Supporting Information

Rich Dynamics underlying Solution Reactions Revealed by Sampling and Data Mining of Reactive Trajectories

Jun Zhang[a,b,1], Zhen Zhang,[a,1], Yi Isaac Yang[a], Sirui Liu[a,b], Lijiang Yang[a,b,2], Yi Qin Gao[a,b,2]

[a]Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

[b]Biodynamic Optical Imaging Center, Peking University, Beijing 100871, China.

[1]J.Z. and Z.Z. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: ljyang@pku.edu.cn or gaoyq@pku.edu.cn

Table of Contents:

Supplementary Text

S-I Simulation Details

The MITS program and the ESoRT were executed at the QM/MM interface on AMBER10 MD platform[1]. The self-consistent charge density functional tight-binding (SCC-DFTB) method[2, 3] was adopted to approximate the quantum mechanical Hamiltonian of the reactant molecule, and the solvent water was treated with the SPCE model. MITS algorithm supports all quantum methods available in the hybrid MD. In this paper we chose SCC-DFTB out of following concerns: i) DFTB generally ensures better accuracy than other semi-empirical methods, performs reliably in polar systems, and is computationally extremely efficient compared to the first-principle methods[4]; ii) the current DFTB parameter library works quite well for peptide and organic compound structures and conformation energies, the geometries of hydrogen-bonded systems are well reproduced, and also include dispersion energies[5]; iii) in recent years, SCC-DFTB has been extensively benchmarked and widely utilized in research involving organic compounds, biomolecules, and materials, giving reliable results[2, 3, 6]; iv) moreover, the output of SCC-DFTB calculations can be properly trained with neural network algorithm to achieve DFT-level accuracy[7], further enhancing the reliability and practicability of this semi-empirical method.

A cubic box was built with dimensions (30*30*30 $Å^3$), containing 880 SPCE-modeled water molecules and one QM-treated solute molecule. The concentration of the reactant molecule is ~0.06 mol/kg, below the saturation limit of the homologs[8]. The system underwent a standard relaxation procedure and equilibrated to an NTP ensemble (300 K, 1atm). To equilibrate the system to the appropriate volume, the pressure of the system was adjusted to 1atm by the Berendsen weak-coupling[9] algorithm (with the relaxation time constants of 0.2 ps) under another 1ns long normal MD. A cutoff of 10.0 Å was applied for calculating nonbonding interactions. Before the production run of sampling, 5-ns long simulation was used to determine the parameters (e.g. [n_k], see *section II*) in MITS. Then the configurational sampling data were collected for over 500 ns in NPT ensemble (300 K, 1 atm). For ESoRT, we first chose the initial structures (see *section IV*) from the MITS-generated configurations, then executed the true-dynamics transition path shootings under NVE ensemble. Totally 525 successfully transited trajectories were collected (out of ~25,000 total shooting attempts). All the simulations were performed with a 1-fs time integration step (no SHAKE on QM-treated molecule) and with periodic boundary condition.

S-II Multilevel Integrated Tempering Sampling (MITS)

In the selective integrated tempering sampling (SITS) method, we divide the simulation system into two parts and potential energy into three parts. For instance, in QM/MM simulation, the system can be separated to the QM part and the MM part, and then the potential energy U can be divided into the energy of QM part E_{QM}, the energy of the MM part E_{MM} and the interactional energy between the QM/MM part $E_{\mathrm{QM/MM}}$ [8]:

$$U = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}} \tag{1}$$

For convenient, we use a ration factor a $(0 \le a \le 1)$ to separate the interactional energy $E_{\mathrm{QM/MM}}$, and incorporate them into the QM and MM parts:

$$U = E_{\mathrm{QM}} + aE_{\mathrm{QM/MM}} + E_{\mathrm{MM}} + (1-a)E_{\mathrm{QM/MM}} = E'_{\mathrm{QM}} + E'_{\mathrm{MM}} \tag{2}$$

where $E'_{\mathrm{QM}} = E_{\mathrm{QM}} + aE_{\mathrm{QM/MM}}$ and $E'_{\mathrm{MM}} = E_{\mathrm{MM}} + (1-a)E_{\mathrm{QM/MM}}$. In our previous studies, in order to use a widely distributed temperature range $\{\beta_i\}$ ($\beta_i = 1/k_B T_i$ and k_B is the Boltzmann constant) to selectively enhance the sampling of the QM part of the system but keep the thermodynamics of the MM part of the system at normal temperature, we use a generalized distribution $f'(U)$:

$$f'(U) = e^{-\beta_0 E'_{\mathrm{MM}}} \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}} \tag{3}$$

where $\beta_0 = 1/k_B T_0$, T_0 is the simulation temperature and n_i is the weighting factor at temperature β_i.

Here, we also want to enhance the sampling of the MM part using another temperature range $\{\beta_j\}$. Analogous to the SITS method, it is easily to design a new generalized distribution $f''(U)$:

$$f''(U) = \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}} \cdot \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}} \tag{4}$$

where m_j is the weighting factor at temperature ranges $\{\beta_j\}$. Notice that when $\beta_i = \beta_j = \beta_0$, $f''(U) = \sum_i \sum_j n_i m_j e^{-\beta_0 U} = Ce^{-\beta_0 U} = Cf(U)$ (C is a constant), which means the thermodynamics of the system can be easily recovered to the normal temperature β_0.

Similar to the ITS and SITS method, we assume that an effective system with a potential U^{eff} at the temperature β_0 has the same distribution as the generalized distribution $f''(U)$:

$$e^{-\beta_0 U^{\mathrm{eff}}} = f''(U) = \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}} \cdot \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}} \tag{5}$$

Therefore, the "effective" potential energy U^{eff} is:

$$U^{\mathrm{eff}} = -\frac{1}{\beta_0}\ln \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}} - \frac{1}{\beta_0}\ln \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}} \tag{6}$$

To calculate the derivate of the effective potential energy U^{eff} with the respect to the atomic coordinates $\boldsymbol{r}_{\mathrm{QM}}$ and $\boldsymbol{r}_{\mathrm{MM}}$ of the QM and MM part of the system, we can obtain the force acting on the atoms of the QM part $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{QM}}$ and MM part $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{MM}}$, respectively:

$$\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{QM}} = \frac{\sum_i \beta_i n_i e^{-\beta_i E'_{\mathrm{QM}}}}{\beta_0 \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}}}\left(\boldsymbol{F}_{\mathrm{QM}\to\mathrm{QM}} + a\boldsymbol{F}_{\mathrm{MM}\to\mathrm{QM}}\right) + \frac{\sum_j \beta_j m_j e^{-\beta_j E'_{\mathrm{MM}}}}{\beta_0 \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}}} \cdot (1-a)\boldsymbol{F}_{\mathrm{MM}\to\mathrm{QM}} \tag{7}$$

$$\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{MM}} = \frac{\sum_i \beta_i n_i e^{-\beta_i E'_{\mathrm{QM}}}}{\beta_0 \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}}} \cdot a\boldsymbol{F}_{\mathrm{QM}\to\mathrm{MM}} + \frac{\sum_j \beta_j m_j e^{-\beta_j E'_{\mathrm{MM}}}}{\beta_0 \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}}} \cdot \left[\boldsymbol{F}_{\mathrm{MM}\to\mathrm{MM}} + (1-a)\boldsymbol{F}_{\mathrm{QM}\to\mathrm{MM}}\right] \tag{8}$$

where $\boldsymbol{F}_{\mathrm{QM}\to\mathrm{QM}} = -\partial E_{\mathrm{QM}}/\partial \boldsymbol{r}_{\mathrm{QM}}$, $\boldsymbol{F}_{\mathrm{MM}\to\mathrm{MM}} = -\partial E_{\mathrm{MM}}/\partial \boldsymbol{r}_{\mathrm{MM}}$, $\boldsymbol{F}_{\mathrm{MM}\to\mathrm{QM}} = -\partial E_{\mathrm{QM/MM}}/\partial \boldsymbol{r}_{\mathrm{QM}}$ and $\boldsymbol{F}_{\mathrm{QM}\to\mathrm{MM}} = -\partial E_{\mathrm{QM/MM}}/\partial \boldsymbol{r}_{\mathrm{MM}}$ are the force acting on the QM part from E_{QM}, on the MM part from E_{MM}, on the QM part from $E_{\mathrm{QM/MM}}$ and acting on the MM part from $E_{\mathrm{QM/MM}}$, respectively. Here we set S_{QM} and S_{MM} as two scale factors:

$$S_{\mathrm{QM}} = \frac{\sum_i \beta_i n_i e^{-\beta_i E'_{\mathrm{QM}}}}{\beta_0 \sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}}} \tag{9}$$

$$S_{\mathrm{MM}} = \frac{\sum_j \beta_j m_j e^{-\beta_j E'_{\mathrm{MM}}}}{\beta_0 \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}}} \tag{10}$$

Then the effective forces $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{QM}}$ and $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{MM}}$ can be simplified as:

$$\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{QM}} = S_{\mathrm{MM}}\boldsymbol{F}_{\mathrm{QM}} + \left(S_{\mathrm{QM}} - S_{\mathrm{MM}}\right)\left(\boldsymbol{F}_{\mathrm{QM}\to\mathrm{QM}} + a\boldsymbol{F}_{\mathrm{MM}\to\mathrm{QM}}\right) \tag{11}$$

$$\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{MM}} = S_{\mathrm{MM}}\boldsymbol{F}_{\mathrm{MM}} + \left(S_{\mathrm{QM}} - S_{\mathrm{MM}}\right) \cdot a\boldsymbol{F}_{\mathrm{QM}\to\mathrm{MM}} \tag{12}$$

where $\boldsymbol{F}_{\mathrm{QM}} = -\partial U/\partial \boldsymbol{r}_{\mathrm{QM}} = \boldsymbol{F}_{\mathrm{QM}\to\mathrm{QM}} + \boldsymbol{F}_{\mathrm{MM}\to\mathrm{QM}}$ is the original force acting on the QM part and $\boldsymbol{F}_{\mathrm{MM}} = -\partial U/\partial \boldsymbol{r}_{\mathrm{MM}} = \boldsymbol{F}_{\mathrm{MM}\to\mathrm{MM}} + \boldsymbol{F}_{\mathrm{QM}\to\mathrm{MM}}$ is the original force acting on the MM part.

Using the effective forces $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{QM}}$ and $\boldsymbol{F}^{\mathrm{eff}}_{\mathrm{MM}}$ to perform the MD simulation, the sampling of the system can be enhanced by the multilevel integrated tempering sampling method. Similar to the ITS and SITS method, we can use the reweighting factor c_0 to recover the thermodynamics of the system to the normal temperature β_0:

$$c_0 = \frac{f(U)}{f''(U)} = e^{-\beta_0(U-U^{\mathrm{eff}})} = \frac{e^{-\beta_0(E'_{\mathrm{QM}}+E'_{\mathrm{MM}})}}{\sum_i n_i e^{-\beta_i E'_{\mathrm{QM}}} \sum_j m_j e^{-\beta_j E'_{\mathrm{MM}}}} = \left[\sum_i n_i e^{(\beta_0-\beta_i)E'_{\mathrm{QM}}} \sum_j m_j e^{(\beta_0-\beta_j)E'_{\mathrm{MM}}}\right]^{-1} \quad (13)$$

Unlike the SITS method, as both the QM and MM parts are enhanced in the MITS method, we can calculate the reweighting factors at the other temperatures β_x that in the overlap area of the two widely distributed temperature ranges $\{\beta_i\}$ and $\{\beta_j\}$:

$$c_x = \frac{f_x(U)}{f''(U)} = \left[\sum_i n_i e^{(\beta_x-\beta_i)E'_{\mathrm{QM}}} \sum_j m_j e^{(\beta_x-\beta_j)E'_{\mathrm{MM}}}\right]^{-1}, \qquad \beta_x \in \{\beta_i\} \cap \{\beta_j\} \quad (14)$$

Finally, the two sets of $[n_k]$'s can be obtained using the iteration method as in ITS and SITS method[10].

We note here that, MITS does not prescribe the RCs or alter the TS properties, which allows thorough sampling of the phase space and automatic search for reaction mechanisms. To further optimize the computational efficiency, we distributed the entire sampling into three windows, and the MITS parameters for each window was set differently and individually (see *section III*).

S-III Calculation of Thermodynamics

In pursuit for best efficiency, we adopted the distributive sampling strategy, that is, dividing the entire sampling into three windows (window 1 for reactant, 2 for barrier and 3 for product). For MM-region, the sampling temperature ranges for all three windows are the same from 270 K to 330 K, with 30 intervals. The temperature ranges of QM-part for window 1 and 3 are relatively narrower (220 K to 650 K) and with fewer temperature intervals (k=50 for $[n_k]$ series in Eq. S4, Table S1) provided that the sampling of reactant/product metastable basins converge much more quickly than on the transition region (250 K to 1000 K, with k= 150). In the fulfillment of distributive sampling, an innocuous restraint force ($F(\Omega(\mathbf{R}))$) was introduced to the effective potential of window 2, in a Woods-Saxon' form:

$$F(\Omega(\mathbf{R})) = \frac{\varepsilon}{1 + e^{-2\alpha\varepsilon(r-r_0)}} \quad (15)$$

where $\Omega(\mathbf{R}) = r$ is a set of chosen order parameters that delimit the reactant and product (we chose r=d(C1-C6) and d(O3-C4) in this paper); ε is the maximum restriction force, α describes the steepness of the restriction boundary, r_0 is the defined boundary distance in the restriction, and r is the actual distance between the two restrained atoms. This restraint effectively manifests only when $r > r_0$ and is almost null when $r < r_0$. We set $r_0 = 2.5$ Å, $\alpha = 5.0$ and $\varepsilon = 10.0$ for both d(C1-C6) and d(O3-C4), which guarantees an extensive overlap between the adjacent sampling windows. The corresponding bias potential for window 2 is:

$$V_2(\Omega(\mathbf{R})) = \ln\{\exp[-2\alpha\varepsilon(r - r_0)] + 1\}/2\alpha \tag{16}$$

As discussed in Ref. [11], this restraint does not affect the properties of transition state, guarantees the authenticity and spontaneity of the transition events, and can be independent of the reaction coordinate(s). Therefore, for each window, we obtained a MITS effective potential $\widetilde{U}_i(\mathbf{R})$, and a possible restraint potential $V_i(\mathbf{R})$ ($V_1 = V_2 = 0$, V_2 in the form of Eq. S16). The overall reweighting procedure is performed according to the method proposed by Yang *et al*[12]. If we treat the combined term $\widetilde{U}_i(\mathbf{R}) - U(\mathbf{R}) + V_i(\Omega(\mathbf{R}))$ as the biasing potential, the weighted histogram analysis method (WHAM)[13, 14] can be used to analyze the simulation data. Following the derivation of WHAM, to evaluate the free energy correction to each sampling window, f_i, can be realized through iterative computing of the following two quantities:

$$\rho(\mathbf{R}) = \sum_{i=1}^{3} \frac{m_i \rho_i^b(\mathbf{R}) {g_i}^{-1}}{\sum_{j=1}^{3} m_j {g_j}^{-1} e^{-\beta_0\left(\left(V_j(\mathbf{R}) - f_j\right) + \left(\widetilde{U}_j(\mathbf{R}) - U(\mathbf{R})\right)\right)}} \tag{17}$$

$$e^{-\beta_0 f_k} = \sum_{i=1}^{3} \sum_{l=1}^{m_i} \frac{e^{-\beta_0\left(V_k(\mathbf{R}_{i,l}) + \left(\widetilde{U}_k(\mathbf{R}_{i,l}) - U(\mathbf{R}_{i,l})\right)\right)} {g_i}^{-1}}{\sum_{j=1}^{3} m_j {g_j}^{-1} e^{-\beta_0\left(\left(V_j(\mathbf{R}_{i,l}) - f_j\right) + \left(\widetilde{U}_j(\mathbf{R}_{i,l}) - U(\mathbf{R}_{i,l})\right)\right)}} \tag{18}$$

where $\rho(\mathbf{R})$ is the unbiased probability distribution, $\rho_i^b(\mathbf{R})$ is the biased distribution for *i*-th window, m_i is the number of samples recorded in the *i*-th simulation window, $\mathbf{R}_{i,l}$ denotes the *l*-th sampled configuration in the *i*-th window, and β_0 is the inverse of the targeted reweighting temperature (can be chosen from anywhere between 280 K to 320 K, as mentioned in above sections). The statistical inefficiency term[13], g_i, is given by $g_i = 1 + 2\tau_i$, where τ_i is the integrated autocorrelation time (IACT)[14] of window *i* (in units of the simulation frame time step). This term assigns lower weights to histograms with longer autocorrelations (in this study $\tau_2 \sim 4000\tau_1$ and $\tau_1 \approx \tau_3$, partly because the samples in widow 2 were recorded in every femtosecond, whereas those in windows 1 and 3 were both recorded in every picosecond). After converged f_i were obtained, the unbiased distribution $\rho(\mathbf{R})$ follows Eq. S17, and the distribution of other quantities can be derived from $\rho(\mathbf{R})$, for instance, the reweighted probability distribution along a collective variable $\Theta(\mathbf{R})$:

$$\rho(\theta) = \int \rho(\mathbf{R}) \delta(\Theta(\mathbf{R}) - \theta) \mathrm{d}\mathbf{R} \tag{19}$$

which can be conveniently evaluated with δ denoting Dirac delta function. The corresponding PMF ($G(\theta)$) along $\Theta(\mathbf{R})$ is simply as:

$$G(\theta) = -\frac{1}{\beta_0}\ln \rho(\theta) \tag{20}$$

Alternatively, one can make proper use of other existing reweighing approaches, e.g. MBAR[15] or TRAM[16] to retrieve the thermodynamics as well.

To calculate the reaction entropy ΔS_{r}, we plotted ΔG_{r} as a function of temperature *T* (Fig. S6B), of which the slope is the $-\Delta S_{\mathrm{r}}$ according to:

$$\left(\frac{\partial \Delta G_{\mathrm{r}}}{\partial T}\right)_p = -\Delta S_{\mathrm{r}} \tag{21}$$

While the reaction enthalpy $\Delta H_{\mathrm{r}}(T_0)$ can be obtained as the slope in the plot of $\Delta G_{\mathrm{r}}/T$-$1/T$ (Fig. S6A) regarding to the Gibbs-Helmholtz equation:

$$\frac{\partial\left(\frac{\Delta G_{\mathrm{r}}}{T}\right)_p}{\partial\left(\frac{1}{T}\right)} = \Delta H_{\mathrm{r}} \tag{22}$$

S-IV Enhanced Sampling of Reactive Trajectories and Calculation of Kinetics

According to fluctuation-dissipation theorem (cf. SI Ref 17), for a single-step rapid transition from state *A* to state *B*, the rate constant can be calculated with a time-dependent function, $k(t) = Z_{AB}(t)/Z_A = \sum_{\text{sus traj}} w_i / \left(\sum_{\text{all traj}} w_i\right)$ (where $Z_{AB}(t)$ is the time correlation function, cf. SI Ref 18, while Z_A is the partition function for state A), which is the probability of finding the system in state *B* a time *t* after it was in state *A*. If *A* and *B* are separated by a single barrier, this probability will increase with an exponential time dependence. $k(t)$ grows linearly in time if *t* is longer than the typical transition time over the barrier (denoted by τ_{mol}), and smaller than the exponential relaxation time (denoted by $\tau_{\mathrm{rxn}} = 1/(k_1 + k_{-1})$). In our case, τ_{mol}<200 fs, so we set *t* to be 2 ps, which is conventionally used in transition path samplings and well satisfies the above conditions.

In reference of the transition path sampling (TPS), the rate constant calculation of chemical reactions requires shooting from the initial phase space points representing the Boltzmann distribution of the reactant and collecting the successful conversions[18]. Based on the reactive-flux-over-population theory, the ratio between the number of successful conversions and that of all trajectories within a given length of time (*t* = 2 ps) can be used to calculate the desired rate constant:

$$k_1 = \frac{\sum_{\text{sus traj,forward}} w_i}{t\left(\sum_{\text{sus traj,forward}} w_i + \sum_{\text{failed traj,forward}} w_j\right)} \tag{23}$$

$$k_{-1} = \frac{\sum_{\text{sus traj,backward}} w_i}{t\left(\sum_{\text{sus traj,backward}} w_i + \sum_{\text{failed traj,backward}} w_j\right)} \tag{24}$$

In Eqs. S23 and S24, w_i is the weighting factor for each initial configuration. Different from the traditional TPS, ESoRT is more efficient since the sampled configurations are biased towards the transition region by MITS, and reside on different spontaneous reactive paths. Out of these sampled configurations, a set of initial structures (with recorded atomic coordinates, velocities, charges, as well as the effective weighting factors) were randomly selected. The neighboring points were assured to be at least 0.1 ps apart in the MITS trajectories to reduce the correlation between the points generated by the same reactive trajectory. With these selected initial configurations, tens of thousands of MD trajectories, each consisting of a pair of forward and backward (for both t=2 ps), were obtained on the original potential energy surface under an NVE ensemble. These trajectories were then categorized based on their destination: a successful transition if the forward and backward-bound trajectories end in different metastable basins, otherwise an unsuccessful attempt. Those successfully converted trajectories will contribute to the rate constant corresponding to its forward direction. While those unsuccessful trajectories will be counted in the denominator in Eqs. S23 and S24 according to which basin they end in. We also tested the linear dependence of $k(t)$ on different choices of *t* (*t*= 1.2 ps, 1.5 ps and 2.0 ps, respectively). The results show that $k(t)/t$ approximates constant within statistical errors (Table S5).

Considering that along one reactive trajectory, the "active" configurations were much rarer than those "inactive" ones, which meant a uniform time interval between selected points might lack of efficiency due to the consequently redundant non-reactive shootings. Some efforts could be taken to enhance the efficiency of the transition path shootings. Our strategy was as follows. A Gaussian selecting bias, denoted as $P_s(i)$, was assigned to the *i*-th configuration, according to the *d*(C1-C6) and *d*(O3-C4):

$$\frac{1}{P_s(i)} = \exp\left(-\frac{(d(\mathrm{C1}-\mathrm{C6})-\mu_1)^2}{\sigma_1^2}\right)\exp\left(-\frac{(d(\mathrm{O3}-\mathrm{C4})-\mu_2)^2}{\sigma_2^2}\right) \tag{25}$$

where μ_1, μ_2, σ_1, σ_2 are constant parameters which estimate the according bond lengths of TS. This P_s means that we picked the configurations more frequently within the proximity of the TS, and should be corrected in the calculation of kinetics by Eq. S26:

$$k = \frac{\sum_{\text{sus traj}} w_i \cdot P_s(i)}{t\left(\sum_{\text{sus traj}} w_i \cdot P_s(i) + \sum_{\text{failed traj}} w_j \cdot P_s(j)\right)} \tag{26}$$

Noteworthy, the independently calculated reaction rates can be linked to equilibrium constant and free-energy change of reaction through Eqs. S27 (where *R* is gas constant, *T* the temperature and *K* the thermodynamics equilibrium constant), which can

be used to validate the methodological consistency of thermodynamics and kinetics calculations:

$$\Delta G_{\mathrm{r}} = -RT\ln K = -RT\ln \frac{k_1}{k_{-1}} \tag{27}$$

One can check the convergence of trajectory sampling simply by projecting the transition path ensemble to the 2D(*d*(C1-C6) and *d*(O3-C4))-heat map and comparing it with the thermodynamics calculations (Fig. S7).

In the calculation of activation enthalpy, we need to compute the kinetics under different temperatures. Since the momentum will re-distribute over different temperatures, the $P_{\mathrm{s}}(i)$ also accounts for a correction for kinetic energy re-distribution, by assuming kinetic energy obeys approximately the Maxwell-Boltzmann distribution:

$$f_E(E_i^K;T) \propto \exp\left[-\frac{\left(E_i^K - \langle E^K\rangle_{T_0}\right)}{k_B T}\right] \tag{28}$$

$$\frac{1}{P_{\mathrm{s}}(i)} \propto f_E(E_i^K;T)/f_E(E_i^K;T_0) \tag{29}$$

where the E_i^K stands for the kinetic energy of the *i*-th configuration, $\langle E^K\rangle_{T_0}$ is the average kinetic energy at temperature T_0, and $f_E(E_i^K;T)$ is the probability density of the kinetic energy E_i^K under temperature *T*.

The activation enthalpy can then be obtained through transition-state theory (Eq. S30) by plotting the rate constant (*k*) against inverse temperature (Fig. S8), where the slope gives $\Delta H^{\ddagger}$:

$$\ln\left(k \cdot h/k_{\mathrm{B}}T\right) \approx -\frac{\Delta H^{\ddagger}}{RT} + \mathrm{Const.} \tag{30}$$

S-V Defining Solvent Coordinates and Transition Paths

We defined several solvent coordinates as follows that can be used to optimize the reaction coordinate and investigate the solvation effects:

1. Exclusion volume (or solvent-excluded cavity volume). The simulation box around the solute (a radius of 10 Å around the center of mass of the solute) was meshed by 0.2×0.2×0.2 Å^3 cubicles. The solvation cavity of the solute was defined as those cubic cells for which the inserting a 2.9-Å-in-radius spherical probe into the center of the cell did not overlap with any water oxygen atom centers[19]. We chose such a radius for the probe so as to avoid the zero overlap insertions in the bulk water[20].

2. Dipole orientation of the solvation shell. We defined the solvation shell as water molecules (water oxygens) within 3-Å-reach of any solute atoms. Given that the dipole-dipole interaction is inversely proportional to the distance (to the third power)

between the two dipoles:

$$V_{\text{dipole-dipole}} \propto -\frac{\mu_1\mu_2}{4\pi r_{12}^3}\cos\theta_{12} \tag{31}$$

(where μ is the strength of the dipole, r_{12} is the distance between two dipole centers, and θ_{12} is the relative angle between the two dipoles), we thus defined each dipole orientation of the solvation water relative to the solute dipole by $\cos\theta_i$, and averaged them regarding their distance to the solute (r_i):

$$\langle\cos\theta\rangle = \frac{\sum \cos\theta_i / r_i^3}{\sum 1/r_i^3} \tag{32}$$

3. Hydrogen-bond length, was defined as the distance between the nearest water hydrogen (Hw, "w" stands for "water") and the O3 atom of reactant, provided that the Ow-Hw-O3 angle is between 120 and 180 degrees.

We also selected the transition path regions out of every successfully reacted trajectory, and defined them as the "transition paths", in order to construct the transition path probability *p*(*r*|TP) (see *section VI*) and calculate the average transition path length (Fig. S9). The transition paths (which were shot both forward and backward from near the separatrix) were terminated once they arrived at the reactant or product basin. We judged the terminus as follows: firstly, project the entire trajectory onto the 2D (*d*(C1-C6) and *d*(O3-C4) map as Fig. 4A in the main text; secondly, divide the entire 2D map into uniform square grids with dimension 0.1×0.1 Å^2 and count the local configurational density in each grid; then, "grow" the transition path from the initial shooting point forward and backward grid by grid until the grid density jumps abruptly. This method works well in that the grids located within the product/reactant basin (Fig. 4A, red and blue symbols) are much more configurationally dense than transition path region (Fig. 4A, green symbols). The final truncated length is termed as transition path length, denoted as τ throughout this paper. According to Hummer, each such transition contributes to the statistics of transition path lengths with a relative weight, *w*:

$$w = \left(\sum_{\text{intersections } k} |v_k|^{-1}\right)^{-1} \tag{33}$$

where the sum is over the points of intersection of the trajectory with the separatrix (or dividing surface, chosen to be *d*(C1-C6)=1.84 Å here), and $v_k = \Delta d(\text{C1}-\text{C6})/\Delta t$ is the velocity normal to the separatrix at the *k*th intersection. This factor is used to correct the over-/underrepresentation of "slow" and "fast" trajectories when calculate the distribution of transition path lengths (Fig. S9)

S-VI Bayesian Learning of Reaction Coordinate(s)

According to Best and Hummer[21], a Bayesian probabilistic relation between the equilibrium and tranistion-path ensembles can be used to define and identify a transition-state ensemble :

$$p(\mathrm{TP}|\mathbf{x})p_{\mathrm{eq}}(\mathbf{x}) = p(\mathbf{x}|\mathrm{TP})p(\mathrm{TP}) \tag{34}$$

where four probability densities are introduced : $p_{\mathrm{eq}}(\mathbf{x})$ and $p(\mathbf{x}|\mathrm{TP})$ are the probability distributions in phase space for the equilibrium emsemble and transition paths, respectively. $p(\mathrm{TP})$ is the probability of a phase-space point being on a transition path, which is a rate-related constant. $p(\mathrm{TP}|\mathbf{x})$ is the probability for being on a transition path given that the system is in configuration **x**. Once a reaction coordinate (RC) $r = r(\mathbf{x})$ is chosen, Eq. S34 is then generalized for the projected dynamics introduced in the main text:

$$p(\mathrm{TP}|r)p_{\mathrm{eq}}(r) = p(r|\mathrm{TP})p(\mathrm{TP}) \tag{35}$$

The shape and maximum value of $p(\mathrm{TP}|r)$ then serve as an indicator of the quality of the chosen RC *r*. During the optimization of *r*, we chose linear combination of different RC candidates (denoted as vector $\boldsymbol{R}$), associated with a linear weight coefficient vector $\boldsymbol{\alpha}$:

$$r(\boldsymbol{R};\boldsymbol{\alpha}) = \sum_i \alpha_i R_i \tag{36}$$

$$\sum_i |\alpha_i| = 1 \tag{37}$$

After choosing a set of $[\alpha_i]$, with the constraint that the absolute values of which sum up to 1, two probability densities, $p(r|\mathrm{TP})$ and $p_{\mathrm{eq}}(r)$, can be calculated along this *r*, and then smoothed by a cubic interpolation procedure. We randomized the choice of $\boldsymbol{\alpha}$, and reserved those results which guaranteed an ascending $p(\mathrm{TP}|r)$ value. We also removed the R_i from $\boldsymbol{R}$ vector if $|\alpha_i|$ is less than one percent out of the concern that a tiny α_i might cause numerical instatibilty during the calculation of $p(\mathrm{TP}|r)$.

It should be noted here that $p(\mathrm{TP})$ is identical for any arbitrary choice of *r*, so it can be evaluated or estimated according to the barrier height of the thermodynamics free-energy profile. For the reaction under invetsigation, we were only able to roughly estimate the order of magnitude of $p(\mathrm{TP})$ (10^{-20} to 10^{-19}) from Fig. 1B in the main text, provided that the barrier is too sharp and

the corresponding peak value is too small when translated into probability. Such a small-valued $p(\mathrm{TP})$ marks the rarity of chemical reactions: intuitively, there are on average only thousands of molecules undergoing transition in a reactor containing one mole of molecules at one moment. Also noteworthy, $p(\mathrm{TP})$ is related to the reaction rate coefficients for the two-state reversible reaction[21]:

$$\frac{p(\mathrm{TP})}{\langle \tau_{\mathrm{TP}} \rangle} \approx \frac{2}{k_1^{-1} + k_{-1}^{-1}} \tag{38}$$

This formula links the thermodynamics ($p(\mathrm{TP})$), kinetics (k_1 and k_{-1}) and dynamics (τ_{TP}) together, so we can employ it to test the self-consistency of our sampling protocal and the calculation results. Based on our computation, $p(\mathrm{TP})$ ≈10^{-20}~10^{-19} and $\langle \tau_{\mathrm{TP}} \rangle$ =88 fs ≈10^{-13} s, so the left hand side of Eq. S38 approximates 10^{-7}~10^{-6} s^{-1}. While $k_1^{-1} + k_{-1}^{-1} \approx k_1^{-1} =$ 0.5×10^{-7} s, so the right hand side is also about 10^{-7}~10^{-6} s^{-1}, suggesting our calculation of dynamics, kinetics and thermodynamics are in very good agreement.

S-VII Recap of Kramers' Theory and Calculation of the Transmission Coefficient

The transmission coefficient in TST (κ_{TST}), after choosing a certain RC, can be evaluated through:

$$\kappa_{\mathrm{TST}} = \frac{k_{\mathrm{abs}}}{k_{\mathrm{TST}}} \tag{39}$$

$$k_{\mathrm{TST}} = \frac{k_{\mathrm{B}}T}{h} \exp(-\Delta^{\ddagger} G/RT) \tag{40}$$

Where k_{abs} is the rate constant calculated using RC-free flux-over-population theory (Eq in the main text), while k_{TST} is based on the transition-state approximation along a chosen RC (here we used the optimized RC, namely r_0). For the forward reaction, $\Delta G_1^{\ddagger} \approx 25$ kcal/mol and $k_{\mathrm{abs}} = 2.0 \times 10^{-7}$ s^{-1}, so the $\kappa_{\mathrm{TST,forward}} \approx 0.05$. In the same fashion, $\kappa_{\mathrm{TST,backward}}$ also estimates to be 0.5, given $\Delta G_{-1}^{\ddagger} \approx 19$ kcal/mol and $k_{\mathrm{abs}} = 4.3 \times 10^{-3}$ s^{-1}, hence so the $\kappa_{\mathrm{TST,backward}} \approx 0.05$.

In Kramers' theory, chemical reaction in condensed phase can be regarded as a thermally activated barrier crossing process along a well-chosen one-dimensional reaction coordinate. The interactions (classically, the collisions) with solvent are treated as a static friction which hinders the system progressing along the RC. There are two commonly used formal expressions for rate constant under different conditions. For the weak friction limit, known as the energy-diffusion-limit, the reaction rate increases with slightly larger friction[22, 23]. While in the other extreme, with overdamped friction, known as the spatial-diffusion-limit, the reaction rate becomes inversely proportional to the friction coefficient[22, 23]. The general form of Kramers' expression for reaction rate (with

moderate-to-strong friction) is:

$$k_{\mathrm{Kramers}} = k_{\mathrm{TST}} \left\{ \left[\left(\frac{\xi}{2\omega_{\mathrm{b}}} \right)^2 + 1 \right]^{\frac{1}{2}} - \frac{\xi}{2\omega_{\mathrm{b}}} \right\} \approx \frac{\omega_{\mathrm{b}}}{\xi} k_{\mathrm{TST}} \tag{41}$$

k_{Kramers} and k_{TST} stand for the theoretical rate constant derived from Kramers' theory and TST respectively; the friction damping frequency is denoted as ξ. It is easy to show that k_{Kramers} is smaller than k_{TST}. In other words, the TST puts an upper limit to the reaction rate, and always overestimates the rate coefficient, which should be corrected by a "transmission coefficient". To this end, Kramers' theory can be used to estimate the transmission coefficient once we know the friction damping rate ξ.

From Stokes-Einstein relation we know that the friction damping rate can be calculated from the barrier diffusion constant D^*:

$$\xi = \frac{k_{\mathrm{B}} T}{D^* M} \tag{42}$$

where k_{B} is Boltzmann constant, T denotes the temperature; M is the effective mass of the reactive particle, here we adopted the effective mass of the C1-C6 bond, namely 24 g/mol. As discussed in the main text, with the barrier diffusion constant $D^* \sim 7 \times 10^{-6}$ $\mathrm{cm^2 \cdot s^{-1}}$, therefore, we obtained that $\xi \sim 3 \times 10^{14}$ $\mathrm{s^{-1}}$. The transmission coefficient (denoted as $\kappa_{\mathrm{Kramers}}$ in the main text) can be further calculated through Eq. 3 in the main text.

S-VIII Inhomogeneous Reaction Mechanisms and Spectral Clustering of Transition Paths

Aiming to examine the heterogeneity of the transition path ensemble, we applied the spectral clustering algorithm described in Ref.[24]. This algorithm presents a general framework for time series clustering based on spectral decomposition of the similarity matrix, and enjoys several merits in dealing with our concerned problems[24, 25]: i) it is compatible with dimensionality reduction techniques (to define the pairwise similarities) so the high dimensionality of time vectors will not affect the clustering efficiency; ii) once with proper similarity metric, it can be used to clustering time vectors with arbitrary length; iii) most importantly, it can automatically determine the optimal cluster number.

The pairwise similarity between two transition paths is defined based on the Longest Common Subsequence (LCSS)[26]. It is

non-metric similarity function, though, one still needs a quantitative measure of the "coordinate" of each data point (i.e. the molecular configuration) along the time series. Provided that the Bayesian-learning algorithm already resulted in a dimensionality reduction of RCs, so we naturally take advantage of this result to calculate the Euclidean distance $dp(i,j)$ between two data points *i*,*j* along different trajectories in the form:

$$dp(i,j) = \sqrt{\alpha\left(d_i(\mathrm{C1C6}) - d_j(\mathrm{C1C6})\right)^2 + \beta\left(d_i(\mathrm{O3C4}) - d_j(\mathrm{O3C4})\right)^2} \tag{43}$$

where α, β correspond to the coefficients of *d*(C1-C6) and *d*(O3-C4) present in the optimized RC, respectively. Then the LCSS between two transition paths, denoted as c_{ij} (Fig. S14), can be computed using dynamic programming algorithm [26], and the distance matrix element d_{ij} between two transition paths *i*, *j* (containing n_i and n_j data points, respectively) adopts the following form[27]:

$$d_{ij} = \frac{n_i + n_j - 2c_{ij}}{2\sqrt{n_i n_j}} \tag{44}$$

In this form, d_{ij} (the distance matrix element, Fig. S15) not only distances transition paths with less common sequence length (c_{ij}), but also distinguishes the two transition paths in terms of the overall length. The similarity matrix $\mathbf{S}_{ij}$ (or affinity matrix)[24] is constructed as:

$$\mathbf{S}_{ij} = \exp\left(-\frac{d_{ij}}{2\sigma^2}\right) \tag{45}$$

where the parameter σ^2 can be determined with a gradient ascent method[24, 28]. A normalization procedure is performed on matrix $\mathbf{S}$ (*M*×*M*) before estimating the optimal cluster number:

$$\mathbf{S}' = \mathbf{U}^{-1}\mathbf{S} \tag{46}$$

$$\mathbf{U} = diag\ (u_{11}, u_{11}, \dots, u_{MM}) \tag{47}$$

$$u_{ii} = \sum_{j=1}^{M} \mathbf{S}_{ij} \tag{48}$$

Ideally, $\mathbf{S}_{ij} = \mathbf{S'}_{ij} = 0$ if trajectories *i* and *j* belong to different clusters, then the normalized similarity matrix $\mathbf{S}'$ will become block-diagonal (if we assume that the data objects are ordered by clusters). It also follows that each diagonal block in $\mathbf{S}'$ has the largest eigenvalue 1[24]. Practically, given that the similarity matrix may not be rigorously block-diagonal, one can choose the

number of eigenvalues closest to 1 by defining a threshold $\delta \in [0,1]$ and counting the eigenvalues λ_i that satisfy:

$$|\lambda_i - 1| < \delta \tag{49}$$

After the spectral decomposition is done, and the optimal number of cluster, say *K*, is obtained, one can treat each row of the final eigenvectors (corresponding to the close-to-unity eigenvalues) as a new point in $\mathbb{R}^K$ space and cluster them into *K* clusters via KMeans algorithm[29].

For reactions in water, we identified *K*=2 with threshold $\delta = 0.1$, which means there exist two categories of transition paths, which can be clearly seen from the large gap in the eigenvalue spectrum (Fig. S16). We also note here that the clustering result is robust against different choices of similarity metrics. For instance, one can define the "transition velocity" for each transition trajectory, that is, the time spent on crossing the transition-state barriers. Two barriers were well defined for *d*(C1-C6) and *d*(O3-C4), respectively, hence we can assign two such velocities ($v(\mathrm{C1C6})$ and $v(\mathrm{O3C4})$) to each transition path, and define an Euclidean distance between two trajectories by $\sqrt{\left(v_i(\mathrm{C1C6}) - v_j(\mathrm{C1C6})\right)^2 + \left(v_i(\mathrm{O3C4}) - v_j(\mathrm{O3C4})\right)^2}$. Using this identity, we again clustered transition paths using the method introduced above, also into two clusters. The overlap between the two clustering results was quantitatively assessed. Let $\mathbf{x}$ denotes a time vector, $R_i(\mathbf{x})$ denote the clusters found with LCSS similarity, and $V_j(\mathbf{x})$ for those found with "transition velocity" metric, then we can calculate the clustering overlap η as:

$$\eta = \frac{\sum_{\mathbf{x}} I(R_i(\mathbf{x}) = V_j(\mathbf{x}))}{M} \tag{50}$$

where $I(\cdot)$ is the indicator function, $R_i(\mathbf{x})$ is the label of the cluster which $\mathbf{x}$ belongs to under the LCSS metric, $V_i(\mathbf{x})$ is the cluster label which $\mathbf{x}$ belongs to under transition-velocity metric, and *M* is the size of the dataset. Under very different similarity metrics, the two clustering results yet share an overlap over 85%, thus showing the robustness of the clustering results and solidifying the discovery of the inhomogeneity of the transition paths.

The transition paths in Cluster 1 are relatively short, and cross the reaction barrier bypassing the "concerted" transition state, and connect to the product basin mainly from entrance/exit 1 (Fig. S18). In contrast, Cluster 2 consist of long trajectories that deviate far from the concerted transition state, and contribute mainly to the other entrance/exit of the product basin. According to the correlation analysis, the hydrogen bond formed between water and the reactant is likely to be the cause of the non-concertedness observed in Cluster 2: shorter hydrogen bond is formed, accompanied by more negative charge enriched at O3 atom (Fig. S19).

In terms of kinetics, the two clusters contribute differently to the forward and backward reaction rates. As shown in Table S2,

forward reaction prefers transition paths in Cluster1, while trajectories in Cluster 2 mainly contribute to backward reaction. Such mechanistic asymmetry of bi-directional reaction is an evidence for the non-Markovian solvation effect. To be specific, the transition event is very fast compared to the relaxation of solute-solvent interactions including hydrogen bonds. Therefore, the solvation state is determined by the initial configuration upon entering the transition passage and "frozen" for the following transition. On the other hand, the occupancy of hydrogen bond between water and the product (aldehyde) is higher than that for the ether reactant (Fig. S20), so the reverse/backward reaction (starting from the aldehyde) is more probable to be accompanied by a well-established hydrogen bond during transition, that is, to undergo a mechanism shared in Cluster 2.

S-IX Fourier Analysis of Characteristic Reaction Modes: Resonance and Dephasing

Several periodicity and vibrational modes were observed for the reaction, including (peak values of) the distribution of transition path length, the C1-C6 and O3-C4 bond lengths, and the hydrogen bond length. The (discrete) fast Fourier transform was then employed in order to identify the characteristic frequency for each observed periodicity, and further check whether there exists resonance. Consider we have a series of real (can be complex in more general cases) signals $[x_n]$ (in this paper, autocorrelation strength or relative probability, both dimensionless), the discrete Fourier transform (DFT) of which is defined by the formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1. \tag{51}$$

And the corresponding frequency for X_k is k/N, which depends on the series length N. The absolute amplitude of X_k is used to represent the strength at each discrete frequency.

The probability density function (PDF) of transition path length distribution, the autocorrelation functions ($C(\tau)$) of *d*(C1-C6), *d*(O3-C4) and *d*(O3-Hw) were chosen to be the input $[x_n]$ signals (Fig. S22) in the time domain:

$$C(\tau) = \frac{\langle (d(t+\tau) - \langle d \rangle)(d(t) - \langle d \rangle) \rangle}{\sigma^2(d)} \tag{52}$$

FFT analysis showed that the observed periodicity in the PDF of the transition path length is closely related to the vibrational frequency of *d*(C1-C6) (Fig. 4B inset in the main text), *d*(O3-C4) (Fig. S22B). Noteworthy, the damping rate of the periodic peak

values of transition path length distribution in Cluster 2 can be well fitted in a single exponential manner, and the damping rate, ~0.024 fs^{-1} (Fig. S24), also echoes with the above-mentioned resonant vibrations. Moreover, a unique vibrational resonance was identified between the *d*(C1-C6), *d*(O3-C4) and *d*(O3-Hw). However, such resonance is absent for the reactant or product regions (Fig. S23). As expected, the product and reactant regions possess distinct vibrational frequency spectrums, but both of them are different from the transition paths. This observation implies that the resonant vibrational modes may be unique and necessary for the chemical transition event.

S-X Re-crossings and Transition Failure

Since the resonance between hydrogen bonding and breaking/forming bonds are characteristic of successful transition events, we postulate that transient dephasing between the hydrogen bonding and bond breakage/formation is likely to cause transition failure and re-crossing events. As illustrated in Fig. S25, when C1-C6 bond is breaking (in a sense it has already broken regarding the defined transition state) but the hydrogen bond is vanishing, that is, the reactant suffers a dynamic loss of hydrogen bond when crossing the barrier, and the re-crossing event did take place. Actually, the correlation between the hydrogen bond and *d*(C1-C6) or *d*(O3-C4) is much weaker for those unsuccessful trajectories (Table S4). Additionally, the coefficient variance (Fig. S26) of the hydrogen-bond length for those trajectories failed to convert is significantly higher than the successfully transited ones.

Additionally, comparing pre-exponential factors with experiments is quite important and very intriguing. However, it is rather difficult since available amount experimental data is limited. On the other hand, transition-state theory is often assumed when the Arrhenius plot is drawn experimentally, thus one cannot directly tell apart the transmission coefficient from the activation entropy. In spite of these difficulties, we referred to the experimentally reported activation enthalpy (ΔE_{exp}) and reaction rate (k_{exp}) under neat condition[30], and tried to estimate the pre-exponential factor (denoted by *A*) for the forward reaction accordingly, $A = k_{\mathrm{exp}} \exp(\Delta E_{\mathrm{exp}}/RT) \approx 4 \times 10^{11}\ \mathrm{s}^{-1}$, smaller than the TST frequency $k_{\mathrm{B}}T/h = 6.2 \times 10^{12}\ \mathrm{s}^{-1}$.

Supplementary Figures

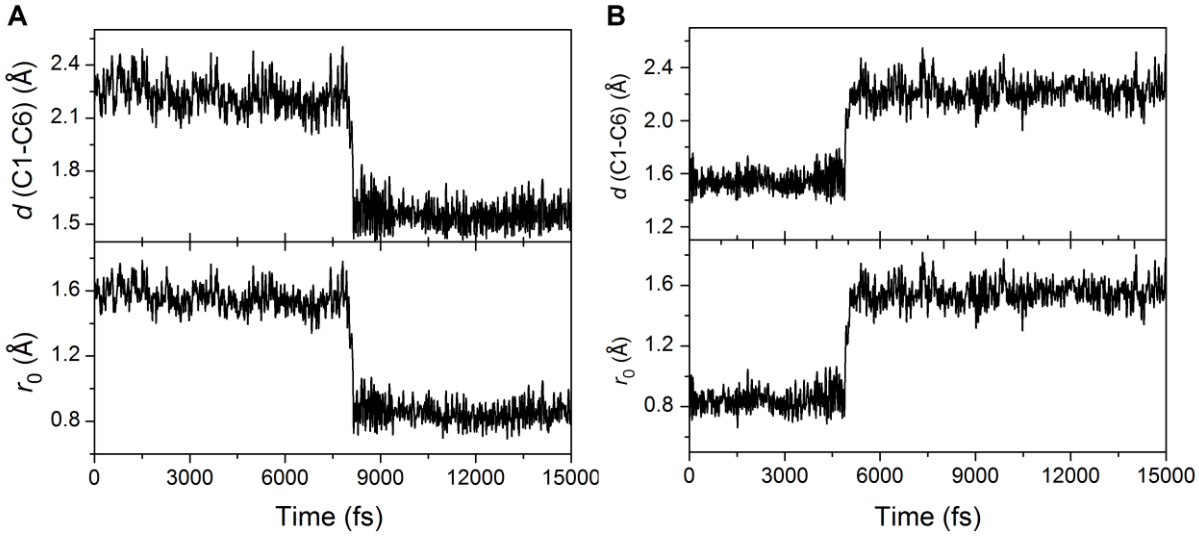

Fig. S1. Reactive trajectories achieved by MITS (window 2). (A) The forward reaction, and (B) the backward reaction are projected on different reaction coordinates: (top) C1-C6 distance (c.f. Fig. 1 in the main text for atom indices), and (bottom) optimized coordinate r_0=0.82*d*(C1-C6)-0.18*d*(O3-C4).

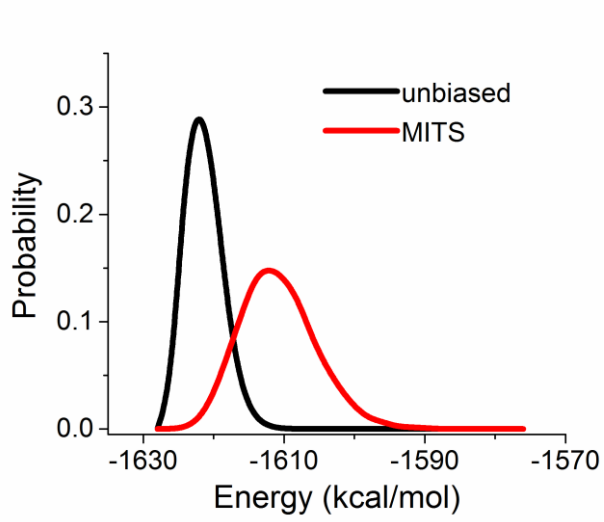

Fig. S2. The broadened distribution of QM-potential energy achieved by MITS (red), and the reweighted Boltzmann distribution (black).

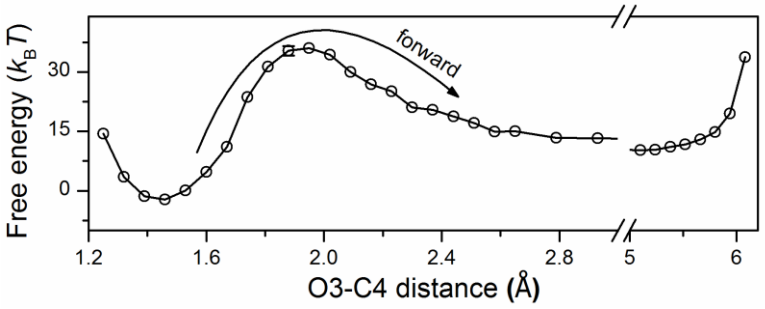

Fig. S3. Free-energy profile along the O3-C4 bond length (or relative atomic distance). The arrow indicates the direction of the forward reaction.

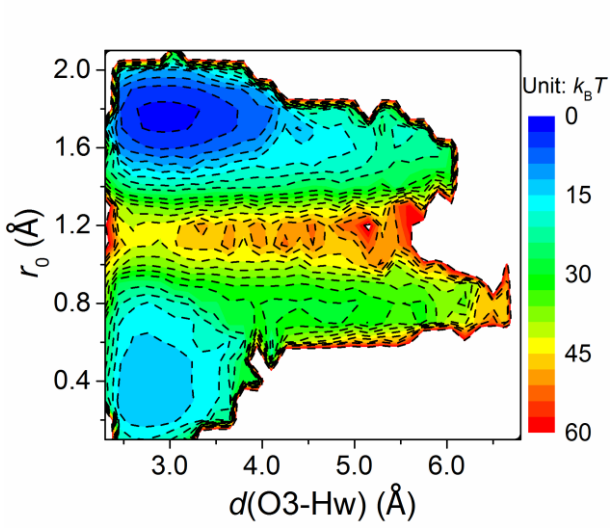

Fig. S4. 2D free-energy projection: (horizontal axis) the hydrogen-bond length between water hydrogen (Hw) and O3 of the reactant; (vertical axis) the optimized RC r_0=0.82d(C1-C6)-0.18d(O3-C4).

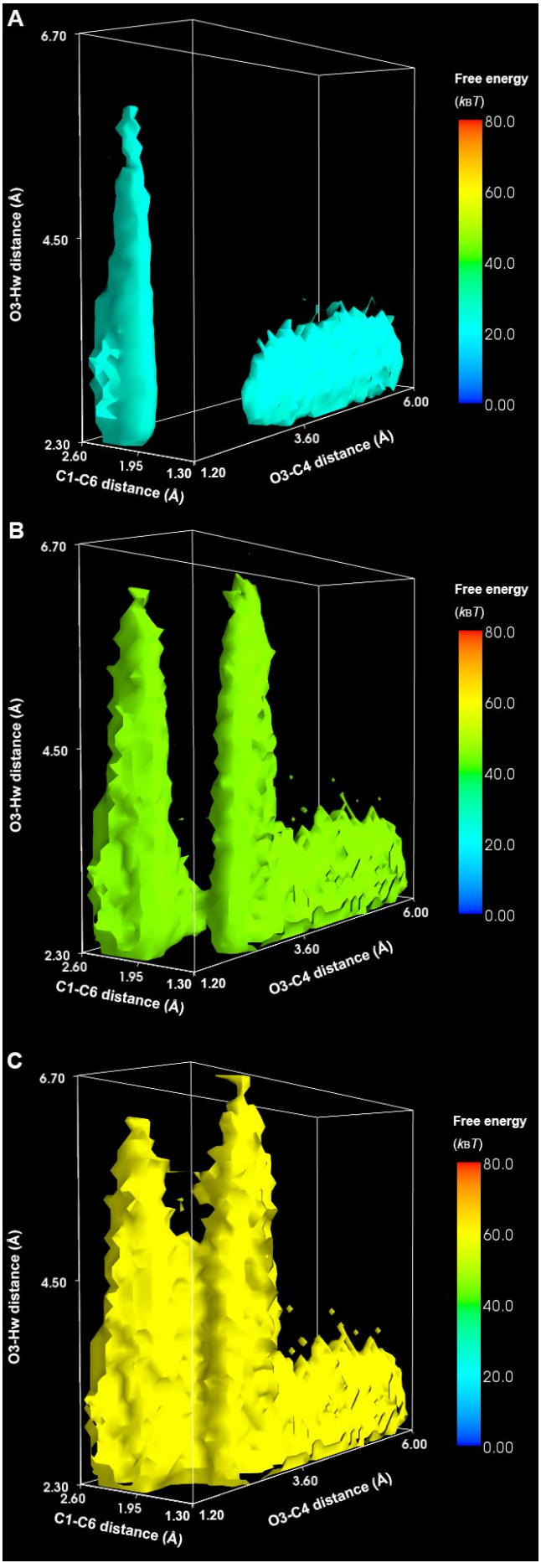
A
6.70
4.50
2.30
O3-Hw distance (Å)
2.60
1.95
1.30
C1-C6 distance (Å)
1.20
3.60
6.00
O3-C4 distance (Å)
Free energy
(kBT)
80.0
60.0
40.0
20.0
0.00
B
6.70
4.50
2.30
O3-Hw distance (Å)
2.60
1.95
1.30
C1-C6 distance (Å)
1.20
3.60
6.00
O3-C4 distance (Å)
Free energy
(kBT)
80.0
60.0
40.0
20.0
0.00
C
6.70
4.50
2.30
O3-Hw distance (Å)
2.60
1.95
1.30
C1-C6 distance (Å)
1.20
3.60
6.00
O3-C4 distance (Å)
Free energy
(kBT)
80.0
60.0
40.0
20.0
0.00

Fig. S5. Free-energy isovalue surfaces in 3D space (d(C1-C6), d(O3-C4) and d(O3-Hw)): with isovalue 25 k_BT (A), 45 k_BT (B) and 60 k_BT (C). The transition passages with lower free-energy is characterized by with short hydrogen-bond length (B).

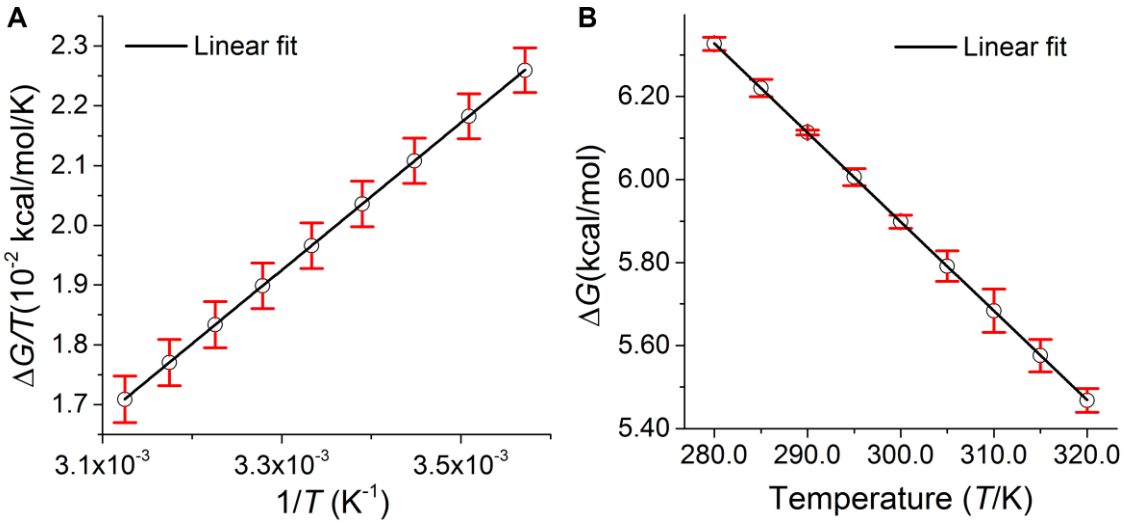

Fig. S6. Calculation of reaction thermodynamics. (A) Reaction enthalpy ΔH_r is obtained as the linear slope according to Gibbs-Helmholtz equation. (B) Reaction entropy ΔS_r is obtained as the negative slope.

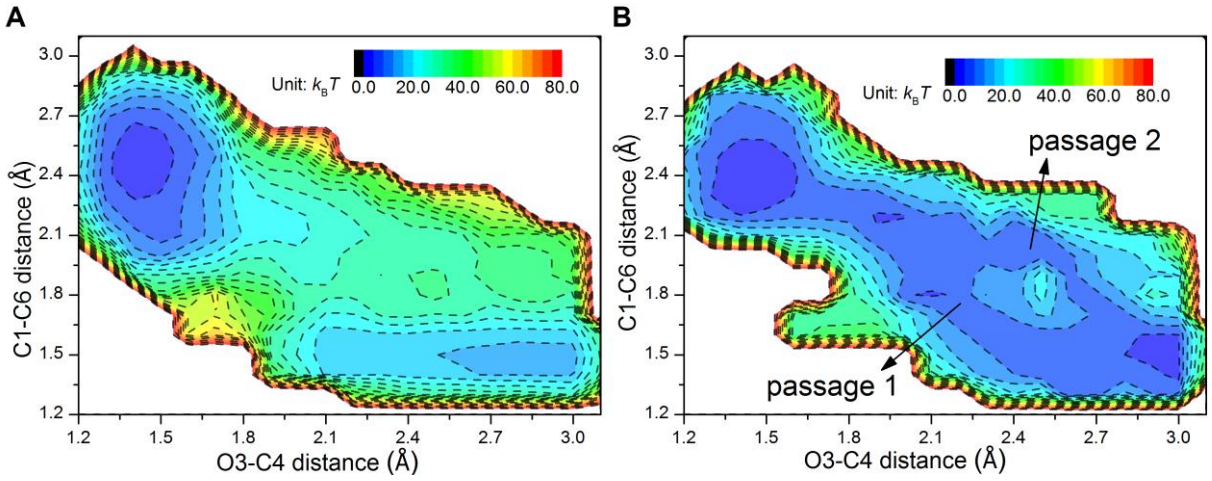

Fig. S7. Free-energy projection of transition path ensembles onto the 2D map. (A) Trajectories of all the attempted shootings

(both successful and unsuccesful ones). (B) Succeeful transition paths, and two passages connecting the product basin with the TS region were identified.

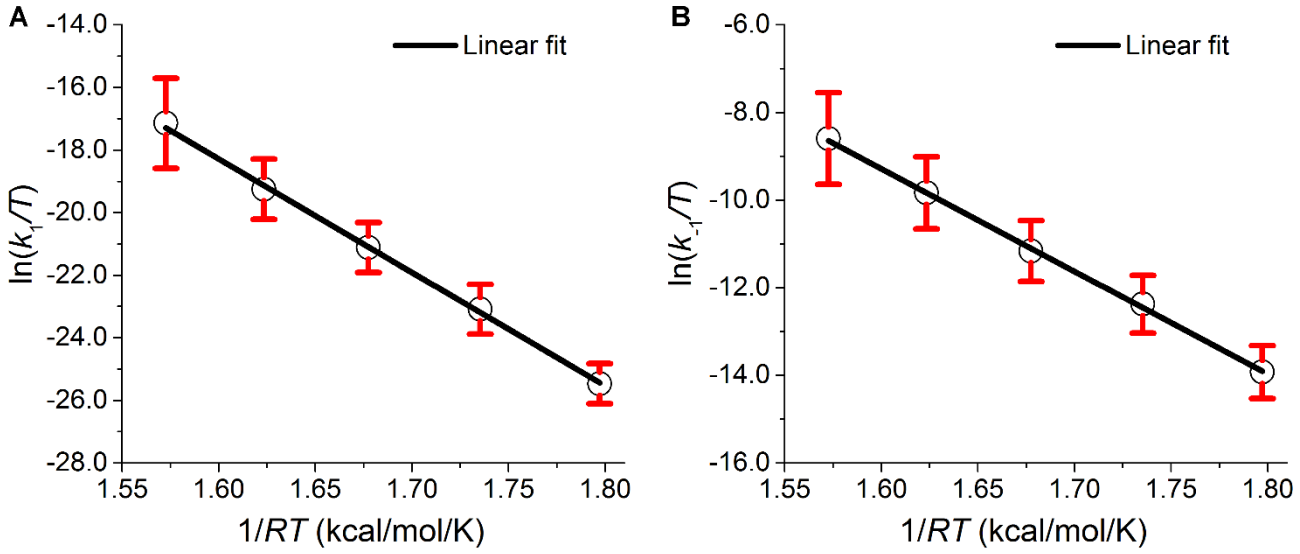

Fig. S8. Calculation of reaction activation enthalpies. (A) The activation enthalpy for the forward reaction, $\Delta H_1^{\ddagger}$, is the negative slope. (B) The activation enthalpy for the backward reaction, $\Delta H_{-1}^{\ddagger}$, is the negative slope.

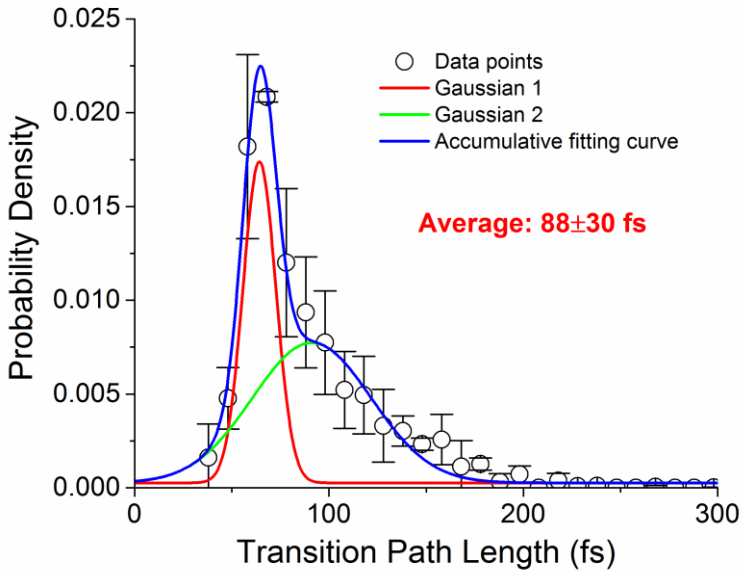

Fig. S9. Probability distribution of transition path length (τ). The overall distribution (blue line) is accumulated by two individual Gaussians (red and green, respectively). The average transition path length is (88±30) fs.

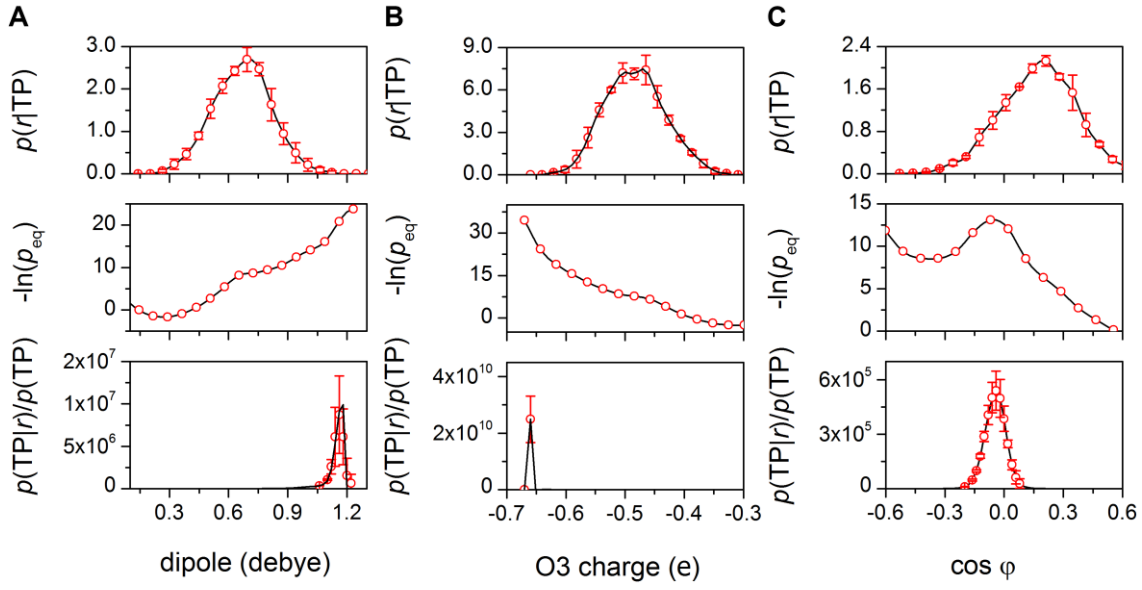

Fig. S10. Bayesian assessment of RC candidates regarding to intrinisc molecular properties : the dipole moment (A), charges on O3 atom (B), and the characteristic dihedral angle φ (C) (see Fig. 1A in the main text for defination). The transistion state is characterized by a large dipole, enriched negative charge on O3 (charge separation) and an almost in-plane dihedral

angle φ.

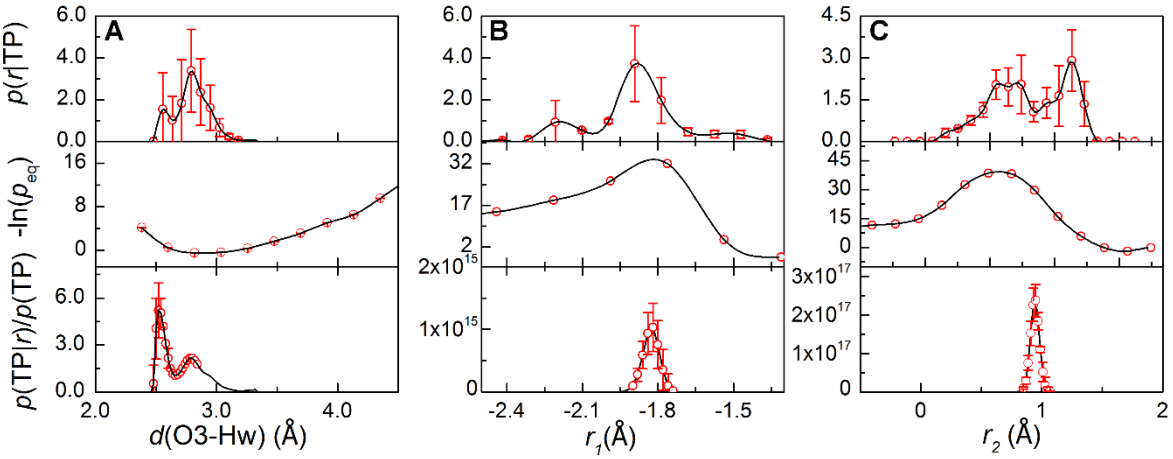

Fig. S11. Bayesian optimization of RC by introducing solvent coordinates. (A) TS is characterized by either a short or a relative long hydrogen bond (indicated by the two peaks in the bottom panel). (B) Mix in hydrogen bond with d(O3-C4) improves the RC defined by O3-4 distance: r_1=0.92d(O3-C4)-0.08d(O3-Hw). (C) Mix in hydrogen bond does not significantly improve the r_0 (the optimized RC on d(C1-C6) d(O3-C4 space), r_2=0.75d(C1-C6)-0.23d(O3-C4)+0.02d(O3-Hw).

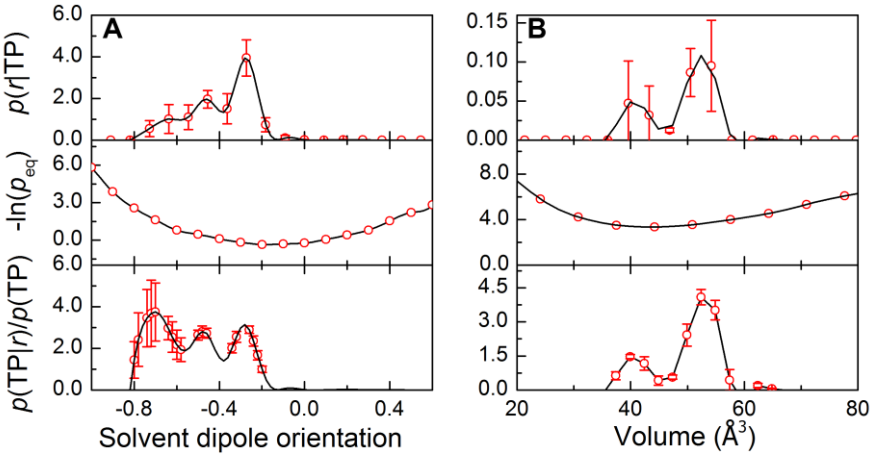

Fig. S12. Bayesian assessment of solvent coordinates. (A) The solvent dipole tends to be antiparallel (with negative values) with

the reactant dipole. (B) The solvent excluded volume (exclusion volume), and the TS is featured by a shrunk solvation shell with small exclusion volume (the “cage effect”).

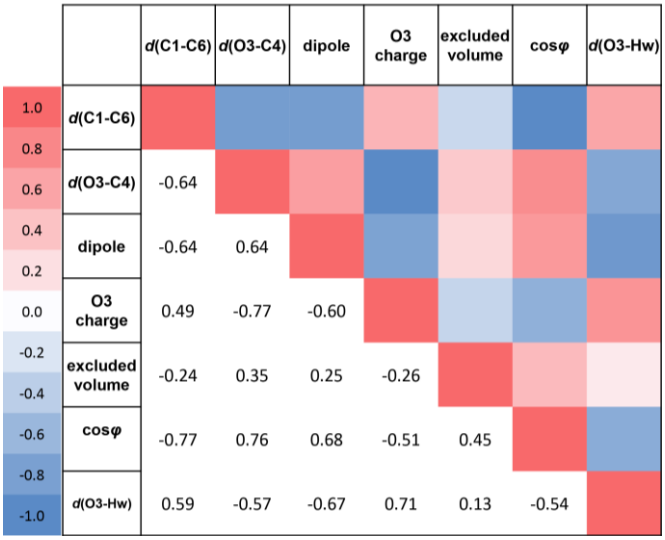

Fig. S13. Correlation coefficient matrix. Cross-correlation coefficients between different molecular properties and solvent properties. (Upper triangle) Positive correlation is rendered in red, negative in blue ; (lower triangle) the corresponding correlation coefficients.

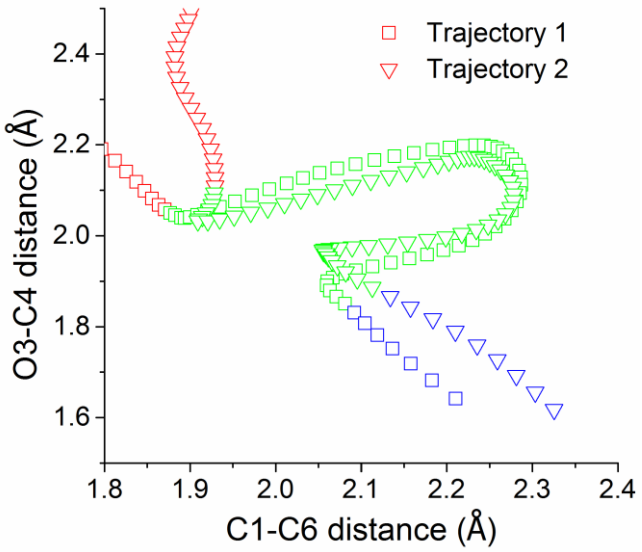

Fig. S14. Finding the longest common subsequences (LCSS) between two trajectories. Two trajectories are symbolized in squares and triangles, respectively. The LCSS are rendered in green.

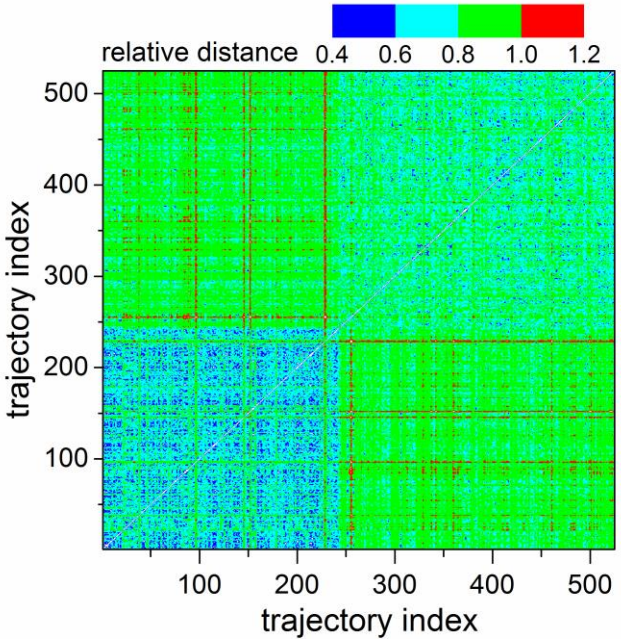

Fig. S15. The similarity (or distance) matrix with trajectories ordered by clusters. There are two diagonal blocks. Blue colors

suggest high similarity while red indicates large distance between trajectories.

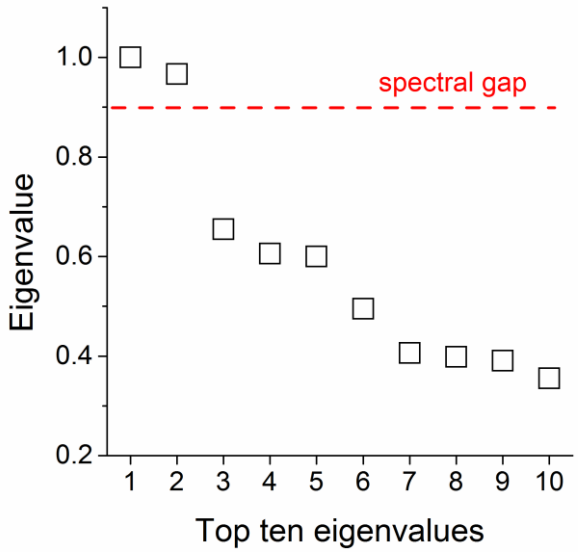

Fig. S16. Eigenvalue spectrum of spectral decomposition. The spectral gap is shown in a red dashed line. There are two principal eigenvalues (close to 1) meaning two clusters of transition paths.

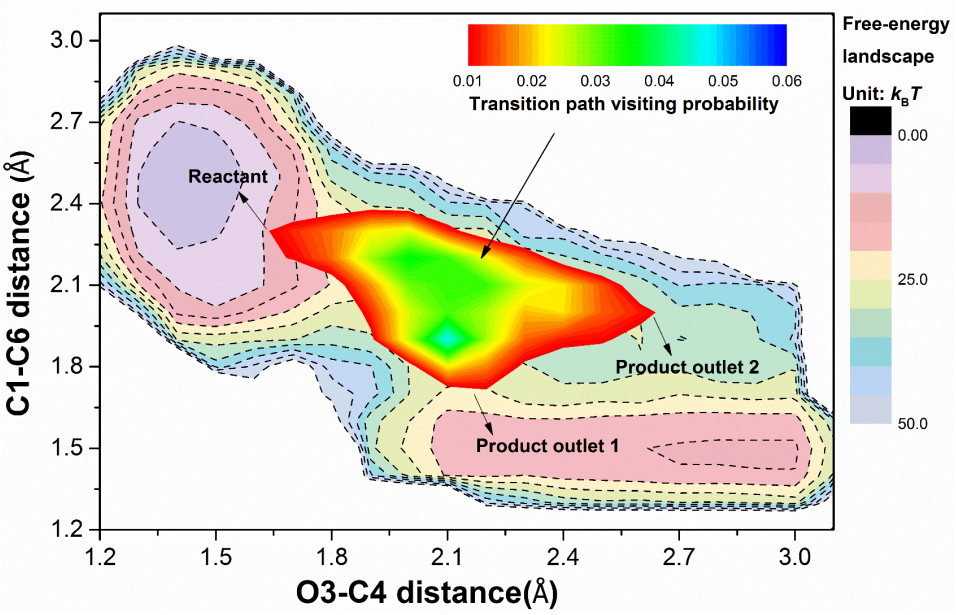

Fig. S17. Transition path visiting probability projected onto the 2D free-energy landscape. The two outlets connecting the TS region to the product basin, along with the one connecting to the reaction, are labelled out.

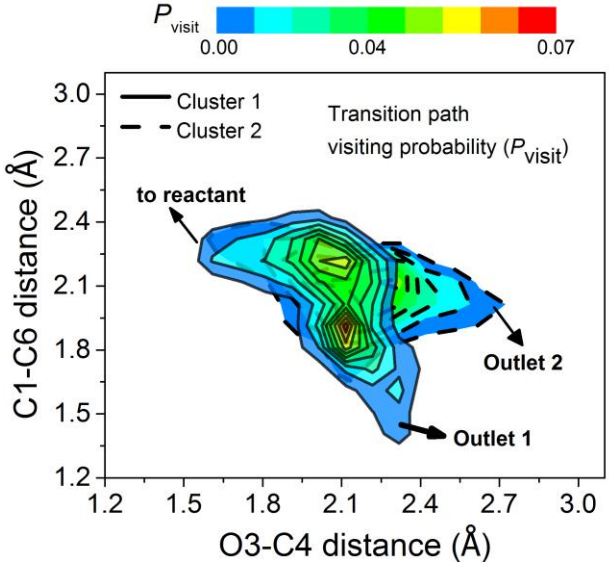

Fig. S18. Transition probability visiting probability (P_{visit}) of two individual clusters projected onto the 2D map. The front layer with solid contour lines corresponds to cluster 1, which occupies the outlet 1; the bottom layer with dashed contour lines corresponds to cluster 2, which mainly transits through outlet 2.

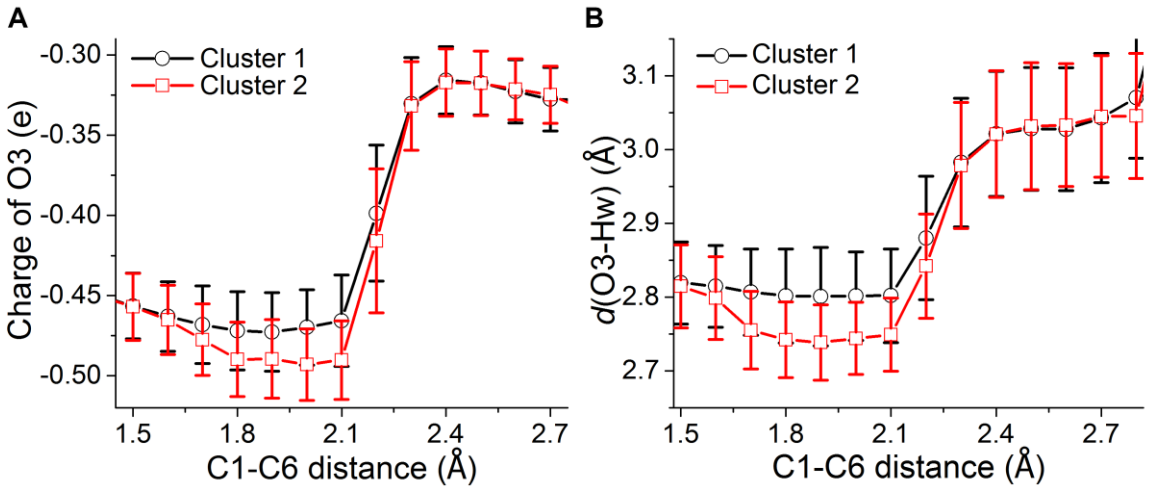

Fig. S19. Mulliken charge on O3 atom (A) and hydrogen-bond length (B) change along the reaction progression calculated for different clusters (cluster 1 in black lines, cluster 2 in red lines.

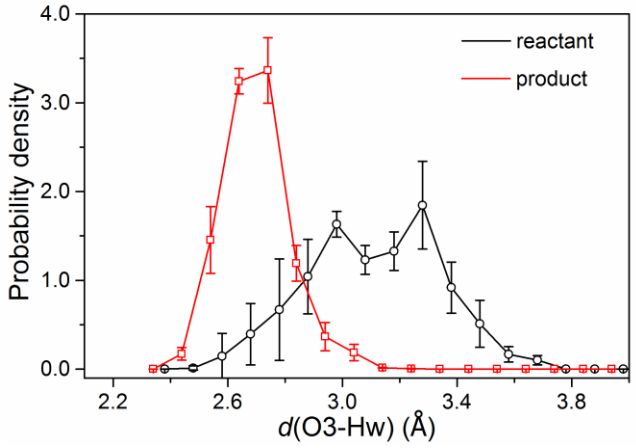

Fig. S20. The distributions of hydrogen-bond length in reactant (black) and product (red) states. Hydrogen-bond length is shorter for product (the aldehyde) than for reactant (ether).

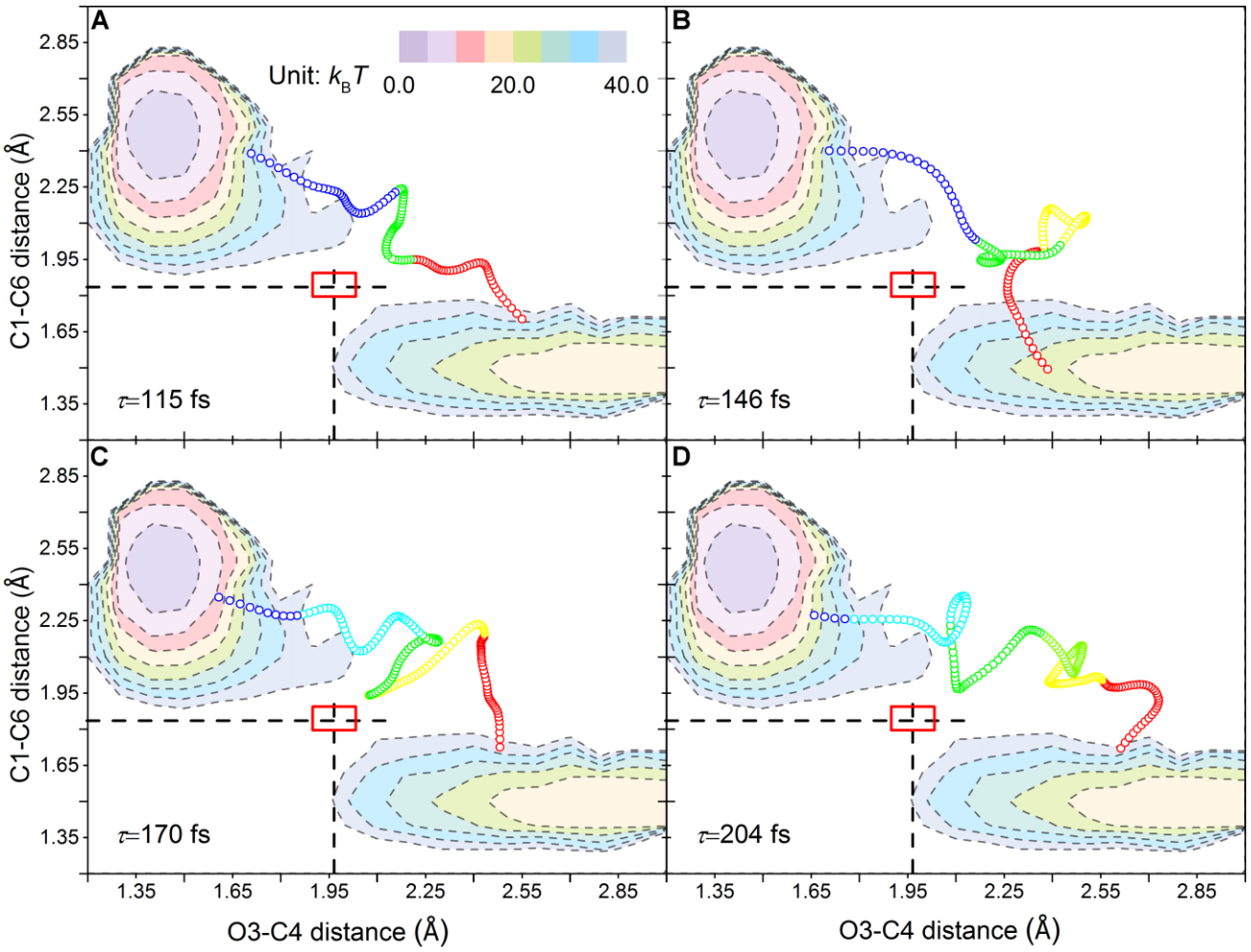

Fig. S21. Example of transition paths with different lengths (τ) and numbers of vibrations: τ=115 fs (A), τ=146 fs (B), τ=170 fs (C), and τ=204 fs (D), respectively. Each purely colored trajectory segment indicates a 40-fs-long vibration (or oscillation). The red rectangular box denotes the concerted TS region, whereas the two dashed lines correspond to the separatrixes defined by *d*(C1-C6) and *d*(O3-C4) as in Fig. 4A in the main text.

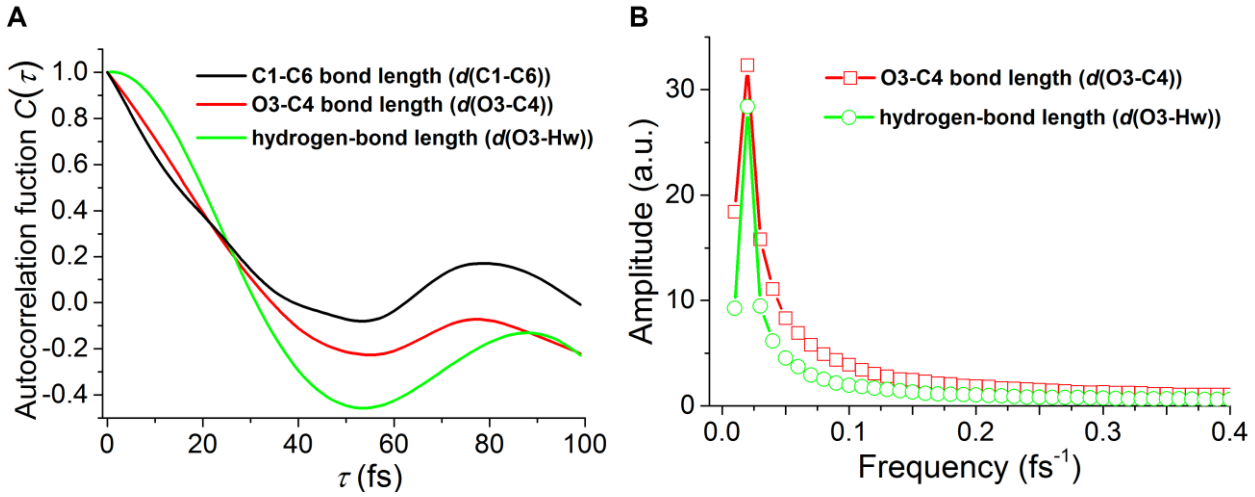

Fig. S22. Autocorrelation functions $C(\tau)$ (A) and corresponding Fourier transforms (B) of: d(C1-C6) (black), d(O3-C4) (red) and d(O3-Hw) (green) in transition path region.

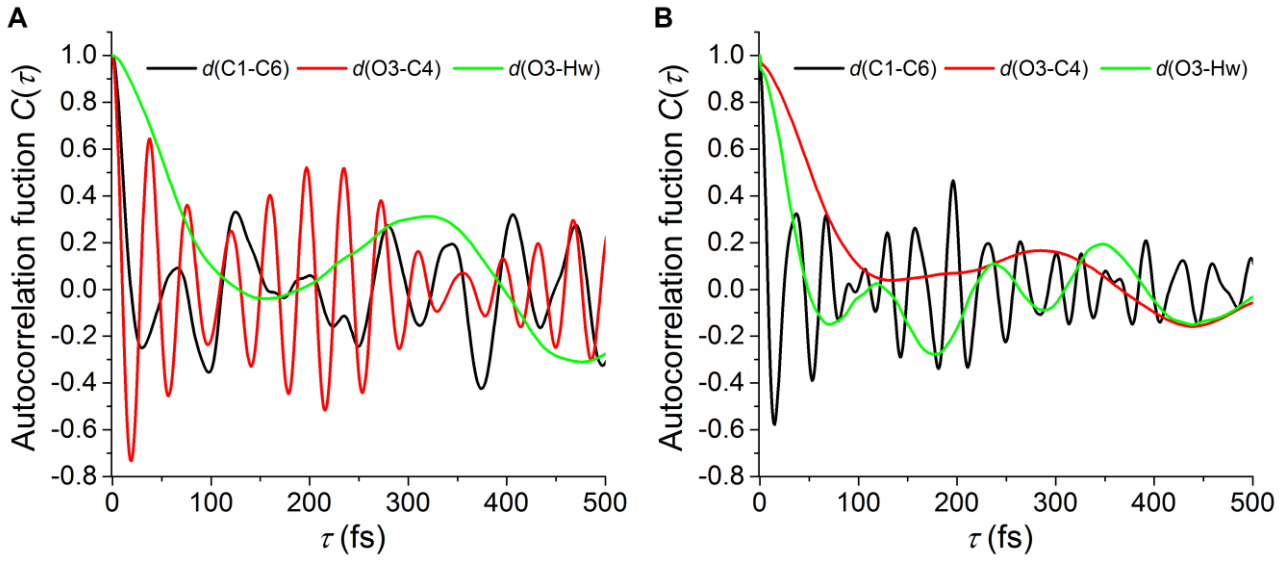

Fig. S23. Autocorrelation functions $C(\tau)$ of different motions for reactant (A) and product (B), respectively. C1-C6 bond length is rendered in black, O3-C4 bond length in red, hydrogen-bond length in green.

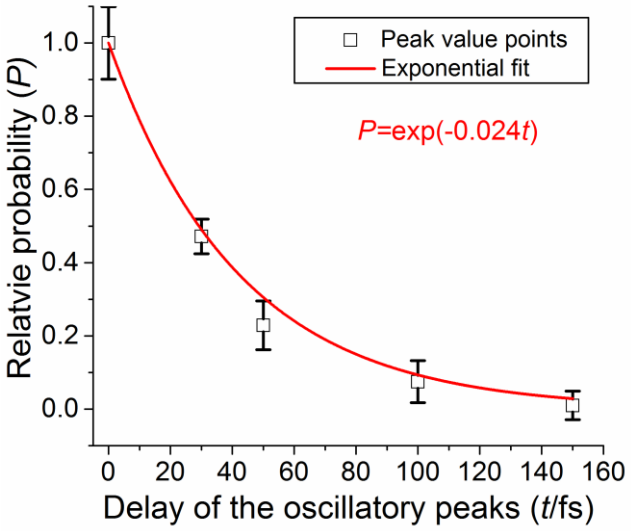

Fig. S24. Exponential fit of the damping peak values observed in the transition path length distribution of cluster 2. The abscissa is the relative delay of the peak values (the first peak in Fig. 4B in the main text is set to zero). The damping rate is about 0.024 fs^{-1}.

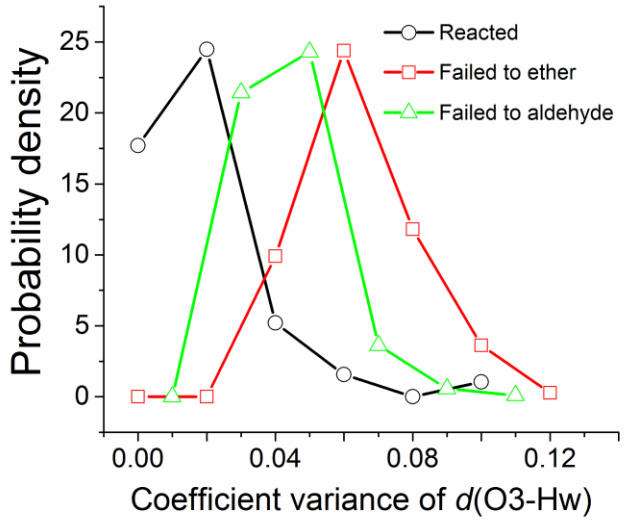

Fig. S25. Distribution of coefficient variance of hydrogen bond length for different trajectories: black for reacted trajectories, green for those started and returned to the aldehyde basin, red for those non-reacted trajectories in ether basin.

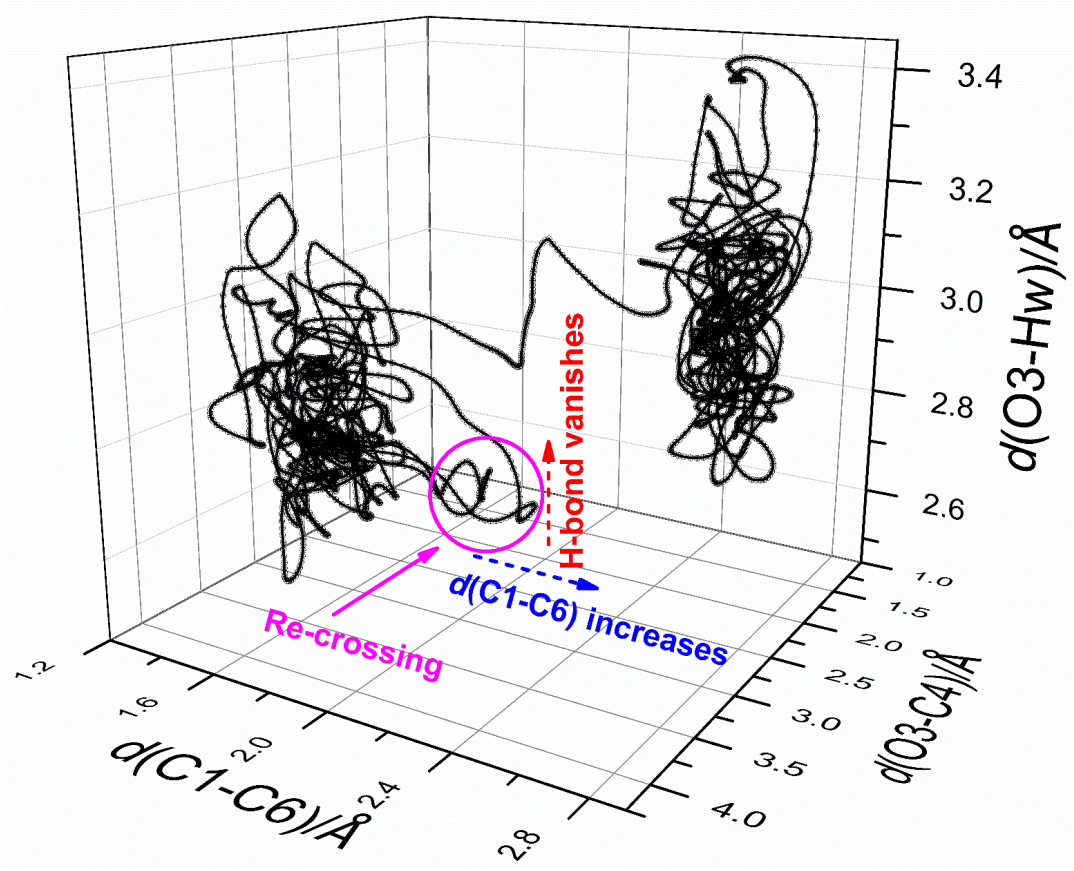

Fig. S26. Example of re-crossing event caused by dephasing of the hydrogen bond. The purple circled region: When the C1-C6 bond is breaking and O3-C4 bond forming in the beginning, the hydrogen bond is vanishing and dephasing, so the trajectory returns to the metastable basin.

Supplementary Tables

Table S1. Sampling parameters (of QM-region) in distributed MITS for different windows

Window index	Temperature downlimit[a]	Temperature uplimit	Temperature intervals (n_k size)	Applied restraints
1	250(275)[b]	650(325)	50(30)	No
2	250(275)	1000(325)	150(30)	Yes[c]
3	250(275)	650(325)	50(30)	No

a. Temperatures are all in unit of Kelvin (K).

b. Numbers with and without parentheses correspond to MITS parameters for QM-Hamiltonian and MM-Hamiltonian, respectively.

c. Corresponding to Eq. S16.

Table S2. Calculation of rate constants and convergence test

Data Block[a]	1	2	3	4	5
N_{sus}[b]	182	271	355	443	525
$N_{attempt}$[c]	15207	19007	23873	28375	32168
k_1/s^{-1}	1.9×10^{-7}	1.9×10^{-7}	2.0×10^{-7}	2.1×10^{-7}	2.0×10^{-7}
k_{-1}/s^{-1}	3.3×10^{-3}	3.8×10^{-3}	4.2×10^{-3}	4.3×10^{-3}	4.3×10^{-3}

a. The convergence was tested against a series of block averages (totally 5 blocks).

b. The successful transitions within the block.

c. The total attempted shootings within the block.

Table S3. Contribution of different clusters to the rate constants

	Contribution to Forward Reaction Rate	Contribution to Backward Reaction Rate
Cluster 1	63%	45%
Cluster 2	37%	55%

Table S4. Correlation coefficient between hydrogen bond and breaking/forming bonds in different trajectories

	d(C1-C6)	*d*(O3-C4)
Reacted	0.59	-0.57
Non-reacted	<0.01	-0.07

Table S5. Test of transition path shooting length *t*

t/ps	$k(t)/t$ (10^{-7} s^{-1})
1.2	1.7±0.3
1.5	1.9±0.3
2.0	2.0±0.2

Supplementary Movie

FR.wmv: An exemplary trajectory of a forward transition is shown in the movie. Carbon atoms are rendered in cyan, oxygens in red, hydrogens in white.

References

1. Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W. *Amber 10*; University of California: 2008.
2. Seifert, G.; Joswig, J. O., Density-functional tight binding—an approximate density-functional theory method. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2* (3), 456-465.
3. Gaus, M.; Cui, Q.; Elstner, M., Density functional tight binding: application to organic and biological molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014,** *4* (1), 49-61.
4. Otte, N.; Scholten, M.; Thiel, W., Looking at self-consistent-charge density functional tight binding from a semiempirical perspective. *J. Phys. Chem. A* **2007,** *111* (26), 5751-5755.
5. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998,** *58* (11), 7260.
6. Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S., Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations. *Chem. Phys.* **2001,** *263* (2), 203-219.
7. Shen, L.; Wu, J.; Yang, W., Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **2016,** *12* (10), 4934-4946.
8. Zhang, J.; Yang, Y. I.; Yang, L.; Gao, Y. Q., Conformational Preadjustment in Aqueous Claisen Rearrangement Revealed by SITS-QM/MM MD Simulations. *J. Phys. Chem. B* **2015,** *119* (17), 5518-5530.
9. Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
10. Gao, Y. Q., An integrate-over-temperature approach for enhanced sampling. *J. Chem. Phys.* **2008,** *128* (6), 064105.
11. Zhang, J.; Yang, Y. I.; Yang, L.; Gao, Y. Q., Dynamics and Kinetics Study of “In-Water” Chemical Reactions by Enhanced Sampling of Reactive Trajectories. *J. Phys. Chem. B* **2015,** *119* (45), 14505-14514.
12. Yang, M.; Yang, L.; Gao, Y.; Hu, H., Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface. *J. Chem. Phys.* **2014,** *141* (4), 044108.
13. Souaille, M.; Roux, B. t., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001,** *135* (1), 40-57.

14. Hub, J. S.; De Groot, B. L.; Van Der Spoel, D., g_wham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010,** *6* (12), 3713-3720.

15. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12), 124105.

16. Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F., Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (23), E3221-E3230.

17. Chandler, D., Introduction to modern statistical mechanics. *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler.* Oxford University Press: Oxford, U.K., Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771, **1987,** *1*.

18. Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L., Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002,** *53* (1), 291-318.

19. Ferguson, A. L.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Kevrekidis, I. G., Systematic determination of order parameters for chain dynamics using diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (31), 13597-13602.

20. Hummer, G.; Garde, S.; Garcia, A.; Paulaitis, M. E.; Pratt, L. R., Hydrophobic effects on a molecular scale. *J. Phys. Chem. B* **1998,** *102* (51), 10469-10482.

21. Best, R. B.; Hummer, G., Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (19), 6732-6737.

22. Hänggi, P.; Talkner, P.; Borkovec, M., Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **1990,** *62* (2), 251.

23. Straub, J. E.; Borkovec, M.; Berne, B. J., Non-Markovian activated rate processes: Comparison of current theories with numerical simulation data. *J. Chem. Phys.* **1986,** *84* (3), 1788-1794.

24. Wang, F.; Zhang, C. In *Spectral clustering for time series*, International Conference on Pattern Recognition and Image Analysis, Springer: Berlin Heidelberg, 2005; pp 345-354.

25. Porikli, F. In *Clustering variable length sequences by eigenvector decomposition using HMM*, Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), Springer: Berlin Heidelberg, 2004; pp 352-360.

26. Das, G.; Gunopulos, D.; Mannila, H. In *Finding similar time series*, European Symposium on Principles of Data Mining and Knowledge Discovery, Springer: Berlin Heidelberg, 1997; pp 88-100.

27. Vlachos, M.; Kollios, G.; Gunopulos, D. In *Discovering similar multidimensional trajectories*, Data Engineering, 2002. Proceedings. 18th International Conference on, IEEE: 2002; pp 673-684.

28. Huang, J.; Yuen, P. C.; Chen, W.-S.; Lai, J.-H. In *Kernel subspace LDA with optimized kernel parameters on face recognition*, Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, IEEE: 2004; pp 327-332.

29. Hartigan, J. A.; Wong, M. A., Algorithm AS 136: A k-means clustering algorithm. *J. R. Statist. Soc. C* **1979,** *28* (1), 100-108.

30. Rhoads, S. J.; Cockroft, R. D., Valence tautomerism in cis-2-vinylcyclopropanecarboxaldehyde. 2, 5-Dihydrooxepin. *J. Am. Chem. Soc.* **1969,** *91* (10), 2815-2816.