

JCTC

Journal of Chemical Theory and Computation

Proton Transfer Studied Using a Combined Ab Initio Reactive Potential Energy Surface with Quantum Path Integral Methodology

Kim F. Wong,^{†,∇} Jason L. Sonnenberg,^{‡,○} Francesco Paesani,^{†,◆} Takeshi Yamamoto,[§]
 Jiří Vaníček,^{||,¶} Wei Zhang,[⊥] H. Bernhard Schlegel,^{*,‡} David A. Case,[⊥]
 Thomas E. Cheatham III,[#] William H. Miller,^{*,||} and Gregory A. Voth^{*,†,⊗}

Center for Biophysical Modeling and Simulations and Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, Department of Chemistry, Wayne State University, Detroit, Michigan 48202, Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan, Department of Chemistry and Kenneth S. Pitzer Center for Theoretical Chemistry, and Chemical Science Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, BioMaPS Institute and Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, and Departments of Medicinal Chemistry, Pharmaceuticals and Pharmaceutical Chemistry, and Bioengineering, 2000 East, 30 South, Skaggs Hall 201, University of Utah, Salt Lake City, Utah 84112

Received November 2, 2009

Abstract: The rates of intramolecular proton transfer are calculated on a full-dimensional reactive electronic potential energy surface that incorporates high-level ab initio calculations along the reaction path and by using classical transition state theory, path-integral quantum transition state theory, and the quantum instanton approach. The specific example problem studied is malonaldehyde. Estimates of the kinetic isotope effect using the latter two methods are found to be in reasonable agreement with each other. Improvements and extensions of this practical, yet chemically accurate framework for the calculations of quantized, reactive dynamics are also discussed.

1. Introduction

The breaking and formation of chemical bonds is fundamental to chemistry. While molecular dynamics (MD)

simulations of large-scale assemblies for hundreds of nanoseconds and even for a few microseconds are permissible using current terascale and emergent petascale high-performance computing infrastructures,^{1–3} the majority of these scientific applications involving conformational sampling or molecular association processes cannot model chemical reactions. Much of current force field research

* Corresponding authors. Phone: 313-577-2562 (H.B.S.), 510-642-0653 (W.H.M.), 773-702-9092 (G.A.V.). Fax: 313-577-8822 (H.B.S.), 510-642-6262 (W.H.M.), 773-795-9106 (G.A.V.). E-mail: hbs@chem.wayne.edu (H.B.S.), millerwh@berkeley.edu (W.H.M.), gavoth@uchicago.edu (G.A.V.).

[†] Center for Biophysical Modeling and Simulations and Department of Chemistry, University of Utah.

[‡] Wayne State University.

[§] Kyoto University.

^{||} University of California, Berkeley.

[⊥] Rutgers University.

[#] Departments of Medicinal Chemistry, Pharmaceuticals and Pharmaceutical Chemistry, and Bioengineering, University of Utah.

[∇] Current Address: Center for Simulation and Modeling, 205 Bellefield Hall, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

[○] Current Address: Department of Chemistry, Stevenson University, 1525 Greenspring Valley Road, Stevenson, Maryland, 21153.

[◆] Current Address: Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093.

[¶] Current Address: Ecole polytechnique fédérale de Lausanne, Institut des sciences et ingénierie chimiques, EPFL SB ISIC LCPT, BCH 3110 (Bât. BCH), CH-1015 Lausanne, Switzerland.

[⊗] Current Address: Department of Chemistry, The University of Chicago, 5735 S. Ellis Avenue, Chicago, Illinois 60637.

focuses on the development of accurate and reliable interactions for describing protein structures, solvation energies, and hybrid bioinorganic material properties.^{4–10} Designing a chemically reactive potential energy surface (PES), nevertheless, is not a new idea.¹¹ During the early half of the twentieth century, chemical physics pioneers, such as Eyring¹² and Evans and Polanyi,^{13–15} proposed procedures for forming surfaces to describe diverse systems involving ionic, S_N2 and Diels–Alder reactions. The common element among these methodologies is the description of the system within a matrix representation for the Hamiltonian:

$$\hat{\mathbf{H}} = \begin{bmatrix} H_{11} & \cdots & H_{1N} \\ \vdots & \ddots & \vdots \\ H_{N1} & \cdots & H_{NN} \end{bmatrix} \quad (1)$$

$$H\Psi = \varepsilon\Psi \quad (2)$$

Here, the system evolves on the lowest energy eigenstate, ε_0 , of the Hamiltonian as a superposition state of multiple nuclear-electronic configurations.¹⁶ Variations of this technique differ primarily in the treatments of the H_{ii} and H_{ij} terms. Notably, Warshel and Weiss utilized a combination of molecular dynamics (MD) force fields and empirical fitting, an empirical valence bond (EVB) approach, to approximate the matrix elements for describing reactions in solutions and within enzyme environments.^{17,18}

Although the empirical approach is a computationally practical strategy for modeling chemistry, a general framework for constructing reactive potential energy surfaces based on *first principles*, ab initio, information is desirable. Empirical procedures typically involve a high startup cost in the form of parameter fitting and calibrations that in some instances may be system-specific. On the other hand, even with current advances in computation infrastructure (both from an algorithms and from a hardware perspective), direct ab initio MD approaches are limited from small to moderate sized systems for a few hundred picoseconds of conformational sampling.^{19–24} These time scales cannot capture the physics of many biologically relevant reactions, typically occurring in the microseconds or on longer time scales. Furthermore, the computation time required for adequately sampling phase space to obtain converged thermodynamic observables precludes the use of direct ab initio MD approaches for complex systems consisting of 10 000 atoms or more. A general hybrid simulation framework combining the computational advantages of MD simulation with the accuracy of electronic structure methods is appealing.

In the present paper, we describe an implementation of such a general framework, the distributed Gaussian (DG) approach,^{25–27} for developing accurate ab initio-based potential energy surfaces capable of modeling chemical reactions. The purpose of this paper is to describe the integration, within the Amber²⁸ biosimulations suite, of the DG-EVB surface generation methodology with molecular dynamics, path integrals,^{29–31} and quantum instanton^{32–35} approaches for the computation of thermal rate constants. This methodology development is tested on a well-established system, the prototype intramolecular proton transfer reaction in malonaldehyde coupled to a thermal bath. Despite the

presumed simplicity of malonaldehyde, it is only recently that full-dimensional quantum calculations of tunneling splitting based on accurate ab initio surfaces were possible.^{36–40} Previous works utilized reduced-dimensional approximations or less accurate potentials that may not fully capture the physics of the reaction.^{41–47} Malonaldehyde thus provides a well-studied yet challenging test for assessing the diverse components of the DG-EVB methodology. Integrating both molecular mechanical and ab initio elements, DG-EVB is similar in strategy to Truhlar's recently developed multi-configuration molecular mechanics (MCMM) method.^{48–52}

The remainder of this paper is organized as follows: To show the relationship between the various methods, the theoretical backgrounds for the DG method, the path integral approach and the QI method are briefly reviewed. The simulation details are described in section 3 followed by the Results and Discussion in section 4. Last, we conclude with an assessment of the method and suggestions for future improvements.

2. Theoretical Background

This section briefly describes the theoretical basis to our algorithmic development for constructing a reactive PES from ab initio information and the incorporation of nuclear quantum effects in molecular dynamics. Dynamical methods for computing thermal rate constants and kinetic isotope effects are described. Our test system of a single malonaldehyde molecule coupled to a thermal reservoir is modeled within the canonical ensemble. Within the canonical ensemble, thermal rate constants are well-defined quantities.⁵³ Under favorable circumstances, tunneling splittings can also be extracted from the imaginary time correlation function from PIMD simulations, but in the present case reliable error bounds on the tunneling splitting could not be obtained using the maximum entropy approach.

2.1. Designing ab Initio Based Chemically Reactive PESs. A number of methods for building an ab initio derived PES using eq 1 has been proposed recently. Among these are Chang–Miller,^{54,55} Minichino–Voth,⁵⁶ multiconfiguration molecular mechanics (MCMM),^{48–52} and our distributed Gaussian EVB (DG-EVB).^{25–27} Since the distributed Gaussian approach is related to the Chang–Miller prescription, we concisely describe Chang–Miller below to motivate the subsequent method development. The Chang–Miller approach attempts to construct an accurate reactive potential energy surface by fitting a superposition of reactant and product configurations using a generalized Gaussian form for the coupling term

$$H_{12}^2(\mathbf{q}) = A \exp[\mathbf{B}^T \cdot \Delta\mathbf{q} - \frac{1}{2}\Delta\mathbf{q}^T \cdot \tilde{\mathbf{C}} \cdot \Delta\mathbf{q}], \Delta\mathbf{q} = \mathbf{q} - \mathbf{q}_{\text{TS}} \quad (3)$$

The diagonal elements describing the reactant state (RS) and product state (PS) are approximated using conventional, nonreactive force fields (FF). This choice was motivated by the expectation that the development of classical force fields will continually improve toward a state where they are accurate *enough*. Formulated to reproduce the ab initio

energy ε_Ψ , gradient, and Hessian at the transition state geometry, \mathbf{q}_{TS} , the parameters A (a scalar), \mathbf{B} (a vector), and $\tilde{\mathbf{C}}$ (a matrix) take the analytical forms

$$A = [H_{11}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})][H_{22}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})] \quad (4)$$

$$\mathbf{B} = \frac{\mathbf{G}_1}{[H_{11}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} + \frac{\mathbf{G}_2}{[H_{22}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} \quad (5)$$

$$\mathbf{G}_N = \left. \frac{\partial H_{NN}(\mathbf{q})}{\partial \mathbf{q}} \right|_{\mathbf{q}=\mathbf{q}_{\text{TS}}} - \left. \frac{\partial \varepsilon_\Psi(\mathbf{q})}{\partial \mathbf{q}} \right|_{\mathbf{q}=\mathbf{q}_{\text{TS}}}$$

$$\tilde{\mathbf{C}} = \frac{\mathbf{G}_1 \mathbf{G}_1^T}{[H_{11}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} + \frac{\mathbf{G}_2 \mathbf{G}_2^T}{[H_{22}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} - \frac{\tilde{\mathbf{K}}_1}{[H_{11}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} - \frac{\tilde{\mathbf{K}}_2}{[H_{22}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} \quad (6)$$

$$\tilde{\mathbf{K}}_N = \left. \frac{\partial^2 H_{NN}(\mathbf{q})}{\partial \mathbf{q}^2} \right|_{\mathbf{q}=\mathbf{q}_{\text{TS}}} - \left. \frac{\partial^2 \varepsilon_\Psi(\mathbf{q})}{\partial \mathbf{q}^2} \right|_{\mathbf{q}=\mathbf{q}_{\text{TS}}}$$

When the system configuration deviates far from the transition state structure (i.e., for large $\Delta \mathbf{q}$) and the matrix $\tilde{\mathbf{C}}$ contains one or more negative frequencies, H_{12}^2 diverges. Although refinements are available for controlling the asymptotic behavior of the Chang–Miller approach,⁵⁵ simply recasting eq 3 in terms of a quadratic polynomial times a spherical Gaussian

$$H_{12}^2(\mathbf{q}) = A[1 + \mathbf{B}^T \cdot \Delta \mathbf{q} + \frac{1}{2} \Delta \mathbf{q}^T \cdot (\tilde{\mathbf{C}}' + \alpha \tilde{\mathbf{I}}) \cdot \Delta \mathbf{q}] \times \exp[-\frac{1}{2} \alpha |\Delta \mathbf{q}|^2] \quad (7)$$

keeps the coupling element bounded at the asymptotes.²⁵ The scalar A and vector \mathbf{B} parameters are identical to those in the Chang–Miller approach, while the matrix assumes a modified form

$$\tilde{\mathbf{C}}' = \frac{\mathbf{G}_1 \mathbf{G}_1^T + \mathbf{G}_2 \mathbf{G}_2^T}{A} + \frac{\tilde{\mathbf{K}}_1}{[H_{11}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} + \frac{\tilde{\mathbf{K}}_2}{[H_{22}(\mathbf{q}_{\text{TS}}) - \varepsilon_\Psi(\mathbf{q}_{\text{TS}})]} \quad (8)$$

Note that $\tilde{\mathbf{I}}$ is the identity matrix and α is a parameter related to the Gaussian width. The distributed Gaussian approach generalizes the above polynomial times a Gaussian prescription [eq 7] to utilize ab initio information not only at the transition state geometry but also at other points on the potential energy surface. Here, $H_{12}^2(\mathbf{q})$ is approximated as an expansion in a set of *distributed Gaussians* centered on a set of molecular configurations \mathbf{q}_K :

$$H_{12}^2(\mathbf{q}) = \sum_K^{N_{\text{cfg}}} \sum_{i \geq j \geq 0}^{N_{\text{dim}}} B_{ijk} g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K) \quad (9)$$

$$g(\mathbf{q}, \mathbf{q}_K, 0, 0, \alpha_K) = (1 + \frac{1}{2} \alpha_K |\Delta \mathbf{q}_K|^2) \exp[-\frac{1}{2} \alpha_K |\Delta \mathbf{q}_K|^2] \quad (10)$$

$$g(\mathbf{q}, \mathbf{q}_K, i, 0, \alpha_K) = (\Delta \mathbf{q}_K)_i \exp[-\frac{1}{2} \alpha_K |\Delta \mathbf{q}_K|^2] \quad (11)$$

$$g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K) = (1 - \frac{1}{2} \delta_{ij}) (\Delta \mathbf{q}_K)_i (\Delta \mathbf{q}_K)_j \times \exp[-\frac{1}{2} \alpha_K |\Delta \mathbf{q}_K|^2] \quad (12)$$

$$\Delta \mathbf{q}_K = \mathbf{q} - \mathbf{q}_K \quad (13)$$

where N_{cfg} is the number of ab initio data points used for the fitting; N_{dim} is the number of system coordinates; $g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K)$ are the s-, p- and d-type Gaussians; and B_{ijk} are the expansion coefficients. The term involving the identity matrix in eq 7 was accumulated into the s-type Gaussian [see eq 10] to precondition the system of linear equations for efficient convergence when utilizing iterative methods. The nonstandard form of the d-type Gaussian is for similar reasons. If the number of Gaussian centers, K , is equal to the number of data points where $H_{12}^2(\mathbf{q})$ is evaluated, eq 9 describes a system of linear equations

$$\mathbf{F} = \tilde{\mathbf{D}} \mathbf{B} \quad (14)$$

that can be solved using singular value decomposition or by an iterative procedure, such as GMRES (generalized minimal residual method).^{57–60} When derivatives for the coupling terms are available, this information can also be utilized for the fitting, i.e.,

$$\left. \frac{\partial H_{12}^2(\mathbf{q})}{\partial \mathbf{q}} \right|_{\mathbf{q}=\mathbf{q}_L} = \sum_K^{N_{\text{cfg}}} \sum_{i \geq j \geq 0}^{N_{\text{dim}}} B_{ijk} \left. \frac{\partial g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K)}{\partial \mathbf{q}} \right|_{\mathbf{q}=\mathbf{q}_L} \quad (15)$$

$$\left. \frac{\partial^2 H_{12}^2(\mathbf{q})}{\partial \mathbf{q}^2} \right|_{\mathbf{q}=\mathbf{q}_L} = \sum_K^{N_{\text{cfg}}} \sum_{i \geq j \geq 0}^{N_{\text{dim}}} B_{ijk} \left. \frac{\partial^2 g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K)}{\partial \mathbf{q}^2} \right|_{\mathbf{q}=\mathbf{q}_L} \quad (16)$$

The \mathbf{F} column vector stores the terms $H_{12}^2(\mathbf{q}_L)$, $\partial H_{12}^2(\mathbf{q})/\partial \mathbf{q}|_{\mathbf{q}=\mathbf{q}_L}$, and $\partial^2 H_{12}^2(\mathbf{q})/\partial \mathbf{q}^2|_{\mathbf{q}=\mathbf{q}_L}$ evaluated at the N_{cfg} ab initio configurations. The $\tilde{\mathbf{D}}$ matrix contains the values of the Gaussian bases, $g(\mathbf{q}_L, \mathbf{q}_K, i, j, \alpha_K)$, $\partial g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K)/\partial \mathbf{q}|_{\mathbf{q}=\mathbf{q}_L}$, and $\partial^2 g(\mathbf{q}, \mathbf{q}_K, i, j, \alpha_K)/\partial \mathbf{q}^2|_{\mathbf{q}=\mathbf{q}_L}$, evaluated at these same configurations \mathbf{q}_L . The \mathbf{B} column vector contains the set of unknown expansion coefficients being solved for within this linear system of equations. Once this solution vector is obtained, we have a general (\mathbf{q} -dependent) analytical approximation to the coupling [eq 9] and corresponding derivatives [eqs 15 and 16], computed as a linear combination of the distributed Gaussian basis set.

For a symmetric 2×2 Hamiltonian matrix, the analytical expression for the coupling and derivatives are given by

$$H_{12}^2(\mathbf{q}) = [H_{11}(\mathbf{q}) - \varepsilon_0(\mathbf{q})][H_{22}(\mathbf{q}) - \varepsilon_0(\mathbf{q})] \quad (17)$$

$$\frac{\partial H_{12}^2(\mathbf{q})}{\partial \mathbf{q}} = \left[\frac{\partial H_{11}(\mathbf{q})}{\partial \mathbf{q}} - \frac{\partial \varepsilon_0(\mathbf{q})}{\partial \mathbf{q}} \right] \left[\frac{\partial H_{22}(\mathbf{q})}{\partial \mathbf{q}} - \frac{\partial \varepsilon_0(\mathbf{q})}{\partial \mathbf{q}} \right] \quad (18)$$

$$\frac{\partial^2 H_{12}^2(\mathbf{q})}{\partial \mathbf{q}^2} = \left[\frac{\partial^2 H_{11}(\mathbf{q})}{\partial \mathbf{q}^2} - \frac{\partial^2 \varepsilon_0(\mathbf{q})}{\partial \mathbf{q}^2} \right] \left[\frac{\partial^2 H_{22}(\mathbf{q})}{\partial \mathbf{q}^2} - \frac{\partial^2 \varepsilon_0(\mathbf{q})}{\partial \mathbf{q}^2} \right] \quad (19)$$

where ε_0 is the lowest eigenvalue of the matrix and H_{11} and H_{22} are the reactant and product valence bond states. H_{11} and H_{22} can be described by a force field potential (as in the spirit of Chang–Miller⁵⁴) or as a Taylor series expansion about the respective ab initio minimum.²⁵ The key idea is to determine $H_{12}^2(\mathbf{q})$ such that the resulting ε_0 surface approximates the ab initio surface, i.e., $\varepsilon_0 = \varepsilon_\Psi$. For the above two-state system, the coupling and corresponding derivatives can be evaluated directly from eqs 17–19 using the ab initio energies (ε_Ψ), gradients ($\partial \varepsilon_\Psi / \partial \mathbf{q}$), and Hessian ($\partial^2 \varepsilon_\Psi / \partial \mathbf{q}^2$) data in conjunction with the corresponding H_{ii} values and derivatives. This procedure provides coupling values and derivatives for the N_{cfg} discrete ab initio geometries. The DG-EVB method provides a prescription for evaluating the coupling at all coordinates as a linear combination of Gaussians expanded about these ab initio configurations, which may be chosen to be distributed along the IRC (intrinsic reaction coordinate), for example. While an analytical expression of the H_{ij} term for a general multistate $N \times N$ system Hamiltonian does not exist, one can make, nevertheless, the pairwise approximation and estimate the couplings using the two-state expression above for unique ij pairs. In this paper, we focus only on the two-state problem.

2.2. Incorporating Nuclear Quantum Effects. When the reactive process involves particles whose thermal de Broglie wavelength is comparable to the characteristic scale of the potential, nuclear quantum effects such as tunneling and zero-point motion are important for describing the chemical rate.^{61–63} Feynman’s path integral (PI)^{29–31} formalism of quantum mechanics provides a general framework for describing equilibrium (i.e., thermodynamic and structural) properties of a quantum many-body system. Briefly, the quantum partition function for the canonical (NVT) ensemble can be written as

$$Q = \left(\frac{P}{2\pi\hbar^2\beta} \right)^{3NP/2} \prod_{i=1}^N m_i^{3P/2} \int d\mathbf{q}^{(1)} \dots d\mathbf{q}^{(P)} \exp[-\beta\Phi] \quad (20)$$

where

$$\Phi(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)}) = \frac{P}{2\hbar^2\beta^2} \sum_{i=1}^N m_i \sum_{s=1}^P (\mathbf{q}_i^{(s)} - \mathbf{q}_i^{(s+1)})^2 + \frac{1}{P} \sum_{s=1}^P V(\mathbf{q}^{(s)}) \quad (21)$$

In this form, Q is isomorphic to the classical configurational partition function of N ring polymers, each comprised of P particles.⁶⁴ Each particle within the polymer interacts harmonically with neighbors along the cyclic path (first term) and with other particles within the same s -indexed PI slice

via $V(\mathbf{q}^{(s)})/P$ (second term). By introducing a set of Gaussian integrals into eq 20, one sees that the quantum partition function

$$Q = \Lambda \left(\frac{P}{2\pi\hbar^2\beta} \right)^{3NP/2} \prod_{i=1}^N m_i^{3P/2} \int d\mathbf{p}^{(1)} \dots d\mathbf{p}^{(P)} \int d\mathbf{q}^{(1)} \dots d\mathbf{q}^{(P)} \times \exp \left[-\beta \left(\sum_{i=1}^N \frac{p_i^{(s)}}{2\mu_i^{(s)}} + \Phi \right) \right] \quad (22)$$

can be evaluated from molecular dynamics based on the Newtonian equations of motion derived from a fictitious classical Hamiltonian of the form

$$H(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(P)}, \mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)}) = \sum_{i=1}^N \sum_{s=1}^P \left(\frac{p_i^{(s)}}{2\mu_i^{(s)}} + \Phi(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)}) \right) \quad (23)$$

This is possible because the inserted fictitious momenta are uncoupled and thus can be integrated analytically such that the prefactor Λ can be defined to give back the correct quantum partition function in eq 20. The purpose of the fictitious Hamiltonian is simply to provide a computationally practical scheme for evaluating the total phase space integral. Finally, to recover Q in the NVT ensemble, we also need to couple the system to a thermostat so as to ensure that the sampled distribution is indeed canonical.⁶⁵

While the path integral formulation is general for any internuclear potential, the results will depend on the accuracy of the electronic surface. In our study, the DG-EVB potential enters into PIMD through the second term of eq 21. The PI results presented here are computed from the normal mode representation of PIMD,⁶⁶ where the thermostat is a set of Nosé–Hoover chains.⁶⁷

2.3. Computing Observables from Molecular Dynamics Trajectories. The computable quantities connecting theory and simulation to experimental measurements are traditionally referred to as *observables*. For a reactive process, the *naturally* measurable quantity is the chemical rate. Transition state theory (TST)^{53,68–70} provides a simple framework for calculating the rate from molecular dynamics trajectories. In TST, the rate along a *reaction path* $\xi(\mathbf{q})$ [synonymous with the *reaction coordinate* (RC)] is given by

$$k_{\text{TST}} = \frac{1}{2} \langle |d\xi/dt| \rangle_{\xi^{\ddagger}} \langle \rho(\xi^{\ddagger}) \rangle \quad (24)$$

where $d\xi/dt$ is the *dynamical frequency factor* and ρ is the normalized density of sampled RC values

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{q} \exp[-\beta H(\mathbf{q})] \delta[\xi(\mathbf{q}) - \xi]}{\int d\mathbf{q} \exp[-\beta H(\mathbf{q})] h[\xi^{\ddagger} - \xi(\mathbf{q})]} \quad (25)$$

We note that the *intrinsic reaction coordinate* (minimum-energy reaction path) is one well-defined prescription for $\xi(\mathbf{q})$, although any RC (subject to the limitations of TST) may be employed. In the above equation, $\beta = 1/k_{\text{B}}T$, h is the Heaviside step function, ξ^{\ddagger} is the location of the *dividing surface* partitioning the reactant and product regions, and ξ

is the value of the RC that was sampled during MD described by the system Hamiltonian $H(\mathbf{q})$. The dynamical frequency factor can be estimated by the velocity of a free particle along the RC direction

$$\langle |d\xi/dt| \rangle_{\xi^{\ddagger}} = \left(\frac{2}{\pi\beta} \right)^{1/2} \left\langle \left[\sum_{i=1}^{3N} \frac{1}{m_i} \left(\frac{\partial \xi(\mathbf{q})}{\partial q_i} \right)^2 \right]^{1/2} \right\rangle_{\xi^{\ddagger}} \quad (26)$$

where $\langle \dots \rangle_{\xi^{\ddagger}}$ denotes the conditional average computed at the dividing surface ξ^{\ddagger} . Both the average distribution of RC values in eq 25 and the gradient of the RC in eq 26 are readily computable from MD using umbrella sampling techniques, for example.

For pedagogical purposes, it is convenient to recast the average distribution function

$$\langle \tilde{\rho}(\xi) \rangle = \frac{\int d\mathbf{q} \exp[-\beta H(\mathbf{q})] \delta[\tilde{\xi}(\mathbf{q}) - \xi]}{\int d\mathbf{q} \exp[-\beta H(\mathbf{q})]} \quad (27)$$

in the perspective of the *potential of mean force* (PMF)⁷¹

$$w(\xi) = w(\xi^0) - \beta^{-1} \ln \left[\frac{\langle \tilde{\rho}(\xi) \rangle}{\langle \tilde{\rho}(\xi^0) \rangle} \right] \quad (28)$$

where ξ^0 and $w(\xi^0)$ are constants that are typically chosen to reflect initial conditions of the chemical system. For example, note that $\langle \rho(\xi) \rangle$ and $\langle \tilde{\rho}(\xi) \rangle$ in the above equations differ only by a normalization factor. This factor, within transition state theory, is chosen to normalize the initial distribution in the reactant state region.

The PMF relates the probability of sampling along the RC to a *free energy* profile. Higher probabilities are associated with relatively low free energy values compared to regions of lower probabilities. At the bottleneck (i.e., at $\min[\langle \tilde{\rho}(\xi) \rangle]$), the PMF corresponds to the barrier for a process under observation. The PMF, therefore, provides an intuitive *free energy surface* perspective to chemical dynamics. In terms of $w(\xi)$, the TST normalized density factor can be rewritten as

$$\langle \rho(\xi) \rangle = \frac{\exp[-\beta w(\xi)]}{\int_{-\infty}^{\infty} d\xi' \exp[-\beta w(\xi')]} \quad (29)$$

where the limits of integration in the denominator are over the reactant state region.

Another experimental measure of reactive chemical dynamics, capable of providing insights into nuclear quantum effects, is the kinetic isotope effect (KIE). The KIE is defined as the ratio of the reaction rate, k , involving the system with a lighter isotope (l) compared to the rate involving the system with a heavier (h) isotopic substitution

$$\text{KIE} = \frac{k^{(l)}}{k^{(h)}} \quad (30)$$

When the isotopic substitution involves a chemical bond pertaining to the rate-limiting step, the KIE is characterized as *primary*. If the substitution does not directly involve chemical bonds that are broken and formed, the KIE is referred to as *secondary*. This paper only addresses the

primary KIE for the intramolecular proton transfer reaction in malonaldehyde; however, the methodology presented here can be applied to estimating the secondary KIE as well.

As described in the previous section, the inclusion of zero-point motion and nuclear tunneling are important for calculating the rate of proton transfer. These nuclear quantum effects can be taken into account by using a *quantum* analog of TST based on the path integral formalism⁷²

$$k_{\text{PI-QTST}} = 1/2 \langle |d\xi_c/dt| \rangle_{\xi_c^{\ddagger}} \langle \rho(\xi_c^{\ddagger}) \rangle \quad (31)$$

where $d\xi_c/dt$ is the dynamical frequency factor and ρ is the density of sampled *centroid* RC values, defined as

$$\langle \rho(\xi_c) \rangle = \frac{\int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta \Phi(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)})] \delta[\tilde{\xi}_c(\mathbf{q}^{(c)}) - \xi_c]}{\int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta \Phi(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)})] h[\xi_c^{\ddagger} - \tilde{\xi}_c(\mathbf{q}^{(c)})]} \quad (32)$$

where $\beta = 1/k_B T$, h is the Heaviside step function, ξ_c^{\ddagger} is the location of the *dividing surface* partitioning the reactant and product regions, and $\tilde{\xi}_c$ is the value of the RC (as a function of the centroid coordinates $\mathbf{q}^{(c)} = 1/P \sum_{s=1}^P \mathbf{q}^{(s)}$) that was sampled during MD described by the effective potential $\Phi(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)})$ defined in eq 21. Similar to classical TST, the dynamical frequency factor can be estimated by the velocity of a free particle along the centroid RC direction

$$\langle |d\xi_c/dt| \rangle_{\xi_c^{\ddagger}} = \left(\frac{2}{\pi\beta} \right)^{1/2} \left\langle \left[\sum_{i=1}^{3N} \frac{1}{m_i} \left(\frac{\partial \xi_c(\mathbf{q}^{(c)})}{\partial q_i^{(c)}} \right)^2 \right]^{1/2} \right\rangle_{\xi_c^{\ddagger}} \quad (33)$$

Furthermore, the centroid density can be recast in terms of a *quantum* PMF $w(\xi_c)$ as

$$\langle \rho(\xi_c) \rangle = \frac{\exp[-\beta w(\xi_c)]}{\int_{-\infty}^{\infty} d\xi'_c \exp[-\beta w(\xi'_c)]} \quad (34)$$

where the integration limits are again over the RS region.

From eq 31, it is possible to compute the KIE for proton transfer in malonaldehyde by directly forming the ratio of the rates of reaction, i.e., $\text{KIE} = k_{\text{PI-QTST}}^{(\text{H})}/k_{\text{PI-QTST}}^{(\text{D})}$, where (H) denotes the system with the hydrogen isotope and (D) indicates the deuterium isotopic substitution. Alternatively, one can estimate the ratio of the hydrogen and deuterium centroid densities using thermodynamic integration (TI) with respect to mass^{73,74}

$$\frac{\langle \rho(\xi_c^{\ddagger}) \rangle^{(\text{H})}}{\langle \rho(\xi_c^{\ddagger}) \rangle^{(\text{D})}} = \exp \left\{ - \int_0^1 d\lambda \frac{\partial}{\partial \lambda} \ln[\rho(\lambda)] \right\} \quad (35)$$

where the parameter λ interpolates between the masses of the hydrogen and deuterium isotopes as

$$\tilde{m}_i(\lambda) = (1 - \lambda)m_i^{(\text{H})} + \lambda m_i^{(\text{D})} \quad (36)$$

Substituting the centroid density expression [eq 32] into the logarithmic derivative, we obtain

$$\frac{\langle \rho(\xi_c^z) \rangle^{(H)}}{\langle \rho(\xi_c^z) \rangle^{(D)}} = \exp \left[-\beta \int_0^1 d\lambda \left(\left\langle \frac{d\Phi(\lambda)}{d\lambda} \right\rangle_{RS} - \left\langle \frac{d\Phi(\lambda)}{d\lambda} \right\rangle_{\xi_c^z} \right) \right] \quad (37)$$

where $\langle \dots \rangle_{RS}$ denotes the canonical average over the reactant state region, $\langle \dots \rangle_{\xi_c^z}$ denotes the conditional average with the system constrained to the dividing surface ξ_c^z , and $d\Phi/d\lambda$ can be computed from the PI virial-like estimator^{65,75}

$$\frac{d\Phi(\lambda)}{d\lambda} \simeq - \sum_{i=1}^N \frac{d\tilde{m}_i/d\lambda}{\tilde{m}_i} \left[\frac{3}{2\beta} + \frac{1}{2P} \left(\sum_{s=1}^P (\mathbf{q}_i^{(s)} - \mathbf{q}_i^{(c)}) \cdot \frac{\partial V(\mathbf{q}^{(s)})}{\partial \mathbf{q}_i^{(s)}} \right) \right] \quad (38)$$

All elements in the above equation are previously defined. While two separate trajectories are required for each value of λ , the savings in computation (compared to explicit construction of the PMFs) can be substantial if one needs to perform multiple umbrella sampling trajectories in order to map out the entire RC range due to an intrinsic high free energy barrier. $d\Phi/d\lambda$ is typically a smooth, slowly varying function and thus can be numerically integrated using the simple trapezoidal rule along uniformly spaced λ points. Because the isotopic ratio of the dynamical frequency factors only involves conditional averages with the system constrained to ξ_c^z , the computation is not expensive, and thus it is not necessary to consider TI for this factor.

An alternative to PI-QTST for incorporating nuclear quantum effects in thermal rate calculations is the quantum instanton (QI) approach.^{32–35} Here, we only briefly sketch the key equations for QI theory and the steps for computing the relevant quantities using PIMD. The derivation of the QI rate equation begins with the formally exact quantum mechanical expression for the thermal rate constant:³²

$$k(T) Q_r(T) \equiv kQ_r = \frac{1}{2\pi\hbar} \int dE \exp(-\beta E) \times \frac{(2\pi\hbar)^2}{2} \text{tr}[\hat{F}_a \delta(E - \hat{H}) \hat{F}_b \delta(E - \hat{H})] \quad (39)$$

where $Q_r(T)$ is the reactant partition function per unit volume at temperature T , β is the inverse temperature $1/k_B T$, and $\hat{F}_\gamma = i\hbar[\hat{H}, h(\xi_\gamma(\mathbf{q}))]$ is the flux operator where the Heaviside step function h defines the location of the dividing surface $\xi_\gamma(\mathbf{q}) = 0$. Both the microcanonical density operator $\delta(E - \hat{H})$ and the integral over E can be evaluated within the steepest descent approximation to give the following approximate expression for the QI rate constant:³⁴

$$k(T) \simeq k_{QI} = \frac{1}{Q_r} C_{ff}(0) \frac{\sqrt{\pi} \hbar}{2 \Delta H} \quad (40)$$

In the above equation, $C_{ff}(0)$ is the zero time value of the flux–flux correlation function generalized to the case of two separate dividing surfaces³⁵

$$C_{ff}(t) = \text{tr}[e^{-\beta\hat{H}/2} \hat{F}_a e^{-\beta\hat{H}/2} e^{i\hat{H}t/\hbar} \hat{F}_b e^{-i\hat{H}t/\hbar}] \quad (41)$$

and ΔH is a particular type of energy variance

$$\Delta H^2 = \frac{\text{tr}[\hat{\Delta}_a e^{-\beta\hat{H}/2} \hat{H}^2 \hat{\Delta}_b e^{-\beta\hat{H}/2}] - \text{tr}[\hat{\Delta}_a e^{-\beta\hat{H}/2} \hat{H} \hat{\Delta}_b e^{-\beta\hat{H}/2} \hat{H}]}{\text{tr}[\hat{\Delta}_a e^{-\beta\hat{H}/2} \hat{\Delta}_b e^{-\beta\hat{H}/2}]} \quad (42)$$

with $\hat{\Delta}_a$ and $\hat{\Delta}_b$ being a modified version of the Dirac Δ function:

$$\hat{\Delta}_\gamma \equiv \Delta[\xi(\mathbf{q}) - \xi_\gamma] = \sqrt{\sum_{i=1}^N m_i^{-1} (\nabla_i \xi_\gamma(\mathbf{q}))^2} \times \delta[\xi(\mathbf{q}) - \xi_\gamma] \quad (43)$$

Rewriting eq 40 in the form

$$k_{QI} = \frac{C_{dd}(0)}{Q_r} \left\{ \frac{C_{ff}(0)}{C_{dd}(0)} \frac{\sqrt{\pi} \hbar}{2 \Delta H} \right\} \quad (44)$$

leads to components that are easily computable using PIMD, where

$$\begin{aligned} & \frac{C_{dd}(0; \xi_a, \xi_b)}{Q_r} \\ & \int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta\Phi(\{\mathbf{q}^{(s)}\})] \Delta \times \\ & \frac{[\xi(\mathbf{q}^{(P)}) - \xi_a] \Delta[\xi(\mathbf{q}^{(P/2)}) - \xi_b]}{\int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta\Phi(\{\mathbf{q}^{(s)}\})] \times} \\ & h[\xi^z - \xi(\mathbf{q}^{(P)})] h[\xi^z - \xi(\mathbf{q}^{(P/2)})] \end{aligned} \quad (45)$$

$$C_{ff}(0)/C_{dd}(0) = \langle f_v(\{\mathbf{q}^{(s)}\}) \rangle_{\xi(\mathbf{p}), \xi(\mathbf{p}/2)} \quad (46)$$

$$\Delta H^2 = \frac{1}{2} \langle F(\{\mathbf{q}^{(s)}\})^2 + G(\{\mathbf{q}^{(s)}\}) \rangle_{\xi(\mathbf{p}), \xi(\mathbf{p}/2)} \quad (47)$$

The conditional average $\langle \dots \rangle_{\xi(\mathbf{p}), \xi(\mathbf{p}/2)}$ is computed from the ensemble sampled with the P and $P/2$ PI slices constrained to the dividing surfaces

$$\begin{aligned} & \langle \dots \rangle_{\xi(\mathbf{p}), \xi(\mathbf{p}/2)} \\ & \int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta\Phi(\{\mathbf{q}^{(s)}\})] \Delta \times \\ & \frac{[\xi(\mathbf{q}^{(P)}) - \xi_a] \Delta[\xi(\mathbf{q}^{(P/2)}) - \xi_b] \times (\dots)}{\int d\mathbf{q}^{(1)} d\mathbf{q}^{(2)} \dots d\mathbf{q}^{(P)} \exp[-\beta\Phi(\{\mathbf{q}^{(s)}\})] \Delta \times} \\ & [\xi(\mathbf{q}^{(P)}) - \xi_a] \Delta[\xi(\mathbf{q}^{(P/2)}) - \xi_b] \end{aligned} \quad (48)$$

where the quantities within the average are defined as follows:

$$\begin{aligned} f_v(\{\mathbf{q}^{(s)}\}) &= \left(\frac{iP}{2\hbar\beta} \right)^2 \frac{\sum_{i=1}^N \nabla_i \xi_a(\mathbf{q}^{(P)}) \cdot (\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P-1)})}{\sqrt{\sum_{i=1}^N m_i^{-1} [\nabla_i \xi_a(\mathbf{q}^{(P)})]^2}} \times \\ & \frac{\sum_{i=1}^N \nabla_i \xi_b(\mathbf{q}^{(P/2)}) \cdot (\mathbf{q}_i^{(P/2+1)} - \mathbf{q}_i^{(P/2-1)})}{\sqrt{\sum_{i=1}^N m_i^{-1} [\nabla_i \xi_b(\mathbf{q}^{(P/2)})]^2}} \end{aligned} \quad (49)$$

$$F(\{\mathbf{q}^{(s)}\}) = -\frac{P}{\hbar^2\beta^2} \left\{ \sum_{s=1}^{P/2} - \sum_{s=P/2+1}^P \right\} \sum_{i=1}^N m_i (\mathbf{q}_i^{(s)} - \mathbf{q}_i^{(s-1)})^2 + \frac{2}{P} \left\{ \sum_{s=1}^{P/2-1} - \sum_{s=P/2+1}^{P-1} \right\} V(\mathbf{q}^{(s)}) \quad (50)$$

$$G(\{\mathbf{q}^{(s)}\}) = \frac{2fP}{\beta^2} - \frac{4P}{\hbar^2\beta^3} \sum_{s=1}^P \sum_{i=1}^N m_i (\mathbf{q}_i^{(s)} - \mathbf{q}_i^{(s-1)})^2 \quad (51)$$

In the expression for G , f is the total number of degrees of freedom (i.e., $f = 3N$ in our malonaldehyde application).

All factors required to calculate the QI rate can be computed using the PIMD facilities within Amber.²⁸ For example, the joint distribution [eq 45] can be computed using umbrella sampling along the reaction coordinates of the P and $P/2$ slices. A set of two-dimensional (2D) biased simulations, each enhancing the sampling near a particular point of the 2D ($\xi_P \times \xi_{P/2}$) configurational space, is required to map out the entire QI joint distribution. Using the weighted histogram analysis method (WHAM),^{71,76,77} the generated biased distributions can be unbiased to form $C_{dd}(0)/Q_r$ on the DG-EVB ground-state surface. The remaining factors, involving the conditional average of f_v , F , and G , can be computed again using umbrella sampling with the P and $P/2$ slices constrained to the dividing surface ξ^z . Now, one can estimate KIEs from QI calculations by directly forming the ratio of the rates or by TI integration with respect to mass. For the latter case, the canonical average is defined over the 2D configurational space of the RS region delineated by the two dividing surfaces, and the conditional average is computed with the P and $P/2$ slices constrained to the dividing surface ξ^z . Also, the modified Dirac Δ functions in the conditional average of eq 48 introduce an additional term in the estimator for the logarithmic derivative of the delta-delta correlation function

$$-\beta^{-1} \frac{d \ln C_{dd}(\lambda)}{d\lambda} \approx -\sum_{i=1}^N \frac{d\tilde{m}_i/d\lambda}{\tilde{m}_i} \left[\frac{3}{2\beta} + \frac{1}{2P} \left(\sum_{s=1}^P (\mathbf{q}_i^{(s)} - \mathbf{q}_i^{(c)}) \cdot \frac{\partial V(\mathbf{q}^{(s)})}{\partial \mathbf{q}_i^{(s)}} \right) - \frac{\sum_{\gamma=a,b} |\nabla_{i,\xi_\gamma}(\mathbf{q})|^2}{2\tilde{m}_i \left\| \sum_{i=1}^N \tilde{m}_i^{-1} (\nabla_{i,\xi_\gamma}(\mathbf{q})) \right\|^2} \right] \quad (52)$$

Thus, whereas TI integration with respect to mass within PIMD uses the same estimator for sampling at the dividing surface and in the RS, QI uses eq 52 for the sampling at the dividing surfaces and eq 38 for sampling in the RS, the difference being the last term in eq 52.

3. Simulation Details

The symmetric intramolecular proton transfer reaction in malonaldehyde is depicted in Figure 1. All classical and quantum nuclear molecular dynamics simulations were performed using the DG-EVB facility within Amber 11.²⁸ The DG-EVB surface was fit using the energy, gradient, and

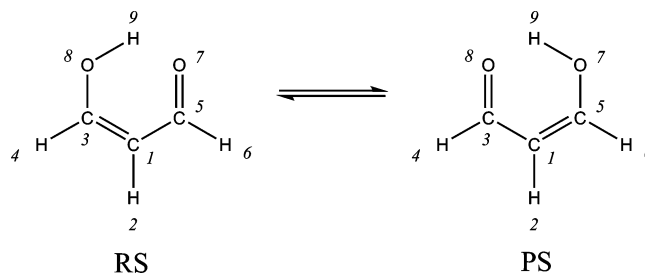


Figure 1. Intramolecular proton transfer reaction in malonaldehyde.

Hessian information at the RS, TS, PS, and off-TS geometries. The geometries at the RS, TS, and PS were optimized according to the WIBD^{78,79} model chemistry using a development version of the Gaussian suite.⁸⁰ The additional nonoptimized off-TS geometry lies *in front* of the TS and corresponds to both OH bonds being 1.5 Å. This point was chosen to improve the surface fit to the repulsive wall in front of the TS. Being able to systematically incorporate additional ab initio points as needed for refining the surface fit is one strength of the DG-EVB approach. The gradients and Hessian tensors were computed from the resulting geometries at the CCSD/cc-pVTZ level of theory. This choice of ab initio methods is comparable to the most accurate approach described in the literature for malonaldehyde, employing frozen-core CCSD(T)/aug-cc-pV5Z single-point calculations at frozen-core CCSD(T)/aug-cc-pVTZ optimized C_s and C_{2v} structures.³⁶ The present calculations yield a barrier height of 4.08 kcal/mol, nearly identical to the barrier of 4.09 kcal/mol obtained by Bowman.

The diabatic state Hamiltonian matrix element is approximated as a harmonic expansion about the ab initio optimized minimum (RS or PS) plus a scaled nonbonding, van der Waals, exponential-6 term from the universal force field (UFF)⁸¹ and Amber force field terms for angles and dihedrals:

$$H_{NN}(\mathbf{q}) = C + \mathbf{G}_N \cdot \Delta \mathbf{q}_r + \frac{\Delta \mathbf{q}_r^i \cdot \tilde{\mathbf{K}}_N \cdot \Delta \mathbf{q}_r}{2} + \eta \sum_i A_{\text{UFF}}^i \exp[-B_{\text{UFF}}^i \Delta \mathbf{q}_r^i] + \sum_{\text{angles}} K_\theta (\mathbf{q}_\theta - \theta_0)^2 + \sum_{\text{dihedrals}} (V_n/2)(1 + \cos[n\mathbf{q}_\phi - \delta]) \quad (53)$$

In the above equation, A_{UFF}^i and B_{UFF}^i are UFF exponential-6 parameters, $\Delta \mathbf{q}_r^i$ is the selected repulsive coordinate (e.g., the distance between the transferring H and the acceptor O in malonaldehyde) for H_{NN} and

$$C = \varepsilon_\psi(\mathbf{q}_N) - \eta \sum_i A_{\text{UFF}}^i \exp[-B_{\text{UFF}}^i \Delta \mathbf{q}_r^i] - \sum_{\text{angles}} K_\theta (\mathbf{q}_\theta - \theta_0)^2 - \sum_{\text{dihedrals}} (V_n/2)(1 + \cos[n\mathbf{q}_\phi - \delta]) \quad (54)$$

is a constant ensuring that the ab initio energy is recovered at the DG-EVB data point, $\mathbf{q} = \mathbf{q}_N$. A set of redundant internal coordinates, $\mathbf{q} = \{\mathbf{q}_r, \mathbf{q}_\theta, \mathbf{q}_\phi\}$, comprised of bond lengths (r), angles (θ), and dihedrals (ϕ) is used in the DG-EVB method to maintain invariance of the resulting energy hypersurface under global rotations. Typically, $H_{12}^2(\mathbf{q})$ does

not depend on the full set of redundant internal coordinates, and a subspace comprised of coordinates that change within a prescribed tolerance between the reactant and product configurations has been shown to provide sufficient accuracy.²⁶ For the malonaldehyde system, the above parameters for the angle (K_θ, θ_0) and dihedral (V_n, n, δ) interactions are taken from GAFF (generalized Amber force field).⁶ The UFF interactions were scaled by $\eta = 0.60$, and an *optimized* average value of 0.85 was used for all Gaussian α_K parameters in this study.²⁶

The *classical* RC ξ , chosen to represent the breaking and formation of a chemical bond, is defined as the difference of bond lengths r between the donor (D), the acceptor (A), and the transferring particle (H) positions

$$\xi(\mathbf{q}) = r(\mathbf{q}_D, \mathbf{q}_H) - r(\mathbf{q}_A, \mathbf{q}_H) \quad (55)$$

$\xi(\mathbf{q}) < 0$, thus, represents the reactant state region, while $\xi(\mathbf{q}) > 0$ represents the product state and $\xi(\mathbf{q}) = 0$ delineates the transition state (TS). Within the umbrella sampling framework,⁷¹ the system Hamiltonian is described by the modified potential

$$\begin{aligned} V_{\text{biased}}^{(n)}(\mathbf{q}) &= \varepsilon_0(\mathbf{q}) + V_{\text{umb}}^{(n)}(\mathbf{q}) \\ &= \varepsilon_0(\mathbf{q}) + \frac{1}{2}k^{(n)}[\xi(\mathbf{q}) - \xi_0^{(n)}(\mathbf{q})]^2 \end{aligned} \quad (56)$$

where \mathbf{q} is the set of system coordinates, k is the harmonic force constant parameter, and $V_{\text{umb}}^{(n)}$ is a biasing potential that is added to the original system potential ε_0 (obtained from diagonalization of the EVB matrix) to enhance the sampling of a predetermined region of configuration space near $\xi_0^{(n)}$. The distributions of $\xi^{(n)}$ from all of the (n) biased simulations are unbiased and combined using WHAM^{71,76} to form the PMF describing the chemical reaction of interest. All biased sampling simulations were performed in the NVT ensemble at a temperature of 300 K using a leapfrog Verlet integration time step of 0.5 fs for a minimum of 3 ns of total sampling per window.⁸² The temperature was maintained via coupling of the system to a Langevin thermostat⁸³ with a collision frequency of 1 ps^{-1} .

With the exception of the modifications described below, similar simulation protocols were followed for the PIMD simulations. The RC chosen to describe the breaking and formation of a chemical bond within the PI-QTST framework is the difference of bond lengths r between the donor, the acceptor, and transferring particle centroid coordinates

$$\tilde{\xi}_c(\mathbf{q}) = r(\mathbf{q}_D^{(c)}, \mathbf{q}_H^{(c)}) - r(\mathbf{q}_A^{(c)}, \mathbf{q}_H^{(c)}) \quad (57)$$

where the centroid position for particle χ is defined as the average of positions over the P path integral slices

$$\mathbf{q}_\chi^{(c)} = \frac{1}{P} \sum_{s=1}^P \mathbf{q}_\chi^{(s)} \quad (58)$$

Substituting $\tilde{\xi}_c$ for ξ in eq 56 allows for enhanced sampling of a prescribed reaction path using the same umbrella sampling procedure as employed for the classical molecular dynamics. The *quantum* PMF then is generated using WHAM^{71,76} from the centroid densities sampled in all the

biased trajectories. In contrast to the classical simulations described above, the PIMD was propagated in the normal mode representation⁶⁶ using a 0.5 fs leapfrog Verlet integration time step and with the temperature maintained at 300 K via coupling to a Nosé–Hoover chain bath of size four. Each region of the reaction path in the vicinity of $\xi_0^{(n)}$ is sampled for a minimum of 3 ns.

Two procedures were utilized for estimating the primary KIE in malonaldehyde. In the direct approach, the PI-QTST⁷² rates for the intramolecular proton transfer of the hydrogen and deuterium isotopes are computed separately using eq 31 to form the ratio $k_{\text{PI-QTST}}^{(\text{H})}/k_{\text{PI-QTST}}^{(\text{D})}$. The centroid density component of the PI-QTST rate was computed using umbrella sampling with $k^{(n)} = 100.0 \text{ kcal/mol } \text{Å}^2$ and $\xi_0^{(n)} = \{-0.60, -0.40, -0.20, 0.0, 0.20, 0.40, 0.60\} \text{ Å}$. The dynamical frequency factor was computed again with umbrella sampling with the RC restrained to the transition state region, i.e., $\xi_0 = 0.0 \text{ Å}$, using a harmonic force constant of $k = 2000.0 \text{ kcal/mol } \text{Å}^2$. A total of 14 independent biased trajectories (seven for each isotopically labeled proton transfer) are required for estimating the ratio of centroid densities, while two independent biased simulations are required for estimating the ratio of the frequency factors. In the TI with respect to mass approach,^{73,74} the ratio of isotopic centroid densities is estimated using eq 37 from a set of $\lambda = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ simulations where the average $\langle \dots \rangle_{\text{RS}}$ is computed from ground-state EVB sampling in the RS region and the average $\langle \dots \rangle_{\xi_c^{\text{TS}}}$ is computed from umbrella sampling with the RC restrained to $\xi_0 = 0.0 \text{ Å}$ via a harmonic force constant, $k = 2000.0 \text{ kcal/mol } \text{Å}^2$. The above TI by mass, therefore, entails a total of $6\lambda \times 2$ trajectories/ $\lambda = 12$ independent trajectories. Although the computational costs appear similar for both alternatives, it is expected that the statistical error for the KIE estimate based on relative rates [from eq 35] will be smaller compared to the estimate based on absolute rates [from eq 31]. For cases where a large number of biased trajectories are required to cover the range of RC values, due to an intrinsically high free energy barrier of the system or because of multiple dividing surfaces, the TI by mass route will be computationally less expensive.

Particularly in the case of QI^{33,34} calculations, TI by mass^{73,74} is a substantially more efficient approach for estimating KIEs. In the direct approach, the conditional averages of f_v , F , and G [eqs 49–51] were computed using umbrella sampling with the P and $P/2$ PI slices restrained to the dividing surface with a harmonic force constant of $k = 2000.0 \text{ kcal/mol } \text{Å}^2$. The 2D joint distribution was computed using umbrella sampling with $k^{(n)} = 100.0 \text{ kcal/mol } \text{Å}^2$ and $\xi_0^{(n)} = \{-1.00, -0.80, -0.60, -0.40, -0.20, 0.0, 0.20, 0.40, 0.60, 0.80, 1.00\} \text{ Å}$ for all combinations of pairs of restraints applied to the P and $P/2$ PI slices. Thus, we have a grid of $11 \times 11 = 121$ independent biased sampling trajectories for each isotope, for a total of 242 trajectories in a single KIE estimate. Compared to the 12 trajectories required for TI by mass, the computation disparity is significant. A protocol similar to the above PI-QTST details was used for the QI TI with respect to the mass procedure, with the only exception being that the conditional average in eq 37 is computed from umbrella sampling restraining the RC of both

Table 1. Computed TST, PI-QTST, and QI Chemical Rates and KIE for the Intramolecular Proton Transfer Reaction in Malonaldehyde^a

$$k_{\text{TST}} = \frac{1}{2} \langle \dot{\xi} \rangle_{\xi^{\ddagger}} \langle \rho(\xi^{\ddagger}) \rangle$$

isotope	$\langle \dot{\xi} \rangle_{\xi^{\ddagger}} [\text{\AA} \cdot \text{s}^{-1}]$	$\langle \rho(\xi^{\ddagger}) \rangle [\text{\AA}^{-1}]$	$k_{\text{TST}} [\text{s}^{-1}]$	KIE
H	$2.48 \times 10^{13} (\pm 1.31 \times 10^9)$	$2.04 \times 10^{-3} (\pm 1.81 \times 10^{-4})$	$2.53 \times 10^{10} (\pm 2.25 \times 10^9)$	1.54 (± 0.16)
D	$1.78 \times 10^{13} (\pm 1.18 \times 10^9)$	$1.84 \times 10^{-3} (\pm 1.04 \times 10^{-4})$	$1.64 \times 10^{10} (\pm 9.28 \times 10^8)$	

$$k_{\text{PI-QTST}} = \frac{1}{2} \langle |\dot{\xi}_c| \rangle_{\xi_c^{\ddagger}} \langle \rho(\xi_c^{\ddagger}) \rangle$$

isotope	$\langle \dot{\xi}_c \rangle_{\xi_c^{\ddagger}} [\text{\AA} \cdot \text{s}^{-1}]$	$\langle \rho(\xi_c^{\ddagger}) \rangle [\text{\AA}^{-1}]$	$k_{\text{PI-QTST}} [\text{s}^{-1}]$	KIE
H	$2.46 \times 10^{13} (\pm 4.53 \times 10^8)$	$5.34 \times 10^{-2} (\pm 2.77 \times 10^{-3})$	$6.56 \times 10^{11} (\pm 3.40 \times 10^{10})$	4.05 (± 0.27)
D	$1.77 \times 10^{13} (\pm 7.94 \times 10^8)$	$1.83 \times 10^{-2} (\pm 7.31 \times 10^{-4})$	$1.62 \times 10^{11} (\pm 6.47 \times 10^9)$	

$$k_{\text{QI}} = \frac{C_{dd}(0)}{Q_r} \left\{ \frac{C_{ff}(0)}{C_{dd}(0)} \frac{\sqrt{\pi} \hbar}{2 \Delta H} \right\}$$

isotope	$\left\{ \frac{C_{ff}(0)}{C_{dd}(0)} \frac{\sqrt{\pi} \hbar}{2 \Delta H} \right\} [\text{\AA}^2 \text{s}^{-1}]$	$\frac{C_{dd}(0)}{Q_r} [\text{\AA}^{-2}]$	$k_{\text{QI}} [\text{s}^{-1}]$	KIE
H	$3.71 \times 10^{12} (\pm 9.61 \times 10^{11})$	$5.98 \times 10^{-2} (\pm 9.76 \times 10^{-4})$	$2.22 \times 10^{11} (\pm 5.76 \times 10^{10})$	2.41 (± 0.86)
D	$3.44 \times 10^{12} (\pm 8.38 \times 10^{11})$	$2.67 \times 10^{-2} (\pm 6.02 \times 10^{-4})$	$9.19 \times 10^{10} (\pm 2.25 \times 10^{10})$	

^a The uncertainties given within parentheses are estimated from the distribution of results as a parameter of the MD sampling intervals.

the P and $P/2$ slices to $\xi_0 = 0.0 \text{ \AA}$ via a harmonic force constant of $k = 2000.0 \text{ kcal/mol \AA}^2$, i.e., $\langle \dots \rangle_{\xi_c^{\ddagger}} \rightarrow \langle \dots \rangle_{\xi_{(P)}}$, $\xi_{(P/2)}$. The relevant estimator used for the constrained sampling on the dividing surfaces is given by eq 52, while sampling within the RS well uses the estimator defined in eq 38.

Estimating the uncertainty in a rate calculation is not straightforward, especially when the quantities are derived from sampling over phase space. The frequency factor, PMF, and TI by mass calculations are subject to uncertainties related to the level of convergence of phase space sampling. To provide an estimate of this type of uncertainty, we compute the various components as a parameter of MD sampling intervals. The uncertainties reported in Table 1, using this approach, give an indication of the variability of the results as a parameter of phase space averaging. Adequate MD sampling should give averages with a relatively small standard deviation.

4. Results and Discussion

To evaluate the effectiveness of the DG-EVB method for constructing ab initio based reactive PESs for modeling chemical dynamics, the classical TST rate, PI-QTST rate, QI rate, and KIE for the prototypical intramolecular proton transfer reaction in malonaldehyde are computed. The calculations of these observables require MD sampling of configurational space to obtain thermodynamic averages. Furthermore, since the chemistry under consideration involves a proton, nuclear quantum effects, such as zero-point motion and tunneling, are important for describing the

reaction rate. These elements have been integrated into the latest release of the Amber biosimulation suite (version 11),²⁸ and the results for malonaldehyde are described below.

The 2D PES for the intramolecular proton transfer reaction in malonaldehyde, employing four Gaussian centers (located at the RS, TS, PS, and off-TS) to fit H_{12}^2 , is depicted in Figure 2. The diabatic state H_{NV} is represented as a quadratic expansion about the ab initio minimum configuration, augmented by a nonbonding, van der Waals, exponential-6 term from the universal force field⁸¹ and Amber angle and dihedral terms [eq 53]. The UFF term was included to prevent an anomalous “swimming hole” located behind the TS region geometry.²⁷ Alternatively, one can fill in this swimming hole by placing additional distributed Gaussians in these anomalous regions. The minima are indicated with a green dot and the TS with an orange dot. The additional configuration (indicated with a blue dot) along the plane orthogonal to the reaction path improves the curvature of the surface in this region. The goal here is to develop a robust, automated method capable of generating a multidimensional surface from ab initio information gleaned from sparse and strategically chosen geometries. For the malonaldehyde system, utilizing a quadratic expansion about the ab initio minimum and empirical FF terms is adequate for generating an accurate smooth PES for exploring chemical dynamics. The \mathbf{B} vector of eq 14 was solved using the GMRES^{57–60} algorithm with a convergence tolerance of 1×10^{-9} . The maximum error between the resulting DG-EVB and ab initio energies for geometries at the fit points was computed to be $7.49 \times 10^{-5} \text{ kcal/mol}$. The correspond-

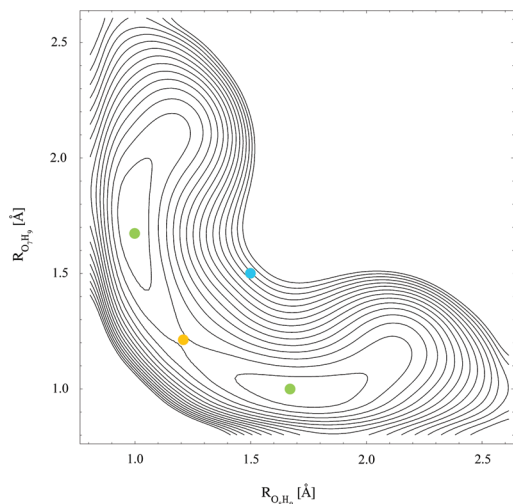


Figure 2. Two-dimensional potential energy surface for the intramolecular proton transfer reaction in malonaldehyde employing a four Gaussian fit for H_{12} and H_{MN} with a UFF repulsive term and Amber angle and dihedral terms. The minima are denoted with a green dot and the TS with an orange dot. An additional off-TS point (depicted by the blue dot) along the plane orthogonal to the reaction path improves the overall fit. A grid of geometries obtained from relaxed scans at the Hartree–Fock level and the cc-pVTZ basis set was used in the DG-EVB energy calculations.

ing maximum root-mean-square deviations between the DG-EVB and ab initio gradients is 5.65×10^{-4} kcal/mol Å, and the maximum difference between elements of the DG-EVB and ab initio internal coordinate Hessian matrices is 4.86×10^{-8} hartree/(internal coordinate)². Our DG-EVB surface exhibits an energy barrier height of 4.08 kcal/mol, which is in excellent agreement with the best estimate of 4.09 kcal/mol published in the literature.³⁶

With a well-defined PES, it is now possible to perform nuclear dynamics on this surface incorporating ab initio data. Since the ab initio energy barrier of 4.08 kcal/mol is well above thermal energy, conventional MD may not adequately sample important TS configurations. The resulting PMF (Figure 3) obtained from ground-state dynamics indicates that malonaldehyde predominantly samples the conformational space of the reactant and product region but not so much the barrier region. Figure 3 is obtained by initiating conventional MD on the reactant minimum and collecting the statistics (indicated by the gray histogram in the background) for sampling a particular value of the RC. At an average temperature of 300 K, the system has sufficient thermal energy to overcome the energy barrier. Sampling near the TS, however, remains statistically insignificant. This intrinsic *rare event* nature of the chemistry, thus, requires enhanced sampling techniques for sampling conformations within the important TS region.

Utilizing the umbrella sampling procedure⁷¹ described in the methodology section, a set of independent biased trajectories was used to map out the distribution of RC values over the entire range of the reactive path. Each trajectory only enhances the sampling about a predefined RC value, $\xi_0^{(n)}$. Using the WHAM^{71,76} procedure, the set of biased distributions (gray curves, Figure 4) are combined to form a

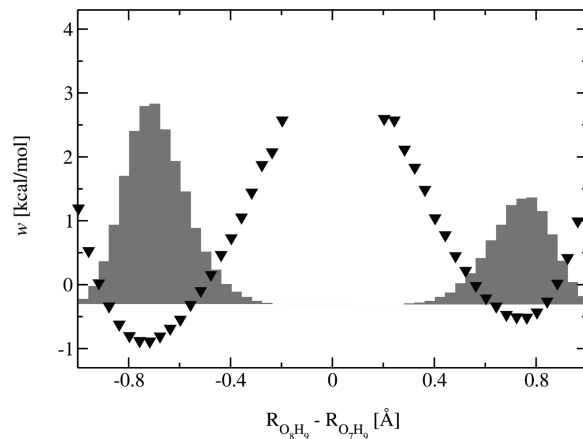


Figure 3. Potential of mean force obtained from direct sampling on the DG-EVB ground-state surface (▼). The gray background depicts the histogram of RC values encountered during the MD.

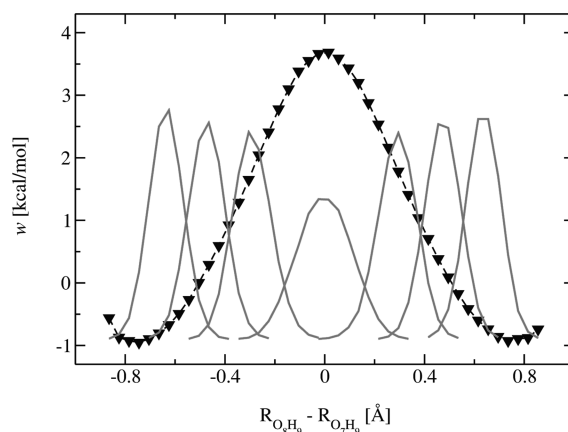


Figure 4. Potential of mean force obtained from umbrella sampling on the DG-EVB ground-state surface. The dashed line is a cubic spline fit through the data points (▼). The gray curves in the background show the histogram of RC values sampled along the biased MD trajectories, with $\xi_0^{(n)} = \{-0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6\}$ Å. These distributions are combined using WHAM.

free energy profile (▼curve, Figure 4) spanning the entire range of the RC for the proton transfer under observation. The resulting normalized density contribution [eq 29] to the TST rate is $2.04 \times 10^{-3} (\pm 1.81 \times 10^{-4})$ Å⁻¹. Again, using umbrella sampling with the RC restrained to the dividing surface ($\xi_0 = 0.0$ Å), the other contribution from the dynamical frequency factor can be estimated from the gradient of the RC. With a value of $2.48 \times 10^{13} (\pm 1.31 \times 10^9)$ Å s⁻¹ for $\langle d\xi/dt \rangle_{\xi^{\ddagger}}$, the classical TST estimate of the proton transfer rate is $2.53 \times 10^{10} (\pm 2.25 \times 10^9)$ s⁻¹.

While the above classical model of the proton experiences a free energy barrier of ~ 3.75 kcal/mol, zero-point motion and nuclear tunneling may effectively lower this reaction barrier. Figure 5 compares the quantum PMF (▼) obtained from PIMD sampling on the DG-EVB surface to the classical PMF (dotted line) from Figure 4. Both curves are normalized to their respective RS partition function. The free energy values at the TS ($w(\xi^{\ddagger})$ and $w(\xi_c^{\ddagger})$), therefore, are relative to absolute energy origins as defined in the classical TST

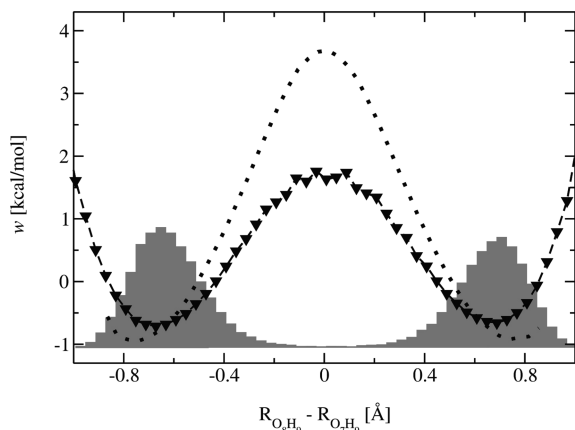


Figure 5. Potential of mean force obtained from direct PIMD sampling on the DG-EVB ground-state surface (▼). The gray background depicts the histogram of RC values encountered during the PIMD. For comparison, the classical PMF from Figure 4 is shown by the dotted curve.

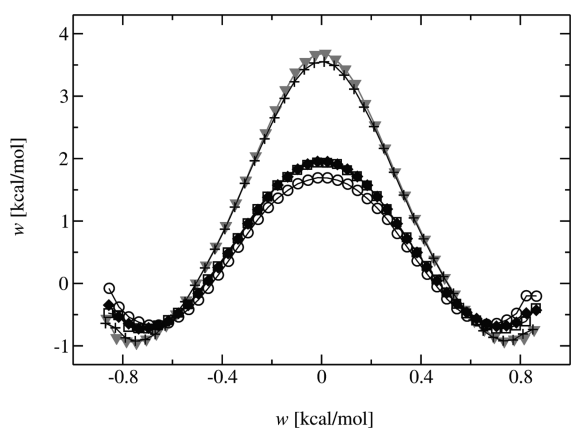


Figure 6. Potential of mean force as a parameter of the level of nuclear quantization: classical nuclei (▼), quantization only of the donor and acceptor oxygens (+), quantization of transferring proton (◆), quantization of donor and acceptor oxygens and transferring proton (◻), full nuclear quantization (○).

and PI-QTST rate equations. It is seen that full nuclear quantization of malonaldehyde lowers the free energy barrier by about 2 kcal/mol. This result is similar to trends observed in earlier studies of malonaldehyde by Tuckerman and Marx⁸⁴ and by Schofield and Iftimie.^{85,86} While inclusion of nuclear quantization sufficiently lowers the free energy barrier to allow sampling of the TS, a general framework for enhancing sampling of a particular region of configurational space is desirable. Reactions in enzyme environments, for example, may encounter barriers typically on the order of ~ 10 kcal/mol or higher. Our implementation of umbrella sampling within the PIMD function in Amber will permit studies of rare event phenomena using quantum dynamics approaches.

From a computational efficiency perspective, it is also worthwhile to consider if nuclear quantization is necessary for the entire system or if certain degrees of freedom can remain in the classical description. Figure 6 shows the PMFs as a parameter of the *level of nuclear quantization*, where the ▼ curve reproduces the classical description from Figure

4 and the ○ curve depicts the fully quantum description. Here, all PMFs are obtained using the umbrella sampling protocol described in the Simulation Details section. The ◻ curve reflects quantization only of the donor, acceptor, and transferring proton, while the ◆ curve shows the impact of quantization only of the transferring proton. Both of these curves are on top of each other and are slightly higher than the fully quantized system. The remaining curve (+) corresponds to quantization of only the donor and acceptor oxygens. As expected from theoretical considerations, nuclear quantization of the light particles (hydrogens) has the most significant effect on the free energy barrier height; whereas, quantization of the heavier oxygen atoms only has a minor effect compared to the classical description. For the case of malonaldehyde, quantizing the transferring proton is sufficient to capture the bulk of the quantum effects associated with our definition of the RC. This trend may not be generally transferrable to other more flexible molecules or to alternative prescriptions of the reaction path. In addition to lowering the reaction barrier, nuclear quantization also shifts the location of the reactant and product minima and changes the curvature in the transition state region. This alteration of both the barrier height and shape of the free energy surface will manifest in KIE measurements.

The estimates of primary KIEs in this study were obtained using both the direct approach of explicitly forming the ratio of the absolute rates and via TI with respect to mass. Table 1 shows all contributions to the chemical rates for the hydrogen and deuterium isotopes as computed from classical TST, PI-QTST, and QI prescriptions. Because experimental measurements of KIEs for malonaldehyde are unavailable, the latter two approximate quantum rates serve to provide ballpark estimates from well-established methods. The KIE of $4.05 (\pm 0.27)$ from PI-QTST is about one and a half times larger than the QI value of $2.41 (\pm 0.86)$. The absolute rates from PI-QTST, however, are approximately two to three times larger than the estimates from QI, due predominantly to the larger frequency factors. The contribution to the rate from the PMF is smaller in PI-QTST than in QI for both proton and deuterium transfers. Since PI-QTS and QI are approximate quantum rate methods, equivalently at the level of transition state theory, the differences in the rates are not due to dynamical recrossing effects. A clear relationship between PI-QTS and the QI method has yet to be established. Furthermore, the results indicate that classical TST is inadequate in providing KIE estimates. Here, isotopic substitution predominantly impacts the frequency factor and leaves the PMF contribution to the rate little changed, resulting in a KIE of 1.54. The effective lowering of the barrier height due to zero-point motion and nuclear tunneling are missing. Nuclear quantization via PI recovers these effects and provides the dominant contributing factor of $2.91 (\pm 0.19)$ to the PI-QTST KIE. The frequency factors obtained from PIMD are similar to those computed from classical sampling, suggesting that the centroid coordinates of the system at the dividing surface for this symmetric transfer can be represented by classical nuclear coordinates. For asymmetric reactions, this correspondence between the centroid and classical variables may not hold, and one cannot

simply assume that the computationally less demanding classical MD sampling will be sufficient for estimating the frequency factor component of the PI-QTST rate.

While a direct comparison of our computed KIEs with the published results of Schofield⁸⁵ is complicated by differences in the underlying PESs, it is worthwhile to identify the contributing factors giving rise to varying estimates of the KIE. In their previous PI-QTST study employing a molecular mechanical EVB potential for describing the proton transfer reaction,⁸⁵ the primary KIEs range from 6.49 to 11.41 and depend on the choice of RC. The sensitivity of KIEs on the prescription of the RC is to be expected, as PI-QTST is a transition state approximation to the quantum rate. Different RCs may result in varying levels of barrier recrossings, and it is this recrossing factor that is missing in the KIE calculations. Furthermore, the molecular mechanical parameters employed in that EVB formulation give rise to an energy difference of 8.75 kcal/mol between the reactant and transition state conformations. This activation energy is about two times larger than the barrier obtained from the present DG-EVB approach, leading to rates that are about 2–3 orders of magnitude smaller. Also, a higher barrier typically leads to a larger KIE, and this trend is observed between the two methods.

Figure 7a shows the 2D PMF associated with the joint distribution of the RC along the P and $P/2$ PI slices for the proton transfer reaction. The contribution to the QI rate is obtained using the dividing surface corresponding to the top of the free energy barrier in Figure 7b with $\xi_a = \xi_b = \xi^{\ddagger} = 0.0$ Å. The deuterium transfer is depicted by ▼, and the proton transfer is depicted by ○. Note that this PMF is normalized with respect to the reactant partition function within the full 2D space, and as such, one cannot simply perform a 1D biased sampling with both dividing surfaces set equal to each other. The 2D distribution requires a total of $11 \times 11 = 121$ biased sampling trajectories in order to map out the entire $\xi_a \times \xi_b$ configurational space. For estimating the KIE, the computation becomes quite expensive, totaling 242 trajectories for the malonaldehyde system.

As noted in the methods section, the computationally expensive PMF calculation can be avoided if one reformulates the ratio of isotope densities as a thermodynamic integration over mass. Figure 8 displays the average values of $-\beta\partial\Phi(\lambda)/\partial\lambda$ sampled during PIMD in the RS well and at the dividing surface as a parameter of λ . Integration over the λ values using eq 37 provides a centroid density value of 3.08 (± 0.01), which is comparable to 2.91 (± 0.19) obtained directly from the PMFs. Similarly, TI by mass for the QI calculations provides an estimate of 4.73 (± 0.02) for $C_{dd}^{(H)}(0)/C_{dd}^{(D)}(0)$ compared to the value of 2.24 (± 0.06) from direct calculation. In the direct calculations, the QI PMF barrier height for proton transfer is 1.68 kcal/mol and is 2.16 kcal/mol for deuterium transfer [see Figure 7b]. To get a $C_{dd}^{(H)}(0)/C_{dd}^{(D)}(0)$ of ~ 4.73 , the proton transfer barrier height needs to be lowered to 1.23 kcal/mol (while keeping the deuterium transfer barrier at 2.16 kcal/mol) or the deuterium transfer barrier height needs to increase to 2.60 kcal/mol (while keeping the proton transfer barrier at 1.68 kcal/mol). This resolution difference of ~ 0.5 kcal/mol is especially

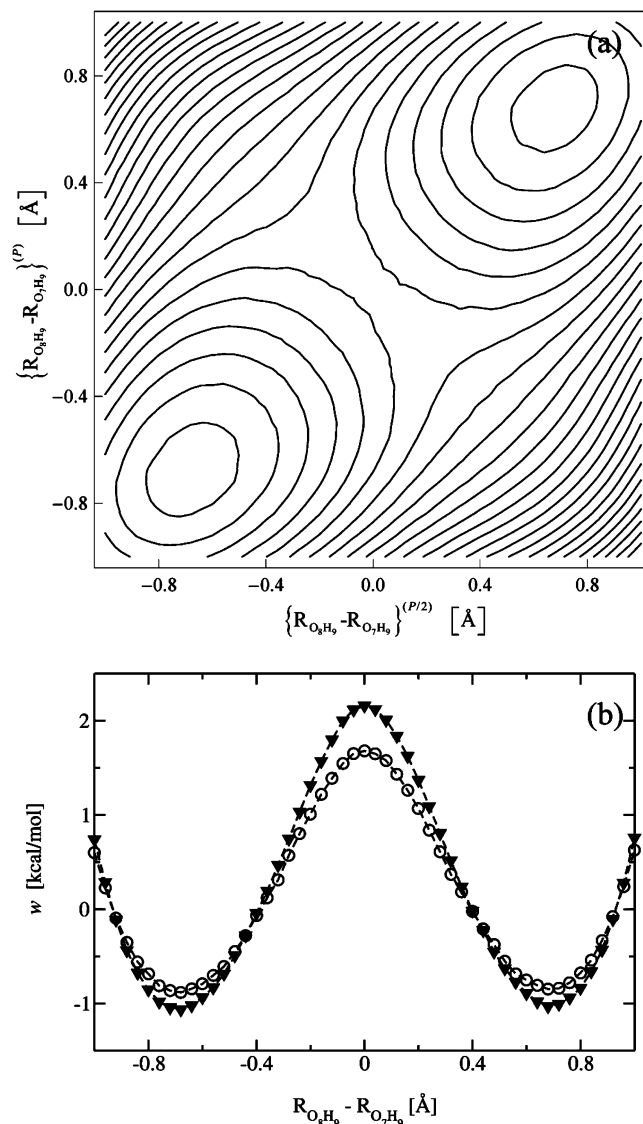


Figure 7. Potential of mean force associated with the joint probability density $C_{dd}(0; \xi_a, \xi_b)/Q_r$ of eq 45. (a) Contours for proton transfer (0.4 kcal/mol each) as a parameter of RC values along the P and $P/2$ PI slices. (b) Slice through the contours at $\xi_a = \xi_b$ that corresponds to a single dividing surface. The deuterium transfer is depicted by ▼, and the proton transfer is depicted by ○.

challenging to achieve in the QI 2D PMFs, where errors in the reweighting of the 121 biased distributions may accumulate and impact the global unbiased distribution. The virial estimator for the average quantities in TI by mass has been shown to converge efficiently.⁷⁴ The errors in the TI by mass estimate of the KIE, thus, are expected to be smaller than errors arising from the direct approach.

Multiplying the ratio of isotopic densities by the corresponding ratio of frequency factors, the PI-QTST estimate of 4.27 (± 0.01) for the KIE is within 80% of the QI value of 5.10 (± 1.81). The error in the QI estimate, however, is much larger than that from PI-QTST because of the variability of the frequency factor estimates. While both approaches average the frequency factor over the same time span (3 ns), QI appears to require more phase space sampling to achieve a higher level of convergence. On the other hand,

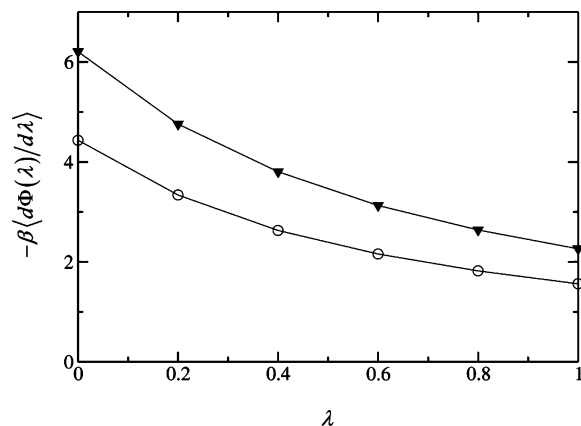


Figure 8. Average value of $-\beta\langle d\Phi(\lambda)/d\lambda\rangle$ sampled during PIMD in the RS region (\blacktriangledown) and at the dividing surface ξ_c^{\ddagger} (\circ) as a parameter of λ .

the phase space averages required for TI by mass converge much faster. A nanosecond of sampling of the virial estimators for $d\Phi/d\lambda$ at each value of λ is sufficient for obtaining the ratio of isotopic densities with relative uncertainties that are less than 1%. The savings in computation as well as a much smaller estimate of the uncertainty is especially evident in QI calculations, where the two sets of computationally demanding PIMD sampling can be avoided altogether by TI integration over the mass.

5. Concluding Remarks

This work presents an application of the distributed Gaussian EVB method for constructing an ab initio-based reactive potential energy surface describing the intramolecular proton transfer reaction in malonaldehyde. Albeit a deceptively simple gas-phase system, malonaldehyde has all the salient features necessary to test the methodology development toward a description of chemical dynamics in complex systems. Extension of the approach to treat nuclear quantum effects is made possible through coupling to path integral methods as well as by the path integral quantum TST and quantum instanton formalisms. Experimental measurements of rates and KIEs are important tools of chemical kinetics for elucidating the mechanism of complex chemical reactions. In situations where the KIE is important, one can either estimate this measurable by directly computing the ratio of the isotopic rates or by TI integration with respect to mass. Our results for malonaldehyde show both approaches to give similar KIEs, although TI integration with respect to mass is computationally less expensive compared to explicit construction of the PMFs. More importantly, the above functionalities of the DG-EVB approach have been made accessible to the broader community via integration within the Amber biosimulation suite.

Some needed improvements to the DG-EVB method, nevertheless, remain. The most apparent of these is our strategy for filling in the swimming hole near the transition state region by including a UFF repulsive interaction and Amber angle and dihedral terms to the diabatic state energy. Although this practical patch appears functional, a more general solution is desirable. Using a more realistic dissociative potential beyond the harmonic approximation, such as

a Morse-type interaction, and using a full force field prescription for the H_{ii} terms will be the subjects of future development of the DG approach. It is important to emphasize that H_{11} and H_{22} need to be higher than ϵ_{ψ} for the EVB approach to function satisfactorily.²⁷ Some solutions to overcome anomalous behavior of EVB include (1) adding more ab initio data points for the fitting procedure, (2) modifying the prescription of H_{ii} , and (3) modifying the prescription of H_{12} . The DG-EVB methods development has focused on options 1 and 2, either individually or in combination, while MCOMM has explored option 3.⁴⁸ In a recent paper, Truhlar and Tishchenko allowed H_{12} to become imaginary (i.e., equivalent to our letting H_{12}^2 become negative).⁵²

Additionally, the current DG implementation requires the user to optimize the Gaussian α_K parameters such that the DG energy, gradient, and Hessian reproduce corresponding ab initio information along IRC geometries. For the case of a single DG data point, this parameter is unique. When multiple Gaussians are used to enhance the fit, the choice of α_K parameters will affect the quality of the surface. Although this parametrization overhead cannot be circumvented, additional application of the DG method may provide some *rule of thumb* intuition as to the better choices of α_K values. For example, is it necessary to assign unique values for each distributed Gaussian or will an average parameter suffice? Is it better to use α_K values that provide more diffuse Gaussians or are more localized Gaussians the optimal choice? These are questions related to parameter sensitivity analysis that will be the subject of future studies. Furthermore, an extension of DG-EVB to treat chemical reactions in condensed environments requires the development of hybrid methods that fit only part of the *active site* region to ab initio data, as electronic structure calculations of the whole condensed phase configuration are prohibitively infeasible.

Acknowledgment. The Office of Naval Research (N00014-05-1-0457) supported this research. J.L.S. is grateful for the computer time on the Wayne State University Grid. K.F.W. is grateful for an allocation of computer time from the Center for High Performance Computing at the University of Utah. A portion of the computational resources for this project have been provided by the National Institutes of Health (Grant #NCRR 1 S10 RR17214-01) on the Arches Metacluster, administered by the University of Utah Center for High Performance Computing.

Supporting Information Available: Details of the electronic structures for malonaldehyde used in the DG-EVB surface fit. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75.
- (2) *Petascale Computing: Algorithms and Applications*; Bader, D., Ed.; Chapman & Hall/CRC Press: Boca Raton, FL, 2008.
- (3) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu, H.;

- Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155.
- (4) Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1.
- (5) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504.
- (6) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
- (7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.
- (8) Xu, Z.; Luo, H. H.; Tieleman, D. P. *J. Comput. Chem.* **2007**, *28*, 689.
- (9) Soares, T. A.; Hünenberger, P. H.; Kastenholz, M. A.; Krätzler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 725.
- (10) Hill, J.-R.; Freeman, C. M.; Subramanian, L. In *Rev. Comput. Chem.*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2007; p 141.
- (11) Balint-Kurti, G. G. *Adv. Chem. Phys.* **1975**, *30*, 137.
- (12) Eyring, H. *Trans. Faraday Soc.* **1938**, *34*, 3.
- (13) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1938**, *34*, 11.
- (14) Ogg, R. A.; Polanyi, M. *Trans. Faraday Soc.* **1934**, *31*, 604.
- (15) Evans, M. G.; Warhurst, E. *Trans. Faraday Soc.* **1938**, *34*, 614.
- (16) Venkatnathan, A.; Voth, G. A. *J. Chem. Theory Comput.* **2005**, *1*, 36.
- (17) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.
- (18) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons Inc.: New York, 1997.
- (19) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149.
- (20) Hutter, J.; Curioni, A. *Parallel Computing* **2005**, *31*, 1.
- (21) Carloni, P.; Rothlisberger, U.; Parrinello, M. *Acc. Chem. Res.* **2002**, *35*, 455.
- (22) Bolton, K.; Hase, W. L. In *Modern Methods for Multidimensional Dynamics*; Thompson, D. L., Ed.; World Scientific: Singapore, 1998; p 143.
- (23) Hase, W. L.; Song, K. H.; Gordon, M. S. *Comput. Sci. Eng.* **2003**, *5*, 36.
- (24) Schlegel, H. B. *Bull. Korean Chem. Soc.* **2003**, *24*, 837.
- (25) Schlegel, H. B.; Sonnenberg, J. L. *J. Chem. Theory Comput.* **2006**, *2*, 905.
- (26) Sonnenberg, J. L.; Schlegel, H. B. *Mol. Phys.* **2007**, *105*, 2719.
- (27) Sonnenberg, J. L.; Wong, K. F.; Voth, G. A.; Schlegel, H. B. *J. Chem. Theory Comput.* **2009**, *5*, 949.
- (28) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vaníček, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber*; University of California: San Francisco, 2010.
- (29) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, 1965.
- (30) Feynman, R. P. *Statistical Mechanics*; Benjamin: Reading, MA, 1972.
- (31) Kleinert, H. *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics*; World Scientific: Singapore, 1995.
- (32) Miller, W. H. *J. Chem. Phys.* **1975**, *62*, 1899.
- (33) Miller, W. H.; Zhao, Y.; Ceotto, M.; Yang, S. *J. Chem. Phys.* **2003**, *119*, 1329.
- (34) Yamamoto, T.; Miller, W. H. *J. Chem. Phys.* **2004**, *120*, 3086.
- (35) Miller, W. H.; Schwartz, S. D.; Tromp, J. W. *J. Chem. Phys.* **1983**, *79*, 4889.
- (36) Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. *J. Chem. Phys.* **2008**, *128*, 224314.
- (37) Viel, A.; Coutinho-Neto, M. D.; Manthe, U. *J. Chem. Phys.* **2007**, *126*, 024308.
- (38) Mil'nikov, G. V.; Yagi, K.; Taketsugu, T.; Nakamura, H.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 5036.
- (39) Mil'nikov, G. V.; Yagi, K.; Taketsugu, T.; Nakamura, H.; Hirao, K. *J. Chem. Phys.* **2003**, *119*, 10.
- (40) Yagi, K.; Taketsugu, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 10647.
- (41) Hazra, A.; Skone, J. H.; Hammes-Schiffer, S. *J. Chem. Phys.* **2009**, *130*, 054108.
- (42) Ben-Nun, M.; Martinez, T. *J. Phys. Chem. A* **1999**, *103*, 6055.
- (43) Sewell, T. D.; Guo, Y.; Thompson, D. L. *J. Chem. Phys.* **1995**, *103*, 8557.
- (44) Shida, N.; Barbara, P. F.; Almolöf, J. E. *J. Chem. Phys.* **1989**, *91*, 4061.
- (45) Makri, N.; Miller, W. H. *J. Chem. Phys.* **1989**, *91*, 4026.
- (46) Ruf, B. A.; Miller, W. H. *J. Chem. Soc., Faraday Trans.* **1988**, *84*, 1523.
- (47) Carrington, T.; Miller, W. H. *J. Chem. Phys.* **1986**, *84*, 4364.
- (48) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790.
- (49) Albu, T. V.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 8465.
- (50) Kim, Y.; Cochado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718.
- (51) Lin, H.; Pu, J.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4112.
- (52) Tishchenko, O.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1454.
- (53) Chandler, D. *Introduction to Modern Statistical Mechanics*; Elsevier: New York, 1991.
- (54) Chang, Y.-T.; Miller, W. H. *J. Phys. Chem.* **1990**, *94*, 5884.
- (55) Chang, Y.-T.; Minichino, C.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 4341.
- (56) Minichino, C.; Voth, G. A. *J. Phys. Chem. B* **1997**, *101*, 4544.
- (57) Saad, Y.; Schultz, M. H. *SIAM J. Sci. Stat. Comput.* **1986**, *7*, 856.
- (58) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393.
- (59) Pulay, P. *J. Comput. Chem.* **1982**, *3*, 556.

- (60) Pulay, P. In *Molecular Quantum Mechanics: Analytic Gradients and Beyond*; Csaszar, A. G., Fogarasi, G., Schaefer, H. F., III, Szalay, P. G., Eds.; ELTE Institute of Chemistry: Budapest, 2007; p 71.
- (61) Arndt, M.; Nairz, O.; Vos-Andreae, J.; Keller, C.; van der Zouw, G.; Zeilinger, A. *Nature* **1999**, *401*, 680.
- (62) Zuev, P. S.; Sheridan, R. S.; Albu, T. V.; Truhlar, D. G.; Hrovat, D. A.; Borden, W. T. *Science* **2003**, *299*, 867.
- (63) McMahon, R. J. *Science* **2003**, *299*, 833.
- (64) Chandler, D.; Wolynes, P. G. *J. Chem. Phys.* **1981**, *74*, 4078.
- (65) Ceperley, D. M. *Rev. Mod. Phys.* **1995**, *67*, 279.
- (66) Berne, B. J.; Thirumalai, D. *Annu. Rev. Phys. Chem.* **1986**, *37*, 401.
- (67) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
- (68) Hänggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251.
- (69) Kapral, R.; Consta, S.; McWhirter, L. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: Singapore, 1998.
- (70) Hinsen, K.; Roux, B. *J. Chem. Phys.* **1997**, *106*, 3567.
- (71) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.
- (72) Voth, G. A.; Chandler, D.; Miller, W. H. *J. Chem. Phys.* **1989**, *91*, 7749.
- (73) Yamamoto, T.; Miller, W. H. *J. Chem. Phys.* **2005**, *122*, 044106.
- (74) Vaníček, J.; Miller, W. H. *J. Chem. Phys.* **2007**, *127*, 114309.
- (75) Cao, J.; Berne, B. J. *J. Chem. Phys.* **1989**, *91*, 6359.
- (76) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (77) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
- (78) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.
- (79) Barnes, E. C.; Petersson, G. A.; Montgomery, J. A., Jr.; Frisch, M. J.; Martin, J. M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2687.
- (80) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Scalmani, G.; Mennucci, B.; Barone, V.; Petersson, G. A.; Caricato, M.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Li, X.; Hratchian, H. P.; Peralta, J. E.; Izmaylov, A. F.; Kudin, K. N.; Heyd, J. J.; Brothers, E.; Staroverov, V.; Zheng, G.; Kobayashi, R.; Normand, J.; Sonnenberg, J. L.; Ogliaro, F.; Bearpark, M.; Parandekar, P. V.; Ferguson, G. A.; Mayhall, N. J.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Burant, J. C.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Chen, W.; Wong, M. W.; Pople, J. A. *Gaussian 03*, Revision G.01; Gaussian, Inc, Wallingford, CT, 2007.
- (81) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (82) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U. K., 1987.
- (83) Pastor, R. W.; Brooks, B. R.; Szabo, A. *Mol. Phys.* **1988**, *65*, 1409.
- (84) Tuckerman, M.; Marx, D. *Phys. Rev. Lett.* **2001**, *86*, 4946.
- (85) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *115*, 5891.
- (86) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *114*, 6763.

CT900579K