Multiscale Approach to Explore the Potential Energy Surface of Water Clusters (H$_2$O)$_n$ $n \leq 8$

Quoc Chinh Nguyen,[†] Yew Soon Ong,[‡] Harold Soh,[§] and Jer-Lai Kuo*,[†]

School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore, School of Computer Engineering, Nanyang Technological University, 639798, Singapore, and Advanced Computing, Institute of High Performance Computing, Agency for Science, Technology and Research (ASTAR), Singapore 117528, Singapore

We propose a multiscale method to explore the energy landscape of water clusters. An asynchronous genetic algorithm is employed to explore the potential energy surface (PES) of OSS2 and TTM2.1-F models. Local minimum structures are collected on the fly, and the ultrafast shape recognition algorithm was used to remove duplicate structures. These structures are then refined at the B3LYP/6-31+G* level. The number of distinct local minima we found (21, 76, 369, 1443, and 3563 isomers for $n = 4-8$, respectively) reflects the complexity of the PES of water clusters.

Introduction

Neutral water clusters have been investigated extensively for a long period because they provide important understanding of properties of water molecules in aqueous media.[1–16] A number of empirical potential models have been developed for estimating the interaction energies and to reproduce the ground-state structures of first-principle calculation. To date, many researchers have focused on using the global minima to validate and compare the different potential models. Lee et al. applied a simulated annealing method with the empirical potential function of Cieplak, Kollman, and Lybrand to optimize water clusters up to $n = 20$.[4] With the use of basin hopping, Wales and Hodges studied the TIP4P[17] and TIP5P[18] potentials for $n \leq 21$ and performed a comparison of the structures and formation energies obtained against the MP2 calculation. More recently, Bandow and Hartke[10] developed a highly parallel evolutionary algorithm to study water clusters up to $n = 34$ on TIP4P and TTM2-F potentials. It is worth noting that these earlier methods generally focus on global optimization with empirical models for the large-sized system in the efforts to bridge the gap from single molecule to bulk materials.

Recently, Maeda and Ohno[11] proposed a full first-principle-based approach for studying the structures and thermodynamic transitions of water clusters by searching for isomers instead of focusing only on the global minima. The authors have developed a scaled hypersphere search (SHS) method that employs an "uphill-walking" technique to locate the isomers sequentially. The set of 168 identified isomers for (H$_2$O)$_8$ at the B3LYP/6-31+G(d,p) level was considered to be relatively large. The shortcoming of the proposed method, however, is its complete reliance on first-principle and second-order derivative calculations. The high computational demand of first-principle calculations renders this method computationally infeasible for exploring the quantum chemistry potential energy surfaces (PES) with a large number of local minima.

In this study, we presented an alternative approach that synergizes empirical model with first-principle calculations. Our aim is to benefit from the low computational cost of empirical models by coupling it with first-principle calculation to explore the PES of water clusters at the quantum mechanical level. Consequently, the search for isomers becomes an efficient yet highly accurate process. We demonstrated the approach with two sophisticated, flexible potential models, OSS2 and TTM2.1-F. The isomers of (H$_2$O)$_n$ in the range of $n = 4-8$ of these models were collected extensively using an asynchronous parallelized genetic algorithm and subsequently refined using B3LYP/6-31+G*. The details are described in the next section, followed by the results and discussions.

Methodology

Empirical Models. TTM2-F is a flexible, polarizable, Thole-type interaction potential developed by Bernham and co-workers.[5–8,19] Although the model was parameterized using water dimer only, it was shown to reproduce the binding energies that are in close agreement to MP2 calculations for (H$_2$O)$_n$ ($n = 2-6$)[8] and (H$_2$O)$_{20}$.[9] In this study, we first considered the TTM2.1-F empirical model,[20] which is a revised version of the original TTM2-F, that reportedly resolves issues relating to the dipole moment of individual water and short intermolecular interactions.

The second sophisticated empirical model considered in this study is the OSS2 potential.[21] It was developed to describe water as a participant in ionic chemistry. Since it is an atomic and polarizable potential, OSS2 is useful for studying the proton-transfer reaction and water disassociation. While originally designed for H$^+$(H$_2$O)$_n$, it was later demonstrated by some researchers to model small-sized pure water clusters well.[22]

The close agreement of both TTM2.1-F and OSS2 with the first-principle method on small-sized water clusters inspired our idea of engaging empirical models to enhance the exploration of PES at the quantum level. Because the PES landscape and the structures of water cluster isomers generated by these two models are sufficiently close to that of first-principle counterparts, it makes good sense to use their local minima as input to first-principle methods, thus reducing the computational efforts required.

* Corresponding author. Fax: +65 6316 2967. E-mail: jlkuo@ntu.edu.sg.
† School of Physical and Mathematical Sciences, Nanyang Technological University.
‡ School of Computer Engineering, Nanyang Technological University.
§ Agency for Science, Technology and Research.

**6258** *J. Phys. Chem. A, Vol. 112, No. 28, 2008*

Nguyen et al.
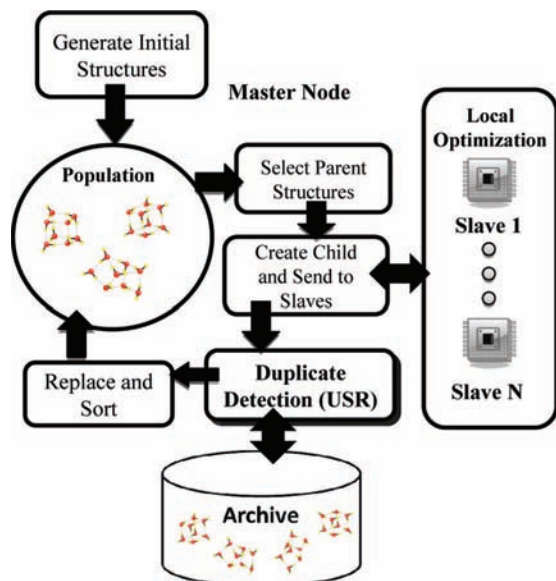
Figure 1. Flowchart of the asynchronous genetic algorithm (GA).

Exploring PES Using a Genetic Algorithm. An asynchronous genetic algorithm (GA) is used for exploring the PES to locate the global and local minima of the empirical models. Similar to prior works,[10,23] our real-coded GA explores directly the configuration space based on a Cartesian representation. The asynchronous, nongenerational-based GA, which was successfully applied elsewhere for optimizing water clusters,[10] is used in this study: the master node manages the population of candidate structures and performs the genetic operators while the slave nodes locally optimize the structures sent by the master.

The flowchart of our GA is depicted in Figure 1. After initialization, parent structures are rank-selected from the population before undergoing crossover and/or mutation to generate the offspring. In crossover, the selected structures are first cut into two equal-sized substructures using a random plane across their centers of mass. The substructures are then swapped to form new, possibly lower energy, structures. In mutation, both molecular and atomic moves are applied to a maximum of 30% of the molecules in the cluster. Further, 10% of the molecules are randomly chosen to be rotated around the mass center. Constraints on bond lengths and bond angles are imposed on the molecular and atomic moves to avoid unnatural geometries.

Each individual offspring structure is sent to a slave node for local optimization. If the local optimization is successful, i.e., the root mean square of force is less than 0.00005 hartree/ Å, the resultant locally optimized structure is then inserted back into the GA population of the master node to compete for reproductive opportunities in the spirit of Lamarckian learning.[24–26] It is also archived for further analysis. This entire process repeats until a maximum number of iterations or the wall-clock time limit is reached.

In many cases, low-lying structures easily dominate the others in the population leading to premature search convergence. In the present context, for example, the consequence is that the crossover operator fails to generate any further improvements when the population of structures becomes homogeneous. To maintain diversity in the population, the ultrafast shape recognition (USR) metric, developed recently by Ballester and Richards[27] was used. USR provides an efficient means of measuring the similarity of cluster geometries. With the use of the Cartesian coordinates of two structures, the USR similarity index is computed. The similarity index is a scalar value in the range of
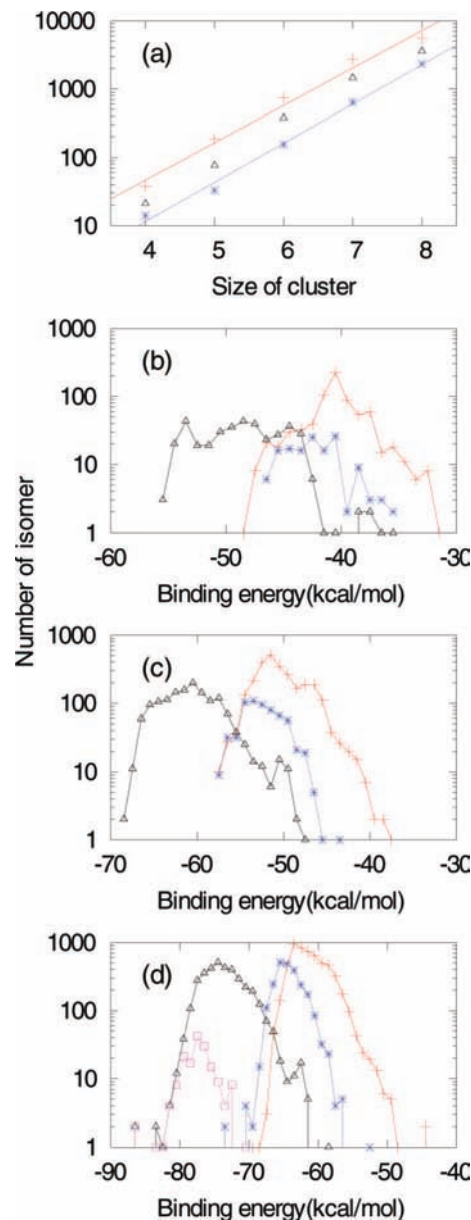
Figure 2. (a) Number of isomers of water clusters in the range of n = 4−8 of OSS2 (red crosses), TTM2.1-F (blue asterisks), and B3LYP/ 6-31+G* (black triangles). The exponential dependence of the number of isomers as a function of n can be estimated as $e^{1.26n}$ (red solid line) and $e^{1.32n}$ (blue dotted line) for OSS2 and TTM2.1F, respectively. The density of states (DOS) of $(H_2O)_6$, $(H_2O)_7$, and $(H_2O)_8$ in terms of binding energy (kcal/mol) are shown in (b−d), respectively. In subplot (d), the DOS of $(H_2O)_8$ originated from ref 11 is included for comparison (shown in pink squares).

[0, 1]. A value of 0 implies that the structures are completely dissimilar, whereas the other extreme indicates identical structures. In the present study, diversity is preserved by ensuring the similarity indices of all pairs of structures in the population are maintained above the threshold level of 0.93.

Throughout the search process, the issue of archiving duplicate local minima is also of concern. Since the number of local minima grows exponentially with cluster size, the local minima archive may become unnecessarily large and, hence, difficult to classify. The number of isomers grows to thousands in the present study. To ensure that only unique structures are archived, local minima with a USR similarity of 0.96 to any structure already in the archive are excluded.

Figure 3. Molecular structures of the five most stable isomers in OSS2, TTM2.1-F, and B3LYP/6-31+G*, at $n = 4-6$. The unit of the binding energy is reported in kcal/mol. Note that the binding energies of the B3LYP/6-31+G* isomers reoptimized using MP2/aug-cc-pvDZ are also reported and enclosed in parentheses.

First-Principle Calculation. Subsequently, all the archived isomers using the two empirical models are refined using a first-principle method. In particular, these archived isomers undergo geometrical optimization using Becke's three-parameter hybrid method[28] with the Lee, Yang, and Parr (B3LYP) functional[29] and the 6-31+G* basis set. The convergence criteria were set as the root mean square and the maximum component of gradient less than 0.0003 and 0.00045 hartree/Å, respectively. Furthermore, selected low-energy isomers (those shown in Figures 3 and 4) have been examined by the MP2 method with the aug-cc-pvDZ basis. All calculations were completed using the GAUSSIAN-03 package.[30]
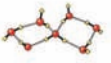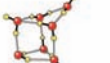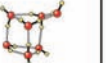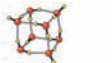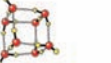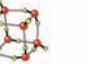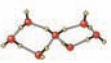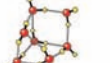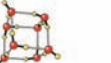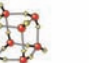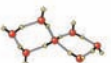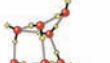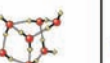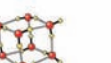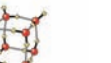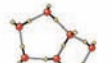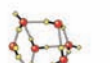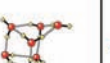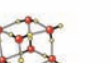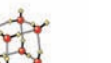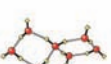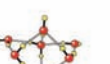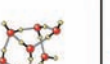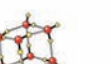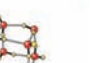
Results and Discussion

The number of distinct isomers is an important index to reflect the complexity of the PES. Although there is no strict rules on the number of isomers, N_{isomer}, with system size, it was demonstrated in LJ clusters and TIP models that N_{isomer} would grow exponentially with system size n, that is, $N_{isomer} \sim \exp(\alpha n)$, where α is a system-dependent constant.[1,12,31,32] Here, the number of isomers identified using our GA search on both OSS2 and TTM2.1F models are tabulated in Table 1. Note that since a value of 0.96 is used as the similarity threshold index to remove duplicate isomers, it is possible for overelimination of unique isomers to happen. Hence, our results may be regarded as a lower bound for the number of unique isomers that exist. Nevertheless, the numbers of isomers we collected increases rapidly from dozens to thousands in the range of $n = 4-8$. The exponential dependence of isomer size, n, is depicted clearly in Figure 2a, and by linear regression we estimated the values of α for OSS2 and TTM2.1F to be 1.26 and 1.32, respectively.

To evaluate the efficacy of using empirical potential models to guide the exploration of PES at the quantum level, the success

rate, the ratio between the number of successfully located isomer using B3LYP/6-31+G* and the initial number of isomers used as starting geometries, was computed. As shown in Table 1, upon relaxation using density functional theory (DFT), approximately 50% of the isomers found in OSS2 successfully converged to equivalent-topology isomers. On the other hand, TTM2.1-F displays a higher efficiency of >60%. It is worth noting that this also highlights the rugged PES of OSS2 over TTM2.1-F on $n = 4-8$. Nevertheless, by cross-referencing both empirical models simultaneously, the exploration on the PES of the first-principle method has led to the identification of far more isomers than would be possible when using only any one of the empirical models in isolation. For example, the set of $(H_2O)_7$ isomers found increased to 1443 isomers compared to 1175 and 406 on OSS2 and TTM2.1F alone, respectively.

The density of states (DOS) for $(H_2O)_6$, $(H_2O)_7$, $(H_2O)_8$ are illustrated in the Figure 2b−d, respectively, to reflect the distribution of isomers on the energy scale. Note that calculations based on B3LYP/6-31+G* lead to a shift in DOS to larger binding energy in comparison with OSS2 and TTM2.1-F. This is as expected since both OSS2 and TTM2.1F have been parameterized based on MP2 binding energies, and it is commonly acknowledged that the B3LYP method tends to overestimate the interaction energies. On the other hand, the overall features of DOS for all three models are similar. For example, in $(H_2O)_8$ the energy gaps separating the most stable cubic structure from the others are of similar values, with a Gaussian distribution of the DOS spanning the scale of ~20 kcal/mol. This suggests thermal simulation of all three models would be comparable. Maeda and Ohno[11] independently examined the energy landscape of $(H_2O)_8$ and found 164 isomers. Here, we reoptimized their isomers using the B3LYP/6-31+G* method, and the resultant DOS is depicted in Figure 2d. The

OSS2	TTM2-F	B3LYP	OSS2	TTM2.1-F	B3LYP
-57.7660	-57.8342	-68.5069 (-60.9106)	-69.1409	-73.3289	-86.7434 (-77.3861)
-57.7409	-57.5960	-68.0896 (-60.4691)	-68.6715	-73.2965	-86.6625 (-77.3663)
-57.5125	-57.5447	-67.7464 (-60.9106)	-67.1448	-70.9212	-83.2212 (-74.3787)
-57.4999	-57.3696	-67.7345 (-60.2692)	-67.0946	-70.8902	-83.2055 (-74.3560)
-57.4541	-57.1882	-67.5632 (-59.7587)	-67.0613	-70.1057	-82.0980 (-73.4505)

Figure 4. Molecular structures of the five most stable isomers in OSS2, TTM2.1-F, and B3LYP/6-31+G*, at $n = 7$ and 8. The unit of the binding energy is reported in kcal/mol. Note that the binding energies of the B3LYP/6-31+G* isomers reoptimized using MP2/aug-cc-pvDZ are also reported and enclosed in parentheses.

TABLE 1: Numbers of Distinct Isomers of OSS2, TTM2.1-F, and B3LYP/6-31+G* for $(H_2O)_n$, $n = 4-8$[a]

	OSS2			TTM2.1-F			
size	I	II	III (%)	I	II	III (%)	B3LYP/6-31+G*
4	38	21	55	14	10	71	21
5	186	68	37	33	23	70	76
6	736	332	45	154	95	62	379
7	2700	1175	44	639	406	64	1443
8	5521	2455	44	2331	1429	61	3563

[a] The "success rate" (shown in column III) is the ratio between the number of successfully located isomers using B3LYP/6-31+G* (column II) and the initial number of isomers used in the empirical model (column I).

figure indicates that Maeda and Ohno have generated a fairly good coverage of the low-energy region. Nevertheless, our results highlight a significantly more complex energy landscape of $(H_2O)_8$ than was revealed earlier, since we discovered at least 2093 isomers.

The structures of the five most stable isomers for OSS2, TTM2.1-F, and B3LYP/6-31+G* are depicted in Figures 3 and 4. For $n = 4$ and 5, the most stable forms for all three models are ring structures. For $n = 6$, the global minima of OSS2 is also of ring form, which is similar to the TIP5P model; the TTM2.1F global minima has a cagelike form similar to the TIP4P model, whereas B3LYP/6-31+G* is of a two-ring-membered form. For $n = 7$, TTM2.1-F and DFT calculations predict the cubelike structure with a missing corner as the most stable state, whereas OSS2 retains the prediction of double ring

structure. Overall, one can conclude that OSS2 tends to favor more open structures rather than compact forms, whereas TTM2.1F displays an opposite trend. From the binding energies (by OSS2, TTM2.1F, B3LLYP/6-31+G*, and MP2/aug-cc-pvDZ methods) of these low-energy isomers, it is clear that the lower part of the PES of water clusters are characterized by many isoenergetic isomers, and it is beyond the scope of this work to make an extensive comparison of the empirical models. Nevertheless, the isomers we have archived will serve as a good starting effort to span greater investigation on using a multiscale approach to explore the PES of water clusters. Furthermore, we have also provided the five most stable structures in this paper with the aim of facilitating future studies in the field.

Conclusions

In this work, we demonstrated a multiscale approach for exploring the PES of water clusters at the quantum chemistry level in the range of $n = 4-8$. Two flexible models, TTM2.1F and OSS2, were used in the prescreening process for identifying the probable locations of the isomers in the PES of the B3LYP/6-31+G* calculation. The numbers of distinct isomers found using the present methodology are much larger than those reported in the literature highlighting the complexity of the PES of water clusters. In this study, the moderate 6-31+G* basis set has been chosen due to its high computational efficiency. Currently, we are extending the present work to a closely coupled multiscale optimization approach where the higher level

Potential Energy Surface of Water Clusters

J. Phys. Chem. A, Vol. 112, No. 28, 2008 **6261**

calculations are directly coupled to mitigate the basis set superposition error which results from the limited basis set method.

References and Notes

(1) Tsai, C. J.; Jordan, K. D. *J. Chem. Phys.* **1991**, *95*, 3850.

(2) Tsai, C. J.; Jordan, K. D. *J. Chem. Phys.* **1993**, *99*, 6957.

(3) Tsai, C. J.; Jordan, K. D. *J. Phys. Chem.* **1993**, *97*, 11227.

(4) Lee, C.; Chen, H.; Fitzgerald, G. *J. Chem. Phys.* **1995**, *102*, 1266.

(5) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 1500.

(6) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 1479.

(7) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 5115.

(8) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. *J. Chem. Phys.* **2002**, *116*, 1493.

(9) Lagutschenkov, A.; Fanourgakis, G. S.; Niedner-Schateburg, G.; Xantheas, S. S. *J. Chem. Phys.* **2005**, *122*, 194310.

(10) Bandow, B.; Hartke, B. *J. Phys. Chem. A* **2006**, *110*, 5809.

(11) Maeda, S.; Ohno, K. *J. Phys. Chem. A* **2007**, *111*, 4527.

(12) Tsai, C. J.; Jordan, K. D. *J. Phys. Chem.* **1993**, *97*, 5208.

(13) James, T.; Wales, D. J.; Rojas, J. H. *J. Chem. Phys.* **2007**, *126*, 054506.

(14) Wales, D. J.; Ohmine, I. *J. Chem. Phys.* **1993**, *98*, 7245.

(15) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758.

(16) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891.

(17) Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett.* **1998**, *286*, 65.

(18) James, T.; Wales, D. J.; Hernandez-Rojas, J. *Chem. Phys. Lett.* **2005**, *415*, 302.

(19) Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. *J. Chem. Phys.* **1999**, *110*, 4566.

(20) Fanourgakis, G. S.; Xantheas, S. S. *J. Phys. Chem. A* **2006**, *110*, 4100.

(21) Ojamae, L.; Shavitt, I.; Singer, S. J. *J. Chem. Phys.* **1998**, *109*, 5547.

(22) Mella, M.; Kuo, J.-L.; Clary, D. C.; Klein, M. L. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2324.

(23) Daven, D. M.; Tit, N.; Morris, J. R.; Ho, K. M. *Chem. Phys. Lett.* **1996**, *256*, 195.

(24) Ong, Y. S.; Keane, A. J. *IEEE Trans. Evol. Comput.* **2004**, *8*, 99.

(25) Turner, G. W.; Tedesco, E.; Harris, K. D. M.; Johnston, R. L.; Kariuki, B. M. *Chem. Phys. Lett.* **2000**, *321*, 183.

(26) Cox, G. A.; Mortimer-Jones, T. V.; Taylor, R. P.; Johnston, R. L. *Theor. Chem. Acc.* **2004**, *112*, 163.

(27) Ballester, P. J.; Richards, W. G. *J. Comput. Chem.* **2007**, *28*, 1711.

(28) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(29) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004).

(31) Stillinger, F. H. *Phys. Rev. E* **1999**, *59*, 48.

(32) Wales, D. J. *Energy Landscape*; Cambridge University Press, Cambridge, UK; New York:2003.

JP802118J