An efficient Monte Carlo method for calculating *ab initio* transition state theory reaction rates in solution

Radu Iftimie
Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada and New York University, Department of Chemistry, New York, New York 10003

Dennis Salahub
University of Calgary, 2500 University Drive N.W., Calgary, Alberta T2N 1N4, Canada

Jeremy Schofield[a)]
Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

(Received 31 July 2003; accepted 5 September 2003)

In this article, we propose an efficient method for sampling the relevant state space in condensed phase reactions. In the present method, the reaction is described by solving the electronic Schrödinger equation for the solute atoms in the presence of explicit solvent molecules. The sampling algorithm uses a molecular mechanics guiding potential in combination with simulated tempering ideas and allows thorough exploration of the solvent state space in the context of an *ab initio* calculation even when the dielectric relaxation time of the solvent is long. The method is applied to the study of the double-proton transfer reaction that takes place between a molecule of acetic acid and a molecule of methanol in tetrahydrofuran. It is demonstrated that calculations of rates of chemical transformations occurring in solvents of medium polarity can be performed with an increase in the cpu time of factors ranging from 4 to 15 with respect to gas-phase calculations.
 [DOI: 10.1063/1.1622653]

I. INTRODUCTION

The concept of reaction mechanism plays a major role in chemistry and represents a synthesis of our understanding of the way in which different topological changes in the bonding structure of a reactant or product are correlated as the reaction proceeds. Recent advances in ultrafast lasers,[1,2] x-ray,[3] and other spectroscopies as well as in computational chemistry[4,5] have made possible the study of most gas-phase and some condensed-phase reactions in molecular detail. However, most experimental investigations of complex reaction mechanisms taking place in liquid environments are still inferred from isotope and solvent (medium) effects on the reaction rate.[6,7] Consequently, the interpretation of the experimental results as well as the reaction mechanisms inferred are more controversial than those of gas-phase reactions.

Computer studies can be useful as a complement to experimental data in cases where experiments alone cannot provide a definitive picture of the mechanism of the chemical process. It is therefore desirable to develop systematic computational approaches to carefully examine the relation between isotope effects and reaction mechanism in condensed-phase systems. However, computational calculations of kinetic isotope effects in condensed-phase reactions can become expensive due to a number of difficulties. Some of the practical challenges involving the calculation of kinetic isotope effects and reaction rates in solution are associated with the fact that accurate descriptions of transformations in which chemical bonds are broken and formed require time-consuming *ab initio* electronic structure methods. Computer time limitations become particularly relevant when investigating ''rare events'' such as chemical reactions, especially when the reactions are accompanied by substantial differences in the structure of the solvent. To further complicate matters, quantum effects such as zero-point vibrations and tunneling effects are important in some chemical processes, such as proton transfer reactions. Another technical problem in the simulations of reactive systems is that the statistical resolution of calculations of the reaction rate depends on how many statistically independent configurations are obtained during the simulation: Simulations in which a large number of successive configurations have similar configurations of the reactive core or of the solvent molecules suffer from large uncertainties in the calculated reaction rates, precluding any definitive interpretation of the reaction mechanism.

It is therefore critical to develop methods which sample statistically independent configurations along the reaction path rapidly and correctly. In the case where the reaction path can be characterized by means of a small number of reaction coordinates, accurate, statistically well-resolved calculations of reaction rates can be performed by developing improved methods for computing reaction free energy profiles along these reaction coordinates. A number of techniques for computing free energy profiles along reaction coordinates have been proposed in the literature, including umbrella sampling,[8] thermodynamic integration in conjunc-

[a)]Author to whom correspondence should be addressed. Electronic mail: jmschofi@chem.utoronto.ca

tion with the blue-moon ensemble method,[9] projection methods,[10] variable transformation approaches,[11] and guiding potentials.[12,13] The use of molecular mechanics-based guiding potentials was proposed simultaneously and independently by Iftimie *et al.*[12] and Vondele *et al.*[13] and implemented in a Monte Carlo and a molecular dynamics framework, respectively. The basic idea of the method consists of using a ''fast'' *molecular mechanics* potential to guide a computationally intensive *ab initio* simulation.

The Monte Carlo version of the ''guiding'' approach in Ref. 12 was called the molecular mechanics based importance function method (MMBIF). It was demonstrated that the utilization of a reasonably accurate molecular mechanics potential as an importance function decreases the correlation of an *ab initio* Monte Carlo calculation by two orders of magnitude. The method was illustrated on a gas-phase formic acid-water system in which the activated processes involved breaking and forming hydrogen bonds,[12] and was successfully applied to calculate the kinetic isotope effects in a model gas-phase intramolecular proton transfer reaction.[14–16]

One of the major challenges in *ab initio* simulations of reactions in condensed phase environments is to thoroughly sample configurations of the system when changes in the solvent occur on long time scales. For instance, in molecular dynamics simulations of proton transfer reactions in which the collective behavior of the solvent can strongly influence the dynamics of the reaction, the sampling efficiency can be limited by long solvent dielectric relaxation time. In essence, an independent configuration of the system requires that the equations of motion be propagated for a time which is longer than the dielectric relaxation time. Even simple organic solvents such as tetrahydrofuran, the relevant solvent in this study, have dielectric relaxation times on the order of 4 ps,[17] so that independent solvent configurations are only obtained after several thousand elementary propagation steps. More structured solvents such as water, with a dielectric relaxation time of roughly 8.3 ps,[18] require even longer propagation for proper sampling of solvent configurations. The long time scale of structural rearrangements in solvents pose a serious challenge to *ab initio* calculations even when the solvent is modeled using molecular mechanics since each propagation step in the dynamics involves a time-consuming *ab initio* calculation. Ideally, successive configurations in a simulation involve drastically different solvent and solute configurations. This is only possible using an artificial dynamics to generate the sequence of configurations. One way to generate relatively uncorrelated successive configurations is to apply importance sampling ideas.

In this article, the molecular mechanics-based importance sampling method is adapted to calculate reaction rates of chemical processes in condensed phases where collective motions of the environment can influence the quantitative features of the chemical process and, in some cases, play a critical role in determining the mechanism of a reaction. The Monte Carlo procedure involves separating the task of sampling the configurations of the condensed phase system into two parts. The first part involves an efficient scheme of updating the solvent configuration while the second focuses on the relatively slow *ab initio* calculation of the reactive core. This approach allows extensive sampling of the molecular mechanical solvent without a significant increase in the overall computational work over a gas phase *ab initio* simulation. The method is applied to study the double-proton transfer reaction in an acetic acid-methanol complex solvated by tetrahydrofuran.

II. METHODOLOGY

A. Motivation for the model system

The computational study of proton transfer reactions can be used to understand the conditions for the validity of a well-known conjecture proposed in the physical organic chemistry literature, stating that the breakdown of the rule of geometric mean for kinetic isotope effects, which is a relation[19] involving ratios of kinetic isotope effects corresponding to different isotopic substitutions at the primary and secondary atoms, is a signature of tunneling for both primary and secondary atoms.[6,20] The consequences of applying the rule of geometric mean to interpret experimentally determined isotope effects are of far-reaching importance: The inferred relationship between the rule of geometric mean and reaction mechanism forms the basis for the proposal that *multiple intramolecular* proton transfer reactions are likely to proceed via a two-step mechanism, in contrast to *multiple intermolecular* proton transfers which are believed to proceed via a synchronous pathway.[20] The same relationship is at the heart of the recent suggestion that tunneling effects have played an important role in the design of the active sites of some proteins.[6,21]

Some of the most striking consequences of the rule of geometric mean[6,20] appear when studying multiple proton transfer reactions in condensed phases. The study of reactions involving the exchange of a pair of protons between two molecules may provide insight into the dynamics of certain types of enzymatic reactions in which several functional groups in the active center are properly aligned so that concerted catalysis can occur. This type of catalysis mechanism is called *bifunctional catalysis* and is the principal mechanism responsible for the several orders of magnitude increase in the reaction rate in several important biochemical transformations.[22]

The double-proton transfer reaction between acetic acid and methanol is one of the simplest examples of reactions involving an intermolecular exchange of protons between two molecules and therefore is a good candidate for computationally investigating general aspects of bifunctional catalysis. The chemical processes occurring in a solution of acetic acid-methanol in tetrahydrofuran (THF) have been studied experimentally by Gerritzen and Limbach.[23] The majority species in the system consist of complexes formed from either a single molecule of acetic acid or a single molecule of methanol hydrogen-bonded with a single molecule of solvent. The minority species in the system consist of linear and cyclic clusters of acetic acid hydrogen-bonded with methanol and solvated by tetrahydrofuran. The double-proton transfer reaction takes place along the hydrogen bonds of the cyclic clusters. When the concentrations of ace-

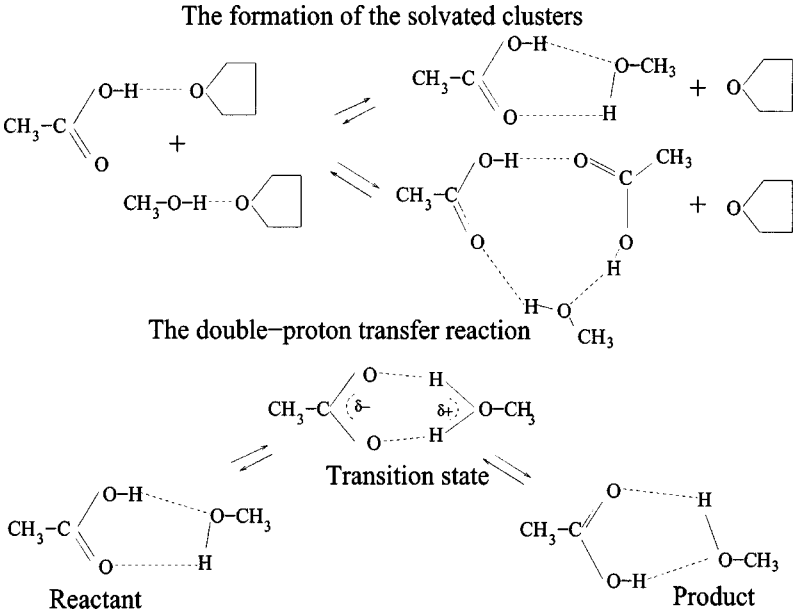

FIG. 1. The formation of two cyclic clusters involving one or two acetic acid molecules and one molecule of methanol. The double-proton transfer reaction studied here takes place along the hydrogen bonds of the cyclic cluster formed from one molecule of acetic acid and one molecule of methanol.

tic acid and methanol are reduced, the only cyclic cluster formed within which double-proton transfer is experimentally observed is formed from one molecule of acetic acid and a single molecule of methanol[23] (see Fig. 1).

In this work, we will focus only on the intermolecular double-proton transfer which takes place in the cyclic cluster formed from one molecule of methanol and a single molecule of acetic acid in a solution of tetrahydrofuran (THF). The quantum nuclear effects due to the small mass of the hydrogen atoms will be neglected here, and we will concentrate only on the sampling issues. Once the sampling method is tested for classical atoms, the method will be developed in a later article to include nuclear quantum effects via centroid transition state theory using sampling ideas similar to those used in our previous studies of malonaldehyde.[14,15]

B. The QM/MM/implicit solvent approach

Since it is impractical to treat large systems quantum mechanically, one is inevitably faced with the decision of how to combine *ab initio* electronic structure methods with more empirical approaches. One alternative to full *ab initio* calculations consists of using a mixed quantum-mechanical/molecular-mechanical (QM/MM) description of the system. However, in general it is problematic to clearly define the physics of the interface region between the quantum and molecular mechanical subsystems. In the best of circumstances, the separation can be made without ''cutting'' a covalent bond. For reactive processes in which the reaction occurs in a small solute which can be simulated in isolation using *ab initio* methods, the logical separation between the quantum and molecular mechanical regions is at the solute-solvent level. That is, the energies of the solute in which the reaction is occurring are calculated using electronic structure methods for the solute in the presence of the solvent, while the solvent energies are described by molecular mechanical potentials.

Even after this separation has been defined, one is faced with another technical issue of how an infinite condensed phase system can be represented in a practical fashion. One common approach utilized in simulations of condensed phase systems is to simulate explicitly a system consisting of a solute surrounded by a small number of solvent molecules periodically replicated in an infinite fashion. Unfortunately, such a periodic replication of the system introduces artifacts. One such artifact is the existence of spurious long-ranged correlations due to the periodicity of the system. In principle, such correlations could have a large impact on the reactive process. The use of periodic boundary conditions also introduces complexity in the electronic structure calculation itself when localized basis sets are used. An alternative approach is to use toroidal boundary conditions in which part of the solvent is explicitly represented while the influence of the solvent from regions far from the solute is implicitly incorporated.

In the present study we have utilized a quantum mechanics/molecular mechanics/implicit solvent (QM/MM/continuum) approach, in which the bond breaking and forming processes taking place in the solute are described by including all solute electron-electron, electron-nuclei, and nuclei-nuclei interactions in an electronic Hamiltonian. The interactions between the solvent molecules which form the first few solvation shells (in practice, 343 THF molecules) are described using a molecular mechanics potential. The electrostatic interactions between the solute and the solvent molecules in the first few solvation shells are incorporated into this approach by solving the *ab initio* electronic structure equations for the solute in the presence of the electric field generated by the molecular mechanics solvent charges. The total energy of the system also includes the Lennard-Jones interaction energy between solute and the solvent molecules in the first solvation shells, and the reaction field energy which accounts for the long-range electrostatic interactions between solute and solvent molecules (see Fig. 2).

The simulations were performed using the toroidal boundary conditions approach,[24] in which a cutoff distance is used that sets the boundaries of the region within which in-

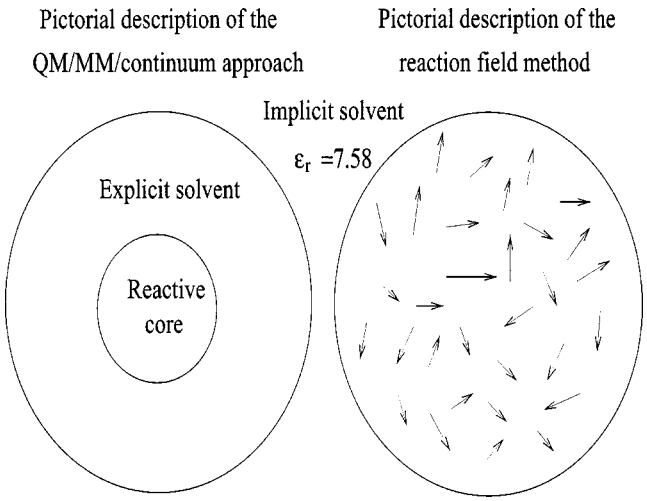

FIG. 2. A pictorial view of the QM/MM/continuum solvent method and of the reaction field approach. The ''reactive core'' region contains a quantum representation (i.e., nuclei+electrons) of the atoms which are involved in the actual covalent bond-breaking and bond-forming events. The ''explicit solvent'' region contains an atomic representation of the first few shells of solvent molecules. The effects of the solvent molecules which are far from the reaction center are included in an implicit manner in the QM/MM/continuum approach using the reaction field method. In our implementation, the reaction field method consists of calculating an effective electrostatic interaction between the dipole moments of the molecules inside the first two regions [see Eq. (1)].

termolecular interactions are explicitly counted. Therefore, the only quantum electron-electron, electron-nuclei, and nuclei-nuclei interactions which were explicitly counted in the present treatment were those which correspond to atoms separated by less than the cutoff distance, here chosen to be 14 angstroms.

The effects of the neglected interactions in the toroidal boundary conditions approach have been approximately accounted for by adding reaction field corrections[24] to the electronic energy:[25]

$$E_{RF} \approx -\sum_i \frac{\epsilon_r - 1}{(2\epsilon_r + 1)R_c^3}\mu_i M_i, \tag{1}$$

where the sum is over all molecules i inside the primitive cell, ϵ_r is the dielectric constant of the solvent, R_c is the radius of the spherical surface, μ_i is the dipole moment of molecule i, and M_i is the total dipole moment inside the spherical surface surrounding the molecule i.

The correct energetics in hydrogen-bonded systems and in systems undergoing proton transfer reactions is difficult to describe even with *ab initio* methods. In particular, DFT studies of weak hydrogen-bonding systems have proved to be particularly difficult and only limited success has been achieved in predicting the geometries and energies for reactant and transition state configurations on the potential energy surface using most exchange-correlation functionals.[26] Care should therefore be exercised when choosing a particular *ab initio* method to calculate proton transfer reaction rates. The nonlocal exchange-correlation schemes developed by Proynov, Vela, and Salahub[27] have shown particular promise for the description of hydrogen-bonded systems. Sirois *et al.*[28] have demonstrated that their kinetic-energy-dependent exchange functionals (BLAP and PLAP) performed better than all GGA options (BP86, PP86, PW91), BLYP, or other hybrid methods (B3LYP, B3PW91) on systems involving intramolecular hydrogen bonds. The predictions for equilibrium and transition state geometries as well as the energetics was in agreement with high-quality post-Hartree–Fock calculations [CCSD(T) and G2].[28] Specifically, for the gas-phase cyclic cluster of acetic acid and methanol in Fig. 1, the activation energy using the PLAP exchange correlation functional was found to be approximately 16.4 kcal/mol, in excellent agreement with the QCISD value of 16.14 kcal/mol.[29]

For the simulations described in the present work, the energies of different configurations were carried out using a modified version of the LCGTO-DFT program deMon-KS3.4[30,31] using the PLAP exchange-correlation functional. The DFT electronic structure calculations were carried out as in Ref. 28, where the application of DFT electronic structure methods to hydrogen-bonding systems is discussed in detail. A double-ζ plus polarization (DZVP) orbital basis set was used for all atoms and the convergence level for the SCF (self-consistent field) energy using the auxiliary fitting basis sets[28] was 0.01 kcal/mol.

C. The molecular mechanics potential describing the interaction between the solvent molecules

In order to implement the QM/MM/continuum approach to compute reaction rates for the double-proton transfer reaction in the acetic acid-methanol cluster solvated by tetrahydrofuran, a sufficiently accurate molecular mechanics description of the interaction between the solvent molecules is needed.[12,14] In this work, the OPLS all atom (OPLS-AA) force field of Jorgensen *et al.*[32] with a modified electrostatic interaction term has been used to describe the interactions between THF solvent molecules. The modifications to the electrostatics were designed to improve the gas-phase distribution of the partial charges in a THF molecule as well as to improve the description of polarization effects in condensed phases. Since the local electrostatic environment as well as long-ranged polarization effects can influence the proton transfer process, it is important to properly account for the permanent and induced charges in the solvent. To describe all such electrostatic phenomena, all solvent molecules have been assigned permanent and induced charges. It can be demonstrated[33] using second-order perturbation theory that the electrostatic interaction energy between two polarizable molecules A and B, each of which carries a set of atomic permanent and induced charges, Q_p^I and Q_{in}^I, with $I=1,\ldots,N$ corresponding to the charges on molecule A, and q_p^i and q_{in}^i, where $i=1,\ldots,n$ are the site charges on molecule B, respectively, can be written to a good approximation as

$$V(Q_p^1,\ldots,Q_p^N,Q_{\mathrm{in}}^1,\ldots,Q_{\mathrm{in}}^N,q_p^1,\ldots,q_p^N,q_{\mathrm{in}}^1,\ldots,q_{\mathrm{in}}^N) \approx \sum_{I=1}^{N}\sum_{i=1}^{n}\frac{Q_p^I q_p^i}{4\pi\epsilon_0 d_{iI}} + \frac{1}{2}\sum_{I=1}^{N}\sum_{i=1}^{n}\left(\frac{Q_p^I q_{\mathrm{in}}^i}{4\pi\epsilon_0 d_{iI}} + \frac{Q_{\mathrm{in}}^I q_p^i}{4\pi\epsilon_0 d_{iI}}\right), \tag{2}$$

where d_{iI} is the distance between sites I and i on molecules A and B. In principle, the induced charge appearing on a given solvent molecule is dependent on its local environment. One simple way of incorporating solvent polarization effects is to assign each solvent molecule the same average induced charge in a ''mean-field'' fashion. More sophisticated methods of including polarization effects either assign site polarizabilities or use fluctuating charges distributed at specific locations on each solvent molecule.[34] Such methods when combined with *ab initio* electronic structure methods either necessitate an iterative solution of the electronic structure and fluctuating charge distributions or involve dynamical methods in an extended Lagrangian system.[35] Unfortunately, each of these approaches has shortcomings which make them impractical to implement in conjunction with importance sampling Monte Carlo methods.

In principle, the iterative minimization of the Kohn–Sham and fluctuating charge functionals can be implemented within a Monte Carlo sampling approach at the cost of additional computational effort. However, provided the solvent is not very polarizable, the variation of the induced charges on the solvent molecules from their mean due to the presence of the solute is expected to be small. For this reason, we have utilized fixed induced charges for all solvent molecules. This corresponds to computing ground state energies for the reactive core $E(\rho_s(\mathbf{x}))$ based on a Kohn–Sham functional[36–38] $F[\rho_s(\mathbf{x})]$ which depends on the ground state electron distribution $\rho_s(\mathbf{x})$ of the solute in the presence of the fixed permanent and induced external solvent charges. Note that this functional includes the electrostatic energy of interaction between the quantum solute and the charges on the molecular-mechanical solvent molecules.

In general, one complication must be considered when discussing electrostatic interactions in mixed QM/MM systems that arises from the fact that the quantum mechanical electron density can become overpolarized by the molecular mechanical point charges due to the absence of considerations of the Fermi repulsion between quantum and molecular mechanical charges.[39] Such effects are particularly severe when using delocalized basis sets to represent the quantum subsystem, but are less significant when Gaussian or other local basis sets are utilized. In the present work, the full Coulomb interaction potential has been used to describe electrostatic interaction terms between the solute and the solvent charges without any screening modifications at short distances since all DFT calculations for the quantum subsystem use localized basis sets.

In our study of the acetic acid-methanol system in a THF solution, the permanent charges have been assigned the value $Q_p^I = q_p^I = 0.887 q_{\text{opls}}$, whereas the induced charges are set to a mean-field value of $Q_{\text{in}}^I = q_{\text{in}}^i = 0.239 q_{\text{opls}}$ for all indices i and I, where q_{opls} are the standard charges in the OPLS force field. It can be verified that the electrostatic interaction energy calculated using Eq. (2) with this set of permanent and average induced charges is precisely the electrostatic energy calculated using the standard set of fixed OPLS charges. On the other hand, by assigning a mean induced charge to each solvent molecule, the computed values for the gas-phase dipole moment and condensed phase dielectric constant are in better agreement with the corresponding experimental values (see Table I).[40]

TABLE I. The values of the gas-phase dipole moment μ_g expressed in Debyes, and of the static dielectric constant ϵ_r, obtained by considering the OPLS-AA charges as permanent charges, obtained from our modified version of the OPLS-AA force field (see text) which accounts approximately for electronic polarization effects, and from experimental data from Ref. 40.

	OPLS-AA	Modified OPLS-AA	Experimental
$\mu_g(D)$	1.97	1.76	1.75
ϵ_r	6.15 ± 0.3	7.61 ± 0.38	7.58

It should be emphasized that a less satisfactory means of describing electrostatic interactions between the solute and the solvent would consist of calculating the electrostatic interactions between the gas phase electron distribution with the solvent charges. In such a scenario, the ground state electron distribution for the solute interacts with the solvent via a Coulomb interaction of the form

$$V = -\frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} \int \frac{\rho_0(\mathbf{x}) q^i}{l_i(x)} d\mathbf{x}, \tag{3}$$

where $\rho_0(\mathbf{x})$ is the gas-phase electron distribution and $l_i(x)$ is defined to be the distance between the ith charge q^i in the solvent and the point x. This approximation corresponds to calculating the solute energy in the condensed phase by a zeroth-order approximation for the electronic distribution of the solute, that is,

$$F[\rho_s(\mathbf{x})] \approx F[\rho_0(\mathbf{x})] + V. \tag{4}$$

Such a description neglects the fact that the ground state solute electron distribution is influenced by the presence of the solvent charges. The influence of the solvent charges on the ground state energy can be accounted for by incorporating a polarization energy of the solute by the solvent.

Although such a crude level of description of the electrostatic interactions may be incomplete, it is useful in developing importance sampling Monte Carlo schemes based on guiding potentials, such as the molecular mechanics based importance function method (MMBIF) described in the next section.

D. The sampling methods

The MMBIF method[12] consists of utilizing an auxiliary Markov chain with a known asymptotic molecular mechanical distribution to propose trial configurations for an *ab initio* based Monte Carlo simulation. In the method, each trial configuration is obtained as the last state in a classical Markov chain generated from the current configuration in the *ab initio* simulation using various updating schemes. The proposed configurations are then accepted or rejected in the *ab initio* chain according to the usual Metropolis–Hastings algorithm.[12] If the previous and new trial configurations in the *ab initio* MC chain are denoted by $\mathbf{x}_{\text{old}}$ and $\mathbf{x}_{\text{new}}$, respectively, the proposed state is accepted with the probability $\min\{1, \exp(-\Delta\Delta E/k_B T)\}$, where $\Delta\Delta E$ is defined to be

$$\Delta\Delta E = (E^{\mathrm{DFT}}(\mathbf{x}_{\mathrm{new}}) - E^{\mathrm{cl}}(\mathbf{x}_{\mathrm{new}})) - (E^{\mathrm{DFT}}(\mathbf{x}_{\mathrm{old}}) - E^{\mathrm{cl}}(\mathbf{x}_{\mathrm{old}})), \tag{5}$$

where $E^{\mathrm{DFT}}(\mathbf{x})$ and $E^{\mathrm{cl}}(\mathbf{x})$ are the potential energies of configuration $\mathbf{x}$ calculated by *ab initio* methods (here density functional theory, abbreviated DFT) and the classical potential, respectively, and k_B is Boltzmann's constant. It is straightforward to show that this acceptance criterion guarantees that the *ab initio* Markov chain has the correct limiting Boltzmann distribution,[12] regardless of the number of classical updates used to generate the proposed configuration.

The efficiency of the MMBIF approach relies on constructing a molecular mechanics potential for the entire system which approximates the true interactions of the system at a qualitative level. At first glance, the construction of a molecular mechanics potential for a condensed phase system appears a daunting task given that the electron distribution of the solute changes considerably during the reactive process and is influenced in a complicated fashion by its local environment. However, it is relatively straightforward to construct a molecular mechanics potential based on simple approximate forms of the interaction potentials, such as that in Eq. (4). For such forms of the potential, the construction of the potential is reduced to modeling the reaction in gas phase and calculating effective charges on the gas-phase solute which mimic the correct electron distribution. The approximate form for the potential can be corrected for using importance sampling methods. For example, a Monte Carlo chain of states generated using an approximate expression for the energy can be manipulated by reweighting the configurations appearing in the chain by an appropriate factor.[14] The efficiency of this approach is highly dependent on the quality of the approximation for the true energy of the system. One might anticipate that the crude of level of description of the electrostatic interactions in Eq. (4) which neglects any polarization effects of the solute by the solvent molecules would introduce large statistical uncertainties at the reweighting step. However, the polarization of the solute by the solvent can be approximated by adjusting the charge of the solvent molecules in the expression for the interaction between the solute and the solvent. As will be discussed, the effective charge on the solvent molecules can be designed to approximate solute polarization effects and thereby improve the statistical resolution of the reweighting procedure.

The task of constructing a molecular mechanics potential for the gas-phase proton transfer reaction is facilitated by using bond evolution theory considerations. Following these lines, the molecular mechanics description of the acetic acid-methanol complex in the absence of the solvent was created as suggested in Ref. 14. The total molecular mechanics energy was decomposed into two components. The first component of the total energy was written as a sum of harmonic potentials representing the variation of the potential with bond length, bond angle or bond dihedral displacements from their minimum energy values at a fixed value of the control parameter *b*, defined by

$$b = d_{\mathrm{O_7H_4}} - d_{\mathrm{O_7H_6}}, \tag{6}$$

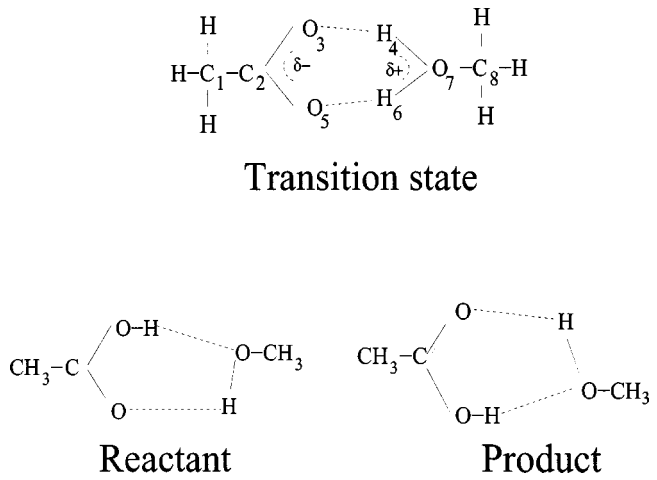

FIG. 3. The structures of the reactant, transition state, and product in the gas-phase double-proton transfer reaction.

where the numbering of the atoms is that from Fig. 3.

The second component of the total energy was written as a sum of four Morse potentials depending on the four O–H bond lengths, plus two effective potentials depending on two parameters, a_1 and a_2, defined as

$$a_1 = d_{\mathrm{O_3O_7}} \quad \text{and} \quad a_2 = d_{\mathrm{O_5O_7}}. \tag{7}$$

These effective potentials account for the flow of electronic charge during the reaction, as well as for the Fermi repulsion between the oxygen atoms O_3 and O_7, and between O_5 and O_7. This second component of the total energy was implemented using the same functional forms as in Ref. 14. As in the case of the malonaldehyde study, no potential which depends explicitly on the angles $O_3\widehat{H_4}O_7$ or $O_5\widehat{H_6}O_7$ was utilized, although some dependence on the angle $H_4\widehat{O_7}H_6$ was explicitly introduced using a functional form which interpolates between a harmonic potential for transition state values of the parameter *b*, and zero for values of *b* characteristic for the reactant or product configurations. The complete details for the construction of the guiding potential for the solute can be found in Ref. 33.

The simplest practical means of incorporating the electrostatic interactions between the solute and the solvent molecules in the molecular mechanics potential is to fit partial charges to atomic sites in the solute to reproduce the gas-phase electronic distribution. However, since the electron distribution of the solute varies appreciably with the configuration of the solute, the guiding potential must incorporate solute charges which vary as the reaction proceeds. In their bond evolution theory analysis of the tautomerization of malonaldehyde, Krokidis *et al.*[41] found that the total charge in the basins of attraction of the proton and oxygen atoms varies approximately linearly with a control parameter similar to the parameter *b*. Therefore, it can reasonably be assumed that most of the variation of the charges on atomic sites in the solute can be explained by a linear variation with *b*.

An alternative approach to incorporate the solute–solvent interactions can be constructed using simulated tempering methods.[42,43] The advantage of this procedure is that it does not rely on any approximation for the variation of the

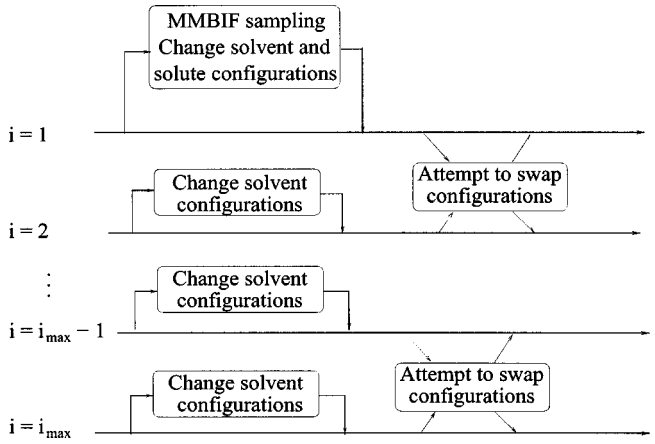

FIG. 4. A schematic representation of the simulated tempering method which uses the MMBIF approach to generate configurations of the system in a simulation where solute polarizability is neglected. The values of the electrostatic scaling parameter λ_i for electrostatic indices $i=1,...,i_{\max}$ are gradually increased from zero to one.

fitted charges on the solute during the reaction. The method essentially consists of using an extended state space to gradually turn on electrostatic interactions between the solute and the solvent. This approach is represented schematically in Fig. 4.

In the simulated tempering method, a Markov chain is constructed whose states $(i,\mathbf{r})$ are defined on the space formed by the direct product between a finite set of ''electrostatic'' indices, $i=1,2,...,i_{\max}$, and the entire solvent plus the solute configurational space. In vector notation, the states $\mathbf{r}$ will be denoted $\mathbf{r}=(\mathbf{r}^s,\mathbf{r}^S)$, where $\mathbf{r}^s$ and $\mathbf{r}^S$ represent the solute and solvent degrees of freedom, respectively. This Markov chain, whose generic state is denoted $(i,\mathbf{r})$, is constructed to produce states asymptotically distributed according to the probability density

$$p(i,\mathbf{r})=w_i p_i(\mathbf{r}), \tag{8}$$

where w_i are constants which will be referred to as the weights for the unnormalized probability densities $p_i(\mathbf{r})$ defined by

$$p_i(\mathbf{r})=\exp(-\beta E_i(\mathbf{r})), \tag{9}$$

where $\beta=1/k_B T$. The potentials $E_i(\mathbf{r})$ contain five components: the *ab initio* potential $E^s(\mathbf{r}^s)$ calculated for the gas-phase solute configuration, the molecular mechanics potential $E^S(\mathbf{r}^S)$ describing the interaction between the explicit solvent atoms, the Lennard-Jones potential $E_{\mathrm{LJ}}^{sS}(\mathbf{r}^s,\mathbf{r}^S)$ describing the dispersive and short-ranged interactions between solute and solvent atoms, the Coulomb electrostatic interaction $E_i^{sS}(\mathbf{r}^s,\mathbf{r}^S)$ between some charges on the solute atoms and the permanent and induced charges on the solvent atoms, and the reaction field energies $E_{\mathrm{RF}}(\mathbf{r}^s,\mathbf{r}^S)$ describing the long-range solute-solvent electrostatic interactions:

$$E_i(\mathbf{r}^s,\mathbf{r}^S)=E^s(\mathbf{r}^s)+E^S(\mathbf{r}^S)+E_{\mathrm{LJ}}^{sS}(\mathbf{r}^s,\mathbf{r}^S)+E_i^{sS}(\mathbf{r}^s,r^S)+E_{\mathrm{RF}}(\mathbf{r}^s,\mathbf{r}^S). \tag{10}$$

The Coulomb solute–solvent electrostatic interaction potential between charges q^s on the solute and q^S on the solvent, given by

$$E_i^{sS}(\mathbf{r}^s,\mathbf{r}^S)=\lambda_i\sum_{J=1}^{N}\sum_{j=1}^{n}\frac{q_J^S q_j^s}{4\pi\epsilon_0 d_{Jj}}, \tag{11}$$

where n and N are the number of charge sites on the solute and solvent molecules, respectively, and λ_i is a scaling factor henceforth called the charge fraction. Note that this interaction potential depends on the choice of charges assigned to the solvent and solute molecules. In our calculations, the charges on the solute were fitted using the Kollman–Singh procedure[44] from the gas-phase electron distribution of the solute. The use of charges fitted from the gas-phase calculation of the electron distribution does not account for the polarization of the solute by the solvent. To partially compensate for the neglect of this effect, the total charge on the solvent molecules used in Eq. (11) is set to the sum of the permanent and induced charges on the solvent, $q=q_p+q_{\mathrm{in}}$. This is in contrast to the calculation of the electrostatic interaction between solvent molecules in which the effective charges are set to $q=q_p+q_{\mathrm{in}}/2$. This approximation is consistent with first-order perturbation theory in which the polarization energy of the solute by the solvent is approximated by the polarization energy of the solvent by the solute.[33]

Note that the only difference between the potentials $E_i(\mathbf{r}^s,\mathbf{r}^S)$ for different electrostatic indices i consists of the form of the electrostatic interaction energy $E_i^{sS}(\mathbf{r}^s,\mathbf{r}^S)$ between the solute and the explicit solvent atoms. The energy $E_1^{sS}(\mathbf{r})$ is chosen to be zero in this work implying that $\lambda_1=0$, which amounts to turning off solute charges. For this system all electrostatic interactions between the solvent and the solute are scaled to zero although the solvent and solute still interact through Lennard-Jones potentials. By design, the last electrostatic index is set to unity, $\lambda_{i\,\max}=1$ so that the energy $E_{i\,\max}^{sS}(\mathbf{r})$ corresponds to calculating the electrostatic interaction between the solvent and solute atoms using the fitted configuration-dependent solute charges obtained via the Kollman–Singh method. The additional dimension of the solute plus solvent state space represented by the electrostatic index $i=1,...,i_{\max}$ ensures that the solvent configuration adapts smoothly to the solute via a stepwise process in which the charges of the solute atoms gradually interact with the other charges in the system. It should be noted that an equivalent result can be achieved by using a parallel tempering scheme in which the label $i=1,...,i_{\max}$ corresponds to a stepwise decrease of a ''sampling temperature'' associated with the solute–solvent electrostatic interaction.

In our implementation of the importance sampling, a Markov chain of extended state space configurations was generated by two types of transitions. In transitions of the first type the electrostatic index i was kept fixed while the configuration of the system $\mathbf{r}=(\mathbf{r}^s,\mathbf{r}^S)$ was updated using a transition matrix which leaves $p_i(\mathbf{r})$ invariant. The way in which the configurations were updated for fixed electrostatic index was dependent on the electrostatic scaling factor. When the scaling factor was zero both the configuration of the solute and the solvent were updated simultaneously using

the MMBIF approach. The background molecular mechanics simulations used to guide the updates were run so as to produce effectively independent but energetically reasonable configurations of the entire system. The simultaneous update of both solute and solvent configuration is possible in the absence of electrostatic interactions between the solute and solvent since the fitted charges are not used in the molecular mechanics auxiliary chain. However, after the calculation of the *ab initio* DFT energy in the acceptance–rejection step of the MMBIF method, the fitted charges for the given solute configuration can be calculated. When the electrostatic scaling factor is nonzero and solute–solvent electrostatic interactions occur, the solvent was allowed to adjust to the charge distribution on the solute by updating the solvent configuration while maintaining the configuration of the core reactive region unchanged.

Transitions of the second type move the system through the auxiliary parameter space by applying a transition matrix that changes the electrostatic index while leaving the configuration of the solute and solvent unchanged. In this study, the method for updating the electrostatic index i consisted of using a Metropolis algorithm, with a proposal distribution in which the proposed indices $i_{\text{new}}=i_{\text{old}}+1$ and $i_{\text{new}}=i_{\text{old}}-1$ are equally probable. The proposed change of electrostatic index is rejected if i_{new} is outside the valid range $i=1,...,i_{\max}$, otherwise it is accepted with probability

$$\min\left[1,\frac{p(i_{\text{new}},\mathbf{r})}{p(i_{\text{old}},\mathbf{r})}\right]=\min\left[1,\frac{w_{i_{\text{new}}}}{w_{i_{\text{old}}}}\exp\left(-\beta(E_{i_{\text{new}}}-E_{i_{\text{old}}})\right)\right]. \tag{12}$$

The marginal distribution of i with respect to the equilibrium distribution is given by

$$p(i)=\int p(i,\mathbf{r})\,d\mathbf{r}=\int w_i p_i(\mathbf{r})\,d\mathbf{r}=w_i Z_i, \tag{13}$$

where the configuration partition function Z_i is defined as

$$Z_i=\int \exp\left(-\beta E_i(\mathbf{r})\right)\,d\mathbf{r}. \tag{14}$$

If the distributions $p(i,\mathbf{r})$ are all to play a useful role in sampling, the weights w_i should be chosen such that a roughly uniform distribution over i is obtained. Since the Z_i are initially unknown, suitable values for the weights are found through a trial and error process using preliminary runs. To do this, an iterative procedure can be used in which the Markov chain is simulated using the current values for w_i, and the frequencies f_i with which each distribution is visited are recorded.

Next, new and improved weights $w_{i'}$ are calculated as $w_{i'}=w_i/f_i$ for electrostatic index i. If some of the frequencies f_i are zero, various elaborations of the estimation procedure can be used and some of them are summarized in Ref. 43. The number $i_{\max}$ of values of the electrostatic scaling parameter λ_i used and the actual values of w_i, $i=1,...,i_{\max}$ are chosen by minimizing the average computer time necessary for a new solute configuration to appear with an electrostatic scaling parameter $\lambda_{i\,\max}=1$. A good starting estimate for the numbers w_i and $i_{\max}$ can be obtained by optimizing the weights and the number of intermediate chains using only the molecular mechanics guiding potential with a reasonable set of atomic charges on the solute atoms.

III. RESULTS

To explore the issue of how importance sampling methods can be effectively utilized in simulations of reactions in condensed phases, two different implementations of the MMBIF sampling method were analyzed. In the first implementation, the dependence of the solute charges on the reactive state of the system was taken to vary linearly with the control parameter b defined in Eq. (6). For this simulation, the electrostatic interactions between the interpolated charges on the solute and the charges in the solvent were taken to be Coulombic. In the second simulation, the simulated tempering method described above was applied to the solvated acetic acid-methanol cluster. Both simulations generate chains of states asymptotically distributed according to a Boltzmann distribution based on Eq. (4) in which the solute–solvent electrostatic interactions are modeled by calculating the Coulomb interaction between the gas phase electron distribution with the charges in the solvent. In both simulations, the desired distribution for the chain of states based on the universal Kohn–Sham functional $F[\rho_s(x)]$ can be recovered at the end of the simulation by reweighting each of the N_T total configurations by a configuration-dependent factor

$$W(x_i)=\frac{e^{-\beta\Delta E_{\text{pol}}(x_i)}}{\sum_{i=1}^{N_T}e^{-\beta\Delta E_{\text{pol}}(x_i)}}, \tag{15}$$

where

$$\Delta E_{\text{pol}}(x_i)=F[\rho_s(x_i)]-(F[\rho_0(x_i)]+V)=F[\rho_s(x_i)]-E^s(r_i^s)-E^{sS}(r_i^s,r_i^S). \tag{16}$$

is the difference in the polarization energy of the solute by the solvent estimated by calculating the energy of the ground state electron distribution in the presence of the solvent charges and the energy of a gas-phase electron distribution interacting with the optimized solvent charges.

In all simulations, the calculations for the solvated cluster (shown in Fig. 3) have been carried out by treating all nuclei as classical point particles. The calculations were conducted in the isobaric-isothermal ensemble at $p=1$ atm and $T=298$ K. In order to improve the sampling along the reaction coordinate, an umbrella potential[8] was constructed for the guiding potential using a self-adaptive scheme. Simulations biased by the converged umbrella potential yielded a uniform sampling of the important regions of the reaction coordinate even though the activation barriers for the proton transfer reaction was on the order of $28\,k_{\text{B}}T$.

Two-thirds of the *ab initio* Markov chain transition steps in the simulation using the linearly interpolated charges on the core were generated using the MMBIF method. The same fraction of base transitions were generated using the MMBIF updates when the electrostatic index is zero in the simulated tempering simulation (see Fig. 4). The rest of the *ab initio* Markov chain transitions were performed utilizing Metropo-

lis single-variable updates using the *ab initio* DFT energy. As demonstrated in a previous study,[12] the role of the Metropolis DFT updates is to prevent the guiding molecular mechanics Markov chain of states from spending a large number of successive steps in those regions of the configuration space where the molecular mechanics density of states in the solute configuration space substantially underestimates the corresponding *ab initio* DFT density of states.

In the previous studies of the formic acid-water[12] and of the malonaldehyde[14] systems it was demonstrated that a useful strategy for optimizing the MMBIF method was to separate the variables to be updated in a classical MC step into several groups, with strongly correlated variables grouped together. In the case of the double-proton transfer in the cyclic cluster formed by acetic acid and methanol, the vibrations of the two methyl groups should be relatively uncoupled from the motions of the other atoms in the cluster. Applying this separation of which variables are updated in the MMBIF method, the percent of rejections of proposed configurations obtained with the MMBIF method was about 30%. The percent of rejections of the transitions which employed single-variable Metropolis updates using the *ab initio* potential was 45%.

It is important to compare the computational effort of performing liquid-phase versus gas-phase simulations of chemical reactions within the MMBIF approach. The potentials of mean force obtained using the reaction coordinate

$$\xi = b = d_{O_7H_4} - d_{O_7H_6}, \tag{17}$$

describing the double proton-transfer in the acetic acid-methanol system in tetrahydrofuran obtained in the first simulation and in a gas-phase simulation are represented in Fig. 5. Although equal cpu times were spent in computing the two potentials of mean force depicted in Fig. 5, the statistical uncertainties for the activation energy calculated for the double-proton reaction rate in the solvated system is approximately two times larger than the error bar for the activation energy calculated from the gas-phase simulation. This decrease in the statistical resolution of the computations in the solvated system with respect to the gas-phase system is due to a larger integrated correlation time of the overall Markov chain as well as to the fact that 50% of the cpu time in the simulation of the solvated system is dedicated to calculating the molecular mechanics interactions between the *solvent* molecules. Although a single *ab initio* calculation is roughly four orders of magnitude slower than a single update of the configuration of the molecular mechanical solvent, the long dielectric relaxation of the solvent required that many updates be carried out on the solvent before an independent solvent configuration was generated. This translated into an approximately equal amount of CPU time for the molecular mechanical and quantum mechanical components of the simulation.

The approximation of the variation of the charges fitted using the Kollman–Singh procedure with the solute configuration by a linear variation with the parameter b in Eq. (6) proved quite accurate. The standard deviation of the differences between the electrostatic interaction energies calculated using these fitted charges obtained via the Kollman–Singh procedure and the same energies calculated using solute charges which vary linearly with the parameter b defined in Eq. (6) is 0.6 kcal/mol. This small deviation should be contrasted with the value of 4 kcal/mol representing the standard deviation of the values of the electrostatic interaction energy between the solute and solvent molecules (see Fig. 6).

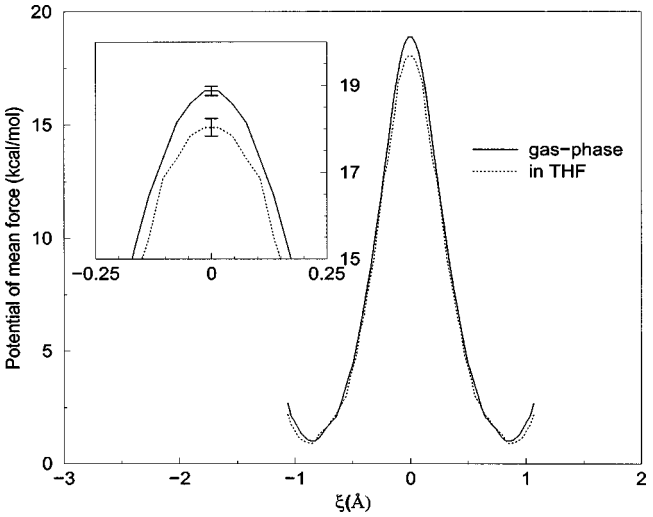

FIG. 5. The calculated potentials of mean force for the double-proton transfer reaction in the acetic acid-methanol cluster in gas-phase and in a solution of tetrahydrofuran using the reaction coordinate ξ defined in Eq. (17). Note that the difference between the activation energies is approximately 0.8 kcal/mol. This difference is larger than the width of the 75% confidence intervals for the activation energies, which have been estimated to be 0.2 kcal/mol for the gas-phase and 0.4 kcal/mol for the liquid-phase simulations (inset).

The distribution of the values of the polarization energy difference ΔE_{pol} defined in Eq. (16) relevant for the final reweighting of data points is plotted in Fig. 7. As expected, the values of ΔE_{pol} are distributed over a small range of energy values of approximately 0.4 kcal/mol. In fact, calculation of the activation energy in the double-proton reaction rate in tetrahydrofuran without reweighting yields an activation energy which differs from the activation energy in Fig. 5 only by a statistically irrelevant value of 0.1 kcal/mol. It should be emphasized that the small values of ΔE_{pol} in Fig. 7 are in part a consequence of approximating the polarization energy of the solute by the solvent by the polarization energy of the solvent by the solute. We estimated that if the solvent charges used in Eq. (11) were set to $q^S = q_p^S + p_{\mathrm{in}}^S/2$, the standard deviation for ΔE_{pol} would have been approximately 1.5 kcal/mol.

The potentials of mean force for the double-proton transfer reaction in tetrahydrofuran calculated using the linearly interpolated charge approach and using the simulated tempering method are identical within error bars. However, if the error bars for the activation energy in the linearly interpolated charge method is 0.4 kcal/mol, they are roughly four times larger in the simulation using the parallel tempering algorithm for comparable cpu times.

The optimal number of electrostatic indices in the simulated tempering approach was found to be $i_{\max}=4$. In the optimized setting for the simulated tempering method in

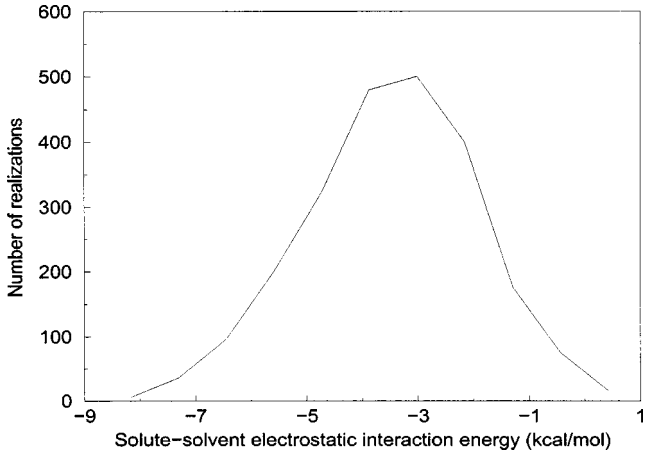

FIG. 6. The values of the electrostatic interaction energy between the solute and solvent molecules obtained for $\lambda=1$ in the simulated tempering method in which the solute charges are the gas-phase Kollman–Singh charges. Note that the values of the solute-solvent electrostatic interaction energies are scattered over an energy range of approximately 4 kcal/mol. In contrast, the differences in the electrostatic interaction energies calculated using the Kollman–Singh charges and the same energies calculated using solute charges which vary linearly with the parameter b defined in Eq. (6) are scattered over an interval of only 0.6 kcal/mol (data not shown).

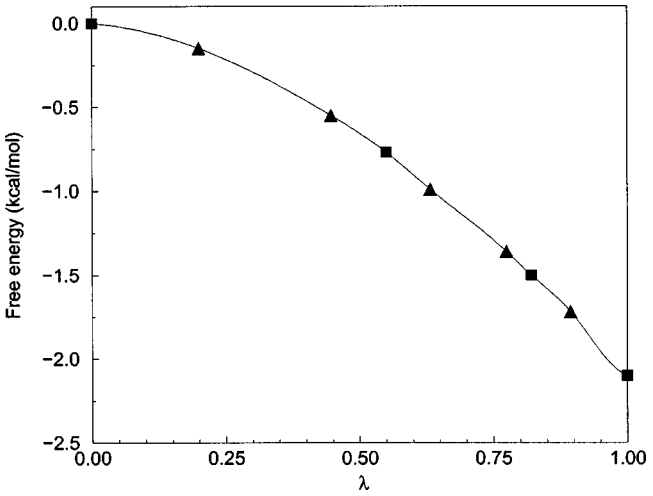

FIG. 8. The free electrostatic interaction energy between an acetic acid-methanol complex and molecules of tetrahydrofuran solvent as a function of the charge state of the complex. $\lambda=0$ and $\lambda=1$ correspond to fully uncharged and charged solute. The stars and diamonds represent the estimated values of the free energy obtained from the preliminary and from the optimized simulated tempering runs, respectively. The continous line represents a spline interpolation between the computed values.

which equal amounts of time were spent at all values of the electrostatic index, the Gibbs free energy differences $\Delta G_{i,i+1}$ computed for consecutive values of the index were found to be approximately 0.7 kcal/mol. The Gibbs free energy $G(\lambda)$ of the electrostatic interaction between the solute and solvent charges have been estimated using Eq. (14) and are plotted in Fig. 8 as a function of the fraction of solute charge λ_i being turned on.

Several comments are worth making about the results in Fig. 8. First, the partition functions Z_i for the umbrella potential-biased ensemble from which the free energies were computed correspond to a particular value of the electrostatic fraction parameter λ_i. It should also be noted that although a difference in free energy between adjacent values of the electrostatic index of the order of k_BT seems to suggest a low probability of rejecting swaps between consecutive values of the index, the actual rejection rate was found to be significantly higher. The relatively large rejection rate is due to the fact that differences in free energy reflect the *average* energetic *and* entropic differences between thermodynamic states with different λ, whereas the acceptance probability in Eq. (12) involves only the difference in the energies of the actual configurations to be swapped. In particular, it was observed that the enthalpic $H(\lambda)$ and entropic $-TS(\lambda)$ contributions to the Gibbs free energy $G(\lambda)$ vary in opposite directions as the fraction of the solute charge λ increases from zero to one. These variations of enthalpic and entropic terms with the charge fraction can be visualized by comparing the results plotted in Figs. 8 and 9. From the two figures it appears that $H(\lambda)$ increases and $-TS(\lambda)$ decreases with λ. The opposite directions in which enthalpic and entropic terms vary with λ is reflected in the mobility of the state $(i,\mathbf{r})$ of the simulated tempering algorithm in the i (or λ) subspace. Turning our attention to Eq. (12), note that an increase of $-TS(\lambda)$ with λ suggests that for a large fraction of configurations $(i,\mathbf{r})$ transitions in which i is *increased* are accepted only if the ratio

$$w_{\uparrow}(i)=\frac{w_{i+1}}{w_i} \tag{18}$$

is approximately unity, $w_{\uparrow}(i)\approx 1$. Such is the case for an important number of transitions from electrostatic index $i=1$ to electrostatic index $i=2$, for example, for which an increase in λ is accompanied only by a small decrease and occasionally even an increase in energy. On the other hand, for a λ near unity, a decrease of $H(\lambda)$ with λ suggests that a large fraction of transitions in which i is *decreased* are rejected unless $w_{\uparrow}(i)\ll 1$. Such is the case in a large number of transitions which are attempted from the final index $i_{\max}$ to $i_{\max}-1$ for example, which are accompanied by a significant increase in energy.

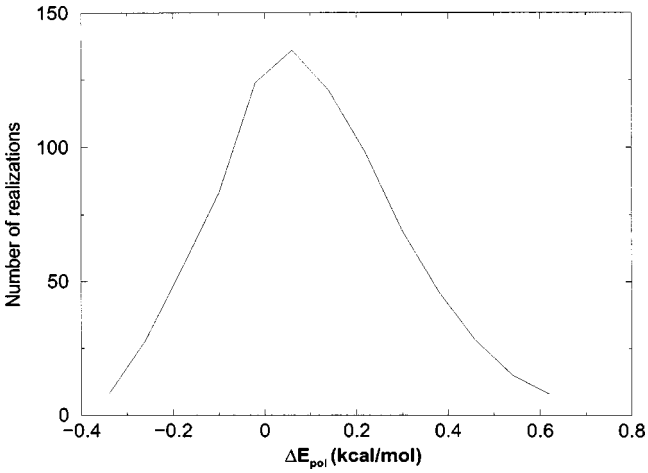

FIG. 7. The values of the polarization energy difference ΔE_{pol} defined in Eq. (16). Note that the distribution of values is centered roughly at 0 kcal/mol and has a small standard deviation of 0.4 kcal/mol.

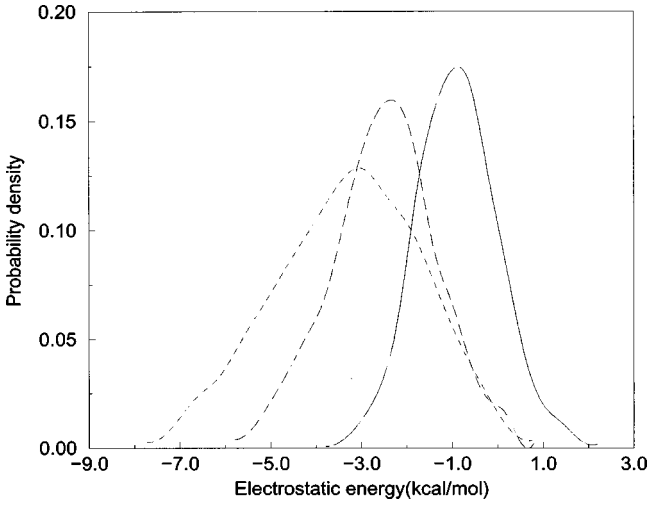

FIG. 9. The normalized probability density of the solvent-solute electrostatic interaction energy on for electrostatic indices $i=2$ (solid curve), $i=3$ (long dashed curve), and $i=4$ (dashed curve). Note that the enthalpies, calculated as the average solute-solvent electrostatic interaction energies, are approximately -1.0, -2.4, and -3.5 kcal/mol. Comparing these results with the corresponding free energies in Fig. 8 one obtains the corresponding entropies as being 0.3, 1.0, and 1.4 kcal/mol. The variations in the mechanical work pV with the number of chain are negligible.

The choice of the weights w_i suggested in Eq. (13) represents a good compromise in the sense that transitions which increase and which decrease i have an equal probability of being accepted on average. Nevertheless, this analysis points to the fact that if one uses the simulated tempering approach for the study of chemical reactions in solution, one should try to avoid turning on the charges of the reactive core all the way from zero to their final values. As enthalpic and entropic terms will always vary in opposite directions during the charging process, the sampling could become quite inefficient, especially when studying reactions in which there is a difference in the net charge between reactants and transition states, and not just in their dipole moments as in the present case.

However, this conclusion does not mean that the MMBIF approach used in the simulation with linearly interpolated charges will always be more efficient than the simulated tempering method. The efficiency of the MMBIF approach relies heavily on the appropriateness of the postulated variation of the charges in a reactive system with the reaction coordinate. In the present study, the approximation of the variation of the fitted Kollman–Singh charges with the solute configuration by a linear variation with the parameter b in Eq. (6) proved quite accurate. However, such a simple relation could break down, especially when studying reactions with a net transfer of charge. In this case, we suggest using an approach which combines the benefits of both methods illustrated in the present study: the use of a simulated tempering approach in which the solute charge is gradually modified from a linear variation with the reaction coordinate to their actual values obtained via the Kollman–Singh approach.

IV. DISCUSSION AND CONCLUSIONS

In this article, two important issues on the applicability of the molecular mechanics based importance sampling method to the study of reactive events in condensed phase environments have been addressed. One major concern in developing a successful implementation of the MMBIF approach is the ease of development of a sufficiently accurate molecular mechanics potential to guide the sampling. Although a fully automated approach to generating guiding potentials for general reactions is still not available, it is encouraging to note that the use of the same principles of bond-evolution theory[41] as in our earlier study of the malonaldehyde system[14] were adequate for designing a molecular mechanics potential for the acetic acid-methanol system. This success is particularly impressive in light of the fact that the malonaldehyde and acetic acid-methanol systems differ substantially not only in their barrier heights (by a factor of 4), but also in the qualitative nature of the chemical event (one proton versus two protons transferred). It suggests that bond-evolution theory guidelines are likely to be practical in developing molecular mechanics potentials for other proton transfer reactions, and for possibly other types of chemical events.

The study of the double-proton transfer reaction between acetic acid and methanol in tetrahydrofuran also demonstrates that the MMBIF method can be applied in reaction rate calculations of chemical transformations in solvents of medium polarity. In particular, an increase in the CPU time of factors of 4 and 15 with respect to gas-phase calculations were obtained using two different sampling methods. Hence, we conclude that it should be possible in many instances to compute solvent mediated reaction rates with statistical accuracies comparable to those obtained in gas-phase calculations, even though the complexity of the calculation is increased enormously by the presence of the solvent.

It should be emphasized that the contribution of the solvent to the total activation energy in the present study is only on the order of a few factors of k_BT at room temperature. As a result, the actual mechanism of the proton transfer event is virtually unchanged from the process in gas phase. This simplification allowed the separation of the reactive *ab initio* core and the solvent degrees of freedom into effectively disjoint sets which were updated in isolation in the parallel tempering implementation of the sampling. In many cases, the structure of the solvent plays a more substantial role in the reactive process. Under such circumstances, care must be taken to devise methods in which the reactive core and the solvent structure are updated in a more correlated fashion. For example, for a reaction in a solvent with a larger dielectric constant, the simulated tempering approach can be implemented by incorporating simultaneous MMBIF updates of the solute and solvent degrees of freedom at each electrostatic index.

It is informative to compare and contrast the advantages and disadvantages of using the MMBIF method versus using present day molecular dynamics based methods for calculating reaction rates in solution using a hybrid QM/MM approach. The practical use of the simple molecular dynamics methods for sampling configurations of a system containing

a reactive core which is described using *ab initio* electronic structure methods and the solvent molecules which are described using a molecular mechanics potential is limited by the fact that calculating the time-evolution of the system necessitates the computation of the time-consuming *ab initio* forces acting on the core atoms. These time-consuming calculations must be performed in traditional molecular dynamics calculations even when only the solvent degrees of freedom change significantly.

Several methods have been proposed for circumventing the inefficiency of the traditional QM/MM molecular dynamics calculations which are based on an ''artificial'' separation of time scales associated with the solute and solvent atoms.[45,46] The central idea of these methods is to use a large mass in conjunction with a high-temperature thermostat for the solute atoms, whereas the solvent atoms have usual masses and are in contact with a room temperature thermostat.[45,46] The large mass of the core atoms is chosen in such a way that the solvent degrees of freedom relax on a time scale which is much smaller than the time scale of the massive core atoms. Multiple-time scale arguments can be utilized to demonstrate that the integration of Hamilton's equations describing the evolution of the solute plus solvent system can be performed by computing the forces acting on the solvent degrees of freedom significantly more often than the *ab initio* forces acting on the solute without altering the asymptotic Boltzmann distribution of the configurations of the system. In addition, the decoupling between the solute and solvent time scales enables the use of a large temperature thermostat coupled to the solute atoms without introducing an irreversible heat flow, thereby ensuring that the average kinetic energy of the core atoms is comparable with the magnitude of the reaction barrier which separates reactant and product configurations on the potential energy surface. Therefore, a relatively small number of *ab initio* calculations must be performed before a reactive event occurs.

Both the MMBIF and the modified molecular dynamics approach succinctly described above have a number of advantages and shortcomings when studying chemical reactions in solution which are essentially driven by fluctuations in the structure of the *solute* using a QM/MM approach. The main disadvantage of the MMBIF method consists of the fact that a reasonable molecular mechanics description of the reactive event in the *gas-phase* solute is required, and this molecular mechanics potential must usually be created from scratch. Nevertheless, once such molecular mechanics potential has been constructed, the cpu time necessary for calculating reaction rates in solution is essentially independent of the characteristics of the solvent such as its dielectric relaxation time. In contrast, the molecular dynamics approach does not necessitate prior information with respect to the potential energy surface of the cluster, although if such information exists, it can be used to improve the efficiency of the sampling.[45,46] However, the average time needed to observe a reaction event increases as the square root of the effective mass of the reactive degree of freedom. On the other hand, given the requirement of the separation of time scales between the solvent and solute degrees of freedom, the characteristic time scale of the solute atom motion and therefore the lower bound for the value of the mass needed for the core atoms is *determined* by the duration of the solvent dielectric relaxation time. Therefore, it appears that the efficiency of the above-mentioned molecular dynamics scheme decreases with the increase in the solvent dielectric relaxation time.

The methodology proposed in the present work can be combined with the ideas presented in Ref. 14 to include nuclear quantum effects via centroid transition state theory with a supplementary increase in the cpu time by a factor of 2 times the number of path-integral beads. Such a combination of procedures provides a rigorous and practical platform for calculation of kinetic isotope effects. An important goal for future work is to clarify the mechanistic origin of the relation between the breakdown of the rule of geometric mean in multiple-proton transfer reactions and tunneling effects.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. R.I. would also like to thank the Ontario Ministry of Education for financial support.

[1] G. R. Fleming and P. G. Wolynes, Phys. Today **43**, 36 (1990).
[2] D. Zhong and A. H. Zewail, J. Phys. Chem. A **102**, 4031 (1998).
[3] B. Perman, V. Srajer, Z. Ren *et al.*, Science **279**, 1946 (1998).
[4] P. J. Rossky and J. D. Simon, Nature (London) **370**, 263 (1994).
[5] D. C. Clary, Science **279**, 1879 (1998).
[6] A. Kohen, R. Cannio, S. Bartolucci, and J. P. Klinman, Nature (London) **399**, 496 (1999); B. J. Bahnson, T. D. Colby, J. K. Chin, B. M. Goldstein, and J. P. Klinman, Proc. Natl. Acad. Sci. U.S.A. **94**, 12797 (1997); Y. Cha, C. J. Murray, and J. P. Klinman, Science **243**, 1325 (1989).
[7] M. Page and A. Williams, *Organic and Bio-organic Mechanisms* (Addison-Wesley Longman, Harlow, England, 1997).
[8] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[9] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).
[10] S. Melchionna, Phys. Rev. E **62**, 8762 (2000).
[11] Z. Zhu, M. E. Tuckerman, and G. J. Martyna, Phys. Rev. Lett. **88**, 100201 (2002).
[12] R. Iftimie, D. Salahub, D. Wei, and J. Schofield, J. Chem. Phys. **113**, 4852 (2000).
[13] J. Van de Vondele and U. Rothlisberger, J. Chem. Phys. **113**, 4863 (2000).
[14] R. Iftimie and J. Schofield, J. Chem. Phys. **114**, 6763 (2001).
[15] R. Iftimie and J. Schofield, J. Chem. Phys. **115**, 5891 (2001).
[16] R. Iftimie and J. Schofield, Int. J. Quantum Chem. **91**, 404 (2003).
[17] A. Chaudhari, P. Khirade, R. Singh, S. N. Helambe, N. K. Narain, and S. C. Mehrotra, J. Mol. Liq. **82**, 245 (1999).
[18] N. Nandi, S. Roy, and B. Bagchi, J. Chem. Phys. **102**, 1390 (1995).
[19] A. Kohen and J. H. Jensen, J. Am. Chem. Soc. **124**, 3858 (2002).
[20] H. H. Limbach, in *Electron and Proton Transfer in Chemistry and Biology*, edited by A. Müler, H. Ratajczak, W. Junge, and E. Diemann (Elsevier, New York, 1992).
[21] M. J. Knapp, K. Rickert, and J. Klinman, J. Am. Chem. Soc. **124**, 3865 (2002). For a simplified treatment of the problem of the rule of geometric mean see also Ref. 19.
[22] H. Dugas, *Bioorganic Chemistry, a Chemical Approach to Enzyme Action*, 3rd ed. (Springer-Verlag, New York, 1996). One of the textbook examples of bifunctional catalysis in solution chemistry is C. G. Swain and J. F. Brown, J. Am. Chem. Soc. **74**, 2534 (1952).
[23] D. Gerritzen and H. H. Limbach, J. Am. Chem. Soc. **106**, 869 (1984).
[24] M. Neumann, Mol. Phys. **50**, 841 (1983).
[25] R. O. Watts, Mol. Phys. **28**, 1069 (1974).
[26] V. Barone and C. Adamo, J. Chem. Phys. **105**, 11007 (1996).
[27] E. I. Proynov, A. Vela, and D. R. Salahub, Chem. Phys. Lett. **230**, 419 (1994); **234**, 462(E) (1995).

[28] S. Sirois, E. I. Proynov, D. T. Nguyen, and D. R. Salahub, J. Chem. Phys. **107**, 6770 (1997).
[29] A. Fernández-Ramos, Z. Smedarchina, and J. Rodríguez-Otero, J. Chem. Phys. **114**, 1567 (2001).
[30] E. I. Proynov, S. Sirois, and D. R. Salahub, Int. J. Quantum Chem. **64**, 427 (1997).
[31] For details on the deMon quantum chemistry package, see http://cvs.demon-software.com/public_html/
[32] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[33] R. Iftimie, thesis, University of Toronto, 2001.
[34] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).
[35] M. C. González Lebrero, D. E. Bikiel, M. Dolores Elola, D. A. Estrin, and A. E. Roitberg, J. Chem. Phys. **117**, 2718 (2002).
[36] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[37] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[38] M. Levy, Proc. Natl. Acad. Sci. U.S.A. **76**, 6062 (1979).
[39] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[40] C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry* (VCH, New York, 1988), Appendix A, Table A-1.
[41] X. Krokidis, V. Goncalves, A. Savin, and B. Silvi, J. Phys. Chem. A **102**, 5065 (1998); see in particular the discussion in page 5070 and Fig. 8; X. Krokidis, S. Nouri, and B. Silvi, *ibid.* **101**, 7277 (1997).
[42] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[43] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[44] U. C. Singh and P. A. Kollman, J. Comput. Chem. **5**, 129 (1984).
[45] L. Russo, P. Mináry, Z. Zhu, and M. E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002).
[46] J. V. Vondele and U. Rothlisberger, J. Phys. Chem. B **106**, 203 (2001).