JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

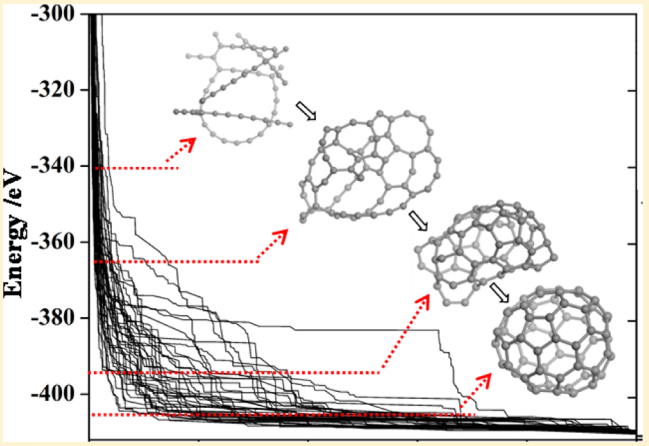
From Atoms to Fullerene: Stochastic Surface Walking Solution for Automated Structure Prediction of Complex Material

Xiao-Jie Zhang, 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: It is of general concern whether the automated structure prediction of unknown material without recourse to any knowledge from experiment is ever possible considering the daunting complexity of potential energy surface (PES) of material. Here we demonstrate that the stochastic surface walking (SSW) method can be a general and promising solution to this ultimate goal, which is applied to assemble carbon fullerenes containing up to 100 atoms (including 60, 70, 76, 78, 80, 84, 90, 96, and 100 atoms) from randomly distributed atoms, a long-standing challenge in global optimization. Combining the SSW method with a parallel replica exchange algorithm, we can locate the global minima (GM) of these large fullerenes efficiently without being trapped in numerous energy-nearly degenerate isomers. Detailed analyses on the SSW trajectories allow us to rationalize how and why the SSW method is able to explore the highly complex PES, which highlights the abilities of SSW method for surmounting the high barrier and the preference of SSW trajectories to the low energy pathways. The work demonstrates that the parallel SSW method is a practical tool for predicting unknown materials.



1. INTRODUCTION

One ultimate goal of computational simulation is to design useful materials that are not yet discovered in experiment and the structure prediction is an essential step for revealing the chemical and physical properties of the new material. Unfortunately, due to the complexity of the potential energy surface (PES) of material, the structure prediction from scratch, i.e. with only limited information such as the number of atoms and the chemical formula, remains highly challenging and the experimental trial-and-error approach is still the dominant theme for material discovery. The difficulty in theory can be exemplified using covalently bonded carbon materials that have complex PES with three representative features, as illustrated in Figure 1a, namely, the high barrier of transformation, the limited/specific reaction channels between isomers (due to the local and directional bonding), and the large dimensionality of degrees of freedom. The C_{60} fullerene was first synthesized in the 1980s,[1] and after that, it became a testing ground for new theoretical methods that aim to predict the GM of unknown materials.[2−7] To date, the unbiased search of large carbon structures (>60 atoms) remains to be a grand challenge for theory[7] and new general-purpose global optimization methods are highly desirable for solving such complex systems to facilitate large-scale computational material design.

Soon after the C_{60} fullerene structure was reported, the theoretical method using simulated annealing (SA) molecular dynamics (MD) has been tested to identify the fullerene structure from random atoms. The SA method[8] mimics the material evolution in nature to find the low energy structures via repeated temperature elevation and subsequent quenching in MD. However, the complexity of the carbon−carbon potential disallows a satisfactory sampling of the PES using the SA method alone, and only with additional constraints (e.g., with a spherical shape) can the result be reasonable for GM locating.[9] An obvious deficiency of SA is the difficulty to overcome the high barrier (E_a in Figure 1a) that requires to break the strong carbon−carbon bonds: the search will inevitably be trapped into numerous metastable local minima.[10]

Obviously, an ideal algorithm for structure prediction requires at least the ability to overcome the high barrier of structural transformation and at the mean time the efficiency should not be sensitive to the temperature applied, if a temperature based criterion is ever required in selecting the low energy structures. The current so-called global optimization methods, such as the basin-hopping (BH)[11−13] and genetic algorithm (GA)[14−17] have these essential features, achieved by

Received: March 25, 2013

   dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

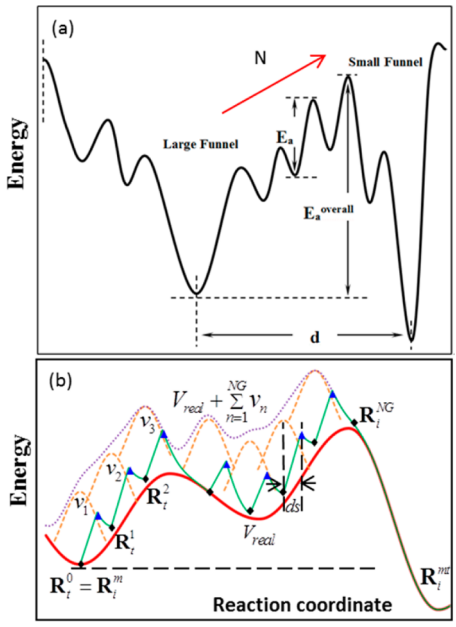


Figure 1. (a) Diagram illustrating the challenges for exploring complex PESs, where (i) a high overall barrier ($E_a^{overall}$) and a large spatial separation (d) are present between two minima; (ii) the limited reaction channels for the interconversion (along the **N** direction); and (iii) the large dimensionality as reflected by the numerous energy wells. (b) Scheme showing one SSW step from one minimum ($\mathbf{R}_0$) to another ($\mathbf{R}_i^{mt}$).

transforming the original PES via aggressive (random) structural deformation (in BH) or the structural crossover/mutation (in GA). They can be generally categorized as surface hopping methods by maximally bypassing the transition region between minima. Indeed, the first successful prediction of C_{60} fullerene from random structures was achieved by using the genetic algorithm (GA) in 1995.[15] In general, the one-go large structural change required to bypass the high-energy transition region is often detrimental for the PES exploration of systems with local bonding, where the desired reaction channels are limited. In fact, this problem arises from the lack of the pathway selection in the above-mentioned GM searching methods where the low energy pathways are not preferred in structural transformation. Consequently, cleverly designed strategies (e.g., additional constraints, special rules in crossover/mutation) are required to treat covalent bonding molecular systems to maintain the preferred bonding motifs. A detailed account of these methods can be found in literature such as ref 13. By overcoming the shortcomings of these surface hopping approaches, we recently developed a general-purpose unbiased PES searching method, namely stochastic surface walking (SSW) method, and applied this method for predicting the structure of the model systems, including Lennard-Jones clusters and short-ranged Morse clusters starting from random structures. One key feature of the SSW method is that the pathway linking the minima is present in the searching trajectory and the low energy pathways are implicitly preferred during the SSW searching.

In this work, we investigate the performance of the SSW method in large realistic systems by applying it to predict the GM of carbon clusters up to 100 atoms starting from random initial structure. In addition to their wide applications, the carbon fullerenes are also ideal testing system of global optimization since the GMs of them can be known a priori according to the isolated pentagon rule (IPR, i.e. the pentagon is surrounded by hexagons)[18] and thus serve as the natural criterion for the convergence of global optimization.

By combing SSW method with a replica exchange parallel algorithm, we assemble successfully the large fullerenes from atoms without using a priori knowledge of the systems. The results outline the key challenges facing for PES searching in real material world, in which the PES can be orders of magnitude more complex than those of Lennard-Jones or Morse model systems. Using C_{20} system, we rationalize how SSW method works in complex PES that is controlled by three SSW parameters. From our results showing that the parallel replica exchange with SSW can expedite the PES searching significantly, we believe that the large scale computational material design is now likely, particularly with reliable empirical reactive potential.

2. METHODS

Carbon−Carbon Interaction Potential. The empirical bond-order type potential, namely Brenner potential[19] as implemented in the GULP package,[20−22] was utilized routinely in all the simulations. The previous work has shown that this potential can reproduce the bulk properties of diamond and graphite[23] as well as small fullerene molecules.[24,25] To verify the empirical Brenner potential for a whole PES exploration, we also utilized the self-consistent-charge density functional-based tight binding as implemented in dftb+ (mio parameters)[26,27] in combination with SSW method for searching GM of selected carbon clusters up to C_{60}. Starting from random structures, the correct C_{60} GM can be obtained within 500 SSW steps in a parallel run with 20 replicas (see below for explanation), the efficiency of which is similar to that using the empirical potential, indicating that the empirical potential is capable to describe the complexity of the PES of carbon systems.

SSW Method. The SSW method has been documented in detail in our recent publication[28] and an illustrative scheme of the overall algorithm can be found in Figure 1b. For completeness, here we briefly introduce the essence of the method. Originating from the bias-potential driven constrained Broyden dimer (BP-CBD) method[29] for transition state (TS) searching of complex catalytic systems,[30−32] the SSW method is based on the idea of bias-potential-driven dynamics[33] and Metropolis Monte Carlo (MC)[34] targeting for a rapid PES exploration. The SSW manipulates smoothly the structural configuration on PES from one minimum to another by adding bias potentials (eq 1) and relies on the Metropolis MC to accept or refuse each move (eq 2, P is the acceptance probability). The bias potentials are utilized to surmount the barrier along the pathway. The PES exploration using the SSW method is fully automated and does not need a priori knowledge on the system such as the structure motif (e.g., bonding patterns, symmetry) of material. The minimum structures, both local and global, can be obtained from SSW searching trajectories, and the information on the reaction pathways connecting minima is also preserved.

$$V_{\text{m-to-H}} = V_{\text{real}} + \sum_{n=1}^{\text{NG}} v_n = V_{\text{real}} + \sum_{n=1}^{\text{NG}} w_n \times \exp[-((R^t - R_t^{n-1})\cdot N_i^n)^2 /(2\text{ds}^2)] \quad (1)$$

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$P = \begin{cases} \exp[E(R_i^m) - E(\mathbf{R}_i^{mt}) / RT], & \text{when } E(\mathbf{R}_i^{mt}) > E(\mathbf{R}_i^m) \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

In one SSW step, i.e. from one minimum to another (denoted using subscript i), the potential is divided into two parts: (i) from the minimum $\mathbf{R}_i^m$ to a high energy configuration $\mathbf{R}_i^{NG}$ (see eq 1), where the potential is modified by adding consecutively the bias Gaussian potentials ($V_{m\text{-to-}H}$); and (ii) from $\mathbf{R}_i^{NG}$ to a second minimum $\mathbf{R}_i^{mt}$ where all bias potentials are removed and only real potential (V_{real}) remains. Each Gaussian potential is added along a direction $\mathbf{N}_i^n$ (eq 1) that is randomly generated and refined by constrained Broyden dimer rotation with bias potential,[29,35] where the superscript n is the index for the current Gaussian potential. Simply speaking, the random mode $\mathbf{N}_i^n$ combines two parts, one with global movement and one with local movement, the contribution of each is determined by a random coefficient λ (see eq 1 of ref 28). The local movement is a bond formation mode between two randomly selected atoms shown in eq 2 of ref 28. As depicted in Figure 1b (green line), the local relaxation is repeatedly performed to relax the forces perpendicular to the $\mathbf{N}_i^n$ direction right after each Gaussian is added. These local relaxations, although increasing the computational demanding per SSW step, disallow too high energy states along a walking path and thus increases the overall probability to locate low energy configurations.

The SSW method has three important parameters, namely, the Gaussian width ds, the total number of Gaussians (NG) added per SSW step, the temperature utilized in Metropolis MC (T, in eq 2). As shown in Figure 1b, ds determines the magnitude per structural perturbation. Our previous test on model systems show that a sensible value of ds is 0.2−0.6 Å, being about 10−40% of typical chemical bond length; NG dictates the overall walking distance of one SSW step and its typical value is 9−15, corresponding to a displacement of ∼5 Å (NG times ds) per SSW step. From eq 2, T in Metropolis MC controls the acceptance rate for new minima, which should be neither too high nor too low in order to explore PES rapidly and locate the low energy structures. In section 3.4, we will analyze how these parameters affect the overall efficiency of PES exploration in the complex systems.

Parallel Replica Exchange Implementation. One major bottleneck of SSW method is the identification of the right mode direction that can lead to the occurrence of the reaction toward better energy configurations. With the random mode generation procedure,[28] this unbiased approach can be time-consuming particularly for large systems with many degrees of freedom. To overcome this problem, in this work we implemented a parallel version of the SSW method learning from the parallel tempering technique of MC/MD[36−38] to speed up the PES exploration. The purpose is to allow the exchange of the structural configurations among different replicas according to the Metropolis criterion as eq 2. Specifically, the higher energy replica will have a chance to adopt a lower energy configuration identified by the other replica during the simulation. This can lead to the accumulation of random walkers (replicas) at the lower energy regions and thus the PES there is explored more rapidly.

In practice, we define N (e.g., $N \geq 10$) copies for one parallel run, each replica starting from a different random structure. During the run, a local-best and a global-best structure are updated to keep the best configurations in each replica and among all replica, respectively. When an exchange is performed at a predefined exchange frequency (e.g., every 10 SSW steps) by trading the local-best structure of the current replica with the global-best structure according to the Metropolis criterion (the exchange temperature is kept as the same as the T of SSW steps), the following two possible actions are taken: (i) if the local-best structure is replaced by the global-best structure, the replica resumes from the global-best configuration and meanwhile its SSW parameters (NG and T) will be updated to a set of new values; and (ii) if the global-best structure is replaced by the local-best structure, the SSW parameters (NG and T) of this replica are reset to the input values (NG_{inp} and T_{inp}). In the implementation, we define the ranges for NG and T values according to NG_{inp} and T_{inp}, which are (NG_{inp}, NG_{inp} + 5) for NG, and (T_{inp}, $5T_{inp}$) for T, and the new set of parameters is randomly picked within these ranges. By this way, the replicas that have adopted the same global-best structure will have different surface walking behavior (different NG and T), which enhances the probability to find lower energy structures nearby.

3. RESULTS

3.1. PES of Fullerene Isomerization. To assemble individual carbon atoms to stable carbon fullerenes, it is expected that the finding of the correct fullerene structures among numerous spherical isomers at the final stage would be the most challenging task since the initial agglomeration of carbon atoms to condensed cluster should be highly exothermic and thus occur rapidly.[39] To this end, we have analyzed the so-called Stone−Wales or "pyracylene" rearrangement, which is the 90° rotation of two carbon atoms with respect to the midpoint of the bond on C_{60} fullerene.[40−42] The Stone−Wales rearrangement occurs commonly in the structural change of sp^2-bonded carbon nanosystems and the fusion process of fullerenes or carbon nanotubes may occur through a sequence of such a rearrangement. Therefore, in order to provide an overview on the corrugated PES of carbon clusters, we first determined the reaction pathway of the Stone−Wales rearrangement by using the BP-CBD method[29] and the reaction profile is shown in Figure 2. The purpose of this preliminary study on the carbon−carbon bond breaking/

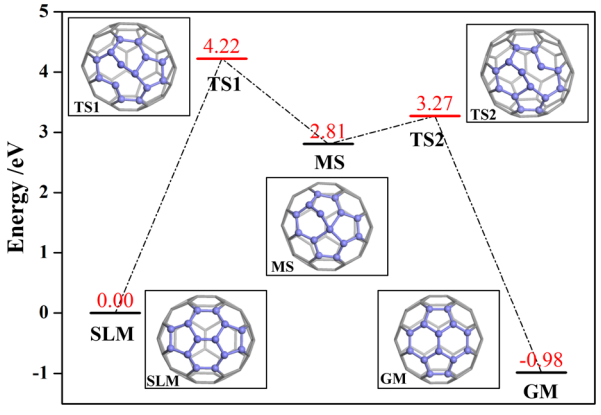


Figure 2. Reaction pathway and structures of the Stone-Wales rearrangement in C_{60} using Brenner potential,[19] illustrating the highly corrugated PES in fullerene isomeration. The atoms in blue highlight the reaction region.

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

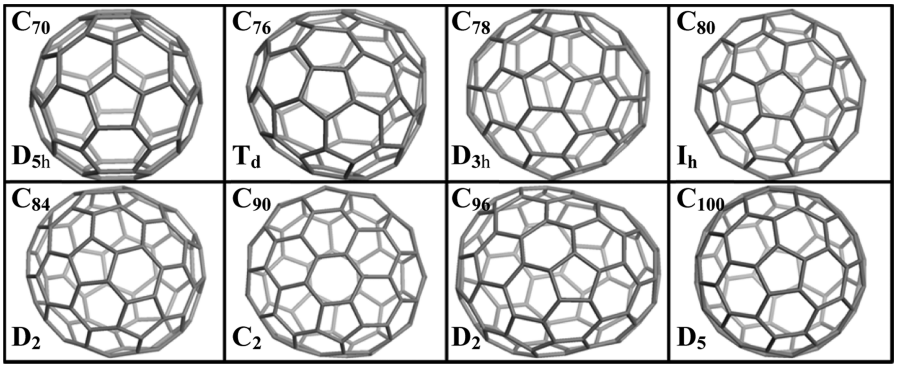


Figure 3. GM structures of C_n fullerenes from Brenner potential.

Table 1. Statistics for the GM Searching of C_{60}–C_{100} Fullerenes Using SSW

			w/o exchange[a]			w exchange	
C_n	E_{GM}/eV	$N_{IPR\text{-}i}/N_{IPR\text{-}t}$	N_{succ}/N_{run}	N_{SSW}[b]	N_{ave}[c]	N_{rep}	N_{SSW}[b]
60	−410.507	1/1	9/10	4397	12354	20	1011
70	−483.177	1/1	4/10	9023	45430	20	1070
76	−527.097	2/2	4/10	19244	23830	20	5730
78	−541.805	3/5	8/10	7003	25174	20	2970
80	−557.124	3/7	1/20	22644	22644	20	14488[d]
84	−584.989	8/24	6/10	27922	61485	20	2144
90	−628.634	20/46	10/20	74927	140775	20	7680
96	−672.543	51/374	3/20	62887	144597	30	27830
100	−702.310	76/900	0/20			30	19340

[a]The allowed maximum number of SSW steps is 10^5 for C_{60}–C_{84} and 2×10^5 for C_{90}–C_{100}. [b]N_{SSW} is the SSW steps in the shortest trajectory. [c]N_{ave} is the SSW steps averaged over the successful trajectories. [d]T is set as 9000 K

making reaction is also to validate the Brenner potential for describing the reaction energy of carbon cluster isomerization.

Figure 2 shows that on going from the initial state (IS), the second lowest minima (SLM) of C_{60} that is a fullerene-like structure with adjacent pentagons, to the final state (FS) that is the GM of C_{60}, the Stone–Wales rearrangement experiences a high overall barrier (4.2 eV) and several key states, including two TSs (**TS1**, **TS2**) and one intermediate state (**MS**). The atoms involved in the rearrangement have been highlighted by the light blue color, illustrating the highly local and directional bonding of carbon: during the reaction process, two σ C–C bonds break involving only six atoms and the other atoms are almost intact. This feature implies that the large random perturbation on the IS structure is unlikely for yielding the FS. The reaction enthalpy (1.0 eV) and the barrier (4.2 eV) from Brenner potential compares reasonably well with previous DFT calculations,[43,44] where the rearrangement is found to be exothermic by ∼1.5 eV (1.5 from LDA, BLYP, PBE and 1.7 from B3LYP) with a barrier of ∼5 eV (4.7 from LDA, BLYP, 5.3 from PBE, and 5.5 from B3LYP).[43]

The high barrier of Stone–Wales rearrangement indicates that the PES of carbon clusters is highly corrugated and each of the isomers forms a deep energy well on the PES. This is the main reason that MD-based techniques are readily trapped in the local minima. From microkinetics, an event with a barrier of 4 eV will require a system temperature of about 1500 K to allow the reaction to occur within a temporal scale of 1 s, which is already far too long for MD simulations. If a high temperature (e.g., 4000 K) is applied instead in MD simulation, although it helps to overcome the barrier within a shorter time period, the configuration will be driven toward liquid-like structures by the entropy term at the high temperature. This problem of the conventional MD approach has been addressed in the previous investigation,[4] which showed that the relaxation to buckminsterfullerene from high-energy isomers can proceed by continuous Stone–Wales rearrangement using a master equation approach. The local structural arrangement, the high barrier, and the large dimensionality are therefore three keywords for the PES of carbon fullerenes, making them a perfect system for testing seriously the global minimization algorithm.

3.2. GM Search of C_{60}–C_{100} Fullerenes. By using the SSW method, we have searched the GM structure of nine carbon clusters C_n (n ranges from 60 to 100) starting from random structures. For each cluster, we both run a number (N_{run}) of independent searches (without the parallel replica exchange, $N_{run} \geq 10$) and also conduct a parallel run with a number of replicas (N_{rep}). The ds and NG are set as 0.6 and 9, respectively, for all the searches, and the T in Metropolis MC is set generally as 3000 K except for the parallel run of C_{80} system where T is set as 9000 K. The lowest energy structure obtained from these runs were then compared with the energy of all the possible IPR structures reported in Yoshida's fullerene library,[45] which confirmed that the GM of these large carbon fullerenes have all been identified using the SSW method. Figure 3 shows our located GM structure of these carbon fullerenes, including C_{70}, C_{76}, C_{78}, C_{80}, C_{84}, C_{90}, C_{96}, and C_{100}. The statistics for the GM searching are also listed in Table 1.

In general, the greater the degrees of freedom, the more complex the PES. For fullerenes, the complexity of the PES can be viewed simply from the total number of the possible IPR isomers, $N_{IPR\text{-}t}$[46] (see Table 1), which increases exponentially with the increase of the number of atoms.[47,48] The I_h symmetry GM is the only isomer that matches the IPR in C_{60}, and the

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

number of IPR isomers increases to 900 in C_{100}. Since the interconversion between IPR isomers must involves a series of high-barrier carbon ring shift on the spherical surface, each IPR isomer can be regarded as a deep energy trap on PES and the increase of IPR isomer numbers directly leads to the increase of the difficulty for GM location.

From our results of independent searches (Table 1 data without exchange), the GM location becomes more and more challenging on going from C_{60} to C_{100} in general and, as expected, all the carbon clusters exhibit multiple funnel PES (none of them can achieve 100% success rate within reasonable steps). For C_{60}, C_{70}, and C_{78}, the GM location is facile with a relatively high success rate (N_{succ}/N_{run} in Table 1, over 40% in 10 independent searches) and the SSW steps in the shortest trajectory is less than 10^4, denoted as N_{ssw} in Table 1. For C_{76} and C_{84}, N_{ssw} is about 2×10^4; for C_{90} and C_{96}, N_{ssw} is more than 6×10^4 and for C_{100}, all the independent searches cannot locate the GM within 2×10^5 steps. The SSW steps averaged over the successful trajectories (N_{ave}) have the similar trend from the small to the large system but is generally at least two times more than N_{SSW}.

It is also obvious that the parallel replica exchange can speed up the GM location: the number of the SSW steps required in the shortest trajectory (N_{SSW}) decreases by 36−92% when the replica exchange is conducted. This is particularly beneficial for large clusters such as C_{100} where only the parallel run can identify the GM within 2×10^5 steps.

Among the systems, C_{80} is special with extremely low success rate for GM location, 5% (1/20) out of 20 independent runs and it is thus worth of close inspection. The GM of C_{80} is a high symmetry I_h structure, and the SLM has a D_{5h} symmetry (see the Supporting Information (SI) S-Figure 1 for the structure). In fact, most of the nonparallel runs (16/20) and the parallel run at low temperatures (e.g., 3000 K) reach to the SLM rapidly (within 10^4 steps). Only by elevating the T to 9000 K, the parallel run can finally identify the correct GM. The question is therefore why the conversion from the D_{5h} SLM to the I_h GM is of low probability using SSW at low temperatures. By inspecting closely their structures, one can conceive that the interconversion of the two structures requires a clockwise rotation of the upper sphere and at the same time an anticlockwise rotation of the bottom sphere around the principal C_5 axis of the SLM. This process needs to break 10 σ C−C bonds, which is energetically too high to occur as one elementary step. From the identified trajectories for SLM to GM, we found that the conversion from the SLM to the GM always go through the other much less stable isomers (>6 eV higher) involving continuous ring shifting. In addition, we also found that the GM-to-SLM conversion is much more facile than the SLM-to-GM conversion: our results show that the success rate of GM-to-SLM conversion is more than 70% within 10^5 steps at 10 000 K. These facts imply that the PES of C_{80} must possess a double-funnel feature, as illustrated in Figure 1a, that is to say, the GM (at the narrow funnel) and SLM (at the large funnel) are separated by a high barrier and a large spatial distance. It might be mentioned that the GM location in C_{80} is at least 1 order of magnitude more difficult than that in Lennard-Jones 75 (LJ_{75}) that is known to possess the similar double-funnel PES.[11,49−51] Our results show that at a 5% success rate the shortest trajectory for SLM-to-GM conversion requires ~10^4 SSW steps in C_{80} and it is below 10^3 in LJ_{75}.

Apart from the GM structures, the SSW method has identified many other less stable structures that match the IPR along the searching trajectories. For example, the SSW method finds 20 IPR-isomers ($N_{IPR\text{-}i}$) out of 46 in total for C_{90} in ten runs (see Table 1). Compared to the GM, these IPR-isomers generally have a lower symmetry and the pentagon rings are closer to each other compared to them in the GM. Interestingly, with the increase of the system size, $N_{IPR\text{-}i}$ increases much slowly compared to that of $N_{IPR\text{-}t}$. This implies simply that in order to identify the lowest energy configurations it is not necessary to explore the whole PES. In general, the low energy structures are in close neighborhood on PES and it is of high probability to identify the GM by sampling the states along the low energy pathways.

Finally, it is of interest to compare the PES of carbon clusters with that of short-ranged Morse clusters in the context of the SSW GM searching from random structures. The unbiased GM location of short-ranged Morse clusters has long been a challenge[52−55] and only until recently we show that the SSW can successfully identify the GM of the short-ranged Morse clusters up to 103 atoms. In Table 2, we list the efficiency for GM location using the SSW method for Morse 60, 70, and 78 and compare them with the carbon fullerenes. It can be seen that the GM location of the fullerene is several times more difficult than that of the same size Morse cluster. For these Morse clusters, the SSW method can achieve a probability of at least 47% for GM location within 10^4 SSW steps and 100% within 10^5 steps, while it is below 20% within 10^4 SSW steps and less than 90% within 10^5 steps for carbon clusters.

Table 2. Probability to Locate GM from Random Structures Using SSW Method for Short-Ranged Morse Clusters[28] and Carbon Fullerenes within 10^4 and 10^5 SSW Steps

	Morse cluster		carbon cluster	
atom number	10^4 steps	10^5 steps	10^4 steps	10^5 steps
60	47%	100%	10%	90%
70	80%	100%	20%	40%
78	50%	100%	10%	80%

3.3. SSW Trajectories of C_{60} Fullerene from Random Structures. In this section, we analyze the trajectories obtained for the assembling of C_{60} fullerene from randomly generated atoms. In Figure 4, we have collected 40 independent trajectories and plotted the trajectories by recording only the minima that is exothermic along the trajectory for the purpose of clarity. Because the total SSW steps of these trajectories vary from 5079 to 9473, we utilize a rescaled step ratio, being the SSW step number divided by the total steps, as the axis to show the gradual energy decrease of the assembling process in Figure 4a. The density of the states for these exothermic minima (DOM, $g(E)$ with $\int[g(E)/N_{tot}]\,dE = 1$, N_{tot} is the total number of minima in all trajectories) is shown in Figure 4b and c. $g(E)$ is calculated as $g(E) = \sum_i N(E)$, where i labels the trajectory; $N(E)$ is the times the minimum structure with energy E appeared in the ith SSW trajectory.

As shown in Figure 4a, most of the trajectories accumulate at the bottom-left corner, indicating that the trajectories are similar in the sense that they go through the similar intermediates at the similar time scale as measured by the step ratio. From these trajectories, we can also distinguish four common stages from the high to the low energy in the assembling process, and the typical structure picked from the

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

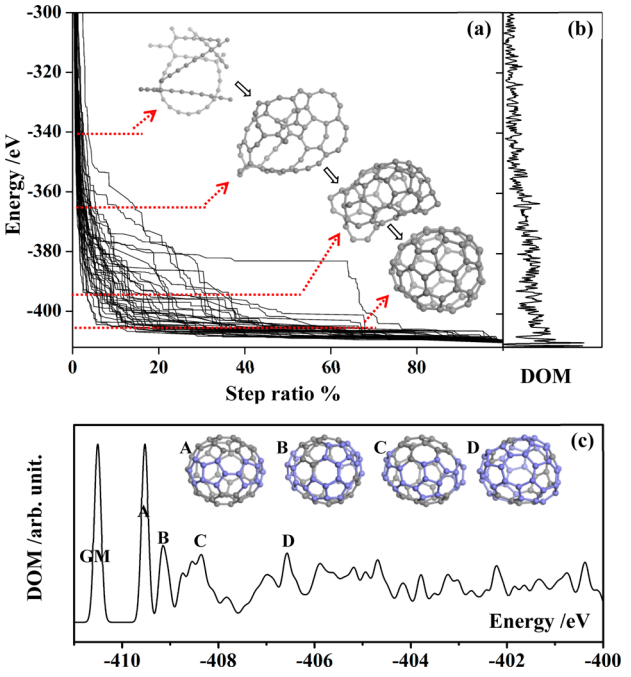


Figure 4. Trajectories of the assembling of C_{60} from random structure (a) and the density of the states for the exothermic minima (DOM) along the trajectories (b and c). The atoms in blue highlight the parts not obeying IPR.

four stages are illustrated in the insert of Figure 4a. Starting from the randomly distributed atoms (high energy states above −300 eV), the first stage reached is the long-chain carbon cluster, which appears at ∼−340 eV. These long chains will then cross-link with each other to reach to the second stage, forming large carbon cycles. The energy window of the second stage is ∼−370 eV. Next, the carbon cycles will condense to form the uncompleted cage structures at ∼−390 eV. The characteristic hexagon also appears at the third stage. Finally, the spherical-like carbon cage featuring mainly pentagons and hexagons emerges at ∼−410 eV, which is no more than 10 eV higher in energy compared to the GM (−410.507 eV). From Figure 4b, it is obvious that many stable isomer structures with close energy spacing are present at the final stage, which leads to a high density of minima. All the trajectories will be trapped at this stage for a long period of up to 90% simulation time before the GM can be eventually identified.

In Figure 4c, we enlarge the density of the exothermic minima along the trajectories at the energy window below −400 eV in order to have a closer look on the possible isomeric structures of the spherical-like shape. The typical low energy structures as indicated by A, B, C, and D are displayed in the inset. These structures all have a hollow spherical shape but with differently arranged rings. The low energy isomers, although they may contain only pentagons and hexagons, do not obey the IPR with closely connected pentagons. The transformation of these structures to GM must go through the Stone−Wales rearrangement. The structure A is the SLM and its pathway to the GM is already shown in Figure 1. Because of the high barrier and the highly local feature of the transformation, a significant computational time is spent on finding the correct movement direction as achieved by the random mode generation procedure in the SSW method.

It should be emphasized that in the gas-phase the C_{60} fullerene is generated by the shrinking mechanism of giant C_n fullerenes ($n > 100$) based on QM molecular dynamics study.[7] The SSW trajectories in Figure 4 are thus not chemical meaningful, mainly for illustrating how the SSW method finds the correct GM from random structure and for identifying the common features among different trajectories, such as the four stages outlined above. The SSW trajectories are generally not the lowest energy pathways since the SSW method focuses on a rapid PES exploration (to identify the low energy minima) and does not intend to find the lowest energy pathway (where accurate TS location is required).

Nevertheless, because the GM and the lowest energy pathways are often correlated properties for a PES, SSW trajectories can also provide important information on the low energy pathways for two reasons: (i) With small perturbations and repeated local relaxations, the SSW trajectory tends to avoid too high energy pathways (e.g., atom collision). (ii) With the Metropolis MC criterion, only those pathways with relatively low energy intermediates will be selected in the searching. This explains that the trajectories identified above for C_{60} assembling exhibit the great similarity.

3.4. Pathway Selection and Overall Efficiency As Influenced by SSW Parameters. Finally, it is of interest to analyze how the tuning of SSW parameters can affect the efficiency of the PES exploration. To this end, we utilize C_{20} as the model system to examine the influence of the parameters on the slowest conversion steps in the SSW trajectories. We select C_{20} because the system is small enough to obtain converged statistics rapidly and the GM location of C_{20} is also representative for larger carbon clusters. With the Brenner potential, C_{20} has a C_{5v} bowl-like GM (see Figure 5), which also

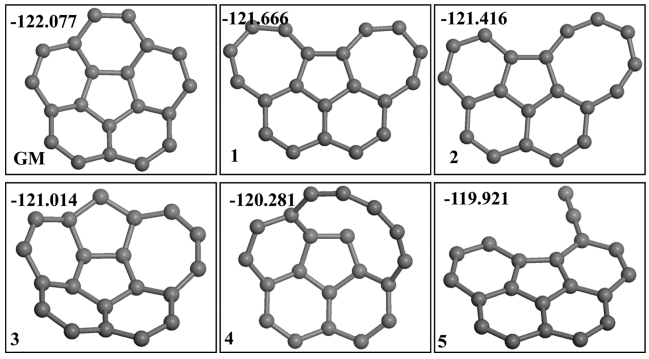


Figure 5. Important structures for the GM location of C_{20} using Brenner potential.

satisfies the IPR as encountered in the large fullerenes. It is noted that C_{20} can have different GM structures depending on the level of theory utilized:[56] the other two common possible GM structures are the chain (D_{10h}) and cage (I_h) structures. These two structures are however 1.30 and 2.60 eV higher than the bowl-like GM structure with Brenner potential, and they are not identified using SSW at the temperature range investigated, which suggests that they belong to different funnels. Indeed, we can identify the chain and the cage structures from random atoms when using other empirical potentials (EDIP[57] and ReaxFF[58] potential).

Starting from randomly distributed carbon atoms, we first run 20 independent SSW searches with T being set as 300 K. The low temperature utilized is to identify the large energy traps before the GM is reached. In these 20 trajectories, we found that the carbon atoms agglomerate rapidly at the initial

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

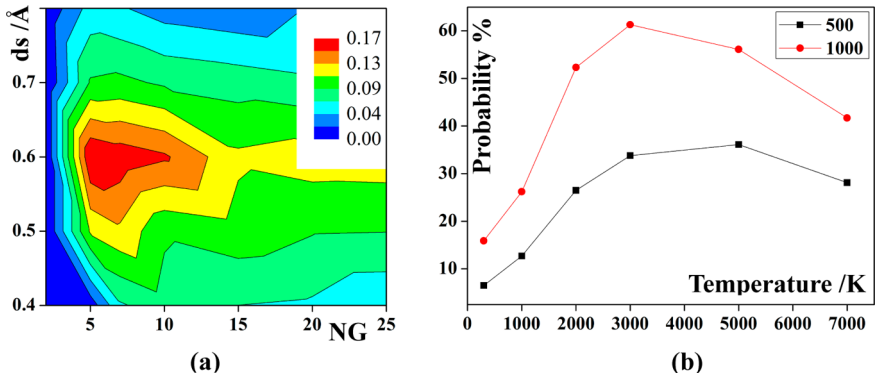


Figure 6. Dependence of the probability of the GM location on the parameters (a) ds and NG and (b) T for C_{20}. In part b, the results with maximum 500 and 1000 SSW steps are both shown.

stage, similar to those found in the assembling of large fullerenes, and all the trajectories will then be trapped at a C_{2v} planar structure for a long period, which consists of 5-, 6-, and 7-member rings, denoted as state **1** (see Figure 5). It is therefore of significance to analyze how the SSW method converts state **1** to the GM in the presence of different SSW parameters.

Effect of ds and NG. The ds and NG are two correlated parameters, which together define the surface walking length of one SSW step. In Figure 6a, we show the contour graph of the probability of the GM location from the state **1** at the different combinations of ds and NG, in which the probability are averaged over 1000 runs per one set of ds and NG (with T being always kept at 300 K). We found that the best performance is achieved when ds is ∼0.6 Å and NG is in the interval of 6−10. This is equivalent to a structural perturbation of 3.6 to 6 Å per SSW step. Apparently, the performance is not very sensitive to NG as long as it is larger than 6 but is rather sensitive to ds. For ds = 0.6 and NG = 24, the performance is only dropped by 22.9% compared to ds = 0.6 and NG = 9. On the other hand, for ds = 0.4 and NG = 9, the performance is dropped by 49.7% compared to ds = 0.6 and NG = 9. It should be mentioned that we also performed the similar ds/NG test by increasing T to 3000 K and the similar result is obtained (see SI S-Figure 2). This indicates that ds being 0.6 and NG larger than 6 are optimum for C_{20} systems.

The effect of *T*. We then studied the temperature effect on the GM location while keeping ds and NG being constant (0.6 and 9). Figure 6b illustrates the variation of the probability to locate GM upon the change of T. The data at each T is averaged over 1000 independent runs with the maximum SSW steps being set as 500 or 1000. It shows that the best performance can be achieved at the interval 3000−5000 K. The probability to locate GM increases rapidly from 300 to 3000 K, and then it declines slowly above 5000 K. While the temperature increases by more than 20 times, the overall efficiency varies less than 7 times. This indicates that the SSW method is not very sensitive to the parameter T, and a good performance can be achieved over a wide temperature window.

The above results can be understood by analyzing the SSW trajectories, which contain many possible pathways leading to the GM. By tuning the parameters ds, NG, and T, the relative percentage of the pathways should differ and therefore the overall efficiency of GM location changes. Below we utilize the SSW trajectories at different T (maximum 1000 SSW steps, Figure 6b red curve) to illustrate how SSW works in this example.

First, we have identified four other intermediates in addition to the state **1** that can reach to the GM in one SSW step, as numbered sequentially from states **2**−**5** with increasing energy shown in Figure 5. On the basis of the large number of SSW trajectories, it is then possible to determine the lowest energy pathways from the state **1** to the GM. This is done by resolving the lowest energy pathways between any of the two states (**1**, **2**, **3**, **4**, **5**, and the GM) where their linkage within one SSW step (e.g., **1**-to-**2**, **2**-to-**3**) is present in SSW trajectories. With the information of two ends (IS and FS), it is straightforward for BP-CBD[29] method to locate the associated TSs and intermediates.

From **1** to **GM**, we found two low energy pathways (pathway-a and pathway-b), as shown in Figure 7, which either

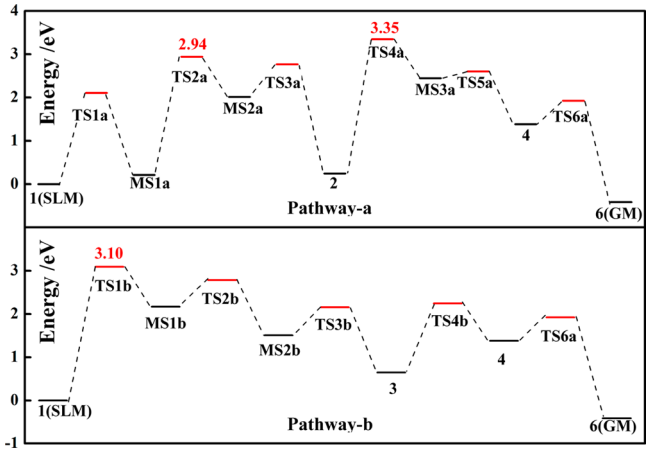


Figure 7. Low energy pathways from the state **1** to the GM of C_{20}. All the intermediate structures are depicted in SI-Figures 3 and 4.

go through the state **2** (pathway-a) or the state **3** (pathway-b). The conversion between states **2** and **3** can occur through the common intermediate state **4**, which is an immediate precursor to the GM. Interestingly, pathway-a has better intermediates (**MS1a** and **2**) with lower energy and pathway-b has an overall lower barrier (3.10 eV). Obviously, except for state **5**, states **2**, **3**, and **4** all locate on these two low energy pathways.

Next, the relative weight of the five states that act as the precursor to the GM (one SSW step before GM being identified) at different temperatures is calculated from SSW trajectories and the results are plotted in Figure 8. It shows that at the low temperature (300 K) all conversions from the state **1** to the GM occur via a single SSW step (the relative weight of

dx.doi.org/10.1021/ct400238j | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

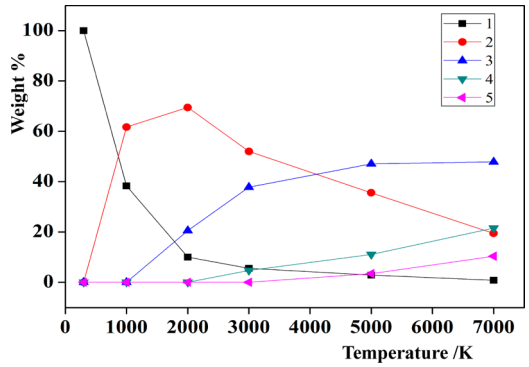


Figure 8. Relative weight of the states **1**−**5** as the precursor to the GM at different T of Metropolis MC.

the state **1** is thus 100%). The overall probability of these conversions is however rather low (<17%), as seen from Figure 6b. As there are at least four elementary steps in the low energy pathways (Figure 7 pathway-b), the one-step conversion must go through the high energy channels and is thus less likely to occur.

By increasing the temperature to 1000−2000 K, the pathway involving state **2** starts contributing to the GM location (the probability for a minima being accepted to replace the state **1** can be evaluated using eq 2 quantitatively, which are, for example, 0.85% and 0.05% at 1000 K for state **2** and **3**). At the low temperature region, the SSW method prefers the pathway-a for converting state **1** to GM because pathway-a contains lower energy intermediates (**MS1a** and **2**) that can be accepted with high probabilities in SSW trajectories. Above 1000 K, the SSW can undergo both pathway-a and pathway-b for the **1**-to-GM conversion and the efficiency of GM location reaches the highest at ∼3000 K. This is obviously due to the fact that the low energy pathways involving the state **2** and **3** that the SSW trajectories prefer can now be accepted with the highest probability with less refusing from the Boltzmann term of Metropolis MC. At high temperatures, 6000 to 7000 K, the unstable states **5** emerge with the reduced probability to locate GM (Figure 6b). The presence of state **5** reflects the opening of the high energy reaction channels (state **5** is not on the two low energy pathways), which can lead to other high energy states instead of the GM.

Therefore, *the SSW's selection of pathways determines first the possible reaction channels*, where the lowest energy pathways are preferred but the high energy pathway is also allowed with a finite probability, and *the Metropolis MC acts as the second guard to refuse the too high energy states that appears most often in the high energy pathways*. The above results on the SSW parameters demonstrate that the PES exploration using the SSW method is not very sensitive on the temperature (of Metropolis MC). As demonstrated in C_{20}, even at the low temperature 300 K, the GM can be located considering that the barrier is at least 3.10 eV. On the other hand, by tracing a large number of SSW trajectories, it is possible to identify the lowest energy pathways because the intermediates along the low energy pathways appear much more frequently than the other high energy states. Overall, the SSW method can provide a large database of minima and their connectivity (as represented in SSW trajectories) in complex systems, and with the help of the TS-location method (BP-CBD) in the same framework, it might be an attractive new tool for investigating material phase transition. Our ongoing work shows that the above approach can be extended to quantify the phase transition of periodic crystal systems (such as the rutile to anatase transition of TiO_2).

4. CONCLUSION

By combining recently developed SSW method with a parallel replica exchange technique, this work assembles nine carbon fullerenes containing up to 100 atoms from random structures without using any a priori knowledge on the systems. These carbon clusters generally have multiple-funnel PES due to the high barrier of elementary reactions, and among them C_{80} is the most challenging case as its GM locates at a much narrower channel compared to the SLM. Within one unified theoretical framework, we show that the GMs of these fullerenes can be identified together with a large data set of trajectories showing the close connectivity between minima, and the parallel replica exchange technique can expedite greatly the PES exploration. With the results presented for the carbon PES, we believe that the SSW method, as a general tool for PES searching, could expand greatly the scope of computational material design in the future.

Detailed analyses on the SSW trajectories demonstrate that the SSW method is able to identify the GM rapidly because of its ability to surmounting high barrier and the preference of SSW trajectories to low energy pathways. Therefore, the key difficulties encountered for material structure prediction are now solved using the SSW algorithm: (i) The high barrier required for structural transformation. For the isomerization of carbon fullerenes, the barriers are generally more than 4 eV. The repeatedly added Gaussian potentials along one particular direction in each SSW step help to walk through the high energy transition region. (ii) The sampling of the desirable reaction channels toward better configurations. The random mode generation together with the biased dimer rotation procedure in SSW help to identify the entrance channel for the desirable reaction pattern. During the surface walking, the low energy pathways along this direction are generally preferred thanks to the repeated local relaxation right after the addition of Gaussian potential, which makes the PES exploration biased toward the low energy regions. In combination with parallel replica exchange, the SSW method can maintain a high efficiency of GM location for systems up to a large dimensionality.

■ ASSOCIATED CONTENT

Supporting Information

SLM of C_{80}; the efficiency of GM location at 3000 K with different NG for C_{20}; the structures for TSs and MSs along the lowest energy pathways (Figure 8) for C_{20} conversion. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail address: zpliu@fudan.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge National Science foundation of China (21173051), 973 program (2011CB808500, 2013CB834603), Science and Technology Commission of Shanghai Municipality

(08DZ2270500), Program for Professor of Special Appointment (Eastern Scholar) at the Shanghai Institute of Higher Learning for financial support.

REFERENCES

(1) Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162−163.
(2) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758−760.
(3) De, S.; Willand, A.; Amsler, M.; Pochet, P.; Genovese, L.; Goedecker, S. *Phys. Rev. Lett.* **2011**, *106*, 225502.
(4) Walsh, T. R.; Wales, D. J. *J. Chem. Phys.* **1998**, *109*, 6691−6700.
(5) Maruyama, S.; Yamaguchi, Y. *Chem. Phys. Lett.* **1998**, *286*, 343−349.
(6) Xu, C.; Scuseria, G. E. *Phys. Rev. Lett.* **1994**, *72*, 669−672.
(7) Irle, S.; Zheng, G.; Wang, Z.; Morokuma, K. *J. Phys. Chem. B* **2006**, *110*, 14531−14545.
(8) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(9) Ballone, P.; Milani, P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1990**, *42*, 3201−3204.
(10) Zhang, B. L.; Wang, C. Z.; Ho, K. M.; Xu, C. H.; Chan, C. T. *J. Chem. Phys.* **1992**, *97*, 5007−5011.
(11) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(12) Doye, J. P. K.; Wales, D. J. *Phys. Rev. Lett.* **1998**, *80*, 1357−1360.
(13) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368−1372.
(14) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed.; Addison-Wesley: Reading, MA, 1989.
(15) Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288−291.
(16) Oganov, A. R.; Glass, C. W. *J. Chem. Phys.* **2006**, *124*, 244704.
(17) Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*; MIT press: Cambridge, MA, 1992.
(18) Kroto, H. *Nature* **1987**, *329*, 529−531.
(19) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2002**, *14*, 783.
(20) Gale, J. D. *Philos. Mag. B* **1996**, *73*, 3−19.
(21) M. Woodley, S.; D. Battle, P.; D. Gale, J.; Richard A. Catlow, C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2535−2542.
(22) Gale, J. D.; Rohl, A. L. *Mol. Simul.* **2003**, *29*, 291−341.
(23) Gao, G. T.; Van Workum, K.; Schall, J. D.; Harrison, J. A. *J. Phys.: Condens. Matter* **2006**, *18*, S1737−S1750.
(24) Yamaguchi, Y.; Maruyama, S. *Chem. Phys. Lett.* **1998**, *286*, 336−342.
(25) Hobday, S.; Roger Smith, a. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 3919−3926.
(26) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678−5684.
(27) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.
(28) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2013**, *9*, 1838−1845.
(29) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2012**, *8*, 2215−2222.
(30) Wang, H.-F.; Liu, Z.-P. *J. Am. Chem. Soc.* **2008**, *130*, 10996.
(31) Shang, C.; Liu, Z.-P. *J. Am. Chem. Soc.* **2011**, *133*, 9938−9947.
(32) Fang, Y.-H.; Liu, Z.-P. *J. Am. Chem. Soc.* **2010**, *132*, 18214−18222.
(33) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
(34) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(35) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2010**, *6*, 1136−1144.
(36) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(37) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910−3916.
(38) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(39) Lee, I.-H.; Kim, H.; Lee, J. *J. Chem. Phys.* **2004**, *120*, 4672−4676.
(40) Stone, A. J.; Wales, D. J. *Chem. Phys. Lett.* **1986**, *128*, 501−503.
(41) Balaban, A. T.; Schmalz, T. G.; Zhu, H.; Klein, D. J. *J. Mol. Struc.:THEOCHEM* **1996**, *363*, 291−301.
(42) Slanina, Z.; Zhao, X.; Uhlík, F.; Ozawa, M.; Ōsawa, E. *J. Organomet. Chem.* **2000**, *599*, 57−61.
(43) Choi, W. I.; Kim, G.; Han, S.; Ihm, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 113406.
(44) Kumeda, Y.; Wales, D. J. *Chem. Phys. Lett.* **2003**, *374*, 125−131.
(45) Yoshida, M.; Fujita, M.; Gotō, H.; Ōsawa, E. *Electron. J. Theor. Ch.* **1996**, *1*, 151−162.
(46) Fajtlowicz, S.; Larson, C. *Chem. Phys. Lett.* **2003**, *377*, 485−490.
(47) Stillinger, F. H.; Weber, T. A. *Science* **1984**, *225*, 983−989.
(48) Wales, D. J.; Doye, J. P. K. *J. Chem. Phys.* **2003**, *119*, 12409−12416.
(49) Bogdan, T. V.; Wales, D. J.; Calvo, F. *J. Chem. Phys.* **2006**, *124*, 044102−044113.
(50) Leary, R. H.; Doye, J. P. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **1999**, *60*, 6320−6322.
(51) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891−908.
(52) Doye, J. P.; Leary, R. H.; Locatelli, M.; Schoen, F. *Informs J. Comput.* **2004**, *16*, 371−379.
(53) Cheng, L.; Yang, J. *J. Phys. Chem. A* **2007**, *111*, 5287−5293.
(54) Miller, M. A.; Doye, J. P.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 328.
(55) Doye, J. P.; Wales, D. J.; Berry, R. S. *J. Chem. Phys.* **1995**, *103*, 4234.
(56) Grossman, J. C.; Mitas, L.; Raghavachari, K. *Phys. Rev. Lett.* **1995**, *75*, 3870−3873.
(57) Nigel, M. *J. Phys.: Condens. Matter* **2002**, *14*, 2901.
(58) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.