JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

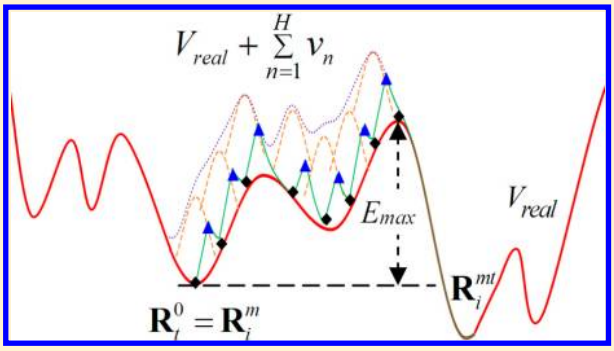
Stochastic Surface Walking Method for Structure Prediction and Pathway Searching

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: We propose an unbiased general-purpose potential energy surface (PES) searching method for both the structure and the pathway prediction of a complex system. The method is based on the idea of bias-potential-driven dynamics and Metropolis Monte Carlo. A central feature of the method is able to perturb smoothly a structural configuration toward a new configuration and simultaneously has the ability to surmount the high barrier in the path. We apply the method for locating the global minimum (GM) of short-ranged Morse clusters up to 103 atoms starting from a random structure without using extra information from the system. In addition to GM searching, the method can identify the pathways for chemical reactions with large dimensionality, as demonstrated in a nanohelix transformation containing 222 degrees of freedoms.

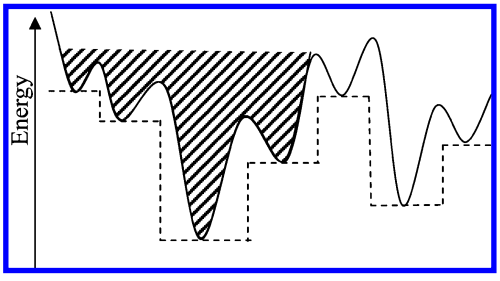


1. INTRODUCTION

Structure prediction and pathway searching are the main theme of modern computational simulation, which are essential for understanding the thermodynamics and kinetics properties of materials. While molecular dynamics (MD) has been widely utilized as a general tool for exploring the PES and simulating chemical reactions, its predictive power drops substantially for systems with large dimensionality or high barriers, such as those in protein folding or carbon nanotube growth. The traditional simulated annealing (SA) method[1] based on MD can overcome the barrier of structural transformation via elevating the system temperature and aims to find the structural minimum by subsequent quenching. However, the SA method is not efficient for identifying energy minima that are not favorable in free energy or even not present at high temperatures. An efficient theoretical approach that can explore PES at low temperatures is thus highly desirable, which must, at least, be able to surmount high barriers.

For studying systems with high kinetic barriers, one popular approach is to determine the associated transition states (TS) and the reaction pathway between the interested local minima.[2−15] This often involves intensive intuition for guessing the likely pathways (according to the initial and the final states)[4−6] and the computation of the Hessian,[9] which becomes impractical for large systems, where gradient-only methods are required.[7,8] To overcome these drawbacks, some alternative approaches have been designed, such as reduced gradient following, the anharmonic downward distortion following, and the artificial force induced reaction.[10−15] A survey of the TS-searching methods was recently reviewed by Schlegel.[2] The other class of methods can also enhance the sampling of the high barrier events by enforcing constraints (e.g., bias potential) along a predefined reaction direction,[16−19] such as metadynamics and umbrella sampling. As seen from Scheme 1, these targeted MD based methods can only visit other faraway minima (the white region) after all the nearby minima (the crossed region) have been filled by added bias potentials. This makes these techniques not ideal for a fast PES sampling over a large dimensionality, and instead they are often utilized for finding the pathway of a single reaction process where *a priori* knowledge on the constraint is known.

Scheme 1. A Schematic 1-D PES[a]



[a]The dotted lines indicate the transformed PES as treated by BH[20] and GA (Lamarckian type[26]) methods; the crossed area indicates the filling of the energy wells by added potentials as utilized in metadynamics.[16]

Received: November 17, 2012
Published: February 5, 2013

By contrast, the so-called global minimum (GM) searching methods, such as basin-hopping (BH)[20−22] and genetic algorithms (GAs),[23−26] are targeted to find the GM, the most stable structure on PES, while discarding totally the pathway information. The BH method introduced by the Wales group is particularly powerful for structure prediction in many applications, ranging from clusters to biomolecules.[20−22,27,28] These methods transform the original PES, with only the information on the minimum configuration being preserved, as depicted by the dotted lines in Scheme 1. In these methods, an aggressive (random) structural deformation is a must to escape from deep energy traps on PES. A natural consequence of the aggressive structural deformation is that it can destroy indiscriminately the bondings in the current structural configuration, including those preferred bonding motifs, and thus extra constraints are often needed as input to treat covalent bonding molecular systems.[29]

In this work, we propose a new unbiased general purpose PES searching method, namely the stochastic surface walking method (SSW), inspired by our previous work for continuous TS searching using the biased potential driven constrained Broyden dimer method (BP-CBD).[30,31] The SSW method proposed here is based on the idea of bias-potential-driven dynamics[17] and Metropolis Monte Carlo (MC),[32] which manipulates smoothly the structural configuration from one minimum to another on PES and relies on Metropolis MC to accept or refuse each move. The method can explore a complex PES with high efficiency without the need for *a priori* knowledge on the system such as the structure motif of materials. The minimum structures, both local and global, can be obtained from SSW searching trajectories, while the information on the reaction pathways connecting the minima is also preserved. We demonstrate the method in both small molecular systems and large nanostructures for structure and pathway prediction using either first principles density functional theory (DFT) or empirical potential calculations.

2. METHOD

Each step in SSW (a MC step) comprises three independent parts, namely, (i) the climbing, (ii) the relaxation, and (iii) the Metropolis MC. One such a step is called a MC step. The illustration in Figure 1 is such a step containing the climbing and the relaxation on a simplified 1D-PES from one minimum to another. The climbing procedure lies at the heart of the SSW method and is elaborated below in detail.

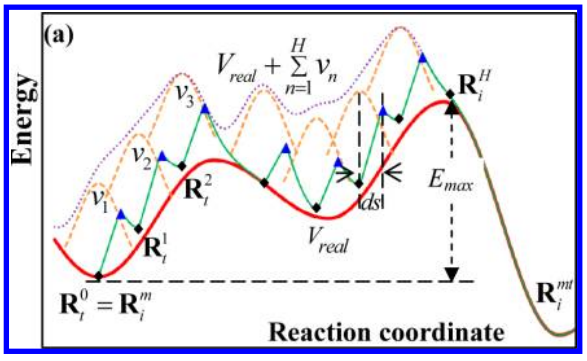


Figure 1. An illustration of the SSW method in 1D-PES. The red, orange, purple, and green curves represent the real PES (V_{real}), the Gaussian functions (v_n, n = 1, 2, ...H), the modified PES, and the searching trajectory, respectively.

The Climbing. The climbing procedure of the SSW is a bias-potential-driven climbing and local optimization process. In the procedure, one needs first to generate a random direction $\mathbf{N}_i^0$ at the current minimum $\mathbf{R}_i^m$, where "i" is the index of the current MC step. Next, according to $\mathbf{N}_i^0$, $\mathbf{R}_i^m$ is gradually dragged to a high energy configuration $\mathbf{R}_i^H$.

To enable an unbiased exploration of PES, the initial direction $\mathbf{N}_i^0$ is a randomly generated function of the ionic coordinates. In the SSW method, we generate $\mathbf{N}_i^0$ by combining two randomly generated direction $\mathbf{N}_i^g$ and $\mathbf{N}_i^l$, as eq 1.

$$\mathbf{N}_i^0 = (\mathbf{N}_i^g + \lambda \mathbf{N}_i^l)/|\mathbf{N}_i^g + \lambda \mathbf{N}_i^l| \tag{1}$$

where the mixing parameter λ controls the relative portion of these two displacement directions. In this work, λ is a random value from 0.1 to 1.5 in each step. Specifically, we set $\mathbf{N}_i^g$ as a randomly generated normalized vector, the distribution of which satisfies the Maxwell−Boltzmann velocity distribution at 300 K, as utilized in the standard molecular dynamics method for generating initial random velocity. As $\mathbf{N}_i^g$ distributes over a group of atoms, it mainly consists of soft global move. On the other hand, $\mathbf{N}_i^l$ describes the stiff local move by design, and in our implementation it was set as a bond formation mode between two non-neighboring atoms as utilized previously.[31] For example, a bond formation mode $\mathbf{N}_i^l$ associated with the atom A (e.g., the first atom in system) and the atom B (the second atom in system) can be derived as eq 2, using their coordinates q_A and q_B. Unlike the previous BP-CBD method where a known atomic pair is provided to indicate the desired reaction direction, in the SSW method the atom pair is chosen randomly provided that the two atoms are not in close contact (i.e., their distance > 3 Å).

$$\mathbf{N}_i^l = \begin{pmatrix} q_B \\ q_A \\ 0 \\ \vdots \end{pmatrix} - \begin{pmatrix} q_A \\ q_B \\ 0 \\ \vdots \end{pmatrix} \tag{2}$$

Once $\mathbf{N}_i^0$ is ready, a biased dimer rotation method is utilized to refine this randomly generated mode. The biased dimer rotation method has been utilized previously for identifying the TS of reaction in BP-CBD method.[30,31] It was found that the biased rotation algorithm is essential to retaining the key message from $\mathbf{N}_i^0$ and can also take into account the other useful information from the real PES. The details of the method can be found in the BP-CBD method,[31] and here it is only briefly introduced, as eqs 3−6.

$$\mathbf{R_1} = \mathbf{R_0} + \mathbf{N}_t \cdot \Delta R \tag{3}$$

$$C = \frac{(\mathbf{F}_0 - \mathbf{F}_1) \bullet \mathbf{N}_t}{\Delta R} \tag{4}$$

$$V_{R1} = V_{real} + V_N \tag{5}$$

$$V_N = -\frac{a}{2} \cdot [(\mathbf{R_1} - \mathbf{R_0}) \bullet \mathbf{N}_i^0]^2 = -\frac{a}{2} \cdot (\Delta R \cdot \mathbf{N}_t \bullet \mathbf{N}_i^0)^2 \tag{6}$$

In the dimer rotation,[33] one needs to define two images separated by a fixed distance of ΔR (e.g., ΔR = 0.005 Å) on PES, namely $\mathbf{R_0}$ and $\mathbf{R_1}$ (eq 3). The rotation of the dimer vector $\mathbf{N}$ according to their force ($\mathbf{F_0}$ and $\mathbf{F_1}$) will provide the local curvature information (C in eq 4). The dimer rotation is a numerical way to identify one eigenvector of the Hessian.[7,8,33] For the biased rotation proposed by us previously,[31] the

potential of $\mathbf{R_1}$ is modified as eq 5, where V_N is the bias potential added to the real PES V_{real} of $\mathbf{R_1}$ that is a quadratic function of coordinates $\mathbf{R_1}$ along $\mathbf{N}_i^0$ (eq 6). As long as the parameter a (eq 6) is large enough, the biased rotation can guarantee the rotation of the dimer will not deviate far away from $\mathbf{N}_i^0$. The force due to the bias potential can be evaluated straightforwardly[31] for constraining the dimer rotation.

On moving from $\mathbf{R}_i^m$ to a high energy configuration $\mathbf{R}_i^H$, a modified PES $V_{\text{m-to-H}}$ is utilized, as shown in eq 7, in which a series of bias Gaussian potential v_n (n is the index of the bias potential, n = 1, 2, ...H) is added one by one consecutively along the direction $\mathbf{N}_i^n$.

$$\begin{aligned} V_{m\text{-to-}H} &= V_{\text{real}} + \sum_{n=1}^{H} v_n \\ &= V_{\text{real}} + \sum_{n=1}^{H} w_n \times \exp[-((\mathbf{R}^t - \mathbf{R}_t^{n-1})\cdot\mathbf{N}_i^n)^2 \\ &\quad /(2 \times ds^2)] \end{aligned} \tag{7}$$

$$\mathbf{F_{tot}} = \mathbf{F_{real}} + \sum_n w_n \cdot \exp\left[-\frac{((\mathbf{R}^t - \mathbf{R}_t^{n-1})\cdot\mathbf{N}_i^n)^2}{2 \times ds^2}\right] \cdot \frac{(\mathbf{R}^t - \mathbf{R}_t^{n-1})\cdot N_i^n}{ds^2}\cdot\mathbf{N}_i^n \tag{8}$$

$\mathbf{N}_i^n$ is always updated from the initial random direction $\mathbf{N}_i^0$ using the biased dimer rotation method mentioned above. In eq 7, the w and the ds are the height and the width of Gaussian function v_n; $\mathbf{R}_t^n$ represents the nth local minima along the movement trajectory on the modified PES that is created after adding n Gaussian functions. The PES at $\mathbf{R}_t^n$ is defined by $V_{m\text{-to-}n} = V_{\text{real}} + \sum_{k=1}^{n} v_k$.

To summarize, the movement from $\mathbf{R}_i^m$ to $\mathbf{R}_i^H$ is a repeated procedure containing (i) updating the direction $\mathbf{N}_i^n$ at $\mathbf{R}_t^{n-1}$, (ii) adding a new Gaussian function v_n and displacing $\mathbf{R}_t^{n-1}$ along the direction $\mathbf{N}_i^n$ by a magnitude of ds ($\mathbf{R}_t^{n-1} + \mathbf{N}_i^n \cdot ds$), and (iii) local relaxing to $\mathbf{R}_t^n$ on the modified PES (the energy minimization is constrained by the added potentials). The force for the local optimization on the modified PES can be evaluated according to eq 8. The other detail on the climbing process has been described in the biased translation algorithm in the BP-CBD method,[31] which is aimed for moving a configuration from IS to a TS region.

The Gaussian width ds is a critical parameter controlling the step length of surface walking. A typical value ranges from 0.2 to 0.6 Å, being 10% to 40% of the bond length of chemical bonding. With a large ds, a large scope of PES can be explored rapidly, but it is at the expense of the resolution on the reaction pathway between minima. The max number of Gaussian potentials (H) is also a system-dependent parameter. In this work, unless specifically mentioned, we set it as 14, which is equivalent to an overall displacement of 4−5 Å per MC step.

Obviously, the overall efficiency of the SSW method depends on the choice of ds and H. As for the SSW method applying to a complex PES (e.g., with multiple funnels and high barrier) with H = 14, typically 70∼80% of computational efforts (energy and gradient evaluation) are spent on the climbing process because of the repeated procedure in adding bias potential and relaxing, and the remaining 20∼30% is in the relaxation step to the minimum. As we will show in the examples (section 3), the smooth perturbation in the climbing, although it is computationally intensive, helps to increase greatly the probability for visiting low energy structures, and thus the method overall is efficient for complex systems.

Overall Algorithm. We have illustrated the whole procedure of the SSW method in Scheme 2 and the algorithm can be described as follows:

Step 1: Generate an initial random direction $\mathbf{N}_i^0$ at the current minimum $\mathbf{R}_i^m$. Set n = 1 and $\mathbf{R}_t^0 = \mathbf{R}_i^m$.

Step 2: Refine $\mathbf{N}_i^0$ to $\mathbf{N}_i^n$ by using biased dimer rotation method.

Step 3: Add the nth Gaussian function to the PES and displace $\mathbf{R}_t^{n-1}$ by a magnitude of ds along $\mathbf{N}_i^n$.

Step 4: Relax locally the structure to a new minimum $\mathbf{R}_t^n$ on the modified PES and make $n = n + 1$. The limited memory BFGS (l-BFGS) method[34] is utilized for the structural optimization.

Step 5: Stop the climbing procedure if (i) the number of the Gaussian, "n," is larger than the maximum Gaussian number (H) or (ii) the energy of the local minimum $\mathbf{R}_t^n$ is lower than the energy of the initial $\mathbf{R}_i^m$; otherwise, repeat steps 2−4.

Step 6: Remove all the added bias potentials and optimize the current structure to a new minimum $\mathbf{R}_i^{mt}$ on the real PES. The l-BFGS method[34] is utilized for the structural optimization.

Step 7: According to the Metropolis Monte Carlo scheme, the probability P of accepting the new configuration $\mathbf{R}_i^{mt}$ by making $\mathbf{R}_{i+1}^m = \mathbf{R}_i^{mt}$ is

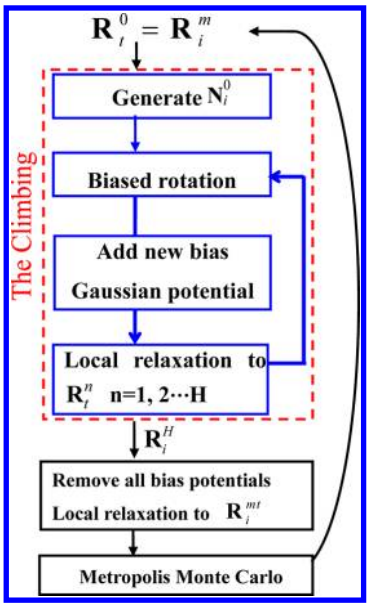
$$P = \begin{cases} \exp[E(\mathbf{R}_i^m) - E(\mathbf{R}_i^{mt}) /RT], & \text{when } E(\mathbf{R}_i^{mt}) > E(\mathbf{R}_i^m) \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

where E(X) is the energy of the structure X.

Step 8: Make $i = i + 1$ and repeat the steps 1−7.

It should be noticed that all saddle points (TSs) are no longer located explicitly in the SSW method. This is because (i) by using reduced ds, the barrier of the pathway can be estimated reasonably well without knowing the TS (see examples below) and (ii) TSs can always be located exactly afterward when

Scheme 2. The Overall Algorithm of the SSW Method



dx.doi.org/10.1021/ct301010b | J. Chem. Theory Comput. 2013, 9, 1838−1845

reaction pathways are known. As only lowest energy pathways are vital kinetically, to determine TSs for all reaction channels is tedious and unnecessary.

The current SSW method has some similarities to the previous methods using TS searching methods (e.g., eigenvector-following[35,36]) to explore the PES in the sense that the pathways linking the minima are present in the searching trajectory. However, the SSW method does not locate TS explicitly. By adding bias potentials consecutively, the PES has been coarse-grained, and the minimum reached in one MC step of SSW searching can be several minima away from the initial minimum. By applying to various examples, we will show that the SSW method is efficient for a range of complex PES systems, including covalent molecules and model clusters such as Lennard-Jones clusters (demonstrated previously in the discrete path sampling framework[37,38]) and short-ranged Morse clusters. The success of the SSW method for exploring PES rapidly lies with two aspects: (i) the SSW saves the computation efforts for not locating the TS exactly, as addressed above; (ii) the SSW can bypass effectively the high energy minima without visiting them explicitly. This is found to be important for finding the GM in a complex PES, e.g., those of double funnels (see below section 3.2 for detail).

Two important features of the SSW method should be pointed out. First, *it can provide directly useful information on kinetics*, including the barrier height and reaction pathway. From the searching trajectory, a series of structures ($\mathbf{R}_t^n$) along the pathway can be obtained (see Figure 1), and the highest energy point (E_{max}, as labeled in Figure 1) is a good indication for the barrier height of the reaction provided that the *ds* of Gaussian is not too large. Second, *the probability for one* reaction to *occur depends no longer exponentially on its barrier*. In the SSW method, whether a reaction occurs depends only on whether its associated normal mode is selected by the random mode generation procedure.

3. RESULTS AND DISCUSSION

3.1. C_4H_6 Conformation. To illustrate the SSW method, we first apply the method to a simple molecular system, C_4H_6 isomers, where first principles DFT calculations at the GGA-PBE level[39] are utilized to obtain the energy and the force. A single run of 90 steps was performed starting from *trans*-butadiene, the most stable isomer.

In total, we identify four isomers in 90 steps, being *trans*- and *cis*-butadiene, cyclobutene, and a high energy radical. The trajectory connecting the lowest three minima is plotted in Figure 2. There are 61 times for trans-to-cis butadiene chain rotation (green curves). The second most common conversion (five times) is the ring-closing reaction from *cis*-butadienes to cyclobutene. The trajectories to form the high energy radical (three times) are only of minor contribution. It is clear that while the low barrier chain rotation still dominates the isomer conversion, the sampling of the high barrier ring-closing reaction is now much more efficient. It is noticed that the lowest E_{max} for cis-to-trans transformation is 0.36 eV and that for cis/trans-to-cyclo transformation is 1.75 eV (shown by the red curve) from the searching trajectory, which are in good agreement with those determined exactly (0.33 and 1.56 eV) by TS location. Within a short run, one can identify all three lowest energy isomers and also estimate the barriers leading to them.

It might be mentioned that there are many other isomers for C_4H_6 that are however much less stable than the three

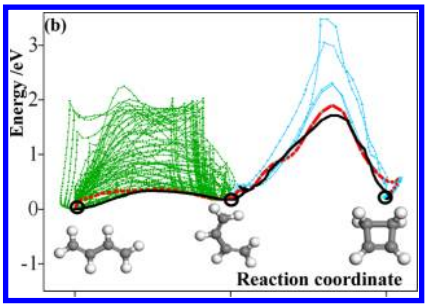


Figure 2. The PES exploration of C_4H_6 isomers. The temperature utilized in Metropolis MC is 1000 K. The *ds* for Gaussian is 0.1 Å, and the max number of Gaussian potentials is 25. The *x* axis is the reaction coordinate along the structural vector defined by the three lowest minima. The lowest energy curve from the searching trajectories is highlighted by the red color. For comparison, the lowest energy pathway is shown as the black curve that is determined by the steepest-descent optimization starting from the located TSs toward nearby minima.

presented in Figure 2. By increasing *ds* to 0.4 Å and the total MC step to 150, we performed a second run and then identified five other isomers, including buta-1,2-diene, methylenecyclopropane, bicyclo[1.1.0]butane, 3-methylcycloprop-1-ene, and 1-methylcycloprop-1-ene. Each of them appears only once out of 150 MC steps, indicating that a high barrier is present to generate these unstable isomers.

3.2. Lennard-Jones 75 Cluster. The Lennard-Jones cluster with 75 atoms (LJ_{75}), the PES of which is known to have double funnels,[20,40] was then utilized to test the ability of our method for surmounting high barrier events while identifying the low energy structures. The potential is described by

$$V_{LJ} = 4\epsilon \sum_{i<j} [(\sigma/r_{ij})^{12} - (\sigma/r_{ij})^{6}] \tag{10}$$

where ϵ and σ are the pair equilibrium well depth and separation, respectively (we set $\epsilon = 1$ and $\sigma = 2.7$). The identification of the GM of LJ_{75} is hard because the funnel associated with GM (a Marks' decahedron) is much narrower than that of the second lowest minimum (SLM, a Mackey icosahedron), and both the high barrier ($\sim 10\ \epsilon$) and the large spatial gap ($\sim 100\ \sigma$) are present between SLM and GM.[37] By starting from the SLM, the trajectories for converting SLM to GM were collected, and the results are shown in Figure 3, which are averaged over 100 independent runs, each with 10^6 steps. The MC temperature is 0.8 and the Gaussian *ds* is 0.6. A minimum is reached when the maximal force on each coordinate is less than $0.04\ \epsilon/\sigma$. We found that 1.10×10^5 steps are required to achieve a 50% probability for finding the GM (0.53×10^5 on average) and 7.43×10^5 steps for ~100% (1.69×10^5 on average). On average, each MC step requires evaluation 378 times for the energy and gradient, in which 78% is spent in the climbing and 22% for relaxing to the minimum. The large number of the averaged simulation steps is mainly due to the repeated SLM-to-SLM trajectories, although the shortest trajectory for the SLM-to-GM conversion contains as low as 23 MC steps. It is intriguing how the SSW method surmounts the high barrier and walks through the large structural gap in between SLM and GM.

To answer this, we selected one short trajectory (23 minima) from SLM to GM, as shown by the red curve in Figure 3b, and compared it with the complete pathway (the blue curve) with

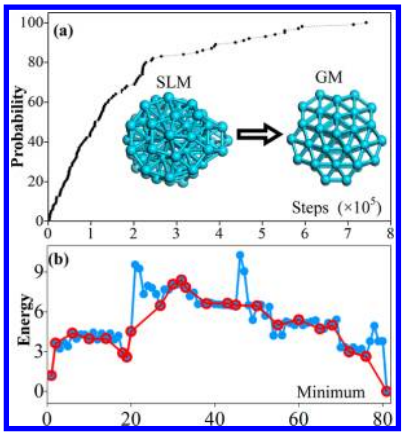


Figure 3. (a) The probability for SLM to GM conversion in LJ_{75}. (b) One searching trajectory from SLM to GM (red curve). All the involved minima along the trajectory are located and also shown in the blue curve.

all the TSs and minima on PES along the trajectory being determined. In the complete pathway, there are in fact 81 minima, significantly more than those identified in the trajectory. The overall probability of conversion $[P_{(SLM\rightarrow GM)}]$ can then be estimated according to Metropolis MC. It turns out that $P_{SLM\rightarrow GM}$ following the red curve is 10 orders of magnitude larger than that of the blue curve. Apparently, this is because the current method has the ability to bypass a number of (high) energy minima and reduce significantly the times for Metropolis MC evaluation.

It should be noticed that the trajectory in Figure 3b is only one possible pathway linking SLM with GM, which should be distinguished from the kinetically fastest pathway or the lowest energy pathway that have already been characterized by Wales et al.[37,38] In the current case, we are aiming to test whether the SSW can overcome the high barrier between SLM and GM, and thus no attempts have been made to sample all important trajectories (100 trajectories from Figure 3a is too limited in the 225 (= 3 × 75) dimensional space) to evaluate the overall conversion rate, which will be far more time-consuming and has already been done by Wales et al. using a discrete path sampling method based on the information on all minima and the TSs linking them. Nevertheless, the trajectory in Figure 3b shows a similar highest energy position (~10 ϵ) along the pathway to that in the lowest energy pathway.[37,38]

3.3. Structure Prediction of Morse Clusters. One potential application of the SSW method is to search for the GM of materials, in particular those with a complex PES. For this purpose, we applied the method for searching the GM of the short-ranged Morse potential clusters, as described by eq 11 with $\rho_0 = 14$ (we set $\epsilon = 1$ and $r_0 = 2.7$). This interatomic potential mimics strong local bonding such as those encountered in carbon materials. The MC temperature is 0.8, and the Gaussian ds is 0.6. A minimum is reached when the maximal force on each coordinate is less than 0.04 ϵ/σ.

$$V_M = \epsilon \sum_{i<j} e^{\rho_0(1-r_{ij}/r_0)}(e^{\rho_0(1-r_{ij}/r_0)} - 2) \quad (11)$$

The GM problem of these Morse clusters has been studied previously, and due to the great complexity of the PES, the unbiased searching of all GMs was not achieved even for small clusters with atom number $N < 80$.[41] Recent studies showed that by including structural motif information, all the minima can be identified; see for example, Doye et al.[42] and Cheng and Yang.[43] In this work, we studied the Morse clusters (N = 5 to 103) and found that all GMs can be identified. For each cluster size, we conducted 100 independent searches, each starting from a random initial structure. The probability for finding GM was calculated and is summarized in Figure 4a. It shows that the

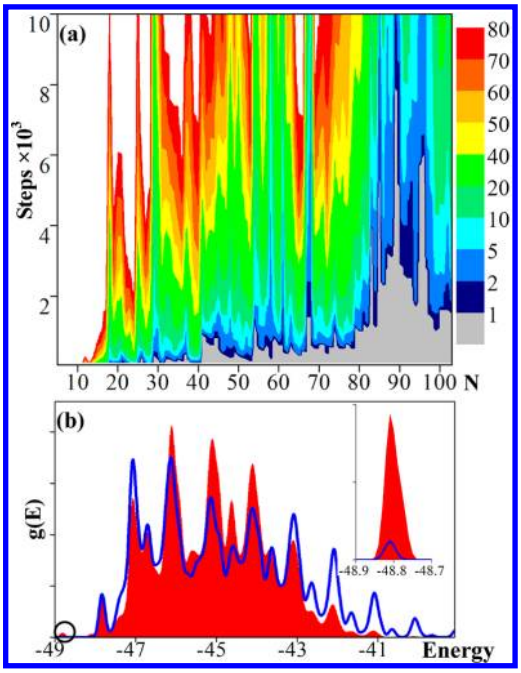


Figure 4. (a) The probability (%) of the trajectories for finding the GM of Morse clusters using the SSW method. (b) The probability density ($g(E)$) for visiting minima with two different methods for the move step. Red: using the SSW method. Blue: using large steps of displacement (see text).

probability drops with the increase of the cluster size in general. The 29-atom cluster (M_{29}) is the first challenging case, in which the probability to find GM is 25% within 10 000 MC steps. The clusters with more than 80 atoms are considerably more difficult: for M_{89}, only one trajectory out of 100 finds the GM within 10 000 steps. On average, the energy and gradient are evaluated 443 ± 20 times in each MC step for the systems with more than 30 atoms, 80% of which is in the climbing and 20% in relaxing to the minimum. We noted that the standard BH methods, although efficient for finding the GM of Lennard-Jones clusters, meet a great difficulty for large short-ranged Morse clusters;[41] for example, for M_{80}, the GM cannot be found within 10 000 steps in 100 trajectories using the standard BH method, while in the SSW the GM can be reached within 2000 steps (8 out of 100 trajectories).

The overall efficiency of the SSW method for finding the GM of M_{80} starting from 100 random structures is reported in Table 1, which has been compared with three typical LJ clusters. Interestingly, we notice that it is easier to find the GM of M_{80} compared to that of LJ_{75} using the SSW method, which indicates that the SSW method is particularly useful for the corrugated PES as presented by Morse clusters. The LJ_{75} cluster is more difficult with the criterion of achieving a high success ratio of finding GM because the GM locates at the narrower funnel, and a huge structural separation and a high barrier are present between two funnels, as addressed above.

Table 1. Typical Efficiency of SSW Method for the Unbiased Searching of GM of Model Systems Starting from Random Structures

		ave. total MC steps[a]			
system	E_{GM}/ϵ	100% suc. rat.	50% suc. rat.	ave. total force calc. times per MC[b]	runs
LJ_{55}	−297.24847	76.5	38.2	364	1000
LJ_{38}	−173.92843	1.71×10^3	5.03×10^2	376	1000
LJ_{75}	−397.49233	1.07×10^5	2.29×10^4	396	100
M_{80}	−340.81137	4.74×10^4	5.84×10^3	460	100

[a]The averaged total MC steps to achieve a 100% or 50% success ratio. [b]The averaged total the energy and gradient evaluation times in one MC step.

Quantitatively, the efficiency of a theoretical method for finding GM can be measured by calculating the probability for visiting a minimum (at energy E_j), $P(E_j)$. In Figure 4b, we showed the calculated probability density, $g(E)$ ($P(E) = g(E)\,dE$), of the M_{16} cluster based on the Wang−Landau (WL) flat histogram algorithm.[44] In the WL algorithm, the density of states (DOS) is generated by performing a set of random walks in energy space with a probability proportional to the reciprocal of the density of states. By modifying the estimated DOSs in a systematic way, a flat histogram over the allowed range of energy can be produced, and the DOSs will simultaneously converge to the true value. In our simulation, we focus on the first 700 minima which are visited more than 10 times in the first WL iteration. Six WL iterations are performed to converge the DOS. Figure 4b compares the DOSs of $g(E)$ using two different methods for generating new structures in the move step. The first is the SSW method that utilizes small but continuous structural perturbations. The second method applies one large random displacement along each coordinate with the maximum step size being 40% of the equilibrium distance, as typically utilized in the BH method. We found that the SSW method has a lower population at the high energy states but a higher population at low energy states compared to that with the large displacement. Specifically, GM is 10 times more frequently visited in our method (Figure 4b insert). It indicates that the small but continuous structure perturbation can bias better toward lower energy structures and consequently increase the probability for finding GM.

We emphasize that the PES of the above Morse potential clusters is among the highest complexity owing to its highly local bonding. The SSW method is thus expected to be able to predict the structure of nanomaterial in general. Indeed, we have examined other systems, including model potential systems such as Lennard-Jones clusters up to $N = 105$, and carbon clusters including C_{60} and C_{70} fullerene cages (with Brenner potential[45] to assemble randomly distributed individual carbon atoms into the fullerene cages). The typical efficiency for finding the GM of LJ clusters (LJ_{55}, LJ_{38}, and LJ_{75}) starting from random structures is reported in Table 1. The animation depicting the formation trajectory of C_{70} starting from a random structure is provided in the Supporting Information. In all cases, the correct global minimum of the systems can be identified within 10^4 MC steps (at least 1% success ratio) starting from a random structure.

3.4. Pathway for Nanostructure Deformation. Because the SSW trajectory follows the landscape of PES, it can naturally be used to identify the most likely reaction channels and predict the product distribution for complex systems. Here, we demonstrate our methods for the phase transformation of a Boerdijk−Coxeter−Bernal (BCB) nanohelix,[46] which can be considered as 17 interlinked tetrahedron clusters, containing 74 particles and 222 degrees of freedom in total, as shown Figure 5, top panel. The BCB helix has been commonly observed for

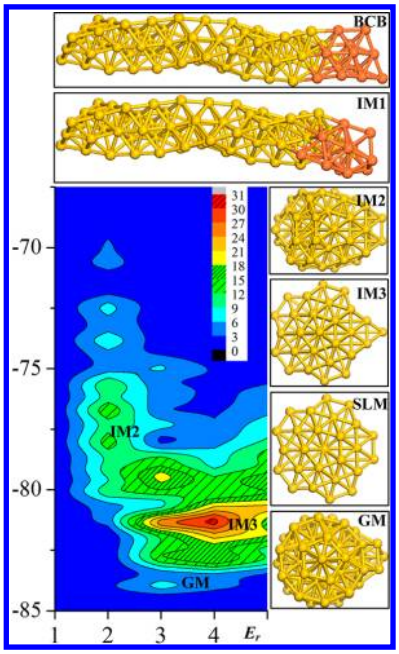


Figure 5. The contour plot (bottom-left) for the density of residence time (%) of the minimum together with snapshots for the structural evolution of the BCB helix to a cluster. In the contour plot, the x axis is E_r and the y axis is the energy with respect to BCB.

the self-assembly of semiconducting nanoparticles or metal atoms into nanowires under specific synthetic conditions.[47,48] A special ligand or solution environment is generally required to stabilize the helix pattern. The interaction potential between particles utilized is the Lennard-Jones potential, as described by eq 10.

To understand the stability of the helix and the kinetics of phase transition, we have run five sets of simulations with varied preset E_r (=1.0, 2.0, 3.0, 4.0, 5.0). E_r serves as a maximum barrier threshold for the reaction to occur, which disallows E_{max} of a move step to be higher than E_r. This can be achieved by stopping the climbing procedure when the real energy of $\mathbf{R}_i^H$ in a trajectory is higher than E_r and starting the energy minimization from $\mathbf{R}_i^{H-1}$ to obtain $\mathbf{R}_i^{mt}$. The purpose of setting a series of E_r is to identify the kinetics sequence of elementary steps. In this case, the pathway is the concern, and thus the new structure $\mathbf{R}_i^{mt}$ is always accepted to obtain continuous trajectories. We run 100 trajectories at each fixed E_r. Each trajectory ends either when no lower energy minimum is found within 2×10^5 steps or the same minimum is trapped for more than 10^5 steps. The residence time at a minimum is calculated as the times for the minimum being visited divided by the total steps. A long residence time indicates the presence of a large basin around this minimum, where a high barrier close to E_r is required in order to leave the trap. The contour plot for the density of the residence time against the potential energy and E_r is plotted in Figure 5 (bottom-left) together with the snapshots for the structural evolution of BCB to a cluster.

Our simulation shows that although the LJ-BCB_{74} structure is thermodynamically very unstable compared to the aggregated clusters (exothermic by 83.97), it is a kinetically trapped state: to destroy the helix structure, an E_r of about 1.0 is required as only 6 out of 100 trajectories can leave the BCB basin with E_r = 1.0. For E_r from 2.0 to 5.0, the BCB structure can be destroyed easily, typically within nine steps. From the trajectories with E_r = 1.0, the minimum (**IM1**) next to the BCB helix with lower energy is as shown in Figure 5, where the ending part of BCB containing three tetrahedra collapses first. Using the trajectories obtained, it is facile to identify exactly the pathway for the initial collapse of the BCB helix by using the BP-CBD method.[31] Our calculation shows that the barrier to destroy the BCB helix is 1.037 ϵ (Figure 6). In the process, the minor groove at the end

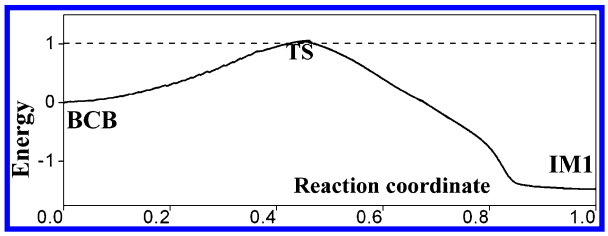


Figure 6. The energy profile for the initial collapse of the BCB helix (from BCB to IM1 in Figure 5). The pathway is obtained from the searching trajectories and refined by the BP-CBD method that determines the TS exactly. The reaction coordinate is along the BCB-IM1 vector by setting the BCB coordinate to 0 and IM1 to 1.

of BCB (highlighted as red atoms) is occupied by new particles, and the sharp corner disappears. It is indicated that the least coordinated particle drives the structure deformation, and the ending tetrahedra are the weakest link in the BCB helix.

By tracing the trapped intermediate states that have a large residence time, the simulation with varied E_r can provide rich information on kinetics. Upon increasing E_r to 2.0, a significant number of minima are found along the pathways from BCB to GM, implying that an energy height of 2.0 is about the barrier of the elementary reactions to clustering and an E_r value in between 1.0 to 2.0 should be used to trap the intermediate structures for the BCB collapse. At E_r = 2.0, the intermediate (**IM2**) at the energy of −78.06 (with respect to BCB) is the major product, being an incomplete icosahedron, structurally similar to GM (Figure 5). The major product (**IM3**) shifts downward in energy to −81.32 when E_r is 3.0 and 4.0, and it is an incomplete decahedron, structurally similar to SLM (Figure 5). Further increasing E_{max}, the simulation becomes less likely to be trapped in any of the minimum as the barrier separating the minima can now be overcome. As a result, above E_r = 4.0, all trajectories can converge to GM. From our analyses, for the BCB to cluster transformation, the SLM is the kinetically preferred product, while the GM is the thermodynamically preferred and can be reached above E_r = 4.0.

4. CONCLUSION

To recap, this work describes an automated and efficient method for the PES exploration of complex systems with high barriers. We demonstrate our method in a number of problems, including the search for minima of nanostructures and molecules, both local and global, and the identification of the reaction pathways leading to them. It should be emphasized that the SSW method requires only the coordinates, the energy, and its first derivative (forces) as the input and thus can be applied generally to any PES, including both *ab initio* and empirical potential. Due to the small structural perturbation per step (*ds* in each Gaussian), the method is particularly suitable for first principles electronic structure calculations. Within the same framework, the current SSW method and the BP-CBD method introduced by us previously can be used together to tackle a number of important issues related to PES, including the stable/metastable structure prediction, the TS, and the pathway identification.

Not limited to small molecules and the model cluster systems presented here, our ongoing work finds that the SSW method is transferable to treating periodic bulk and surface systems without extra coding, the PESs of which are in fact often simpler due to the confinement. As an unbiased searching method starting from one single structure, it should be possible to further improve the SSW method for particular applications by, for example, incorporating extra structural motif information (e.g., symmetry constraint, biased initial mode direction), performing parallel replica exchange, and combining with other global minimization approaches (e.g., GA method). These will be the future directions to explore. We hope that this new method can expand greatly the scope of computational simulation toward the rational design of new materials.

■ ASSOCIATED CONTENT

Supporting Information

The animation depicting the formation trajectory of C_{70} starting from a random structure is provided. This information 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

We acknowledge National Science foundation of China (20825311, 21173051), 973 program (2011CB808500, 2013CB834603), Science and Technology Commission of Shanghai Municipality (08DZ2270500) for financial support.

■ REFERENCES

(1) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(2) Schlegel, H. B. *WIREs Comput. Mol. Sci.* **2011**, *1*, 790.
(3) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385.
(4) Wang, H. F.; Liu, Z. P. *J. Am. Chem. Soc.* **2008**, *130*, 10996.
(5) Fang, Y. H.; Liu, Z. P. *J. Am. Chem. Soc.* **2010**, *132*, 18214.
(6) Shang, C.; Liu, Z. P. *J. Am. Chem. Soc.* **2011**, *133*, 9938.
(7) Kumeda, Y.; Wales, D. J.; Munro, L. J. *Chem. Phys. Lett.* **2001**, *341*, 185.
(8) Munro, L. J.; Wales, D. J. *Phys. Rev. B: Condens. Matter* **1999**, *59*, 3969.
(9) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800.
(10) Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. *J. Comput. Chem.* **1998**, *19*, 1087.
(11) Bofill, J. M.; Anglada, J. M. *Theor. Chim. Acta* **2001**, *105*, 463.
(12) Ohno, K.; Maeda, S. *Chem. Phys. Lett.* **2004**, *384*, 277.
(13) Hirsch, M.; Quapp, W. *J. Mol. Struct.: THEOCHEM* **2004**, *683*, 1.
(14) Maeda, S.; Morokuma, K. *J. Chem. Theory Comput.* **2011**, *7*, 2335.

(15) Crehuet, R.; Bofill, J. M.; Anglada, J. M. *Theor. Chim. Acta* **2002**, *107*, 130.
(16) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
(17) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
(18) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(19) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737.
(20) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111.
(21) Doye, J. P. K.; Wales, D. J. *Phys. Rev. Lett.* **1998**, *80*, 1357.
(22) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368.
(23) Woodley, S.; Battle, P.; Gale, J.; Richard, A.; Catlow, R. A. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2535.
(24) Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288.
(25) Oganov, A. R.; Glass, C. W. *J. Chem. Phys.* **2006**, *124*, 244704.
(26) Turner, G. W.; Tedesco, E.; Harris, K. D. M.; Johnston, R. L.; Kariuki, B. M. *Chem. Phys. Lett.* **2000**, *321*, 183.
(27) Doye, J. P. K.; Wales, D. J. *Science* **1996**, *271*, 484.
(28) Doye, J. P. K.; Wales, D. J.; Branz, W.; Calvo, F. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *64*, 235409.
(29) Kusumaatmaja, H.; Whittleston, C. S.; Wales, D. J. *J. Chem. Theory Comput.* **2012**, *8*, 5159.
(30) Shang, C.; Liu, Z. P. *J. Chem. Theory Comput.* **2010**, *6*, 1136.
(31) Shang, C.; Liu, Z. P. *J. Chem. Theory Comput.* **2012**, *8*, 2215.
(32) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(33) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 7010.
(34) Liu, D. C.; Nocedal, J. *Math. Program.* **1989**, *45*, 503.
(35) Davis, H. L.; Wales, D. J.; Berry, R. S. *J. Chem. Phys.* **1990**, *92*, 4308.
(36) Doye, J. P. K.; Wales, D. J. *Z. Phys. D: At., Mol. Clusters* **1997**, *40*, 194.
(37) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *111*, 8417.
(38) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891.
(39) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(40) Bogdan, T. V.; Wales, D. J.; Calvo, F. *J. Chem. Phys.* **2006**, *124*, 044102.
(41) Wales, D. J.; Dullweber, A.; Hodges, M. P.; Naumkin, F. Y.; Calvo, F.; Hernández-Rojas, J.; Middleton, T. F. *The Cambridge Cluster Database*. http://www-wales.ch.cam.ac.uk/CCD.html (accessed January 06, 2013)
(42) Doye, J. P. K.; Leary, R. H.; Locatelli, M.; Schoen, F. *Informs J. Comput.* **2004**, *16*, 371.
(43) Cheng, L. J.; Yang, J. L. *J. Phys. Chem. A* **2007**, *111*, 5287.
(44) Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050.
(45) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2002**, *14*, 783.
(46) Boerdijk, A. H. *Philips Res. Rep.* **1952**, *7*, 303.
(47) Chen, Q.; Whitmer, J. K.; Jiang, S.; Bae, S. C.; Luijten, E.; Granick, S. *Science* **2011**, *331*, 199.
(48) Wang, Y.; Wang, Q.; Sun, H.; Zhang, W.; Chen, G.; Wang, Y.; Shen, X.; Han, Y.; Lu, X.; Chen, H. *J. Am. Chem. Soc.* **2011**, *133*, 20060.