

Refined metadynamics through canonical sampling using time-invariant bias potential: A study of polyalcohol dehydration in hot acidic solutions

Tomomi Kondo^{*†}, Takehiko Sasaki^{*},
Sergi Ruiz-Barragan^{†‡}, Jordi Ribas-Ariño[§], Motoyuki Shiga[†]

October 9, 2020

Abstract

We propose a canonical sampling method to refine metadynamics simulations a posteriori, where the hills obtained from metadynamics are used as a time-invariant bias potential. In this way the statistical error in the computed reaction barriers is reduced by an efficient sampling of the collective variable space at the free energy level of interest. This simple approach could be useful particularly when two or more free energy barriers are to be compared among chemical reactions in different or competing conditions. The method was then applied to study the acid dependence of polyalcohol dehydration reactions in high-temperature aqueous solutions. It was found that the reaction proceeds consistently via an S_N2 mechanism, whereby the free energy of protonation of the hydroxyl group created as an intermediate is affected significantly by the acidic species. Although demonstration is shown for a specific problem, the computational method suggested herein could be generally used for simulations of complex reactions in the condensed phase.

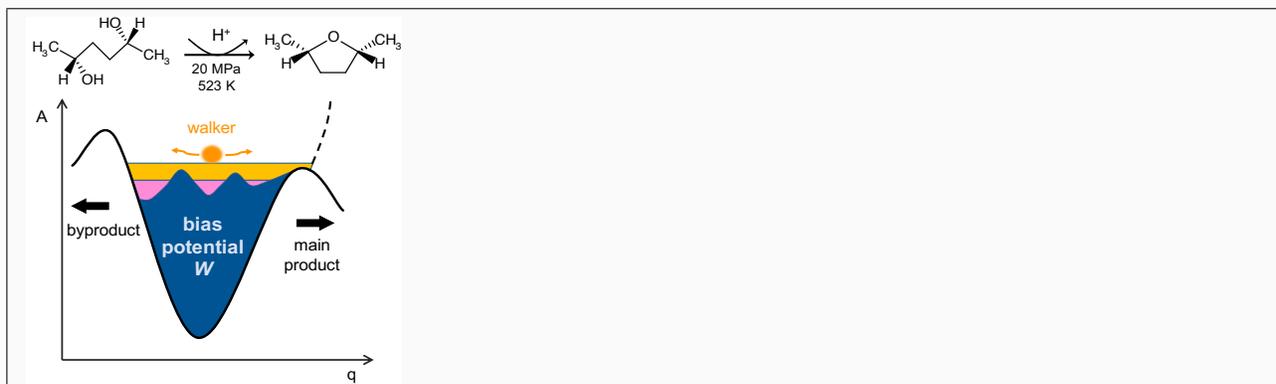
Keywords: refined metadynamics, bias sampling, polyalcohol dehydration, high-temperature water, density functional tight binding ■

^{*}Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa 277-8561, Japan

[†]Center for Computational Science and e-Systems, Japan Atomic Energy Agency, 178-4-4, Wakashiba, Kashiwa, Chiba, 277-0871, Japan

[‡]Present Address: Lehrstuhl für Theoretische Chemie Ruhr-Universität Bochum, 44780 Bochum, Germany

[§]Departament de Ciència dels Materials i Química Física and IQTCUB, Universitat de Barcelona, Martí i Franquès 1, Barcelona, E-08028, Spain



We propose a canonical sampling method to refine metadynamics simulations where the hills obtained from metadynamics are used as a time-invariant bias potential. The method was applied to study the acid dependence of polyalcohol dehydration reactions in high-temperature aqueous solutions. It was found that the reaction proceeds via an S_N2 mechanism, whereby the free energy of protonation of the hydroxyl group created as an intermediate is affected significantly by the acidic environment.

INTRODUCTION

The study of complex reactions in the condensed phase is an important task in computational chemistry. As chemical reactions are generally rare event phenomena that only occur by overcoming high free energy barriers, conventional molecular dynamics (MD) simulations are not proper for studying them. Many different methods have been developed to study rare events in an efficient manner.^{1–18}

Modern computational techniques have enabled qualitative estimates of high reaction barriers. For a typical organic reaction with a half-life of 1 hour, the free energy barrier according to transition state theory (TST) is as high as 30–40 times the thermal energy $k_{\text{B}}T$, where k_{B} is the Boltzmann constant and T is the temperature. However it is usually more difficult to compare the relative difference of two or more reactive processes, e.g., the dependence on different conditions, such as temperature, pressure and acidity. In this case, a free energy difference of the order of few $k_{\text{B}}T$ may become important. When it comes to a competitive reaction, a small difference in the free energy barrier would be crucial to determine the product ratio. Therefore, there is a need for an accurate means of determining free energy barriers.

Metadynamics (MTD)^{8,19–29} is very useful for complex reactions in the condensed phase, when the reaction can be characterized by a small number of degrees of freedom N_{cv} called collective variables (CVs), $\mathbf{q} = (q_1, \dots, q_{N_{\text{cv}}})$. In this method, the landscape of the free energy surface $W(\mathbf{q})$ can be computed by exploring the CV space efficiently by applying a history-dependent bias potential that discourages the system from returning to visited places in the configuration space. For a proper description of chemical bond exchange of reactive processes, *ab initio* or semi-empirical MTD, i.e., metadynamics combined with on-the-fly *ab initio* or semi-empirical electronic structure calculations, is usually required. For the computational efficiency of such simulations, the run could be terminated after reaching the product basin. While this approach does not give access to any information regarding the free energy landscape of the product, the free energy barrier can be estimated from the difference in values between the saddle point and the reactant minimum. The escape process from the reactant basin allows one to identify the reaction mechanism via the change of CVs

in going from the reactant minimum to the saddle point.

Theoretically, MTD is an exact methodology in the limit of infinitesimally low frequency and height of the deposited hills. However, in practice, these values are finite and, often, large enough for an efficient sampling of configuration space. This gives rise to the characteristic roughness of MTD free energy surfaces, which in turn might lead to a significant source of error when it comes to determining free energy barriers or differences. For this reason, several variants of MTD have been recently proposed with the goal of improving the accuracy of the obtained free energy surfaces. These variants include cleverly designed adaptive schemes in which the increments of the bias potential become smaller as the simulation goes on,²³⁻²⁶ reweighting techniques applied a posteriori to the MTD simulations,^{27,30-36} and methods in which the reweighting is applied on-the-fly and the result of such reweighting is used to define the adaptive bias potential.²⁸ Improvement of MTD has been also suggested using the combination with umbrella sampling^{37,38}, adiabatic free energy dynamics³⁹, and temperature accelerated molecular dynamics⁴⁰. Despite all these valuable recent efforts, converging the free energy surface with an error lower than $k_B T$ is still computationally demanding in general.

In this paper, we propose a method of refining the free energy estimate using the approximate $W(\mathbf{q})$ obtained from conventional MTD simulations. As shown in Figure 1 schematically, it is a post-process for conventional MTD to upgrade the bias potential by canonical sampling. We start from the $W(\mathbf{q})$ set from upon reaching the transition state or from just prior to reaching the transition state. This allows us to explore the CV space \mathbf{q} at the free energy level where the reaction takes place. Next, we perform a biased sampling in the reactant basin where $W(\mathbf{q})$ is regarded as a time-invariant bias potential. Here we add an artificial wall potential that has high values at the entrance of the product basin and zero values in the reactant basin, so as not to affect the sampling. The sampling can be performed by any kind of equations of motion that generates the canonical ensemble, i.e., thermostatted MD or MTD in the absence of the hills deposit. The resulting density distribution $\rho_W(\mathbf{q})$ is then used to upgrade $W(\mathbf{q})$. After some iterations of this procedure, the bias potential should converge to the negative value of the free energy within the region of \mathbf{q} from the reactant basin to the transition state. Eventually, the sampling becomes almost

uniform in this region from which one can judge the convergence. Unlike conventional MTD, the sampling is done always with a fixed $W(\mathbf{q})$ at the last upgrade, so the unbiasing can be done without any approximate assumptions. Thus the final error could be less than $k_{\text{B}}T$ after sufficient sampling, in principle.

In this study, we use the refined MTD to estimate the free energy barrier for dehydration of polyalcohols in high-temperature aqueous solutions. The dehydration reaction of polyalcohols is seen as an important step for the conversion of biomass into cyclic ethers. High-temperature water has attracted attention from a viewpoint of green chemistry because of the advantages of being less toxic, cheaper, and easier to handle than organic solvents.^{41–43} The reaction barriers in aqueous media are sensitive to temperature, pressure and acidity, which opens up the possibility of optimizing the reaction conditions. In previous experiments, it was reported that the rate of polyalcohol dehydration can be accelerated significantly under acidic conditions.^{44–46} Recently, the reaction rates of polyalcohol dehydration in high-temperature water and carbonated water has been studied systematically by an experimental group of Shirai, Yamaguchi, et al.^{47–52} Accordingly our research group has been working on elucidating the mechanism of this reaction from a standpoint of computational chemistry.^{53,54} From the analysis of dehydration of simple polyalcohols, i.e., 2,5-hexanediol (HDO) and 1,2,5-pentanetriol (PTO), we proposed that the main reaction is dominated by a protonation-assisted $S_{\text{N}}2$ pathway. However we have not yet proved the acidity dependence that has been observed experimentally. In this study, our aim is to compute the free energy barriers for the dehydration reaction of 2,5-HDO in different acidic solutions in order to verify that the proposed $S_{\text{N}}2$ reaction mechanism shows such an acidity dependence.

THEORY

Consider a situation where the negative value of the free energy surface, $W(\mathbf{q})$, has been obtained approximately after a conventional MTD simulation, and we wish to refine the quality of that $W(\mathbf{q})$. In this case, running an isothermal simulation in the presence of a

time-invariant bias potential $W(\mathbf{q})$ would generate a canonical distribution

$$\begin{aligned}\rho_W(\mathbf{q}) &\propto \int d\mathbf{r} e^{-\beta(V(\mathbf{r})+W(\mathbf{q}))} \delta(\mathbf{Q}(\mathbf{r}) - \mathbf{q}) \\ &\propto e^{-\beta W(\mathbf{q})} \rho(\mathbf{q})\end{aligned}\tag{1}$$

where $\beta = k_B T$, $\mathbf{Q}(\mathbf{r})$ is the set of CVs given as a function of atomic coordinates \mathbf{r} , $V(\mathbf{r})$ is the potential energy of the system, and

$$\rho(\mathbf{q}) \propto \int d\mathbf{r} e^{-\beta V(\mathbf{r})} \delta(\mathbf{Q}(\mathbf{r}) - \mathbf{q})\tag{2}$$

is the canonical distribution in the absence of the bias potential. Now suppose that the biased simulation gave $\rho_W(\mathbf{q})$ as the probability distribution of finding the CVs at $\mathbf{q} = \mathbf{Q}(\mathbf{r})$. Then, the Helmholtz free energy surface defined by $A(\mathbf{q}) \equiv -\beta^{-1} \ln \rho(\mathbf{q})$ can be estimated as

$$A(\mathbf{q}) \approx -W(\mathbf{q}) - \Delta W(\mathbf{q})\tag{3}$$

where the unbiasing correction is, apart from a redundant constant,

$$\Delta W(\mathbf{q}) = \beta^{-1} \ln \rho_W(\mathbf{q}).\tag{4}$$

The relationship in Equation (3) becomes exactly equal when the amount of sampling is infinitely large.

In conventional MTD, the free energy is described as a sum of Gaussian functions (hills) to obtain a smooth surface. In the same way it would be convenient also to describe the free energy correction $\Delta W(\mathbf{q})$ as a sum of hills from the histogram of distributions, which will be shown below. In this way the refinement of $\Delta W(\mathbf{q})$ could be done in an iterative manner by carrying out biased simulations with the upgraded bias potential, as shown by the schematic in Figure 2. Note here that the bias potential is shifted incrementally until convergence is attained. This meets our purpose to estimate the free energy barrier since it allows sampling at the same free energy level for the reaction of interest. This is in contrast to conventional MTD where the free energy level is continuously elevated during the simulation. The more sampling we perform, the better quality we obtain for the free energy surface.

We now describe our implementation to compute the free energy correction described by N_{grid} grid points assuming the form

$$\Delta W(\mathbf{q}) = \sum_{j=1}^{N_{\text{grid}}} h_j e^{-\frac{1}{2}t(\mathbf{q}-\mathbf{q}_j)\sigma^{-2}(\mathbf{q}-\mathbf{q}_j)}\tag{5}$$

where h_j is the height of the hill centered at the j th grid point in the CV space. For simplicity it is assumed that the hills have a common width for all the grid points, and they are the elements of the diagonal matrix, $\boldsymbol{\sigma} = \text{diag}(\sigma_1, \dots, \sigma_{N_{\text{cv}}})$. Let $\Delta W_i = \Delta W(\mathbf{q}_i)$ be the histogram obtained at the i th grid point \mathbf{q}_i . To optimize the set of heights of the hills, $\mathbf{h} = (h_1, \dots, h_{N_{\text{grid}}})$, we impose a condition that the mean-square deviation

$$E = \frac{1}{N_{\text{grid}}} \sum_{i=1}^{N_{\text{grids}}} \left(\sum_{j=1}^{N_{\text{grid}}} h_j e^{-\frac{1}{2}t(\mathbf{q}_i - \mathbf{q}_j)\boldsymbol{\sigma}^{-2}(\mathbf{q}_i - \mathbf{q}_j)} - \Delta W_i \right)^2 \quad (6)$$

is minimized. Then $dE/d\mathbf{h} = 0$ leads to a linear equation

$$\sum_{i=1}^{N_{\text{grid}}} C_{ki} h_i = b_k \quad (7)$$

where we have defined

$$b_k = \sum_{i=1}^{N_{\text{grid}}} (\Delta W_i) e^{-\frac{1}{2}t(\mathbf{q}_i - \mathbf{q}_k)\boldsymbol{\sigma}^{-2}(\mathbf{q}_i - \mathbf{q}_k)} \quad (8)$$

and

$$C_{ki} = \sum_{j=1}^{N_{\text{grid}}} e^{-\frac{1}{2}(t(\mathbf{q}_i - \mathbf{q}_j)\boldsymbol{\sigma}^{-2}(\mathbf{q}_i - \mathbf{q}_j) + t(\mathbf{q}_j - \mathbf{q}_k)\boldsymbol{\sigma}^{-2}(\mathbf{q}_j - \mathbf{q}_k))}. \quad (9)$$

In conventional MTD, $W(\mathbf{q})$ is expressed as a sum of Gaussian hills, see Equation (12) in the next section. Thus, according to Equation (3), adding $\Delta W(\mathbf{q})$ of Equation (5) results in a new set of Gaussian hills.

Equation (7) can be solved numerically, e.g., using the LAPACK library⁵⁵ (DSYSV or DGESV). The grid points should cover the CV space where the histogram is non-zero, but they do not need to be regular. The value of $|\boldsymbol{\sigma}|$ should be set greater than half of the grid spacing to make the $W(\mathbf{q})$ surface smooth. For large grid sizes with high-dimensional CVs, evaluation of C_{ki} matrix could be a computational bottleneck. In that case, the summation of C_{ki} matrix can be done only for the grid points j where \mathbf{q}_j are close to both \mathbf{q}_k and \mathbf{q}_i , within a certain cutoff distance that is set much larger than $|\boldsymbol{\sigma}|$. For angular CVs, periodic boundary should be applied to Equations (8) and (9) to obtain the set of heights \mathbf{h} .

SIMULATION METHOD

The refined MTD has been applied to the dehydration reaction of $2R,5R$ -HDO in high-temperature aqueous solutions. The simulations were carried out in the thermodynamic conditions of 523 K and 20 MPa in accordance with experiments using pure and carbonated water. The simulated system contained one $2R,5R$ -HDO ($C_6H_{14}O_2$) molecule and 30 H_2O molecules, either with or without acidic species (H_2CO_3 , H_2SO_4 or HCl) added to it. It was confirmed previously that the system size effect on the free energy barrier is small for this system since the reaction takes place locally involving several water molecules in the vicinity of HDO.⁵³ The atoms were placed in a cubic box with a volume of $(11.38 \text{ \AA})^3$, which was preliminarily determined from classical molecular dynamics simulations in the NPT ensemble at 20 MPa and 523 K using the OPLS all atom force field.⁵⁶ Periodic boundary condition was applied to the system to represent the bulk aqueous solution. On-the-fly electronic structure calculations of the Born-Oppenheimer forces were done with a semi-empirical approach based on the self-consistent-charge density-functional tight-binding (SCC-DFTB) method⁵⁷⁻⁵⁹ using the 3ob Slater-Koster parameter set.^{60,61} The massive Nosé-Hoover chain (MNHC) thermostat technique⁶²⁻⁶⁵ was used to control the system efficiently. The simulations were conducted using a combination of the PIMD^{66,67} and DFTB+^{68,69} software packages.

Three CVs ($N_{cv} = 3$, $q_1 = d$, $q_2 = \phi$, $q_3 = n$) were chosen based on minimum energy path analysis of a microsolvated model of $2R,5R$ -HDO.⁵³ As shown in Figure 3, d is the difference in C_2-O_7 and C_2-O_8 distances, which represents the exchange of bond breaking and formation upon the ether ring formation, ϕ is the dihedral angle of $O_7-C_2-C_3-C_4$ atoms which characterizes the conformation of $2R,5R$ -HDO molecule, and n is the hydrogen coordination number of the O_7 atom which represents the protonation of the O_7H hydroxyl group. Note that the C_2-O_7 bond is not included in n . We used the sum of rational functions for this

$$n = \sum_{j \in H} \frac{1 - \left(\frac{r_{j7}}{r_b}\right)^6}{1 - \left(\frac{r_{j7}}{r_b}\right)^{12}} \quad (10)$$

where the bond length parameter is $r_b = 1.4 \text{ \AA}$. Unlike our earlier works,^{53,54} we did not employ artificial wall potentials in the large d regions to restrict the conformational sampling. Instead, the whole sampling, including the conformational change of $2R,5R$ -HDO,

was undertaken using the refined MTD. Artificial wall potential at $d < 0$ was employed in the refinement stage to prevent sampling from the product region.

The pseudo-Hamiltonian of conventional MTD was given by

$$H = \sum_{k=1}^{N_w} H_{\text{sys}}(\mathbf{r}^{(k)}, \mathbf{p}^{(k)}) + \sum_{i=1}^{N_{\text{cv}}} \sum_{k=1}^{N_w} \left[\frac{(p_i^{(k)})^2}{2\mu_i} + \frac{\kappa_i}{2} (q_i^{(k)} - Q_i(\mathbf{r}^{(k)}))^2 \right] + W(\mathbf{q}(t), t) + H_{\text{bath}} \quad (11)$$

where the first term is the Hamiltonian with respect to the N_w copies of the molecular system, $H_{\text{sys}}(\mathbf{r}^{(k)}, \mathbf{p}^{(k)})$, where k is the index of the copy. Each of them represents the sum of kinetic energy of atoms and the Born-Oppenheimer potential calculated with the DFTB method, $V(\mathbf{r}^{(k)})$. The second term is the kinetic energy of fictitious particles (N_w walkers), where $p_i^{(k)}$, $q_i^{(k)}$ and μ_i are the momentum, the position and the mass of the walker, respectively, for the CV component i and the walker k . The third term is a harmonic potential that binds the position of the walker to the actual CV value, $Q_i(\mathbf{r}^{(k)})$, where κ_i is the force constant. The fourth term is a history-dependent bias potential expressed as the sum of hills deposited to the places where all the walkers have visited until time t , i.e.,

$$W(\mathbf{q}(t), t) = \sum_{\tau < t} \sum_{k,l=1}^{N_w} h \exp \left[- \sum_{i=1}^{N_{\text{cv}}} \frac{1}{2} \left(\frac{q_i^{(k)}(t) - q_i^{(l)}(\tau)}{\sigma_i} \right)^2 \right], \quad (12)$$

where h and σ_i are the hill height and the hill width, respectively. The fifth term is the contribution from the MNHC thermostats. In our implementation, a Nosé-Hoover chain is coupled to each degree of freedom of the atoms and for each walker, and it is also coupled to each degree of freedom of the fictitious particle on the CV space for each walker. Thus, the total number of MNHC thermostats is $(3N + N_{\text{cv}})N_w L$ where $L(= 4)$ is the chain length.

First, conventional MTD simulations were undertaken using $N_w = 12$ walkers with the length 94–348 ps each (i.e., 1.1–4.2 ns in total) with the step size $\Delta t = 0.25$ fs until the reaction basin was filled. To speed up the calculation, the hill size was set initially to $(h, \sigma_1, \sigma_2, \sigma_3) = (1.0 \text{ kcal mol}^{-1}, 0.05 \text{ \AA}, 10^\circ, 0.1)$, and was diminished finally to $(0.2 \text{ kcal mol}^{-1}, 0.05 \text{ \AA}, 10^\circ, 0.1)$. Following the setup by Ensing et al.²⁰ the fictitious mass was set as $\mu_i = k_B T \left(\frac{\tau_{\text{cv}}}{1.5\sigma_i} \right)^2$ where $\tau_{\text{cv}} = 40$ fs, and the force constants were set as $\kappa_i = \mu_i \left(\frac{2\pi}{\tau_{\text{fc}}} \right)$ where $\tau_{\text{fc}} = 10$ fs. A new hill was deposited at a time interval of τ_{cv} for each walker.

In the refinement stage, the histogram in the CV space was collected from the DFTB MD simulations in the presence of the MTD bias potential. MNHC thermostats were employed for the efficient sampling in the canonical ensemble. The MD simulations were conducted for 12 individual trajectories each being 309–828 ps long (i.e., 3.7–9.9 ns in total), with the step size $\Delta t = 0.25$ fs. Equation (7) was solved using the regular grid points of $31 \times 30 \times 31$ with the spacing of 1.5 times the width parameters $\sigma = (0.095 \text{ \AA}, 8.0^\circ, 0.067)$ covering all the visited places in the CV space within the reactant basin.

RESULTS

Numerical test

Before proceeding to production simulations, we tested the performance of the refined MTD introduced in the previous section. Since we would like to use a simple diatomic model that imitates the C-O dissociation involved in the polyalcohol dehydration reaction, we designed a Lennard-Jones potential function scaled by a cosine function,

$$V(r) = \epsilon_{\text{LJ}} \left[\left(\frac{\sigma_{\text{LJ}}}{r} \right)^{12} - 2 \left(\frac{\sigma_{\text{LJ}}}{r} \right)^6 \right] \left[-\cos \left(\frac{\pi r}{\sigma_{\text{LJ}}} \right) \right] + \epsilon_{\text{LJ}}, \quad (13)$$

with $\epsilon_{\text{LJ}} = 40.0$ kcal/mol and $\sigma_{\text{LJ}} = 1.40$ Å. As depicted in Figure 4(a), this function has a minimum at 1.40 Å and a maximum at 2.43 Å with a potential barrier of 41.9 kcal/mol. Assuming one-dimensional CV with respect to the interatomic distance $q = r = |\mathbf{r}_{\text{C}} - \mathbf{r}_{\text{O}}|$, the free energy function is exactly given by

$$A(r) = V(r) - \beta^{-1} \ln(4\pi r^2). \quad (14)$$

with a free energy barrier of 40.8 kcal/mol at 523 K, as shown in Figure 4(a). The second term of the Equation (14) arises from the volume element with respect to r .

First, a conventional MTD calculation using $N_w = 4$ walkers was carried out at 523 K with the hill-width $\sigma = 0.026$ Å, the hill-height $h = 0.2$ kcal/mol, and the time scales $\tau_{\text{cv}} = 40$ fs, $\tau_{\text{fc}} = 10$ fs. The run was terminated after about 137 ps when one of the walkers first crossed the maximum point in $A(r)$ at $r = 2.39$ Å. This is to focus on the sampling at the free energy level around 40 kcal/mol near the transition state of the reaction. The free

energy curve estimated as the sum of hills at that time, $A(r) \approx -W_0(r)$, is shown in Figure 4(b). Then isothermal MD calculation at 523 K was conducted for 4 trajectories with 75 ps each in the presence of the bias potential $W_0(r)$ and an artificial wall to prevent from visiting the product region, $r > 2.39 \text{ \AA}$. From the histogram of $\rho_{W_0}(r)$ the refined free energy curve $A(r) \approx -W_1(r)$ was computed according to Equation (3), which is shown in Figure 4(c). The MD calculation was continued with the upgraded bias potential $W_1(r)$ for an additional 1.5 ns. The free energy curve $A(r) \approx -W_2(r)$, which was further refined by the histogram of $\rho_{W_1}(r)$, is shown in Figure 4(d).

Figure 4(b)-(d) clearly shows that the refinement works well as we expected. When compared with the exact solution, the root-mean-square error (RMSE) and the maximum absolute error (MaxAE) were found 1.15 and 2.37 kcal/mol, respectively, in conventional MTD, while the RMSE and MaxAE dropped significantly to 0.43 and 1.05 kcal/mol (0.16 and 0.50 kcal/mol), respectively, in the first (second) refinement of MTD. Note that the errors after final refinement became smaller than the thermal energy, $k_B T = 1.04 \text{ kcal/mol}$ at 523 K, at least for this test case. The performance in the production simulation will be also confirmed in the next section.

Simulation results

Three-dimensional (3D) free energy profiles $A(d, \phi, n)$ of the aqueous solutions of *2R,5R*-HDO were obtained from the DFTB MTD simulations. For the clarity of our results, they are displayed by means of a two-dimensional (2D) subspace. Here we take the subspace of the free energy minimum with respect to the reduced CV as

$$A_\phi(d, n) = \min_{\phi} A(d, \phi, n), \quad A_n(d, \phi) = \min_n A(d, \phi, n). \quad (15)$$

These subspaces contains both the minimum and saddle points on the 3D free energy surface since the primary reactive mode is parallel to the CO bond-exchange variable, d . In Figures 5 and 6, $A_\phi(d, n)$ and $A_n(d, \phi)$ are shown where the origin is chosen to the value at the transition state (the minimum on the $d = 0$ lines). Thus, the darker colors on the free energy surface mean that the energy basin is deeper.

In Figure 5, the free energy surfaces $A_\phi(d, n)$ and $A_n(d, \phi)$ are shown for the case of

2*R*,5*R*-HDO in H₂CO₃ solutions. The two figures, 5(a) and 5(b), are displayed to compare the quality of the free energy surfaces obtained before and after the refinement of conventional MTD. The smoothness of the free energy surfaces clearly reflects the improvement by the refinement. The refined MTD was able to remove the roughness with less spurious peaks and valleys in the surface. The difference of free energy profiles, Figures 5(a) and (b), is shown in Figure 5(c). In the unrefined MTD, the RMSE (MaxAE) with respect to negative free energy region of $A_\phi(d, n)$ and $A_n(d, \phi)$ were estimated to be 1.5 and 1.6 kcal/mol (3.8 and 7.3 kcal/mol), respectively. Figure 6(a-c) shows the results of the 2D subspaces, $A_\phi(d, n)$ and $A_n(d, \phi)$, in the HCl solution, in the H₂SO₄ solution and in pure water, respectively, obtained from the refined MTD. Compared with our earlier works using conventional MTD,^{53,54} the refined MTD has significantly improved the quality of free energy surface.

The regions colored in black/brown are the deepest regions in the free energy, which correspond to the most stable structures of 2*R*,5*R*-HDO in the aqueous solution. Those regions are located at $d \simeq 4.1 \text{ \AA}$, meaning that the HDO molecule is in the extended conformation, and at $n \simeq 1.2$, meaning that the hydroxyl group of the molecule is in the neutral form ($n \geq 1$ because the hydroxyl group creates hydrogen bonds with surrounding water molecules). It can be seen that the 2*R*,5*R*-HDO favors the conformations with three dihedral angles about $\phi \simeq 60^\circ, 180^\circ, \text{ and } 300^\circ$. A shallow minima found at $d \simeq 2.3 \text{ \AA}$ indicates the contracted conformation of the 2*R*,5*R*-HDO. As time evolves, the MTD simulation explores from the lower free energy regions colored in orange to the higher free energy regions colored in yellow. Conventional MTD runs were terminated when we observe an escape from the CV range ($d \leq 0 \text{ \AA}$) after visiting the highest free energy region in grey. After that, the free energy surface was refined by the MD runs sampling the whole CV space in color. The saddle point of the free energy surface thus obtained at $d = 0 \text{ \AA}$ corresponds to the transition state whereby the $\text{O} \cdots \text{C}-\text{O} \rightarrow \text{O}-\text{C} \cdots \text{O}$ bond exchange occurs before forming an ether ring of cis-2,5-dimethyltetrahydrofuran (DMTHF). The transition state at specific conformation $90^\circ < \phi < 270^\circ$, indicate that the reaction takes place via the S_N2 reaction. This result is consistent with the high stereoselectivity of this reaction confirmed experimentally.⁴⁸ For the acidic solutions, the coordination number $n \simeq 2.0$ at the transition state implies that the reaction involves the protonation of hydroxyl group.

As to the reaction process just mentioned, the DFTB simulations were able to reproduce the results of our earlier Car-Parrinello MD (CPMD) simulation based on density functional theory (DFT).⁵³ However, a difference between DFT and DFTB was seen in the protonation mechanism of the hydroxyl group. The protonation occurs via a Grotthuss relay mechanism among water in the case of DFT, while the proton was mostly captured from the acid group in the case of DFTB. We believe this is an artifact of DFTB3 with the 3ob parameter set, which unfortunately could not also be removed by applying the D3 dispersion correction.⁷⁰ In spite of this, the DFTB simulations were able to capture qualitatively correct features of the free energy profiles as compared with the DFT simulations. This was the case in our previous work on the dehydration of PTO as well.⁵⁴

In Table 1, we summarize the results of the computed free energy barriers ΔA for the dehydration reaction of *2R,5R*-HDO. Here the statistical error is estimated from the mean-square deviation of the block averages of ΔA with respect to three thirds of the MD trajectories in the refinement stage. The ΔA_{dftb} values obtained from the present DFTB simulation of the aqueous HCl solutions was 35.4 ± 0.4 kcal/mol, in reasonable agreement with the value $\Delta A_{\text{cpmd}} = 36$ kcal/mol obtained in our previous CPMD simulation. Note here that ΔA_{cpmd} is computed for a system of protonated solution, where the effect of counter anions were replaced by a uniform background charge. On the other hand, the DFTB simulation predicts a substantially lower barrier of $\Delta A_{\text{dftb}} = 26.3 \pm 1.0$ kcal/mol for the aqueous H_2SO_4 solution. The low barrier is consistent with the finding of an early experiment⁴⁴ that the *2R,5R*-HDO dehydration proceeded in sulfuric acid at a temperature as low as 293 K. In fact, at this temperature the experimental barrier roughly estimated from TST with the rate of 1 per hour becomes as low as 22.4 kcal/mol. However, in the same experiment it was reported that the reaction in sulfuric acid lead to a racemic mixture of *cis*- and *trans*-2,5-DMTHF suggesting an $\text{S}_{\text{N}}1$ mechanism. This is in contrast to the $\text{S}_{\text{N}}2$ mechanism found in this computational study, in spite of the difference in temperature and pressure conditions (573 K, 20 MPa). The peculiarity of sulfuric acid is an issue to be addressed in the future.

The computed values were compared with the experimental free energy barrier, ΔA_{exptl} . This was estimated from the measurement of the half life of *2R,5R*-HDO, $\tau_{\text{exptl}} \approx 30$ minutes,

combined with the TST reaction rate,

$$k_{\text{exptl}} = 2 \frac{k_{\text{B}}T}{h} \exp\left(-\frac{\Delta A_{\text{exptl}}}{k_{\text{B}}T}\right) = \frac{\ln 2}{\tau_{\text{exptl}}}, \quad (16)$$

where h is the Planck constant. The factor 2 in the second equality arises from the fact that the reaction can proceed in two equivalent pathways leading to the same *cis*-2,5-DMTHF product with either O₇ or O₈ oxygen. In the case of the H₂CO₃ solution, $\Delta A_{\text{dftb}} = 39.6 \pm 0.3$ kcal/mol agrees surprisingly well with $\Delta A_{\text{exptl}} = 40.0$ kcal/mol. The ΔA_{dftb} values for the HCl solution was lower than that for the H₂CO₃ solution by 4.1 ± 0.4 kcal/mol. This is consistent with experimental results from different research groups which showed that the reaction of 2,5-hexanediol and 1,4-butanediol is accelerated by aqueous solutions as the acidity is stronger.^{44–46,48,50} On the other hand, ΔA_{dftb} values for pure water, 50.6 ± 0.8 kcal/mol, was much higher than that for the H₂CO₃ solution. This is also consistent with the experimental observation that the reaction in the H₂CO₃ solution is at least 50 times faster than that in pure water.⁴⁸

Why is the free energy barrier lower for strong acidic solutions? Comparing the free energy profiles in Figures 5 and 6, a notable difference can be found in the regions $(d, n) \approx (2.0 \text{ \AA}, 2.0)$ and $(d, n) \approx (4.0 \text{ \AA}, 2.0)$, which correspond to the protonation of the hydroxyl group in the confined and extended conformations, respectively. The free energy barrier is significantly damped in the H₂SO₄ solution when those regions become stable minima. Thus, the value of free energy barrier is probably linked to the stability of the protonated state of HDO in the acidic aqueous solutions.

At the end of this section, we show the typical molecular configurations taken from the present DFTB MTD simulations. Figures 7 and 8 are the results from the simulations of H₂CO₃ solutions and pure water, respectively. In the aqueous H₂CO₃ solution, it is found that the reaction proceeds in the following manner. At the equilibrium state, *2R,5R*-HDO is in its extended conformation. Then at the transition state *2R,5R*-HDO is protonated in the confined conformation. Finally the ether ring is formed to become *cis*-2,5-DMTHF. The reaction processes of the aqueous HCl and H₂SO₄ solutions were found to be in the S_N2 mechanism, and were similar to the case of the aqueous H₂CO₃ solution.

The transition state configuration in pure water is different from that in acidic water.

Specifically, the C–OH bond is broken without protonation in pure water. This can be seen in Figure 6(c) that the coordination number is $n \approx 1.4$ when $d = 0$, in contrast to the case of acidic solutions, $n \approx 2.0$. Yet, the transition state in pure water leads to the same product, namely cis-2,5-DMTHF. These results demonstrate the possibility that polyalcohol dehydration in high-temperature water follows different pathways depending on the presence of excess protons. Thus we speculate that the anomalous acidity dependence on the reaction kinetics as discussed in experiments^{45,46} is ascribed to the multiple pathways competing with each other.

CONCLUSION

In this paper, we have proposed a refinement of MTD simulations a posteriori to help reduce the error of free energy barriers of reactive processes. Making use of the hills obtained from the MTD as a time-invariant bias potential, a canonical sampling is performed at the same free energy level as the reaction barriers. This method was used to study the dehydration reaction of HDO in high-temperature water with a focus on the acid dependence. It was found that the acid species affects significantly the free energy profile of the S_N2 reaction pathway, which can be traced back to the protonation free energy of hydroxyl group in hydrated environment. This result provides a theoretical foundation for the experimental observations that the rate of this reaction is accelerated by the acidity without losing the R/S stereoselectivity.

Finally, it is worth pointing out that the refined MTD was very helpful in this study. For the calibration of the acid dependence, it is essential to estimate accurately the free energy barriers which can differ from each other only by few $k_B T$. This method works as a post-processing, making a full use of the data from conventional MTD simulations that usually have taken a large computation effort to obtain already. Importantly, the method is simple but robust, and does not require any changes to existing software for conventional MTD simulations. Our method allows the refinement of the bias obtained from any other variants of MTD, such as the well-tempered metadynamics.

ACKNOWLEDGMENTS

The computations in this work were mostly done using the supercomputer facilities at the Japan Atomic Energy Agency and the Institute for Solid State Physics, The University of Tokyo. M.S. thanks financial support from JSPS KAKENHI (18H05519, 18H01693, 18K05208) and MEXT Program for Promoting Researches on the Supercomputer Fugaku (Fugaku Battery & Fuel Cell Project). J.R.A. acknowledges financial support from MINECO through Grant CTQ2017-87773-P/AEI/FEDER, UE and Catalan DURSI (2017SGR348 Grant). The authors thank Yong Lik Chang for the cooperation at early stage of this work, Dr. Bo Thomsen in JAEA for helping us to prepare the molecular figures, and Dr. Alex Malins in JAEA for proofreading the text.

References

1. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 2017.
2. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, Vol. 1; Elsevier, 2001.
3. Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer, 2007.
4. Tuckerman, M. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press, 2010.
5. Awasthi, S.; Nair, N. N. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2019**, *9*(3), e1398.
6. Nagaoka, M.; Okuyama-Yoshida, N.; Yamabe, T. *The Journal of Physical Chemistry A* **1998**, *102*(42), 8202–8208.
7. Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Peptide Science: Original Research on Biomolecules* **2001**, *60*(2), 96–123.

8. Laio, A.; Parrinello, M. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*(20), 12562–12566.
9. Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. E. *The Journal of Chemical Physics* **2002**, *116*(11), 4389–4402.
10. Héning, J.; Chipot, C. *The Journal of Chemical Physics* **2004**, *121*(7), 2904–2914.
11. Maragliano, L.; Vanden-Eijnden, E. *Chemical Physics Letters* **2006**, *426*(1-3), 168–175.
12. Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *The Journal of Chemical Physics* **2008**, *128*(14), 144120.
13. Kästner, J. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*(6), 932–942.
14. Morishita, T.; Itoh, S. G.; Okumura, H.; Mikami, M. *Physical Review E* **2012**, *85*(6), 066702.
15. Itoh, S. G.; Okumura, H. *Journal of Chemical Theory and Computation* **2013**, *9*(1), 570–581.
16. Chen, M.; Yu, T.-Q.; Tuckerman, M. E. *Proceedings of the National Academy of Sciences* **2015**, *112*(11), 3235–3240.
17. Shiga, M.; Tuckerman, M. E. *The Journal of Physical Chemistry Letters* **2018**, *9*(21), 6207–6214.
18. Mitsuta, Y.; Shigeta, Y. *Journal of Chemical Theory and Computation* **2020**, *16*(6), 3869–3878.
19. Iannuzzi, M.; Laio, A.; Parrinello, M. *Physical Review Letters* **2003**, *90*(23), 238302.
20. Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *The Journal of Physical Chemistry B* **2005**, *109*(14), 6676–6687.
21. Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *The Journal of Physical Chemistry B* **2006**, *110*(8), 3533–3539.

22. Laio, A.; Gervasio, F. L. *Reports on Progress in Physics* **2008**, *71*(12), 126601.
23. Barducci, A.; Bussi, G.; Parrinello, M. *Physical Review Letters* **2008**, *100*(2), 020603.
24. Branduardi, D.; Bussi, G.; Parrinello, M. *Journal of Chemical Theory and Computation* **2012**, *8*(7), 2247–2254.
25. Dama, J. F.; Parrinello, M.; Voth, G. A. *Physical Review Letters* **2014**, *112*(24), 240602.
26. Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. *Journal of Chemical Theory and Computation* **2014**, *10*(9), 3626–3633.
27. Tiwary, P.; Parrinello, M. *The Journal of Physical Chemistry B* **2015**, *119*(3), 736–742.
28. Invernizzi, M.; Parrinello, M. *The Journal of Physical Chemistry Letters* **2020**, *11*(7), 2731–2736.
29. Bussi, G.; Laio, A. *Nature Reviews Physics* **2020**, *2*, 200–212.
30. Smiatek, J.; Heuer, A. *Journal of Computational Chemistry* **2011**, *32*(10), 2084–2096.
31. Mones, L.; Bernstein, N.; Csányi, G. *Journal of Chemical Theory and Computation* **2016**, *12*(10), 5100–5110.
32. Donati, L.; Keller, B. G. *The Journal of Chemical Physics* **2018**, *149*(7), 072335.
33. Marinova, V.; Salvalaglio, M. *The Journal of Chemical Physics* **2019**, *151*(16), 164115.
34. Ono, J.; Nakai, H. *Chemical Physics Letters* **2020**, *751*, 137384.
35. Schäfer, T. M.; Settanni, G. *Journal of Chemical Theory and Computation* **2020**, *16*(4), 2042–2052.
36. Giberti, F.; Cheng, B.; Tribello, G. A.; Ceriotti, M. *Journal of Chemical Theory and Computation* **2020**, *16*(1), 100–107.
37. Zhang, Y.; Voth, G. A. *Journal of Chemical Theory and Computation* **2011**, *7*(7), 2277–2283.

38. Awasthi, S.; Kapil, V.; Nair, N. N. *Journal of Computational Chemistry* **2016**, *37*(16), 1413–1424.
39. Chen, M.; Cuendet, M. A.; Tuckerman, M. E. *The Journal of Chemical Physics* **2012**, *137*(2), 024102.
40. Awasthi, S.; Nair, N. N. *The Journal of Chemical Physics* **2017**, *146*(9), 094108.
41. Akiya, N.; Savage, P. E. *Chemical Reviews* **2002**, *102*(8), 2725–2750.
42. Savage, P. E. *The Journal of Supercritical Fluids* **2009**, *47*(3), 407–414.
43. Kruse, A.; Dahmen, N. *The Journal of Supercritical Fluids* **2015**, *96*, 36–45.
44. Molnár, Á.; Felföldi, K.; Bartók, M. *Tetrahedron* **1981**, *37*(11), 2149–2151.
45. Nagai, Y.; Matubayasi, N.; Nakahara, M. *Bulletin of the Chemical Society of Japan* **2004**, *77*(4), 691–697.
46. Hunter, S. E.; Ehrenberger, C. E.; Savage, P. E. *The Journal of Organic Chemistry* **2006**, *71*(16), 6229–6239.
47. Yamaguchi, A.; Hiyoshi, N.; Sato, O.; Bando, K. K.; Shirai, M. *Green Chemistry* **2009**, *11*(1), 48–52.
48. Yamaguchi, A.; Hiyoshi, N.; Sato, O.; Shirai, M. *ACS Catalysis* **2011**, *1*(1), 67–69.
49. Yamaguchi, A.; Hiyoshi, N.; Sato, O.; Shirai, M. *Green Chemistry* **2011**, *13*(4), 873–881.
50. Yamaguchi, A.; Hiyoshi, N.; Sato, O.; Shirai, M. *Catalysis Today* **2012**, *185*(1), 302–305.
51. Sato, O.; Yamaguchi, A.; Shirai, M. *Catalysis Communications* **2015**, *68*, 6–10.
52. Yamaguchi, A.; Muramatsu, N.; Mimura, N.; Shirai, M.; Sato, O. *Physical Chemistry Chemical Physics* **2017**, *19*(4), 2714–2722.
53. Ruiz-Barragan, S.; Ribas-Ariño, J.; Shiga, M. *Physical Chemistry Chemical Physics* **2016**, *18*(47), 32438–32447.

54. Chang, Y. L.; Sasaki, T.; Ribas-Ariño, J.; Machida, M.; Shiga, M. *The Journal of Physical Chemistry B* **2019**, *123*(7), 1662–1671.
55. Anderson, E.; Bai, Z.; Bischof, C.; Blackford, L. S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Ostrouchov, S.; Sorensen, D. *LAPACK Users' guide*; SIAM, 1999.
56. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*(45), 11225–11236.
57. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*(11), 7260.
58. Elstner, M. *The Journal of Physical Chemistry A* **2007**, *111*(26), 5614–5621.
59. Gaus, M.; Cui, Q.; Elstner, M. *Journal of Chemical Theory and Computation* **2011**, *7*(4), 931–948.
60. Gaus, M.; Goez, A.; Elstner, M. *Journal of Chemical Theory and Computation* **2013**, *9*(1), 338–354.
61. Kubillus, M.; Kubar, T.; Gaus, M.; Rezac, J.; Elstner, M. *Journal of Chemical Theory and Computation* **2015**, *11*(1), 332–342.
62. Nosé, S. *The Journal of Chemical Physics* **1984**, *81*(1), 511–519.
63. Hoover, W. G. *Physical Review A* **1985**, *31*(3), 1695–1697.
64. Martyna, G. J.; Klein, M. L.; Tuckerman, M. *The Journal of Chemical Physics* **1992**, *97*(4), 2635–2643.
65. Shinoda, W.; Shiga, M.; Mikami, M. *Physical Review B* **2004**, *69*(13), 134103.
66. PIMD version 2.4.0. <https://ccse.jaea.go.jp/software/pimd/index.en.html>.
67. Ruiz-Barragan, S.; Ishimura, K.; Shiga, M. *Chemical Physics Letters* **2016**, *646*, 130–135.

68. Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Di Carlo, A.; Suhai, S. *Journal of Physics: Condensed Matter* **2002**, *14*(11), 3015.
69. Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshayé, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C.; Sternberg, M.; Stöhr M, and Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T. *The Journal of Chemical Physics* **2020**, *152*(12), 124101.
70. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*(15), 154104.

Figure 1: A schematic figure for conventional MTD and its refinement. The bias potentials are depicted for those after (a) conventional MTD in blue, (b) the first refinement in pink, and (c) the second refinement in orange.

Figure 2: Computational procedure of the refined MTD method.

Figure 3: The structure of *2R,5R*-HDO molecule in the extended conformation. The atoms related to the CVs, d , n and ϕ , are colored in yellow, green and blue, respectively.

Figure 4: Tests on a diatomic model. (a) The functions $V(r)$ and $A(r)$, and the comparison of $A(r)$, and the free energy curves (b) obtained from conventional MTD, (c) refined at the first iteration and (d) refined at the second iteration.

Figure 5: Calculated free energy surfaces $A_n(d, \phi)$ (left panels) and $A_\phi(d, n)$ (right panels) of *2R,5R*-HDO in the H_2CO_3 solution obtained from (a) conventional MTD and (b) refined MTD. Figure (c) shows the difference of Figures (a) and (b).

Figure 6: Calculated free energy surfaces $A_n(d, \phi)$ (left panels) and $A_\phi(d, n)$ (right panels) of *2R,5R*-HDO in (a) the HCl solution, (b) the H_2SO_4 solution, and (c) pure water obtained from refined MTD.

Figure 7: Dehydration process of *2R,5R*-HDO in H_2CO_3 solution: (1) HDO is in its extended conformation. (2) HDO is its confined conformation and the OH group is protonated. (3) The C-O and O-C bonds are exchanged. (4) The ether ring of *cis*-2,5-DMTHF is formed. After this, the proton bonded to the ether will be released to water to complete the reaction.

Figure 8: Dehydration process of *2R,5R*-HDO in pure water: (1) HDO is its confined conformation. (2) The CO bond is elongated when the OH group is surrounded by water. (3) The OH^- ion is eliminated by water. (4) The ether ring of *cis*-2,5-DMTHF is formed. After this, the proton bonded to the ether will be released to water to complete the reaction.

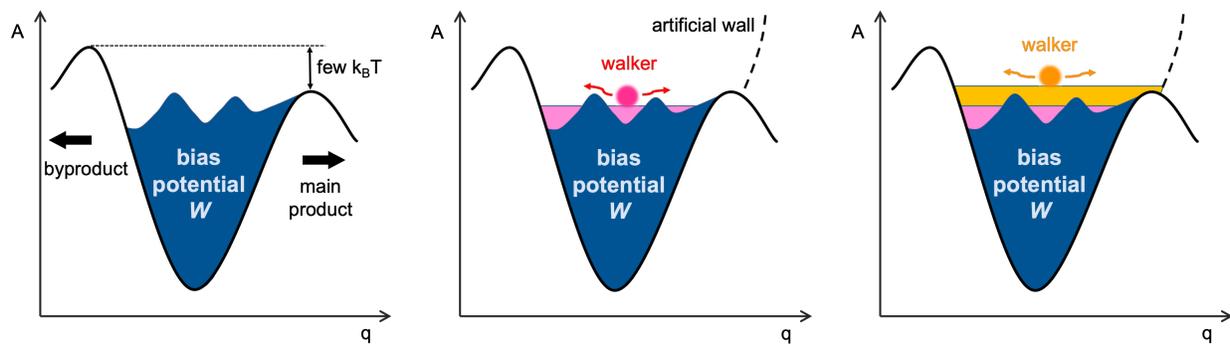


Figure 1, Kondo et al., J. Comput. Chem.

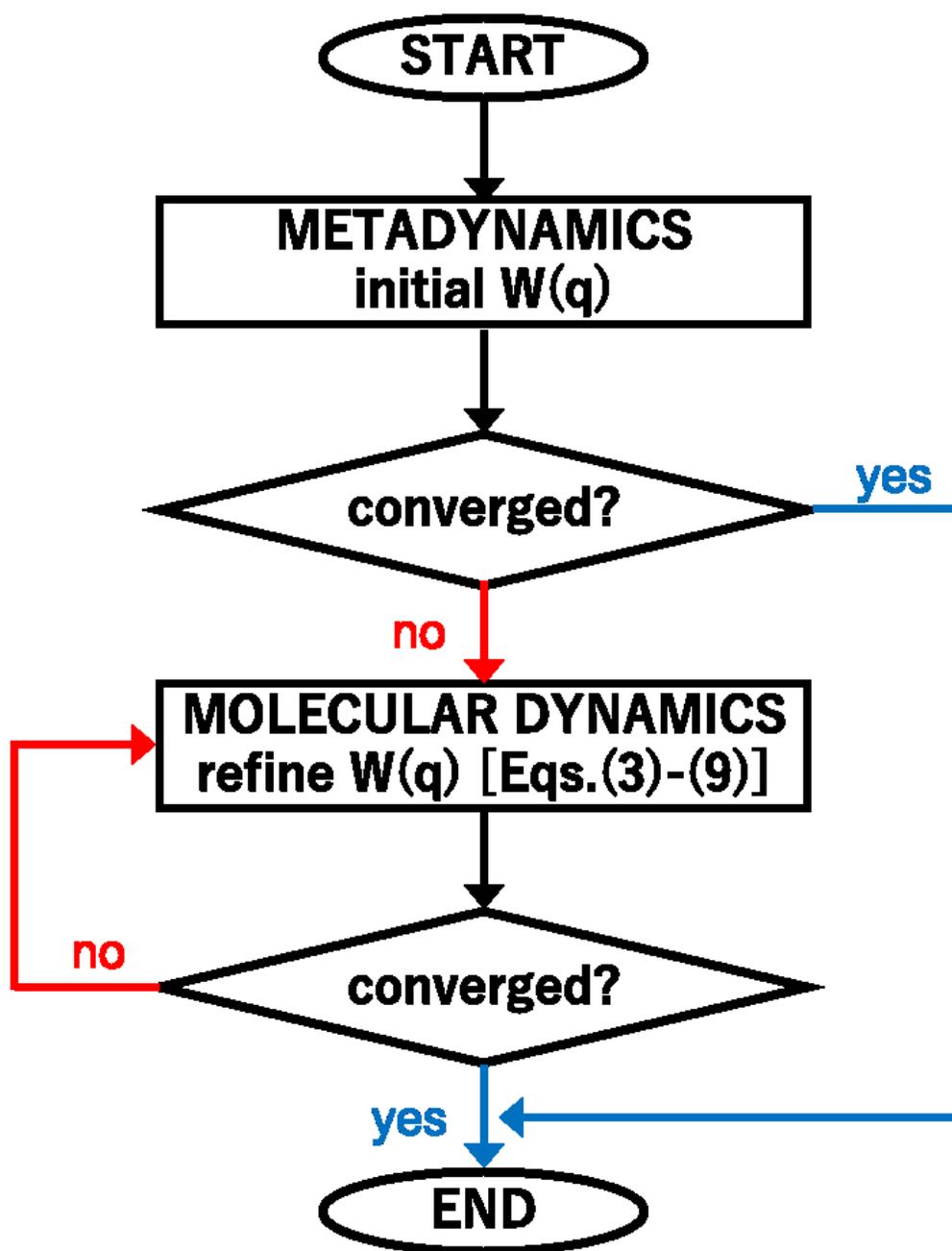


Figure 2, Kondo et al., J. Comput. Chem.

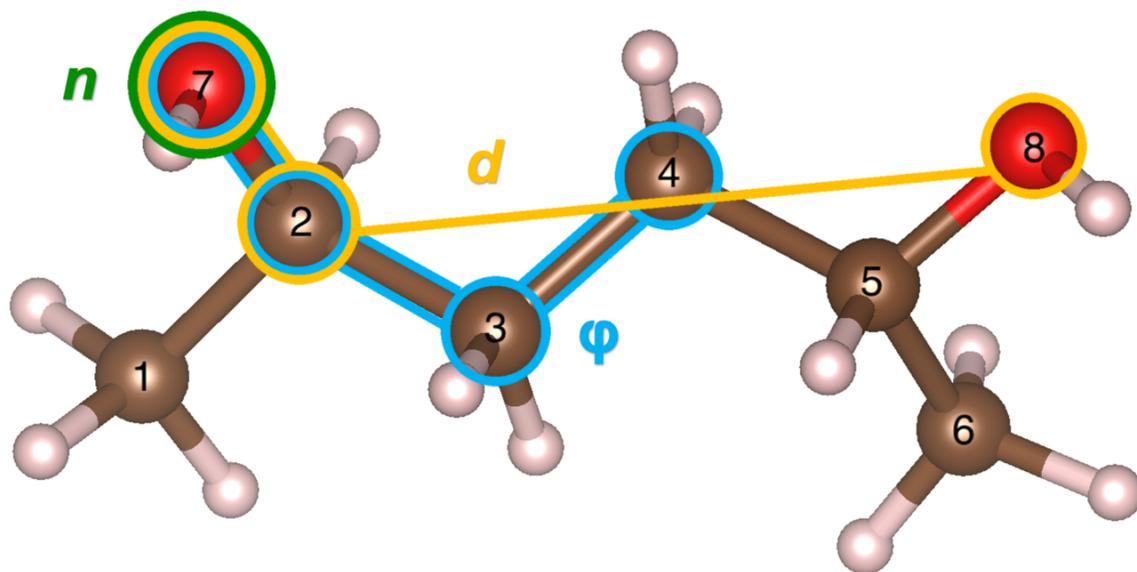


Figure 3, Kondo et al., J. Comput. Chem.

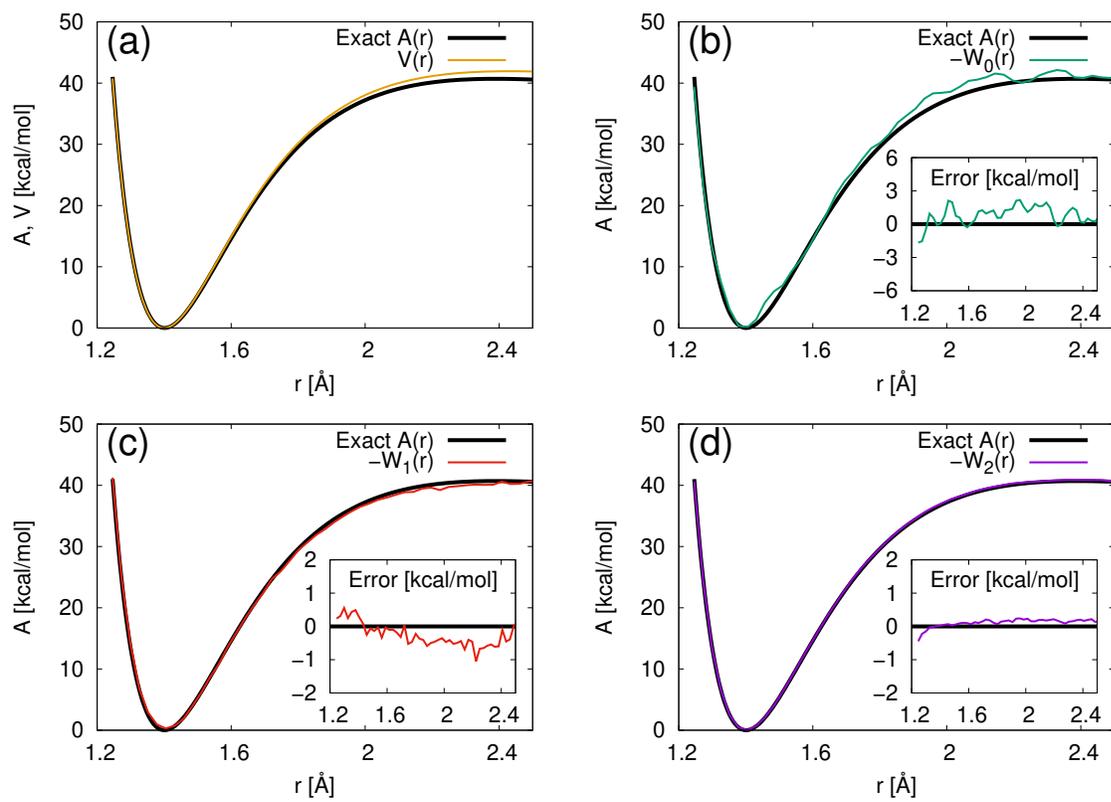


Figure 4, Kondo et al., J. Comput. Chem.

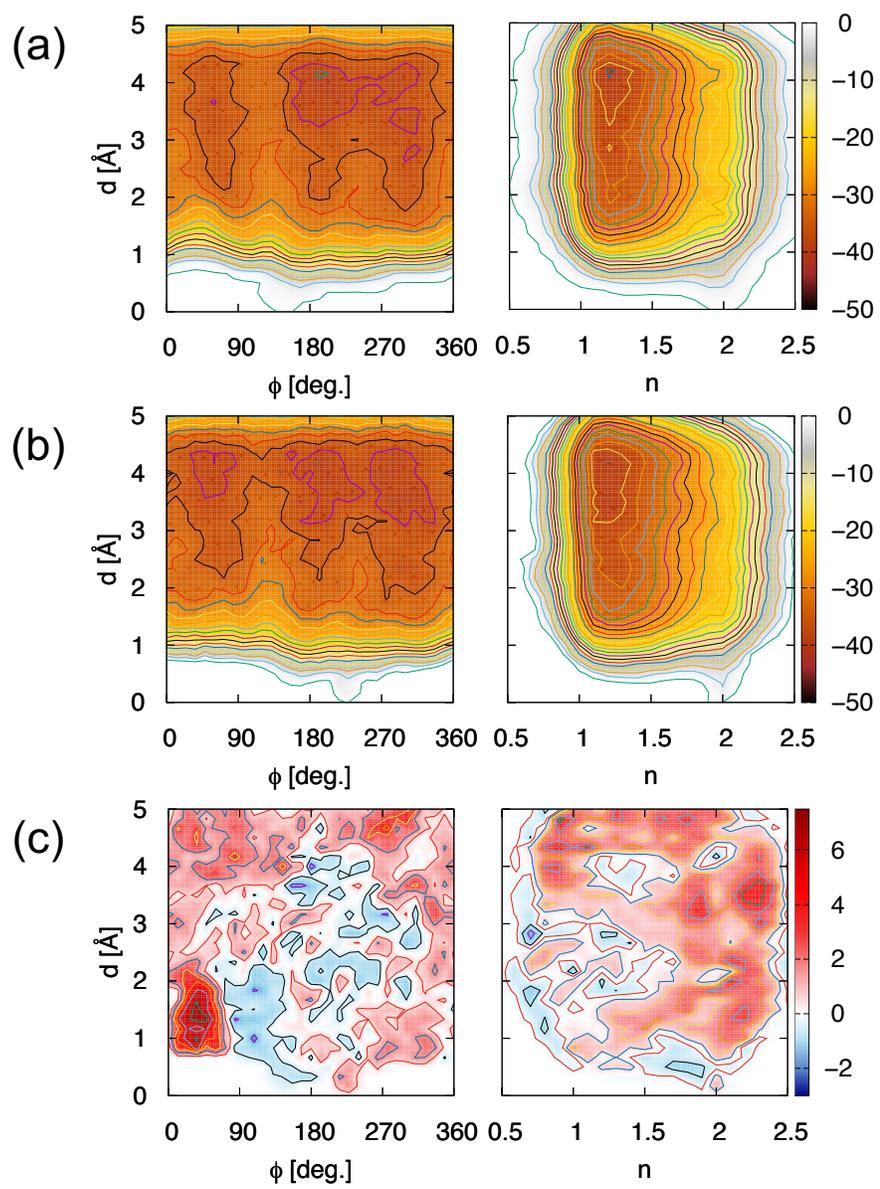


Figure 5, Kondo et al., J. Comput. Chem.

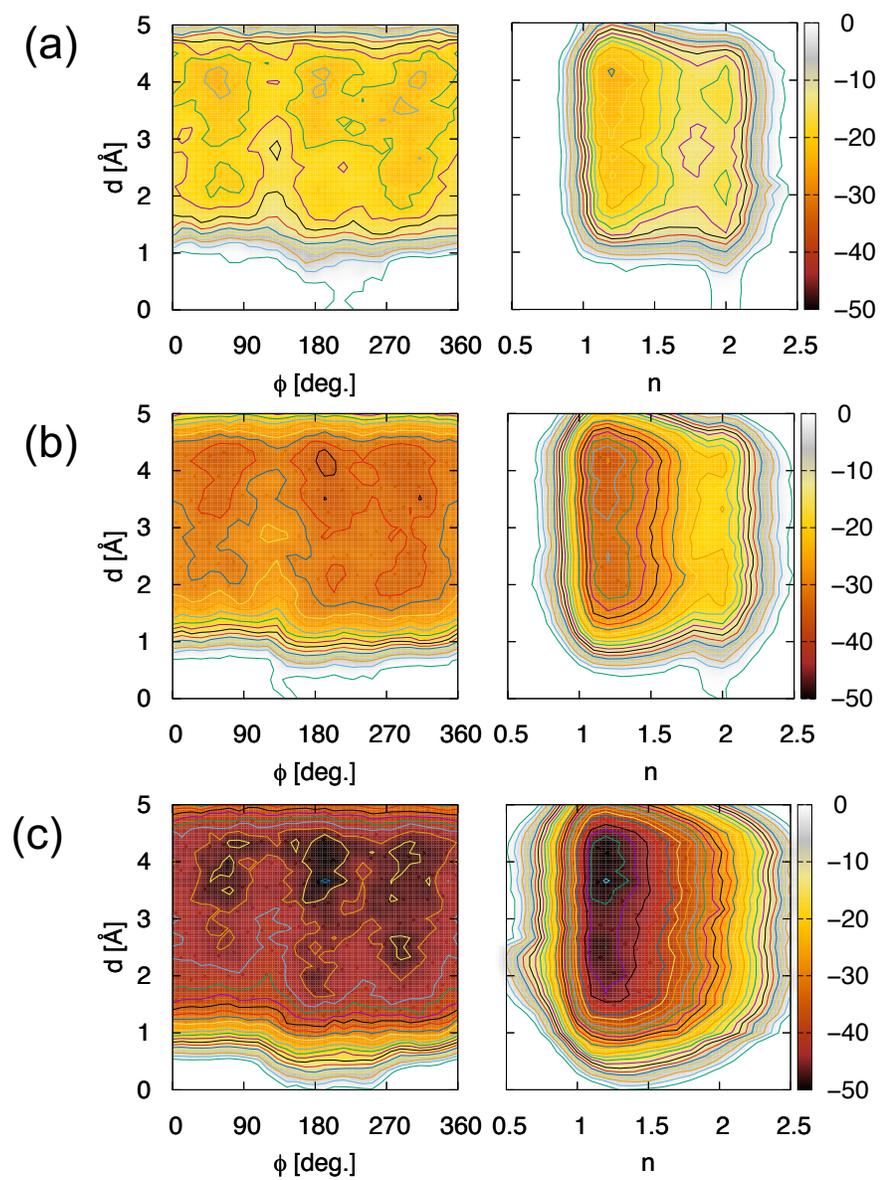


Figure 6, Kondo et al., J. Comput. Chem.

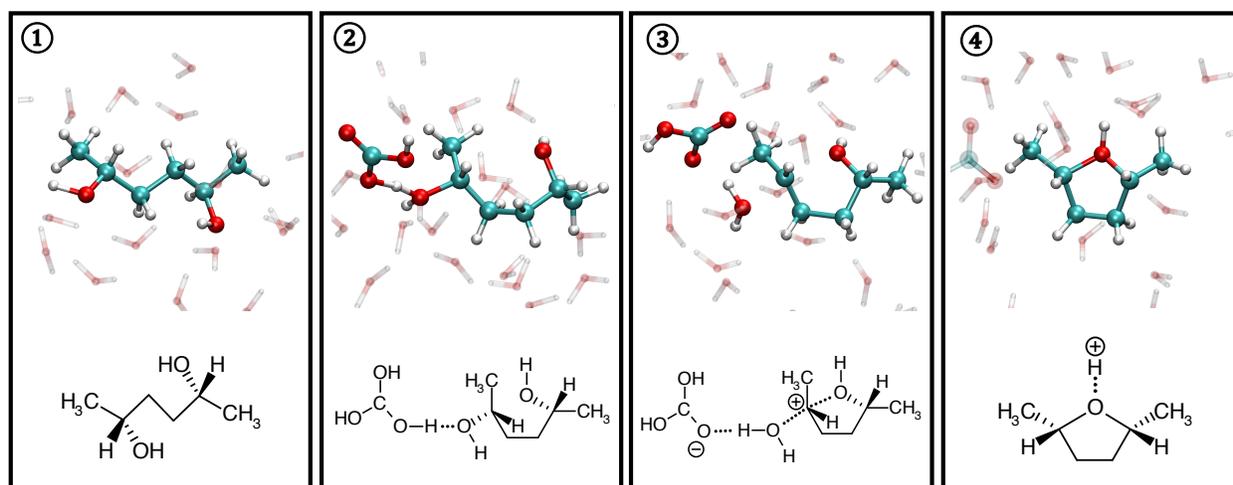


Figure 7, Kondo et al., J. Comput. Chem.

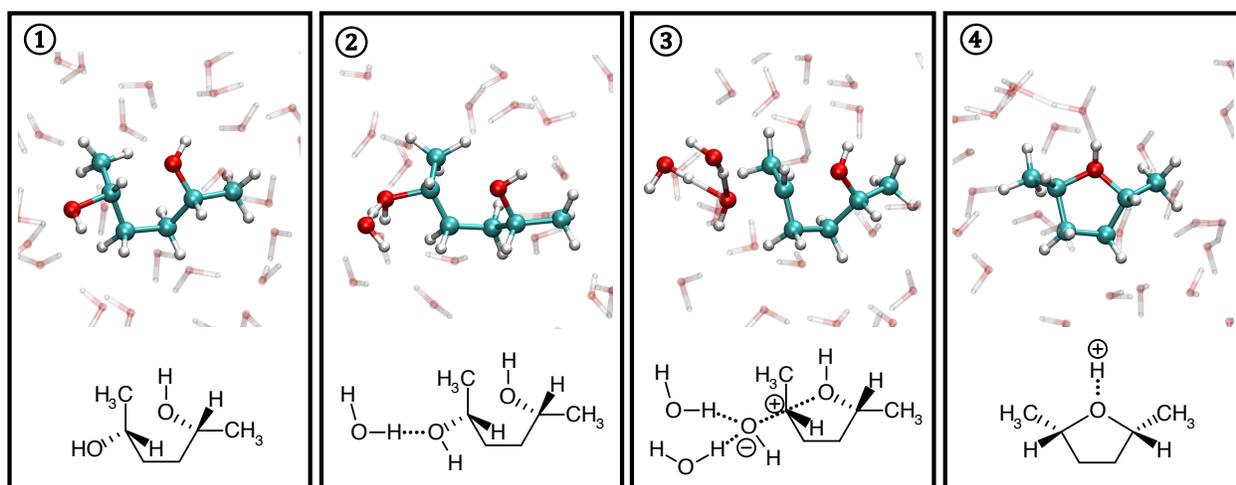


Figure 8, Kondo et al., J. Comput. Chem.

Table 1. Free energy barrier ΔA of $2R,5R$ -HDO dehydration in kcal/mol.

solution	DFTB (this work)	CPMD ¹	experiment ²
H ₂ SO ₄	26.3 ± 1.0		
HCl	35.4 ± 0.4		
H ₂ CO ₃	39.6 ± 0.3		40.0
pure water	50.6 ± 0.8		44.1 <
protonated water		36	

¹CPMD calculation from Reference⁵³.

²Measurement of half life of HDO from Reference⁴⁸ (30 minutes in H₂CO₃, and more than 25 hours in pure water) is combined with the TST, Eq.16.