Ab initio calculations of free-energy reaction barriers

2008 J. Phys.: Condens. Matter 20 064211

(http://iopscience.iop.org/0953-8984/20/6/064211)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **20** (2008) 064211 (9pp) doi:10.1088/0953-8984/20/6/064211

Ab initio calculations of free-energy reaction barriers

T Bucko

Fakultät für Physik and Center for Computational Materials Science, Universität Wien, Sensengasse, Wien 1090, Austria

E-mail: tomas.bucko@univie.ac.at

Received 27 September 2007, in final form 20 December 2007
Published 24 January 2008
Online at stacks.iop.org/JPhysCM/20/064211

Abstract

The theoretical description of chemical reactions was until recently limited to a 'static' approach in which important parameters such as the rate constant are deduced from the local topology of the potential energy surface close to minima and saddle points. Such an approach has, however, serious limitations. The growing computational power allows us now to use advanced simulation techniques to determine entropic effects accurately for medium-sized systems at *ab initio* level. Recently, we have implemented free-energy simulation techniques based on molecular dynamics, in particular on the blue-moon ensemble technique and on metadynamics, in the popular DFT code VASP. In the thermodynamic integration (blue-moon ensemble) technique, the free-energy profile is calculated as the path integral over the restoring forces along a parametrized reaction coordinate. In metadynamics, an image of the free-energy surface is constructed on the fly during the simulation by adding small repulsive Gaussian-shaped hills to the Lagrangian driving the dynamics. The two methods are tested on a simple chemical reaction—the nucleophilic substitution of methyl chloride by a chlorine anion.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Computer simulations based on a quantum mechanical description of the interatomic forces allow us to study atomistic details of chemical processes. This allows us to identify reaction mechanisms and important factors affecting the reaction kinetics which are usually not accessible by experimental techniques.

Until recently, *ab initio* studies of chemical reactions were restricted to structural relaxations in which the important stationary points on the potential energy surface (PES) such as minima and saddle points were identified. In this approach, the activation energy for a reaction is approximated by the potential energy difference between the saddle point and the corresponding minimum. Very often conclusions on the reaction mechanisms have been made solely on the basis of this parameter. The entropic effect is estimated using harmonic transition state theory [1]. This can be adequate for simple molecular reactions where only a few degrees of freedom are modified during the reaction. In the case of complex systems where soft degrees of freedom such as hindered rotations and translations also undergo transformations during the reaction, a more accurate treatment of entropic effects is necessary.

Many techniques to compute free-energy profiles of a reaction based either on Monte Carlo or molecular dynamics have been proposed. An incomplete list of free-energy methods includes umbrella sampling [2], thermodynamic integration [3–7], metadynamics [8, 9], adiabatic molecular dynamics [10], transition path sampling, [11] and transition interface sampling [12]. In this paper we focus on two methods recently implemented into the popular periodic DFT code VASP [13–16]—thermodynamic integration and metadynamics. The two methods are tested on a simple chemical reaction—the nucleophilic substitution of methyl chloride by a chlorine anion.

2. Methodology

2.1. Electronic structure calculations

Periodic *ab initio* DFT calculations have been performed using the VASP code [13–16]. The Kohn–Sham equations have been solved variationally in a plane-wave basis set using

Scheme 1. Nucleophilic substitution of methyl chloride by a chlorine anion.

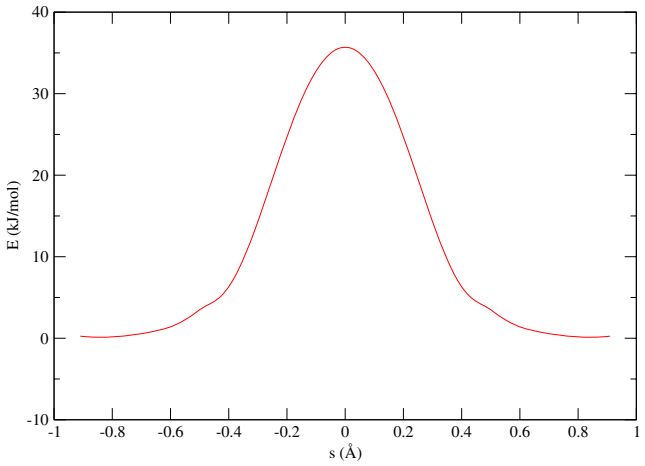

Figure 1. Potential energy profile along the intrinsic reaction coordinate as a function of the path length s.

the projector-augmented-wave (PAW) method of Blöchl [17], as adapted by Kresse and Joubert [18]. The exchange–correlation energy was described by the PW91 generalized gradient approximation (GGA) functional [19, 20]. Brillouin-zone sampling was restricted to the Γ-point. The plane-wave cut-off was set to 205 eV.

2.2. *Molecular dynamics simulations*

Born–Oppenheimer molecular dynamics simulations were carried out in the canonical NVT ensemble; the temperature was controlled by a Nosé–Hoover thermostat [21, 22]. The equations of motion were integrated with a time step of 1 fs. The reactants were placed into the center of a cubic supercell with an edge length of 9.0 Å. In order to avoid overall translations and rotations of the molecule, six Cartesian coordinates (x, y, and z coordinates of carbon, x and y coordinates of one and x coordinate of the other hydrogen atom) have been fixed. The mass of the hydrogen atoms was that of a tritium isotope so as to minimize thermal decoupling of the high-and low-frequency modes.

3. Model reaction

The nucleophilic substitution (S_N2) of chloromethane (scheme 1) has been studied as a model reaction to test the performance of the free-energy methods. This chemical reaction has been the subject of many theoretical studies [23–34]. The calculated reaction barrier depends strongly on the particular method and basis set used. The values reported in literature vary between ∼22 kJ mol^{-1} [29] and ∼90 kJ mol^{-1} [33]. The potential energy profile along the intrinsic reaction coordinate

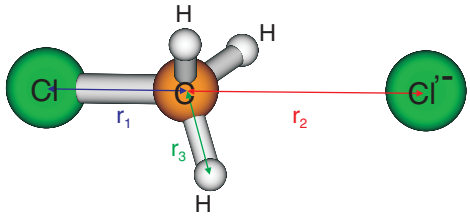

Figure 2. Van der Waals complex between chloromethane and chlorine anion. Three interatomic distances used to study the free-energy surface of the S_N2 reaction are shown.

Table 1. Selected internal parameters for the reactant and transition state.

	$T = 0$ K		$T = 300$ K
	Reactant	Transition state	Reactant
r_1	1.85	2.35	1.84 ± 0.02
r_2	3.15	2.35	3.40 ± 0.10
r_1–r_2	0.00	1.30	1.55 ± 0.10
r_3	1.10	1.08	1.10 ± 0.01

(IRC) [35, 36], defined as the steepest descent path in mass weighted coordinates that connects the transition state to reactant and product, is shown in figure 1. The intrinsic reaction coordinate has been determined using the damped velocity Verlet algorithm [37]. The calculated reaction barrier, defined as the potential energy difference between the transition state and the stable reactant (van der Waals complex), is 36 kJ mol^{-1}. This value is close to the B3LYP calculations reported by Glukhovtsev *et al* [26] (31–36 kJ mol^{-1}) and the MP2 calculations of Streitwieser *et al* [27] (32 kJ mol^{-1}). We stress, however, that we have studied this reaction merely as a simple, yet realistic model for testing the performance of free-energy methods. Therefore, we do not address the question of the accuracy of the density functional method used in this study.

The selected geometry parameters for reactant and transition state as defined in figure 2 are collected in table 1. At the minimum, the two C–Cl separations are $r_1 = 1.85$ Å and $r_2 = 3.15$ Å, respectively, whereas the length of the C–H bonds (r_3) is 1.10 Å. The transition state is symmetric, with $r_1 = r_2 = 2.35$ Å. The length of the C–H bond is slightly shortened during the reaction; r_3 decreases to 1.08 Å in the transition state. In order to appreciate the effect of temperature on the structural parameters and also to obtain a reference for further discussion we have performed an NVT molecular dynamics simulation at a temperature of 300 K. The total simulation time was 65 ps, with the initial 5 ps serving for equilibration. The finite temperature analogues of the geometric parameters of the reactant were obtained as the maxima of histograms of the data collected in the MD simulation (see table 1). Due to increased

temperature, r_2 increases to ∼3.4 Å, whereas r_1 and r_3 remain almost unchanged. This result is consistent with calculations of Yang *et al* [30], who found that the difference between the two C···Cl interatomic separations grows from ∼1.3 Å for $T = 0$ K to ∼1.5 Å for $T = 300$ K.

4. Thermodynamic integration

Suppose that the reaction coordinate can be approximated by a single parameter ξ. The free-energy difference between two states can be calculated using

$$\Delta A_{1\to 2} = \int_{\xi(1)}^{\xi(2)} \mathrm{d}\xi \left(\frac{\partial A}{\partial \xi}\right)_{\xi^*} \tag{1}$$

as an integral over $(\frac{\partial A}{\partial \xi})_{\xi^*}$, the free-energy gradient evaluated at a fixed value ξ^* of the reaction coordinate. The free energy along the reaction path $A(\xi^*)$ is related to the partition function $Q(\xi^*)$ by $A(\xi^*) = -RT\ln[Q(\xi^*)]$. Therefore, a calculation of the free-energy gradient requires first the evaluation of the derivative of the partition function, $\frac{\partial Q}{\partial \xi}$. The partition function is defined by

$$Q(\xi^*) = \int \mathrm{d}q \int \mathrm{d}p_q \, \mathrm{d}p_\xi \exp(-\beta H), \tag{2}$$

where the dynamical variables of the Hamiltonian for the system with M degrees of freedom have been split into the active coordinate ξ defining the reaction path, the inactive coordinates $\mathbf{q} = \{q_i; i = 1, \ldots, M-1\}$ and the associated momenta p_ξ and p_q. In the MD simulations, the reaction coordinate is constrained to remain constant and equal to ξ^*, and this requires the additional constraint $\dot{\xi} = 0$. Therefore, during the MD simulation p_ξ is not sampled; a constrained ensemble average over a quantity $\mathcal{O}$ is evaluated as

$$\langle \mathcal{O} \rangle_{\xi^*} = \frac{\int \mathrm{d}q \int \mathrm{d}p_q \, \mathcal{O} \exp(-\beta H^{\mathrm{c}}_{\xi^*})}{\int \mathrm{d}q \int \mathrm{d}p_q \exp(-\beta H^{\mathrm{c}}_{\xi^*})}, \tag{3}$$

with the Hamiltonian

$$H^{\mathrm{c}}_{\xi^*} = \tfrac{1}{2}\mathbf{p}_q^t \mathbf{X} \mathbf{p}_q + V(\mathbf{q}, \xi). \tag{4}$$

The mass-metric tensor X is defined as

$$X_{\alpha,\beta} = \sum_{i=1}^{i=M} \frac{1}{m_i} \frac{\partial q_\alpha}{\partial x_i} \frac{\partial q_\beta}{\partial x_i}, \qquad \alpha = 1, \ldots, M-1, \quad \beta = 1, \ldots, M-1. \tag{5}$$

The unconstrained average is

$$\langle \mathcal{O} \rangle = \frac{\int \mathrm{d}q \, \mathrm{d}\xi \int \mathrm{d}p_{q\xi} \, \mathcal{O} \exp(-\beta H)}{\int \mathrm{d}q \, \mathrm{d}\xi \int \mathrm{d}p_{q\xi} \exp(-\beta H)}, \tag{6}$$

and the unconstrained and constrained Hamiltonians are related via

$$H = H^{\mathrm{c}}_{\xi^*} + p^t_\xi (\mathbf{Y} \cdot \mathbf{p}_q) + \tfrac{1}{2}\left(p^t_\xi Z p_\xi\right), \tag{7}$$

with

$$Y_\alpha = \sum_{i=1}^{i=M} \frac{1}{m_i} \frac{\partial \xi}{\partial x_i} \frac{\partial q_\alpha}{\partial x_i}, \qquad \alpha = 1, \ldots, M-1, \tag{8}$$

and

$$Z = \sum_{i=1}^{i=M} \frac{1}{m_i} \left(\frac{\partial \xi}{\partial x_i}\right)^2. \tag{9}$$

Following Carter *et al* [4], constrained and unconstrained ensemble averages are related through a 'blue-moon' correction

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O} \, Z^{-1/2} \rangle_{\xi^*}}{\langle Z^{-1/2} \rangle_{\xi^*}}, \tag{10}$$

where the angular brackets $\langle \cdots \rangle_{\xi^*}$ denote conditional thermal averages with constrained coordinate ξ.

To constrain the system to remain on the reaction path during the MD simulation, a modified Lagrangian with the Lagrange multiplier λ associated with the reaction coordinate is used:

$$\mathcal{L}^*(\mathbf{x}, \xi, \dot{\mathbf{x}}) = \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) + \lambda(\xi(\mathbf{x}) - \xi), \tag{11}$$

where ξ is the desired value of geometric parameter $\xi(\mathbf{x})$.

The SHAKE algorithm [38] can be used to determine the Lagrange multipliers. It can be shown [4] that the free-energy gradients $(\frac{\partial A}{\partial \xi})_{\xi^*}$ can be calculated using the Lagrange multipliers via

$$\left(\frac{\partial A}{\partial \xi}\right)_{\xi^*} = \frac{1}{\langle Z^{-1/2} \rangle_{\xi^*}} \left\langle Z^{-1/2} \left[-\lambda_\xi + k_{\mathrm{B}} T (Z^{-1}) \sum_{i=1}^{i=M} \frac{1}{m_i} \frac{\partial \xi}{\partial x_i} \frac{\partial Z}{\partial x_i} \right] \right\rangle_{\xi^*}. \tag{12}$$

The choice of a proper order parameter which reasonably approximates the reaction coordinate is a crucial point in the blue-moon ensemble technique. For the model reaction we are dealing with, a possible choice is the difference between the two Cl···C interatomic separations, $\xi = r_1 - r_2$, (see figure 2). In order to determine the free-energy profile, gradients for several different values of ξ covering the interval between the reactant and transition states must be calculated. The free-energy gradients from a simulation at $T = 10$ K are displayed in the upper panel of figure 3. The free-energy minimum and the saddle point are identified by vanishing free-energy gradients. The minimum is located at $\xi \sim -1.3$ Å, whereas the transition state occurs at $\xi \sim -0.1$ Å. The small asymmetry of the transition state is most likely due to the fixed orientation of the molecule in the simulation cell. As the molecule also interacts with its periodically repeated images, the two chlorine atoms feel slightly different environments and the symmetry is broken.

In the lower panel of figure 3, free-energy profiles for $T = 10$ and 300 K are compared with the potential energy profile along $\xi = r_1 - r_2$. As the entropy contribution is negligible at 10 K, the low temperature free-energy and potential energy profiles almost coincide. At a temperature of 300 K, the entropic effect becomes significant: the barrier increases to ∼45 kJ mol^{-1} and the position of the minimum shifts from $\xi = \sim 1.3$ to ∼1.5 Å.

In this simple reaction it was not too difficult to choose a single parameter which approximates the reaction coordinate reasonably well. On the other hand, for more complicated reactions the reaction coordinate is more complex and it

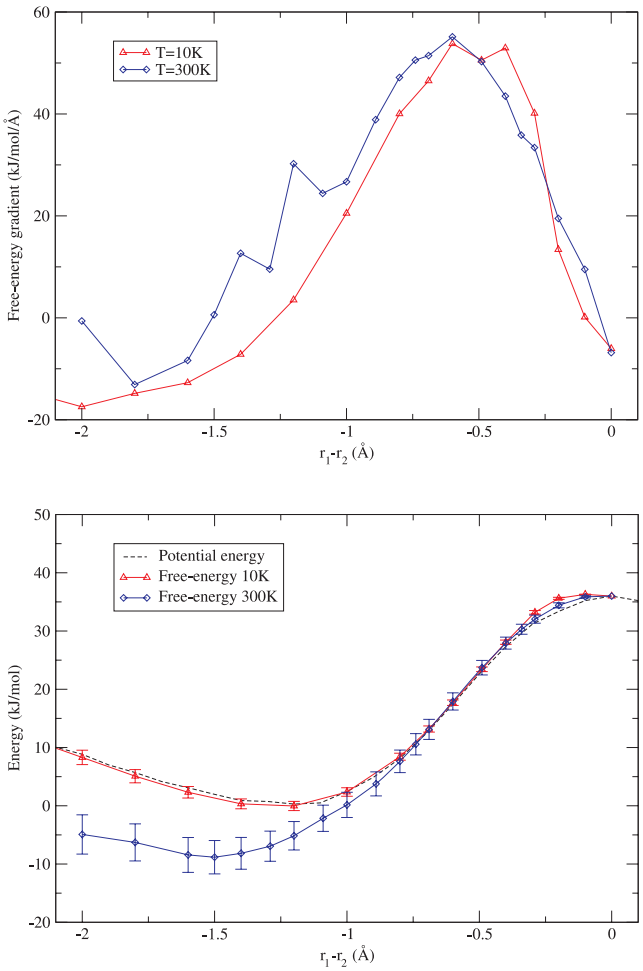

Figure 3. Free-energy gradients (above) and free-energy profiles (below) for the S_N2 reaction obtained for two different simulation temperatures. For the sake of comparison, the potential energy profile along the $r_1 - r_2$ coordinate is shown.

is usually difficult to approximate it by just one internal coordinate. Hence, it might be advantageous to define the thermodynamic state via a vector of parameters $\boldsymbol{\xi} = \{\xi_k; k = 1, \ldots, r\}$. The formula for the free-energy gradients (equation (12)) can be generalized to [5–7]

$$\left(\frac{\partial A}{\partial \xi_k}\right)_{\xi^*} = \frac{1}{\langle |\mathbf{Z}|^{-1/2}\rangle_{\xi^*}} \left\langle |\mathbf{Z}|^{-1/2} \left[-\lambda_{\xi_k} + k_B T \sum_{j=1}^{j=r} (\mathbf{Z}^{-1})_{kj} \sum_{i=1}^{i=M} \frac{1}{m_i} \frac{\partial \xi_j}{\partial x_i} \frac{\partial |\mathbf{Z}|}{\partial x_i} \right] \right\rangle_{\xi^*}, \tag{13}$$

where $\mathbf{Z}$ is a mass-metric tensor:

$$Z_{\alpha,\beta} = \sum_{i=1}^{i=M} \frac{1}{m_i} \frac{\partial \xi_\alpha}{\partial x_i} \frac{\partial \xi_\beta}{\partial x_i}, \qquad \alpha = 1, \ldots, r, \quad \beta = 1, \ldots, r. \tag{14}$$

As suggested by Fleurad-Lessard *et al* [7], the free-energy gradients can be used in free-energy optimizations in a similar way as the potential energy gradients in static relaxations, i.e. the standard optimization algorithms such as rational function optimization (RFO) [39] and direct inversion of the iterative subspace (DIIS) [40] can be used to determine free-energy minima and transition states. The free-energy barrier is then calculated as a path integral along an arbitrary path connecting the two states:

$$\Delta A_{1\to 2} = \int_{\xi(1)}^{\xi(2)} \mathrm{d}\xi \cdot \left(\frac{\partial A}{\partial \xi}\right)_{\xi^*}. \tag{15}$$

To demonstrate this method, we used the distances of the two Cl ions from carbon, r_1 and r_2 (see figure 2), as components of the reaction coordinate. We started the search for the free-energy minimum from the parameters estimated using standard unconstrained MD simulations at $T = 300$ K: $r_1 = 1.84$ Å, $r_2 = 3.40$ Å, see section 3. By calculating the corresponding free-energy gradients (equation (13)) and performing a steepest descent optimization step we obtained a configuration with $r_1 = 1.86$ Å and $r_2 = 3.40$ Å with gradients well below 0.1 eV Å^{-1} that we have used as the optimization criterion. The parameter $r_1 - r_2$ is -1.54 Å, which is consistent with the one-dimensional case described above. To determine the saddle point we used the partitioned RFO method [39]. The optimization has been initialized with the parameters and Hesse matrix corresponding to the zero-temperature saddle point (see table 1). During the optimization, the Hesse matrix has been updated using a weighted combination [41] of Powell–symmetric-Broyden (PSB) [42, 43] and symmetric rank one (SR1) [44] formulas. In only two relaxation steps the calculated free-energy gradients decreased below the relaxation threshold of 0.1 eV Å^{-1}. The parameters r_1 and r_2 for the finite temperature transition state are 1.37 and 1.39 Å, respectively. As in the one-dimensional case, the transition state is slightly asymmetric due to the fixed orientation of the molecule in the simulation cell. Our results are in agreement with Yang *et al* [30], who found $r_1 = r_2 = 1.39$ Å for the free-energy transition state at $T = 300$ K.

Having identified two stationary points, we need to define several states along a path connecting the minimum with the transition state and calculate free-energy gradients (equation (13)) and the path integral (equation (15)). As the free energy is a state quantity, the result of the integral in equation (15) is independent of integration path. Hence the integration path can be chosen arbitrarily. We have calculated the free-energy profiles by integrating along two different paths (see figure 4, upper panel). As shown in figure 4, the calculated free-energy barrier is ~40 kJ mol^{-1} at $T = 300$ K, independent of the integration path. On the other hand, the free-energy barrier obtained using two collective variables is lower compared to the one-dimensional case. In fact, the two simulations are not expected to yield exactly the same results, as the number of degrees of freedom contributing to the entropy is different. The free-energy difference between two states becomes more similar to the potential energy difference with increasing dimensionality of the calculated free-energy surface. In the extreme case when the dimensionality of the free-energy surface reaches the number of degrees of freedom, i.e. $r = M$, the free-energy and potential energy surfaces become identical. This fact is sometimes overlooked in the literature [7, 31].

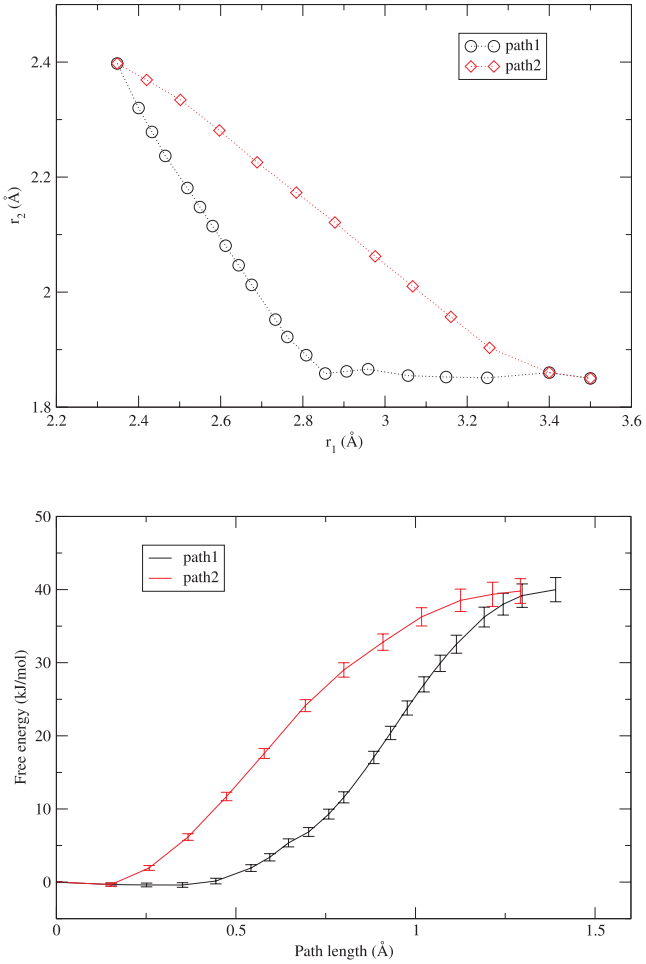

Figure 4. Two different paths in two-dimensional collective variable space (above), and corresponding free-energy profiles for $T = 300$ K (below).

5. Metadynamics

Metadynamics [8, 9] is a powerful method allowing to determine multidimensional free-energy surfaces for chemical reactions. Compared to the Lagrangian $\mathcal{L}_0$ for the standard molecular dynamics, the Lagrangian for metadynamics includes additional degrees of freedom driving the reaction:

$$\mathcal{L} = \mathcal{L}_0 + \sum_\alpha \tfrac{1}{2}\mu_\alpha \dot{s}_\alpha^2 - \sum_\alpha \tfrac{1}{2}k_\alpha (S_\alpha(\mathbf{x}) - s_\alpha)^2 - V(t, \mathbf{s}), \quad (16)$$

where s_α is the coordinate and $\dot{s}_\alpha$ the velocity of an extra variable α with 'mass' μ_α. The fictitious coordinates are coupled to the relevant geometrical parameters of the system $S_\alpha(\mathbf{x})$ (collective variables) via harmonic springs with force constants k_α. The collective variables $S_\alpha(\mathbf{x})$ are chosen so as to have significant components along the reaction coordinate. The time-dependent bias potential $V(t, \mathbf{s})$ driving the system towards the product state consists of a superposition of small Gaussian-shaped 'hills' with height H and width ΔW. Gaussians are added to the bias potential with a time increment t_G, which is typically one or two orders of magnitude greater than the time step used in the MD simulation:

$$V(t, \mathbf{s}) = \sum_{i=1}^{\lfloor t/t_G \rfloor} H \exp\left[-\frac{|\mathbf{s} - \mathbf{s}^i|^2}{2(\Delta W)^2}\right]. \quad (17)$$

The bias potential is, in the limit of infinite simulation time, related to the free energy [8, 9] via

$$A(\mathbf{S}(\mathbf{x})) = -\lim_{t\to\infty} V(t, \mathbf{s}) + \text{const.} \quad (18)$$

Practical hints as how to adjust the five parameters used in metadynamics (H, ΔW, t_G, k_α, μ_α) can be found in [31, 45].

In the first example we examine metadynamics with one collective variable $\xi = r_1 - r_2$. The force constant and fictitious particle mass were set to $k_\alpha = 20$ eV Å^{-2} and $\mu_\alpha = 100$ amu, respectively. The total simulation time was 300 ps. In order to avoid an undesired translation of chloride anion too far from chloromethane, two large repulsive Gaussians with $H = 868$ kJ mol^{-1} and $\Delta W = 0.5$ Å, centered at $\xi = -3.0$ Å and $\xi = 3.0$ Å, respectively, were added to $V(t, \mathbf{s})$ at the beginning of the simulation. In this way, the sampling of the collective variable space was restricted to the region -2.0 Å $< \xi < 2.0$ Å.

In principle, metadynamics is a non-equilibrium method. It depends on the choice of the adjustable parameters, in particular ΔW, H, and t_G, to what extent the system is driven out of equilibrium. It has been shown [45–47] that the statistical error of the free-energy difference in metadynamics is approximately proportional to $\sqrt{\Delta W H / t_g}$. Taking this fact into account we can design two different strategies. One can either update the bias potential frequently with small Gaussians so that the system is often disbalanced, but always close to equilibrium, or, alternatively, use large hills but leave the system for a longer time to equilibrate. We have tested three computational setups with fixed values of $\Delta W = 0.0794$ and $H/t_G = 6.65 \times 10^{-3}$ kJ mol^{-1} fs^{-1}, but with different values of H (0.328, 1.312, and 6.560 kJ mol^{-1}) and t_G (50, 200, and 1000 fs). The free-energy difference between the expected transition state ($\xi = 0.0$) and the reactant state ($\xi = -1.5$) calculated for different simulation times is shown in figure 5. In all three cases, the free-energy barrier does not converge smoothly toward the expected value of 45 kJ mol^{-1} obtained using thermodynamic integration; rather, it oscillates around this value with fluctuations growing with H. Hence, in order to minimize the statistical error it is necessary to keep the system close to equilibrium. This example also shows that a straightforward metadynamics simulation is not particularly well suited for accurate free-energy calculations. Even in the case of a simulation with the lowest value of H, the fluctuations lead to time-dependent and usually asymmetric reaction profiles. This is illustrated in figure 6, where several estimates of the free-energy profiles computed using different simulation times are shown. It has been suggested to terminate simulation immediately after first recrossing the free-energy maximum [48], i.e. immediately after accessing both the reactant and the product states. In our best simulation ($H = 0.328$ kJ mol^{-1}), the first recrossing event takes place after adding 1292 hills. The corresponding free-energy profile is

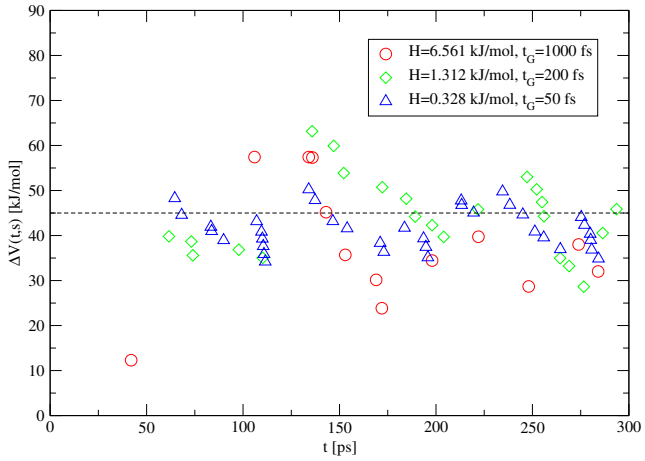

Figure 5. Metadynamics, time evolution of the estimated free-energy difference $\Delta V(t, s)$ between states $\xi = 1.5$ Å and $\xi = 0.0$ corresponding to the free-energy minimum and transition state for three different computational setups. The dashed line indicates the value of the free-energy barrier calculated with thermodynamic integration (45 kJ mol^{-1}).

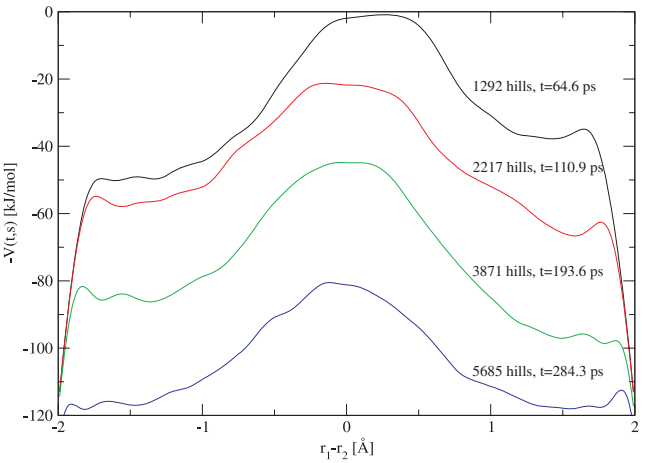

Figure 6. Metadynamics, estimated free-energy profiles upon adding different numbers of hills with $H = 0.328$ kJ mol^{-1}, $\Delta W = 0.0794$ Å, and $t_G = 50$ fs.

shown in figure 6. Obviously, this rather arbitrary strategy not only does not solve the problem with an inaccurate free-energy profile, it also introduces another source of error due to the insufficient sampling of the region close to the dividing surface (see figure 6). As shown by Ensing *et al* [31], metadynamics can be efficiently combined with the umbrella sampling method [2] to improve the accuracy of calculated free-energy profiles. In this strategy, the potential $V(t, \mathbf{s})$ from metadynamics is used as a bias potential which enhances sampling of regions with low statistical weights.

The metadynamics can also be formulated such that the bias potential acts directly on the geometric parameters of the system $S(\mathbf{x})$, i.e.

$$V\,(t, \mathbf{S}(\mathbf{x})) = \sum_{i=1}^{\lfloor t/t_G \rfloor} H \exp\left[-\frac{|\mathbf{S}(\mathbf{x}) - \mathbf{S}^i(\mathbf{x})|^2}{2(\Delta W)^2}\right], \qquad (19)$$

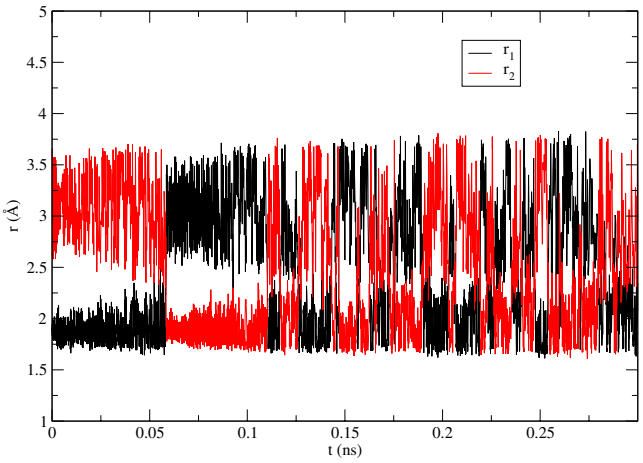

Figure 7. Time evolution of the parameters r_1 and r_2 chosen as the collective variables in the 2D metadynamics. The parameters for Gaussians added to the bias potential are $H = 0.328$ kJ mol^{-1}, $\Delta W = 0.0794$ Å, $t_G = 50$ fs.

and the Lagrangian is

$$\mathcal{L} = \mathcal{L}_0 - V\,(t, \mathbf{S}(\mathbf{x}))\,. \qquad (20)$$

In this formulation, the number of adjustable parameters is reduced to three (H, ΔW, and t_G). In the next example we determine the free-energy surface for two collective variables r_1 and r_2 (see figure 2). The adjustable parameters were set to the same values as in the one-dimensional example ($H = 0.328$ kJ mol^{-1}, $\Delta W = 0.079\,40$ Å, $t_G = 50$ fs). The simulation temperature was 300 K. Once again, repulsive Gaussians were added to $V(t, \mathbf{S}(\mathbf{x}))$ in order to avoid an undesired collapse of the van der Waals complex. Because of the repulsive potentials, the parameters r_1 and r_2 were smaller than 3.8 Å during the whole simulation time. The time evolution of r_1 and r_2 is shown in figure 7. Both the reactant and the product well were filled completely after adding 2215 hills, almost twice as many as needed in the one-dimensional case with the same values of adjustable parameters. The reason is that the volume of the collective variable space that has to be sampled increases with growing dimensionality of the problem. Hence the simulation time necessary to explore collective variable space increases. A contour plot of the free-energy surface obtained after a simulation time of 300 ps is shown in figure 8. The nudged elastic band method [49] was used to obtain the free-energy analog of the intrinsic reaction coordinate: Using a steepest descent algorithm two minima on the free-energy surface were identified. Next, we defined a 'string' of 50 states in the collective variable space forming a line connecting the two minima (figure 8, left). For each image the system was relaxed in the subspace perpendicular to the string until the sum of forces perpendicular to the string was smaller than 10^{-5} eV Å^{-1}. The relaxed string is shown in the right panel of figure 8. The free-energy profile along the relaxed string is shown in figure 9. Similarly as in the simulation with one collective variable, the resulting free-energy profile is slightly asymmetric.

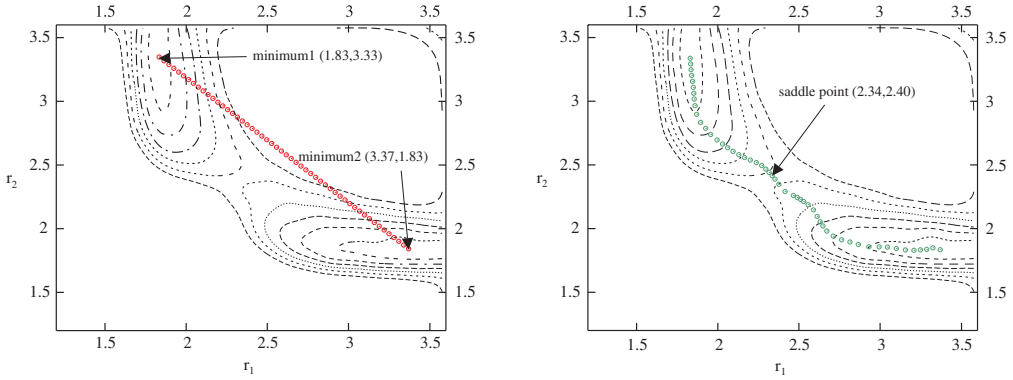

Figure 8. Contour plot of the free-energy surface obtained from metadynamics with two collective variables r_1 and r_2 (see figure 2). Initial string obtained as linear interpolation between two minima (left) and converged string (right).

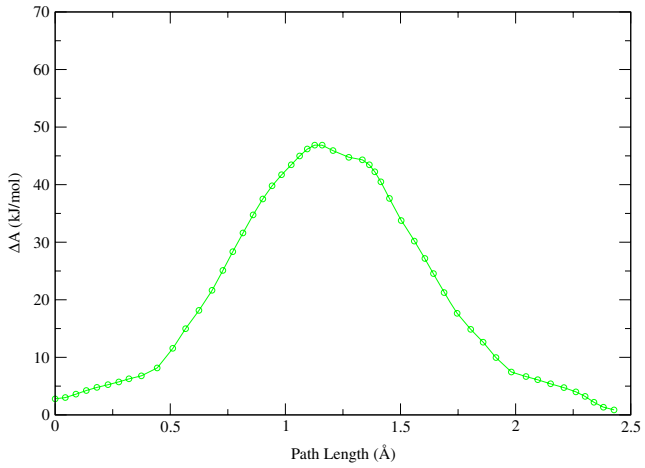

Figure 9. Free-energy profile along minimum free-energy path on the two-dimensional free-energy surface.

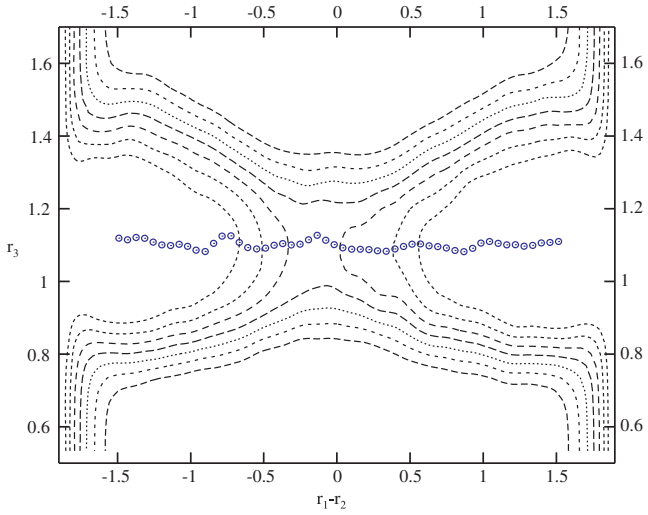

Figure 10. Contour plot of the free-energy surface obtained using metadynamics with two collective variables $r_1 - r_2$ and r_3 (see figure 2). The converged minimum free-energy path is shown.

In section 4 we have briefly discussed the quantitative difference between barriers obtained from the thermodynamic integration with one and two collective variables. Similarly, the barrier found by analysis of the two-dimensional free-energy profile obtained using metadynamics is not, due to the different number of degrees of freedom contributing to entropy, exactly equivalent to the simulation with only one collective variable. In contrast to thermodynamic integration, metadynamics offers a simple way to relate results from simulations with different numbers of collective variables. Using the relation between the $n-1$- and n-dimensional partition functions,

$$Q(q_2, \ldots, q_n) = \int_{-\infty}^{\infty} \mathrm{d}q_1 \, Q(q_1, q_2, \ldots, q_n), \tag{21}$$

the correct lower-dimensional free-energy surface can be obtained:

$$A(q_2, \ldots, q_n) = -k_\mathrm{B}T \times \ln\left(\int_{-\infty}^{\infty} \mathrm{d}q_1 \, \exp\left[-\frac{A(q_1, q_2, \ldots, q_n)}{k_\mathrm{B}T}\right]\right). \tag{22}$$

In order to make this procedure useful, it is necessary that all 'important' regions of the collective variable space are sufficiently sampled.

As we have seen, the efficiency of metadynamics decreases with an increasing number of collective variables used in the simulation. On the other hand, as the true reaction coordinate is often difficult to estimate for complex chemical reactions, it is a useful property of metadynamics that it allows us to identify those coordinates among the collective variables which are inactive. In order to demonstrate this property we have performed a simulation with two collective variables—one being active ($r_1 - r_2$) and another inactive (r_3), see figure 2. The resulting contour plot of the free-energy surface is shown in figure 10. Evidently, the minimum free-energy path is, except for small fluctuations, a straight line parallel to the $r_1 - r_2$ coordinate. The coordinate r_3 only oscillates around its equilibrium value of ∼1.10 Å and does not change significantly during the reaction. Upon integrating components perpendicular to the $r_1 - r_2$ coordinate,

$$A(r_1 - r_2) = -k_\mathrm{B}T \ln\left(\int_{-\infty}^{\infty} \mathrm{d}r_3 \, \exp\left[-\frac{A(r_1 - r_2, r_3)}{k_\mathrm{B}T}\right]\right), \tag{23}$$

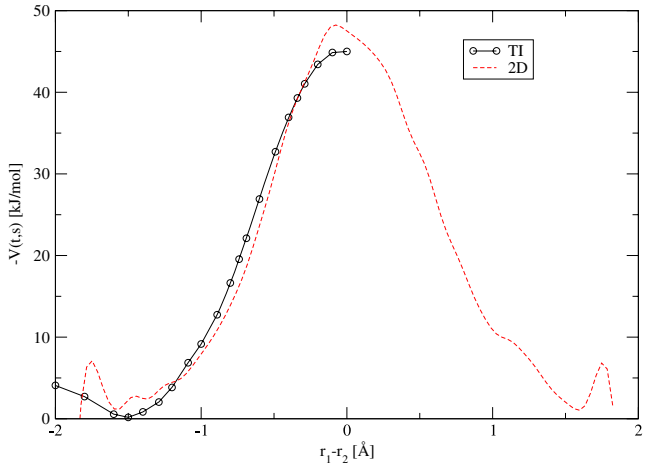

Figure 11. Comparison of the free-energy profiles along the $r_1 - r_2$ coordinate obtained using the thermodynamic integration (TI), and metadynamics with two collective variables $r_1 - r_2$ and r_3 (2D).

we obtain a free-energy profile equivalent to the one-dimensional case. The resulting free-energy profile is compared with that from thermodynamic integration in figure 11.

6. Conclusions

We have examined two popular free-energy methods used to determine reaction barriers. If the reaction mechanism is known, thermodynamic integration allows an efficient calculation of the reaction barrier as the collective variable space has to be sampled only along the reaction coordinate. The free-energy gradients can be used to determine minima and saddle points in the collective variable space. The statistical error can easily be controlled and simulations can be designed such as to reduce the total error in the free-energy difference below a prescribed threshold. The statistical error for the free-energy barriers consists of two parts: (i) the error due to the discretization of integral in equation (15) and (ii) the error due to imperfect convergence of the free-energy gradients. The first type of error can be reduced by use of a sufficiently dense integration grid, especially in regions where integrand changes rapidly. The error due to imperfect convergence of the free-energy gradients can be estimated as the variance of the gradients (note that the correlation between subsequent steps in the MD simulation should be taken into account [50, 51]) and can be reduced by performing sufficiently long MD simulations.

Metadynamics is a robust method which identifies the path of least resistance in the collective variable space. Unlike thermodynamic integration, metadynamics is able to explore new mechanisms, provided an adequate set of collective variables is defined. The statistical error in the free-energy difference determined by metadynamics is related to the setting of the adjustable parameters. Although qualitatively analyzed by Laio *et al* [45, 47], to our knowledge, a reliable error estimator which would allow us to quantify error at an arbitrary stage of simulation is not available.

Acknowledgments

This work has been supported by the Austrian Science Funds under project P19983. The use of facilities at the Computing Center of Vienna University is gratefully acknowledged. I thank Professor Jürgen Hafner and Dr Thomas Bligaard for very useful discussions.

References

[1] Jensen F 1999 *Introduction to Computational Chemistry* (Chichester: Wiley) p 301
[2] Torrie G M and Valleau J P 1977 *J. Comput. Phys.* **23** 187
[3] Kirkwood J G 1935 *J. Chem. Phys.* **3** 300
[4] Carter E A, Ciccotti G, Hynes J T and Kapral R 1989 *Chem. Phys. Lett.* **156** 472
[5] den Otter W K and Briels W J 2000 *Mol. Phys.* **98** 773
[6] Darve E, Wilson M A and Pohorille A 2002 *Mol. Simul.* **28** 113
[7] Fleurat-Lessard P and Ziegler T 2005 *J. Chem. Phys.* **123** 084101
[8] Laio A and Parinello M 2002 *Proc. Natl Acad. Sci. USA* **99** 12562
[9] Ianuzzi M, Laio A and Parinello M 2003 *Phys. Rev. Lett.* **90** 238302
[10] Rosso L, Mináry P, Zhu Z and Tuckerman M E 2002 *J. Chem. Phys.* **116** 4389
[11] Dellago C, Bolhuis P G and Geissler P 2002 *Adv. Chem. Phys.* **123** 1
[12] van Erp T S, Moroni D, Bolhuis P G and Geissler P 2003 *J. Chem. Phys.* **118** 7762
[13] Kresse G and Hafner J 1993 *Phys. Rev.* B **48** 13115
[14] Kresse G and Hafner J 1994 *Phys. Rev.* B **49** 14251
[15] Kresse G and Furthmüller J 1996 *Comput. Mater. Sci.* **6** 15
[16] Kresse G and Furthmüller J 1996 *Phys. Rev.* B **54** 11196
[17] Blöchl P E 1994 *Phys. Rev.* B **50** 17953
[18] Kresse G and Joubert D 1999 *Phys. Rev.* B **59** 1758
[19] Perdew J P, Chewary J A, Vosko S H, Jackson K A, Pedersen M R, Singh D J and Fiolhais C 1992 *Phys. Rev.* B **46** 6671
[20] Perdew J P and Wang Y 1992 *Phys. Rev.* B **45** 13244
[21] Nosé S 1984 *J. Chem. Phys.* **81** 511
[22] Hoover W G 1985 *Phys. Rev.* A **31** 1695
[23] Vetter R and Zülicke L 1990 *J. Am. Chem. Soc.* **112** 5136
[24] Deng L, Branchadell V and Ziegler T 1994 *J. Am. Chem. Soc.* **116** 10645
[25] Glukhovtsev M N, Pross A and Radom L 1995 *J. Am. Chem. Soc.* **117** 2024
[26] Glukhovtsev M N, Bach R D, Pross A and Radom L 1996 *Chem. Phys. Lett.* **260** 558
[27] Streitwieser A, Sik-Cheung Choy G and Abu-Hasanayn F 1997 *J. Am. Chem. Soc.* **119** 5013
[28] Botschwina P 1998 *Theor. Chem. Acc.* **99** 426
[29] Ensing B, Meijer E J, Blöchl P E and Baerends E J 2001 *J. Phys. Chem.* A **105** 3300
[30] Yang S-Y, Fleurat-Lessard P and Ziegler T 2004 *J. Phys. Chem.* A **108** 9461
[31] Ensing B, Laio A, Parinello M and Klein M L 2005 *J. Phys. Chem.* B **109** 6676
[32] Gonzales J M, Allen W D and Scheffer H F III 2005 *J. Phys. Chem.* A **109** 10613
[33] Song L, Wu W, Hiberty P C and Shaik S 2006 *Chem. Eur. J.* **12** 7458
[34] Ebrahimi A, Habibi M and Amirmijani A 2007 *J. Mol. Struct.: Theochem* **809** 115
[35] Fukui K 1970 *Phys. Chem.* **74** 4161
[36] Fukui K 1981 *Acc. Chem. Res.* **14** 363

[37] Hratchian H P and Schlegel H B 2002 *J. Phys. Chem.* A **106** 165
[38] Ryckaert J P, Ciccoti G and Berendsen H J C 1977 *J. Comput. Phys.* **23** 327
[39] Banerjee A, Adams N, Simons J and Shepard R 1985 *J. Chem. Phys.* **89** 52
[40] Càszàr P and Pulay P 1984 *J. Mol. Struct.* **114** 31
[41] Bofill J M 1994 *J. Comput. Chem.* **15** 1
[42] Powell M J D 1970 *Nonlinear Programing* ed J B Rosen, O L Mangasarian and K Ritter (New York: Academic)
[43] Powell M J D 1971 *Math. Prog.* **1** 26
[44] Murtagh B and Sargent R W H 1972 *Comput. J.* **13** 185
[45] Laio A, Rodriguez-Fortea A, Luigi Gervasio F, Ceccarelli M and Parinello M 2005 *J. Phys. Chem.* B **109** 6714
[46] Raiteri P, Laio A, Gervasio F L, Micheletti C and Parinello M 2006 *J. Phys. Chem.* B **110** 3533
[47] Bussi G, Laio A and Parinello M 2006 *Phys. Rev. Lett.* **96** 090601
[48] Gervasio F L, Laio A and Parrinallo M 2005 *J. Am. Chem. Soc.* **127** 2600
[49] Mills G, Jónsson H and Schenter G K 1995 *Surf. Sci.* **324** 305
[50] Flyvbjerk H and Petersen H G 1989 *J. Chem. Phys.* **91** 461
[51] Frenkel D and Smit B 2002 *Understanding Molecular Simulation* (London: Academic) pp 529–32