PCCP

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2015, **17**, 2757

Reaction sampling and reactivity prediction using the stochastic surface walking method†

Xiao-Jie Zhang and Zhi-Pan Liu*

The prediction of chemical reactivity and thus the design of new reaction systems are the key challenges in chemistry. Here, we develop an unbiased general-purpose reaction sampling method, the stochastic surface walking based reaction sampling (SSW-RS) method, and show that the new method is a promising solution for reactivity prediction of complex reaction systems. The SSW-RS method is capable of sampling both the configuration space of the reactant and the reaction space of pathways, owing to the combination of two recently developed theoretical methods, namely, the stochastic surface walking (SSW) method for potential energy surface (PES) exploration and the double-ended surface walking (DESW) method for building pathways. By integrating with first principles calculations, we show that the SSW-RS method can be applied to investigate the kinetics of complex organic reactions featuring many possible reaction channels and complex hydrogen-bonding networks, as demonstrated here using two examples, epoxypropane hydrolysis in aqueous solution and β-D-glucopyranose decomposition. Our results show that simultaneous sampling of the soft hydrogen-bonding conformations and the chemical reactions involving hard bond making/breaking can be achieved in the SSW-RS simulation, and the mechanism and kinetics can be predicted without *a priori* information on the system. Unexpected new chemistry for these reactions is revealed and discussed. In particular, despite many possible pathways for β-D-glucopyranose decomposition, the SSW-RS shows that only β-D-glucose and levoglucosan are kinetically preferred direct products and the 5- or 7-member ring products should be secondary products derived from β-D-glucose or levoglucosan. As a general tool for reactivity prediction, the SSW-RS opens a new route for the design of rational reactions.

Received 2nd October 2014,
Accepted 1st December 2014

DOI: 10.1039/c4cp04456h

www.rsc.org/pccp

1. Introduction

The chemical reactivity of matter is a subject of fundamental importance in chemistry. It has long been a goal to predict the reactivity of a given reactant without recourse to experiment. Mathematically, this requires the exploration of the PES of the reaction system to identify the low energy reaction pathways that can occur under certain experimental conditions.[1] The proper sampling of both the minima and the barrier regions on the PES is essential in order to quantify the reaction kinetics. The reactivity prediction is however long regarded as a formidable task, although a chemical reaction may involve only a few molecules at the reaction center. The outstanding difficulties include the large configuration space of molecules interacting with the environment, the high barriers separating the minima and a large number of possible reaction channels towards different products. How to reduce the dimensionality in reaction space and how to improve the efficiency of locating the favorable reaction channels are open questions in theoretical chemistry.

The enhanced sampling techniques based on molecular dynamics (MD) methods, such as string MD,[2] umbrella-sampling MD,[3] metadynamics[4,5] and replica exchange MD,[6,7] were most often utilized to investigate the PES of reactions.[8–10] A predefined reaction coordinate is critical to bias the simulation towards the desired product, but it may yield an inaccurate transition state (TS), free energy and rate constant when the chosen reaction coordinate does not capture the molecular mechanism. In order to sample reaction channels and identify the most favorable one in kinetics, alternative methods must be utilized, such as kinetics Monte Carlo,[11–14] transition path sampling,[15–17] discrete path sampling[18,19] and minima hopping guided pathway searching method,[20,21] which however are computationally much more demanding. In general, these path sampling methods still require

Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, Key Laboratory of Computational Physical Science (Ministry of Education), Fudan University, Shanghai 200433, China. E-mail: zpliu@fudan.edu.cn

† Electronic supplementary information (ESI) available: Ten lowest energy pathways for vinyl alcohol and formaldehyde recombination; reaction snapshots for Epo-P2, Epo-P3 and Epo-P4 in epoxypropane hydrolysis; the conformational spectrum and typical structures of β-D-glucopyranose; the structures of 15 possible products from β-D-glucopyranose decomposition. See DOI: 10.1039/c4cp04456h

some basic knowledge of the PES, *e.g.* an initial path that connects the reactant and product and a database of minima in the discrete path sampling method. For the more widely utilized kinetics Monte Carlo method, it relies on a database of the rate of all elementary reactions (events), from which a rate list can be established and updated dynamically to describe the time evolution of the system. The transition path sampling method requires a path ensemble, where a number of reaction pathways are present to connect the predefined reactant and product. Monte Carlo importance sampling is then utilized to sample the path ensemble by, for example, performing short MD simulations from the barrier region that is distinguishable using the pre-designed order parameter. Due to the computational cost, the transition path sampling and discrete path sampling methods are most often applied to systems that can be described by a classical potential (*e.g.* no explicit chemical bond making/breaking), such as the protein conformation change[22] and the particle morphology transformation.[23,24]

In recent years, new methods have been proposed aiming to explore the PES of chemical reactions starting from a given reactant, for example, gradient extremal following method,[25] reduced gradient following method,[26] anharmonic downward distortion following method[27–30] and artificial force induced reaction (AFIR) method. In general, these methods involve the pathway (or pseudopathway) collection with the explicit TS location. As a representative, the AFIR method integrated with first principles calculations was utilized to establish the reaction network of some organic molecules,[29,31] demonstrating the great potential for reactivity prediction. These methods focus on the sampling of possible reaction pathways by forcing the reaction to occur towards different products. The sampling of the conformation space of the reactant could be a problem for these methods, but it can be amended by combining with other efficient stochastic conformation-generation methods.

Very recently, we developed a new global optimization method, SSW method,[32] to explore PES unbiasedly *via* the smooth structure perturbation (surface walking), where the second derivative information is taken into account. The method has been utilized to identify the global minimum (GM) of complex systems, including short-ranged Morse clusters,[32] carbon fullerene up to 100 atoms[38] and Boron clusters.[33,34] One key strength of the SSW method is the ability to sample the minima on PES by following likely pathways, which enables the applicability of the method for both structure prediction and pathway searching.

Inspired by the recent progress of reaction pathway sampling, here we develop a general-purpose method, SSW-RS, for automated reaction sampling and reactivity prediction. The new method combines our recently developed SSW method and DESW method, and can identify low energy pathways without the need of *a priori* information on the reaction. The SSW-RS method is naturally parallel and thus can be applied to large and complex systems with modern computing facilities. By integrating with first principles calculations, we apply the SSW-RS method for the reactivity prediction of two organic reactions featuring many possible reaction channels and complex hydrogen-bonding networks. We show that the SSW-RS method has a high efficiency of reactivity prediction and can cope with reaction systems with large degrees of freedom and high barriers involved in chemical bond making/breaking.

2. Methodology

2.1 Theoretical background

The algorithm for the reaction sampling in the SSW-RS method is rooted in the continuous time Markov chain model for reaction kinetics[16,19] and here we briefly outline the theoretical background. The transition rate to a state α from the other states on the PES can be described by the Master equation,

$$\frac{\mathrm{d}p_\alpha}{\mathrm{d}t} = \sum_{\beta\neq\alpha}\left[k_{\alpha\beta}p_\beta(t) - k_{\beta\alpha}p_\alpha(t)\right] \quad (1)$$

where $k_{\alpha\beta}$ is the rate constant of transition from state β to state α and the $p_\alpha(t)$ is the population of state α, being a function of time t. Obviously, the construction of the transition rate for each pair of states (the transition rate matrix) is critical in kinetics, but can be extremely expensive in practice as the system size grows.[19,35,36]

As illustrated in Fig. 1a, we may consider a complex reaction network that contains a number of major intermediates, each represented by a large circle labeled using a capital letter from A to D. For each intermediate, there are a number of geometrically distinguishable configurations (microstates) represented by the small squares inside the large circle. Specifically, in molecular systems, these microstates may represent the same molecules at different configurations, *e.g.* different H-bonding networks and *cis–trans* isomers, and each intermediate could represent the molecules with different bonding patterns.

The equilibrium assumption can often be introduced to simplify the computation of the transition rate matrix by separating the fast reactions from the slow reactions. If we consider that the interconversion rate between microstates (inside the circle, Fig. 1a) is much faster than the conversion rate between intermediates (the arrow connections, Fig. 1a),

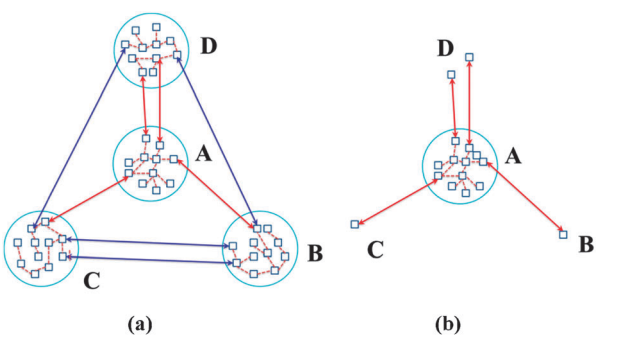


Fig. 1 (a) Illustration of a complex reaction network, containing a number of major intermediates, labeled as A, B, C and D (large circles), with each intermediate containing a number of possible configurations (small squares). (b) Illustration of SSW-RS starting from one configuration in A. The purpose of SSW-RS is to identify the important low energy configurations of A (including the GM) and the connectivity to other products in B, C and D.

one can derive the rate of transition between intermediates, say, A and B, as shown in eqn (2) with the equilibrium condition.

$$\frac{\mathrm{d}p_{\mathrm{A}}}{\mathrm{d}t} = k_{\mathrm{AB}}p_{\mathrm{B}}(t) - k_{\mathrm{BA}}p_{\mathrm{A}}(t) \tag{2}$$

Here $p_{\mathrm{A/B}}$ is the overall population of the intermediate A/B. The rate constant k_{AB} (similarly k_{BA}) can be further expressed as

$$k_{\mathrm{AB}} = \frac{1}{p_{\mathrm{B}}^{\mathrm{eq}}}\sum_{\mathrm{a}\in\mathrm{A}}\sum_{\mathrm{b}\in\mathrm{B}} k_{\mathrm{ab}}p_{\mathrm{b}}^{\mathrm{eq}} \tag{3}$$

where k_{ab} is the rate constant of transition between two microstates (belonging to A and B intermediates, respectively). Because it is an elementary reaction, k_{ab} can be calculated using transition state theory as in eqn (4):

$$k_{\mathrm{ab}} = \mathrm{A}\exp\left(\frac{-\Delta G}{kT}\right) \tag{4}$$

In eqn (3), the population ratio, $p_{\mathrm{b}}^{\mathrm{eq}}/p_{\mathrm{B}}^{\mathrm{eq}}$ (the superscript eq represents the equilibrium condition), can in principle be calculated from statistic mechanics using the partition function (Z) of all microstates, as described by eqn (5). Because the number of microstates is extremely large for systems even with a few atoms, one has to use approximations to compute the population ratio. For reactions where the vibrational component of the partition function is similar for the microstates concerned, the population ratio can be reduced to a function of the configuration component of the partition function, as shown in eqn (5):

$$\frac{p_{\mathrm{b}}^{\mathrm{eq}}}{p_{\mathrm{B}}^{\mathrm{eq}}} = \frac{Z_{\mathrm{b}}}{\sum_{\mathrm{i}\in\mathrm{B}} Z_{\mathrm{i}}} \approx \frac{\exp\left(-\dfrac{E_{\mathrm{b}}-E_0}{kT}\right)}{\sum_{\mathrm{i}\in\mathrm{B}}\exp\left(-\dfrac{E_{\mathrm{i}}-E_0}{kT}\right)} \tag{5}$$

where E_{b} and E_0 are the energy of the reactant at the microstate b and at the GM of the intermediate B.

From eqn (3)–(5), one can see that the rate constant of transition k_{AB} between two intermediates can be computed once the equilibrium distribution of the microstates and the free energy barrier linking the microstates at the two intermediate regions are known. Therefore, for reactivity prediction, it is essential to achieve the proper sampling for both the microstates of the reactant and the pathways connecting to other products. In the following, we will show that the SSW-RS method can be utilized to calculate these quantities by simultaneously sampling the configuration space of the reactant and the reaction space of pathways, as illustrated in Fig. 1b.

2.2 SSW and biased SSW methods

The recently developed SSW global optimization method[32,37] is the central part of the SSW-RS. The SSW method can explore a PES unbiasedly by gradually perturbing the structure from one minimum to another. Unlike the other global minimum search methods, the SSW method features the random direction generation, the soft mode following and the smooth surface walking. The details of the SSW algorithm can be found in our previous papers[32,37] and here we briefly describe the algorithm, focusing on the initial random direction in the SSW and the biased SSW PES exploration.

The SSW method is an unbiased PES exploration method, originating from the bias-potential driven constrained Broyden dimer (BP-CBD) method[38,39] for TS location developed by the group. Each step in SSW simulation contains a climbing procedure and a relaxation procedure to perturb the structure from one minimum to another, where the climbing procedure involves the consecutive Gaussian addition and repeated local relaxation, which is followed by the relaxation procedure to remove all the bias potentials and carry out unconstrained structure optimization. At the end of each SSW step, a structure selection module, *e.g.* using Metropolis Monte-Carlo scheme, is applied to accept/refuse the new minimum. The SSW method makes no attempt to locate the TS explicitly during PES exploration.

In the SSW method (also in BP-CBD), the barrier of the reaction can be surmounted by adding consecutively Gaussian bias potentials. The reaction direction where Gaussians are added is refined from a randomly generated direction using the biased constrained Broyden dimer method. The randomly generated initial direction N_{i}^0 (i labels the index of the SSW step) is thus critical for the efficiency of PES exploration, which is designed to include two components related to the global structure deformation ($N_{\mathrm{i}}^{\mathrm{g}}$) and the local bond formation ($N_{\mathrm{i}}^{\mathrm{l}}$),[32,37] as shown in eqn (6) and (7):

$$N_{\mathrm{i}}^0 = \frac{N_{\mathrm{i}}^{\mathrm{g}} + \lambda N_{\mathrm{i}}^{\mathrm{l}}}{\left\|N_{\mathrm{i}}^{\mathrm{g}} + \lambda N_{\mathrm{i}}^{\mathrm{l}}\right\|} \tag{6}$$

$$N_{\mathrm{i}}^{\mathrm{l}} = \begin{pmatrix} 0 \\ \vdots \\ q_{\mathrm{A}} \\ \vdots \\ q_{\mathrm{B}} \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ \vdots \\ q_{\mathrm{B}} \\ \vdots \\ q_{\mathrm{A}} \\ \vdots \\ 0 \end{pmatrix} \tag{7}$$

where λ controls the ratio between the global and the local components; and q_{A} and q_{B} are the coordinates of two randomly selected atoms A and B, respectively, which should not be in close contact.[32,37] Both $N_{\mathrm{i}}^{\mathrm{g}}$ and $N_{\mathrm{i}}^{\mathrm{l}}$ are important for PES exploration to prevent the long-time trapping in one particular minimum, since different reaction channels, including the global structure deformation and the local chemical bond making/breaking, can be involved from one minimum to another. In practice, λ is often set as a random number in between 0.1 and 15.

In this work, we also designed a biased initial reaction direction, $N_{\mathrm{i}}^0(\mathrm{b})$, by grouping a third component, $N_{\mathrm{i}}^{\mathrm{r}}$, with the unbiased N_{i}^0, as shown in eqn (8):

$$N_{\mathrm{i}}^0(\mathrm{b}) = \frac{N_{\mathrm{i}}^0 + \lambda' N_{\mathrm{i}}^{\mathrm{r}}}{\left\|N_{\mathrm{i}}^0 + \lambda' N_{\mathrm{i}}^{\mathrm{r}}\right\|} \tag{8}$$

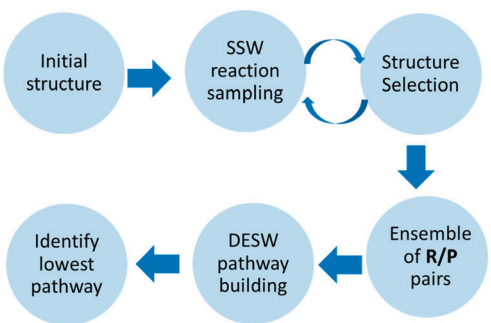
where the parameter λ' controls the ratio between the unbiased direction N_i^0 and the biased direction N_i^r. N_i^r can be inputted to contain the information on a desirable reaction pattern, *e.g.* a bond formation mode between two predefined atoms. We will show in Sections 3 and 4b that the biased search can be a useful tool to expedite the reaction sampling, especially when the reaction pattern is partly known.

2.3 SSW-RS method

a. Reaction sampling. The reaction sampling is carried out by using the SSW method, as illustrated in Scheme 1, which aims to generate an ensemble of reactant–product pairs for pathway building. The two structures of the reactant–product pair are geometrically close on PES (*e.g.* coordinate matched) thanks to the smooth structure perturbation of the SSW method, and they can be utilized later for TS location and pathway determination using the double-ended TS searching method, *e.g.* DESW method in this work.

The SSW reaction sampling starts from one random configuration of the reactant, which is the input structure and may well not be the GM of the reactant. During the simulation, the structures nearby this reactant will be visited, including the conformation isomers of the same reactant (the same phase region) and the likely products with different bonding patterns (the other phase regions). The region of the SSW sampling can be better visualized in Fig. 1b, where the starting structure is a conformation of intermediate A.

Different from the Metropolis MC scheme in the global structure search, a special structure selection module (to decide whether a new minimum is accepted or refused) needs to be designed in reaction sampling, which is described as follows. If a new minimum in the other phase regions is identified, we record/output the reactant structure and the product structure of the current SSW step. Then the program will return back to the reactant by rejecting the new minimum to continue the PES exploration; on the other hand, if the new minimum identified is still at the same reactant phase, the simulation will accept the new isomeric structure and continue the structure exploration. The whole procedure will be repeated until a certain number of reactant–product (R–P) pairs are collected, typically a few hundreds of R–P pairs when the first principles calculations are utilized for sampling (see Scheme 1).



Scheme 1 The flow chart of the SSW-RS method.

The structure selection module for reaction sampling therefore needs to determine whether a chemical reaction occurs (inside or outside the reactant phase). The probability for accepting a new minimum is described by eqn (9)

$$P_r = \begin{cases} 0, \text{reaction occurs} \\ 1, \text{otherwise} \end{cases} \quad (9)$$

where the occurrence of the reaction is characterized by the making–breaking of chemical bonds or the change of the local chirality of the atom, which are generally associated with reactions of high barrier. P_r can be computed by comparing the bond matrix and the chirality of the new minimum with the starting minimum. With eqn (9), the sampling of the reactant conformation can be automatically carried out along with the search of reaction pathways. When P_r equals 1, the new minimum must be an isometric configuration and its energy and structure are recorded for conformation analysis of the reactant region, as described by the partition function in eqn (5). On the other hand, the probability P for recording a R–P pair is further given in eqn (10) and (11),

$$P = (1 - P_r)\cdot P_{mc} \quad (10)$$

$$P_{mc} = \begin{cases} \exp\left[\dfrac{E(R^m) - E(R^{mt})}{kT}\right], & \text{when } E(R^{mt}) > E(R^m) \\ 1, \text{otherwise} \end{cases} \quad (11)$$

where P_{mc} is the Metropolis Monte Carlo probability for accepting the new minimum R^{mt} or reverting to the starting minimum R^m. The Metropolis Monte Carlo scheme in reaction sampling helps to screen the highly endothermic reactions, which will not be considered for pathway building.

b. Pathway building. Once enough R–P pairs are collected, we then need to find the reaction pathways connecting these R–P pairs, as illustrated in Fig. 2. Because the reactants and the products obtained from the SSW reaction sampling may be

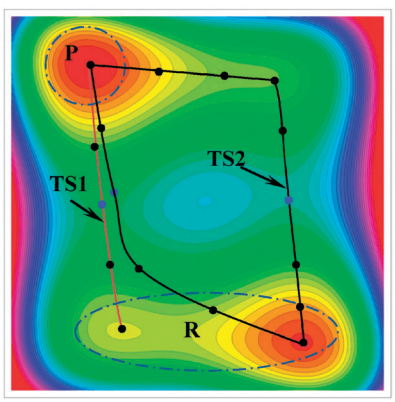


Fig. 2 Illustration of building DESW pseudo pathways to connect two minima regions (R and P) on a two-dimensional PES. The purpose of pathway building is to identify the lowest energy pathway (the red curve). Black dots: the points along the DESW pseudo pathway; blue dots: the maximum energy point along the DESW pseudo pathway.

separated by more than one minimum, the highest energy TS along the pathway will be first identified and the barrier calculated with respect to the GM energy of the reactant will be utilized as a quantitative measure to compare different pathways and to select the low energy pathways. The exact knowledge of the minimum energy pathway (MEP) at this stage is in fact not essential (MEP will be calculated only for a few lowest energy pathways that are important for kinetics). We utilize the DESW[40] method to identify the highest energy TS for a R–P pair and the procedure is described as follows.

The DESW method operates with two structural images starting from the reactant and the product, respectively, to walk stepwisely towards each other. The surface walking involves the repeated bias potential addition and local relaxation with the biased-CBD method to correct the walking direction, in a manner similar to the SSW and BP-CBD method. Since the TS location is the concern, the Gaussian width utilized in the DESW method is generally much smaller, *e.g.* 0.1–0.2 Å, compared to that (0.6 Å) in the SSW method for PES exploration. The DESW method can fast build a pseudo pathway to connect two minima since the added Gaussian functions effectively smooth the corrugated PES. The TS location is then performed from the highest energy image point (E_{max}) at the pseudo pathway using the single-ended CBD TS searching method.[38]

To speed up the pathway screening, the E_{max} of the pseudo pathway in DESW (see blue dots in Fig. 2) can be a good estimation for the barrier height of the real pathway since the Gaussian width utilized in DESW is generally small (*e.g.* 0.1 Å). This is because the pseudo pathway tends to converge to the real pathway at the TS region where the frequency (eigenvalue) of the reaction coordinate is negative.[40] Because of the large number of R–P pairs in the database, in practice we exploit the E_{max} of the pseudo pathway as the estimated barrier and only those pathways that have E_{max} lower than a preset energy value (*e.g.* 3.0 eV) will be selected to locate the TS explicitly. After the TSs are identified for these pathways, the lowest energy pathways can be identified by comparing the barrier of these pathways. The MEP will finally be retrieved using the Gonzalez–Schlegel IRC method[41] as utilized in our previous work,[40] by extrapolating the TS structure towards the reactant and the product.

It should be emphasized that the above procedure completes the PES exploration nearby one particular reactant, as illustrated in Fig. 1b. For multiple-step reaction systems with multiple intermediates (Fig. 1a), the same procedure has to be carried out by starting from other intermediates, and the whole reaction network can be established eventually. Methods such as Dijkstra's algorithm[42] can be utilized to analyze the network and identify the reaction mechanism. On the other hand, if the overall rate of the reaction system is the concern, the transition rate matrix can be constructed using eqn (3)–(5) based on the database from SSW-RS (including the reactant structures and the located pathways), and methods such as kinetic Monte Carlo can be utilized to simulate the reaction kinetics.

While the key elements of the SSW-RS for reactivity prediction, *e.g.* the pseudopathway building and the TS location, are similar to many of the elegant approaches proposed in recent years,[21,28,43] the SSW-RS has its own advantages using the SSW/DESW method as the reaction-sampling/pathway-building techniques. For example, the SSW-RS is a general purpose method for both association and dissociation reactions. The SSW-RS has the ability to explore different types of PESs, flat PES or corrugated PES, without using the history-dependent parameters to prevent local trapping. This guarantees the easy implementation and high transferability of the SSW-RS method for different applications, from molecules to clusters and to periodic crystals. An immediate benefit is that the SSW-RS can sample simultaneously the reactant conformations, a flat PES with very low barriers of conformation transformation, and the reaction pathways that have high barriers to unexpected products. These features of the SSW-RS method will be demonstrated in the following examples.

2.4 Calculation setups

In the following sections, all the SSW-RS simulations were carried out in the framework of first principles DFT as implemented in the SIESTA package,[44] which was utilized to evaluate the energy and force. The exchange–correlation functional utilized is the GGA-PBE[45] functional and the optimized numerical atomic orbital[46] with double-ξ polarization is utilized as the basis set. The force convergence criterion is set as 0.1 eV Å^{-1} for each degree of freedom in the SSW reaction sampling to collect all the R–P pairs. The more stringent criterion of 0.05 eV Å^{-1} for each degree of freedom is utilized to converge the structures (including the reactant, product and TS) at the low energy pathways.

For β-D-glucopyranose decomposition, we also utilized the hybrid M06-2X functional[47] together with the aug-cc-pVTZ basis set[48,49] as implemented in the Gaussian09 program[50] to refine the reaction kinetics for all lowest energy reaction channels, which were utilized to compare with the results in the literature.

In the SSW-RS, the Gaussian width, ds, is set as 0.5 Å and the number of Gaussians per SSW step, H, is 8, which are reduced compared to the previous work for global structure search (*e.g.* ds = 0.6 and H = 10). This is because the purpose of reaction sampling is to identify low energy pathways, the efficiency of which will decrease with too large ds and H (the geometry perturbation should not be too aggressive). The temperature utilized in the Metropolis Monte Carlo scheme is generally high, *e.g.* 3000 K, to remove only too high energy products (and pathways) that are kinetically unlikely.

All SSW-RS simulations were run in parallel to speed up the collection of R conformations together with R–P pairs. At least a few thousands of SSW steps are required in order to achieve a satisfied statistics for identifying enough low energy R conformations, *e.g.* a few hundreds of structures with energy ~1 eV above the lowest energy structure obtained (*i.e.* the regarded GM), and a significant number of low reaction pathways, *e.g.* ~10 times for each low energy pathway (see examples in Tables 1 and 2). One merit of SSW-RS is that each SSW run is independent and there is no requirement for communication between parallel runs. The DESW pathway building is a post processing technique, which can also be separated into many

small jobs, each containing a portion of R–P pairs. By parallelizing the SSW-RS, the overall computational time required for reactivity prediction can be reduced by more than one order of magnitude.

3. Comparison with the AFIR method for vinyl alcohol and formaldehyde recombination

In order to provide the first idea of the utility of the SSW-RS method for automated reaction prediction, we have applied the method to a model reaction, the recombination between vinyl alcohol and formaldehyde, which was previously studied by Maeda and Morokuma[51] using the AFIR method. It would be interesting to compare these two methods in terms of their efficiency, including the reactant minimum search and the reaction pathway finding. For this system, if as many as possible low energy pathways are concerned, the AFIR method takes 59 208 force (gradient) and 1465 Hessian evaluations to identify 22 distinct reaction pathways, 8 of them with the barrier below 2.5 eV,[51] as summarized in Table 1. If only the lowest energy pathway is concerned, AFIR can finish in 3490 force and 87 Hessian calculations by setting suitable simulation parameters. Since the Hessian calculation is not required in the SSW-RS method, we will focus on the times of the force evaluations *versus* the low energy reaction pathway identified.

Unbiased SSW-RS

First, we carried out the unbiased SSW-RS simulation in parallel for ten jobs starting from a random initial structure (0.5 eV above GM). It takes in total 57 591 force evaluations (379 SSW steps) to obtain 51 R–P pairs (see Section 2.4 for calculation details). The DESW method was then utilized to connect all the 51 R–P pairs, and 44 reaction pathways were obtained with the TS being successfully located (the TS search is given up if the DESW pseudopathway is too long in distance, *e.g.* >8 Å). Among them, we identified 12 distinct reaction pathways, 9 of them with the barrier below 2.5 eV. All these pathways were refined using Gaussian 09 at the level of RB3LYP/6-31G and confirmed using IRC analysis in order to compare with the AFIR results. We noted that SSW-RS already identifies the lowest energy pathway at the very beginning of the simulation: the average force evaluation times for finding this pathway is 2200 from all the parallel jobs and the minimum times is 442 (two SSW steps). We have summarized our main results in Table 1 and compared with the AFIR results. The detailed pictures of these pathways are shown in Fig. S1, ESI.†

From Table 1, it can be found that within roughly the same force evaluation times, the SSW-RS has obtained 9 out of 10 low reaction pathways (below 2.5 eV), while the AFIR method has identified 8 out of 10. Importantly, the SSW-RS method identified a new low energy pathway, *i.e.* the second lowest reaction pathway (barrier 1.47 eV), that was not identified in AFIR. In Fig. 3, we have highlighted the reaction snapshots for this reaction pathway, including IS, TS and final state (FS), which features collective two hydrogen transfer from vinyl alcohol to formaldehyde, leading to the final product of ethenone and methanol. Pathway 7 and pathway 8 have the same final product and a very similar barrier: the SSW-RS identifies pathway 8 and the AFIR method identifies pathway 7. Table 1 also shows that the SSW-RS method exhibits a clear preference for the low barrier pathways, *e.g.* the lowest barrier channels (identifies 20 times) dominating the pathways searched. This explains why the total pathways (12 pathways) identified from SSW-RS are less than those from the AFIR method (22 pathways) but the low energy pathways identified are even one more than that from the AFIR method.

For the vinyl alcohol and formaldehyde recombination, it is noted that the lowest energy pathway has a barrier much lower than that of the other pathways. If only this lowest energy pathway is concerned, SSW-RS still performs well: it takes on average 2200 force evaluation steps to identify the lowest energy channel, which is also comparable with the AFIR method (3490 force and 87 Hessian calculations; note that the exact comparison is not possible because the two methods have a different theoretical framework and are implemented in different computational packages).

In addition to the reaction pathway searching, the SSW-RS method can also simultaneously explore the reactant configurations

Table 1 Comparison of the identified ten low energy pathways from the SSW-RS method and AFIR (the names, aldol-high-PX, listed are the same as those reported[51]) for vinyl alcohol and formaldehyde recombination (these pathways are shown in Fig. S1, ESI)[a]

Pathway	AFIR	N_{path}(SSW-RS)[b]	IS energy/eV	TS energy/eV	FS energy/eV
1	Aldol-high-P1	20	−0.36	0.28	−1.23
2	—	2	−0.29	1.18	−0.28
3	Aldol-high-P5	1	−0.36	1.30	−0.62
4	Aldol-high-P4	4	−0.29	1.31	−0.54
5	Aldol-high-P2	4	−0.36	1.42	−0.74
6	Aldol-high-P13	1	−0.12	1.84	−0.85
7	Aldol-high-P6	0	−0.11	1.85	−0.49
8	—	1	−0.29	1.87	−0.48
9	Aldol-high-P8	2	−0.11	2.07	−1.00
10	Aldol-high-P12	3	−0.36	2.15	−1.28

[a] SSW-RS method: 57 591 force evaluations (379 SSW steps) to identify 12 distinct pathways, 9 lowest shown in the table. AFIR method: 59 208 force and 1465 Hessian evaluations to identify 22 distinct reaction pathways, 8 lowest shown in the table. All energies are calculated using Gaussian 09 at the level of RB3LYP/6-31G. [b] N_{path}: the times of the same pathway being identified by the SSW-RS method.

Table 2 SSW-RS results for epoxypropane hydrolysis in aqueous solution and β-D-glucopyranose decomposition[a]

Product	N_{path}[b]	N_{path}/N_{tot}[c] (%)	N_{path}/N_{SSW}[a] (%)	Minimum barrier/eV	Reaction energy/eV
Epoxypropane hydrolysis					
PG-C1	8	12	0.21	1.11	−0.50
PA	13	19	0.34	1.50	−0.52
PG-C2	27	40	0.71	1.58	−0.62
AA	5	7	0.13	1.86	0.16
Others	14	22	0.37	>3.00	—
β-D-Glucopyranose decomposition					
D-Glucose	16	10	0.47	1.15(1.52)[d]	0.84(0.73)[d]
Levoglucosan	10	6	0.30	1.90(2.23)	0.29(0.14)
β-D-Glucofuranose	11	7	0.32	2.04(2.50)	0.53(0.35)
β-D-Glucoseptanose	3	2	0.09	2.24(2.38)	0.36(0.10)
2,5-Anhydro-D-mannose	14	9	0.41	2.41(3.14)	0.51(0.43)
Others	102	66	3.02	>2.5	—

[a] Epoxypropane hydrolysis, SSW-RS method: 765 929 force evaluations (N_{SSW} = 3812 SSW steps) to identify 4 distinct pathways (<3.0 eV), 4 lowest shown in the table. β-D-Glucopyranose decomposition, SSW-RS method: 584 381 force evaluations (N_{SSW} = 3376 SSW steps) to identify 14 distinct pathways (<3.0 eV), 5 lowest shown in the table. [b] N_{path}: the times of the same pathway being identified by the SSW-RS method. [c] N_{tot} is the total number of pathways obtained from SSW-RS, *i.e.* 67 for epoxypropane hydrolysis and 156 for β-D-glucopyranose decomposition. [d] The data in parentheses are calculated using the M06-2X functional[47] with the aug-cc-pVTZ basis set (see the calculation details in Section 2.4).

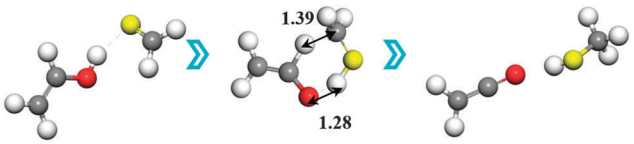


Fig. 3 The second-lowest energy pathway of vinyl alcohol and formaldehyde recombination identified from the SSW-RS method. From left to the right are the IS, TS and FS structures. C: gray; O: red; H: white, except that the O atom in formaldehyde is highlighted in yellow. Key distances (in Å) at the TS are labeled.

within the same global optimization scheme. By contrast, the reactant sampling is carried out by random structure generation in the AFIR method, which is expected to be demanding in order to sample enough reactant configuration space for complex multiple molecular systems (*e.g.* see examples in the following section). By investigating the configuration space of the reactant from SSW-RS simulation, we found that the GM of the reactant has an energy of −0.36 eV (with respect to the isolated molecules) identified within 10 minima visited, corresponding to the initial state (IS) of the aldol-high-P1 pathway also identified in the AFIR method. The SSW-RS identified the second lowest initial state structure (−0.29 eV) within 7 minima visited.

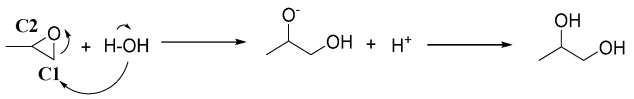
Biased SSW-RS

It might be mentioned that the AFIR method involves artificial force to "push" two molecules together. This in fact corresponds to a biased reaction direction. Following a similar idea, we therefore also performed the biased SSW-RS search by adding a biased direction N_i^r (eqn (7) and (8)), which is generated by randomly selecting two atoms, each from one molecule, to provide an initial bond formation mode (eqn (7)). In the biased SSW-RS simulation, we found that the reaction probability increases largely: within 146 288 force evaluations, 226 R–P pairs were collected and finally 165 reaction pathways were obtained. The overall efficiency increases by 1.7 times compared to the unbiased SSW-RS method. The biased SSW-RS is thus more efficient than the AFIR method.

To recap, we show that the SSW-RS can be utilized for automated reaction pathway searching and the efficiency of locating the low energy pathways (*e.g.* below 2.5 eV) is comparable with that of the existing AFIR method. The SSW-RS has some additional advantages. (i) It has a clear preference for the low energy pathways, which are kinetically more relevant. In the case of vinyl alcohol and formaldehyde recombination, an important low energy pathway, *i.e.* the second lowest energy pathway, is identified by the SSW-RS method but is missed in the AFIR search. (ii) It can simultaneously sample the reactant configurations using the global optimization method and thus can be straightforwardly applied for reaction rate prediction using eqn (5). (iii) It does not require the calculation of Hessian and the incorporation of the biased reaction direction can further enhance the pathway sampling. This allows the application of SSW-RS to much large systems where the analytical Hessian is often difficult to obtain or compute.

4. Epoxypropane hydrolysis

To further examine the performance of the SSW-RS method, we have applied the method in different types of reactions, namely the association reaction (epoxypropane hydrolysis in aqueous solution) and the dissociation reaction (β-D-glucopyranose decomposition). The epoxypropane hydrolysis occurs in solution and involves a complex intermolecular hydrogen bonding network in the reaction, being a good representative of associative organic reactions. The epoxypropane hydrolysis is also a relatively simple reaction of high selectivity with propylene glycol (PG) as the major product. The β-D-glucopyranose decomposition is the initiating step in glucose pyrolysis occurring at high temperatures (*e.g.* 800 K), featuring high barrier reactions and multiple reaction channels. The detailed reaction mechanism remains highly debated[52] and



Scheme 2 The generally regarded mechanism for epoxypropane hydrolysis without acid/base catalysts.

we aim to utilize the unbiased SSW-RS method to provide an insight into the pyrolysis kinetics.

a. Unbiased sampling

While both acid and base are good catalysts for epoxypropane hydrolysis, in this work we focused on the reaction under the neutral pH condition for simplicity, which is a good starting case for testing the ability of SSW-RS to sample the high barrier reactions (without catalysts) together with the low barrier isomeric transformation in solution. The generally regarded mechanism is depicted in Scheme 2, where a water molecule attacks the carbon atom (C1) of epoxypropane and leads to the ring opening. In aqueous solution, it is expected that the hydrogen-bonding network helps to stabilize the polarized reaction intermediates. Therefore, a main challenge here is to identify the flexible hydrogen-bonding network in the low energy reaction pathways.

In the SSW-RS simulation, we contained three explicit water molecules and one epoxypropane in a cubic box ($10 \times 10 \times 10$ Å) with the periodic boundary condition. These molecules were immersed in an implicit periodic continuum solvation background that describes the long-range solvation polarization effects.

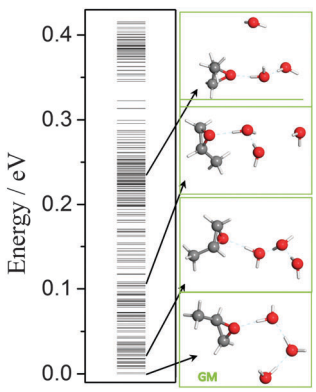


Fig. 4 The configuration energy spectrum of the epoxypropane/water structures in solution (continuum solvation model) obtained from SSW-RS. Grey: C atom, red: O atom.

The periodic continuum solvation model based on the modified Poisson–Boltzmann equation, namely CM-MPB method, was utilized for implicit solvation, which has been applied recently by us for describing a number of reactions at the solid–liquid interface.[53–55]

Our unbiased SSW-RS simulations were carried out in parallel for eight jobs. In total, 3812 SSW steps were performed (765 929 force evaluations), and 102 R–P pairs were collected to yield an ensemble of R–P pairs. We noticed that SSW-RS spends most of the simulation time (3710 out of 3812 SSW steps) in exploring the conformation space of epoxypropane in solution (reactant conformation) and only less than 3% of the time in exploring the reaction space of pathways.

In Fig. 4, we show the configuration energy spectrum of the reactant structures collected from the SSW trajectories at the low energy end (within 0.5 eV above the GM conformation). It is obvious that the energy spacings of the reactant conformations are quite small, indicating a flat PES of the reactant conformations. This is obvious because the major difference between the conformations is the hydrogen-bonding network and the presence of the implicit continuum solvation further flattens the PES. The representative structures from the spectrum, including the GM, are also shown in Fig. 4.

The GM conformation is identified from the SSW-RS trajectories within 50 SSW steps, which is close to a four-member hydrogen-bonding ring (note that the exact GM structure may well be different under a different calculation setup due to the very flat nature of the PES). The other minima differ from the GM structure in the hydrogen-bonding network. For example, a linear H-bonding network appears only $\sim$0.03 eV above the GM. The H-bonding network starts to break down, leading to separate clusters, at $\sim$0.1 eV above the GM.

The DESW method was utilized to connect all the 102 R–P pairs, and 67 reaction pathways were obtained with the TS being successfully located (the TS search is given up if the DESW pseudopathway is too long in distance, *e.g.* >8 Å). Among them, 53 have the reaction barrier lower than 3.0 eV. These pathways can be grouped into four different reaction channels, named from Epo-P1 to Epo-P4, with only three different final products, being PG, propionaldehyde (PA) and allyl alcohol (AA). Epo-P1 and Epo-P3 produce the same final product, PG, but the reaction mechanism differs in the carbon atom of the epoxypropane that is attacked by water: it is the C1 in Epo-P1 but the C2 in Epo-P2 (see Scheme 2). The located structures for the IS, TS and FS of the lowest energy pathway are shown in Fig. 5, and the numerical results on the reaction

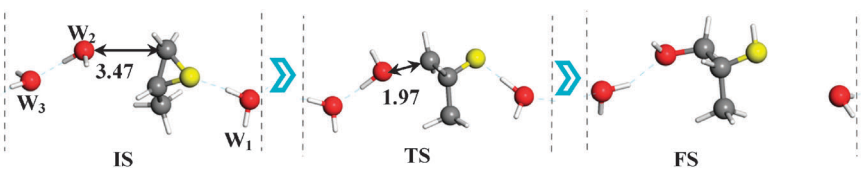


Fig. 5 The lowest energy pathway for epoxypropane hydrolysis obtained from SSW-RS. Owing to the periodic boundary condition (the dotted line), the reaction features a hydrogen-bonding chain, where the H_2O (W_1) loses its H to epoxy O but accepts another H from the neighboring H_2O (the periodic image of W_3). Grey: C atom, red: O atom; and yellow: the highlighted epoxy O. Key distances (in Å) are labeled.

channels are summarized in Table 2. The energy profile of the lowest energy pathways to form the three different products, PG, PA and AA, is summarized in Fig. 6. The main results are elaborated in the following.

Epo-P1. This reaction channel has the lowest energy pathway with the calculated barrier being 1.11 eV. In the pathway, a water molecule attacks the C1 atom of epoxypropane, leading to PG (PG-C1). Interestingly, at the located TS (see Fig. 4) the molecules align up as a chain, yielding a proton-passing path by exploiting the periodic boundary condition, where the H_2O (W_1 in Fig. 5) loses its H to epoxy O but accepts another H from the neighboring H_2O (the periodic image of W_3).

Epo-P2. This reaction channel involves the 1,2-H transfer in epoxypropane that leads to PA. The calculated barrier is 1.50 eV. The nearby water molecules help to stabilize the TS by forming hydrogen-bonding with the epoxy O (see Fig. S2 of ESI† for the reaction snapshots).

Epo-P3. This reaction channel is initiated by a water molecule attacking the C2 atom of epoxypropane and finally also yields PG (PG-C2) (see Fig. S3 of ESI† for the reaction snapshots). The calculated barrier is 1.58 eV, much higher than that in Epo-P1, obviously owing to the larger steric repulsion at the C2 position.

Epo-P4. This reaction channel produces AA, involving the hydrogen transfer from the methyl group of epoxypropane to the epoxy O mediated by the surrounding water molecules. The calculated barrier is 1.86 eV, the highest among the four pathways. In the reaction, the hydrogen atom of a water molecule first passes to the epoxy O of epoxypropane, which facilitates the ring opening of epoxypropane. Simultaneously, a hydrogen atom linked to the methyl group shifts to the nearby water molecule and a proton-passing chain is formed among the water molecules (see Fig. S4 of ESI† for the reaction snapshots).

Based on the energetics of the pathways shown in Fig. 6, we can conclude that PG is the only likely product under normal conditions since the barrier of Epo-P1 is considerably lower than that of the other reaction pathways. The identified reaction mechanism from the unbiased SSW-RS is also consistent with the generally regarded two-step mechanism as depicted in Scheme 2, except that it is a one step reaction *via* a collective proton transfer process as revealed from theory. Apart from the energetics, the computational simulation provides a clear atomic level picture of the TS, which features the collective proton transfer network involving more than one water molecule nearby epoxypropane. It is important to notice that the geometry of the TS (a chain structure) in the lowest energy pathway is distinct from that of the GM at the IS (a ring structure). This example shows that the predictive power of SSW-RS is particularly useful in complex reactions, where the structure of low energy TSs is hard to be pre-guessed according to the IS conformations.

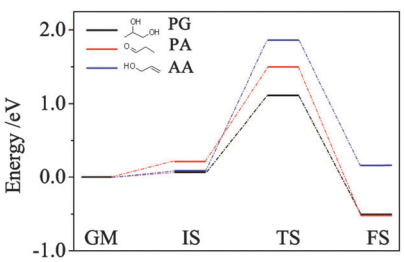


Fig. 6 The energy profile of the lowest energy pathways to form the three major products in epoxypropane hydrolysis obtained from SSW-RS. The energy is with respect to the GM of epoxypropane/water in solution (continuum solvation).

b. Biased sampling

It is possible to expedite the reaction sampling by adding predefined reaction patterns in the SSW-RS, as described in eqn (8). While the guess of the exact TS or the reaction coordinate can be challenging in complex reaction systems, we notice that the guess of an approximate reaction direction is in fact likely in many reactions, considering that the functional group of the reactant is often known, such as the C1, C2 and the epoxy O atoms in epoxypropane. By defining the reaction patterns according to the functional group of the reactant, one may reduce significantly the degrees of freedom in the reaction space of pathways and thus expedite the reaction search.

To demonstrate this idea, we also examined the performance of the biased SSW-RS for epoxypropane hydrolysis and the performance was compared with that of the unbiased SSW-RS. For the biased reaction direction, we designed two types of bond formation modes centering on a pair of atoms (eqn (7)) – type-I, the C1 atom with a neighboring O atom of water, and type-II, the C2 atom with a neighboring O atom of water molecules (also see Fig. 7, inset) – and selected one of them stochastically as N_i^r (eqn (7)) in each SSW step. The λ' parameter in eqn (8) is set as a random number ranging from 0.1 to 10.

In total, 450 SSW steps were carried out in the biased SSW-RS starting from the same reactant conformation as in the unbiased SSW-RS. 75 R–P pairs were collected from the trajectories and 55 pathways were consequently established with the TS being located using the DESW method. Not surprisingly, we

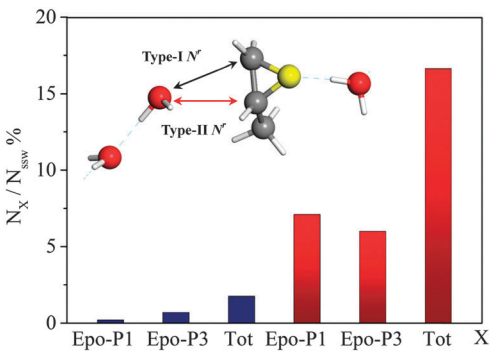


Fig. 7 Comparison of the efficiency to identify the reaction pathways in the biased SSW-RS (red) and the unbiased SSW-RS (blue) in epoxypropane hydrolysis. N_{tot}, N_{path} and N_{SSW} are as those defined in Table 2. The inset illustrates two types of biased reaction directions (N_i^r in eqn (8)). On the Y axis, N_x represents N_{tot} or N_{path} referring to Epo-P1 or Epo-P2; N_{tot}, N_{path} and N_{SSW} are as those defined in Table 2.

found that the reaction channels represented by Epo-P1 and Epo-P3 now dominate the identified pathways, and no new low energy reaction channels other than the four reaction channels presented above appear. The overall efficiency using the biased and the unbiased SSW-RS is compared in Fig. 7.

Fig. 7 shows that there is a significant increase in the probability to sample the pathways compared to that to sample the reactant conformations: the probability to sample reaction pathways is now 17% in the biased SSW-RS, 10 times higher than that in the unbiased sampling (0.17%). Consistently, the probability to identify the reaction channels belonging to Epo-P1 increases by more than 35 times, from 0.2% to 7%, and for those belonging to Epo-P3 it increases by about 10 times, from 0.7% to 6%. The probability to identify the reaction channels belonging to Epo-P2 and Epo-P4 remains quite constant, being about 0.4% in the biased sampling. It is also noticed that the GM conformation of the reactant is identified after 100 SSW steps, which takes more efforts than that in the unbiased sampling (50 SSW steps).

It is obvious that the biased pathway sampling has a higher efficiency to identify pathways, in particular those associated with the predefined reaction directions. This is however at the expense of the efficiency to sample the reactant configurations. Apparently, the ratio of the pathway sampling to the reactant configuration sampling could be tuned by the value of λ': when $\lambda' = 0$, the biased sampling reverts back to the unbiased sampling. The biased SSW-RS provides a useful mechanism to fast resolve some important reaction patterns when the information on the reaction, *e.g.* the reaction sites, is (partly) known. Due to the parallel nature of the SSW-RS sampling, it is beneficial to combine the unbiased and the biased sampling for resolving a complex reaction network.

5. β-D-Glucopyranose decomposition

Not limited to the association reaction, the SSW-RS method can also be applied to the dissociation reaction without extra coding, as demonstrated here using the decomposition reaction of β-D-glucopyranose. β-D-Glucopyranose is a main building block of cellulose, a widely available biomass chemical.[56,57] Pyrolysis of cellulose can produce many different chemicals,[58] including useful biofuels[59] and high value fine chemicals.[60] The selectivity is the major concern. The mechanism and kinetics of β-D-glucopyranose pyrolysis have been the hot topic in biomass research.[61–63] However, due to the presence of many possible reaction channels, the reaction network of β-D-glucopyranose pyrolysis is not established yet. Here with the unbiased SSW-RS, we explore the initial reaction channels of β-D-glucopyranose decomposition, which is known to be the key event in cellulose pyrolysis.

The unbiased SSW-PS simulation starts from a β-D-glucopyranose molecule in the chair conformation, which is contained in a large vacant cubic cell (10 Å length). In total, 3376 SSW steps were carried out (584 381 force evaluations) and 207 R–P pairs were collected from eight parallel runs of

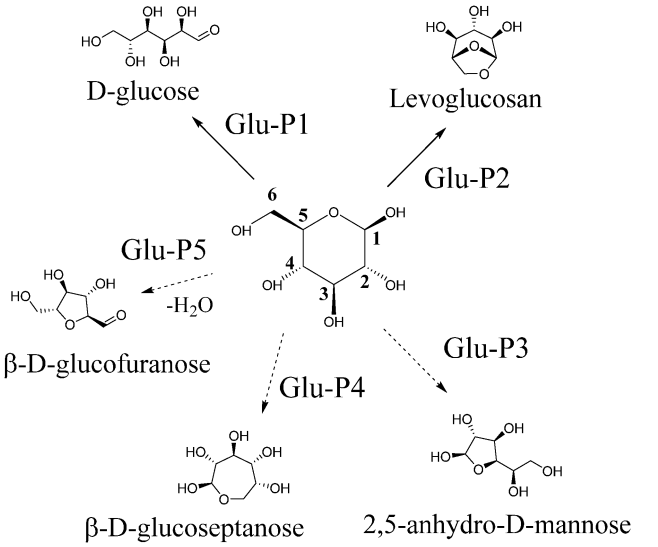


Fig. 8 Five lowest energy pathways of β-D-glucopyranose decomposition leading to five different products from SSW-RS.

SSW-RS simulation. In total, 354 conformations of β-D-glucopyranose were collected, from which the GM structure was identified (these are shown in Fig. S5, ESI†). Among the R–P pairs, 156 reaction pathways were established with the TS being explicitly located, 68 of them with the barrier below 3.0 eV. By inspecting these low energy barrier reaction pathways, we identified 15 possible reaction products (listed in Table S1, ESI†) and 5 of them can be produced with relatively low barrier (<2.5 eV from PBE), *i.e.* D-glucose, levoglucosan, β-D-glucofuranose, β-D-glucoseptanose, and 2,5-anhydro-D-mannose, as summarized in Fig. 8, and the reaction channels associated with these five products are named from **Glu-P1** to **Glu-P5**. The reaction snapshots, including the located IS, TS and FS, for the five lowest energy pathways are shown in Fig. 9. The probability of SSW-RS to identify the pathways and the energetic data are also listed in Table 2. Our main results on the pathways are detailed in the following.

Glu-P1 is the ring opening of β-D-glucopyranose to produce D-glucose. This reaction has the lowest barrier (1.15 eV from PBE, 1.52 eV from M06-2X, see Table 2) among all the pathways, but is endothermic by $\sim$0.7 eV, indicating that the reaction favors the close-ring form thermodynamically. At the TS, the bridging O (O_{br}) accepts a proton from the neighboring O6 (the O atom connected with C6, see Fig. 8), which simultaneously accepts a proton from the O1 (the O atom connected with C1). The OH groups on C3 and C6 help to stabilize the nascent terminal O in acetaldehyde *via* hydrogen-bonding.

Glu-P2 is the dehydration of β-D-glucopyranose to levoglucosan. This pathway has the second lowest barrier among all the pathways, being 1.90 eV from PBE and 2.23 eV from M06-2X. The relatively low barrier of the reaction is consistent with the fact that levoglucosan is a common product found in the pyrolysis of cellulose.[62,64] The TS identified by SSW-RS features a hydrogen-bonding network among the O(H) on C1, C3 and C6, which is similar to that reported in the previous theoretical study.[65]

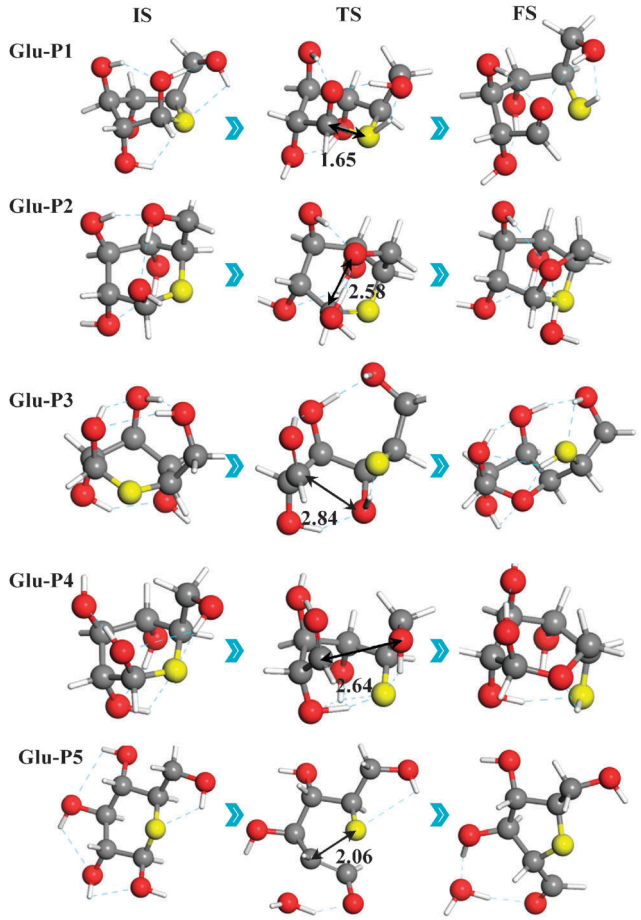


Fig. 9 Reaction snapshots of the five lowest reaction pathways in the β-D-glucopyranose decomposition obtained from SSW-RS. Key distances (in Å) are labeled.

Compared to the two lowest energy pathways, **Glu-P3**, **Glu-P4** and **Glu-P5** have much higher barriers (>2.0 eV from PBE and >2.5 eV from M06-2X), where the **Glu-P3** and **Glu-P5** produce the 5-member ring products β-D-glucofuranose and 2,5-anhydro-D-mannose and **Glu-P4** produces the 7-member ring β-D-glucoseptanose. The TSs in these pathways are generally complex in geometry, involving both the 6-member ring opening (of β-D-glucopyranose) and the new 5- or 7-member ring formation, where the C–O bond between C1 and O_{br} is lengthened significantly in order to free the C1 to form new bonding with the other atoms. Taking the **Glu-P3** as an example (see Fig. 9), the located TS features a hydrogen-bonding network of O1H···O3H···O6H and a long C1–O_{br} distance of 2.62 Å.

It is of interest to compare our identified pathways with those studied previously. The **Glu-P1** and **Glu-P2** pathways are theoretically studied recently.[66] According to the structures reported, we recalculated these pathways (using the same calculation setup here) and found that the barriers are 2.22 eV for **Glu-P1** and 2.37 eV for **Glu-P2**. The first one is much higher than that (1.69 eV) obtained from the SSW-RS method, and the second is the same as what we calculated (2.37 eV). We noticed that the difference in **Glu-P1** is due to the more favorable TS structure (Fig. 9) identified in this work. Similarly, the previous reported barrier for the formation of 2,5-anhydro-D-mannose (in **Glu-P5**) from β-D-glucofuranose,[65] 4.05 eV using M06-2X, is also 0.90 eV higher than the barrier found in this work. This can be attributed to the absence of two important hydrogen-bonding networks at O6H···O_{br} and O1H···O2 at the TS in the previous study.[65] From these comparisons, we conclude that the SSW-RS is capable of identifying more stable TS conformation in these complex organic reactions.

By computing the preexponential factors and free energy barriers, we have estimated the rate of the above five pathways according to microkinetics with the aim to provide an insight into the mechanism of β-D-glucopyranose pyrolysis. At the typical pyrolysis temperature, 800 K, the first two pathways, **Glu-P1** and **Glu-P2**, are the most relevant initial reaction channels, where the computed rate constants are 58.4 and 9.92×10^{-2} s^{-1}. The other pathways are at least three orders of magnitude slower compared to the **Glu-P2**. Therefore, we conclude that the 5- and 7-member ring products, 2,5-anhydro-D-mannose and β-D-glucoseptanose, should be the secondary products from either the D-glucose or levoglucosan (1,6-anhydro-β-D-glucose). This conclusion agrees with the suggestion from previous experimental studies on cellulose and glucose pyrolysis. For example, using a mass spectrometer to detect the pyrolysis products, Lin *et al.*[62] and Vinu *et al.*[52] suggested that the 5-member ring molecule 5-(hydroxymethyl) furfural (5-HMF) cannot be produced directly from β-D-glucofuranose (β-D-glucofuranose is the precursor of 5-HMF) or β-D-glucoseptanose, but should be produced from the chain D-glucose after three consecutive dehydration reactions.

Finally, we may add a few remarks on the potential applications of the SSW-RS method in the future. The above examples indicate that the combination of SSW-RS and the first principles DFT calculations, SSW-RS/DFT, is a powerful tool to resolve the mechanism and the kinetics of complex reaction systems, which can now be routinely carried out for systems below 30 atoms, including organic molecules and solid crystals.[67] The efficiency of the reaction sampling for the two systems, as measured by N_{path}/N_{ssw} in Table 2, is generally in the range of 0.1% to 1% for a single product, the value being also quite constant for all low barrier reaction channels despite that fact that the difference of the barrier between them can be more than a few tens of eV. Obviously, the SSW reaction sampling does not follow the physical law of kinetics as obeyed in the traditional MD simulation, which enables the fast sampling of the conformation space and the reaction space. However, because of the poor scaling of first principles calculations and the exponential increase in the number of microstates for large systems, we noticed that the application of the unbiased SSW-RS/DFT to large systems (*e.g.* beyond 50 atoms) is still of great challenge, where more than a few thousands of SSW steps can be prohibitively expensive in computation (each SSW step contains ~300 energy/force evaluation steps[32,67]).

In these circumstances, other measures must be taken to expedite the reaction search, for example, by developing and utilizing reactive force fields for an initial SSW sampling, which should at least provide the information on the likely

reaction patterns. Based on this *a priori* knowledge, it is possible to increase the efficiency of reaction sampling *via* the biased SSW-RS as demonstrated for epoxypropane hydrolysis. With the help of the massive parallel computing and the biased sampling, our ongoing work shows that the SSW-RS/DFT can be applied to understand the heterogeneous catalytic reaction at the solid–liquid interface (systems with more than 100 atoms),[54,68,69] where the effects of the metal surface, the water solution and the coverage of coadsorbates on the reaction can be considered in one unified framework.

6. Conclusion

This work develops an unbiased general-purpose method, the SSW-RS method, for reaction sampling and reactivity prediction. The new method is based on two recently developed techniques, the SSW method for PES exploration and the DESW method for pathway building and TS location. We show that, as derived from the Master equation in kinetics theory, the reactivity prediction requires the proper sampling on both the conformation space of reactants and the reaction space of pathways. The SSW method is such a suitable sampling tool for reactivity prediction. The DESW method, being in the same theoretical framework as the SSW method, is utilized for the fast assessment of a large number of pathways in a complex reaction. By combining the SSW-RS method with first principles DFT calculations, we have first compared the performance with the existing methods. We showed that (i) the unbiased SSW-RS is particularly efficient for identifying the low energy pathways. It revealed a new (the second lowest energy) pathway for vinyl alcohol and formaldehyde recombination, which is not found previously; (ii) the SSW-RS is able to sample reactant configurations during the pathway sampling within the global optimization scheme; (iii) the biased SSW-RS simulation can be incorporated readily in the framework and it increases the performance dramatically.

Two new example reactions with complex PES, the epoxypropane hydrolysis in aqueous solution and the β-D-glucopyranose decomposition, are further examined using SSW-RS. For the epoxypropane hydrolysis in aqueous solution, the SSW-RS simulation reveals four different reaction channels with the barrier below 2.0 eV, and the lowest energy pathway is initiated *via* a water molecule attacking the C1 atom of epoxypropane, which agrees with the generally regarded mechanism. Importantly, the SSW-RS is able to resolve the complex hydrogen-bonding network for both the IS and the TS. In the lowest energy pathway, three water molecules nearby epoxypropane are found to take part in the reaction to produce PG. We also show that the biased SSW-RS can further speed up the reaction sampling towards the biased reaction direction by more than 10 times.

For β-D-glucopyranose decomposition, the SSW-RS simulation reveals 15 possible products, indicating the presence of a complex reaction network in this system. Five of them have relatively low barriers (below 2.5 eV). By analyzing the five lowest energy pathways leading to these five products, we found that only β-D-glucose and levoglucosan are kinetically preferred as the direct products from β-D-glucopyranose. The other products, such as the 5- or 7-member ring molecules, β-D-glucofuranose and β-D-glucoseptanose, should be secondary products. In one word, the one-step ring rearrangement from the 6-member ring to 5- or 7-member rings is kinetically not preferred, even though the SSW-RS reveals the presence of such reaction channels involving complex collective movement of atoms. Our results provide new insights towards the understanding of the selectivity of cellulose pyrolysis.

Acknowledgements

This work is supported by the National Science Foundation of China (21173051, 21361130019), the 973 program (2011CB808500, 2013CB834603), the Science and Technology Commission of Shanghai Municipality (08DZ2270500), and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institute of Higher Learning.

References

1 H. B. Schlegel, *J. Comput. Chem.*, 2003, **24**, 1514–1527.
2 Y. Kanai, A. Tilocca, A. Selloni and R. Car, *J. Chem. Phys.*, 2004, **121**, 3359–3367.
3 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
4 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.
5 B. Ensing, M. De Vivo, Z. Liu, P. Moore and M. L. Klein, *Acc. Chem. Res.*, 2005, **39**, 73–81.
6 R. Zhou, in *Protein Folding Protocols*, ed. Y. Bai and R. Nussinov, Humana Press, 2006, vol. 350, pp. 205–223.
7 Y. M. Rhee and V. S. Pande, *Biophys. J.*, 2003, **84**, 775–786.
8 Y. Wang and P. B. Balbuena, *J. Phys. Chem. B*, 2005, **109**, 14896–14907.
9 J. Hénin and C. Chipot, *J. Chem. Phys.*, 2004, **121**, 2904–2914.
10 D. Hamelberg, J. Mongan and J. A. McCammon, *J. Chem. Phys.*, 2004, **120**, 11919–11929.
11 W. M. Young and E. W. Elcock, *Proc. Phys. Soc., London*, 1966, **89**, 735.
12 L. Xu and G. Henkelman, *J. Chem. Phys.*, 2008, **129**, 114104.
13 A. B. Bortz, M. H. Kalos and J. L. Lebowitz, *J. Comput. Phys.*, 1975, **17**, 10–18.
14 D. T. Gillespie, *J. Comput. Phys.*, 1976, **22**, 403–434.
15 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Adv. Chem. Phys.*, John Wiley & Sons, Inc., 2003, pp. 1–78.
16 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
17 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964–1977.
18 D. J. Wales, *Mol. Phys.*, 2002, **100**, 3285–3305.
19 D. J. Wales, *Int. Rev. Phys. Chem.*, 2006, **25**, 237–282.
20 S. Goedecker, *J. Chem. Phys.*, 2004, **120**, 9911–9917.
21 B. Schaefer, S. Mohr, M. Amsler and S. Goedecker, *J. Chem. Phys.*, 2014, **140**, 214102.

22 P. G. Bolhuis, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12129–12134.
23 T. F. Middleton, J. Hernández-Rojas, P. N. Mortenson and D. J. Wales, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2001, **64**, 184201.
24 T. F. Middleton and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 8134–8143.
25 H. B. Schlegel, *Theor. Chim. Acta*, 1992, **83**, 15–20.
26 W. Quapp, M. Hirsch, O. Imig and D. Heidrich, *J. Comput. Chem.*, 1998, **19**, 1087–1100.
27 Y. Luo, S. Maeda and K. Ohno, *J. Comput. Chem.*, 2009, **30**, 952–961.
28 S. Maeda, K. Ohno and K. Morokuma, *J. Phys. Chem. A*, 2009, **113**, 1704–1710.
29 S. Maeda, K. Ohno and K. Morokuma, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3683–3701.
30 Vikas and G. Kaur, *J. Chem. Phys.*, 2013, **139**, 224311.
31 M. Hatanaka, S. Maeda and K. Morokuma, *J. Chem. Theory Comput.*, 2013, **9**, 2882–2886.
32 C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2013, **9**, 1838–1845.
33 Q. Chen, G. Wei, W.-J. Tian, H. Bai, Z. Liu, H.-J. Zhai and S. Li, *Phys. Chem. Chem. Phys.*, 2014, **16**, 18282.
34 H.-J. Zhai, Y.-F. Zhao, W.-L. Li, Q. Chen, H. Bai, H.-S. Hu, Z. A. Piazza, W.-J. Tian, H.-G. Lu, Y.-B. Wu, Y.-W. Mu, G.-F. Wei, Z.-P. Liu, J. Li, S.-D. Li and L.-S. Wang, *Nat. Chem.*, 2014, **6**, 727.
35 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2001, **115**, 9657–9666.
36 T. P. Schulze, *J. Comput. Phys.*, 2008, **227**, 2455–2462.
37 X.-J. Zhang, C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2013, **9**, 3252–3260.
38 C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2010, **6**, 1136–1144.
39 C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2012, **8**, 2215–2222.
40 X.-J. Zhang, C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2013, **9**, 5745–5753.
41 C. Gonzalez and H. B. Schlegel, *J. Phys. Chem.*, 1990, **94**, 5523–5527.
42 E. W. Dijkstra, *Numer. Math.*, 1959, **1**, 269–271.
43 S. Maeda, T. Taketsugu and K. Morokuma, *J. Comput. Chem.*, 2014, **35**, 166–173.
44 M. S. José, A. Emilio, D. G. Julian, G. Alberto, J. Javier, O. Pablo and S.-P. Daniel, *J. Phys.: Condens. Matter*, 2002, **14**, 2745.
45 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
46 J. Junquera, ó. Paz, D. Sánchez-Portal and E. Artacho, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2001, **64**, 235111.
47 Y. Zhao and D. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
48 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
49 K. B. Wiberg, C. M. Hadad, T. J. LePage, C. M. Breneman and M. J. Frisch, *J. Phys. Chem.*, 1992, **96**, 671–679.
50 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford, CT, USA, 2009.
51 S. Maeda and K. Morokuma, *J. Chem. Theory Comput.*, 2011, **7**, 2335–2345.
52 R. Vinu and L. J. Broadbelt, *Energy Environ. Sci.*, 2012, **5**, 9808–9826.
53 H.-F. Wang and Z.-P. Liu, *J. Phys. Chem. C*, 2009, **113**, 17502–17508.
54 Y.-F. Li, Z.-P. Liu, L. Liu and W. Gao, *J. Am. Chem. Soc.*, 2010, **132**, 13008–13015.
55 Y.-H. Fang, G.-F. Wei and Z.-P. Liu, *Catal. Today*, 2013, **202**, 98–104.
56 L. R. Lynd, J. H. Cushman, R. J. Nichols and C. E. Wyman, *Science*, 1991, **251**, 1318–1323.
57 C. Şerbănescu, *Chem. Pap.*, 2014, **68**, 847–860.
58 H. Yang, R. Yan, H. Chen, D. H. Lee and C. Zheng, *Fuel*, 2007, **86**, 1781–1788.
59 P. L. Dhepe and A. Fukuoka, *ChemSusChem*, 2008, **1**, 969–975.
60 Z.-G. Guo, S.-R. Wang, Y.-Y. Zhu, Z.-Y. Luo and K.-F. Cen, *J. Fuel Chem. Technol.*, 2009, **37**, 49–52.
61 M. J. Antal Jr. and G. Varhegyi, *Ind. Eng. Chem. Res.*, 1995, **34**, 703–717.
62 Y.-C. Lin, J. Cho, G. A. Tompsett, P. R. Westmoreland and G. W. Huber, *J. Phys. Chem. C*, 2009, **113**, 20097–20107.
63 Z. Luo, S. Wang, Y. Liao and K. Cen, *Ind. Eng. Chem. Res.*, 2004, **43**, 5605–5610.
64 P. R. Patwardhan, J. A. Satrio, R. C. Brown and B. H. Shanks, *J. Anal. Appl. Pyrolysis*, 2009, **86**, 323–330.
65 V. Seshadri and P. R. Westmoreland, *J. Phys. Chem. A*, 2012, **116**, 11997–12013.
66 H. B. Mayes, M. W. Nolte, G. T. Beckham, B. H. Shanks and L. J. Broadbelt, *ACS Sustainable Chem. Eng.*, 2014, **2**, 1461–1473.
67 C. Shang, X.-J. Zhang and Z.-P. Liu, *Phys. Chem. Chem. Phys.*, 2014, **16**, 17845–17856.
68 Y.-H. Fang and Z.-P. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 18214–18222.
69 W.-N. Zhao and Z.-P. Liu, *Chem. Sci.*, 2014, **5**, 2256–2264.