Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemical Physics Letters 389 (2004) 309–314
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

A connectivity table for cluster similarity checking in the evolutionary optimization method

Longjiu Cheng, Wensheng Cai, Xueguang Shao *

Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, 230026, PR China

Received 2 March 2004; in final form 25 March 2004
Available online 15 April 2004

Abstract

A novel and effective cluster similarity checking method using the connectivity table (CT) is proposed in this study. Because CT contains the topological information of a cluster, configurations at different funnels on the potential energy surface (PES) show great difference in their CTs. An adaptive immune optimization algorithm (AIOA) is utilized for optimization of Lennard–Jones (LJ) clusters up to LJ_{110} using the CT for similarity checking. It is proved that the method is very efficient, and the method is also capable for optimizations of larger clusters, e.g., LJ_{200}.

1. Introduction

The global optimization problem [1] is an active field with rapid growth in structural optimization, chemical engineering design and molecular biology. In chemical fields, the applications of global optimization include finding the lowest energy configuration of a molecular system or finding the lowest energy conformation of molecular or atomic clusters [2–4]. These problems are notoriously difficult because there are many local minima in the energy landscape, and also a large number of parameters in the force field need to be optimized. As a result, a global optimization algorithm for this kind of problem should have good performance in solving the high-dimensional continuous problems. Different optimization methods have been applied to solve the problem, such as genetic algorithms (GAs) [5–7], simulated annealing algorithm (SA) [8–10], basin-hopping and its variants [1,11–13], fast annealing evolutionary algorithm (FAEA) [14,15], random tunneling algorithm (RTA) [16], potential deformation [17], simple linkage [18], and modeling methods [19,20], etc.

The determination of the global minima of Lennard–Jones (LJ) clusters by numerical global optimization methods is one of these problems, which has been intensely studied. LJ clusters consist of identical atoms interacting by the LJ potential:

$$V(r) = 4\varepsilon \sum_{i<j} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], \quad (1)$$

where r_{ij} represents the distance between atoms i and j, and the reduced units are generally used, i.e., $\varepsilon = \sigma = 1$. The LJ potential is not only interesting as a model for heavy inert gases but also serves as a popular benchmark for evaluating optimization algorithms. Most of the known global minima for LJ clusters [21] containing fewer than 110 atoms are based on icosahedral packing. The exceptions, global minima of LJ_{38} (truncated octahedron), $LJ_{75–77}$, $LJ_{102–104}$ (Marks decahedron) and LJ_{98} (Leary tetrahedron [12]) serve as particularly interesting test cases. At these magic numbers, the global minima lie in a very deep funnel on the LJ potential energy surface (PES), while the lowest-energy icosahedral ones act as a trap in a much wider funnel.

For an evolutionary method, to guarantee efficiency, the key technique is to keep diversity of the population for speeding up its convergence. The number of local minima on the LJ PES is very huge, and it increases

* Corresponding author. Fax: +86-551-3601592.
E-mail address: xshao@ustc.edu.cn (X. Shao).

0009-2614/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2004.03.125

exponentially with the number of atoms. Furthermore, the number of parameters (coordinates) of the LJ cluster is very large, and the same structure of a cluster may correspond to different coordinates after rotation or moving operations. Therefore, determination of the distance of two local minimum structures directly by their coordinates is with great difficulty. In recent years, to guarantee the diversity in evolutionary methods for cluster optimization, many attempts have been carried out to measure the distances between local minimum structures. In [7], a GA method is applied to the LJ clusters, and the distance of two local minimum structures is measured directly by the gap of their potential energy. The global minimum of LJ_{38} is reproduced by this method, and the global minimum of LJ_{88} is found for the first time, but it failed for the other difficult cases ($LJ_{75–77}$ and LJ_{98}). In [5], a novel similarity checking method for the non-icosahedral cluster using the concept of niches is presented, and most of the known global minima up to LJ_{150} are located, except for LJ_{98}. There also have been some cluster similarity checking methods by directly comparing the difference of their coordinates [16,17]. More recently, a powerful evolutionary global optimization method, named as conformational space annealing (CSA) [22], was applied for LJ clusters, where the concept of the histogram is presented and used for cluster similarity checking. All known global minima up to LJ_{201} are located by CSA with a high success rate.

According to the methods in the literature and our experiences, cluster similarity checking method should only be based on the topological information of structures instead of the coordinates. Also, to measure the distance of two local minimum structures easily, the similarity should be expressed by a numerical value.

In this study, a novel and efficient cluster similarity checking method is proposed, which uses the connectivity information of the atoms in a cluster, and the topological structure of a cluster can be characterized by a connectivity table (CT). To test the performance of the method, it is adopted in an evolutionary algorithm, named as adaptive immune optimization algorithm (AIOA), and applied to the optimization of LJ clusters. A high performance of the method is proved.

2. Connectivity table for cluster similarity checking

The number of the nearest neighbor contacts [23] is an important property of clusters, and generally, the number for the clusters with decahedral packing and close packing is smaller than that for the icosahedral packing. The number of nearest neighbor contacts, n_{nn}, is given by

$$n_{\mathrm{nn}} = \sum_{i<j} \delta_{ij}, \tag{2}$$

where

$$\delta_{ij} = \begin{cases} 1 & r_{ij} < r_0, \\ 0 & r_{ij} \geqslant r_0, \end{cases}$$

and r_0 is a nearest neighbor criterion.

To study the topological information, we defined that an atom is connected with its nearest neighbors, and the connectivity number of each atom is the number of its nearest neighbors. In our experiences, the connectivity number of an atom in a minimized LJ cluster is no more than 12 (the number could be larger than 12 for some other clusters, such as Ni–Al alloy clusters [24]), and the atom with full connectivity number always lies in the inside of a cluster. The atoms on the surface of a cluster will have connectivity number less than 12, and the connectivity information will be much different for different packing. Therefore, a CT is defined to characterize the connectivity information of a cluster by using a vector $\mathrm{CT}(i)$, where $\mathrm{CT}(i)$ ($i = 1$–12) means the number of the atoms having i nearest neighbors in a cluster.

To verify whether the CTs can measure the difference of different configurations, we studied the CTs of LJ_{98} with various packing: Leary tetrahedron (TE), Mackay icosahedron (IC), anti-Mackay icosahedron (FC) [25], 101-atom Marks decahedron (ID), 75-atom anti-Marks decahedron (FD) [20], and close packing (CP) [23]. Fig. 1 shows the two-dimensional projections of these local minimum structures on the funnel bottoms. It can be found that, there are great differences in topological structure among them. Table 1 shows the number of nearest neighbor contacts, potential energy and CTs of these local minimum structures. To test the effect of CTs for similar configurations, two similar TEs and two similar ICs are also listed in Table 1. It can be seen that the CTs of the two TEs are very similar, and also for the two ICs. In fact, the two TEs and the two ICs are both differed by one atom location. This indicates that, similar structures have similar CTs. From Table 1, it also can be seen that the structures with different packing have great difference with their CTs, while the potential energy and the nearest neighbor contacts may be similar. The comparison indicates that CT can be utilized as a good tool for cluster similarity checking.

Generally, the nearest neighbor contacts of FC are larger than that of IC, and much larger than that of other packing. Therefore, we define the distance measure of two local minimum structures v and w as follows:

$$D(v,w) = (1 + |n_{\mathrm{nn}}^{v} - n_{\mathrm{nn}}^{w}|)^{1/2} \sum_{i=1}^{12} [\mathrm{CT}^{v}(i) - \mathrm{CT}^{w}(i)]^2, \tag{3}$$

where n_{nn} is the number of nearest neighbor contacts obtained by Eq. (2). The equation is a good distance measure for similar and dissimilar clusters. For example, according Eq. (3), the distance between the two similar

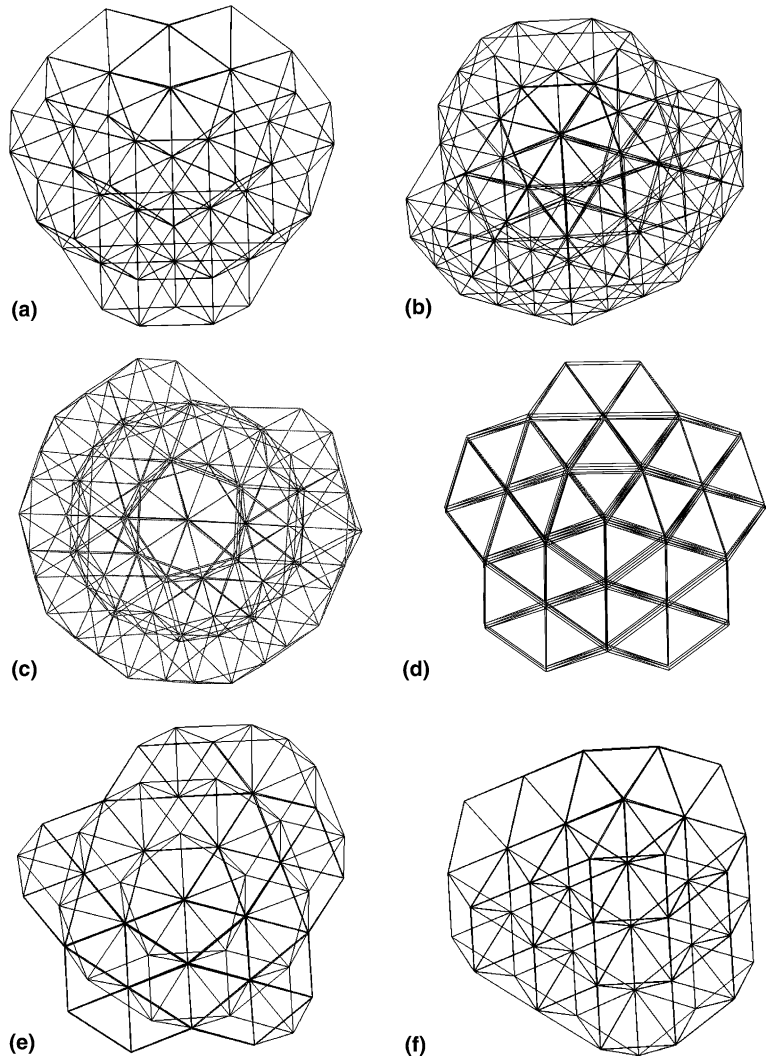

Fig. 1. Two-dimensional projections of the configurations at the funnel bottoms of LJ_{98} with various packing: (a) Leary tetrahedron (TE); (b) Mackay icosahedron (IC); (c) anti-Mackay icosahedron (FC); (d) 101-atom Marks decahedron (ID); (e) 75-atom anti-Marks decahedron (FD); (f) close packing (CP).

Table 1
Connectivity table for various local minimum structures of LJ_{98}

Packing	n_{nn}	V	Connectivity table CT(i)											
			1	2	3	4	5	6	7	8	9	10	11	12
TE	432	−543.665341	0	0	0	0	0	24	24	0	4	18	0	28
	430	−541.405607	0	0	0	1	1	24	20	2	5	16	2	27
IC	437	−543.642957	0	0	0	0	0	19	9	26	12	0	3	29
	437	−543.546726	0	0	0	0	0	20	7	27	12	0	3	29
FC	441	−539.720452	0	0	0	0	0	29	19	13	3	7	4	32
ID	428	−541.894959	0	0	0	0	0	23	22	5	11	9	1	27
FD	430	−541.436260	0	0	0	0	2	21	17	12	11	4	2	29
CP	429	−541.869748	0	0	0	0	0	24	15	18	3	9	0	29

TEs listed in Table 1 is 55.4 only, while the distance between the TE and the IC is 3243.

To investigate the performance of the cluster similarity checking method, it was applied in an evolutionary optimization method of the AIOA.

3. Adaptive immune optimization algorithm

In recent years, the study of novel algorithms based on biological immune mechanisms has become an active research field [26,27]. The system is an efficient natural

protection system that can generate multiple antibodies from antibody gene libraries. The primary immune theory model is the immune density regulation mechanism. The theory shows that the biological immune system can regulate the generation of antibodies and balance the quantity of the multiple kinds of antibodies. When antigens invade, the antibodies that match these antigens are activated and generate more antibodies to restrain the antigens. Then the immune system reaches a new balanceable state.

An adaptive immune optimization algorithm is established and the CT is adopted in the algorithm for cluster similarity checking. Furthermore, the basic idea of simulated annealing is adopted in this study. The algorithm has a higher mutation rate at a higher temperature, and the fitness of individuals is also related to the temperature. The annealing schedule of temperature $T(k)$ is calculated by

$$T = T_0 \exp[-2.0(k/K)^{0.25}], \tag{4}$$

where T_0 is the starting temperature, k is the iteration number, and K is a constant. A larger K means the temperature goes down more slowly. In this study, $T_0 = 1.5$–3.0 and $K = 300$–1000.

AIOA adopts similar frame of GAs, and regards the evolving individuals as antibodies. But it is different from GAs, there is no crossover operation in AIOA, and antibodies are selected from a gene library by an adaptively generated selection rate at each generation. Each member of the population is minimized after every step, using the limited memory quasi-Newton algorithm (L-BFGS) [28]. The detail steps of the AIOA are as following:

(1) Antigen invades, which means a problem to be solved. Initialize an empty *dead library*. The function of the dead library is to record failures. The best individual of the failed convergences will be put into the dead library and used to control the current antibodies to be dissimilar with them by similarity checking.

(2) Initialize a certain number (n_{lib}) of initial configurations whose energy is subsequently minimized. We call the set of these configurations the *gene library*. The configurations in the library are updated in later stages, and the number of configurations in the library is kept unchanged when it is updated. The average distance (D_{av}) between the configurations in the gene library is evaluated by Eq. (3).

(3) For the kth generation, the fitness of the individual in the gene library is calculated by

$$\text{fit}(v) = \exp[-V/(T(k)A)], \tag{5}$$

where V is the potential energy, $T(k)$ is the temperature obtained by Eq. (4), and A is a constant ($\sqrt[3]{N}$ is used in this study and N is the number of atoms). From the equation, it can be seen that, individuals with lower potential energy will have higher fitness, and the fitness difference between individuals will be small at a high temperature. This indicates that the algorithm guarantees more in diversity with a small k, and guarantees more in convergence speed with a large k.

(4) Calculate the density (C_i) of each individual in the gene library. The value of C_i is the number of configurations similar to individual i. In this study, if the distance between v and w, $D(v, w)$, is smaller than $D_{\text{av}}/3$, the two configurations are thought to be similar. The density is no less than 1, for a configuration is similar to itself.

(5) Select n_{pop} individuals from the gene library to compose kth child antibodies by the immune selection procedure. After calculating the density of each individual in the gene library, the regulation of activating and suppressing of different genes can be achieved by immune selection mechanism. On the basis of the traditional selection mechanism of the fitness proportion, by increasing the regulation probability factor based on density, the selection probability of individual v, e_v, is determined by two sections, the fitness and the density:

$$e_v = \frac{\text{fit}(v)}{C_v}. \tag{6}$$

This equation indicates that the greater the individual fitness, the higher the selection probability it possesses; the greater the density of an individual, the lower the selection probability it possesses. Thus AIOA can not only maintain the individuals of high affinity, but also guarantee the diversity.

(6) Perform energy-based mutation on the antibodies. Based on our knowledge, in a real atomic cluster system, the atoms with higher energy will have higher activity. Also, atoms have higher activity at a higher temperature. Generally, atoms with lower connectivity number will have higher potential energy. Therefore, the mutation rate of an atom in a cluster is defined as

$$R_{\text{mut}} = \exp[(4.5 - nn)/T(k)], \tag{7}$$

where nn is the connectivity number of the atom. If an atom is selected, it will be moved to a random location on the surface of the cluster. The mutation number of an antibody is set to be no less than 2. From the equation, it can be seen that the bad atoms with connectivity number no more than four are always mutated. The energy-based mutation can greatly improve the convergence speed of the algorithm. Also, the higher mutation rate in smaller k guarantees more in diversity, and the lower mutation rate in larger k guarantees more in convergence speed. Each antibody is minimized with the local minimization procedure after the mutation operation.

(7) If the dead library is not empty, each antibody is compared with the individuals in the dead library. Let w be the individual in the dead library. For antibody v, if $D(v, w) < D_{\text{dead}}$ ($D_{\text{dead}} = D_{\text{av}} \sim 2D_{\text{av}}$ in this study), the

antibody is thought to be in the same funnel on the PES with the individual in the dead library, and discard the antibody directly. For general cases, the dead library is not necessary. But for the non-icosahedral magic numbers, the dead library can make the problem easier after a failure.

(8) Each remaining antibody is compared with individuals in the gene library to decide how the library should be updated. For antibody v, let w be the most similar one to v in the gene library according to Eq. (3). If $D(v,w) < D_{\text{sim}}$, where D_{sim} is a threshold, v is considered as similar to w. In this case, the one with lower energy among v and w is kept in the library. However, if $D(v,w) > D_{\text{sim}}$, v is regarded as distinct from all individuals in the library. In this case, the individual with the highest energy among the library plus v is discarded, and the rest are kept in the library. D_{sim} is an important parameter. A larger D_{sim} corresponds to a more severe criterion for similarity checking, and the diversity is emphasized more than the convergence speed. In this study, D_{sim} is related to the temperature: $D_{\text{sim}} = D_{\text{av}}T(k)/2$, and D_{sim} is set to be no smaller than $D_{\text{av}}/4$.

(9) If the iteration reaches the preset number *LOOP*, the algorithm is thought to converge at a local funnel bottom on the PES, and the energy lowest individual in the gene library is put into the dead library. The iteration number is reset to zero, and go to step (2). The value of *LOOP* is set by experience. Generally, a more severe criterion of similarity checking needs a larger *LOOP*.

(10) If the global minimum is obtained, stop the calculation. Otherwise, go to step (3). Because the global minima of the LJ clusters in this study are already known, to verify the performance of the proposed method, the algorithm is controlled to terminate when it reaches the known global minima.

4. Results and discussion

All the LJ clusters up to LJ_{110} were investigated by the proposed method. It was found that the method successfully located all the known global minima. To measure the overall efficiency of our method, ten independent runs were performed for each cluster size. It was found that all known global minima were obtained for all ten runs, without an exception. For a general case, the size of the gene library (n_{lib}) is 30–100, and the population size $n_{\text{pop}} = 20$–60. But for the magic numbers with non-icosahedral global minima (LJ_{75-77}, LJ_{98} and $LJ_{102-104}$), $n_{\text{lib}} = 200$ and $n_{\text{pop}} = 100$ were used.

Fig. 2 shows the average number of local minimizations per hit of global minima for the clusters with $20 \leqslant N \leqslant 110$ of ten independent runs, the average CPU time is also shown in the figure. The computations were carried out on an HP cluster with Intel Itanium2 Mad-

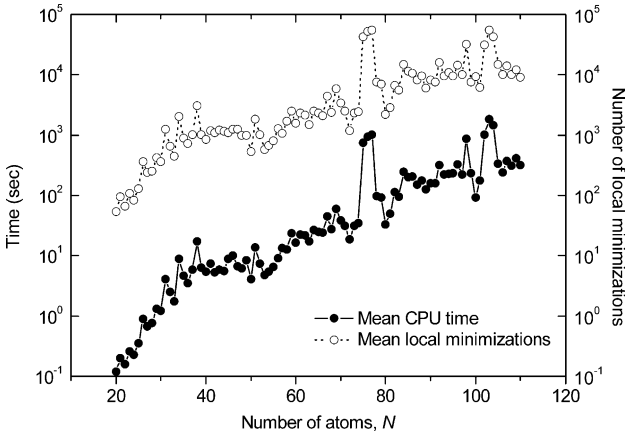

Fig. 2. Average CPU time (s) and average number of local minimizations per hit of the global minima for clusters with $20 \leqslant N \leqslant 110$.

sion processors (1.5 GHz). It can be seen that, due to the effective similarity checking method, the magic numbers do not show too much difficulty, as compared to their neighbors. For example, the average number of local minimizations consumed by LJ_{75-77} is no more than ten times of their neighbor's, but in [29] the value is about 100. The same local minimization method was used as in CSA [22]. By comparison of the average CPU time in Fig. 2 with that in Fig. 2 of [22] (obtained with a 1.667 GHz Athlon processor), it can be found that our method is faster than CSA for the magic numbers. For example, the average CPU time for one hit of the global minima of LJ_{75} and LJ_{98} in [22] is nearly 10^4 s. However, in this work, the corresponding time is only 755 and 862 s, respectively.

The average number of local minimizations is usually taken as a criterion to measure a global optimization algorithm. Furthermore, the success rate is another important criterion to evaluate an evolutionary algorithm. The average number of local minimizations needed by one hit of global minima and the success rate in the optimization of some selected clusters is compared with the monotonic sequence basin-hopping (MSBH) [29] and RTA [16] in Table 2. It can be seen that, average number of local minimizations of our method is larger than that of MSBH for the general cases with small size, but for the magic numbers (especially for LJ_{75} and LJ_{98}) and larger cases (LJ_{110}), our method is much better than MSBH. It also can be seen that, the success rate of our method is much higher than that of MSBH and RTA. The results indicate that connectivity table is a good tool for cluster similarity checking and AIOA is a good global optimization method which can guarantee the individual diversity and convergence speed adaptively.

To further test the performance of the proposed method, it is applied to a cluster with larger size, LJ_{200}, with $n_{\text{lib}} = 300$ and $n_{\text{pop}} = 150$. The known global minimum was also reproduced in all ten independent

Table 2
Mean number of local minimizations per hit of global minima and success rate for selected clusters of CT-AIOA and some other unbiased global optimization methods

N	MSBH		RTA	CT-AIOA	
	N_{LM}[a]	R_{hit}[b]	R_{hit}[b]	N_{LM}[a]	R_{hit}[b]
30	739	0.387	4/10	360	10/10
38	2875	0.124	9/10	3046	10/10
40	279	0.849	10/10	844	10/10
50	460	0.868	5/10	537	10/10
60	388	0.948	1/10	1575	10/10
70	1526	0.630	3/10	3420	10/10
75	152000	0.004	1/50	43000	10/10
80	2009	0.420	4/10	2231	10/10
90	4699	0.206	1/10	8185	10/10
98	180000	0.006	2/20	32000	10/10
100	9128	0.122	2/10	9304	10/10
102	36028	0.031		31000	10/10
110	40420	0.031		9200	10/10

[a] N_{LM}: average number of local minimizations per hit of the global minima.
[b] R_{hit}: success rate of hitting the global minima.

runs, and the average number of local minimizations is 1 40 000 with about 10^4 s. This indicates the proposed method is also capable for clusters with larger size.

Acknowledgements

This study is supported by the outstanding youth fund (No. 20325517) from National Natural Scientific Foundation of China (NNSFC), and the Teaching and Research Award Program for Outstanding Young Teachers (TRAPOYT) in higher education institutions of the Ministry of Education (MOE), PR China. Thanks to USTC-HP High Performance Computing Joint Lab (HPCJL) for affording HP-cluster.

References

[1] D.J. Wales, H.A. Scheraga, Science 285 (1999) 1368.
[2] F. Baletto, C. Mottet, R. Ferrando, Phys. Rev. Lett. 84 (2000) 5544.
[3] C.L. Cleveland, U. Landman, J. Chem. Phys. 94 (1991) 7376.
[4] J.E. Hearn, R.L. Johnston, J. Chem. Phys. 107 (1997) 4674.
[5] B. Hartke, J. Comput. Chem. 20 (1999) 1752.
[6] M.D. Wolf, U. Landman, J. Phys. Chem. A 102 (1998) 6129.
[7] D.M. Deaven, N. Tit, J.R. Morris, K.M. Ho, Chem. Phys. Lett. 256 (1996) 195.
[8] G.L. Xue, J. Global Optim. 4 (1994) 425.
[9] S. Schelstraete, H. Verschelde, J. Phys. Chem. A 101 (1997) 310.
[10] R.F. Gutterres, M.A. de Menezes, C.E. Fellows, O. Dulieu, Chem. Phys. Lett. 300 (1999) 131.
[11] D.J. Wales, J.P.K. Doye, J. Phys. Chem. A 101 (1997) 5111.
[12] R.H. Leary, J.P.K. Doye, Phys. Rev. E 60 (1999) R6320.
[13] J. Pillardy, A. Liwo, H.A. Scheraga, J. Phys. Chem. A 103 (1999) 9370.
[14] W.S. Cai, X.G. Shao, J. Comput. Chem. 23 (2002) 427.
[15] W.S. Cai, Y. Feng, X.G. Shao, Z.X. Pan, J. Mol. Struct. (Theochem) 579 (2002) 229.
[16] H.Y. Jiang, W.S. Cai, X.G. Shao, Phys. Chem. Chem. Phys. 4 (2002) 4782.
[17] L. Piela, J. Kostrowicki, H.A. Scheraga, J. Phys. Chem. 93 (1989) 3339.
[18] F. Schoen, Eur. J. Oper. Res. 119 (1999) 345.
[19] J.A. Northby, J. Chem. Phys. 87 (1987) 6166.
[20] D. Romero, C. Barrün, S. Gümez, Comput. Phys. Commun. 123 (1999) 87.
[21] The complete up-to-date list of the global minima of LJ cluster can be found at http://brian.ch.cam.ac.uk/ for $N \leqslant 150$ and at http://www.vcl.uh.edu/~cbarron/LJ_cluster/researchpot.html for $147 \leq N \leq 309$.
[22] J. Lee, I.H. Lee, J. Lee, Phys. Rev. Lett. 91 (2003) 080201.
[23] J.P.K. Doye, D.J. Wales, Chem. Phys. Lett. 247 (1995) 339.
[24] M.S. Bailey, N.T. Wilson, C. Roberts, R.L. Johnston, Eur. Phys. J. D 25 (2003) 41.
[25] S.C. Hendy, J.P.K. Doye, Phys. Rev. B 66 (2002) 235402.
[26] J.S. Chun, M.K. Kim, H.K. Jung, IEEE T. Magn. 33 (1997) 1876.
[27] S.J. Huang, Int. J. Electron. Power 21 (1999) 245.
[28] D.C. Liu, J. Nocedal, Math. Program. B 45 (1989) 503.
[29] R.H. Leary, J. Global Optim. 18 (2000) 367.