JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Constrained Broyden Dimer Method with Bias Potential for Exploring Potential Energy Surface of Multistep Reaction Process

Cheng Shang and Zhi-Pan Liu*

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

Supporting Information

ABSTRACT: To predict the chemical activity of new matter is an ultimate goal in chemistry. The identification of reaction pathways using modern quantum mechanics calculations, however, often requires a high demand in computational power and good chemical intuition on the reaction. Here, a new reaction path searching method is developed by combining our recently developed transition state (TS) location method, namely, the constrained Broyden dimer method, with a basin-filling method via bias potentials, which allows the system to walk out from the energy traps at a given reaction direction. In the new method, the reaction path searching starts from an initial state without the need for preguessing the TS-like or final state structure and can proceed iteratively to the final state by locating all related TSs and intermediates. In each elementary reaction step, a reaction direction, such as a bond breaking, needs to be specified, the information of which is refined and preserved as a normal mode through biased dimer rotation. The method is tested successfully on the Baker reaction system (50 elementary reactions) with good efficiency and stability and is also applied to the potential energy surface exploration of multistep reaction processes in the gas phase and on the surface. The new method can be applied for the computational screening of new catalytic materials with a minimum requirement of chemical intuition.

1. INTRODUCTION

Transition state (TS) searching and activity prediction based on TS theory is a major theme in modern theoretical simulation of chemical reactions. As a chemical process involves generally multiple elementary steps where many TSs and intermediate states along a likely reaction path are present, the efficiency to locate TS is a key issue for chemists to predict chemical activity (e.g., determine the rate-determining step). The algorithms for locating TSs can be generally divided into two classes, namely, (i) the chain-of-states methods and (ii) the surface-walking methods. The former class locates the TS by simultaneously optimizing a few connected structure images on the potential energy surface (PES) to identify the reaction path.[1−11] The latter class requires only one image on the PES, and local information such as the gradient (force) or the second derivative (Hessian) of the PES is utilized to manipulate the image toward the TS.[12−23] Despite the huge progress made in the field, the current theoretical methods for simulating reaction pathways are still not sufficiently efficient, and both computational power and human experiences are highly demanding. Specifically, the chain-of-states methods require a knowledge of the exact final state, and the success of the surface-walking method relies heavily on the input (TS-like) structures. Good chemical intuition is therefore often indispensable in simulating reactions, which hinders the large-scale computational screening of catalytic materials for new applications.

Among all of the methods for TS searching, the Hessian-involved methods such as the P-RFO approach are perhaps the most efficient when the (analytic) Hessian is cheaply available.[12−15] By modifying and following the eigenvalue of the Hessian (eigenvector following), these methods can maximize energy in one degree of freedom while minimize energy in all the others. To reduce the computational cost in calculating the Hessian, the Quasi-Newton-based methods have been utilized to update the Hessian, such as Powell-symmetric-Broyden (PSB) and Symmetric rank 1 (Murtagh−Sargent),[24−26] and the hybrid approach developed by Bofill.[27−32] In practice, a constraint on the step length is often implemented to deal with the overstepping problem.[33] A comprehensive survey for methods in this category has been reviewed by Schlegel.[34] With the exact Hessian for TS location, new algorithms have also been developed for automated reaction path finding without any preknowledge on the PES, such as the reduced gradient following (RGF), the anharmonic downward distortion following (ADDF), and the artificial force induced reaction (AFIR), which have been successfully applied for reactions of small organic molecules.[32,35−39]

Different from the P-RFO approach, the dimer method, proposed by Henkelman and Jonsson[19] initially and improved by several other groups later,[20−23] can locate the TS without the need of a Hessian. This is particularly advantageous for the cases (i) when the Hessian is not cheaply available, as in many modern quantum mechanics packages with non-Gaussian basis sets, and (ii) where a complex reaction system with many degrees of freedoms is of interest.[21,22] The dimer method involves two structure images (defined as a dimer ($\mathbf{R}_1$, $\mathbf{R}_2$)) on the PES, the linkage between which creates a unit vector $\mathbf{N}$ separated by a prefixed distance $2\Delta R$(e.g., $\Delta R = 0.005$ Å). The whole algorithm of the dimer method is constituted by two independent parts. First, it is the rotation of the dimer to identify a relevant normal mode and its curvature (C) according to the dimer rotational force $\mathbf{\Delta}F^{\perp}$ in eq 1 (computationally, this is much less demanding

Received: March 27, 2012
Published: May 23, 2012

than solving the whole Hessian numerically). The vectors $\mathbf{F}_2$ and $\mathbf{F}_1$ are the forces acting on each image of the dimer ($\mathbf{R}_1$, $\mathbf{R}_2$). The curvature C corresponding to the dimer direction $\mathbf{N}$ (the normal mode) is calculated via finite difference using eq 2. Second, it is the translation of the middle point of the dimer ($\mathbf{R}_0$) toward the TS according to the translational force in eq 3 ($\mathbf{F}_{tran}$), where λ_1 and λ_2 are 0 and 1 at the $C > 0$ regions and are 1 and 1 at the $C < 0$ regions. In the equations, $\mathbf{F}_0$ is the total force acting on the middle point of the dimer. The parallel force $\mathbf{F}^{\parallel}$ is defined as the force component parallel to the dimer direction $\mathbf{N}$ ($\mathbf{F}^{\parallel} = (\mathbf{F}_0 \cdot \mathbf{N})\cdot\mathbf{N}$), and the vertical force $\mathbf{F}^{\perp}$ is defined by $\mathbf{F}^{\perp} = \mathbf{F}_0 - \mathbf{F}^{\parallel}$. To improve the efficiency, Olsen et al.[20] suggested that the force on the image 2 ($\mathbf{F}_2$) can be approximated as $\mathbf{F}_2 = 2\mathbf{F}_0 - \mathbf{F}_1$, which can save two steps of energy and gradient calculations in the determination of a rotation angle.

$$\Delta\mathbf{F}^{\perp} = (\mathbf{F}_1 - \mathbf{F}_2) - [(\mathbf{F}_1 - \mathbf{F}_2)\cdot\mathbf{N}]\cdot\mathbf{N} = 2(\mathbf{F}_1 - \mathbf{F}_0) - 2[(\mathbf{F}_1 - \mathbf{F}_0)\cdot\mathbf{N}]\cdot\mathbf{N} \quad (1)$$

$$C = \frac{(\mathbf{F}_2 - \mathbf{F}_1)\cdot\mathbf{N}}{2\Delta R} = \frac{(\mathbf{F}_0 - \mathbf{F}_1)\cdot\mathbf{N}}{\Delta R} \quad (2)$$

$$\mathbf{F}_{tran} = \lambda_1\mathbf{F}^{\perp} - \lambda_2\mathbf{F}^{\parallel} \quad (3)$$

Recently, we developed a constrained Broyden dimer (CBD) method by constraining the Quasi-Newton Broyden optimization trajectory with approximate normal mode as obtained from a constrained Broyden dimer rotation.[23] The CBD method can greatly reduce the rotation and translation steps required in TS searching without the need of a Hessian as input. Compared to the original dimer method, the CBD method (i) eliminates most of the rotation steps with the searching direction depending on only the approximate mode from the previous dimer rotation and (ii) carries out the geometry relaxation with multiple Broyden steps at a fixed approximate mode, while the original dimer method takes only one move (e.g., $\leq$ 0.2 Å) at each translation step. A key feature of the CBD method is that λ_1 and λ_2 in eq 3 are adjustable parameters: λ_1 is not zero even at the $C > 0$ region, and λ_2 can be lower than 1 at the $C < 0$ region. These two features allow the CBD method to find typical reaction paths with good efficiency and stability, including those in heterogeneous catalysis.[23,40,41] More details of the original dimer method and the improved versions can be found in previous publications.[19−23]

However, the CBD method also has limitations that are intrinsic to the dimer-based methods: (i) The efficiency of the algorithm is sensitive to the input, i.e., the initial structure and the initial mode. As a reasonable guess of these relies on good knowledge (chemical intuition) of the reaction, it becomes increasingly difficult for complex, multistep reactions when the exact final state is often not known *a priori*. (ii) The surface walking at the regions near the initial state (IS), final state (FS), or intermediate state basin (basin region) is not stable enough and can lead to the failure of TS searching. This is often due to a poor guess of the initial structure that locates at the basin region. The rotation of the dimer at the basin region (curvature $C > 0$) can converge to the softest modes, such as those related to rotation and translation, and thus loses the predefined reaction coordinate.

In this work, we aim to develop a better approach to remove the major limitations of the CBD method while retaining its superiority in locating TS at the regions near the TS (TS region). Indeed, we find that this can be achieved by applying bias potential in the dimer rotation and translation at the basin region. The idea of adding bias potential onto real PES has been widely practiced for sampling the PES more efficiently, e.g., in the metadynamics method.[42−45] Specifically, our new method can achieve the following: (i) The reaction path searching starts from the IS, and the guess of the initial structure is no longer required. (ii) The algorithm is stable in both the positive and negative curvature regions, and the predefined reaction direction can be maximally preserved at the positive curvature region. (iii) Multiple reactions can be explored in one single computational run. By testing our method in 50 Baker reactions and heterogeneous catalytic reactions, we show that the new algorithm is efficient and stable for locating TSs and identifying the reaction pathway of multiple elementary steps. The current approach can greatly speed up the identification of the reaction mechanism of complex reactions.

2. METHOD

The current method introduces bias potential on the real PES that allows the image to walk out from the basin region and then searches for the TS using the CBD method in the TS region. In the following, we describe our algorithm in three sections, namely, the biased rotation, the biased translation, and the CBD searching. The first two modules manipulate the structure image from the basin region to the TS region, and the third module is used to locate the TS accurately.

2.1. Biased Rotation. In the surface walking methods, the exact Hessian (e.g., P-RFO) or at least one normal mode (e.g., dimer methods) is required to manipulate the image toward the TS. The rotation algorithm in the CBD method is an efficient way to find a minimum in the normal mode space ($3N$ degrees of freedom). However, when all eigenvalues of the Hessian are positive, i.e., at the basin regions, the rotation algorithm (unbiased rotation) will often converge to the softest translation and rotational modes (as will be illustrated in Figure 1). Obviously, the TS searching would fail after losing the desired reaction direction.

To solve this problem, we add a bias potential V_N onto the PES of image $\mathbf{R}_1$ as in eq 4 to constrain the mode near the initial defined one. While a variety of mathematic functions, such as Gaussian, quadratic, and logarithmic functions, can be used as bias potential, here we chose a quadratic form as in eq 5 for simplicity, where a single parameter a determines the magnitude of the function. The force $\mathbf{F}_N$ due to the bias potential V_N, the rotational force $\Delta\boldsymbol{F}^{\perp}$, and the curvature C_{rot} of the modified PES of V_{R1} can then be calculated as in eqs 6−9.

$$V_{R1} = V_{real} + V_N \quad (4)$$

$$V_N = -\frac{a}{2}\cdot[(\mathbf{R}_1 - \mathbf{R}_0)\cdot\mathbf{N}_{init}]^2 = -\frac{a}{2}\cdot(\Delta R\cdot\mathbf{N}_t\cdot\mathbf{N}_{init})^2 \quad (5)$$

$$\mathbf{F}_N = -\frac{\partial V_N}{\partial q} = -\frac{\partial V_N}{\partial[(\mathbf{R}_1 - \mathbf{R}_0)\cdot\mathbf{N}_{init}]}\cdot\frac{\partial[(\mathbf{R}_1 - \mathbf{R}_0)\cdot\mathbf{N}_{init}]}{\partial q} = a\cdot\Delta R\cdot(\mathbf{N}_t\cdot\mathbf{N}_{init})\cdot\mathbf{N}_{init} \quad (6)$$

$$\Delta\mathbf{F}^{\perp} = \Delta\mathbf{F}^{\perp}_{real} + \Delta\mathbf{F}^{\perp}_{N} = 2(\mathbf{F}_1 + \mathbf{F}_N - \mathbf{F}_0) - 2[(\mathbf{F}_1 + \mathbf{F}_N - \mathbf{F}_0)\cdot\mathbf{N}_t]\cdot\mathbf{N}_t \quad (7)$$

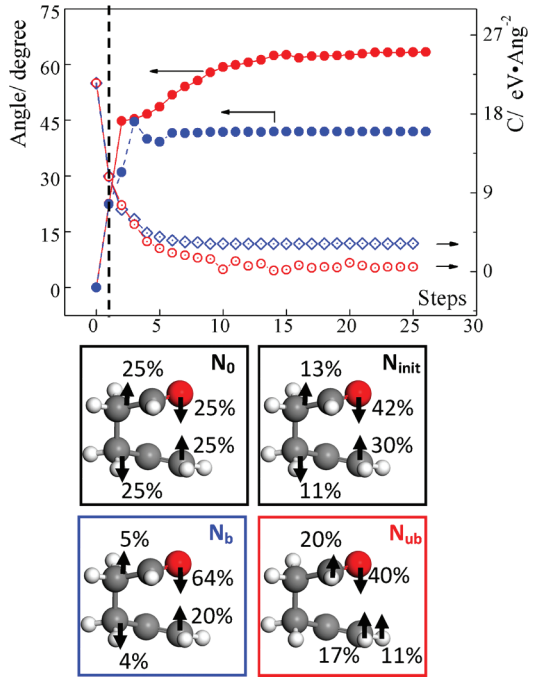


Figure 1. An illustration of the change of the rotation angle (solid symbols, left axis) and the curvature (hollow symbols, right axis) from the unbiased rotation (red curves) and the biased rotation (blue curves) in the basin region. The rotation starts from the IS of the Claisen rearrangement reaction ($CH_2CHCH_2CH_2CHO \rightarrow CH_2CHOCH_2CHCH_2$) with the input raw mode $\mathbf{N_0}$ and converges to $\mathbf{N_b}$ and $\mathbf{N_{ub}}$ for the biased and unbiased rotation, respectively. The dominant components (the quadratic sum of three dimensions) of $\mathbf{N_0}$, $\mathbf{N_{init}}$, $\mathbf{N_b}$, and $\mathbf{N_{ub}}$ on the atoms are also shown with the arrows indicating the displacement direction. The first two steps (before the dashed line) in rotation are the unbiased rotation to obtain $\mathbf{N_{init}}$. The rotation angle [$\theta = \arccos(\mathbf{N_t}\cdot\mathbf{N_0})$] measures the deviation of the mode at the current step ($\mathbf{N_t}$) with respect to the initial mode $\mathbf{N_0}$.

$$C_e = \frac{(\mathbf{F_0} - \mathbf{F_I})\cdot\mathbf{N_t}}{\Delta R} \tag{8}$$

$$\begin{aligned} C_{rot} &= C_e + C_N \\ &= \frac{(\mathbf{F_0} - \mathbf{F_N} - \mathbf{F_I})\cdot\mathbf{N_t}}{\Delta R} \\ &= \frac{(\mathbf{F_0} - \mathbf{F_I})\cdot\mathbf{N_t}}{\Delta R} - a\cdot(\mathbf{N_t}\cdot\mathbf{N_{init}})^2 \end{aligned} \tag{9}$$

In the equations, q represents the generalized nuclei position; $\mathbf{N_t}$ is the normal mode at the current iteration; $\mathbf{N_{init}}$ is the initial mode in the biased rotation, the generation of which is described later; C_e is the estimated curvature of real PES (V_{real}) at the current iteration. Since $C_{rot} = C_e - a$ for $\mathbf{N_t} = \mathbf{N_{init}}$ (at the first iteration of biased rotation), the value of "a" should at least be C_e ($a = C_e$) to make the value of C_{rot} negative ($C_{rot} < 0$) in the subsequent rotation iterations. In other words, by adding a bias potential V_N, we can localize the dimer rotation space near the initial mode. This can also be seen from the expression of $\mathbf{\Delta F}^{\perp}$. According to eq 7, the bias potential will generate a rotational force $\mathbf{\Delta F_N^{\perp}} = 2\mathbf{F_N} - 2(\mathbf{F_N}\cdot\mathbf{N_t})\cdot\mathbf{N_t}$ that has a component pointing to $\mathbf{N_{init}}$. When the rotation converges ($|\mathbf{\Delta F}^{\perp}|$ close to 0), the optimized mode $\mathbf{N_t}$ can retain the key message from $\mathbf{N_{init}}$ ($\mathbf{N_t}$ locates nearby $\mathbf{N_{init}}$ with the angle [θ = arccos ($\mathbf{N_{init}}\cdot\mathbf{N_t}$)] typically below 30°) and also take into account the other useful information from the real PES.

$$\mathbf{N_{bf}} = \begin{pmatrix} q_B \\ q_A \\ 0 \\ \vdots \end{pmatrix} - \begin{pmatrix} q_A \\ q_B \\ 0 \\ \vdots \end{pmatrix} \tag{10}$$

$$\mathbf{N_{rot}} = \begin{pmatrix} 0 \\ 0 \\ (q_C - q_A) \times (q_C - q_B) \\ 0 \\ \vdots \end{pmatrix} \tag{11}$$

In practice, $\mathbf{N_{init}}$ of chemical reactions can be created in two steps. First, one needs to specify the atoms that are related to the chemical reaction, for example, those being involved in a chemical bond breaking/making or in a rotation transformation. A raw mode $\mathbf{N_0}$ can then be generated as input using the coordinate of the atoms. For example, a bond formation $\mathbf{N_0}$ ($\mathbf{N_{bf}}$) associated with atom A (e.g., the first atom in system) and atom B (the second atom in system) can be derived as eq 10, using their coordinates q_A and q_B. A rotational mode ($\mathbf{N_{rot}}$) can be derived as eq 11 that represents the mode of the atom C (e.g., the third atom in system) rotating around the axis defined by atoms A and B (FORTRAN code is given in the Supporting Information). Second, $\mathbf{N_0}$ will be further refined via limited steps of unbiased dimer rotation (i.e., without bias potential) to yield a more realistic mode as $\mathbf{N_{init}}$. The force criterion for this initial sweep is generally rather large, e.g., 10 times larger than that required for the typical dimer rotation to converge a mode.

As an illustration, we compare the unbiased rotation and the biased rotation in the basin region in Figure 1. In this example, $\mathbf{R_0}$ is the IS of the Claisen rearrangement reaction, $CH_2CHCH_2CH_2CHO \rightarrow CH_2CHOCH_2CHCH_2$, and the input raw mode $\mathbf{N_0}$ comprises only information of the cleavage of a C−C bond and the formation of an O−C bond. Starting from $\mathbf{N_0}$, an initial sweep of unbiased rotation (only two steps in this example) is utilized to generate $\mathbf{N_{init}}$, which is 22.6° away from $\mathbf{N_0}$. In the unbiased rotation, $\mathbf{N_{init}}$ is converged to a mode ($\mathbf{N_{ub}}$) that is mainly the rotational movement of O and nearby H atoms. $\mathbf{N_{ub}}$ is 63° away from $\mathbf{N_0}$ with a curvature of 0.51 eV/Å^2, suggesting that the reaction information defined by $\mathbf{N_0}$ has diminished largely. In contrast, the biased rotation converges to a mode ($\mathbf{N_b}$) that is 42° away from $\mathbf{N_0}$, and the optimized mode is still dominated by the C and the O bond formation. The final curvature (i.e., C_e in eq 8) is 3.21 eV/Å^2 (note that the overall C_{rot} is negative, −5.34 eV/Å^2), which is significantly larger than that from the unbiased rotation.

2.2. Biased Translation. With the normal mode obtained from the biased rotation, we can use the information to translate the image uphill to the TS. In this part, we also utilize bias potentials to drag the image out from the basin region. The potential well will be gradually filled by the added potentials, as shown in Figure 2, and the structure image is pushed away from the bottom of the basin. The basic idea of the basin filling is similar to that in the metadynamics method[42−45] (although the free energy is not the concern in our case).

In our implementation, we continuously add bias potentials along the reaction coordinate, which will generate a number of minima labeled as $\mathbf{R_0^i}$ (i = 1, 2... counting the iteration number)

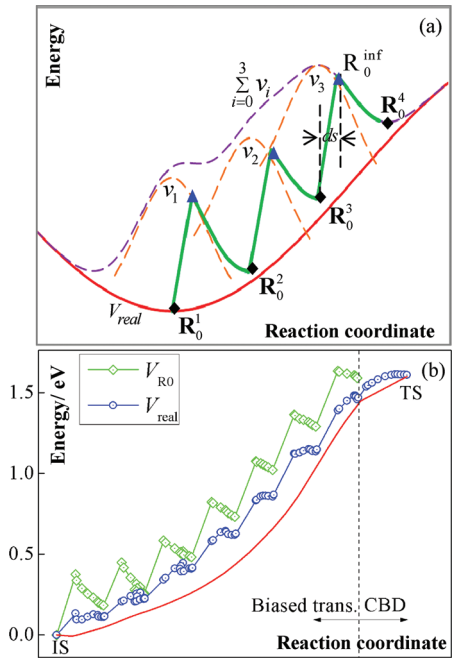


Figure 2. (a) Illustration of the biased translation. The red, orange, purple, and green curves represent the real PES (V_{real}), the Gaussian functions (v_1, v_2, and v_3), the modified PES (V_{R0}), and the TS searching trajectory, respectively. $\mathbf{R}_0^i$ labels the local minima (black dots) on a modified PES. The inflection points of Gaussian functions ($\mathbf{R}_0^{inf} = \mathbf{R}_0^i + ds \cdot \mathbf{N}_i$) are also indicated (blue dots). (b) The TS searching trajectory in the Claisen rearrangement reaction. The *x* axis is along the IS−FS vector by setting IS x-coordinate at 0 and FS at 1. The red curve is the optimization trajectory from TS to IS using the L-BFGS algorithm. The dashed line indicates the end of the biased translation.

on the modified PES in Figure 2a. This allows us to utilize the traditional minimization technique (e.g., quasi-Newton Broyden as utilized in this work) for structural relaxation toward these local minima. The PES of $\mathbf{R}_0$ is modified as eq 12, where the bias potential V_G is a sum of a series of Gaussian functions v_i that added sequentially along the reaction coordinate (see eqs 12−16). The forces are evaluated according to eqs 17 and 18, where $\mathbf{F}_{R0}$ is the force on the modified PES (V_{R0}) and $\mathbf{F}_G$ is that due to the bias potential (V_G).

$$V_{R0} = V_{real} + V_G \tag{12}$$

$$V_G = \sum_i v_i \ (i = 1, 2, 3...k) \tag{13}$$

$$v_i = w_i \times \exp\left[-\frac{(s_t^i - s_0^i)^2}{2 \times ds^2}\right] \tag{14}$$

$$s_t^i = \mathbf{R}_0^t \cdot \mathbf{N}_i \tag{15}$$

$$s_0^i = \mathbf{R}_0^i \cdot \mathbf{N}_i \tag{16}$$

$$\mathbf{F}_{R0} = -\frac{\partial V_{R0}}{\partial q} = \mathbf{F}_{real} + \mathbf{F}_G \tag{17}$$

$$\boldsymbol{F}_G = -\frac{\partial V_G}{\partial s} \cdot \frac{\partial s}{\partial q} = \sum_i w_i \cdot \exp\left[-\frac{(s_t^i - s_0^i)^2}{2 \times ds^2}\right] \cdot \frac{s_t^i - s_0^i}{ds^2} \cdot \mathbf{N}_i \tag{18}$$

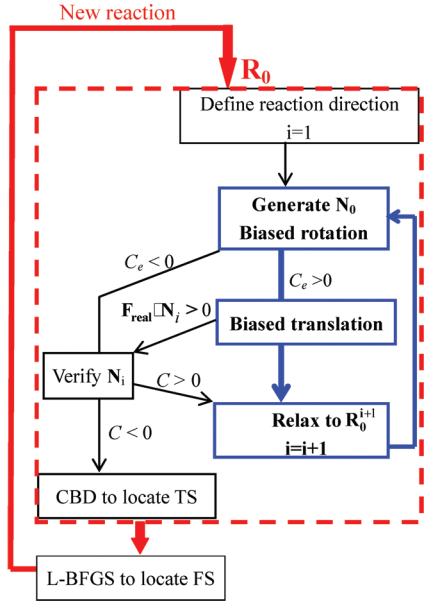
In the equations, $\mathbf{R}_0^t$ is the current structure during the translation; $\mathbf{N}_i$ is the normal mode at $\mathbf{R}_0^i$ obtained by the biased-rotation. The Gaussian function v_i has two parameters, *w* and *ds*, the height and the width of Gaussian. In practice, *ds* is set as a constant, such as 0.1 Å. As the basic role of the Gaussian functions is to generate the proper local minima along the reaction coordinate, the choice of the Gaussian parameters must always satisfy $\mathbf{F}_{R0} \cdot \mathbf{N}_i \geq 0$ when $\mathbf{R}_0^t = \mathbf{R}_0^k + ds \cdot \mathbf{N}_k$ at the *k*th iteration (i.e., *ds* away from the *k*th Gaussian v_k maximum along $\mathbf{N}_k$) where $\mathbf{F}_G$ reaches the maximum (the inflection point $\mathbf{R}_0^{inf}$ of the Gaussian). This will guarantee that $\mathbf{F}_{R0}$ is pointing toward the TS instead of the IS.

The procedure of the biased translation is summarized as follows: (i) Move the image from the local minimum ($\mathbf{R}_0^k$) to $\mathbf{R}_0^k + ds \cdot \mathbf{N}_k$ and evaluate $\mathbf{F}_{real}$. (ii) Derive the form of $\mathbf{F}_{R0} \cdot \mathbf{N}_i$, a function of *w*, and determine the value of *w* to let $\mathbf{F}_{R0} \cdot \mathbf{N}_i = 0.1$. (iii) According to *w* and *ds*, we add Gaussian onto the real PES and utilize the minimization technique to identify the local minimum $\mathbf{R}_0^{k+1}$. To demonstrate this in the real reaction, in Figure 2b we show the trajectory of Claisen rearrangement from the IS to the TS with the biased translation. As shown, seven Gaussian functions (v_i) are added to the real PES before the image reaches the TS region ($C < 0$).

2.3. The Location of TS Using the CBD Method. The CBD method developed in our previous work is utilized for locating the TS once a negative curvature along the reaction coordinate is identified ($C_e < 0$) in the biased rotation and confirmed ($C < 0$) by the unbiased rotation. In practice, we also met the situation that the addition of a new Gaussian moves the image directly over the TS and even into the FS basin. To avoid the failure of the TS location in these situations, the $\mathbf{F}_{real} \cdot \mathbf{N}_i < 0$ criterion is always examined in the biased translation. As soon as $\mathbf{F}_{real} \cdot \mathbf{N}_i > 0$ (the force pointing to the FS), indicating the image may be over the TS, the biased translation is terminated and an unbiased rotation is invoked to verify whether a negative curvature is present. If not, the current $\mathbf{N}_i$ is reverted in direction, and the image can go uphill again (from the FS basin) by repeating the biased rotation and biased translation.

2.4. The Overall Algorithm for Reaction Pathway Searching. The current approach for reaction pathway searching is summarized in Scheme 1, which comprises four

Scheme 1. The Overall Algorithm for Continuous Reaction Pathway Searching



dx.doi.org/10.1021/ct300250h | J. Chem. Theory Comput. 2012, 8, 2215−2222

major modules, namely, the biased rotation, the biased translation, the TS location using CBD, and the FS location. Because of the simple architecture of the method, it is trivial to extend the algorithm for pathway searching of multiple reactions, in which we set the FS as the new IS and provide a new reaction direction. The overall procedure is described as follows:

(i) A reaction direction (such as atomic pairs) and the coordinate of the energy minimum (IS) $\mathbf{R}_0^i$ (i = 1) are provided as input.
(ii) $\mathbf{N}_0$ is generated and refined by the biased rotation until $|\Delta \boldsymbol{F}^{\perp}| < \tau_1$ (e.g., 0.1 eV/Å). After convergence, $\mathbf{N}_i$ (i = 1, 2, 3...) is obtained and C_e is calculated.
(iii) If C_e is positive, we utilize the biased translation to drag the image uphill. The local structural relaxation toward the minimum $\mathbf{R}_0^{i+1}$ is terminated until $|\mathbf{F}_{R0}| < \tau_2$ (e.g., 0.15 eV/Å).
(iv) If C_e is negative in ii or $\mathbf{F}_{real}\cdot\mathbf{N}_i > 0$ in iii, the curvature C will be verified by the unbiased rotation.
(v) Repeat steps ii and iii until a negative C is met in iv.
(vi) Utilize the CBD module to locate the TS.
(vii) Displace the TS toward FS and optimize the structure downhill to the FS. The limited memory BFGS (L-BFGS) method[46] is utilized for the structural optimization.

3. RESULTS AND DISCUSSION

3.1. Efficiency for TS Location. To test the efficiency of our new method, we first chose the Baker reaction system[47] as the testing example for TS location, which contains 25 chemical reactions, as listed in Table 1. The same reactions have been utilized to test the modified dimer methods and the CBD method previously.[23] For every reaction, both the forward reaction and the reverse reaction are investigated. For clarity, the endothermic reaction is defined as the forward reaction and the exothermic reaction as the reverse reaction. All calculations were performed using the density functional theory SIESTA package[48] with the numerical double-ζ polarization basis set[49,50] at the GGA-PBE exchange-correlation functional level.[51] The rotation criterion τ_1 for $|\Delta \boldsymbol{F}^{\perp}|$ convergence is 0.1 eV/Å. The criterion τ_2 for $\mathbf{F}_{R0}$ convergence is 0.15 eV/Å. The criterion in the CBD method for TS location is 0.1 eV/Å for the maximum force on each degree of freedom, except for two rotation reactions (11, 21) where the criterion is set as 0.02 eV/Å (very flat PES for the rotation reactions). All of the determined TSs have been checked with the literature[47] to ensure that the correct TS is identified. Our main results are listed in Table 1. On average, the forward reactions require 126.8 steps (energy and force evaluation) to locate the TS; the reverse reactions need 96.0 steps. Consistently, the number of Gaussian functions (nG) utilized in the biased translation is 6.24 and 5.24 on average for the forward and reverse reactions, respectively.

We notice that the forward (endothermic) reaction does not necessarily take more calculation effort than the reverse (exothermic) reaction, which implies that the barrier height E_a may not be the key factor that determines the computational cost. This can be seen clearly in Figure 3, which plots the total steps against E_a (Figure 3a) and against $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$, the absolute distance between IS and TS (Figure 3b), respectively [two rotation reactions (11 and 21) are not taken into account because

Table 1. Results of our Method for Locating TS of Baker Reaction System[a]

		forward reactions				reverse reactions			
no.	system	steps	$\lvert R^{IS} - R^{TS}\rvert$	E_a	nG	steps	$\lvert R^{IS} - R^{TS}\rvert$	E_a	nG
1	HCN → HNC	67	1.00	1.86	3	140	1.21	1.25	7
2	HCCH → CCH_2	79	1.33	1.72	3	26	0.41	0.03	2
3	H_2CO → H_2+CO	125	0.79	3.57	6	139	1.45	2.83	9
4	CH_3O → CH_2OH	56	0.68	1.32	4	44	0.70	1.29	4
5	ring-opening cyclopropyl	292	2.50	2.08	11	82	0.85	1.11	2
6	bicyclo110 butane TS1	147	1.56	1.92	3	48	0.08	0.01	1
7	bicyclo110 butane TS2	212	2.69	1.19	9	42	0.79	0.17	1
8	β-(formyloxy) ethyl	134	1.82	0.51	5	134	1.82	0.51	3
9	parent Diels–Alder	126	1.07	2.64	4	52	0.72	0.28	3
10	s-tetrazine → 2HCN + N_2	163	1.76	2.20	8	65	0.78	2.02	3
11	rotational TS in butadiene	102	2.24	0.33	5	87	1.88	0.19	5
12	H_3CCH_3 → H_2CCH_2 + H_2	133	1.36	4.84	9	153	1.76	2.98	10
13	H_3CCH_2F→ H_2CCH_2 + HF	80	1.04	2.20	6	105	1.09	1.53	7
14	H_2CCHOH → H_3CCHO	191	1.38	2.65	8	158	1.41	2.11	9
15	HCOCl → HCl + CO	58	0.82	1.64	4	86	1.27	1.13	6
16	H_2O + PO_3^- → $H2PO_4^-$	75	1.05	0.96	5	122	1.73	0.78	6
17	Claisen rearrangement	122	1.45	1.61	7	72	1.14	0.56	4
18	silylene insertion	126	1.41	2.71	9	65	1.00	0.72	4
19	HNCCS → HNC + CS	79	1.12	1.48	1	125	2.09	0.20	7
20	$HCONH3^+$ → $NH4^+$ + CO	183	1.41	1.76	7	114	1.04	1.44	5
21	rotational TS in acrolein	60	1.76	0.40	7	88	1.85	0.30	6
22	HCONHOH → HCOHNHO	119	1.07	1.59	7	115	1.12	1.36	6
23	HNC + H_2 → H_2CNH	115	0.89	3.75	7	173	1.89	2.03	10
24	H_2CNH → $HCNH_2$	89	1.05	3.43	6	75	0.84	1.89	4
25	$HCNH_2$ → HCN + H_2	238	1.88	3.32	12	90	0.80	2.91	7
	average	126.8	1.41	2.07	6.24	96.0	1.19	1.19	5.24

[a] $|R^{IS} - R^{TS}|$ is the absolute distance between IS and TS (unit: Å); E_a is the reaction barrier (unit: eV); nG is the number of Gaussian functions added in the biased translation.

dx.doi.org/10.1021/ct300250h | J. Chem. Theory Comput. 2012, 8, 2215–2222

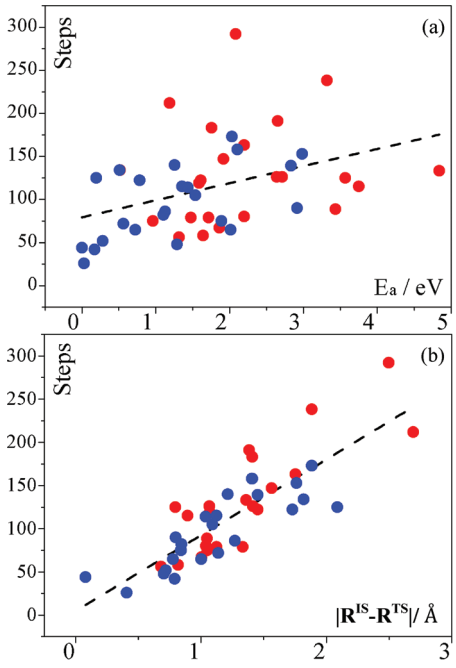


Figure 3. The plots of the total steps against E_a (a) and $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$ (b) of the Baker reaction system. Red and blue dots indicate the forward and the reverse reactions, respectively. The data are from Table 1. Two rotation reactions (11 and 12) are not taken into account.

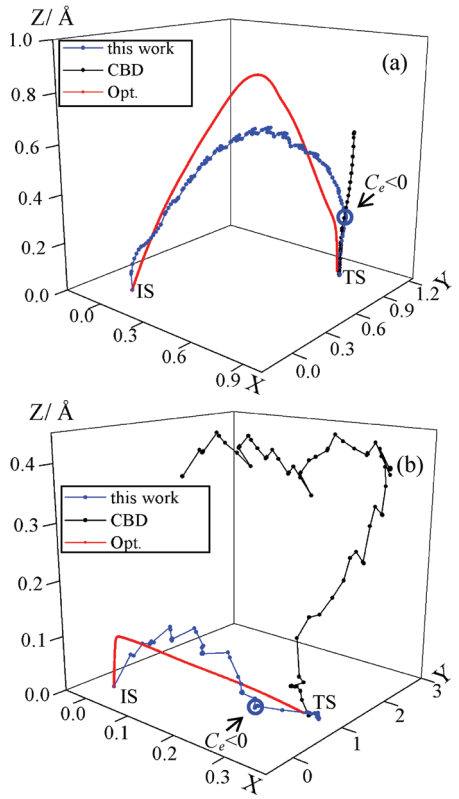


Figure 4. 3D trajectory of TS searching in (a) the Claisen rearrangement reaction and (b) the HCOCl → HCl + CO reaction. The meaning of the axes are as defined in eqs 19−21. The blue trajectory, the current method (starting from the IS); the black trajectory, the CBD method alone (starting from a TS-like guess structure) as reported in the previous work; and the red trajectory, the steepest descent curve from TS to IS.

their PESs are too flat and thus different from the others]. Clearly, the number of steps can be better correlated with $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$ than with E_a. For the reactions investigated, $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$ is 1.35 Å on average for the forward reaction (blue dots) and 1.13 Å for the reverse reaction (red dots). The ratio of the average $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$ (1.35/1.13 = 1.20) is similar to the ratio of the average step number (131/97 = 1.35) and the ratio of the average number of Gaussian function (6.26/5.22 = 1.20). These indicate that the efficiency of the current method is more related to the geometrical position of TS along the reaction path. In other words, the TS searching in the endothermic reactions that have small $|\mathbf{R}^{IS} - \mathbf{R}^{TS}|$ can be more efficient than in their reverse exothermic reactions. In the Baker system, there are 12 such reactions, namely, 1, 3, 4, 12−16, 19, 21−23. Obviously, the traditional thought that the TS is more FS-like for an endothermic reaction[52] does not always hold. This reflects the great difficulty in making a good guess of TS-like structure in general for the surface walking methods.

In order to see more clearly how our new method works, we generate a three-dimensional diagram to trace the searching trajectory, as shown in Figure 4, where the x axis is along the IS−FS vector by setting IS at (0,0,0) and FS at (1,0,0), and the y axis is determined by setting the TS at (x^{TS},1,0) where x^{TS} is the projection of TS onto the IS−FS vector. The z axis is the direction perpendicular to the IS−TS−FS plane. A large z value of a structure would imply that the structure is far away from the reaction plane and thus is not desirable in TS searching. The unit vectors $\mathbf{X}$ and $\mathbf{Y}$ and the z value are defined in eqs 19−21. The same 3D plot has been utilized in our previous work.[23]

$$\mathbf{X} = \frac{\mathbf{R}^{FS} - \mathbf{R}^{IS}}{|\mathbf{R}^{FS} - \mathbf{R}^{IS}|} \tag{19}$$

$$\mathbf{Y} = \frac{(\mathbf{R}^{TS} - \mathbf{R}^{IS}) - [(\mathbf{R}^{TS} - \mathbf{R}^{IS})\cdot\mathbf{X}]\cdot\mathbf{X}}{|(\mathbf{R}^{TS} - \mathbf{R}^{IS}) - [(\mathbf{R}^{TS} - \mathbf{R}^{IS})\cdot\mathbf{X}]\cdot\mathbf{X}|} \tag{20}$$

$$z = |(\mathbf{R} - \mathbf{R}^{IS}) - [(\mathbf{R} - \mathbf{R}^{IS})\cdot\mathbf{X}]\cdot\mathbf{X} - [(\mathbf{R} - \mathbf{R}^{IS})\cdot\mathbf{Y}]\cdot\mathbf{Y}| \tag{21}$$

Using Claisen rearrangement ($CH_2CHCH_2CH_2CHO$ → $CH_2CHOCH_2CHCH_2$) and the HCOCl → HCl + CO reaction (Figure 4a and b) as examples, we have plotted three trajectories for each reaction: the blue curve with the current method (starting from the IS), the black one with the CBD method alone (starting from a TS-like guess structure as utilized in the previous work[23]), and the red one as a mimic of the reaction path (optimization starting from TS using steepest descent method). Figure 4a shows that when a good guess is available initially (i.e., close to the TS), the CBD method is the most efficient and can identify TS staightforwardly. The current method (blue curve) can identify the TS by following the reaction path closely.

On the contrary, when a good initial guess is not available such as that in the HCOCl→HCl+CO reaction shown in Figure 4b, the CBD method alone struggles to locate TS with a large detour (black curve) as reported in our previous work.[23] Apparently, the efficiency of the CBD method is sensitive to the initial structure. In such cases, the current method (blue curve) shows the advantage as it does not require the initial guess. In fact, the total steps in the blue curve (58 steps) is resasonbly larger than the red curve (21 steps), being much lower than that in the black curve (109 steps). It is noticed that when the CBD module is invoked in the current method, the image is already rather close to TS.

3.2. Reaction Pathway Searching with Multiple Elementary Reactions. We then applied our method for

searching the reaction pathway where multiple elementary reactions are present. One gas phase reaction, i.e. *cis*- to *trans*-butadiene and to cyclobutene conversion, and one surface catalytic process, i.e. glycerol initial dehydrogenations on Ru(0001), are investigated, and the results are discussed in the following.

For the *cis*- to *trans*-butadiene and to cyclobutene conversion, there are two TSs and three minima. In one reaction cycle, the molecule precedes the following consecutive steps: (i) trans to cis transformation, (ii) ring closing to form cyclobutene, (iii) ring opening to form *cis*-butadiene, and finally (iv) cis to trans transformation. Figure 5 shows the trajectory of the reaction

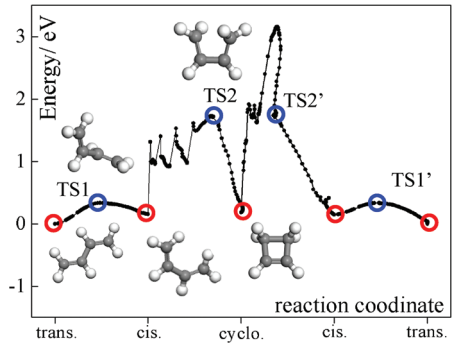


Figure 5. Illustration of our method for the searching pathway of multiple reactions in one single trajectory. The example reaction is butadiene to cyclobutene conversion. The *x* axis is along IS−FS1−FS2... vectors by setting the IS (trans-butadiene) *x* coordinate at 0, the FS1 (cis-butadiene) at 1, the FS2 (cyclobutene) at 2, and so on. The obvious detour in the profile (for the reaction from cyclobutene to *cis*-butadiene) is due to the CBD searching in the TS region.

pathway searching using our method with four TSs and four FSs identified in one single computational run, which takes 1297 steps in total and 162 steps on average for finding each state (the force criterion for the TS location is 0.01 eV/ Å due to the flat PES of cis to trans conversion). Some interesting features in the reaction pathway can thus be revealed. For example, the PES is very flat for the cis-to-trans transformation, and the TS sits at the middle of the IS and FS. For the cis to the cyclobutene conversion, the PES is very corrugated, and the TS is closer to the FS, cyclobutene, despite the reaction being almost thermal neutral. The calculated reaction barriers are 0.33, 1.56, 1.53, and 0.18 eV for these consecutive reactions, which are consistent with previous calculations.[44]

In the second example, we considered a heterogeneous catalytic reaction on the surface, namely, glycerol decomposition/conversion on metal surfaces that has attracted much attention in recent years for its potential application in biomass energy.[53−55] One of the key challenges in glycerol chemistry is to understand the catalytic selectivity since a great number of possible reaction channels are likely (glycerol has 13 bonds in total and thus 13! (10^9) possible reaction pathways in theory). We have applied our method for the initial dehydrogenation of this complex reaction system, and the results are shown in Figure 6. We start from the adsorbed glycerol on Ru(0001) and investigate its first three dehydrogenation reactions (1-OH, 2-OH, and 3-CH). All of the intermediate structures and TSs are located in one single computational run. In total, 264 calculation steps are required to locate all of the TSs and FSs, i.e., 44 steps on average to identify each state. It is interesting to notice that the steps to locate TSs (43) and FSs (45) are similar. The high

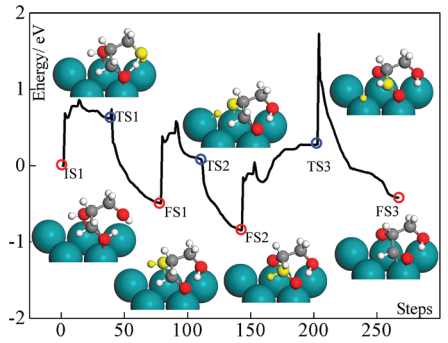


Figure 6. Glycerol initial dehydrogenation on the Ru(0001) surface. The inserts show the located structures of TSs and FSs (H, white; C, gray; O, red; Ru, cyan). The atoms involved in the bond-breaking are highlighted in yellow. After each TS, a simple strategy is applied to eliminate the interference of the dissociated H on the next reaction, i.e., by displacing the dissociating H (3 Å) away from its TS position (in the case of TS3, it induces an immediate sharp increase in energy).

efficiency demonstrated here provides the hope to resolve the kinetics puzzles of such complex reaction systems.

Finally, it should be emphasized that while the current method is capable of connecting multiple minima on the PES by locating the associated TSs, the identified reaction channel may not necessarily be the minimum energy channel to connect these states. This is not least because the intermediate states (i.e., ISs, FSs) optimized using quasi-Newton methods only represent the local minima on the PES, and the global minimization techniques such as the basin-hopping method with Metropolis Monte Carlo[56] would certainly be the better choice for exploring all of the likely structures. For example, in heterogeneous catalytic reactions on surfaces, there are many possible adsorption sites for adsorbates, e.g., hollow, bridge, and top sites, and thus a number of likely ISs, TSs, and FSs are present for one elementary reaction. Technically, the combination of the current method with the global minimization techniques is straightforward and can even be generalized to explore the whole PES with random generated initial modes. The major concern is however the computational demand and the efficiency in the global minimization, which is beyond the scope of the current work.

4. CONCLUSION

This work developed a new algorithm for reaction path searching by combining the recently developed CBD method with a basin-filling method via bias potentials. The method is designed to explore the PES starting from a minimum structure (IS) without the need for preguessing the TS-like or FS structures, which eliminates the reliance upon *a priori* knowledge of the reaction in traditional TS-searching methods and allows the continuous pathway searching of multiple reactions in one single computational run. We test our method in the Baker reaction system (50 elementary reactions), which shows that good stability and efficiency can be achieved for TS location. The ability for the pathway searching of multiple reactions is also demonstrated in a gas phase reaction, i.e., butadiene to cyclobutene conversion, and a heterogeneous catalytic surface reaction, i.e., glycerol initial dehydrogenation. As the new method is largely intuition free, it would be especially useful for the large-scale computational screening of new catalytic materials.

■ ASSOCIATED CONTENT

Supporting Information

FORTRAN code for generating initial raw mode $\mathbf{N_0}$; the coordinates and the related reaction variables of the Baker reaction system, the *cis*- to *trans*-butadiene and to cyclobutene conversion, and glycerol initial dehydrogenation on the Ru(0001) surface. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Fax: (+86) 21-6564-2400. E-mail: zpliu@fudan.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by National Nature Science Foundation of China (20825311, 21173051), 973 program (2011CB808500), Science and Technology Commission of Shanghai Municipality (08DZ2270500).

■ REFERENCES

(1) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978−9985.
(2) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901−9904.
(3) Mills, G.; Jonsson, H. *Phys. Rev. Lett.* **1994**, *72*, 1124−1127.
(4) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.
(5) Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375−380.
(6) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 7820−7820.
(7) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082−2094.
(8) Koslover, E. F.; Wales, D. J. *J. Chem. Phys.* **2007**, *127*, 134102.
(9) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
(10) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 7877−7886.
(11) E, W. N.; Ren, W. Q.; Vanden-Eijnden, E. *Phys. Rev. B: Condens. Matter* **2002**, *66*, 052301.
(12) Simons, J.; Jorgensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745−2753.
(13) Khait, Y. G.; Puzanov, Y. V. *THEOCHEM* **1997**, *398*, 101−109.
(14) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800−2806.
(15) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385−395.
(16) Munro, L. J.; Wales, D. J. *Phys. Rev. B: Condens. Matter* **1999**, *59*, 3969−3980.
(17) Kumeda, Y.; Wales, D. J.; Munro, L. J. *Chem. Phys. Lett.* **2001**, *341*, 185−194.
(18) Wang, H. F.; Liu, Z. P. *J. Am. Chem. Soc.* **2008**, *130*, 10996−11004.
(19) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 7010−7022.
(20) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jonsson, H. *J. Chem. Phys.* **2004**, *121*, 9776−9792.
(21) Heyden, A.; Bell, A. T.; Keil, F. J. *J. Chem. Phys.* **2005**, *123*, 224101.
(22) Kaestner, J.; Sherwood, P. *J. Chem. Phys.* **2008**, *128*, 014106.
(23) Shang, C.; Liu, Z. P. *J. Chem. Theory Comput.* **2010**, *6*, 1136−1144.
(24) Powell, M. J. D. *Math. Program.* **1971**, *1*, 26−57.
(25) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214−218.
(26) Culot, P.; Dive, G.; Nguyen, V. H.; Ghuysen, J. M. *Theor. Chim. Acta* **1992**, *82*, 189−205.
(27) Bofill, J. M. *J. Comput. Chem.* **1994**, *15*, 1−11.
(28) Anglada, J. M.; Besalu, E.; Bofill, J. M.; Rubio, J. *J. Math. Chem.* **1999**, *25*, 85−92.
(29) Bofill, J. M.; Comajuan, M. *J. Comput. Chem.* **1995**, *16*, 1326−1338.
(30) Bofill, J. M. *Chem. Phys. Lett.* **1996**, *260*, 359−364.
(31) Anglada, J. M.; Bofill, J. M. *J. Comput. Chem.* **1998**, *19*, 349−362.
(32) Bofill, J. M.; Anglada, J. M. *Theor. Chim. Acta* **2001**, *105*, 463−472.
(33) Besalu, E.; Bofill, J. M. *Theor. Chem. Acc.* **1998**, *100*, 265−274.
(34) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.
(35) Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. *J. Comput. Chem.* **1998**, *19*, 1087−1100.
(36) Ohno, K.; Maeda, S. *Chem. Phys. Lett.* **2004**, *384*, 277−282.
(37) Hirsch, M.; Quapp, W. *THEOCHEM* **2004**, *683*, 1−13.
(38) Maeda, S.; Morokuma, K. *J. Chem. Theory Comput.* **2011**, *7*, 2335−2345.
(39) Crehuet, R.; Bofill, J. M.; Anglada, J. M. *Theor. Chim. Acta* **2002**, *107*, 130−139.
(40) Shang, C.; Liu, Z. P. *J. Am. Chem. Soc.* **2011**, *133*, 9938−9947.
(41) Fang, Y. H.; Liu, Z. P. *J. Am. Chem. Soc.* **2010**, *132*, 18214−18222.
(42) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(43) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(44) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
(45) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826−843.
(46) Liu, D. C.; Nocedal, J. *Math. Program.* **1989**, *45*, 503−528.
(47) Baker, J.; Chan, F. R. *J. Comput. Chem.* **1996**, *17*, 888−904.
(48) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745−2779.
(49) Junquera, J.; Paz, O.; Sanchez-Portal, D.; Artacho, E. *Phys. Rev. B: Condens. Matter* **2001**, *64*, 235111.
(50) Anglada, E.; Soler, J. M.; Junquera, J.; Artacho, E. *Phys. Rev. B: Condens. Matter* **2002**, *66*, 205101.
(51) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(52) Schlegel, H. B. *WIREs Comput. Mol. Sci.* **2011**, *1*, 790−809.
(53) Zope, B. N.; Hibbitts, D. D.; Neurock, M.; Davis, R. J. *Science* **2010**, *330*, 74−78.
(54) Zhou, C. H. C.; Beltramini, J. N.; Fan, Y. X.; Lu, G. Q. M. *Chem. Soc. Rev.* **2008**, *37*, 527−549.
(55) ten Dam, J.; Hanefeld, U. *Chemsuschem* **2011**, *4*, 1017−1034.
(56) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.