\xi(\mathbf{x}) - \lambda)^2}{2\sigma^2} + A^k(\lambda)\right] d\mathbf{x} \quad (1.25)$$

By using $\int \delta(\xi(\mathbf{x}) - z) dz = 1$ and $A^k(\lambda) = -\beta^{-1} \ln \rho^k(\lambda)$ one can obtain the relationship between unbiased and biased z -distributions:

$$\tilde{\rho}(z) = \rho(z) \int \frac{\exp\left(-\frac{(\lambda-z)^2}{2\sigma^2}\right)}{\int \exp\left(-\frac{(\lambda-z')^2}{2\sigma^2}\right) \rho(z') dz'} d\lambda \quad (1.26)$$

For the tight coupling limit (high k , low σ) the unbiased distribution $\rho(z)$ is recovered and eABF recovers the behavior of standard ABF. In this case $A^k(\lambda)$ approximates the physical free energy $A(z)$ and the ΔA_z can be approximated by integrating over the converged bias forces on λ , $\bar{F}(\lambda_i, k)$, which will be referred to as "naive estimator" (eABF/naive) hereafter.

An asymptotically unbiased estimator of the free energy can be derived by correcting the free energy gradient obtained from the eABF-biased distribution $\tilde{\rho}(z)$ with the average biasing force on z

$$\frac{\partial A(z)}{\partial z_i} = -\beta^{-1} \frac{\partial \ln \tilde{\rho}(z)}{\partial z_i} + k(\langle \lambda_i \rangle_z - z_i) \quad (1.27)$$

which is called "Corrected z -averaged restraint" (CZAR) and can be calculated numerically from the time trajectory (z_i, λ_i) in an post-processing step.^[11]

1.2.3 Meta-eABF

1.3 Standard vs geometric free energy

List of Tables

List of Figures

1.1	Illustration of FEM method	7
-----	--------------------------------------	---

Bibliography

- [1] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826–843.
- [2] A. Barducci, G. Bussi, M. Parrinello, *Physical review letters* **2008**, *100*, 020603.
- [3] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, C. Chipot, *The Journal of Physical Chemistry B* **2015**, *119*, 1129–1151.
- [4] J. G. Kirkwood, *The Journal of chemical physics* **1935**, *3*, 300–313.
- [5] R. W. Zwanzig, *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- [6] E. Darve, D. Rodriguez-Gómez, A. Pohorille, *The Journal of chemical physics* **2008**, *128*, 144120.
- [7] J. Nocedal, S. Wright, *Numerical optimization*, Springer Science & Business Media, **2006**.
- [8] E. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chemical Physics Letters* **1989**, *156*, 472–477.
- [9] W. K. den Otter, *The Journal of Chemical Physics* **2000**, *112*, 7283–7292.
- [10] G. Ciccotti, R. Kapral, E. Vanden-Eijnden, *ChemPhysChem* **2005**, *6*, 1809–1814.
- [11] A. Lesage, T. Lelièvre, G. Stoltz, J. Hénin, *The Journal of Physical Chemistry B* **2017**, *121*, 3676–3685.