# Using a classical potential as an efficient importance function for sampling from an *ab initio* potential

Radu Iftimie
*Chemical Physics Theory Group, Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario M5S 3H6, Canada*

Dennis Salahub
*Steacie Institute for Molecular Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada and CERCA-Centre for Research on Computation and its Applications, 5160 Décarie Boulevard, Montréal, Québec, Canada*

Dongqing Wei
*CERCA-Centre for Research on Computation and its Applications, 5160 Décarie Blvd., Montréal, Québec, Canada*

Jeremy Schofield[a)]
*Chemical Physics Theory Group, Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario M5S 3H6, Canada*

(Received 19 May 2000; accepted 28 June 2000)

In this paper the *ab initio* potential of mean force for the formic acid–water system is calculated in a Monte Carlo simulation using a classical fluctuating charge molecular mechanics potential to guide Monte Carlo updates. The *ab initio* energies in the simulation are calculated using density-functional theory (DFT) methods recently developed by Salahub *et al.* [J. Chem. Phys. **107**, 6770 (1997)] to describe hydrogen-bonded systems. Importance sampling methods are used to investigate structural changes and it is demonstrated that using a molecular mechanics importance function can improve the efficiency of a DFT simulation by several orders of magnitude. Monte Carlo simulation of the system in a canonical ensemble at $T=300$ K reveals two chemical processes at intermediate time scales: The rotation of the $H_2O$ bonded to HCOOH, which takes place on a time scale of 3 ps, and the dissociation of the complex which occurs in 24 ps. It is shown that these are the only important structural ''reactions'' in the formic acid–water cluster which take place on a time scale shorter than the double transfer of the proton. © *2000 American Institute of Physics.* [S0021-9606(00)50936-2]

## I. INTRODUCTION

The calculation of approximate rate constants via transition state theory (TST) can be done in the framework of molecular simulations using Monte Carlo (MC) and molecular-dynamics (MD) techniques. Molecular simulations with classical potentials are very fast for reasonably large systems, and very good statistics for ensemble averages of properties of the system can be obtained. However, in many cases one is interested in systems which undergo chemical reactions or contain hydrogen bonds which are known to be only approximately described by such classical potentials. In such systems *ab initio* descriptions are often more appropriate. *Ab initio* methods have been used extensively to calculate energies of different configurations for molecules of several atoms and, in some cases, MD simulations have been performed using density functional theory (DFT) for systems of up to 100 atoms.[3] However, since *ab initio* calculations are computationally intensive, in many cases good sampling of the state space can be obtained only for processes which equilibrate quickly and are free of large energy barriers. The difficulty of proper sampling is more evident when one cannot constrain all fast degrees of freedom due to important coupling of the fast motions to quantities of interest in the study. It is generally very difficult to correctly sample multimodal distributions corresponding to energy surfaces where several minima are separated by regions of high energy.

One remedy for these problems consists of using umbrella functions[4,5] to enhance the sampling in regions of very low probability and to overcome barriers on the energy surface. Unfortunately the direct application of an MC or MD method in a quantum framework requires good knowledge of the free energy landscape, as the efficiency of the umbrella functions often depends on detailed features of the system. Simulations utilizing umbrella functions often rely on a time consuming self-consistent procedure to iteratively improve the sampling procedure.

Another alternative is to use multiple Markov chain (MMC) methods[6] in which Markov chains running at a higher temperature are used to promote transitions among different regions of high probability density. However, if the activation barriers are big, a large number of chains is needed as the probability of accepting the swapping between

[a)]Electronic mail: jmschofi@chem.utoronto.ca

 

the chains depends critically on the overlap of the distributions at different temperatures.

In this article, we propose the use of a simple molecular mechanics potential function of the formic acid–water cluster as the basis of a distribution from which updates are proposed in a simulation where the accurate (*ab initio*) calculation of the energy of a hydrogen-bonded system is necessary. Since the classical potential is several orders of magnitude faster to calculate than a DFT potential for any given configuration, it is possible to sample from the classical distribution in a fraction of the time needed for the calculation of an *ab initio* step. It is demonstrated that the use of an importance function based on the classical potential leads to improvements of several orders of magnitude in the efficiency of sampling from an *ab initio* based Boltzmann distribution, giving integrated correlation times less than 10 MC updates.

An outline of this paper is as follows. Section II describes the transition state theory (TST) for activated processes, correlation-reduction techniques, the model system, the procedure used to construct the classical potential and the simulation details. Section III presents the statistical results obtained from a long MC simulation and their detailed analysis. Conclusions are discussed in Sec. IV.

## II. THEORY AND METHOD

### A. Classical rate theory

The calculation of reaction rates has been an active field of research for several decades.[7,8] It was recognized early on that rate processes are events taking place on a time scale which is long compared to time scales characterizing the dynamics around states of local stability. Consider two domains of attraction, $A$ and $B$ in the phase space of the system. If the reaction coordinate $q$, which characterizes the dynamics of the transition between $A$ and $B$, is of lower dimensionality than the full system, then $q(t)$ obeys a kind of Stochastic dynamics in which the other degrees of freedom can either donate or remove energy. In order to make an escape from well $A$, the stochastic variable $q(t)$ must acquire energy to become activated toward the barrier and, upon reaching the barrier top, it must again lose energy to become trapped in the neighboring well $B$. In the following discussion we assume that there is a clear separation of time scale between the reactive process and the other time scales in the system. These time scales include the time of relaxation in either of the two locally stable wells, the correlation time of the noise, the time for a trajectory to cross the barrier region and the time to lose (or gain) the energy from the noise. We also assume that an initial transient time necessary for the degrees faster than the reaction to equilibrate has elapsed, and that the distribution for the slower degrees of freedom is essentially unchanged on the typical time scale of the reaction.

Under these conditions, simple rate expressions for the dynamical processes can be obtained by projection operator techniques.[9] Provided all these requirements are satisfied, the classical TST rate constant $k_{\mathrm{TST}}$ can be expressed as[10]

$$k_{\mathrm{TST}}=\frac{\langle \delta(q)\,\theta(\dot{q})\dot{q}\rangle}{\langle \theta(q)\rangle}, \tag{1}$$

where it has been assumed that the reaction coordinate function $q(\mathbf{x})$ characterizing the progress of the reaction depends only on the nuclear coordinates $\mathbf{x}$, and that the dividing surface between the reactant $A$ and the product regions $B$ of the phase space is chosen at $q(\mathbf{x})=0$. $\theta$ in Eq. (1) is the Heaviside step function which implies the reactant region is characterized by positive values of $q$. In Eq. (1), all averages $\langle\cdots\rangle$ correspond to integrals over the classical coordinates $\mathbf{x}$ and momenta $\mathbf{p}$ of the system weighted by the Maxwell–Boltzmann distribution characteristic of a canonical ensemble in equilibrium at temperature $T$

$$P_{\mathrm{eq}}(\mathbf{x},\mathbf{p})\propto\exp[-(K+U)/k_BT], \tag{2}$$

where $K$ and $U$ are the kinetic and potential energy of the classical system and $k_B$ is Boltzmann's constant. Because TST assumes that all trajectories with $\dot{q}>0$ or $\dot{q}<0$ at the transition state are reactive from left to right or from right to left, respectively, the TST is an upper bound estimate which neglects dynamical recrossings near the top of the barrier. The microscopic transition rate constant is expressed as[11]

$$k_r=\kappa k_{\mathrm{TST}}, \tag{3}$$

where the quantity $\kappa$ is called the transmission coefficient and has a value between 0 and 1. The transmission coefficient $\kappa$ is given by the plateau value of the reactive flux, $\kappa(t)$, where

$$\kappa(t)=\frac{\langle \theta(q(t))\dot{q}\,\delta(q)\rangle}{\langle \theta(\dot{q})\dot{q}\,\delta(q)\rangle}. \tag{4}$$

The reactive flux can be calculated by running dynamical trajectories that start at the top of the barrier at $t=0^+$. The recrossing factor $\kappa$ is essentially the fraction of trajectories which are stabilized after a transient relaxation in the state to which they were initially directed. For a chemical process, the TST transition rate in Eq. (1) can be written, after integration over the momenta, as

$$k_{\mathrm{STS}}=\sqrt{\frac{k_BT}{2\pi m_H}}\,\frac{\int d\mathbf{x}\exp(-U(\mathbf{x})/k_BT)w(\mathbf{x})\delta(q(\mathbf{x}))}{\int d\mathbf{x}\exp(-U(\mathbf{x})/k_BT)[q(\mathbf{x})\geqslant 0]}, \tag{5}$$

where the integral extends over the entire configurational space of the system, the prefactor for the fraction of averages is the thermal velocity of a hydrogen atom at temperature $T$, and the denominator of the ratio is the partition function of the reactant. In Eq. (5) we have used Iverson's convention,[12] which consists of placing a boolean expression in square brackets and requiring that the result is 1 if the expression is *True* and 0 if it is *False*. Hence, in Iverson's notation the Heaviside function $\theta(q(\mathbf{x}))$ is $[q(\mathbf{x})\geqslant 0]$. Note that in Eq. (5) the numerator has the dimension of $(q/x)$ and the denominator the dimension of $q$, which implies that the fraction has the overall dimension of the inverse of a distance. The weighting factor $w(x)$ arises from the momentum integration, and is given by

$$w(\mathbf{x}) = \sqrt{\sum_i \frac{m_H}{m_i}\left(\frac{\partial q}{\partial x_i}\right)^2}, \tag{6}$$

for a one-dimensional reaction coordinate. Note that $w(x)$ depends on the dimensionless mass ratios $m_i/m_H$ due to the prefactor in Eq. (5). In the case where the reaction coordinate is one of the Cartesian spatial coordinates $x_i$ of the system, $w(x)=1$ and the TST transition rate may be written as

$$k_{\mathrm{STS}} = \sqrt{\frac{k_B T}{2\pi m_i}} \frac{\exp(-\phi(0)/kT)}{\int_0^\infty dq \exp(-\phi(q)/k_B T)}, \tag{7}$$

where $\phi(q) = -kT \log\langle \delta(q(\mathbf{x})-q)\rangle$ is the potential of mean force. Equation (7) demonstrates that the potential of mean force is intimately related to the expression for the reaction rate in the TST approximation. As is evident from Eq. (5), the calculation of TST rate requires only straightforward configurational averages in a classical system.

### B. Correlation-reduction techniques

From a practical point of view an estimator for the fraction in Eq. (5) is given by

$$\frac{\int d\mathbf{x} \exp(-U(\mathbf{x})/k_B T) w(\mathbf{x}) \delta(q(\mathbf{x}))}{\int d\mathbf{x} \exp(-U(\mathbf{x})/k_B T)[q(\mathbf{x}) \geqslant 0]} = \hat{J} = \frac{\hat{J}_1}{\hat{J}_2}, \tag{8}$$

where

$$\hat{J}_1 = \frac{1}{\Delta q} \sum_{\mathbf{x}^i} w(\mathbf{x}^i)[0 \leqslant q(\mathbf{x}^i) \leqslant \Delta q], \tag{9}$$

$$\hat{J}_2 = \sum_{\mathbf{x}^i} [0 \leqslant 1(\mathbf{x}^i)], \tag{10}$$

where $\mathbf{x}_i = \{x_1^i, x_2^i, \ldots, x_n^i\}$ are the $n$ spatial coordinates of the system for configuration $i$ in the Markov Chain constructed in the simulation. In Eq. (9), $\Delta q$ should be chosen to be small enough so that $(1/\Delta q)[0 \leqslant q(\mathbf{x^i}) \leqslant \Delta \mathbf{q}]$ is a good approximation of the delta function $\delta(q)$, but large enough such that the variance of the estimator $\hat{J}_1$ is not too big in order to guarantee that there are enough points inside the interval. Note that $\hat{J}_1$ is an unbiased estimator for the quantity in Eq. (9).

A simulation can give only a confidence interval for the expected value of an estimator. Usually this is given in terms of 96% confidence intervals

$$\{E[\hat{J}] - 2\sqrt{Var[\hat{J}]}, E[\hat{J}] + 2\sqrt{Var[\hat{J}]}\}, \tag{11}$$

where $E[\hat{J}]$ is the estimated value from the simulation. The variance $Var[\hat{J}]$ of $\hat{J}$ depends on the number of independent points, or equivalently on the integrated correlation time. In general, if $\hat{J}$ is an estimator of a quantity obtained using the $N$ configurations which are the output of a Markov chain after eliminating the burn-in length of the simulation,[13] then

$$f = \frac{\sqrt{Var[\hat{J}]}}{E[\hat{J}]} \sim \frac{1}{\sqrt{n}}, \tag{12}$$

where $n$ is the number of independent points in the region of interest and $f$ is the ratio between the standard error and the estimated value. The accuracy of the estimation of the classical rate constant $k_{\mathrm{TST}}$ obtained from the simulation therefore increases as the square root of the number of independent points in the activated region.

In general, activation energies are much greater than the average thermal energy so that even if there were no correlation in the data from a MC simulation there would be few independent points at the top of the barrier. Usually two or more simulations are done, where the first simulation estimates the probability of finding the reaction coordinate near the dividing surface and the second simulation is used to obtain a better estimate for the potential of mean force very close to the dividing surface.

Several correlation-reduction techniques can be used to reduce the uncertainty in the estimation of $\hat{J}$. MC techniques have been reviewed extensively both in the physical and statistical literature.[14,15] They typically consist of employing either different updating techniques to suppress the random walk nature of the usual single-variable updating Metropolis method[16] for sampling within a single mode, or in the case of a multimodal distribution, introducing generalized distributions to overcome barriers. Suppressing the random walk behavior is important when sampling from a distribution with strongly correlated variables. When the variables are strongly correlated motion must proceed in small steps, and the difference in efficiency between diffusive and systematic exploration of the distribution can be very large. Although generalized distribution methods can be utilized to sample from a multimodal distribution where usual techniques demonstrate quasi-ergodic behavior, a good knowledge of the energy landscape is required in order to fully benefit from their use.

Dynamical sampling, consisting of methods of simulating the Hamiltonian dynamics, can be viewed as a variant of an ''auxiliary variable'' MC method where the conjugate momenta are the auxiliary variables. The resulting configurational space distribution function over the target ensemble is the marginal distribution over the extended state space spanned by the variables and their conjugate momenta. The advantage of the dynamical methods, especially hybrid Monte Carlo (HMC), is that the random walk nature of the sampling within a mode is in general suppressed.[17,18] One difficulty related to the use of *ab initio* Hamilton dynamical methods is that analytical derivatives of the energy with respect to the nuclear coordinates must be available to calculate the Hellmann–Feynman forces required for adiabatic MD. This also increases the cpu time for one dynamical update.

Another disadvantage of dynamical methods is that the integration time step must be smaller than the characteristic time of the fastest degree of freedom for the total energy to be maintained constant by the classical leap-frog algorithm. However, small integration time steps produce successive configurations which are highly correlated. On the other hand, too large a step size is disastrous, since the dynamical simulation becomes unstable, and very few changes are accepted. In general this problem can be solved using constrained dynamics[19] or multiple-time scale molecular

dynamics.[20] However, the use of an *ab initio* potential, which cannot be cast into one potential for the fast and another for the slow motions, and for which the fast O–H vibrations are correlated with the relative motion inside the cluster $HCOOH–H_2O$, precludes the use of both techniques.

A conclusion can be drawn from all these methods. In order to decrease the correlation of the output data one should use a distribution which mimics the *ab initio* density of states and from which configurations can be easily drawn. This general technique is called the importance function method.[13,21] If the probability density function for the equilibrium distribution of $\mathbf{x}$ is $f(\mathbf{x})$, then this procedure consists of using an auxiliary density function $g(\mathbf{x})$ corresponding to a different distribution, similar to the first and from which we can easily sample points according to their correct weight. In the context of Markov chain Monte Carlo simulations, this means that if the actual state of the chain is $\mathbf{x_i}$, the proposed point $\mathbf{x_f}$ drawn from the distribution $g(\mathbf{x})$ should be accepted with the probability

$$\min\left\{1, \frac{f(\mathbf{x}_f)}{g(\mathbf{x}_f)} \frac{g(\mathbf{x}_i)}{f(\mathbf{x}_i)}\right\}, \tag{13}$$

to ensure that the configurations in the Markov chain are asymptotically distributed according to $f(x)$. When trial states $\mathbf{x_f}$ are drawn from a distribution very similar to the target distribution, nearly all the points will be accepted and the successive configurations will be essentially uncorrelated provided that the trial states $\mathbf{x_f}$ is independent of $\mathbf{x_i}$. For complicated importance functions $g(x)$, the simplest means of sampling new trial configurations is to utilize an auxiliary Markov chain which has limiting distribution $g(x)$. Provided the trial states are selected as the output of sufficiently long intervals on the auxiliary Markov chain, each proposed state for the target Markov chain is essentially independent of the current state. In an *ab initio* MC simulation, drawing trial configurations from a classical distribution using a supplementary Markov chain will increase the total cpu time by a very small factor since calculating the classical energy of a configuration is several orders of magnitude faster than the same *ab initio* calculation.

### C. The model system

The sampling method outlined above has been tested on a model system composed of a cluster of formic acid and water, where the energies of different configurations were calculated using high quality DFT methods. *Ab initio* DFT calculations have proven to be very reliable in calculating properties for systems including transition metals where other *ab initio* methods are inadequate. One of the most stringent tests for contemporary DFT is the correct treatment of weak interactions. Such interactions play a crucial role in biological systems involving polypeptides and nucleic acids with a variety of internal hydrogen bonds. The necessity of high quality nonlocal exchange-correlation functionals in DFT studies of weak hydrogen bonding has been addressed by various studies.[22]

The nonlocal exchange-correlation schemes developed by Proynov, Vela, and Salahub[23] have shown particular promise for the description of hydrogen-bonded systems. In these schemes, the functional involves the Laplacian of the electron density (for each spin direction) and the kinetic energy density as ingredients reflecting inhomogeneity. This functional was combined and carefully synchronized with the generalized gradient approximation (GGA) exchange functionals of Becke[24] (BLAP) and Perdew[25] (PLAP scheme). Salahub *et al.* have demonstrated that the kinetic energy-dependent XC functionals (BLAP and PLAP) perform better than all GGA options (BP86, PP86, PW91), BLYP, or other hybrid methods (B3LYP, B3PW91) on systems involving intramolecular hydrogen bonds.[1] Predictions of all structural parameters obtained using the PLAP functional agree very well with experimental results. These predictions are also in agreement with high-quality post Hartree–Fock calculations [CCSD(T) and G2].

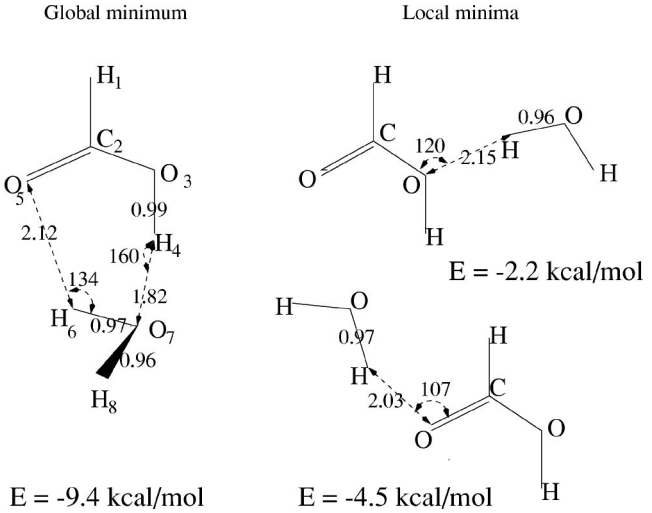


FIG. 1. The three conformations representing minima in the potential-energy surface. $E = 0.0$ Kcal/mol corresponds to the dissociated formic acid water complex.

In spite of its apparent simplicity, the formic acid–water system exhibits surprising complexity due to the hydrogen bond interactions. From geometry optimization calculations using DFT with a PLAP functional, the configuration of minimum energy was found to be more than 9 kcal/mol below the dissociation energy and involves a cycle of two strained hydrogen bonds (Fig. 1). Another two minima can be identified on the energy surface, corresponding to configurations containing single hydrogen bonds with different lone pairs of the oxygens. Although the energies of these minima are much higher, the structures have favorable free energies at higher temperatures due to entropic contributions. The competition between energetic and entropic factors and the presence of two types of hydrogen bonds, one involving an $sp^3$ and the other an $sp^2$ oxygen, makes the study important not only for a methodological test, but in its own right.

In addition to transitions from one minimum energy well to another, a dynamical process involving the exchange of protons between the formic acid and water is possible. Calculations of the energy of the transition-state structure for this process yielded an energy barrier of $\sim$17 kcal/mol.[26] A comparison with other systems with similar geometries and energy barriers allows us to infer that the tunneling of the

protons is essential in the double transfer reaction, and we estimate the reaction rate would be on the order of $10^7\ s^{-1}$ at room temperature.[27,28] The corresponding time scale of the proton transfer process is much larger than the inverse of the calculated rate constant for the dissociation of the cluster. Therefore, the assumption of the separation of the time scales, involving the slower degrees of freedom is satisfied.

### D. The classical potential

An important part of the sampling procedure outlined later consists of constructing simple, but qualitatively accurate, classical or semi-classical models for the system under study. The intramolecular part of the classical potential used in this study contains the usual molecular mechanical bond, angle and dihedral potentials. The values for the equilibrium bond distance, bond angles and the force constants were fitted using the *ab initio* potential. Although these values are similar to those found in typical molecular mechanics potentials, small differences in the equilibrium values for the internal coordinates produce a large variation in the total energy. Consequently, using the standard values from a molecular mechanics (MM) potential would produce great discrepancies between the MM and the *ab initio* potentials. The parameters for the Lennard-Jones interactions between the atoms were taken from the PROSIS[29] potential.

The classical description of the hydrogen bonds is one of the most difficult in molecular mechanics. Two classes of models have been proposed to describe the hydrogen bonds.[30] The first consists of using Morse or Lennard-Jones potentials. This emphasizes the isotropic intermolecular bonding, but lacks an orientational term. A remedy to this is to add a potential taking into account the linearity of the hydrogen bonds and the hybridization of the heteroatom.

Another class of models describes the hydrogen bond using electrostatic interactions. In simple molecular force fields the intermolecular Coulombic interaction is often modeled by point charges fixed on well-defined sites in the molecular frame. As charge induction effects are not additive, improved models include many-body potentials.

Fluctuating charge models, an alternative to explicitly devising many-body potentials, have been shown to give important improvements over the fixed charge models.[31] Itskowitz and Berkowitz have shown how one can derive a generalized formulation of the electronegativity equalization principle from DFT by leaving out a term due to the nonadditivity of the kinetic and exchange-correlation functional.[32] By choosing a spherical form for the perturbation of the density of an atom due to the surrounding molecular environment, they have transformed the energy functional into an energy function. Their final result gives the electrostatic energy as a function of the net charges, $Q_i$

$$E(Q_i)=E_0+\sum X_iQ_i+\sum J_{ii}Q_i^2+\tfrac{1}{2}\sum_{i\neq j}\sum J_{ij}Q_iQ_j, \tag{14}$$

where $E_0$ is a collection of terms independent of $Q_i$, $X_i$ is the electronegativity of the atom $i$, and $J_{ii}$ and $J_{ij}$ are the hardness of the atom $i$ and the screened coulomb interaction between atoms $i$ and $j$. Minimization of the function in Eq. (14) subject to the constraint of total charge conservation in the system

$$\sum_i Q_i=0, \tag{15}$$

gives the net charges $Q_i$. In general, although the electronegativities $X_i$ depend on the configuration of the system, they change little as the system evolves, provided no bond breaking or formation occurs.

The term $E_0$ contains the variation of the total energy with the other degrees of freedom not described by the electrostatic interaction. In particular it contains the variation of the energy with the internal degrees of freedom $\mathbf{R}$, which include bond lengths and angles as well as proper and improper dihedral angles. Care must be exercised in order not to overcount the electrostatic interactions. If we consider a single molecule $A$, the variation of the energy with respect to small deviations of the internal degrees of freedom from their equilibrium values is, to the second order

$$E_A=E_{\mathrm{eq}}+(\mathbf{R}-\mathbf{R}_{\mathrm{eq}})^T\left(\frac{\partial^2 E}{\partial\mathbf{R}\partial\mathbf{R}}\right)_{\mathbf{R}_{\mathrm{eq}}}(\mathbf{R}-\mathbf{R}_{\mathrm{eq}}). \tag{16}$$

In principle, this expression already incorporates the electrostatic energy arising from neighboring atoms and atoms separated by one or two atoms through bond length, bond angle and dihedral angle potentials. For example, in the case of the O–H bond, if we consider $E_A$ from Eq. (16) equal to $E_0$ in Eq. (14), the vibrations which shorten the bond will be favored by the electrostatic interaction as there are opposite charges on the atoms. Thus, in the absence of any internal field, the equilibrium values for the internal degrees of freedom will be modified.

In order to remedy this problem we have taken the net electrostatic energy in Eq. (14) to be

$$E_0=E_A-E' \tag{17}$$

with

$$E'=\sum_{\text{on molecule } l} E'_l, \tag{18}$$

where the summation in Eq. (18) is over individual molecules. $E'_l$ is the electrostatic energy of the molecule $l$ given by Eq. (14) with $J_{ij}=0$ for $i, j$ separated by more than two atoms. It represents the contribution to the electrostatic potential of the Coulombic interaction already counted in the mechanical potential in Eq. (16).

In general another choice has to be made for the conservation of the total charge in the case of a system containing more than one molecule. However, in the formic acid–water system we study only the relative motion of the water with respect to the formic acid and do not allow a chemical reaction to take place. Therefore, a reasonable choice is to impose the electroneutrality separately for HCOOH and $H_2O$.

The electronegativities and hardnesses of the atoms were taken to be constant during the simulation. We have used the

Klopman–Ohno approximation as proposed by Bakowies and Thiel[33] for the calculation of the mixed terms in the hardness kernel

$$J_{ij}=\frac{e^2}{\sqrt{\left(r_{ij}^2+\left(\frac{e^2}{2J_{ii}}+\frac{e^2}{2J_{jj}}\right)^2\right)}}, \tag{19}$$

where $e$ is the charge of the electron and $r_{ij}$ is the distance between the atoms $i$ and $j$. The Klopman–Ohno approximation incorporates screening effects for small atomic distances and smoothly converges to the classical Coulomb potential at large distances. Equation (19) was obtained for two spherically symmetric distributions of the charge (i.e., two $s$ atoms). Special care must be taken to describe the electrostatic interactions which involve O atoms due to the nonspherical symmetry of their lone pairs. We have chosen to represent each of the lone pairs on the oxygens by charges according to their hybridization. Both the $sp^2$O atom of the carbonyl group and the $sp^3$O of the hydroxyl were represented with a fluctuating charge on each lone pair as in the ST2 model of the water.[34]

The location of the fluctuating charges representing the lone pairs together with the values for the electronegativities and hardnesses of the atoms were optimized using the calculated *ab initio* energy of a small number of configurations. The values given by Bakowies and Thiel were taken as initial guess and were subsequently optimized iteratively. In all cases the iterative procedure introduced minor modifications of the initial values. It should be emphasized that the goal of this optimization procedure was not to obtain the best possible classical potential for this particular system, but to show that by including the usual interactions which exist in a molecular mechanics potential one can substantially improve the *ab initio* sampling.

### E. Computational methodology and simulation details

The Kohn–Sham (KS)-DFT calculations were carried out using a modified version of the LCGTO-DFT program deMon-KS (deMon-KS3.4).[1,2] A double-$\zeta$ plus polarization (DZVP) orbital basis set was used for all the atoms. The contraction pattern for the C and O atoms was (621/41/1*), and for the hydrogen atoms was (41/1*).[1] Auxiliary charge density (CD) and exchange-correlation (XC) fitting basis sets consisting of five $s$ functions and two sets of $s$, $p$, and $d$ functions with common exponents were used for C and O [denoted as (5, 2; 5, 2)] in combination with the DZVP orbital basis. Similarly, (5, 1; 5, 1) auxiliary patterns were employed for all H atoms. Using the auxiliary basis sets, the charge density was fitted analytically, while the XC potential was fitted numerically on a grid. The convergence level for the SCF (self-consistent field) energy using the auxiliary basis set was 0.01 Kcal/Mol. Everywhere in this study we have used a grid consisting of 64 radial shells (denoted as 64/extrafine grid option). The converged energy obtained for all test configurations using a larger grid of 128 radial shells differed from the value obtained using the smaller grid by less than the convergence level for the energy.

The MC simulation was carried in the canonical ensemble at $T=300\,\mathrm{K}$. Newly proposed configurations which are essentially statistically independent were drawn from an auxiliary Markov chain with the asymptotic classical distribution. Each trial configuration was obtained as the last state in a classical Markov chain generated from the current configuration using simple Metropolis single-variable updates. The integrated correlation time of this ''background'' classical simulation was 1000 MC steps. As the calculation of the classical potential function is ~50 000 times faster than the corresponding *ab initio* computation, only 2% of the total cpu time was spent to obtain the proposed configurations for the DFT Markov chain. Each configuration proposed by the classical MC chain was then accepted or rejected in the *ab initio* MC chain according to the usual Metropolis–Hastings algorithm.[13,14,21] If we denote $\mathbf{X}_{\mathrm{old}}$ and $\mathbf{X}_{\mathrm{new}}$ to be the previous and new trial configuration in the *ab initio* MC chain, the proposed state is accepted with the probability $\min\{1,\exp(-\Delta\Delta E/k_BT)\}$, where $\Delta\Delta E$ is the difference

$$\Delta\Delta E=(E_{\mathrm{DFT}}(\mathbf{X}_{\mathrm{new}})-E_{\mathrm{cl}}(\mathbf{X}_{\mathrm{new}}))-(E_{\mathrm{DFT}}(\mathbf{X}_{\mathrm{old}})-E_{\mathrm{cl}}(\mathbf{X}_{\mathrm{old}})). \tag{20}$$

$E_{\mathrm{DFT}}(\mathbf{X})$ and $E_{\mathrm{cl}}(\mathbf{X})$ are the potential energies of configuration $\mathbf{X}$ calculated by DFT methods and the classical potential, respectively. In the Appendix, it is shown this acceptance criterion guarantees that the *ab initio* Markov chain has the correct limiting Boltzmann distribution.

The classical simulation provides us with independent configurations within a fraction of the time necessary for a single DFT energy calculation. As each proposed configuration is essentially statistically independent from the other configurations, the only correlation in the *ab initio* Markov chain comes from the possibility of rejecting the proposed configuration and is a measure of the similarity between the classical and the quantum potential.

## III. THE RESULTS OF THE SIMULATION

The classical importance sampling method outlined in the previous section was tested on the formic acid–water cluster based on two different classical models of the system. The first model (model I) consisted of standard intramolecular bond length, bond angle, and dihedral potentials fitted from the DFT energies of a small set of configurations. Intermolecular interactions were treated with Lennard-Jones and fluctuating charge interactions, as described in the previous section. The second model (model II) contained an additional potential to encourage the formation of linear hydrogen bonds. There is no clear reason *a priori* to expect that either of these models is more accurate than the other.

The global minimum of the potential-energy surface for model I has a geometry which is very similar to that obtained from minimizing the *ab initio* potential energy, shown in Fig. 1. Model II, on the other hand, exhibits two low-energy minima which have slightly different geometries than that of the *ab initio* global minimum. The first minimum corresponds to a configuration in which the hydroxyl hydrogen bond is shortened ($d_{\mathrm{H_4O_7}}=1.73\,\text{Å}$) and has almost linear geometry ($\mathrm{O_3\widehat{H_4}O_7}=170°$), while the carbonyl hydrogen bond

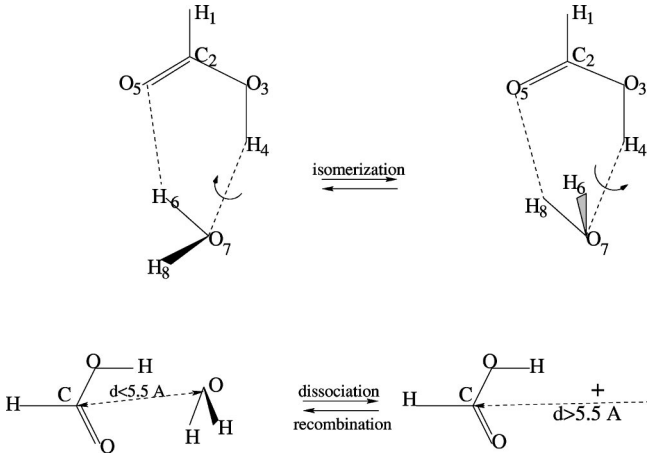


FIG. 2. The isomerization and dissociation–recombination reactions. In the isomerization reaction, the water rotates around the $O_7H_4$ hydroxyl hydrogen bond, which is also the direction of a lone pair on $O_7$. The carbonyl hydrogen bond $O_5H_6$ is broken and the hydrogen bond $O_5H_8$ is formed. The dissociation consists of breaking both carbonyl and hydroxyl hydrogen bonds.

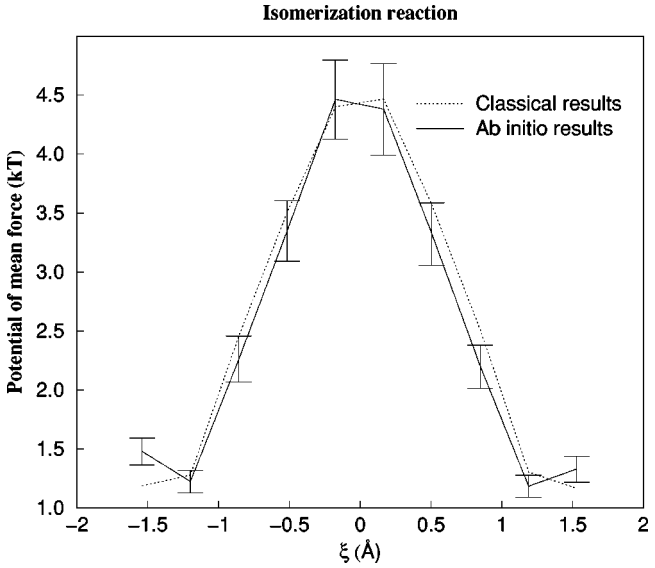


FIG. 3. The weighted potential of mean force for the isomerization reaction for model II.

interaction is very weak ($d_{H_6O_5}=2.4$ Å, $O_5\widehat{H_6}O_7=120°$). The second minimum ($d_{H_4O_7}=2.35$ Å, $O_3\widehat{H_4}O_7=144°$, $d_{H_6O_5}=2.1$ Å, $O_5\widehat{H_6}O_7=160°$) has a bent hydroxyl hydrogen bond and a carbonyl hydrogen bond closer to linear.

Although it is immediately apparent from the considerations detailed above that model II has qualitatively incorrect features incorporated into its potentials, both classical models were applied in separate simulations in order to study how the performance of the molecular mechanics based importance function method depends on the quality of the classical description.

The two activated processes found to take place on a time scale smaller than the double transfer of the proton are depicted in Fig. 2. Although the integrated correlation time at $T=300$ K varies with the quantity estimated, all the random variables we have studied had correlation lengths on the order of 10 MC steps on average for both models, with slightly shorter correlation lengths for the simulation based upon model I. In particular, the correlation times for the dissociation reaction and isomerization reaction obtained using the classical potential of model II as an importance function were 8 and 16 MC steps, respectively for the isomerization and the dissociation reactions (see Fig. 2).

The integrated correlation lengths of a simple *ab initio* Metropolis MC simulation, on the other hand, is estimated to be roughly the same as that obtained from a pure classical simulation, which is on the order of 2000 MC updates. Thus the importance sampling method reduces the correlation length by two orders of magnitude. Therefore it is evident that a simulation performed using the importance sampling obtains estimates for the rate constants which are comparable in accuracy to unbiased MC or MD simulations which are effectively 100 times longer.

The two hydrogen bonds present in the system, formed by the carbonyl C═O group and by the hydroxyl O–H group with the water molecule, were observed to have different strengths in the simulation. The hydroxyl hydrogen bond is stable on the time scale of the isomerization reaction, while the hydrogen bond with the carbonyl group dissociates. The isomerization reaction consists of breaking and reforming the C═O hydrogen bond, possibly with a different proton. Several order parameters for this process were investigated and the largest TST estimate for the reaction rate was obtained using the reaction coordinate

$$\xi = d_{O_5H_6} - d_{O_5H_8}, \tag{21}$$

where $d_{ij}$ is the distance between the atoms $i$ and $j$ in the cluster. For this reaction coordinate the weighting factor from Eq. (6) is

$$w_\xi = \sqrt{\left(2+\frac{2m_H}{m_O}(1-\cos\theta)\right)}, \tag{22}$$

where $\theta$ is the angle formed by the vectors $\overrightarrow{\mathbf{O_5H_6}}$ and $\overrightarrow{\mathbf{O_5H_8}}$. The weighted potential of mean force corresponding to $\xi$ is depicted in Fig. 3. The estimated TST rate constant obtained without an auxiliary sampling of the transition region is $28.9\pm4.3\cdot10^{10}\,s^{-1}$. This rate constant was calculated considering only the configurations for which the $O_3$–$H_4\cdots O_7$ hydrogen bond was intact (i.e., the distance $H_4\cdots O_7$ smaller than 2.4 Å). As the isomerization process requires that the water be bound to the formic acid, the isomerization in the cluster is a constrained reaction. Therefore, the TST description applies for the isomerization reaction only if the time scale for the isomerization is well-separated from the time scale for the breaking of the $O_3$–$H_4\cdots O_7$ hydrogen bond, which corresponds to the dissociation of the cluster.

The second reaction consists of the dissociation of the cluster in which both C═O and O–H hydrogen bonds are broken. The order parameter to describe this process is

$$\eta = d_{C_2O_7}. \tag{23}$$

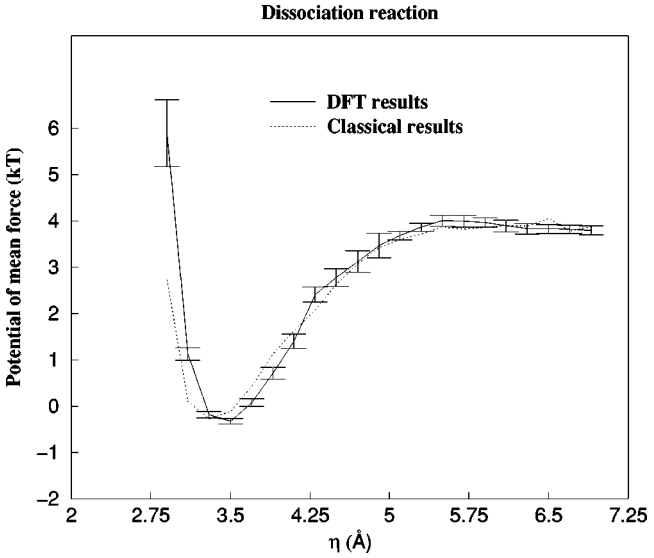


FIG. 4. The weighted potential of mean force for the dissociation reaction for model II. To increase the accuracy of the simulation in the transition region, a short simulation constrained near the top of the barrier was carried out.

The associated weighting factor $w_\eta$ is the square root of the inverse dimensionless reduced mass of C and O

$$w_\eta = \sqrt{\left(\frac{m_H}{m_O} + \frac{m_H}{m_C}\right)}. \tag{24}$$

Because the error bars are quite large in the transition region due to the small probability to obtain points in this region, the location of the top of the potential is not certain. To improve the statistics in the transition region ($5.0\,\text{Å} < \eta < 7.0\,\text{Å}$), a short simulation of 4000 points constrained to the transition region was performed and the improved statistics is shown in Fig. 4. The TST estimation for the rate of the dissociation reaction is $4.14 \pm 0.33 \cdot 10^{10}\,\text{s}^{-1}$, where the uncertainties are expressed as 96% confidence intervals.

The ratio between the two rate constants in the TST approximation is $\sim 7$. This means that even if corrections due to recrossing factors are computed it is likely that the time scales are well separated, and hence the approximations in the previous section are justified *a posteriori*. Although no attempt to calculate the recrossing probabilities in the two processes described above was made, for each activated process different choices of reaction coordinate were considered. The two choices specified above correspond to the reaction coordinates which yield the largest TST estimate for the reaction rates, and therefore, provide a good approximation to the true microscopic rate constants since these choices should minimize the corresponding recrossing factors.

We have carefully analyzed other possible activated processes using the results from the simulation and found no other reactions. In particular, the two minima on the potential-energy surface have almost the same probability as the dissociated configurations, confirming the importance of the entropic factors at this temperature.

It should be emphasized that in constructing the classical potential the van der Waals repulsions between the valence electrons of nonbonded atoms were less strong than suggested by the *ab initio* potential. Care has been exercised to obtain a broader distribution for the classical potential than the *ab initio* distribution to insure that all relevant regions of state space have been sampled. The potential of mean force for the two processes described above using the second classical potential of model II as importance function are also depicted in Figs. 3 and 4.

Another point which should be emphasized is that importance sampling enables *all* the relevant state space to be sampled in one simulation, allowing activated processes taking place on time scales which differ by one order of magnitude or more to be identified from one simulation. This is to be contrasted with umbrella sampling where the system is constrained to one region of the state space and different simulations are needed to study different activated processes identified at the outset of the simulation.

Although the agreement between the classical and the *ab initio* potential of mean force describing the processes in Figs. 3 and 4 is very good, the classical density of states does not mirror exactly the *ab initio* canonical probability density. In Fig. 6 a histogram of the differences $\Delta\Delta E$, defined in Eq. (20), obtained in the simulation at $T = 300\,\text{K}$ is shown for model II. The average error is $\sim 2kT$ and the standard deviation is $\sim 5kT$. The magnitude of the average error suggests that the excellent agreement between classical and quantum results in Figs. 3 and 4 is somewhat fortuitous and can be attributed to a cancellation of errors due to the averaging procedure involved in calculating potentials of mean force.

Because the classical potential for model II gives different densities of states for the low-energy conformations involving both hydroxyl and carbonyl hydrogen bonds, a simulation of the undissociated cluster at a lower temperature of $T = 200\,\text{K}$ was carried out to examine these differences. The classical potential based importance sampling method previously described using the first classical potential as importance function (model I) produced almost the same integrated correlation times (10 MC steps) as in the simulation at $T = 300\,\text{K}$. On the other hand, the qualitative difference between the classical potential of model II and the *ab initio* system made the direct application of the importance sampling technique somewhat difficult at the lower temperature. Global updating of all the variables in the ''background'' classical simulation led to proposal configurations which were rejected 90% of the time by the Metropolis–Hasting test in the *ab initio* simulation. In the work of Roberts, Gelman, and Gilks,[37] the problem of optimal scaling of the acceptance probability for random-walk Metropolis algorithms was considered. It was proved that for certain target distributions, the asymptotic acceptance probability should be tuned to be $\sim 0.2$. However, this result does not apply to more general Hastings algorithms such as importance sampling methods. If the proposal density makes use of the structure of the target density, intuition suggests that a higher acceptance probability is likely to be preferred. Thus, 80% rejection rate can be viewed as an upper bound for efficient importance sampling. Although essentially independent configurations were accepted 10% of the time on average in the simulation of model II, long periods of several hundred re-

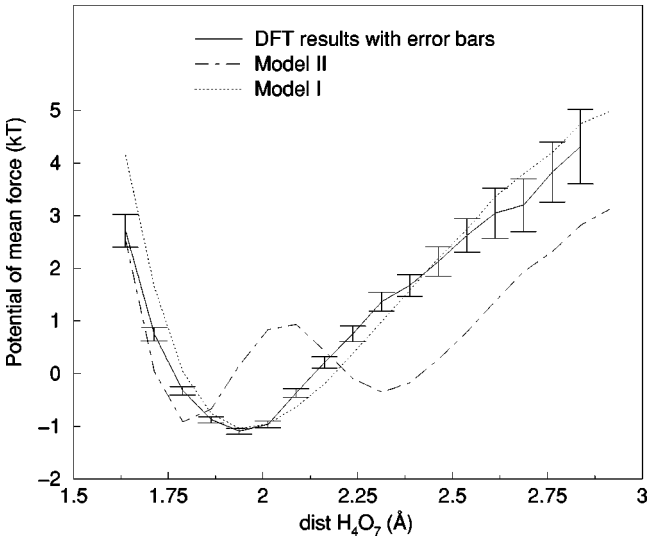


FIG. 5. The weighted potential of mean force for the distance $H_4O_7$ for models I and II along with that calculated from the *ab initio* simulation.

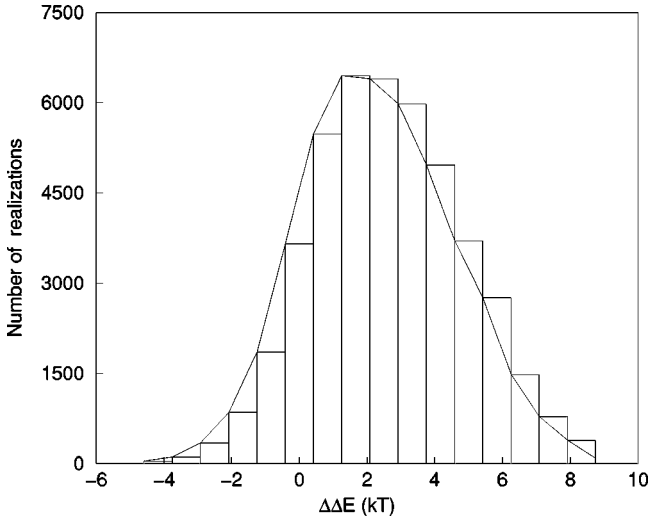


FIG. 6. Histogram of difference $\Delta\Delta E$ defined in Eq. (20) for model II. Note the average error $\sim 2kT$ and the standard deviation $\sim 5kT$.

jections occurred, which lead to significantly longer correlation lengths than those observed in the simulation of model I. In principle, as noted above, the sampling can be substantially improved if long periods of rejections are avoided at the cost of introducing correlation into the proposed configurations. To improve the acceptance ratio and hence the mobility of the simulation, the configurational variables were divided into three groups, representing the internal degrees of freedom, the variables describing the relative position of the center of mass of $H_2O$ with respect to the center of mass of HCOOH, and the Euler angles describing the relative orientation of $H_2O$ plane with respect to HCOOH. Although the overall acceptance of updates was improved to 30%, periods of rejections of 100 configurations were still observed when the system was sampling regions of the state space where the classical potential underestimated the *ab initio* probability density function by a large amount (see Fig. 5). In order to prevent the system from becoming trapped in regions of the state space where the difference $\Delta E = E_{\mathrm{DFT}} - E_{\mathrm{cl}}$ was very negative, the proposal states were obtained by a combination of importance sampling using the classical density of states and single-variable Metropolis updates of the *ab initio* Markov chain. The frequency and nature of the Metropolis updating was selected to give an asymptotic acceptance of 40%. The inclusion of the single-variable Metropolis updates reduced the maximum rejection lengths to around 20 configurations. As can be seen in Fig. 5, the double-well potential characteristic of the second classical potential utilized led to few proposed updates in the region of large *ab initio* density of states. In compensation, the accepted configurations had a small probability to leave this region. The role of the Metropolis sampling with the *ab initio* potential was to improve the mobility in those regions. It is also noteworthy that, in general, the region where $\Delta E$ is very negative does not coincide with the region of high *ab initio* density of states. In this case the *ab initio* single-Metropolis updating would drive the simulation away from the region where the classical and the quantum probabilities are very different. Another point worth noting is that the second classical potential is not reflected in the *ab initio* potential of mean force, which demonstrates that the importance sampling method performs well even where there is a qualitative difference between the classical and the *ab initio* potentials.

The second classical potential (model II) has the property that the region of high *ab initio* probability density corresponds to a region of very low classical density of states. Hence the improvements due to the combining with the *ab initio* Metropolis method are somewhat limited and the simulation using the second classical potential provides an upper bound estimate to the increase in the correlation time when a different but reasonable classical importance function is utilized in conjunction with *ab initio* Metropolis sampling. The integrated correlation time for the distance $H_4 \cdots O_7$ for model II was on the order of 90 MC updates, reflecting the time needed for the system to leave the region of large *ab initio* density of states. This is about nine times greater than the integrated time obtained at the same temperature with the first classical potential. However the integrated correlation times for the other variables in the simulation using model II, including $\xi$ from Eq. (21) describing the isomerization reaction, were increased only by a factor of 3 with respect to the simulation using the first classical potential.

## IV. CONCLUSION

In summary, an MC molecular mechanics based distribution was utilized to propose updates in an *ab initio* MC study of the dynamics of a cluster containing hydrogen bonds. The utilization of the molecular-mechanics-based importance function decreases the correlation time of the *ab initio* MC calculation by two orders of magnitude. Simulations performed without importance sampling require much more computational time to obtain comparable levels of accuracy. Furthermore, the MC importance sampling method enables a thorough sampling of the relevant configurational space which allows accurate estimates of rate constants to be obtained from relatively short simulations. In comparison,

based on the results for the reaction rates reported in the previous section, a typical unbiased MD simulation of this system would require, on average, several thousand configurations before one of the reactions studied could be observed at $T=300$ K. Umbrella sampling techniques can be used to reduce the correlation time of the *ab initio* MD simulation substantially, but usually a separate simulation is required for each activated process of interest. Although the umbrella function is simple to construct for reactive processes in which the density of states is relatively constant along the reaction coordinate, significant variation of the density of states is often observed, particularly for reactions involving conformational changes or solvent effects. For such systems, one must either resort to a computationally intensive iterative procedure to construct and refine the umbrella function or explicitly provide a description of the variation of the density of states along the reaction coordinate based on knowledge of the system. Umbrella sampling methods can also be incorporated into the molecular mechanics based importance sampling method to improve sampling of rare events. In fact simulations of the classical model can be used to construct the umbrella function for the *ab initio* Monte Carlo simulation very rapidly.

If the agreement between the classical and the *ab initio* densities of states is only moderate, separating the variables to be updated in a classical MC step into several groups will improve the mobility of the simulation. If possible, one group should contain variables which are strongly correlated. In general, separating the variables into groups should permit efficient sampling of configurational space when the distribution of states which are poorly estimated by the guiding potential is relatively random, even for classical potentials which overestimate or underestimate the energy by a few $kT$.

If long periods of MC rejections still exist, this is an indication that there is an important disagreement between the classical and the *ab initio* density of states which is very likely localized in some region of the state space. A remedy for this problem is to combine the importance sampling described above with another method, which generates a different Markov chain dynamics. For example, configurations proposed with a different classical potential, including umbrella sampling potentials, could be used to move the simulation away from the problematic region of state space. Another means to avoid becoming trapped in phase space would be to use *ab initio* Metropolis updates and the usual Metropolis criteria for acceptance. An equally good solution is to combine importance sampling with *ab initio* MD when accurate calculation of the forces is possible. We have demonstrated that these two simple techniques are enough to obtain integrated times which are at most one order of magnitude bigger than in the case where a very good classical potential is available.

If the classical potential differs from the *ab initio* potential by significant amounts, neither of the methods described above works efficiently and the classical description is probably missing an important interaction. In general it should be possible to construct a sampling potential of adequate quality given that many studies in the literature have demonstrated that a classical description of the intermolecular forces containing all important interactions is possible.[35,36] The classical potential based importance sampling method described in this article provides an easy way of performing *ab initio* sampling, provided the classical description constitutes a reasonable approximation to the real interactions. It should be emphasized that the importance sampling function need not be entirely classical in nature, so long as it requires little computational time.

Another study related to the double transfer of the protons in the same cluster is in progress. The construction of classical potentials describing chemical reactions in reasonably good agreement with *ab initio* results is more challenging, however, we expect that the quality of many of the potentials proposed in the literature would be adequate to be used as importance functions in a quantum simulation for a variety of situations where chemical bonds are broken and formed. More work on the subject would be of clear interest.

## ACKNOWLEDGMENTS

The authors would like to thank Raymond Kapral for many interesting discussions. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

## APPENDIX A

In this Appendix, it is shown that the acceptance criterion given by Eq. (20) guarantees that the *ab initio* Markov chain has the correct limiting Boltzmann distribution. This follows if the Markov chain satisfies microscopic reversibility for consecutive states $\mathbf{X}_0$ and $\mathbf{X}_n$:

$$P(\mathbf{X}_0)T(\mathbf{X}_0\rightarrow\mathbf{X}_n)=P(\mathbf{X}_n)T(\mathbf{X}_n\rightarrow\mathbf{X}_0), \tag{A1}$$

where $T(\mathbf{X}_i\rightarrow\mathbf{X}_j)$ is the transition matrix from state $\mathbf{X}_i$ to state $\mathbf{X}_j$ for the Markov process. The transition matrix $T(\mathbf{X}_i\rightarrow\mathbf{X}_j)$ can be written as a product of the probability $P_g(\mathbf{X}_i\rightarrow\mathbf{X}_j)$ of generating configuration $\mathbf{X}_j$ from configuration $\mathbf{X}_i$ and the conditional probability $P_A(\mathbf{X}_j|\mathbf{X}_i)$

$$T(\mathbf{X}_i\rightarrow\mathbf{X}_j)=P_g(\mathbf{X}_i\rightarrow\mathbf{X}_j)P_A(\mathbf{X}_j|\mathbf{X}_i). \tag{A2}$$

Equation (A1) follows immediately from Eq. (A2) provided $P_A$ is defined to be

$$P_A(\mathbf{X}_n|\mathbf{X}_0)=\min\left(1,\frac{P(\mathbf{X}_n)P_g(\mathbf{X}_n\rightarrow\mathbf{X}_0)}{P(\mathbf{X}_0)P_g(\mathbf{X}_0\rightarrow\mathbf{X}_n)}\right). \tag{A3}$$

The auxiliary classical Markov chain importance sampling procedure consists of taking the current configuration $\mathbf{X}_0$ and applying the classical transition matrix $T_{\mathrm{cl}}k+1$ times, where $k$ is a large number. The classical transition matrix $T_{\mathrm{cl}}$ is required to satisfy microscopic reversibility for the classical model

$$P_{\mathrm{cl}}(\mathbf{X}_i)T_{\mathrm{cl}}(\mathbf{X}_i\rightarrow\mathbf{X}_j)=P_{\mathrm{cl}}(\mathbf{X}_j)T_{\mathrm{cl}}(\mathbf{X}_j\rightarrow\mathbf{X}_i), \tag{A4}$$

where $P_{\mathrm{cl}}$ is the canonical distribution function for the classical system. The application of $T_{\mathrm{cl}}$ starting from $\mathbf{X}_0$ generates a sequence of states $\{\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k,\mathbf{X}_n\}$ with *path* probability

$$P_g(\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k,\mathbf{X}_n)=T_{\mathrm{cl}}(\mathbf{X}_0\rightarrow\mathbf{X}_1)\times T_{\mathrm{cl}}(\mathbf{X}_1\rightarrow\mathbf{X}_2)\cdots T_{\mathrm{cl}}(\mathbf{X}_k\rightarrow\mathbf{X}_n), \tag{A5}$$

while the probability for the reverse path is

$$P_g(\mathbf{X}_n,\mathbf{X}_k,\ldots,\mathbf{X}_1,\mathbf{X}_0)=T_{\mathrm{cl}}(\mathbf{X}_n\rightarrow\mathbf{X}_k)\cdots T_{\mathrm{cl}}(\mathbf{X}_1\rightarrow\mathbf{X}_0). \tag{A6}$$

Since microscopic reversibility holds for the classical Markov chain, the ratio of the probabilities of the reverse and forward paths is

$$\begin{aligned}\frac{P_g(\mathbf{X}_n,\mathbf{X}_k,\ldots,\mathbf{X}_1,\mathbf{X}_0)}{P_g(\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k,\mathbf{X}_n)}&=\frac{T_{\mathrm{cl}}(\mathbf{X}_1\rightarrow\mathbf{X}_0)}{T_{\mathrm{cl}}(\mathbf{X}_0\rightarrow\mathbf{X}_1)}\cdots\frac{T_{\mathrm{cl}}(\mathbf{X}_n\rightarrow\mathbf{X}_k)}{T_{\mathrm{cl}}(\mathbf{X}_k\rightarrow\mathbf{X}_n)}\\&=\frac{P_{\mathrm{cl}}(\mathbf{X}_0)}{P_{\mathrm{cl}}(\mathbf{X}_1)}\cdots\frac{P_{\mathrm{cl}}(\mathbf{X}_k)}{P_{\mathrm{cl}}(\mathbf{X}_n)}\\&=\frac{P_{\mathrm{cl}}(\mathbf{X}_0)}{P_{\mathrm{cl}}(\mathbf{X}_n)}.\end{aligned} \tag{A7}$$

Using Eq. (A7) into Eq. (A3), one obtains the result that the proposed state $\mathbf{X}_n$ should be accepted with probability

$$P_A(\mathbf{X}_n|\mathbf{X}_0)=\min\left\{1,\exp\left(-\frac{\Delta\Delta E}{k_BT}\right)\right\}, \tag{A8}$$

where $\Delta\Delta E$ is defined in Eq. (20) in the text. Note that the procedure outlined above shows that microscopic reversibility holds for all paths with intermediate states $\mathbf{X}_1,\ldots,\mathbf{X}_k$. Since the intermediate states are arbitrary and

$$P_g(\mathbf{X}_0\rightarrow\mathbf{X}_n)=\sum_{\mathbf{X}_1,\ldots,\mathbf{X}_k}P_g(\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k,\mathbf{X}_n), \tag{A9}$$

the weaker condition (A1) holds as well. The path sampling methods outlined here are easily generalized to allow for annealing, quenching, and generalized multiple Markov chain Monte Carlo methods.[38]


[1] S. Sirois, E. I. Proynov, D. T. Nguyen, and D. R. Salahub, J. Chem. Phys. **107**, 6770 (1997); E. I. Proynov, S. Sirois, and D. R. Salahub, Int. J. Quantum Chem. **64**, 427 (1997).
[2] For details on the deMon quantum chemistry package, see http://www.cerca.umontreal.ca/deMon/Welcom_f.html.
[3] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, J. Chem. Phys. **99**, 9080 (1993).
[4] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[5] A. Gelman and X. L. Meng, Stat. Sci. **13**, 163 (1998).
[6] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[7] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes*, 1st ed. (McGraw-Hill, New York and London, 1941).
[8] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[9] H. Mori, Prog. Theor. Phys. **33**, 423 (1965); R. Zwanzig, Phys. Rev. **124**, 983 (1961).
[10] H. Eyring, J. Chem. Phys. **3**, 107 (1935).
[11] D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
[12] K. E. Iverson, *A Programming Language* (Wiley, New York, 1962), p. 11; see also D. E. Knuth, Am. Math. Monthly **99**, 403 (1992).
[13] G. S. Fishman, *Monte Carlo* (Springer, New York, 1996).
[14] *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman and Hall, London, 1996).
[15] R. M. Neal, Technical Report No. 9722, Department of Statistics, University of Toronto, 1997.
[16] R. M. Neal, Technical Report No. 9508, Department of Statistics, University of Toronto, 1995.
[17] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[18] R. M. Neal, J. Comput. Phys. **111**, 194 (1994).
[19] G. Ciccotti and M. Ferrario, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1997).
[20] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).
[21] P. Bratley, B. L. Fox, and L. E. Schrage, *A Guide to Simulation* (Springer, New York, 1987).
[22] V. Barone and C. Adamo, J. Chem. Phys. **105**, 11007 (1996).
[23] E. I. Proynov, A. Vela, and D. R. Salahub, Chem. Phys. Lett. **230**, 419 (1994); **234**, 462(E) (1995).
[24] A. D. Becke, Phys. Rev. A **38**, 3098 (1998).
[25] J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Academic, Berlin, 1991).
[26] J. F. Truchon, D. Wei, and D. R. Salahub (in preparation).
[27] A. Fernandez-Ramos, Z. Smedarchina, W. Siebrand, M. Z. Zgierski, and M. A. Rios, J. Am. Chem. Soc. **121**, 6280 (1999).
[28] Z. Smedarchina, W. Siebrand, A. Fernandez-Ramos, L. Gorb, and J. Leszczynski, J. Chem. Phys. **112**, 566 (2000).
[29] J. Kolafa, PROSIS project developed at the Center for Applied Mathematics, Odense University, Denmark.
[30] A. R. Leach, *Molecular Modeling* (Longman, Singapore, 1996).
[31] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).
[32] P. Itskowitz and M. L. Berkowitz, J. Phys. Chem. A **101**, 5687 (1997).
[33] Bakowies and Thiel, J. Comp. Chem. **17**, No. 1, 88 (1996).
[34] F. H. Stillinger and A. Rahman, J. Chem. Phys. **60**, 1545 (1974).
[35] J. Bentzien, R. P. Muller, J. Florian, and A. Warshel, J. Phys. Chem. B **102**, 2293 (1998).
[36] D. E. Sagnella and M. E. Tuckerman, J. Chem. Phys. **108**, 2073 (1998).
[37] G. O. Roberts, A. Gelman, and W. R. Gilks, Research Rep. No. 94-16, Statistical Laboratory, University of Cambridge.
[38] S. Opps and J. Schofield (in preparation).