JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Efficient Basin-Hopping Sampling of Reaction Intermediates through Molecular Fragmentation and Graph Theory

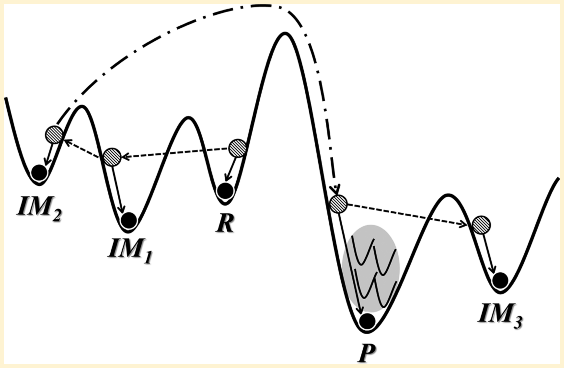
Yeonjoon Kim,[†] Sunghwan Choi,[†] and Woo Youn Kim*,[†,‡]

[†]Department of Chemistry, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea

[‡]KAIST Institute for NanoCentury, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea

S *Supporting Information*

ABSTRACT: Basin-hopping sampling has been widely used for searching local minima on a potential energy surface. Reaction intermediates including reactants and products are also local minima composed of a reaction path, but their brute-force sampling is too demanding because of large degrees of freedom. We developed an efficient Monte Carlo basin-hopping method to sample reaction intermediates through the fragmentation of molecules and a postanalysis scheme using the graph theory with a matrix representation of molecular structures. The former greatly reduces the dimension of a given potential energy surface, while the latter offers not only the effective screening of resulting local minima toward desirable intermediates but also their automatic ordering along a reaction path. We combined it with the density functional tight binding method for rapid calculations and tested its performance for organic reactions.



1. INTRODUCTION

Exploration of a potential energy surface (PES) has been an important subject for understanding molecular structures. A variety of methods and algorithms have been developed to find local minima on a PES.[1−4] Among them, the basin-hopping (BH) sampling was extensively used for diverse systems[5] such as Lennard-Jones particles,[6−10] atomic clusters,[11−13] molecular clusters,[14−18] biomolecules,[19−21] and metallic nanoclusters.[22−28] To our best knowledge, however, it has never been used to find intermediate structures of chemical reactions, though they are also essentially local minima on a PES.

In case of chemical reactions, local minima may correspond to key intermediates which compose a reaction path. Hence, their correct determination is crucial to elucidate full reaction mechanisms. So far theoretical study of reaction paths has been mainly focused on investigating energetics for intermediates given by experiments or chemical intuitions. However, such a conventional approach is problematic, if the collection of proposed intermediates is incomplete. For example, intermediates involved in a rapid reaction step can hardly be detected experimentally. In this regard, state-of-the-art computational methods for automatic exploration of reaction paths are being developed actively.[29−35]

For sampling reaction intermediates, the BH method should take formation and dissociation of chemical bonds into accounts. Therefore, it needs to be combined with a quantum mechanical method for electronic and geometrical structure calculations. Furthermore, various structures due to different geometrical representations of molecules produce lots of local minima on a PES as illustrated in Figure 1, which makes their brute-force searching impractical for large molecules. The PES of chemical reactions has two types of local minima; superbasins corresponding to molecules with different bond connectivity's and their sub-basins (shaded region in Figure 1)

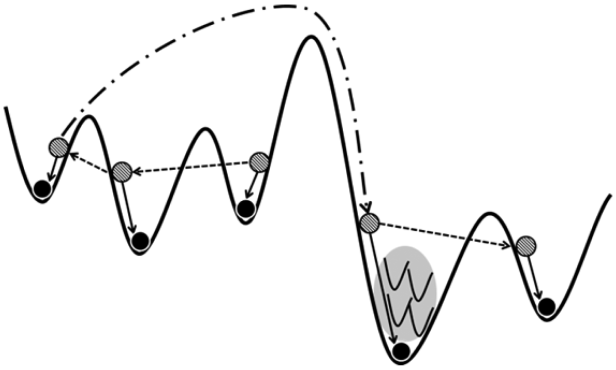


Figure 1. Schematic illustration of the potential energy surface of chemical reactions. A dashed−dotted line indicates a long-range hopping between two superbasins. The dashed lines indicate a short-range hopping between neighboring basins. The molecule at each superbasin may be an intermediate of chemical reactions, while its sub-basins correspond to its different conformers, as shown in the shaded region.

Received: February 17, 2014
Published: May 12, 2014

corresponding to different conformers with the same bond connectivity's. The BH algorithm should be modified to efficiently search both types of molecules as illustrated by the arrows in Figure 1.

We note that in many chemical reactions only a few reactive atoms directly participate in formation and dissociation of chemical bonds, while the others are just relaxed toward the nearest local minimum to form a stable structure. Thus, reactant molecules can be fragmented into minimal pieces in which no more bond dissociation is allowed. These fragments are subjected to the BH algorithm to produce a new chemical bond between them. The resulting structures are further relaxed using a conventional geometry optimization scheme. In this way, the dimension of a given PES greatly reduces down to tractable size even for large molecules. For instance, the PES of a chemical reaction involving a transition-metal catalyst with bulky ligands has very large dimension, but using the fact that the ligands are inactive the whole complex can be regarded as an intact single fragment. However, there would be still a number of local minima due to various structural representations of a molecule on the Cartesian coordinate, while they are considered as a single intermediate on the reaction coordinate. Namely the graph-theoretic approach is useful to classify all the resulting local minima into groups with the same connectivity's. In fact, the graph theory in mathematics has been applied to various fields of chemistry[36,37] such as the description of molecules,[38−40] generation of isomers,[41−43] investigation of kinetics and mechanisms of reactions,[44−46] and study of quantitative structure−activity relationship/quantitative structure−property relationship (QSAR/QSPR).[47−49]

We hereby introduce a powerful method to automatically search a comprehensive set of intermediates of chemical reactions using a modified BH sampling and to find a reaction path connecting from given reactants to target products through the Dijkstra algorithm. In the following sections, we first describe the method in detail and then illustrate with example studies how it can be used to find a reaction path.

2. METHOD

The present method is composed of four major steps. First, reactant molecules are fragmented into several pieces as explained below. Second, a Monte Carlo BH (MCBH) routine is carried out to form new chemical bonds between the fragments and to find their various conformers. The resulting structures are relaxed by a standard geometry optimization routine, yielding molecular structures at local minima. Third, a postscreening scheme classifies the local minima into groups with the same connectivities. They are further filtered by several criteria to obtain only desirable intermediate structures fit with a given chemical reaction. Finally, the Dijkstra algorithm is used to find a reaction pathway from a database of connected intermediates. The following subsections explain each step in detail.

2.1. Fragmentation of Reactant Molecules.

Unfortunately there is no systematic way of the fragmentation of reactant molecules, since it is mainly based on chemical intuition through the analysis of reactant and product molecules. Thus, each reaction has different ways of fragmentation. In general we apply the following rules: (i) ligands in a metal complex, backbone of molecules, and solvent molecules remain intact, (ii) covalent bonds directly taking part in chemical reactions are dissociated for fragmentations, and (iii) terminal hydrogens are typically regarded as an independent fragment. Though breaking those fragments is forbidden during geometry relaxations, some of fragments are allowed optionally to be broken so that natural bond formation or dissociation is also accounted. In principle, however, the fragmentation allows the formation of all possible chemical bonds for given fragments, because the criteria can be mitigated until the number of resulting intermediates is saturated.

Figure 2 shows an example of the fragmentation of reactants for the Buchwald−Hartwig amination.[50,51] One readily notes that the bond between the aryl group and halogen should be dissociated in order to produce the product, so they are fragmented into two pieces. The terminal hydrogen of amine also involves in the reaction so that it is dissociated into two fragments. According to the above rules the Pd complex should be regarded as a single fragment. However, in this reaction, the role of ligands is not elucidated clearly and thus the bond between Pd and one of the ligands is allowed to be broken during geometry relaxations, though they form a single fragment.

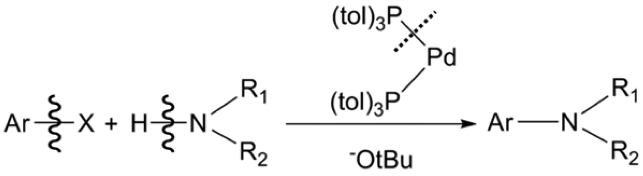


Figure 2. Fragmentation example for the Buchwald−Hartwig amination. The wavy lines indicate the bond dissociation for fragmentation. The dotted line indicates that it can be either fragmented or allowed to be broken during geometry relaxations.

2.2. Monte Carlo Basin-Hopping Sampling.

Figure 3 illustrates the flowchart of the present MCBH method. It begins with M number of input fragments and the cycle is repeated N_{MCBH} times. For every $(pM + 1)$th cycle ($p = 0, 1, 2, ...$), all the fragments are redistributed randomly without the Metropolis algorithm, leading to a hopping between two superbasins with long distance in a reaction coordinate as indicated by the dashed−dotted line in Figure 1. The inner-fragment bond rotation is performed to find various conformations of input fragments. (See the Supporting Information for details). Each fragment is randomly distributed with respect to the others to form a new chemical bond with them or to find its stable orientations. Then, it is rotated at their positions. The relative positions of $M - 1$ fragments are represented by the spherical coordinate (ρ, θ, ϕ) whose origin is taken at the center of mass of a prior-chosen fragment (usually the largest one). θ and ϕ of each fragment are determined randomly, and then, ρ of one selected fragment is gradually increased by 0.08−0.8 Å until it satisfies namely the toughness criterion given in eq 1 to avoid its overlap with the others after the relocation.

$$r_{ij} > c_{tough}(R_i + R_j) \quad (1)$$

where r_{ij} is the distance between the ith atom of one fragment and the jth atom of another one, $R_{i/j}$ is the covalent radius of atom i/j,[52] and c_{tough} is the toughness coefficient which is taken as input. The coefficient is controlled between 0.6 and 1.0 for efficient bond formations. The same action on ρ of all the remaining fragments is repeated one by one. The resulting geometry is optimized and saved with its energy value.

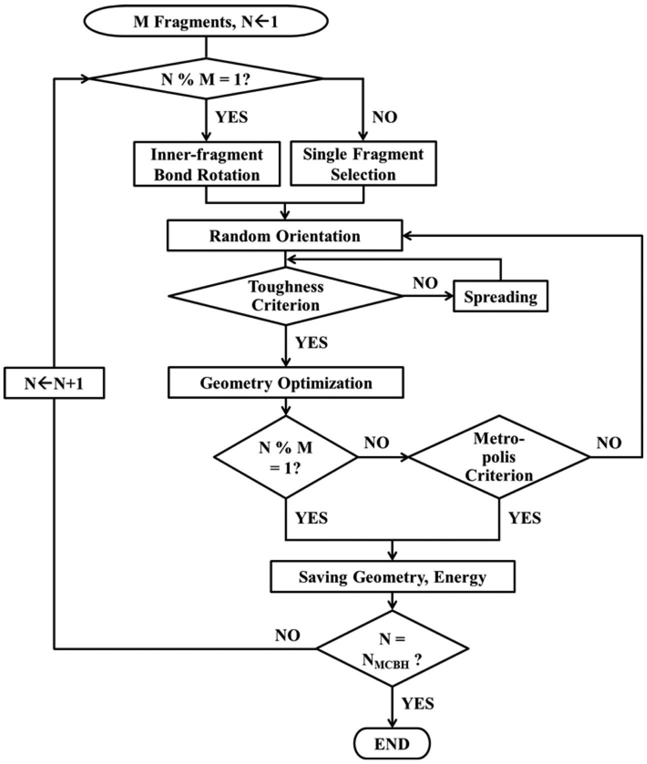


Figure 3. Flowchart of the present Monte Carlo basin-hopping method.

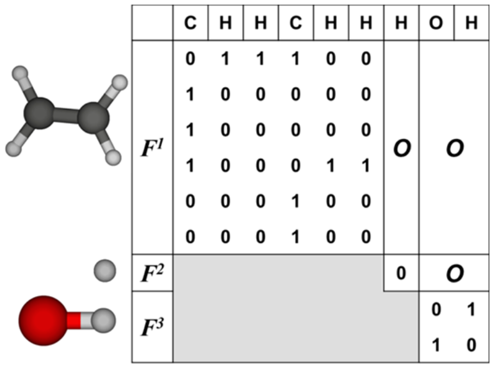


Figure 4. Example of the adjacency matrix for three fragments. The lower triangular blocks are shaded because the matrix is always symmetric.

For every ($pM + k$)th cycle ($2 \leq k \leq M$), only a single fragment is relocated along the same procedure with that of the ($pM + 1$)th cycle but with the standard Metropolis algorithm[53] as described in Figure 3. This step aims to closely search local minima through a short-range hopping between neighboring basins or between sub-basins as denoted by the dashed lines in Figure 1.

2.3. Postscreening. The resulting N_{MCBH} molecular structures after the MCBH cycles are filtered out by an energy criterion in the first place. If their energies are higher than the reactant energy plus a given tolerance energy ($E_{\mathrm{react}} + E_{\mathrm{tol}}$), they are discarded. However, there are yet many local minima to be removed. To this end, we transform molecular structures into a matrix form based on their bond connectivities that are used to analyze and classify the remaining ones. First, we define the so-called adjacency matrix of a molecule as the following:

$$A_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } r_{ij} \leq 1.1(R_i + R_j) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

where i and j denote atom indices of a given molecule. It shows the bond connectivity of the molecule. For a system with M molecular fragments, the adjacency matrix consists of M diagonal block matrices (F^{α}) and off-diagonal block matrices ($B^{\alpha\beta}$), ($\alpha, \beta = 1, 2, ... M, \alpha \neq \beta$) as illustrated in Figure 4. Then, we impose the following three criteria on each matrix of local minima to determine whether they are discarded or not.

The first criterion is that all atoms in a molecule should have equal or less number of chemical bonds than their maximum allowed values, for example, four for C, two for O, and one for H. The number of chemical bonds of an atom can be counted as the sum of matrix elements of the corresponding row in the adjacency matrix. If this number is larger than the given value, it is regarded as hypervalent and thus discarded. The second criterion is to determine whether given input fragments are broken or not. It can be discerned by comparing between the adjacent matrices of reactants and local minima. Their block diagonal matrices should be identical because the matrix itself denotes the internal bond connectivities of each fragment. Otherwise input fragments are broken during simulations and hence they are discarded. The third criterion is that new chemical bonds between input fragments must be formed. Otherwise the products are the same with the input reactants. Off-diagonal block matrices of the adjacent matrix are null for input fragments, but they have nonzero elements once a new chemical bond between the fragments is formed. Therefore, local minima are discarded if all off-diagonal block matrices are still zero after simulations.

Remaining local minima after the screening are classified in a way that different conformers or stereoisomers of the same molecule belong to the same group. For each group, local minima are sorted according to their energies and only the one with the lowest energy is put into namely a comprehensive set of intermediates of a given chemical reaction, which will be used to construct a reaction path.

We note that the adjacency matrix of molecular structures is useful for the postscreening, but it produces namely the permutational isomers, because the same molecule can have different adjacency matrices due to different numbering of atoms which is a basis of the matrix. Its simple example is shown in Figure 5. To rule out such permutational isomers, we define a revised Coulomb matrix as

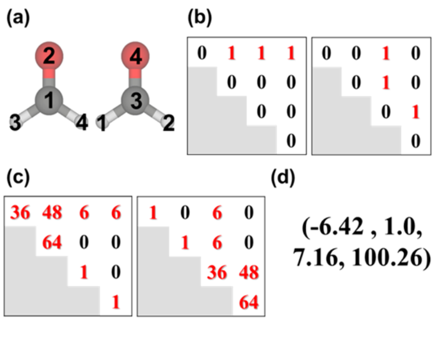


Figure 5. Simple example of permutational isomers. (a) Different atom numbering of the same molecule and (b) their corresponding adjacency matrices and (c) alternative Coulomb matrices, respectively. (d) Eigenvalues of the two alternative Coulomb matrices.

$$C_{ij} = \begin{cases} A_{ij} \cdot Z_i \cdot Z_j & \text{if } i \neq j \\ Z_i^{\,2} & \text{otherwise} \end{cases} \tag{3}$$

where A_{ij} is the element of the adjacency matrix and Z_i is the atomic number of atom i. The concept of the Coulomb matrix was inspired by its utilization in machine learning.[54,55] Moussa and Sadeghi have shown that the original Coulomb matrix does not allow identifying a molecular structure in a unique way.[56,57] However, the alternative Coulomb matrix in eq 3 is sufficient to distinguish molecules with different bond connectivity, because it explicitly includes the bond connectivity rather than the interatomic distances and hence recognizes molecules with the same bond connectivity but with different bond lengths as the same one. The proof of one-to-one correspondence between the alternative Coulomb matrix and the bond connectivity of molecules is available in the Supporting Information.

The alternative Coulomb matrix itself cannot distinguish permutational isomers, but the matrices of permutational isomers share the same eigenvalues (Figure 5). Thus, we compare the eigenvalues of the alternative Coulomb matrix for a given set of intermediates so as to screen out the permutational isomers. In fact, the alternative Coulomb matrix has all the information on the adjacency matrix. However, comparison of elements between two matrices is usually faster than their diagonalization. For the sake of computational efficiency, therefore, molecules are first screened by the comparison of their adjacency matrices and only the remaining ones after that are further analyzed by the alternative Coulomb matrix. After the postscreening, only $N_{\mathrm{PP}}^{\mathrm{IM}}$ local minima are left out of N_{MCBH} and they are finally used to construct a reaction path.

2.4. Automatic Ordering of Intermediates along a Reaction Path.

Automatic ordering of resulting intermediates along a reaction path from given reactants to target products is essential for our method to be practical. Though information on the energy barriers between the intermediates is key factors for finding energetically favored reaction path, it is too demanding to evaluate them for all possible pairs. Thanks to the aid of graph-theoretic analysis such as the Dijkstra algorithm, we are able to extract only a few plausible reaction paths without such information, which will be eventually confirmed by additional energy barrier calculations using conventional methods.

To this end, we first regard individual $N_{\mathrm{PP}}^{\mathrm{IM}}$ local minima as vertices in a graph. Then, we define the distance D_{XY} between two vertices X and Y as the Frobenius norm:

$$D_{XY} = \sqrt{\sum_{ij} (A_{ij}^{X} - A_{ij}^{Y})^2} \tag{4}$$

where $A_{ij}^{X/Y}$ is the adjacency matrix of X/Y. The more the bonds are broken or formed for the chemical transformation between X and Y, the larger the Frobenious norm. Therefore, the distance in the graph implies the kinetic energy barrier for the chemical transformation and thus can be used as a weight factor in the Dijkstra algorithm which finds the shortest pathway from given reactants to target products. All possible pairs from a database of intermediates are initially connected, but after calculating the distances between them, we remove edges (i.e., connection) whose distances are longer than D_{MAX} that is a given threshold value. Then, the Dijkstra algorithm[58] is applied to find the shortest path from the reactants to the products.

The above procedures were implemented using the Python 2.7[59] with common mathematics libraries such as NumPy and SciPy.[60] It is a standalone code that can be linked with any programs for electronic structure calculations. In the present study we adopted specifically the DFTB+ program[61] for geometry optimizations and energy calculations of molecules, because it enables one to perform rapid calculations for many systems.

3. RESULTS

As example studies, we applied the present method to the intramolecular aldol reaction of triketones, the $HCo(CO)_3$-catalyzed hydroformylation, and the $[Ti^{IV}]$-catalyzed hydroamination of alkynes. The geometries of input fragments were optimized in advance using the GAUSSIAN 09 program suite.[62] The B3LYP hybrid functional[63] was employed with the 6-311g(d,p) basis set for the organic elements (C, H, O, and N), and the LANL2DZ basis set with effective core potential (ECP)[64] for Ti. The same method was used to investigate the energy profile of a reaction path constructed from the MCBH samplings. For DFTB calculations, the mio-1−1 parameter set[65] was used for the pairwise potential parameters between C, H, O, and N atoms, and the trans3d-0-1 parameter set[66] was employed for those between Ti and the organic elements. The maximum numbers of self-consistent charge (SCC) and geometry relaxation cycles were set to 100, respectively. The SCC tolerance of 10^{-5} was used and the maximum force value for the geometry convergence was 0.005 hartree/Bohr.

3.1. Intramolecular Aldol Reaction of Triketones.

Conformers of a molecule via bond rotations form local minima within a superbasin on a PES as denoted by the shaded region in Figure 1. They share the same bond connectivity but with different geometric representations. They can hardly be found by conventional MCBH samplings which are rather suitable for the random distribution of individual objects. Some methods have been developed for conformational analysis of molecules.[67,68] To find the most stable conformer, we modified the MCBH method in a way that it samples local minima within the conformational space of a given molecule through random intramolecular bond rotations. Intramolecular aldol reaction is a good example to test the present method for the conformational analysis, because their products can be obtained merely through intramolecular bond rotations. The nonane-2,3,8-trione molecule was chosen, because it yields six products. The MCBH routine randomly rotates the C−C bonds of the reactant **R** indicated by the arrows in Figure 6. The molecule was not allowed to be broken, but protons can be dissociated during geometry relaxations. The tolerance energy E_{tol} was set to 50 kcal/mol. The MCBH cycle was repeated 30 000 times (1.17×10^6 number of gradient evaluations) and 43 local minima including all the products were finally obtained after postscreening, which is extremely effective. Figure 6 shows the resulting six products formed via bond rotations and their relative energies with respect to the reactant. This result confirms that our method is useful to search local minima in a conformational space through random bond rotations.

3.2. $HCo(CO)_3$-Catalyzed Hydroformylation.

The full catalytic cycle of the $HCo(CO)_3$-catalyzed hydroformylation was proposed by Heck and Breslow,[69] and recently, it was further elucidated by the artificial force induced reaction (AFIR) method.[70] This reaction is ideal to evaluate the efficiency of our method, because the system has only 18 atoms but the known reaction path involves 4 intermediates between

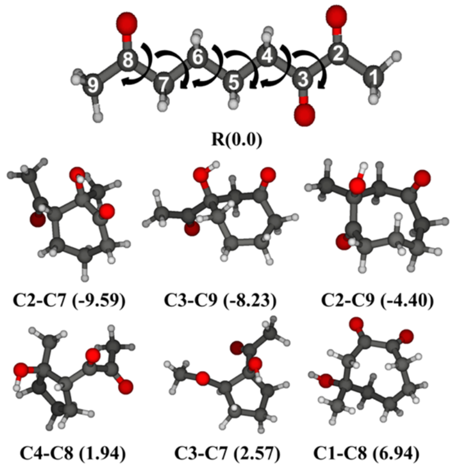


Figure 6. Molecular structures of the reactant (**R**) and the products obtained from the MCBH method. C*i*−C*j* means the product of the aldol reaction between the *i*th and *j*th C atoms of the reactant **R**. The values in the parentheses indicate the relative energy of each molecule with respect to that of the reactant (kcal/mol).

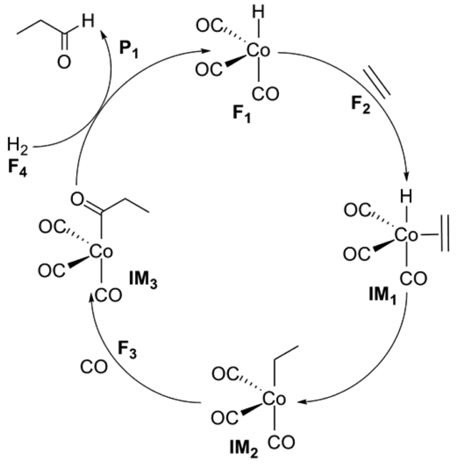


Figure 7. Catalytic cycle of the hydroformylation. $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, and $\mathbf{F}_4$ represent the four fragments of reactants. $\mathbf{IM}_1$, $\mathbf{IM}_2$, $\mathbf{IM}_3$, and $\mathbf{P}_1$ are the known intermediates and the product, respectively.

the reactants and the products as depicted in Figure 7. For comparison, we repeated identical calculations with different options for the molecular fragmentation. We note that our method without any fragmentation step is equivalent to the standard basin-hopping. For each case, we investigated the number of final local minima ($N_{\mathrm{PP}}^{\mathrm{IM}}$), the ratio between the number of converged geometries and the total number of initial geometries ($N_{\mathrm{Opt}}^{\mathrm{Conv}}/N_{\mathrm{Opt}}$), the number of gradient evaluations (N_{Grad}), and whether or not the sampling found the four intermediates and the products. In all the calculations, C_{tough}, the temperature for Metropolis selection, and E_{tol} were set to 0.7, 298.15 K, and 20 kcal/mol, respectively.

Table 1. Efficiency of MCBH Samplings for the Hydroformylation Reaction with Various Input Fragments[a]

input fragments	N_{MCBH}	$N_{\mathrm{PP}}^{\mathrm{IM}}$	$N_{\mathrm{Pot}}^{\mathrm{Conv}}/N_{\mathrm{Opt}}$	N_{Grad}	all IMs found?
$Co(CO)_3H$ + C_2H_4 + CO + H_2	1000	102	393/2304	2.26 × 10^5 (226)	yes
$Co(CO)_3H$ + C_2H_4 + CO + H_2[b]	1000	165	394/2250	2.21 × 10^5 (221)	yes
Co + 6C + 4O + 7H	1000	283	22/2908	2.91 × 10^5 (291)	no
Co + 6C + 4O + 7H	10 000	2663	250/29662	2.96 × 10^6 (296)	no

[a] The values in the parentheses of the N_{Grad} column mean $N_{\mathrm{Grad}}/N_{\mathrm{MCBH}}$. [b] All bonds of fragments were allowed to be broken.

Table 1 summarizes the results. The first column shows the input fragments for each sampling. The first and second samplings had the same input fragments, but the former only allowed for the bond dissociation of hydrogen, while the latter allowed for the dissociation of all bonds. Both cases found all the intermediates as well as the products. On the contrary, the third sampling used individual atoms as the input fragments. After 1000 cycles of the MCBH routine, the third one produced 283 local minima, but none of them corresponds to the known intermediates and the products. Moreover, only 22 out of 2908 geometry optimizations were properly converged. At the fourth sampling, we extended the standard basin-hopping to 10 000 cycles but still failed to find any known intermediates. The number of gradient calculations reflects total computational costs.[20] The simulation without fragmentation (the third and fourth rows) requires more gradient calculations due to smaller $N_{\mathrm{Opt}}^{\mathrm{Conv}}/N_{\mathrm{Opt}}$ as compared to the cases with fragmentation (the first and second rows). Moreover, the values of $N_{\mathrm{Grad}}/N_{\mathrm{MCBH}}$ for the former cases are significantly larger than those for the latter cases. We expect that if the maximum number of geometry relaxation cycles increases, such a difference becomes larger. This example manifests that appropriate fragmentation of reactants is essential not only to reduce computational costs but also to efficiently find key intermediates relevant to a reaction path.

3.3. [Ti^{IV}]-Catalyzed Hydroamination of Alkynes.

The zirconium[71] and the titanium complexes[72−74] have been used for the hydroamination reaction between alkynes and amines, whose mechanism was introduced by Straub and Bergman.[75] The proposed azatitanacyclobutene intermediate ($\mathbf{IM}_1$) in Figure 8a was isolated experimentally and verified through DFT calculations.[76] We examined whether or not our MCBH method finds the proposed reaction path shown in Figure 8a. Using the optimized input fragments $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$, the basin-hopping sampling was performed. Different alkyl groups were assigned for $\mathbf{R}_1$ and $\mathbf{R}_2$ of the fragment $\mathbf{F}_2$ in order to check if the samplings can yield two distinct sets of intermediates. The fragment $\mathbf{F}_3$ was allowed to be broken during the simulation. Different sets of input parameters were used to evaluate its performance, as summarized in Table 2.

We first carried out the MCBH cycling 100 000 times for each set of the tolerance energy E_{tol}, the toughness coefficient C_{tough}, and the temperature given in the top three rows of Table 2 and obtained 390, 607, and 543 local minima after postscreening, respectively. $\mathbf{P}_1$ and $\mathbf{IM}_1$ in Figure 8a were found for all three cases, but none of them produced $\mathbf{IM}_2$, which can be understood from much less probability of finding a stable structure by combining three fragments at once than the case with two fragments. In particular, $\mathbf{F}_1$ has a complicated structure because of the bulky ligands which may hinder a new bond formation of $\mathbf{F}_1$ with the other fragments. Therefore, we performed the secondary simulation with $\mathbf{IM}_1$ and $\mathbf{F}_3$ for $\mathbf{R}_1$ = **Et** and $\mathbf{R}_2$ = **Me** as new input fragments, yielding 129 local minima out of 15 000 samples after postscreening (the fourth

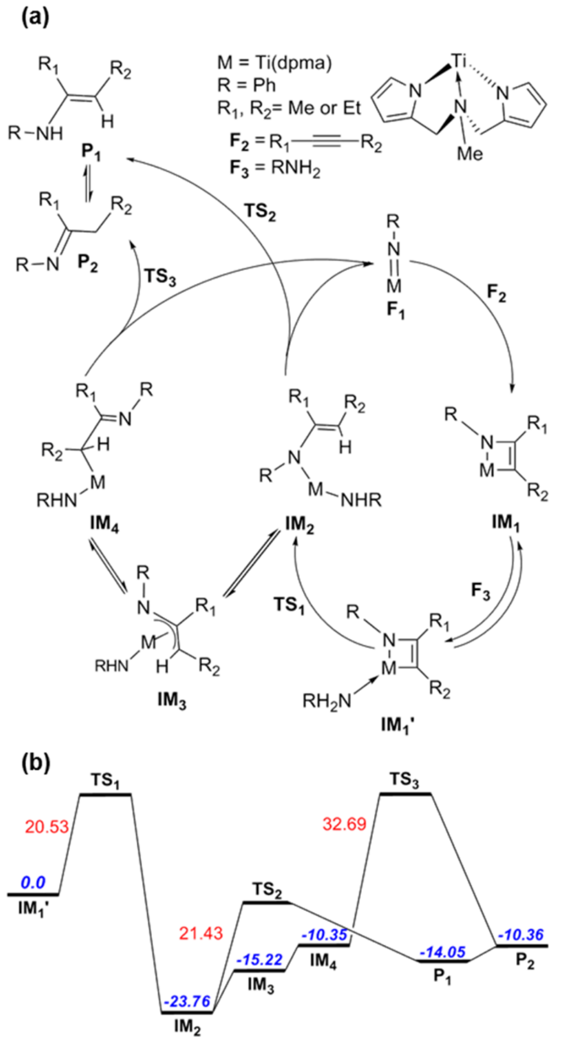


Figure 8. (a) Extended Bergman mechanism of the hydroamination based on the result from the MCBH samplings combined with the SCC-DFTB. (b) Energy profile calculated at the level of B3LYP-DFT for the two paths from IM_1' to the products P_1 and P_2. The relative energy values with respect to that of IM_1' are given in kilocalories per mole.

row of Table 2). In this case, $\mathbf{IM_2}$ was found 5 times, which means that in average 3000 samplings are sufficient to find it.

In addition, we found $\mathbf{IM_1'}$, $\mathbf{IM_3}$, $\mathbf{IM_4}$, and $\mathbf{P_2}$ depicted in Figure 8a from the secondary simulation, which seem to be good candidates of intermediates along a reaction path. Accordingly, we extended the original Bergman mechanism using them and investigated the energy profile along the pathways (Figure 8b). $\mathbf{IM_1'}$ is formed as F_3 approaches to $\mathbf{IM_1}$ without energy barrier between them. Then, the original mechanism starting from $\mathbf{IM_1}$ (or $\mathbf{IM_1'}$) to $\mathbf{IM_2}$ to $\mathbf{P_1}$ shows the energy barrier of 20.53 and 21.43 kcal/mol, respectively. In experiments, the major product is $\mathbf{P_2}$, while the original mechanism produces $\mathbf{P_1}$. Therefore, there should be a rapid conversion from $\mathbf{P_1}$ to $\mathbf{P_2}$.[73] We found that $\mathbf{IM_4}$ provides a direct way to yield $\mathbf{P_2}$ as shown in Figure 8a. $\mathbf{IM_4}$ can be formed via $\mathbf{IM_3}$ from $\mathbf{IM_2}$. We were not able to locate transition states between these intermediates, but we guess that these TSs are very low and not important in this reaction. The energy barrier between $\mathbf{IM_4}$ and $\mathbf{P_2}$ is 32.69 kcal/mol, which is higher than that of the original path, indicating that the original pathway is favorable than the new one.

To illustrate how a comprehensive set of intermediates obtained from the MCBH sampling can be used to find a reaction path without energy barrier calculations, we used the Dijkstra algorithm. We constructed a graph from the 40 intermediates at the last row in Table 2 and the products ($\mathbf{P_1}$ and $\mathbf{P_2}$) according to the procedure introduced in section 2.4. The D_{MAX} value was set to $\sqrt{8} \approx 2.8284$. The Sage mathematics program was used to draw the graph[77] that is shown in Figure 9. The thickness of edges indicates the size of the Frobenius norm defined in eq 4. The white vertices denote the intermediates or the products appeared in Figure 8. The arrows in the lower panel indicate the shortest paths from $\mathbf{IM_1}$ to each vertex found by the Dijkstra algorithm. Surprisingly, the shortest paths from $\mathbf{IM_1}$ to $\mathbf{P_1}$ and $\mathbf{P_2}$ pass through the intermediates in Figure 8a, respectively, as shown in the lower

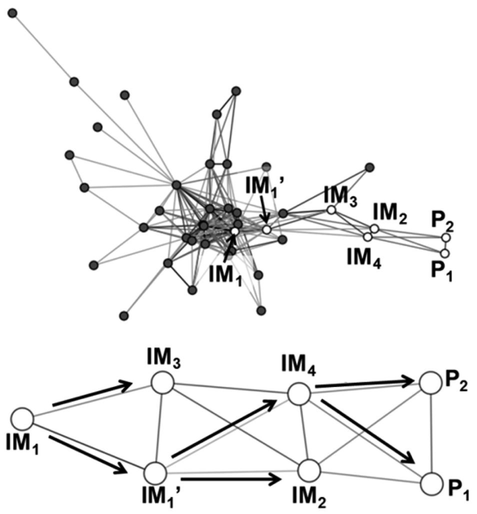


Figure 9. Graph for the 42 intermediates including the two products P_1 and P_2 obtained from MCBH samplings. One isolated vertex was omitted from the graph. The white vertices indicate the intermediates and the products appeared in Figure 8. The thickness of edges represents the size of the Frobenius norm defined in eq 4. The arrows in the lower panel indicate the shortest paths from IM_1 to each vertex found by the Dijkstra algorithm.

Table 2. Results of MCBH Samplings for the Hydroamination Reaction[a]

input fragments	N_{MCBH}	N_{PP}^{IM}	c_{tough}	E_{tol}	temperature	N_{Grad}
$F_1 + F_2 + F_3$	100000	390	0.7	20.0	298.15	2.18×10^7 (218)
$F_1 + F_2 + F_3$	100000	607	0.7	50.0	1000.0	2.05×10^7 (205)
$F_1 + F_2 + F_3$	100000	543	0.6	50.0	1000.0	2.15×10^7 (215)
IM_1-EtMe[b] + F_3	15000	129	0.6	50.0	1000.0	2.45×10^6 (164)
IM_1-MeEt[b] + F_3	6500	40/765[c]	0.6	50.0	1000.0	1.00×10^6 (155)

[a] The values in the parentheses of the N_{Grad} column mean N_{Grad}/N_{MCBH}. [b] Secondary MCBH. [c] All bonds of fragments were allowed to be broken.

panel of Figure 9. Furthermore, we stress that any paths toward the two products must include a subset of the white vertices by all means, which shows the reliability of the present approach.

Finally, we discuss the efficiency of molecular fragmentation. To evaluate actual computational costs, we counted the number of gradient calculations for each sampling as shown in Table 2. The first three cases required about 210 gradient calculations per each MCBH cycle, while the secondary sampling needed 164 times. Comparing the total computational costs, the secondary MCBH found all the intermediates within 2.45 × 10^6 N_{Grad}, while the first three trials still failed to find $\mathbf{IM_2}$ even with nine times more gradient calculations, indicating that the secondary MCBH sampling would be essential for complicated systems. The fifth row of Table 2 shows the input parameters of another secondary MCBH samplings with $\mathbf{IM_1}$ and $\mathbf{F_3}$ for $\mathbf{R_1}$ = **Me** and $\mathbf{R_2}$ = **Et**. When each fragment is not allowed to be broken during the samplings, 40 local minima including all intermediates in Figure 8a are obtained out of 6500 samples, whereas the case for allowing for the bond dissociation of all atoms produces 765 local minima. These results clearly demonstrate the importance of the fragmentation scheme for computational efficiency.

4. CONCLUSIONS

The basin-hopping sampling is suitable for searching local minima of various systems such as atomic and molecular clusters. It can be applied to finding reaction intermediates, since they are also local minima which consist of a reaction path. However, degrees of freedom of reaction intermediates are so large that their stochastic sampling is impractical, when the basin-hopping method is combined with quantum mechanical methods to deal with formation and dissociation of chemical bonds. In this regard, we adopted the following strategies to dramatically enhance its computational efficiency. First of all, we fragment reactant molecules to reduce the dimension of a given potential energy surface. Chemical bonds between atoms in a fragment are typically not allowed to be broken, but one can also allow it to incorporate natural bond formation or dissociation. This fragmentation is particularly effective for transition metal catalysts with bulky ligands as can be seen from the third example of this study. Despite the reduced degrees of freedom, it still requires many samplings to obtain a comprehensive set of intermediates necessary for the construction of a reaction pathway. To extract only meaningful local minima from many samples, we devised an efficient postprocessing scheme using the graph theory with a matrix representation of molecular structures. It enables us to readily classify molecules according to their bond connectivities and thus to help the effective screening of resulting local minima toward desirable intermediates. For rapid samplings, we combined the basin-hopping method with a density functional tight binding code, but more reliable computational methods can also be used for accuracy.

The three example studies show that our method is useful to search conformers of a molecule via stochastic bond rotations as well as to sample a comprehensive set of intermediates for chemical reactions. Sampling with two fragments is much more efficient than that with three fragments. Therefore, the secondary simulation with the results from a first simulation may be required for complicated reactions involving more than two input fragments to reduce computational costs. The resulting intermediates after postscreening can be used to find possible reaction pathways using the graph theory such as the Dijkstra algorithm, which are subsequently verified through additional computations for transition state analysis. The third example study shows that the distance definition based on the simple intuition for kinetic energy barrier (eq 4) was very successful to find a correct reaction path. We believe that it can be tuned by incorporating more chemical knowledge for more effective and reliable calculations.

However, any stochastic sampling method like the present one inherently requires many trials no matter how effective it is, because it cannot guarantee 100% probability to find a designated target structure within a finite number of samplings. Therefore, a deterministic sampling method would be desirable. In future publications, we will introduce such a deterministic method based on the graph theory which is currently under development.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Numerical procedure of bond rotation and random orientation of fragments; proof of one-to-one correspondence between the alternative Coulomb matrix and the bond connectivity of molecules. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: wooyoun@kaist.ac.kr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Basic Science Research Programs (NRF-2012R1A1A1004154), the EDISON [EDucation-research Integration through Simulation On the Net] Program (NRF-2012M3C1A6035359), and MIR Center (20090083525) through the NRF funded by the Korea government [MSIP]. The authors would like to acknowledge the support from KISTI supercomputing center through the strategic support program for the supercomputing application research. (No. KSC-2012-C1-026, KSC-2013-C1-028). W.Y.K. is also grateful for financial support from Chung-Am Fellowship.

■ REFERENCES

(1) Goedecker, S. *J. Chem. Phys.* **2004**, *120*, 9911.
(2) Wales, D. J. *Phys. Biol.* **2005**, *2*, S86.
(3) Wales, D. J.; Bogdan, T. V. *J. Phys. Chem. B* **2006**, *110*, 20765.
(4) Woodley, S. M.; Catlow, R. *Nat. Mater.* **2008**, *7*, 937.
(5) Wales, D. J. *Science* **1999**, *285*, 1368.
(6) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111.
(7) Leary, R. H.; Doye, J. P. K. *Phys. Rev. E* **1999**, *60*, 6320.
(8) Doye, J. P. K.; Wales, D. J.; Miller, M. A. *J. Chem. Phys.* **1998**, *109*, 8143.
(9) Iwamatsu, M.; Okabe, Y. *Chem. Phys. Lett.* **2004**, *399*, 396.
(10) Zhan, L.; Piwowar, B.; Liu, W.-K.; Hsu, P. J.; Lai, S. K.; Chen, J. Z. Y. *J. Chem. Phys.* **2004**, *120*, 5536.
(11) Li, F.; Jin, P.; Jiang, D.; Wang, L.; Zhang, S. B.; Zhao, J.; Chen, Z. *J. Chem. Phys.* **2012**, *136*, 074302.
(12) Wang, J.; Zhou, X.; Wang, G.; Zhao, J. *Phys. Rev. B* **2005**, *71*, 113412.
(13) Yoo, S.; Zeng, X. C. *J. Chem. Phys.* **2003**, *119*, 1442.
(14) Bandyopadhyay, P. *J. Chem. Phys.* **2008**, *128*, 134103.
(15) Do, H.; Besley, N. A. *J. Chem. Phys.* **2012**, *137*, 134106.
(16) Hodges, M. P.; Wales, D. J. *Chem. Phys. Lett.* **2000**, *324*, 279.
(17) Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett.* **1998**, *286*, 65.

(18) White, R. P.; Mayne, H. R. *Chem. Phys. Lett.* **1998**, *289*, 463.
(19) Verma, A.; Schug, A.; Lee, K. H.; Wenzel, W. *J. Chem. Phys.* **2006**, *124*, 044515.
(20) Kusumaatmaja, H.; Whittleston, C. S.; Wales, D. J. *J. Chem. Theory Comput.* **2012**, *8*, 5159.
(21) Cho, Y.; Min, S. K.; Yun, J.; Kim, W. Y.; Tkatchenko, A.; Kim, K. S. *J. Chem. Theory Comput.* **2013**, *9*, 2090.
(22) Hamad, S.; Catlow, C. R. A.; Woodley, S. M.; Lago, S.; Mejías, J. A. *J. Phys. Chem. B* **2005**, *109*, 15741.
(23) Zhan, L.; Chen, J. Z. Y.; Liu, W.-K.; Lai, S. K. *J. Chem. Phys.* **2005**, *122*, 244707.
(24) Barcaro, G.; Fortunelli, A.; Rossi, G.; Nita, F.; Ferrando, R. *J. Phys. Chem. B* **2006**, *110*, 23197.
(25) Harding, D.; Mackenzie, S. R.; Walsh, T. R. *J. Phys. Chem. B* **2006**, *110*, 18272.
(26) Gao, Y.; Bulusu, S.; Zeng, X. C. *ChemPhysChem* **2006**, *7*, 2275.
(27) Hsu, P. J.; Lai, S. K. *J. Chem. Phys.* **2006**, *124*, 044711.
(28) Kim, H. G.; Choi, S. K.; Lee, H. M. *J. Chem. Phys.* **2008**, *128*, 144702.
(29) Maeda, S.; Ohno, K. *J. Phys. Chem. A* **2005**, *109*, 5742.
(30) Ohno, K.; Maeda, S. *J. Phys. Chem. A* **2006**, *110*, 8933.
(31) Berente, I.; Náray-Szabó, G. *J. Phys. Chem. A* **2006**, *110*, 772.
(32) Maeda, S.; Morokuma, K. *J. Chem. Theory Comput.* **2011**, *7*, 2335.
(33) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2012**, *8*, 2215.
(34) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2013**, *9*, 1838.
(35) Lankau, T.; Yu, C.-H. *J. Chem. Phys.* **2013**, *138*, 214102.
(36) García-Domenech, R.; Galvez, J.; de Julian-Ortiz, J. V.; Pogliani, L. *Chem. Rev.* **2008**, *108*, 1127.
(37) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334.
(38) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(39) Pogliani, L. *Chem. Rev.* **2000**, *100*, 3827.
(40) Shelley, C. A. *J. Chem. Inf. Model.* **1983**, *23*, 61.
(41) Hansen, P. J.; Jurs, P. C. *J. Chem. Educ.* **1988**, *65*, 661.
(42) Hu, C.-Y.; Xu, L. *Anal. Chim. Acta* **1994**, *298*, 75.
(43) Benecke, C.; Grund, R.; Hohberger, R.; Kerber, A.; Laue, R.; Wieland, T. *Anal. Chim. Acta* **1995**, *314*, 141.
(44) Sinanoğlu, O. *J. Am. Chem. Soc.* **1975**, *97*, 2309.
(45) Temkin, O. N.; Zeigarnik, A. V.; Bonchev, D. G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 729.
(46) Ratkiewicz, A.; Truong, T. N. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 36.
(47) Hansen, P. J.; Jurs, P. C. *J. Chem. Educ.* **1988**, *65*, 574.
(48) Randić, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
(49) Pogliani, L.; de Julián-Ortiz, J. V. *Chem. Phys. Lett.* **2004**, *393*, 327.
(50) Wolfe, J. P.; Wagaw, S.; Marcoux, J.-F.; Buchwald, S. L. *Acc. Chem. Res.* **1998**, *31*, 805.
(51) Lam, K. C.; Marder, T. B.; Lin, Z. *Organometallics* **2007**, *26*, 758.
(52) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. *Dalton Trans.* **2008**, 2832.
(53) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(54) Montavon, G.; Hansen, K.; Fazli, S.; Rupp, M.; Biegler, F.; Ziehe, A.; Tkatchenko, A.; von Lilienfeld, O. A.; Müller, K. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 449.
(55) Hansen, K.; Biegler, F.; Fazli, S.; Rupp, M.; Sche, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Mu, K. *J. Chem. Theory Comput.* **2013**, *9*, 3404.
(56) Moussa, J. E. *Phys. Rev. Lett.* **2012**, *109*, 059801.
(57) Sadeghi, A.; Ghasemi, S. A.; Schaefer, B.; Mohr, S.; Lill, M. a; Goedecker, S. *J. Chem. Phys.* **2013**, *139*, 184118.
(58) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269.
(59) Rossum, G. *Python reference manual*; CWI(Centre for Mathematics and Computer Science): Amsterdam, The Netherlands, 1995.
(60) Oliphant, T. E. *Comput. Sci. Eng.* **2007**, *9*, 10.
(61) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678.
(62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc., Wallingford CT, 2009.
(63) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(64) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.
(65) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(66) Zheng, G.; Witek, H. a.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Köhler, C.; Frauenheim, T. *J. Chem. Theory Comput.* **2007**, *3*, 1349.
(67) Chang, C.-E.; Gilson, M. K. *J. Comput. Chem.* **2003**, *24*, 1987.
(68) Brain, Z. E.; Addicoat, M. A. *J. Chem. Phys.* **2011**, *135*, 174106.
(69) Heck, R. F.; Breslow, D. S. *J. Am. Chem. Soc.* **1961**, *83*, 4023.
(70) Maeda, S.; Morokuma, K. *J. Chem. Theory Comput.* **2012**, *8*, 380.
(71) Baranger, A. M.; Walsh, P. J.; Bergman, R. G. *J. Am. Chem. Soc.* **1993**, *115*, 2753.
(72) Cao, C.; Shi, Y.; Odom, A. L. *J. Am. Chem. Soc.* **2003**, *125*, 2880.
(73) Odom, A. L. *Dalton Trans.* **2005**, 225.
(74) Banerjee, S.; Shi, Y.; Cao, C.; Odom, A. L. *J. Organomet. Chem.* **2005**, *690*, 5066.
(75) Straub, B. F.; Bergman, R. G. *Angew. Chem.* **2001**, *113*, 4768.
(76) Vujkovic, N.; Fillol, J. L.; Ward, B. D.; Wadepohl, H.; Mountford, P.; Gade, L. H. *Organometallics* **2008**, *27*, 2518.
(77) Stein, W. A. et al. *Sage Mathematics Software*, version 6.1.1. 2014; http://www.sagemath.org (accessed March 31, 2014).