Article

Efficient Computation of Free Energy Surfaces of Diels–Alder Reactions in Explicit Solvent at Ab Initio QM/MM Level

Pengfei Li [1], **Fengjiao Liu [1]**, **Xiangyu Jia [1,†]**, **Yihan Shao [2]**, **Wenxin Hu [3,*]**, **Jun Zheng [3]** and **Ye Mei [1,2,4,*]**

1   State Key Laboratory of Precision Spectroscopy, School of Physics and Materials Science,
     East China Normal University, Shanghai 200062, China; lipengfei_mail@126.com (P.L.);
     fjliu208@gmail.com (F.L.); xj6@nyu.edu (X.J.)
2   Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019, USA;
     yihan.shao@ou.edu
3   The Computer Center, School of Computer Science and Software Engineering,
     East China Normal University, Shanghai 200062, China; jzheng@cc.ecnu.edu.cn
4   NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China
*   Correspondence: wxhu@cc.ecnu.edu.cn (W.H.); ymei@phy.ecnu.edu.cn (Y.M.)
†   Current address: NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China.

check for updates

Abstract: For Diels–Alder (DA) reactions in solution, an accurate and converged free energy (FE) surface at ab initio (*ai*) quantum mechanical/molecular mechanical (QM/MM) level is imperative for the understanding of reaction mechanism. However, this computation is still far too expensive. In a previous work, we proposed a new method termed MBAR+wTP, with which the computation of the *ai* FE profile can be accelerated by several orders of magnitude via a three-step procedure: (I) an umbrella sampling (US) using a semi-empirical (SE) QM/MM Hamiltonian is performed; (II) the FE profile is generated using the Multistate Bennett Acceptance Ratio (MBAR) analysis; and (III) a weighted Thermodynamic Perturbation (wTP) from the SE Hamiltonian to the *ai* Hamiltonian is performed to obtain the ai QM/MM FE profile using weight factors from the MBAR analysis. In this work, this method is extended to the calculations of two-dimensional FE surfaces of two Diels–Alder reactions of cyclopentadiene with either acrylonitrile or 1-4-naphthoquinone at *ai* QM/MM level. The accurate activation free energies at the *ai* QM/MM level, which are much closer to the experimental measurements than those calculated by other methods, indicate that this MBAR+wTP method can be applied in the studies of complex reactions in condensed phase with much-enhanced efficiency.

Keywords: Diels–Alder reaction; free energy surface; ab initio; reference-potential method; umbrella sampling

1. Introduction

The Diels–Alder (DA) reactions are one category of organic chemical reactions (specifically, a [4+2] cycloaddition) between a conjugated diene and a dienophile, which involve a dual carbon–carbon bond-forming process. Among them, the DA reactions of cyclopentadiene (CP) with acrylonitrile (ACR) or 1-4-naphthoquinone (NAP) have attracted much attention from experimental [1–3] and

computational scientists [4–8]. It has been observed that the reaction rate is very sensitive to the solvent [2,3,5–8]. Therefore, the solvation effect should be explicitly considered in order to unveil the reaction mechanism, in which the polarization effect and the reorganization of the solvent molecules may play a critical role.

Hybrid quantum mechanical/molecular mechanical (QM/MM) method, which was proposed by Warshel and Levitt in 1976 [9], is now a mature method that can be used to study chemical reactions taking place in condensed phase such as aqueous solution or enzymatic environment [10–13]. In this approach, only reactive region is treated quantum mechanically, and the remaining part is described by molecular mechanical force field. However, when the reaction barrier is much larger than $k_B T$, which is pervasive for reactions under mild condition, a direct ab initio (*ai*) QM/MM simulation is still notoriously time consuming, if feasible at all. Despite the continuous development in computer technology and enhanced sampling methods, investigation of reaction mechanism in solution or enzymes using direct *ai* QM/MM simulations is still a daunting task.

To reduce the computational expense, Jorgenson et al. proposed a new approach for the calculations of free energy changes in chemical reactions in solution by combining gas phase QM calculations with free energy simulations (QM-FE method) [14–16]. Kollman et al. extended this method to the studies of enzymatic reactions [17]. However, in the QM-FE approach, the QM Hamiltonian and the MM Hamiltonian are separated, which is not a rigorous QM/MM approach. In other words, the impact of the solvent or the enzyme environment on the electronic structure of the QM region is not considered. However, solvent or enzyme environment often has a remarkable impact on the reaction process. Zhang and Yang [18] were motivated by the QM-FE approach and developed a more practical method, which combines *ai* QM/MM calculations with free energy perturbation (FEP) [19,20]. Later, Thiel et al. named this method "QM/MM-FEP" [21]. In this method, an efficient iterative optimization procedure was developed to determine the optimized structures and the minimum energy paths for a large-sized system on an *ai* QM/MM potential energy surface. However, this still requires arduous computations.

To further reduce the computational expense, Jorgensen et al. used semi-empirical methods such as AM1 and PDDG/PM3 in the free energy calculations of these DA reactions [6–8], followed by some high-level correction to the stationary points in the reaction path including the reactant, the product and the transition state (TS) [7]. They found that the solvent-sensitivity originated from a significant nonhydrophobic component stemming from enhanced polarization of the transition state, which leads to strengthened hydrogen bonds [6,7,22,23]. They also found that the DA reaction between CP and ACR was an asynchronous and concerted process [4], while that between CP and NAP has a synchronous and concerted feature [6,7]. This observation was evidenced in a two-dimensional free energy landscape for these DA reactions, and a one-dimensional reaction coordinate for these reactions would lead to potential artifacts and uncertainty in the locations of transition states, which in the end leads to an ambiguous reaction mechanism [6,7]. Although semi-empirical (SE) QM/MM calculations have been widely used [12,24–29], unfortunately, these semi-empirical methods may lead to large errors in the results due to the approximations adopted. Thus, high-level quantum mechanical methods at ab initio levels are necessary for a reliable depiction of electron redistribution during the reaction, which can be critical to the energetics such as activation and reaction free energies.

Based on the idea of probability-reweighting, Gao developed a method termed the dual-Hamiltonian method, also known as the reference-potential method, and applied it in a study of hydration free energy [30,31]. Utilizing an empirical valence bond (EVB) method [32,33] as the reference-potential, Warshel et al. developed a dual-Hamiltonian approach for calculating the free energy (FE) profiles of chemical or enzymatic reactions, from which an *ai* FE profile can be obtained with much lower computational expense than a direct approach [26,34–36]. Rod and Ryde calculated the activation free energy of a methyl transfer reaction in enzyme using the dual-Hamiltonian approach where the free energy was found to be overestimated [37]. Recently, Jia et al. used the dual-Hamiltonian approach to calculate the solvation free energies of the molecules in the SAMPL4 competition by an

alchemical decoupling method, which yielded the globally minimal variance for the QM/MM free energies [38]. Liu et al. used this dual-Hamiltonian approach to calculate protein–ligand binding affinity at an *ai* QM/MM level [39].

In our previous work [40], in the spirit of reference-potential method, a new method termed MBAR+wTP was proposed to obtain the *ai* QM/MM FE profiles with much less computational expense. In this method, a weighted thermodynamic perturbation (wTP) [19] correction is applied to the semi-empirical profile, which is generated by the Multistate Bennett Acceptance Ratio (MBAR) [41,42] analysis of the trajectories from umbrella sampling (US) [43]. The raw *ai* QM/MM FE profile was then smoothed via Gaussian process regression [44]. This MBAR+wTP method had been validated by calculating the FE profiles of one quasi-chemical reaction and three chemical reactions in aqueous solution. The results showed that even if the SE FE profiles deviated from the *ai* ones by several kcal/mol in terms of activation free energy and reaction free energy, after the SE-to-*ai* correction the FE profiles agree much better with the direct QM/MM simulated ones with errors below 1 kcal/mol.

In this work, we applied the MBAR+wTP method to calculate the FE surfaces of two Diels–Alder reactions of cyclopentadiene with either acrylonitrile or 1-4-naphthoquinone as mentioned above and investigated the applicability of this method to the study of reactions with two-dimensional (2D) reaction coordinates (RC).

2. Methods

2.1. Multistate Bennett Acceptance Ratio and Weighted Thermodynamic Perturbation (MBAR+wTP) Method

In our previous work [40], all the reactions studied were characterized by a one-dimensional generalized reaction coordinate. It is well-known that the reference-potential method suffers from numerical difficulty when the reference Hamiltonian and the target Hamiltonian have no significant overlap in phase space. With the increasing complexity of the molecule, this difficulty becomes more and more severe, and the applicability of this method is challenged. This work aims to investigate the applicability of this method to the calculations of the *ai* FE surfaces of two Diels–Alder reactions with two-dimensional reaction coordinates.

The two-dimensional reaction coordinate is denoted by $\boldsymbol{\eta} \equiv (\eta_1(\mathbf{x}), \eta_2(\mathbf{x}))$, where $\eta_1(\mathbf{x})$ and $\eta_2(\mathbf{x})$ are functions of the collective atomic coordinates $\mathbf{x}$. In two-dimensional US simulations [43], two harmonic restraining potentials $W_i^1(\eta_1) = \frac{1}{2}k_{1,i}(\eta_1 - \eta_{1,i})^2$ and $W_i^2(\eta_2) = \frac{1}{2}k_{2,i}(\eta_2 - \eta_{2,i})^2$ are added to the original potential energy surface $U_0(\mathbf{x})$ of the system, where $\eta_{1,i}$ and $\eta_{2,i}$ are the target values of RC, and $k_{1,i}$ and $k_{2,i}$ are the strengths of the restraints on the first and second dimensions, respectively, in the *i*th biased window. A set of two-dimensional biased window simulations indexed by i ($i = 1, 2, \ldots, S$) are carried out with the potential energy surfaces $U_i^{(b)}(\mathbf{x}) = U_0(\mathbf{x}) + W_i(\boldsymbol{\eta})$, where $W_i(\boldsymbol{\eta}) = W_i^1(\eta_1) + W_i^2(\eta_2)$ is the restraint potential. The trajectories are then post-processed using the MBAR method to obtain the unbiased thermodynamic properties on the original potential energy surface $U_0(\mathbf{x})$. The nonlinear equations are solved using the pyMBAR program.

The ensemble average of any physical operator $\hat{O}$ under Hamiltonian $\mathcal{H}$ can always be computed via

$$\langle \hat{O} \rangle_{\mathcal{H}} = \sum_l \omega_{\mathcal{H}}^l O_l, \tag{1}$$

where O_l is the value of the operator $\hat{O}$ for configuration l, $\omega_{\mathcal{H}}^l$ is the *normalized* weight of this configuration under Hamiltonian $\mathcal{H}$, and the summation is over all the configurations. Strictly speaking, we should use thermodynamic state instead of Hamiltonian. The ensemble average is also a function of other macroscopic thermodynamic parameters such as temperature. In this work, the only difference between two thermodynamic states is the Hamiltonian, while the temperature is kept the same. Therefore, we use Hamiltonians to differentiate the thermodynamic states. In the reference-potential method, simulations are carried out under an inexpensive Hamiltonian $\mathcal{L}$, sometimes with biasing

potential W as, for instance, in umbrella sampling. With the samples harvested in the simulations, ensemble averages under an expensive Hamiltonian $\mathcal{H}$ can be calculated via energy reweighting, which will be explained in the following. In this way, the expensive sampling under Hamiltonian $\mathcal{H}$ is avoided.

Suppose we have carried out an umbrella sampling simulation with S windows in total. For the ith window, the potential is $U_i^{(b)}(\mathbf{x}) = U_{\mathcal{L}}(\mathbf{x}) + W_i(\boldsymbol{\eta})$, with $U_{\mathcal{L}}(\mathbf{x})$ being the unbiased potential and $W_i(\boldsymbol{\eta})$ the biasing potential. For the lth configuration from the ith window $\mathbf{x}_{i,l}$, the weight under the unbiased Hamiltonian $\mathcal{L}$ can be written as [40,42]

$$w_{\mathcal{L}}(\mathbf{x}_{i,l}) = \frac{e^{-\beta\left[U_{\mathcal{L}}(\mathbf{x}_{i,l}) - f_{\mathcal{L}}\right]}}{\sum\limits_{k=1}^{S} N_k e^{-\beta\left[U_k^{(b)}(\mathbf{x}_{i,l}) - f_k^{(b)}\right]}}, \tag{2}$$

where $f_k^{(b)}$ is the free energy of the biased window k and β is the reciprocal of the thermodynamic temperature. $f_k^{(b)}$ can be obtained by iteratively solving the core equations in the MBAR method [40–42].

$$e^{-\beta f_k^{(b)}} = \sum_{i=1}^{S} \sum_{l=1}^{N_i} \frac{e^{-\beta\left[W_k(\mathbf{x}_{i,l})\right]}}{\sum\limits_{t=1}^{S} N_t e^{-\beta\left[W_t(\mathbf{x}_{i,l}) - f_t^{(b)}\right]}}. \tag{3}$$

The numerator $e^{\beta f_{\mathcal{L}}}$ in Equation (2) is constant for a given Hamiltonian $\mathcal{L}$, which can be canceled during the normalization. By eliminating $U_{\mathcal{L}}(\mathbf{x})$ from both the numerator and denominator in Equation (2), the *unnormalized* weight of sample $\mathbf{x}_{i,l}$ under the unbiased Hamiltonian $\mathcal{L}$ is

$$w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) = \frac{1}{\sum\limits_{k=1}^{S} N_k e^{-\beta\left[W_k(\mathbf{x}_{i,l}) - f_k^{(b)}\right]}}, \tag{4}$$

in which the superscript 0 is added to emphasize that this is the unbiased weight. The unbiased two-dimensional FE surface, with the reaction coordinate $\boldsymbol{\eta}$ represented as a vector, can thus be expressed as

$$F_{\mathcal{L}}(\boldsymbol{\eta}) = -\beta^{-1} \ln \sum_{i=1}^{S} \sum_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}), \tag{5}$$

in which the delta function picks out the samples falling into bin $\boldsymbol{\eta}$.

The variance of the estimated free energy difference between bin η_i and bin η_j can be obtained by

$$\delta^2 \Delta F_{\mathcal{L}}(\eta_{ij}) = \Theta_{ii} - 2\Theta_{ij} + \Theta_{jj}, \tag{6}$$

where the covariance matrix is obtained from Equation D8 in Ref. [41] by

$$\Theta = \left[\left(\mathbf{W}^T \mathbf{W}\right)^{-1} - \mathbf{N} + \mathbf{1}_{S+M} \mathbf{1}_{S+M}^T / N\right]^{-1}, \tag{7}$$

where $\mathbf{N} = diag(N_1, N_2, \ldots, N_S, 0_1, 0_2, \ldots, 0_M)$; M is the number of bins; $N_1, N_2, \ldots, N_S$ are, respectively, the number of samples in the 1st, 2nd, ..., Sth simulation window; N is the total number of samples collected from all the simulation windows; and $\mathbf{W}$ is a matrix with a dimension of $N \times (S + M)$, whose elements are depicted in detail in Ref. [40].

By combining the binned configurations from the two-dimensional US simulations with $U_i^{(b)}(\boldsymbol{\eta}) = U_{\mathcal{L}}(\mathbf{x}) + W_i(\boldsymbol{\eta})$ and their weights from the MBAR analysis, the FE surface at an arbitrary

high level Hamiltonian $\mathcal{H}$ can be calculated by wTP. For a certain two-dimensional histogram bin, the free energy difference between the low-level and the high-level Hamiltonians can be obtained via weighted TP as

$$
\Delta F(\boldsymbol{\eta}) = -\frac{1}{\beta} \ln \frac{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}}{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta})}, \tag{8}
$$

where the subscripts $\mathcal{H}$ and $\mathcal{L}$ denote the high-level and the low-level Hamiltonians, respectively. Again, the delta function picks out the samples falling into bin $\boldsymbol{\eta}$. Because the two-dimensional US samplings were performed at the low-level Hamiltonian, the FE surface $F_{\mathcal{L}}(\boldsymbol{\eta})$ corresponding to this low-level Hamiltonian $\mathcal{L}$ can be obtained via Equation (5). Then, the FE surface of the high-level Hamiltonian $\mathcal{H}$ can be calculated by

$$
F_{\mathcal{H}}(\boldsymbol{\eta}) = F_{\mathcal{L}}(\boldsymbol{\eta}) + \Delta F(\boldsymbol{\eta}). \tag{9}
$$

It is noted that any physical quantity $\langle O(\boldsymbol{\eta}) \rangle_{\mathcal{H}}$ at the high-level Hamiltonian $\mathcal{H}$ can be obtained via the reweighting method

$$
\langle O(\boldsymbol{\eta}) \rangle_{\mathcal{H}} = \frac{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} O(\mathbf{x}_{i,l}) w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}}{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}}, \tag{10}
$$

and its variance is computed via

$$
\begin{aligned}
\delta^2 \langle O(\boldsymbol{\eta}) \rangle_{\mathcal{H}} = {} & \frac{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} O^2(\mathbf{x}_{i,l}) w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}}{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}} \\
& - \left(\frac{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} O(\mathbf{x}_{i,l}) w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}}{\sum\limits_{i=1}^{S} \sum\limits_{l=1}^{N_i} w_{\mathcal{L}}^0(\mathbf{x}_{i,l}) \, \delta(\boldsymbol{\eta}(\mathbf{x}_{i,l}) - \boldsymbol{\eta}) e^{-\beta[U_{\mathcal{H}}(\mathbf{x}_{i,l}) - U_{\mathcal{L}}(\mathbf{x}_{i,l})]}} \right)^2.
\end{aligned} \tag{11}
$$

2.2. Gaussian Process Regression for FE Surfaces Smoothing

Following the same way as in our previous work [40], a nearly model-free method called Gaussian processes regression (GPR) [44] was utilized to smooth the FE surface after the wTP correction, removing the statistical noise in the wTP correction process as shown in Figure S3. Given a set of observations $\{F_1, F_2, \ldots, F_n\}$, it can be viewed as a single sample from a Gaussian distribution with n variates. Here, due to the two-dimensional reaction coordinates of the reactions studied in this work, each variate has two features which can be labeled by vector $\boldsymbol{\eta}$. Since the observations are noisy, each observation F is related to an underlying function $f(\boldsymbol{\eta})$ through a Gaussian noise model

$$
F = f(\boldsymbol{\eta}) + \mathcal{N}(0, \sigma_n^2). \tag{12}
$$

Then, the covariance function kernel k was defined using the squared exponential as

$$k(\eta, \eta') = \sigma_f^2 \exp\left[\frac{-(\eta - \eta')^2}{2l^2}\right] + \alpha\sigma_n^2\delta(\eta, \eta'), \tag{13}$$

where l is the length-scale and σ_f^2 is the signal variance, $\delta(\eta, \eta')$ is the Kronecker delta function and σ_n^2 is the noise variance, which was set to the reciprocal of the exponential of the reweighting entropy [45] value ($e^{-\mathcal{S}}$) corresponding to each observation in this work as done in our previous work [40]. The "hyperparameters" $\{l, \sigma_f, \alpha\}$ are optimized to maximize the likelihood of the observations. For any point η_* along the reaction coordinate, the free energy can be calculated, with the existence of n training data $\{\eta_i, f(\eta_i)\}$, $i = 1, 2, \dots, n$, via

$$F(\eta_*) = k_*^T K^{-1}(\eta, \eta)\mathbf{f}(\eta), \tag{14}$$

and its variance via

$$\delta^2 F(\eta_*) = k(\eta_*, \eta_*) - k_*^T K^{-1}(\eta, \eta)k_*, \tag{15}$$

where

$$K = \begin{bmatrix} k(\eta_1, \eta_1) & k(\eta_1, \eta_2) & \cdots & k(\eta_1, \eta_n) \\ k(\eta_2, \eta_1) & k(\eta_2, \eta_2) & \cdots & k(\eta_2, \eta_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\eta_n, \eta_1) & k(\eta_n, \eta_2) & \cdots & k(\eta_n, \eta_n) \end{bmatrix}, \tag{16}$$

$$k_* = \begin{bmatrix} k(\eta_*, \eta_1) & k(\eta_*, \eta_2) & \cdots & k(\eta_*, \eta_n) \end{bmatrix}^T. \tag{17}$$

and

$$\mathbf{f}(\eta) = \begin{bmatrix} f(\eta_1) & f(\eta_2) & \cdots & f(\eta_n) \end{bmatrix}^T. \tag{18}$$

Gaussian process regression was performed by using the scikit-learn package [46].

2.3. Locating the Transition State on the Free Energy Surface

The transition state is located on the smoothed free energy surface by satisfying two conditions

$$\frac{\partial \mathbf{F}}{\partial \eta_i} = 0 \tag{19}$$

and the Hessian $\partial^2 \mathbf{F}/\partial\eta_i\partial\eta_j$, $i, j \in \{1, 2\}$ has one positive and one negative eigenvalues, where $\mathbf{F}$ and η_i are the free energy and the reaction coordinate, respectively. The gradient and the Hessian were computed using the finite central-difference method.

2.4. Solvent-Assisted Charge Transfer on the Transition State

The solvent-assisted charge transfer on the transition state can be approximately delineated as interactions between the transferred charges within the newly formed C–C bonds generating a local dipole and the electric field on the middle point of the bonds (M1 or M2) contributed from all the solvent molecules as depicted in Figure 1. The electric field $\vec{E}$ on point j is a sum over the contributions from all the solvent molecules with atomic partial charge q_i by

$$\vec{E}_j = \sum_i \frac{q_i(\vec{r}_j - \vec{r}_i)}{|\vec{r}_j - \vec{r}_i|^3}, \tag{20}$$

where $\vec{r}_i$ is the coordinate of atom i from solvent molecules, $\vec{r}_j$ is the coordinates of the middle points of newly formed C-C bonds (M1 or M2) as shown in Figure 1. The projections of the electric fields along the unit vectors $\vec{e}_1$ and $\vec{e}_2$, respectively, may facilitate or impede the charge transfer between the reactants. The ensemble averages of these physical quantities are computed via Equation (10) by

substituting $\hat{O}$ with the corresponding operators. The interaction between the local dipole moment and the solvent-generated electric field can be computed via

$$\Delta E = -0.5(\delta q_2 - \delta q_1) \times l_{C-C} \times \hat{E}, \tag{21}$$

where δq_1 and δq_2 are the changes of the atomic Mulliken charge from the reactant to the transition state for C_1(or C_2) and C_3 (or C_4), l_{C-C} is the bond length of $C_1 - C_3$ (or $C_2 - C_4$), and $\hat{E}$ is the projection of the solvent-generated electric field along the $C_1 - C_3$ (or $C_2 - C_4$) bond.
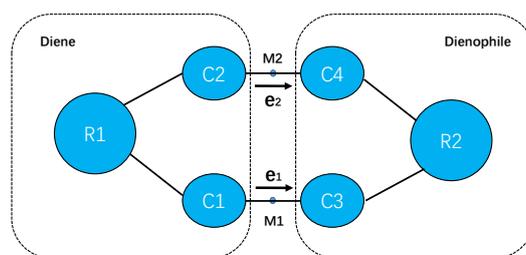
Figure 1. The diagram representation of active sites for Diels–Alder reactions in this work, where C1, C2, C3 and C4 are the four atoms shown in Figure 2, and the remaining atoms in diene are grouped together as R1, and those in dienophile as R2. The unit vector $\vec{e}_1$ points from atom C1 to atom C3 and the unit vector $\vec{e}_2$ points from atom C2 to atom C4. M1 and M2 are the two middle points of C1–C3 and C2–C4 bonds, on which the solvent electric fields are computed.

2.5. Gibbs Free Energies in Implicit Water Solvent

To save computational cost, implicit solvent model is frequently used during the calculations of stationary structures, activation free energy and reaction free energy. In this method, instead of an ensemble of structures, a unique structure is used to represent each of the stationary points (such as reactant, transition state, and product), thus avoiding expensive sampling in the phase space. Reaction free energy and activation free energy can be computed under the rigid-rotor/harmonic-oscillator (RRHO) approximation via frequency analyses for the reactant, transition state and product. In this work, the calculations were carried out at the same ab initio levels as in explicit solvent (B3LYP/6-31G(d) or B3LYP-D3/6-31G(d) level), and the integral equation formalism of the polarizable continuous solvent model (IEF-PCM) was adopted for the solvation effect. In addition, we also chose a higher level *ai* method (MP2/6-311 + g(2d,p)/IEF-PCM), which is similar to that used in Ref. [7] by Jorgensen et al. For the optimization of the transition structures, the option opt = QST3 implemented in Gaussian 16 [47] was used, which requires three molecular specifications corresponding to the reactant, the product, and an initial guess for the transition state structure. All the transition state structures are verified via intrinsic reaction coordinate (IRC) analysis, in which both the reactant and the product can be identified along the IRC path starting from transition state structure.

2.6. Molecular Dynamics Simulations

Two Diels–Alder reactions of cyclopentadiene with either acrylonitrile or 1-4-naphthoquinone were studied in this work, which are shown in Figure 2. Only the reactant molecules including cyclopentadiene and either acrylonitrile or 1-4-naphthoquinone were defined as the semi-empirical QM (SE QM) or the QM region, and the remaining of the system (including only the water molecules) were defined as the MM region. A TIP3P water sphere with a radius of 25 Å was added to solvate the reactive molecules centering on the heavy atom closest to the center-of-mass of the QM regions and was restrained by a soft half-harmonic potential with a force constant of 10 kcal/mol/Å^2 to prevent evaporation, as done by Thiel et al. [48]. There were 2007 water molecules for the ACR system and 1996 water molecules for the NAP system. The nonbonded interactions were fully considered without a truncation and the general AMBER force field [49] was assigned to the solute molecules.

PM6 was used as the low-level Hamiltonian, and B3LYP(-D3)/6-31G(d) was chosen as the high-level QM Hamiltonian. Here, the electrostatic embedding scheme was used to explicitly take into account the polarization effect from the MM region on the QM region. The two-dimensional umbrella sampling simulations were performed at PM6/MM level. The indirect FE surfaces at B3LYP(-D3)/6-31G(d)/MM level was computed by reweighting from the PM6/MM level.

(**a**)

(**b**)
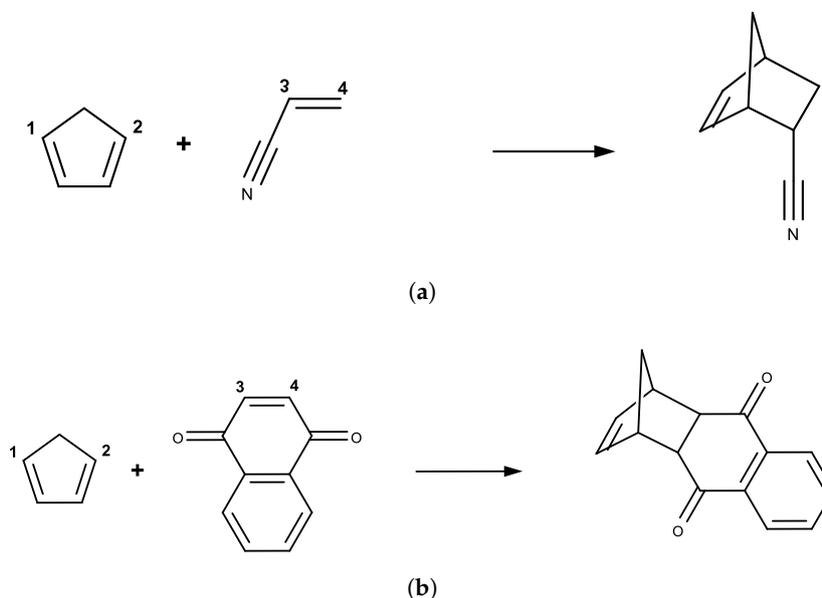
Figure 2. Diels–Alder reactions between: (**a**) cyclopentadiene and acrylonitrile; and (**b**) cyclopentadiene and 1-4-naphthoquinone, where $\eta_1 = d_{C1C3}$ and $\eta_2 = d_{C2C4}$ were chosen as the two-dimensional reaction coordinates.

The endo addition mode for these reactions was chosen, because it corresponds to the preferred transition state from ab initio calculations [4,7] and experimental stereoselectivity preferences [1,5]. The $\eta_1 = d_{C1C3}$ and $\eta_2 = d_{C2C4}$ were chosen as two-dimensional reaction coordinates in both cases. Two-dimensional umbrella samplings were conducted centering on $\boldsymbol{\eta} \equiv (\eta_1, \eta_2)$ ranged from 1.50 to 4.00 Å with increments of 0.05 Å for each dimension. The reactant state in both cases was defined as η_1 = 4.00 Å and η_2 = 4.00 Å, where the FE surfaces were rather flat in the vicinity. To reduce the computational cost, only the important region on the FE surface was sampled.

For each two-dimensional US window under the PM6/MM Hamiltonian, the system was energy-optimized for 500 steps using the steepest decent optimization method followed by 500 steps of the conjugate gradient method with the solute molecules restrained. Then, the same optimization procedure continued with the restraint removed. The system was heated up to 298.15 K in 50 ps and was equilibrated for 100 ps. A 1-ns production molecular dynamics (MD) simulation was performed for each window. The integration time step was set to 1 fs and the configurations were saved every 1 ps. The temperature was regulated at 298.15 K with the Andersen temperature coupling scheme [50]. Then, single point energies under the PM6/MM and B3LYP/6-31G(d)/MM Hamiltonians for the one thousand configurations saved in each US simulation were calculated and taken into the TP reweighting calculations. All the simulations were performed by the AmberTools 17 program package [51], and the QM/MM calculations were carried out by interfacing with Gaussian 16 package [47].

3. Results and Discussion

In the MBAR+wTP method, there are two key factors that are critical to the reliability of the results. The first one is the similarity between the Hamiltonians for neighboring windows in the two-dimensional US. It can be characterized by the overlap of the samples in the phase space, which

can be quantitatively measured by, for instance, the overlap matrix proposed by Klimovich et al. [52]. As shown in Figure S1 in the Supplementary Materials, for both DA reactions, the overlap between neighboring windows are larger than 0.03, which is the lower-limit suggested by Klimovich et al. It indicates that the phase space overlap is sufficient for the subsequent MBAR analysis. Therefore, all the PM6/MM FE surfaces calculated by the MBAR method are statistically reliable. The other factor is the similarity between the PM6/MM Hamiltonian and the B3LYP/MM Hamiltonian, which determines the reliability of the weighted TP and can be measured by reweighting entropy [45]. As shown in Figure S2, overall reweighting entropy values are large enough for yielding reliable results in the weighted TP calculations. Because of the statistical noise in the wTP correction process, as shown in Figure S3, Gaussian processes regression (GPR) method [44] was used to smooth the FE surface after the wTP correction.

3.1. DA Reaction between CP and ACR

As shown in Figure 3, the product state defined by $d_{C1C3} = 1.58$ Å and $d_{C2C4} = 1.58$ Å is a reference state with zero free energy and the reactant state is defined with $d_{C1C3} = 4.00$ Å and $d_{C2C4} = 4.00$ Å. From the FE surfaces obtained by our method, the transition state can be located at $(\eta_1, \eta_2) = (2.41$ Å, 1.92 Å$)$ at the PM6/MM level, which is different from $(\eta_1, \eta_2) = (2.28$ Å, 2.00 Å$)$ at the PM3/MM level obtained by Jorgensen et al. and indicates stronger asynchronism of this reaction at the PM6/MM level than that at the PM3/MM level. Meanwhile, the transition state at the B3LYP/MM level is located at $(\eta_1, \eta_2) = (2.49$ Å, 2.05 Å$)$, which shows a small difference from that obtained at the PM6/MM level and also manifests the strong asynchronism of this reaction. As listed in Table 1, Jorgensen and coworkers showed that the activation free energy for this DA reaction is 34.0 ± 0.5 kcal/mol under PM3/MM Hamiltonian [7], and 24.7 kcal/mol under AM1/MM Hamiltonian [6]. In the current work, we found the activation free energy to be 30.9 ± 0.1 kcal/mol under PM6/MM Hamiltonian. These semi-empirical methods significantly overestimate the activation barrier for this DA reaction, which has an experimental value of 22.2 kcal/mol [3]. The large deviations in the activation barrier lie in the limited accuracy of these semi-empirical methods, as pointed out in Ref. [7]. An activation barrier of 20.5 ± 0.6 kcal/mol is obtained at the B3LYP/MM level via the weighted TP correction from the PM6/MM free energy landscape, which is much closer to the experimental value of 22.2 kcal/mol [3] than those semiempirical methods. It is interesting to note that B3LYP/IEF-PCM overestimates the activation free energy by 8.1 kcal/mol, whereas MP2/IEF-PCM underestimates the activation free energy by 4.6 kcal/mol. Both of them had worse performance than B3LYP/MM in the explicit solvent model. In terms of the reaction free energy, the B3LYP/MM method yielded a value of -15.7 ± 0.6 kcal/mol, which is much closer to the value of -15.8 kcal/mol at the MP2/IEF-PCM level, compared to -26.2 kcal/mol at the AM1/MM level, -16.7 ± 0.6 kcal/mol at the PM3/MM level and -17.5 ± 0.1 kcal/mol at the PM6/MM level. Surprisingly, B3LYP/IEF-PCM significantly underestimates the exothermicity of this reaction, which may come from an inadequate description of the solvent by the continuous solvent model.

As listed in Table 2, the projection of the solvent electric field on M1 (the midpoint of C1–C3 bond) has a magnitude of -14.6×10^{-4} a.u. at the B3LYP/MM level. This electric field facilitates the charge transfer and stabilizes the transition state. As listed in Table 3, from the reactant to the transition state, the charge of C1 atom increases by $0.12e$, and that of C3 atom decreases by $0.18e$, which generates a large dipole moment change along this C1–C3 bond. The projection of the solvent electric field has a magnitude of -19.7×10^{-4} a.u. on M2, which is even stronger than that along the C1–C3 bond. Although this electric field also facilitates the charge transfer from the diene to the dienophile, we observed an inverse flow of electron from C4 to C2. The charge of C2 atom decreases by $0.02e$, and that of C4 increases by $0.09e$, which also generates a dipole moment change along this C2–C4 bond. These two dipole–electric field interactions stabilize the transition state by about 0.4 kcal/mol in total.
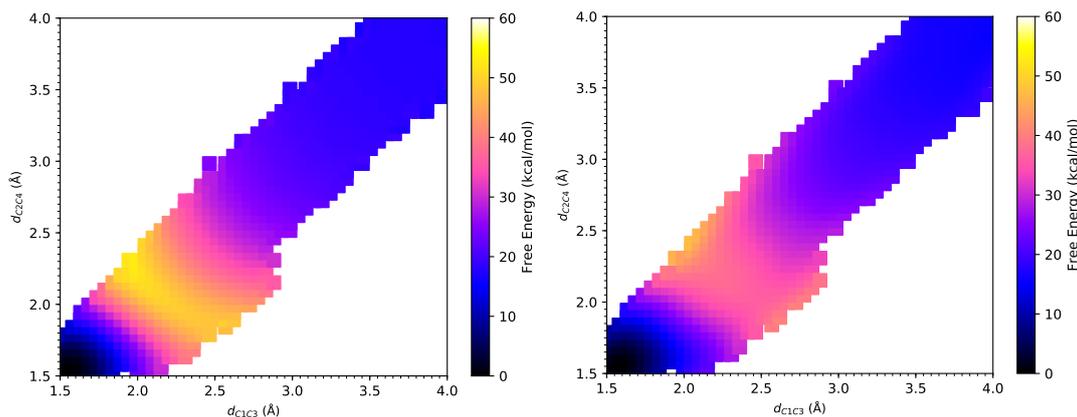
Figure 3. The free energy surfaces under the PM6/MM Hamiltonian (**left**) and under the B3LYP/MM Hamiltonian after Gaussian process regression (**right**) for Diels–Alder reaction between cyclopentadiene and acrylonitrile.

Table 1. Activation free energy $\Delta G^{\ddagger}$, reaction free energy ΔG (in kcal/mol) and the structure of the transition state at 298.15 K for the Diels–Alder reactions between cyclopentadiene and acrylonitrile (ACR), 1-4-naphthoquinone (NAP).

Dienophile	Method	$\Delta G^{\ddagger}$	ΔG	(η_1, η_2) [a]
	PM3/MM [b]	34.0 ± 0.5	-16.7 ± 0.6	(2.28, 2.00)
	AM1/MM [c]	24.7	−26.2	
	PM6/MM	30.9 ± 0.1	-17.5 ± 0.1	(2.41, 1.92)
ACR	B3LYP/MM	20.5 ± 0.6	-15.7 ± 0.6	(2.49, 2.05)
	B3LYP/IEFPCM	30.3	−2.69	(2.54, 2.04)
	MP2/IEFPCM	17.6	−15.8	(2.40, 2.18)
	Exp. [d]	22.2		
	PM3/MM [b]	26.0 ± 0.5	-20.1 ± 0.6	(2.25, 2.22)
	AM1/MM [c]	27.6	−4.4	
	PM6/MM	29.6 ± 0.1	-16.7 ± 0.1	(2.14, 2.18)
NAP	B3LYP-D3/MM	14.3 ± 0.7	-11.5 ± 0.7	(2.23, 2.19)
	B3LYP-D3/IEFPCM	20.7	−2.62	(2.18, 2.18)
	MP2/IEFPCM	6.6	−13.9	(2.26, 2.26)
	Exp. [e]	16.6		

[a] Transition state position in Å; [b] Ref. [7] by Jorgensen et al. using PDDG/PM3/MM/MC; [c] Ref. [6] with one-dimensional reaction coordinate; [d] Ref. [3] at 303.15 K; [e] Ref. [2].

Table 2. The projections of electric fields on points M1 and M2, respectively, along unit vectors $\vec{e}_1$ and $\vec{e}_2$ (in 10^{-4} a.u. where 1 a.u. = 51.42 V/Å) for the Diels–Alder reactions between cyclopentadiene and acrylonitrile (ACR), 1-4-naphthoquinone (NAP).

| Dienophile	Method	Locations	Projection of E-Field
ACR	B3LYP/MM	M1	-14.6 ± 0.3
		M2	-19.7 ± 0.4
NAP	B3LYP$-$D3/MM	M1	-28.0 ± 0.5
		M2	-36.2 ± 0.7

Table 3. The Mulliken partial charges (in a.u.) of all atoms shown in Figure 2 for the Diels–Alder reactions between cyclopentadiene and acrylonitrile (ACR), 1-4-naphthoquinone (NAP).

Dienophile	Method	Atoms	Reactant State	Transition State	Product State
ACR	B3LYP/MM	C1	−0.13	−0.01	−0.05
		C2	−0.15	−0.17	−0.05
		C3	−0.18	−0.36	−0.13
		C4	−0.20	−0.11	−0.28
		R1	0.29	0.35	0.22
		R2	0.38	0.29	0.28
NAP	B3LYP−D3/MM	C1	−0.15	−0.06	−0.04
		C2	−0.14	−0.04	−0.03
		C3	−0.21	−0.26	−0.25
		C4	−0.23	−0.26	−0.26
		R1	0.30	0.34	0.26
		R2	0.44	0.29	0.32

3.2. DA Reaction between CP and NAP

From the FE surfaces, as shown in Figure 4, the transition state can be located at $(\eta_1, \eta_2) =$ (2.14 Å, 2.18 Å) at the PM6/MM level, which has a slight deviation from $(\eta_1, \eta_2) =$ (2.25 Å, 2.22 Å) at the PM3/MM level obtained by Jorgensen et al. The transition state is located at $(\eta_1, \eta_2) =$ (2.23 Å, 2.19 Å) at the B3LYP-D3/MM level [53]. These small differences in these two RC come from statistical noise in the samples. PM3/MM agrees with B3LYP-D3/MM better than PM6/MM does. All the transition state positions obtained by different methods indicate a symmetrical and synchronous process of this reaction [7]. As listed in Table 1, the activation free energy for this DA reaction is 26.0 ± 0.5 kcal/mol under the PM3/MM Hamiltonian [7], and 27.6 kcal/mol under the AM1/MM Hamiltonian [6] according to Jorgensen et al. In contrast, our results was 29.6 ± 0.1 kcal/mol under the PM6/MM Hamiltonian. Thus, similar to the CP-ACR reaction, all semi-empirical methods again overestimate the activation barrier for this DA reaction, which has an experimental value of 16.6 kcal/mol [2]. Encouragingly, an activation barrier of 14.3 ± 0.7 kcal/mol at the B3LYP-D3/MM level was obtained via weighted TP from the PM6/MM free energy landscape, which is very close to the experimental value. With IEF-PCM, B3LYP-D3 overestimates the activation free energy by 4.1 kcal/mol, while MP2 significantly underestimates the activation free energy by 10.0 kcal/mol. In the meantime, B3LYP-D3/MM gives a reaction free energy of −11.5 ± 0.7 kcal/mol, which is also very close to the value of −13.9 kcal/mol calculated at the MP2/IEF-PCM level. Those from AM1/MM, PM3/MM and PM6/MM calculations are −4.4 kcal/mol, −20.1 ± 0.6 kcal/mol, and −16.7 ± 0.1 kcal/mol, respectively, showing larger deviations. Again, B3LYP-D3/IEF-PCM significantly underestimates the exothermicity of this reaction.

As listed in Table 2, the projection of the solvent electric field on M1 has a large magnitude of $−28 \times 10^{-4}$ a.u. at the B3LYP-D3/MM level. This electric field facilitates the charge transfer and stabilizes the transition state. As listed in Table 3, from the reactant to the transition state, the charge of C1 atom increases by 0.09*e*, and that of C3 atom decreases by 0.05*e*. This local dipole moment stabilizes the transition state by interacting with the solvent-generated electric field. The projection of the solvent electric field on M2 has a magnitude of $−36.2 \times 10^{-4}$ a.u., which is even stronger than that along the C1–C3 bond. The charge of C2 atom increases by 0.10*e*, and that of C4 decreases by 0.03*e*. This dipole moment also lowers the energy of the transition state in the electrostatic environment of the solvent molecules. These two dipole–electric field interactions stabilize the transition state by about 1.1 kcal/mol in total.
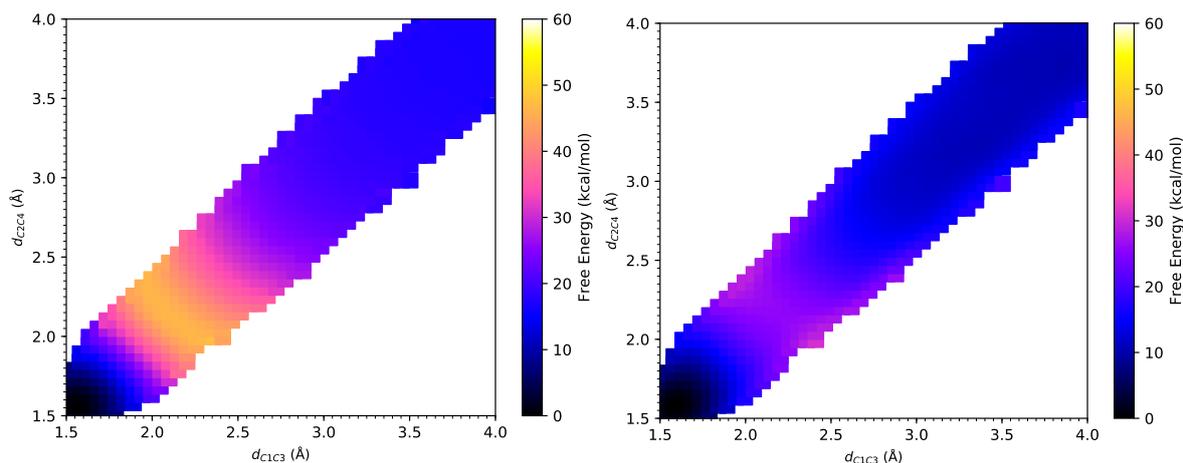
Figure 4. The free energy surfaces under the PM6/MM Hamiltonian (**left**) and under the B3LYP-D3/MM Hamiltonian after Gaussian process regression (**right**) for Diels–Alder reaction between cyclopentadiene and 1-4-naphthoquinone.

3.3. Computational Expense

The estimated wall-clock time for the computations of the QM/MM FE surfaces at the B3LYP/MM level are listed in Table 4. For the calculations of the QM/MM FE surfaces via weighted TP, the wall-clock time includes both the time for generating the SE QM/MM trajectories and the time for the single-point energy calculations at the B3LYP/MM level. The estimated wall-clock time for a direct free energy calculation at the B3LYP/MM level is also listed. It can be seen that the calculation efficiency is enhanced by about 139 and 376 times for CP/ACR and CP/NAP reactions, respectively, via this indirect free energy calculation utilizing a dual-Hamiltonian approach. Because one out of every 1000 configurations are required for energy calculations at the ab initio level in this approach (1000 configurations for each 1-ns window in the indirect approach vs. 1,000,000 configurations in the direct approach), the efficiency enhancement can never exceed 1000. Besides, the computational cost at the PM6/MM level is non-negligible relative to the B3LYP/MM calculations for such small systems. It can also be seen that the efficiency enhancement is greater for the larger reaction system (CP/NAP), because the cost for the low-level sampling becomes less significant relative to that for the single-point energy evaluations at the *ai* level for larger molecules. Overall, an enhancement of two orders of magnitude in efficiency is quite satisfactory.

Table 4. Estimated wall-clock time in a unit of hours for the computations of the QM/MM free energy surfaces at the B3LYP/6-31G(d) level. Assuming one node with 16 cores of Intel Xeon CPU E5-2660 2.20 GHz was used.

Dienophile	PM6/MM to B3LYP/MM Indirect			Direct B3LYP/MM [a]
	Sampling	Energy Evaluation	Total	
ACR	5762	1459	7221	1,006,041
NAP	5945	3907	9852	3,704,062

[a] Using the same number of windows as in the semi-empirical simulations, and one 1-ns simulation for each window.

4. Conclusions

For the Diels–Alder reactions in solution, the computation of converged free energy (FE) surfaces at ab initio (*ai*) QM/MM level is still far from being affordable. In this work, we applied our recently proposed MBAR+wTP method to calculate the two-dimensional FE surfaces of two Diels–Alder reactions of cyclopentadiene with either acrylonitrile or 1-4-naphthoquinone at *ai* QM/MM level with much less computational expense. Due to some approximation lying in the semi-empirical (SE)

method, the FE surfaces at SE QM/MM level deviate from experimental values by several kcal/mol in terms of the activation and the reaction free energies. However, our method can yield the FE profile at the *ai* QM/MM level without performing the expensive *ai* QM/MM MD simulations. Besides the FE surface, other ensemble-averaged properties such as the amount of charge transferred and the external electric potential/field are also readily available. The results agree much better with the experimental measurements than those obtained by other methods for these two Diels–Alder reactions in terms of the activation free energy. Care must be taken when using implicit solvent models, especially when calculating the properties of transition state. Further validation of this method to systems of much higher complexity such as enzymatic reactions will be carried out in future studies.

Supplementary Materials: The following are available online at http://www.mdpi.com/1420-3049/23/10/2487/s1. Figure S1: The overlap matrix elements at the PM6/MM level for the Diels–Alder reactions between: (a) cyclopentadiene and acrylonitrile; and (b) cyclopentadiene and 1-4-naphthoquinone.; Figure S2: The reweighting entropy values in reweighting from the PM6/MM Hamiltonian to the B3LYP(-D3)/MM Hamiltonian for the Diels–Alder reactions between: (a) cyclopentadiene and acrylonitrile; and (b) cyclopentadiene and 1-4-naphthoquinone.; Figure S3: The raw free energy surfaces under the B3LYP(-D3)/MM Hamiltonian for the Diels–Alder reactions between: (a) cyclopentadiene and acrylonitrile; and (b) cyclopentadiene and 1-4-naphthoquinone before Gaussian Process Regression.

Abbreviations

The following abbreviations are used in this manuscript:

DA	Diels–Alder
FE	Free Energy
ai	ab initio
QM	Quantum Mechanical
MM	Molecular Mechanical
US	Umbrella Sampling
MD	Molecular Dynamics
SE	Semi-Empirical
MBAR	Multistate Bennett Acceptance Ratio
wTP	weighted Thermodynamic Perturbation
CP	Cyclopentadiene
ACR	Acrylonitrile
NAP	1-4-Naphthoquinone
FEP	Free Energy Perturbation
TS	Transition State
EVB	Empirical Valence Bond
SAMPL	Statistical Assessment of the Modeling of Proteins and Ligands
RC	Reaction Coordinate
GPR	Gaussian Processes Regression
RRHO	Rigid-Rotor/Harmonic-Oscillator
IRC	Intrinsic Reaction Coordinate
IEF-PCM	Integral Equation Formalism of the Polarizable Continuous Solvent Model
MP2	2nd-Order Møller–Plesset Perturbation Theory

AM1   Austin Model 1
PM3   Parameterization Method 3
PM6   Parameterization Method 6

References

1.   Breslow, R.; Maitra, U.   On the Origin of Product Selectivity in Aqueous Diels-Alder Reactions. *Tetrahedron Lett.* **1984**, *25*, 1239–1240. [CrossRef]

2.   Engberts, J.B. Diels-Alder Reactions in Water: Enforced Hydrophobic Interaction and Hydrogen Bonding. *Pure Appl. Chem.* **1995**, *67*, 823–828. [CrossRef]

3.   Rideout, D.C.; Breslow, R. Hydrophobic Acceleration of Diels-Alder Reactions. *J. Am. Chem. Soc.* **1980**, *102*, 7816–7817. [CrossRef]

4.   Jorgensen, W.L.; Lim, D.; Blake, J.F. Ab initio Study of Diels-Alder Reactions of Cyclopentadiene with Ethylene, Isoprene, Cyclopentadiene, Acrylonitrile, and Methyl Vinyl Ketone. *J. Am. Chem. Soc.* **1993**, *115*, 2936–2942. [CrossRef]

5.   Cativiela, C.; Garcia, J.; Gil, J.; Martinez, R.; Mayoral, J.; Salvatella, L.; Urieta, J.; Mainar, A.; Abraham, M. Solvent Effects on Diels-Alder Reactions. The Use of Aqueous Mixtures of Fluorinated Alcohols and the Study of Reactions of Acrylonitrile. *J. Chem. Soc. Perkin Trans. 2* **1997**, *4*, 653–660, doi:10.1039/A601717G. [CrossRef]

6.   Chandrasekhar, J.; Shariffskul, S.; Jorgensen, W.L. QM/MM Simulations for Diels-Alder Reactions in Water: Contribution of Enhanced Hydrogen Bonding at the Transition State to the Solvent Effect. *J. Phys. Chem. B* **2002**, *106*, 8078–8085. [CrossRef]

7.   Acevedo, O.; Jorgensen, W.L. Understanding Rate Accelerations for Diels-Alder Reactions in Solution Using Enhanced QM/MM Methodology. *J. Chem. Theory Comput.* **2007**, *3*, 1412–1419. [CrossRef] [PubMed]

8.   Thomas, L.L.; Tirado-Rives, J.; Jorgensen, W.L. Quantum Mechanical/Molecular Mechanical Modeling Finds Diels-Alder Reactions Are Accelerated Less on the Surface of Water Than in Water. *J. Am. Chem. Soc.* **2010**, *132*, 3097–3104. [CrossRef] [PubMed]

9.   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. [CrossRef]

10. 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. [CrossRef]

11. Senn, H.M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229. [CrossRef] [PubMed]

12. van der Kamp, M.W.; Mulholland, A.J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728. [CrossRef] [PubMed]

13. Hayashi, S.; Uchida, Y.; Hasegawa, T.; Higashi, M.; Kosugi, T.; Kamiya, M. QM/MM Geometry Optimization on Extensive Free-Energy Surfaces for Examination of Enzymatic Reactions and Design of Novel Functional Properties of Proteins. *Annu. Rev. Phys. Chem.* **2017**, *68*, 135–154. [CrossRef] [PubMed]

14. Chandrasekhar, J.; Smith, S.F.; Jorgensen, W.L. Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1985**, *107*, 154–163. [CrossRef]

15. Chandrasekhar, J.; Jorgensen, W.L. Energy Profile for a Nonconcerted S_N2 Reaction in Solution. *J. Am. Chem. Soc.* **1985**, *107*, 2974–2975. [CrossRef]

16. Jorgensen, W.L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189. [CrossRef]

17. Stanton, R.V.; Peräkylä, M.; Bakowies, D.; Kollman, P.A. Combined ab initio and Free Energy Calculations To Study Reactions in Enzymes and Solution: Amide Hydrolysis in Trypsin and Aqueous Solution. *J. Am. Chem. Soc.* **1998**, *120*, 3448–3457. [CrossRef]

18. Zhang, Y.; Liu, H.; Yang, W. Free Energy Calculation on Enzyme Reactions with An Efficient Iterative Procedure to Determine Minimum Energy Paths on A Combined ab initio QM/MM Potential Energy Surface. *J. Chem. Phys.* **2000**, *112*, 3483–3492. [CrossRef]

19. Zwanzig, R.W.   High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426. [CrossRef]

20. Zhang, Y.; Lee, T.S.; Yang, W. A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods. *J. Chem. Phys.* **1999**, *110*, 46–54. [CrossRef]

21. Kästner, J.; Senn, H.M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction. *J. Chem. Theory Comput.* **2006**, *2*, 452–461. [CrossRef] [PubMed]

22. Blake, J.F.; Jorgensen, W.L. Solvent Effects on a Diels-Alder Reaction from Computer Simulations. *J. Am. Chem. Soc.* **1991**, *113*, 7430–7432. [CrossRef]

23. Blake, J.F.; Lim, D.; Jorgensen, W.L. Enhanced Hydrogen Bonding of Water to Diels-Alder Transition States. ab initio Evidence. *J. Org. Chem.* **1994**, *59*, 803–805. [CrossRef]

24. Cunningham, M.A.; Ho, L.L.; Nguyen, D.T.; Gillilan, R.E.; Bash, P.A. Simulation of the Enzyme Reaction Mechanism of Malate Dehydrogenase. *Biochemistry* **1997**, *36*, 4800–4816. [CrossRef] [PubMed]

25. Alhambra, C.; Wu, L.; Zhang, Z.Y.; Gao, J. Walden-Inversion-Enforced Transition-State Stabilization in a Protein Tyrosine Phosphatase. *J. Am. Chem. Soc.* **1998**, *120*, 3858–3866. [CrossRef]

26. Bentzien, J.; Muller, R.P.; Florián, J.; Warshel, A. Hybrid ab initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin. *J. Phys. Chem. B* **1998**, *102*, 2293–2301. [CrossRef]

27. Qian, P.; Guo, H.B.; Yue, Y.; Wang, L.; Yang, X.; Guo, H. Understanding the Catalytic Mechanism of Xanthosine Methyltransferase in Caffeine Biosynthesis from QM/MM Molecular Dynamics and Free Energy Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 1755–1761. [CrossRef] [PubMed]

28. Liu, F.; Yang, Z.; Mei, Y.; Houk, K.N. QM/QM′ Direct Molecular Dynamics of Water-Accelerated Diels-Alder Reaction. *J. Phys. Chem. B* **2016**, *120*, 6250–6254. [CrossRef] [PubMed]

29. Zheng, Y.; Thiel, W. Computational Insights into an Enzyme-Catalyzed [4+2] Cycloaddition. *J. Org. Chem.* **2017**, *82*, 13563–13571. [CrossRef] [PubMed]

30. Gao, J. Absolute Free Energy of Solvation from Monte Carlo Simulations Using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem.* **1992**, *96*, 537–540. [CrossRef]

31. Gao, J.; Xia, X. A Priori Evaluation of Aqueous Polarization Effects through Monte Carlo QM-MM Simulations. *Science* **1992**, *258*, 631–635. [CrossRef] [PubMed]

32. Warshel, A.; Weiss, R.M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226. [CrossRef]

33. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, NY, USA, 1991.

34. Plotnikov, N.V.; Kamerlin, S.C.L.; Warshel, A. Paradynamics: An Effective and Reliable Model for ab initio QM/MM Free-Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115*, 7950–7962. [CrossRef] [PubMed]

35. Plotnikov, N.V.; Warshel, A. Exploring, Refining, and Validating the Paradynamics QM/MM Sampling. *J. Phys. Chem. B* **2012**, *116*, 10342–10356. [CrossRef] [PubMed]

36. Lameira, J.; Kupchencko, I.; Warshel, A. Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J. Phys. Chem. B* **2016**, *120*, 2155–2164. [CrossRef] [PubMed]

37. Rod, T.H.; Ryde, U. Quantum Mechanical Free Energy Barrier for an Enzymatic Reaction. *Phys. Rev. Lett.* **2005**, *94*, 138302. [CrossRef] [PubMed]

38. Jia, X.; Wang, M.; Shao, Y.; König, G.; Brooks, B.R.; Zhang, J.Z.H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J. Chem. Theory Comput.* **2016**, *12*, 499–511. [CrossRef] [PubMed]

39. Liu, W.; Jia, X.; Wang, M.; Li, P.; Wang, X.; Hu, W.; Zheng, J.; Mei, Y. Calculations of the Absolute Binding Free Energies for Ralstonia solanacearum Lectins Bound with Methyl-α-l-fucoside at Molecular Mechanical and Quantum Mechanical/Molecular Mechanical Levels. *RSC Adv.* **2017**, *7*, 38570–38580. [CrossRef]

40. Li, P.; Jia, X.; Shao, Y.; Mei, Y. Computation of Free Energy Profile at ab initio QM/MM Level Made Orders of Magnitude Faster via the Reference-Potential Approach Using Weighted Thermodynamics Perturbation. *arXiv* **2018**, arxiv:1801.03674.

41. Shirts, M.R.; Chodera, J.D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105. [CrossRef] [PubMed]

42. Shirts, M.R. Reweighting from the Mixture Distribution as a Better Way to Describe the Multistate Bennett Acceptance Ratio. *arXiv* **2017**, arxiv:1704.00891.

43. Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [CrossRef]

44. Rasmussen, C.; Williams, C. Regression. In *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; pp. 7–32.

45. Wang, M.; Li, P.; Jia, X.; Liu, W.; Shao, Y.; Hu, W.; Zheng, J.; Brooks, B.R.; Mei, Y. Efficient Strategy for the Calculation of Solvation Free Energies in Water and Chloroform at the Quantum Mechanical/Molecular Mechanical Level. *J. Chem. Inf. Model.* **2017**, *57*, 2476–2489. [CrossRef] [PubMed]

46. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

47. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Revision B.01*; Gaussian Inc.: Wallingford, CT, USA, 2016.

48. Vasilevskaya, T.; Thiel, W. Periodic Boundary Conditions in QM/MM Calculations: Implementation and Tests. *J. Chem. Theory Comput.* **2016**, *12*, 3561–3570. [CrossRef] [PubMed]

49. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef] [PubMed]

50. Andersen, H.C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393. [CrossRef]

51. Case, D.A.; Cerutti, D.S.; Cheatham, T.E., III; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Greene, D.; Homeyer, N.; et al. AMBER *2017*; University of California: San Francisco, CA, USA, 2017.

52. Klimovich, P.V.; Shirts, M.R.; Mobley, D.L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2015**, *29*, 397–411. [CrossRef] [PubMed]

53. 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. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are not available from the authors.