

# Affordable *Ab Initio* Path Integral for Thermodynamic Properties via Molecular Dynamics Simulations Using Semiempirical Reference Potential

Yuanfei Xue,<sup>†</sup> Jia-Ning Wang,<sup>†</sup> Wenxin Hu,<sup>‡</sup> Jun Zheng,<sup>‡</sup> Yongle Li,<sup>¶</sup> Xiaoliang  
Pan,<sup>§</sup> Yan Mo,<sup>\*,†,||,⊥</sup> Yihan Shao,<sup>§</sup> Lu Wang,<sup>\*,#</sup> and Ye Mei<sup>\*,†,||,⊥</sup>

<sup>†</sup>*State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China  
Normal University, Shanghai 200062, China*

<sup>‡</sup>*The Computer Center, School of Data Science & Engineering, East China Normal University, Shanghai  
200062, China*

<sup>¶</sup>*Department of Physics, International Center of Quantum and Molecular Structure, and Shanghai Key  
Laboratory of High Temperature Superconductors, Shanghai University, Shanghai 200444, China*

<sup>§</sup>*Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, United  
States*

<sup>||</sup>*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

<sup>⊥</sup>*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

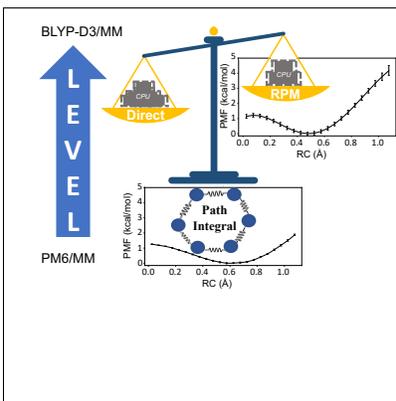
<sup>#</sup>*Department of Chemistry and Chemical Biology, Institute for Quantitative Biomedicine, Rutgers  
University, Piscataway, New Jersey 08854, United States*

E-mail: ymo@phy.ecnu.edu.cn; lwang@chem.rutgers.edu; samuel.y.mei@gmail.com

## Abstract

Path integral molecular dynamics (PIMD) is becoming a routinely applied method for the incorporation of the nuclear quantum effect in computer simulations. However, direct PIMD simulations at an *ab initio* level of theory are formidably expensive. Using the protonated 1,8-bis(dimethylamino)naphthalene molecule as an example, we show in this work that the computational expense for the intra-molecular proton transfer between the two nitrogen atoms can be remarkably reduced by implementing the idea of reference-potential methods. The simulation time can be easily extended to a scale of nanosecond while maintaining the accuracy on an *ab initio* level of theory for thermodynamic properties. In addition, the post-processing can be carried out in parallel on massive computer nodes. A 545-fold reduction in the total CPU time can be achieved in this way as compared to a direct PIMD simulation at the same *ab initio* level of theory.

## TOC Graphic



# Introduction

Hybrid QM/MM is now the method of choice for the studies of enzymatic reactions and chemical reactions in the condensed phase.<sup>1-13</sup> However, the application of QM/MM methods is often plagued with extremely poor computational scaling of the *ab initio* QM methods, as well as the long time scales of molecular dynamics propagation that are usually required before any essential dynamic processes for the degrees-of-freedom (DoF) orthogonal to the boosted one can be observed.<sup>14</sup> Furthermore, exploring the nuclear quantum effect (NQE), such as tunneling, nuclear delocalization phenomena, and zero-point energy of molecules, gains in popularity over the years.<sup>15-22</sup> One of the most appealing approaches to incorporating NQE is path integral molecular dynamics (PIMD).<sup>23</sup> PIMD is based on the isomorphism to approximately transform the representation of quantum effect in single-particle systems into a classical system with discrete finite number of pseudo-particles (or beads) along a cyclic path.<sup>24</sup> Being one of the most commonly used simulation methods, PIMD can be used to obtain thermostistical equilibrium properties and dynamic properties, with free energy profiles of reactions and reaction rate coefficients as two representative examples.<sup>25-28</sup>

However, applications of PIMD simulations in the studies of chemical reactions of complex systems are extremely demanding. On the one hand, PIMD requires  $P$  times more computational time than that of the classical molecular dynamics (MD), where  $P$  is the number of beads, onto which each quantum particle is mapped.  $P$  determines the convergence of PIMD simulations, and it needs to satisfy  $P > \hbar\omega_{max}/k_B T$ , where  $\omega_{max}$  is the maximal vibrational frequency of the system.  $P$  surges as  $\omega_{max}$  increases or  $T$  decreases. It is thus computational expensive to use PIMD to study a reaction in a sufficiently long timescale at an acceptable accuracy with convergence-ensured number of beads. To reduce the computational cost, some progresses have been achieved. Tuckerman et al combined a noncanonical transformation on the quadratic part of the action with multiple time scale integration techniques, and proposed a very efficient algorithm.<sup>29</sup> Markland and Manolopoulos proposed an approach to reduce the computational effort by separating the short-range interaction from the long-

range one.<sup>30</sup> Cheng et al proposed the multiple-timestep molecular dynamics (MTS-MD) algorithm to accelerate the propagation by dividing the force into slowly oscillating part and quickly varying component,<sup>31</sup> but the former factor is still the efficiency bottleneck. Large integration time step can also be made possible by applying proper transformations, so the analytic integration over the high-frequency modes is allowed.<sup>32</sup> After the transformation, the number of the beads can be reduced via ring-polymer contraction, in which some highest normal modes can be ignored.<sup>33-35</sup> Marsalek and Markland proposed a ring polymer contraction approach that can dramatically reduce the computational cost of PIMD at an *ab initio* level.<sup>36</sup> On the other hand, the accessible simulation time is further limited when it is necessary to use high-level electronic structure methods such as post-Hartree-Fock theories and density functional theory (DFT). Approximated levels of theory, such as semi-empirical (SE) methods, can improve computational efficiency and allow sufficient exploration in the phase space for systems in condensed phase. However, these methods are crude in the treatment of electronic structures, which may lead to inexact results for certain physical properties.

In addition, ergodicity in the configurational space is critical for the convergence of the simulated macrostate properties. Enhanced sampling methods, such as umbrella sampling (US),<sup>37</sup> adaptively biased molecular dynamics (ABMD),<sup>38</sup> and metadynamics,<sup>39,40</sup> are now frequently employed to boost one or two DoF and thereby accelerate the reaction. For the orthogonal DoF, the exploration can still be hindered by hidden barriers, and a ns-timescale simulation can be inevitable. Besides, length of the simulations determines the effective number of samples, because independent and identically distributed (i.i.d.) samples are required for the calculations of ensemble averages. For simulations in aqueous solutions, the energy correlation time is on a scale around 1 ps. If the samples are saved too frequently, sampling inefficiency must be considered.<sup>41</sup> Affordable studies are limited to relatively simple systems,<sup>42</sup> low levels of theory,<sup>43</sup> or with the aid of artificial intelligence.<sup>44</sup>

Therefore, it is desirable to develop a practical method for the study of thermodynamics properties based on PIMD that can dramatically reduce the computational expense while

maintain the accuracy. The multistate thermodynamic perturbation (MsTP) is the method of choice, of which the efficiency and reliability has been examined by Li et al. in an early study of some model reactions.<sup>45</sup> MsTP is a variant of the reference-potential method, in which the ensemble averages of any time-independent physical properties at a high level of theory are obtained indirectly from simulations at a lower level of theory. With further improvement,<sup>46–48</sup> this method is becoming more robust for the studies of chemical processes in the condensed phase.

In the current work, we combine PIMD with the reference potential method for the first time and demonstrate its application in the study of the intramolecular proton transfer in the protonated 1,8-bis(dimethylamino)naphthalene (DMANH) molecule. The free energy profile at the PIMD/BLYP-D3/6-31G(d)/MM level of theory along the proton transfer pathway was calculated leveraging a more efficient sampling at the PM6/MM level. A 545-fold enhancement in the computational efficiency was achieved as compared to a direct PIMD simulation at the same *ab initio* level of theory.

## Theory

The partition function of a system with  $N$  distinguishable particles is defined as,<sup>49</sup>

$$Z_P(\beta) = \mathcal{N} \int d^{NP} \mathbf{p} \int d^{NP} \mathbf{x} e^{-\beta \hat{H}_p(\mathbf{x}, \mathbf{p})}, \quad (1)$$

where  $\beta = 1/k_B T$ ,  $\mathcal{N}$  is the total normalization function,  $\mathbf{x}$  represents the position of particles, and  $\mathbf{p}$  the momenta conjugate to  $\mathbf{x}$ . The isomorphic primitive Hamiltonian operator  $\hat{H}_p(\mathbf{x}, \mathbf{p})$  of this  $N$ -atom  $P$ -bead cyclic model is

$$\hat{H}_p(\mathbf{x}, \mathbf{p}) = \sum_{i=1}^P \left\{ \sum_{n=1}^N \left[ \frac{\mathbf{p}_n^{i2}}{2\widetilde{m}_n^i} + \frac{1}{2} m_n \omega_p^2 (\mathbf{x}_n^{i+1} - \mathbf{x}_n^i)^2 \right] + \frac{1}{P} U(\mathbf{x}_1^i, \dots, \mathbf{x}_N^i) \right\}_{\mathbf{x}_n^{i+1} = \mathbf{x}_n^i}, \quad (2)$$

under the constraint  $\mathbf{x}_n^{i+1} = \mathbf{x}_n^i$ . This quantum Hamiltonian operator is made up of  $P$  group of beads (also known as replicas or time-slices), and each forms a classical representation of the original system. Beads of each atom are connected by harmonic potentials. In this way, a  $P$ -membered classical “necklace” model is constructed, and the simulation of quantum systems is transformed into an expanded classical calculation.  $m_n$  is the mass of the particle,  $U$  is the potential energy,  $\widetilde{m}_n^i$  and  $\mathbf{p}_n^i$  are the fictitious mass of the  $i$ th bead of the  $n$ th particle and its corresponding momentum. This fictitious mass can be set to an arbitrary value, because this term disappears in the thermodynamical statistical average. The  $\omega_p = \sqrt{P}/\beta\hbar$  is the harmonic constraint connecting the two most adjacent beads. For the ideal quantum problem, the full quantum Hamiltonian  $\hat{H}$  is established when  $P \rightarrow \infty$ . The introduction of this isomorphic relationship also means the quantum mechanical sampling can be mapped onto classical simulations by either Monte Carlo (MC)<sup>50,51</sup> or MD<sup>52</sup> methods. The potential of mean force with a collection of PIMD samples can thus be written as

$$F(\boldsymbol{\xi}_{PI}) = -\frac{1}{\beta} \ln \left\langle \frac{1}{P} \sum_{i=1}^P \delta(\boldsymbol{\xi}_{PI} - \boldsymbol{\xi}(\mathbf{x}_i)) \right\rangle. \quad (3)$$

To accelerate computation, we incorporate of the PIMD and the reference-potential method (RPM) for the first time to calculate the statistical quantum mechanical properties at a level-of-interest Hamiltonian while simulating at a lower level Hamiltonian state. Additionally, we adopt the US method for enhancing the sampling. From  $K$  simulations, we can collect  $N_k$  configurations from the  $k$ th simulation. Each simulation is characterized by its specific potential function  $U_k$ , for  $k \in 1 \dots, K$ . In the umbrella sampling method,

$$U_k(\mathbf{r}_n) = U_0(\mathbf{r}) + W_k(\boldsymbol{\eta}(\mathbf{r})) \quad (4)$$

where  $U_0(\mathbf{r})$  is the unbiased Hamiltonian and  $W_k(\mathbf{r})$  is the biasing potential acting on some chosen collective variables (CV)  $\boldsymbol{\eta}(\mathbf{r})$ . The unnormalized weight of each configuration under

another Hamiltonian  $U_t$  is

$$\omega_t(\mathbf{r}_n) = \frac{e^{-\beta U_t(\mathbf{r}_n)}}{\sum_{k=1}^K N_k e^{-\beta[U_k(\mathbf{r}_n) - f_k]}} = \frac{e^{-\beta \Delta U_t(\mathbf{r}_n)}}{\sum_{k=1}^K N_k e^{-\beta[W_k(\boldsymbol{\eta}(\mathbf{r}_n)) - f_k]}} \quad (5)$$

in which  $f_k$  is the free energy of state  $k$  and is calculated by iteratively solving the Multistate Bennett Acceptance Ratio (MBAR)<sup>53</sup> equations

$$f_i = -\beta^{-1} \ln \sum_{n=1}^N \frac{e^{-\beta U_i(\mathbf{r}_n)}}{\sum_{k=1}^K N_k e^{-\beta[U_k(\mathbf{r}_n) - f_k]}} \quad \forall i = 1, \dots, K. \quad (6)$$

Here,  $N = \sum_{k=1}^K N_k$  is total number of snapshots collected in all the simulations, and  $\Delta U_t(\mathbf{r}) = U_t(\mathbf{r}) - U_0(\mathbf{r})$ .  $\omega_t(\mathbf{r}_n)$  can be understood as the weighting factor for each configuration in a thermodynamic perturbation from the mixed reference potentials ( $U_k$ ) to the target Hamiltonian ( $U_t$ ). Therefore, this method is referred to as the multistate thermodynamic perturbation (MsTP). Thermodynamic properties that depend only on atomic coordinates under  $U_t$  can thus be obtained as

$$\langle \hat{\mathbf{A}} \rangle_t = \frac{\sum_{n=1}^N \omega_t(\mathbf{r}_n) \hat{\mathbf{A}}(\mathbf{r}_n)}{\sum_{n=1}^N \omega_t(\mathbf{r}_n)}. \quad (7)$$

If  $\mathbf{A}$  is an indication function  $\delta$  of some chosen CV  $\boldsymbol{\xi}(\mathbf{r})$

$$\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r})) = \begin{cases} 1, & \text{if } -\Delta\boldsymbol{\xi}/2 < \boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}) < \Delta\boldsymbol{\xi}/2 \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

the potential of mean force (PMF) under  $U_t(\mathbf{r})$  can be written as

$$F_t(\boldsymbol{\xi}_m) = -\beta^{-1} \ln \sum_{n=1}^N \omega_t(\mathbf{r}_n) \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) \quad (9)$$

defined up to an additive constant. Similarly, the PMF under  $U_0(\mathbf{r})$  can be written as

$$F_0(\boldsymbol{\xi}_m) = -\beta^{-1} \ln \sum_{n=1}^N \omega_0(\mathbf{r}_n) \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)), \quad (10)$$

where

$$\omega_0(\mathbf{r}_n) = \frac{e^{-\beta U_0(\mathbf{r}_n)}}{\sum_{k=1}^K N_k e^{-\beta [U_k(\mathbf{r}_n) - f_k]}} = \frac{1}{\sum_{k=1}^K N_k e^{-\beta [W_k(\boldsymbol{\eta}(\mathbf{r}_n)) - f_k]}} \quad (11)$$

is the unnormalized weight for configuration  $\mathbf{r}_n$  under  $U_0(\mathbf{r})$ . To ameliorate the numerical instability originated from incomplete sampling, the weights  $\omega_t(\mathbf{r}_n)$  under the target Hamiltonian are scaled by Gaussian-smoothing on the density-of-states.<sup>47</sup>

In this work, we take  $U_0$  as the PM6/MM<sup>54</sup> Hamiltonian, and  $U_t$  as the BLYP-D3/6-31G(d)/MM<sup>55-57</sup> level of theory. For simplicity, they will be labeled as the SQM and DFT level of theories. Please also note that  $\boldsymbol{\eta}$  is unnecessarily  $\boldsymbol{\xi}$ , but we set them equal in this work.

## Simulations

One DMANH molecule, shown in Fig. 1,<sup>58,59</sup> was solvated in a TIP3P water<sup>60</sup> sphere with a radius of 25 Å using the LEaP module in the AmberTools19 package.<sup>61</sup> One chlorine ion was added for neutralization. The QM region contained only the DMANH molecule, and the MM region comprised all the solvent molecules and the counter ion. In the MM region, the SHAKE algorithm<sup>62</sup> was applied to constrain all the bonds involving hydrogen atoms. The system was energy-minimized for 2000 steps and then heated up to 298 K in 100 ps. Finally, a 100-ps classical MD simulation was conducted for further relaxation of the system. Temperature was regulated using the Langevin thermostat with a collision frequency of 2.0 ps<sup>-1</sup>.<sup>63</sup>

The CV was defined as the difference in the N–H bond lengths, i.e.  $d_{N_a\text{H}} - d_{N_d\text{H}}$ , where  $d_{N_a\text{H}}$  is the distance between the proton  $\overset{\oplus}{\text{H}}$  and the acceptor nitrogen atom  $N_a$ , and  $d_{N_d\text{H}}$

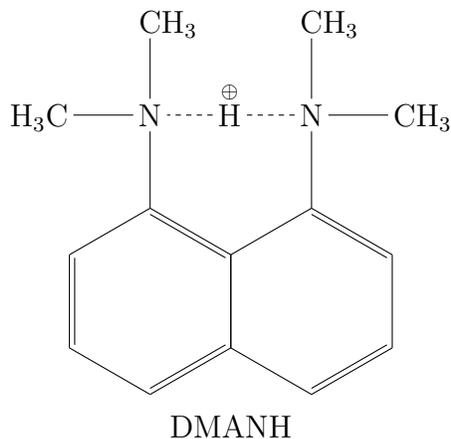


Figure 1: Molecular structure of DMANH

the distance between the proton  $\overset{\oplus}{\text{H}}$  and the donor nitrogen atom  $\text{N}_d$ . The CV, ranged from  $-0.15 \text{ \AA}$  to  $1.10 \text{ \AA}$ , was discretized into 26 windows in total. Harmonic potentials  $W_k(\mathbf{x}) = \frac{1}{2}k_\xi(\boldsymbol{\xi} - \boldsymbol{\xi}_k)^2$  were used with  $k_\xi = 100 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  for all the windows. We first performed the primitive approximation PIMD (PRIMPIMD)<sup>50</sup> with 16 beads for each QM atom for 1-ns per window, and then an additional 1-ns simulation was performed for each of the first 16 windows (with RC between  $-0.15 \text{ \AA}$  and  $0.60 \text{ \AA}$ ) to improve the convergence in this region. The temperature was regulated using the “middle scheme” Langevin thermostat.<sup>64,65</sup> The step size was set to 0.5 fs. Configurations were saved every 1000 steps (500 fs) for single point energy calculations at the DFT level, as well as the SQM level. Due to the symmetry of the molecule, we only calculated the PMF along the CV with positive values. The potential of mean force at the SQM level was generated via the MBAR analysis, and that at the DFT level was calculated with the MsTP and was smoothed using the Gaussian process regression.<sup>66</sup>

## Results and Discussion

Reweighting entropy (RE), is a measure of the reliability of the RPM. We have previously shown that when RE is greater than 0.3 in the MsTP calculations, the extrapolation process is reliable.<sup>45</sup> When it is below 0.3, large biases may reside in the free energy profile. As shown

in Fig. 2, the free energy profile under the DFT level is reliable when the CV is smaller than 0.8 Å. When the CV is greater than 0.8 Å, RE plunges. This is the consequence of the large difference in the preferred structures under these two Hamiltonians, i.e. SQM and DFT. Shown in Fig. 3 are the means and fluctuations of the local structures for the proton transfer reaction. With the CV increases from 0.00 Å to 0.80 Å,  $d_{N_dH}$  decreases from around 1.34 Å to about 1.05 Å, and  $d_{N_aH}$  increases from 1.34 Å to 1.85 Å. These distance variations are very similar to those under the SQM Hamiltonian as shown in Fig. S1. However,  $d_{N_dH}$  shows a sudden drop when the CV is greater than 0.8 Å, and deviates from the mean bond length under the SQM Hamiltonian, indicating an insufficient overlap in the sub-phase space under these two Hamiltonians. The difference is more obvious for the  $N_d-H-N_a$  angle. Therefore, thermodynamic properties with CV greater than 0.8 Å have not reached convergence, and more samples are required in this region. For the PMF analysis at the DFT level in the following, we can only trust the results for the region with CV below 0.8 Å.

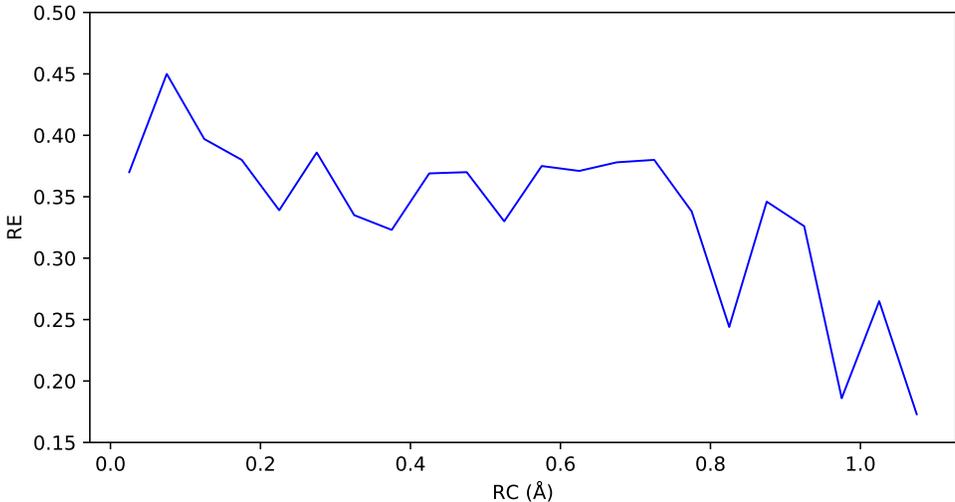
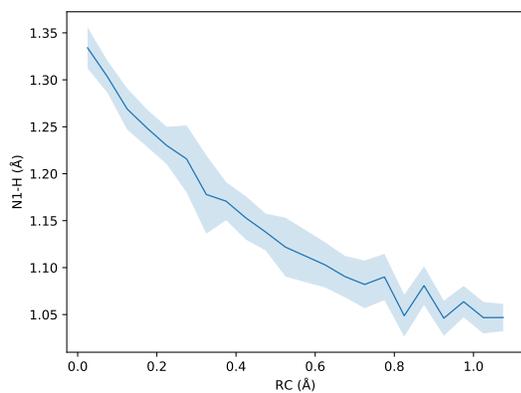
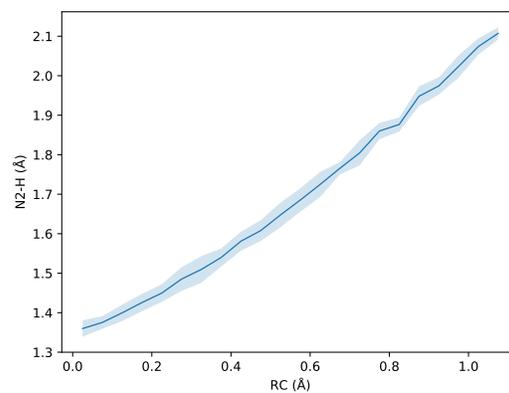


Figure 2: Reweighting entropy

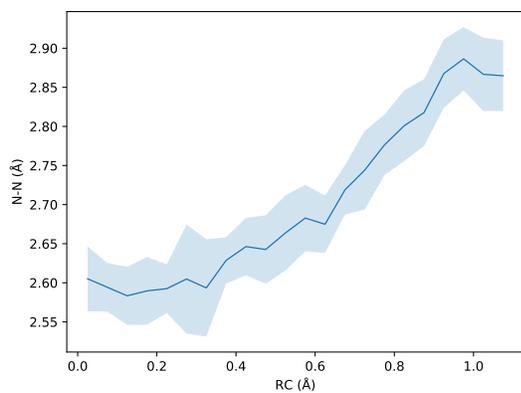
The free energy profiles under the SQM and DFT levels are shown in Fig. 4. Under the SQM level, the uncertainties are much smaller than their counterparts under the extrapolated DFT level. The large magnitude in the uncertainties under the DFT level is the consequence



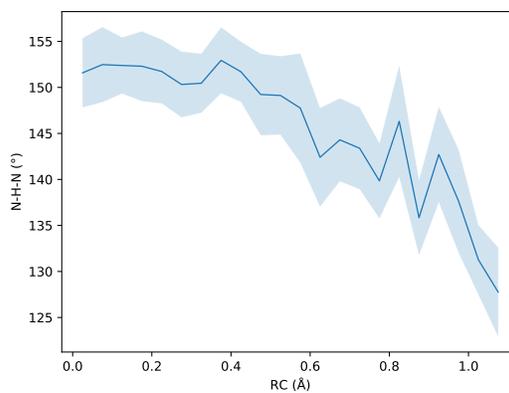
(A)  $N_d$ -H distance



(B)  $N_a$ -H distance



(C)  $N_d$ - $N_a$  distance



(D)  $N_d$ -H- $N_a$  angle

Figure 3: Structural alternations during the proton transfer process at the DFT/MM level from MsTP. (A) Distance between the proton (H) and the  $N_d$  atom. (B) Distance between the proton (H) and the  $N_a$ . (C) Distance between the  $N_d$  and the  $N_a$  atoms. (D) Angle of  $N_d$ -H- $N_a$ . The shaded areas are the standard error.

of the sampling inefficiency when the MBAR weights are assigned to the samples. As shown in Eq. 5, the weight decays exponentially as the energy difference between the two Hamiltonians increases. Precisely,  $\Delta U_t$  is the difference in the deviations of the potential energies from their respective means, since the difference in the means can be canceled in normalization. Under the SQM level, the minimum in the free energy profile locates at  $CV \approx 0.6 \text{ \AA}$ , which is about  $0.1 \text{ \AA}$  larger than the minimum under the DFT level. This indicates that in the optimal structure under the SQM level, the proton prefers to be closer to one of the nitrogen atoms, while under the DFT level the proton is more diffusive. This can be explained by the difference in the basis sets used in the SQM and DFT calculations. PM6 used the minimum basis set, while a larger basis set 6-31G(d) was used in the DFT calculations. Both of the profiles show a reaction free energy barrier of about 1.2 kcal/mol. This picture is different from the results in a recent study by Zhou and Wang, in which the reaction is barrierless.<sup>67</sup> The disagreement comes from the difference in the Hamiltonians. In their studies, all the atoms were quantum particles, while in the current study only the DMANH molecule was treated quantum mechanically. Quantum-mechanically described water molecules can stabilize the delocalized proton, and thereby lowers the reaction barrier. In the current work, we are not aiming at reproducing the results of a full-QM PIMD model. However, the central idea developed in the current work can be applied to a full-QM PIMD model, if a proper reference Hamiltonian can be found. This is beyond the interest of the current work. For the CV to the right side of the minimum, the potential of mean force increases very fast w.r.t the CV under the DFT level than under the SQM level, which indicates a higher probability of a shared proton between the nitrogen atoms under the DFT level.

The estimated CPU times of obtaining PMF at the BLYP-D3 level with direct QM/MM and RPM are shown in Table 1. The CPU time of the RPM includes two parts, i.e. the CPU time for collecting the trajectories at the PM6/MM level and the CPU time for the single point energy calculations at both the PM6/MM and BLYP-D3/MM level. While the direct QM/MM method at the BLYP-D3/MM level only requires the time for the simulations.

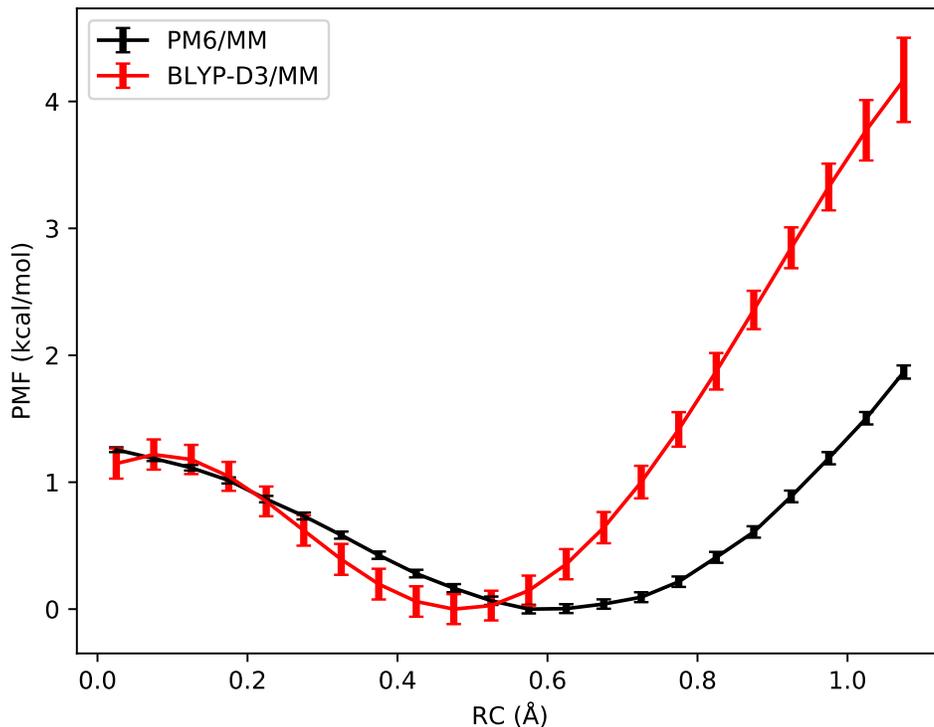


Figure 4: Free energy profiles of the proton transfer under different levels of theory

Since the PM6/MM level is several orders of magnitude more efficient than BLYP-D3/MM level and the configurations were saved once every 1000 steps of MD propagation for the single point energy calculations at the DFT level of theory, a 545-fold enhancement in the efficiency was observed for the RPM method as compared to a direct PIMD simulations at the same level of theory.

**Table 1: Estimated CPU Time in a unit of Days. Assuming one node with 28 cores of Intel Xeon Gold 6132 CPU 2.60 GHz was used.**

RP method			Direct DFT <sup>a</sup>
Sampling	Energy Evaluation	Total	
1.04 K	1.25 K	2.29 K	1250 K

<sup>a</sup> Estimated CPU time assuming the number of windows and simulation length for each window

## Conclusion

The calculations of thermodynamic properties from computer simulations require independent and identically distributed samples. The long correlation time of chemical processes in condensed phases plagues the simulations with slow convergence in thermodynamic properties and poses a challenge for the computational studies at ab initio levels of theory. This difficulty is more severe for PIMD, in which the representation of quantum effect in single-particle systems is transformed into a classical problem of discrete cyclic model systems with a number of beads for each atom. The reference-potential methods can be a remedy for ameliorating the large expense by leveraging an inexpensive sampling method and a post-processing process on the harvested samples. In this work, the proton transfer within a protonated 1,8-bis(dimethylamino)naphthalene molecule was studied, and the free energy profile at the BLYP-D3/6-31G(d)/MM level of theory was calculated using a more efficient sampling at the PM6/MM level. With a finite number of samples, the reliability of this reference-potential method heavily depends on the overlap in the phase space between the target (BLYP-D3/6-31G(d)/MM) Hamiltonian and the reference (PM6/MM) Hamiltonian. The results show that when the CV, which is defined as the difference in the distances from the proton to the acceptor and to the donor, is below 0.8 Å, the potential of mean force and the structural properties can be reliably produced with a 545-fold enhancement in the efficiency.

## Acknowledgement

Y. Mei owes Dr. Jian Liu many thanks for the helpful discussions. Y. Mei is supported by the National Natural Science Foundation of China (Grant No. 22073030). W.H. is supported by the Fundamental Research Funds for the Central Universities. Y. Mo is supported by the National Natural Science Foundation of China (Grant No. 21973030). Y.L. is supported by the National Natural Science Foundation of China (Grant No. 11674212). Y.S. is supported

by the National Institutes of Health (Grant No. R01GM135392). CPU time was supported by the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation No. 001).

## References

- (1) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (2) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (3) Monard, G.; Merz, K. M. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999**, *32*, 904–911.
- (4) Tresadern, G.; Nunez, S.; Faulder, P. F.; Wang, H.; Hillier, I. H.; Burton, N. A. Direct Dynamics Calculations of Reaction Rate and Kinetic Isotope Effects in Enzyme Catalysed Reactions. *Faraday Discuss.* **2003**, *122*, 223–242.
- (5) Friesner, R. A.; Guallar, V. *Ab Initio* Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.
- (6) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3188–3209.
- (7) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2006**, *117*, 185.

- (8) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with *Ab Initio* Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.
- (9) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (10) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.
- (11) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.
- (12) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004**, *303*, 186–195.
- (13) Fernanda, D.; Amrein, B. A.; David, B. N.; Kamerlin, S. C. L. Recent Advances in QM/MM Free Energy Calculations using Reference Potentials. *Biochim. Biophys. Acta.* **2015**, *1850*, 954–965.
- (14) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM Simulations: What Are Some of the “Burning Issues”? *J. Phys. Chem. B* **2021**, *125*, 689–702.
- (15) Tuckerman, M. E. *Ab Initio* Molecular Dynamics: Basic Concepts, Current Trends and Novel Applications. *J. Phys. Condens. Matter* **2002**, *14*, R1297–R1355.
- (16) Marx, D.; Chandra, A.; Tuckerman, M. E. Aqueous Basic Solutions: Hydroxide Solvation, Structural Diffusion, and Comparison to the Hydrated Proton. *Chem. Rev.* **2010**, *110*, 2174–2216.

- (17) Sumner, I.; Iyengar, S. S. Combining Quantum Wavepacket *Ab Initio* Molecular Dynamics with QM/MM and QM/QM Techniques: Implementation Blending ONIOM and Empirical Valence Bond Theory. *J. Chem. Phys.* **2008**, *129*, 054109.
- (18) Marx, D. Proton Transfer 200 Years after von Grothuss: Insights from *Ab Initio* Simulations. *Chem. Phys. Chem.* **2010**, *8*, 1848–1870.
- (19) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges. *Chem. Rev.* **2016**, *116*, 7529–7550.
- (20) Markland, T. E.; Ceriotti, M. Nuclear Quantum Effects Enter the Mainstream. *Nat. Rev. Chem.* **2018**, *2*, 0109.
- (21) Dammak, H.; Hayoun, M.; Briec, F.; Geneste, G. Nuclear Quantum Effects in Molecular Dynamics Simulations. *J. Phys. Conf. Ser.* **2018**, *1136*, 012014.
- (22) Berger, A.; Ciardi, G.; Sidler, D.; Hamm, P.; Shalit, A. Impact of Nuclear Quantum Effects on the Structural Inhomogeneity of Liquid Water. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 2458–2463.
- (23) Feynman, R. P. Atomic Theory of the  $\lambda$  Transition in Helium. *Phys. Rev.* **1953**, *91*, 1291–1301.
- (24) Chandler, D.; Wolynes, P. G. Exploiting the Isomorphism between Quantum Theory and Classical Statistical Mechanics of Polyatomic Fluids. *J. Chem. Phys.* **1981**, *74*, 4078–4095.
- (25) Craig, I. R.; Manolopoulos, D. E. Quantum Statistics and Classical Mechanics: Real Time Correlation Functions from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2004**, *121*, 3368–3373.

- (26) Braams, B. J.; Manolopoulos, D. E. On the Short-Time Limit of Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, 124105.
- (27) Cao, J.; Voth, G. A. The Formulation of Quantum Statistical Mechanics Based on the Feynman Path Centroid Density. I. Equilibrium Properties. *J. Chem. Phys.* **1994**, *100*, 5093–5105.
- (28) Jang, S.; Voth, G. A. A Derivation of Centroid Molecular Dynamics and Other Approximate Time Evolution Methods for Path Integral Centroid Variables. *J. Chem. Phys.* **1999**, *111*, 2371–2384.
- (29) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J.; Klein, M. L. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *J. Chem. Phys.* **1993**, *99*, 2796–2808.
- (30) Markland, T. E.; Manolopoulos, D. E. An Efficient Ring Polymer Contraction Scheme for Imaginary Time Path Integral Simulations. *J. Chem. Phys.* **2008**, *129*, 024105.
- (31) Cheng, X.; Herr, J. D.; Steele, R. P. Accelerating *Ab Initio* Path Integral Simulations via Imaginary Multiple-Timestepping. *J. Chem. Theory Comput.* **2016**, *12*, 1627–1638.
- (32) Ceriotti, M.; Parrinello, M.; Markland, T. E.; Manolopoulos, D. E. Efficient Stochastic Thermostatting of Path Integral Molecular Dynamics. *J. Chem. Phys.* **2010**, *133*, 124104.
- (33) Markland, T. E.; Manolopoulos, D. E. A Refined Ring Polymer Contraction Scheme for Systems with Electrostatic Interactions. *Chem. Phys. Lett.* **2008**, *464*, 256–261.
- (34) Kapil, V.; VandeVondele, J.; Ceriotti, M. Accurate Molecular Dynamics and Nuclear Quantum Effects at Low Cost by Multiple Steps in Real and Imaginary Time: Using Density Functional Theory to Accelerate Wavefunction Methods. *J. Chem. Phys.* **2016**, *144*, 054111.

- (35) John, C.; Spura, T.; Habershon, S.; Kühne, T. D. Quantum Ring-polymer Contraction Method: Including Nuclear Quantum Effects at No Additional Computational Cost in Comparison to *Ab Initio* Molecular Dynamics. *Phys. Rev. E* **2016**, *93*, 043305.
- (36) Marsalek, O.; Markland, T. E. *Ab Initio* Molecular Dynamics with Nuclear Quantum Effects at Classical Cost: Ring Polymer Contraction for Density Functional Theory. *J. Chem. Phys.* **2016**, *144*, 054112.
- (37) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (38) Babin, V.; Roland, C.; Sagui, C. Adaptively Biased Molecular Dynamics for Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, 134101.
- (39) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (40) Quhe, R.; Nava, M.; Tiwary, P.; Parrinello, M. Path Integral Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1383–1388.
- (41) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 26–41.
- (42) Marx, D.; Tuckerman, M.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601–604.
- (43) Kosugi, K.; Nakano, H.; Sato, H. SCC-DFTB-PIMD Method to Evaluate a Multidimensional Quantum Free-Energy Surface for a Proton-Transfer Reaction. *J. Chem. Theory Comput.* **2019**, *15*, 4965–4973.
- (44) Cendagorta, J. R.; Shen, H.; Bačić, Z.; Tuckerman, M. E. Enhanced Sampling Path

- Integral Methods Using Neural Network Potential Energy Surfaces with Application to Diffusion in Hydrogen Hydrates. *Adv. Theory Simul.* **2021**, *4*, 2000258.
- (45) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 5583–5596.
- (46) Pan, X.; Li, P.; Ho, J.; Pu, J.; Mei, Y.; Shao, Y. Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanical Accuracy via a Semi-empirical Reference Potential. II. Recalibrating Semi-empirical Parameters with Force Matching. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20595–20605.
- (47) Hu, W.; Li, P.; Wang, J.-N.; Xue, Y.; Mo, Y.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 3. Gaussian Smoothing on Density-of-States. *J. Chem. Theory Comput.* **2020**, *16*, 6814–6822.
- (48) Wang, J.-N.; Liu, W.; Li, P.; Mo, Y.; Hu, W.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 4. Adaptive QM/MM. *J. Chem. Theory Comput.* **2021**, *17*, 1318–1325.
- (49) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, 1965.
- (50) Ceperley, D. M. Path Integrals in the Theory of Condensed Helium. *Rev. Mod. Phys.* **1995**, *67*, 279–355.
- (51) Barker, J. A. A Quantum-statistical Monte Carlo Method: Path Integrals with Boundary Conditions. *J. Chem. Phys.* **1979**, *70*, 2914–2918.

- (52) Parrinello, M.; Rahman, A. Study of an F Center in Molten KCl. *J. Chem. Phys.* **1984**, *80*, 860–867.
- (53) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- (54) Stewart, J. J. P. Optimization of Parameters for Semi-Empirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (55) Becke, A. D. Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (56) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (57) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (58) White, P. B.; Hong, M. <sup>15</sup>N and <sup>1</sup>H Solid-State NMR Investigation of a Canonical Low-Barrier Hydrogen-Bond Compound: 1,8-Bis(dimethylamino)naphthalene. *J. Phys. Chem. B* **2015**, *119*, 11581–11589.
- (59) Wozniak, K.; He, H.; Klinowski, J.; Barr, T. L.; Hardcastle, S. E. ESCA, Solid-State NMR, and X-Ray Diffraction Studies of Perisubstituted Naphthalene Derivatives. *J. Phys. Chem. B* **1996**, *100*, 11408–11419.
- (60) Jorgensen, W. L.; Chandross, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Chem. Phys.* **1983**, *79*, 926–935.

- (61) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 19, University of California, San Francisco. 2019.
- (62) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327 – 341.
- (63) Langevin, P. Sur la Théorie du Mouvement Brownien. *C. R. Acad. Sci. (Paris)* **1908**, *146*, 530–533.
- (64) Zhang, Z.; Liu, X.; Chen, Z.; Zheng, H.; Yan, K.; Liu, J. A Unified Thermostat Scheme for Efficient Configurational Sampling for Classical/Quantum Canonical Ensembles via Molecular Dynamics. *J. Chem. Phys.* **2017**, *147*, 034109.
- (65) Liu, J.; Li, D.; Liu, X. A Simple and Accurate Algorithm for Path Integral Molecular Dynamics with the Langevin Thermostat. *J. Chem. Phys.* **2016**, *145*, 024103.
- (66) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press, 2006; pp 7–32.
- (67) Zhou, S.; Wang, L. Symmetry and <sup>1</sup>H NMR Chemical Shifts of Short Hydrogen Bonds: Impact of Electronic and Nuclear Quantum Effects. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4884–4895.

# Supporting Information

## Affordable *Ab Initio* Path Integral for Thermodynamic Properties via Molecular Dynamics Simulations Using Semiempirical Reference Potential

Yuanfei Xue,<sup>†</sup> Jia-Ning Wang,<sup>†</sup> Wenxin Hu,<sup>‡</sup> Jun Zheng,<sup>‡</sup> Yongle Li,<sup>¶</sup> Xiaoliang  
Pan,<sup>§</sup> Yan Mo,<sup>\*,†,||,⊥</sup> Yihan Shao,<sup>#</sup> Lu Wang,<sup>\*,@</sup> and Ye Mei<sup>\*,†,||,⊥</sup>

<sup>†</sup>*State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China  
Normal University, Shanghai 200062, China*

<sup>‡</sup>*The Computer Center, School of Data Science & Engineering, East China Normal University, Shanghai  
200062, China*

<sup>¶</sup>*Department of Physics, International Center of Quantum and Molecular Structure, and Shanghai Key  
Laboratory of High Temperature Superconductors, Shanghai University, Shanghai 200444, China*

<sup>§</sup>*Department of Chemistry and Biochemistry, University of Oklahoma, Norman Oklahoma 73019, United  
States*

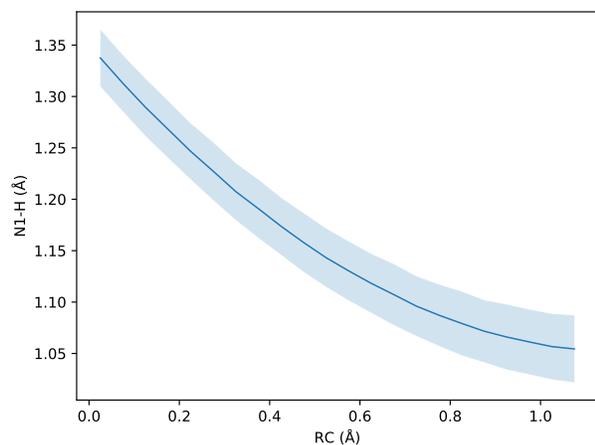
<sup>||</sup>*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

<sup>⊥</sup>*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

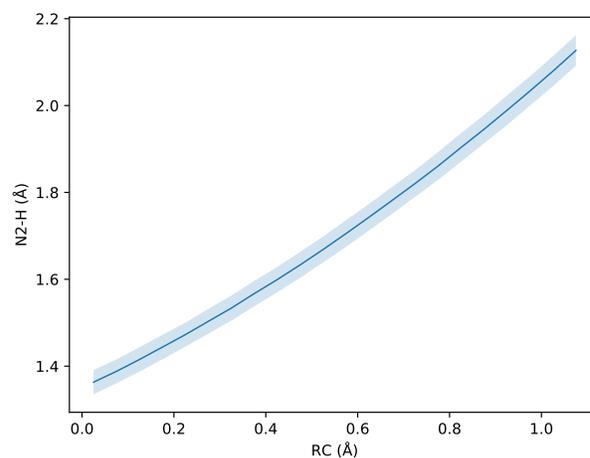
<sup>#</sup>*Department of Chemistry and Biochemistry, University of Oklahoma, Norman Oklahoma 73019, United  
States of America*

<sup>@</sup>*Department of Chemistry and Chemical Biology, Institute for Quantitative Biomedicine, Rutgers  
University, Piscataway, New Jersey 08854, United States*

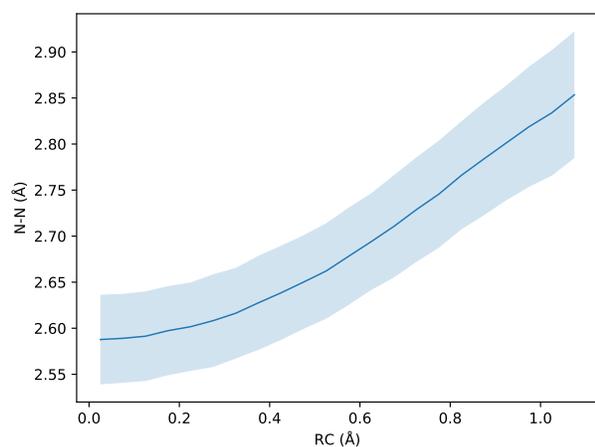
E-mail: ymo@phy.ecnu.edu.cn; lwang@chem.rutgers.edu; samuel.y.mei@gmail.com



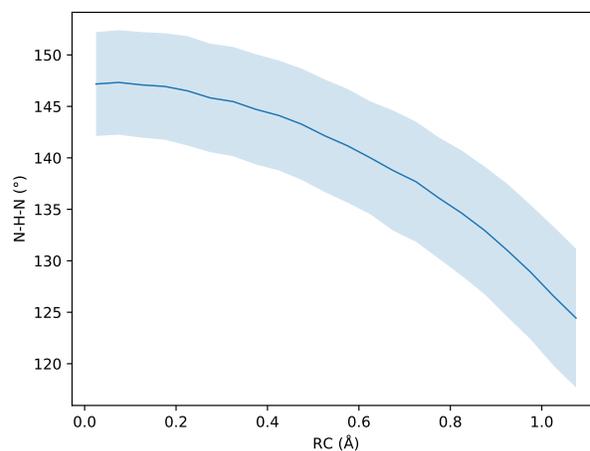
(A)  $N_d$ -H distance



(B)  $N_a$ -H distance



(C)  $N_d$ - $N_a$  distance



(D)  $N_d$ -H- $N_a$  angle

Figure S1: Structural alternations during the reaction at the PM6/MM level. A) Distance between the transferred proton (H) and the donor nitrogen ( $N_d$ ) atom. B) Distance between the transferred proton (H) and the acceptor nitrogen atom ( $N_a$ ). C) Distance between the donor ( $N_d$ ) and the acceptor nitrogen ( $N_a$ ) atoms. D) Angle of  $N_d$ -H- $N_a$ .

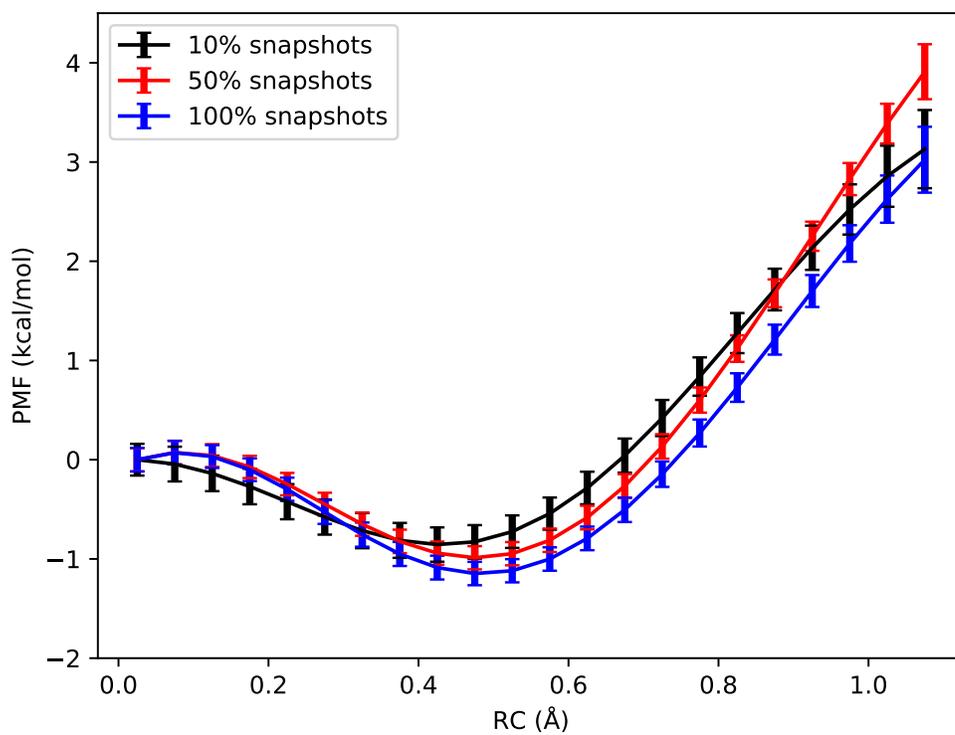


Figure S2: Free energy profiles with Gaussian process regression at the BLYP-D3 level using 200 snapshots, 1000 snapshots and 2000 snapshots from each window simulation.

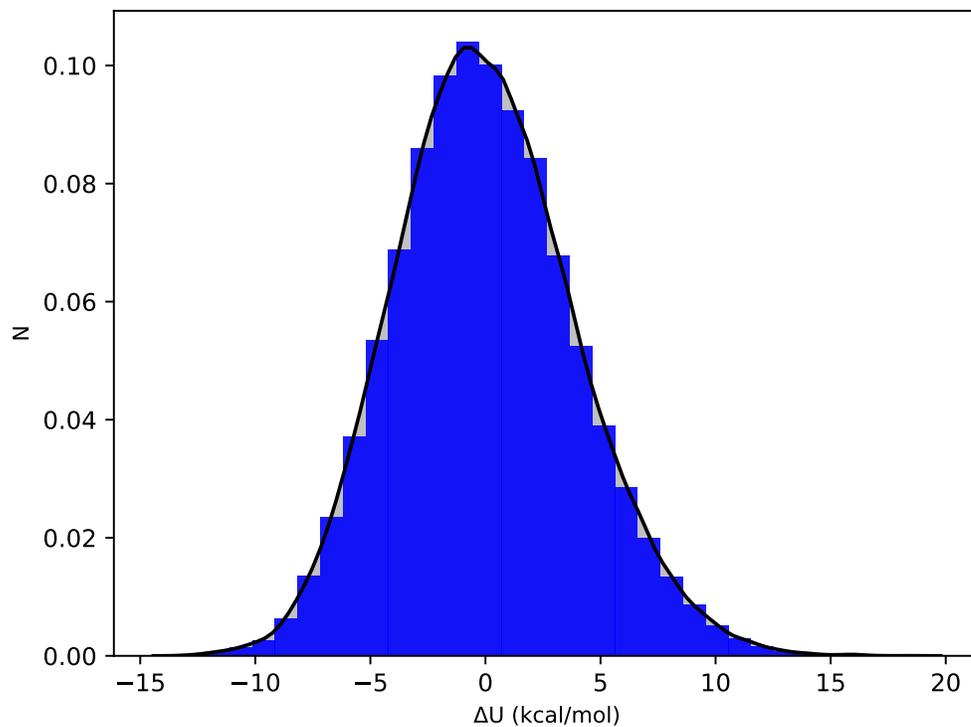


Figure S3: Distribution of  $\Delta U$