28 August 1998

ELSEVIER

Chemical Physics Letters 293 (1998) 202–208

CHEMICAL PHYSICS LETTERS

A symbiotic algorithm for finding the lowest energy isomers of large clusters and molecules

K. Michaelian

Instituto de Física, Universidad Nacional Autónoma de México, A.P. 20-364, 01000 México D.F., Mexico

Received 8 April 1998; in final form 18 June 1998

Abstract

A novel 'symbiotic' algorithm, based on the genetic algorithm, is presented for finding the structure and energy distribution of the lowest energy isomers of large clusters and molecules. This approach takes advantage of the strong coupling of nearest neighbor atoms through the fitness function (the binding energy) directing the selection process, due to the short range of the interatomic potential in comparison to the cluster size. Evolving locally in cells and then forming and evolving a symbiosis of the cells is substantially more efficient than employing the genetic algorithm on the full cluster. Application is made to Lennard-Jones clusters of 6, 18, 23, 38 and 55 atoms.

1. Introduction

The determination of the structure of the ground state and low energy isomers of clusters of atoms, or molecules, is important in many practical applications from nano-electronics [1] to the design of antibiotics [2]. The inherent difficulty in experimentally isolating and probing such fragilely bound systems has put the onus on theoretical calculations to predict their structural, electronic, chemical and thermodynamic properties. There are now various approaches to this end based on the optimization, leading to low energy structures, of, either a basis set of wavefunctions or densities (density functional theory) in quantum calculations, or atom coordinates in semiempirical approaches. Ab initio and density functional studies are generally considered to be the most informative but are computationally too expensive to implement in the global optimization of large clusters. Semi-empirical approaches employing potentials with parameters adjusted to either experimental data or selected ab initio results [3] are an effective alternative. Such potentials can often include effects beyond the scope of the formalism [4] and, in any case, results obtained through the semiempirical optimization can be used as configurational input to a full ab initio or a density functional calculation.

An effective, and therefore popular, optimization approach is simulated annealing invoking either random displacements (Monte Carlo) [5] or molecular dynamics [6,7]. However, recent investigations of large clusters with complex potential energy surfaces

0009-2614/98/$ - see front matter
PII: S0009-2614(98)00733-7

(e.g. the number of local, energy minima of a Lennard-Jones (LJ) cluster increases exponentially with the number of atoms [8]) have demonstrated that the genetic algorithm can significantly outperform simulated annealing and most other currently available global optimization techniques [1] in finding the ground state structure and that of the low energy isomers [10–13].

The genetic algorithm, proposed by Holland [14], is based on nature's efficient problem solving method of evolving a microscopic genetic code through mutation and crossover, and selection based on the fitness of the corresponding individuals. It is an intelligent and information efficient approach [11] to multi-variable, global optimization and has been successfully applied to a large number of complex problems from the physical sciences [15]. Judson et al. [10] first applied the genetic algorithm to the configuration of a cluster of atoms confined to two dimensions and bound through a Lennard-Jones (LJ) potential. Recent application to the well studied, three dimensional LJ [16] and ionic [17] clusters has yielded new ground states and low energy isomers for even relatively small systems. A general introduction to genetic algorithms can be found in Sutton and Boyden [18]. For an introduction to the genetic algorithm approach applied to cluster physics see Michaelian [17] and, to molecular conformation see Judson et al. [19].

The intention of this Letter is to demonstrate that the efficiency of this evolutive approach to cluster or molecular configuration can be substantially improved by taking advantage of a peculiarity of the fitness function directing the selection process. For optimization at zero temperature, the fitness function is naturally chosen as being directly related to the binding energy of the configuration as a whole. However, for large clusters or molecules the interatomic potential is often of short range in comparison to the size of the system and the contribution to the binding energy due to interactions between nearest neighbors is of significantly larger magnitude than that for interactions between atoms separated by larger distances. This is particularly true for the LJ potential used in modeling rare gas clusters. Genetic fitness is thus largely determined by the nearest neighbor structure. This means that, to a first approximation, the multi-variable, global problem can be reduced to a linear combination of various local problems of fewer variables, resulting in a considerable reduction in the complexity and in the number of local minima in the corresponding potential energy surfaces.

The local sensitivity in the fitness function also adversely affects the efficiency of the genetic operator of crossover, regardless of the ordering of the atomic coordinates in the genetic string. If neighboring atom coordinates are located close together on the string, crossover is not very effective locally where it is most essential for improving the fitness. If they are located far apart, or at random positions on the string, unrestrained crossover has the effect of disrupting the acquired beneficial characteristics of the population [20]. Restraining crossover to specific sections of the cluster, for example to cluster halfs [21], is a compromise but far from optimal and is likely to favor evolution towards corresponding symmetries in the final configuration.

Here we consider evolving localized solutions in overlapping cells of the cluster separately. The atom positions within the cell are mutually optimized, the rest of the atoms are fixed and provide a constant energy contribution. Optimizing sequentially from cell to cell, a symbiosis is forced between the evolved cell and the rest of the cluster. Aside from reducing the complexity of the potential energy surface, this scheme makes optimal use of the crossover operation, constraining it to act locally, only within each cell.

In the following sections the symbiotic algorithm is described in detail and an evaluation of the relative efficiency (symbiotic/genetic) is made for optimizing the structures of LJ clusters of 6, 18, 23, 38 and 55 atoms. These sizes were chosen because they present elusive ground states [22,23,16,24].

[1] In an interesting alternative, Wales and Doye [9] have applied the Monte Carlo search to the LJ cluster after transforming the potential energy surface into a collection of interpenetrating staircases, the steps corresponding to basins of attraction of the local minima. This transformation removes the transition state regions while leaving the global minimum unchanged. They report new global minima not found in previous searches (including those employing the genetic algorithm).

2. Symbiotic optimization of the Lennard-Jones cluster

The atoms in the LJ cluster are bound through a potential of the form

$$V_{ij} = \frac{1}{r_{ij}^{12}} - \frac{2}{r_{ij}^{6}} \tag{1}$$

giving an energy of -1 at an equilibrium distance of 1, for the dimer. All energy and distance values reported in this work are in these units and the energy values are per atom. This system has been extensively studied through various traditional seed and growth techniques [8,25], with Monte Carlo and molecular dynamics [5,7,9] and, with the genetic algorithm [10,22,26,16] (see Ref. [9] for a complete list of first reported global minima and references therein for the details of the methods used).

The n atoms of the cluster are first generated at random positions within a sphere of radius

$$r = 1.2 \cdot \left(\frac{n}{13}\right)^{1/3} \tag{2}$$

centered at the origin of a three dimensional coordinate system. This relation gives a radius somewhat larger than that of a sphere enclosing the ground state icosahedral structure for the 13 atom LJ cluster. A ''template'' genetic string is composed for the configuration [17] by locating the three spatial coordinates of each atom together, and the sets of coordinates of each atom are ordered along the string in the order in which they are generated. An eight bit binary Gray coding [27] of the coordinate variables was chosen.

A first spherical cell is defined of radius 1.4, containing on average, $m = 8$ atoms, centered on the location of the first atom appearing in the string. The variables corresponding to atom coordinates within this cell are evolved with the standard genetic algorithm technique [17] employing mutation and crossover, with selection based on a reduced fitness function for the cell

$$E_c = \sum_{l<k}^{m} V_{kl}. \tag{3}$$

The sum on k and l is only over the atoms in cell one, m being the number these. Each of the m atoms is thus simultaneously optimized with respect to the other $m-1$ atoms of the cell. The rest of the cluster ($n-m$ atoms) presents, in this first approximation, a fixed contribution to the total energy.

Evolution within the cell proceeds until the best energy for the cell has not changed during 7 consecutive generations. This stopping criterion is simple to implement and has been shown to work significantly better than the usual one based on bit convergence of the genetic strings [19]. The best fit solution then forms a symbiosis with the rest of the cluster by updating the template if this action gives a lower total energy

$$E = \sum_{j<i}^{n} V_{ij} \tag{4}$$

for the entire cluster.

Next, a second cell is defined of equal radius centered on atom 2 of the genetic string and the evolutive and symbiotic process repeated. After evolving and completing the symbiosis for n cells the algorithm is repeated beginning with cell one. The symbiotic algorithm is continued in this manner until there is little improvement in the binding of the entire cluster. At this point a local conjugate gradient optimization is performed to refine the structural details most dependent on the long range part of the potential.

3. Results

The proposed symbiotic technique was compared with a standard genetic algorithm. The genetic algorithm was obtained from the symbiotic algorithm simply by fixing the radius of the cell to a value sufficiently larger than the size of the whole cluster so that complete global optimization was performed within one cell only. This procedure ensures a fair evaluation of the two different approaches by avoiding significant algorithm coding difference effects.

Fig. 1 shows, for three different sized LJ clusters, the lowest energy obtained for the cluster as a function of the number of evaluations of the potential (Eq. (1)). The curves are averages of 20 runs with distinct initial configurations. Both evolutive algorithms performed best at a small population size of 5 genetic strings (see Fig. 1). For both algorithms, a new generation was produced by creating four new solutions through crossover reproduction between pairs selected from the three best solutions of the

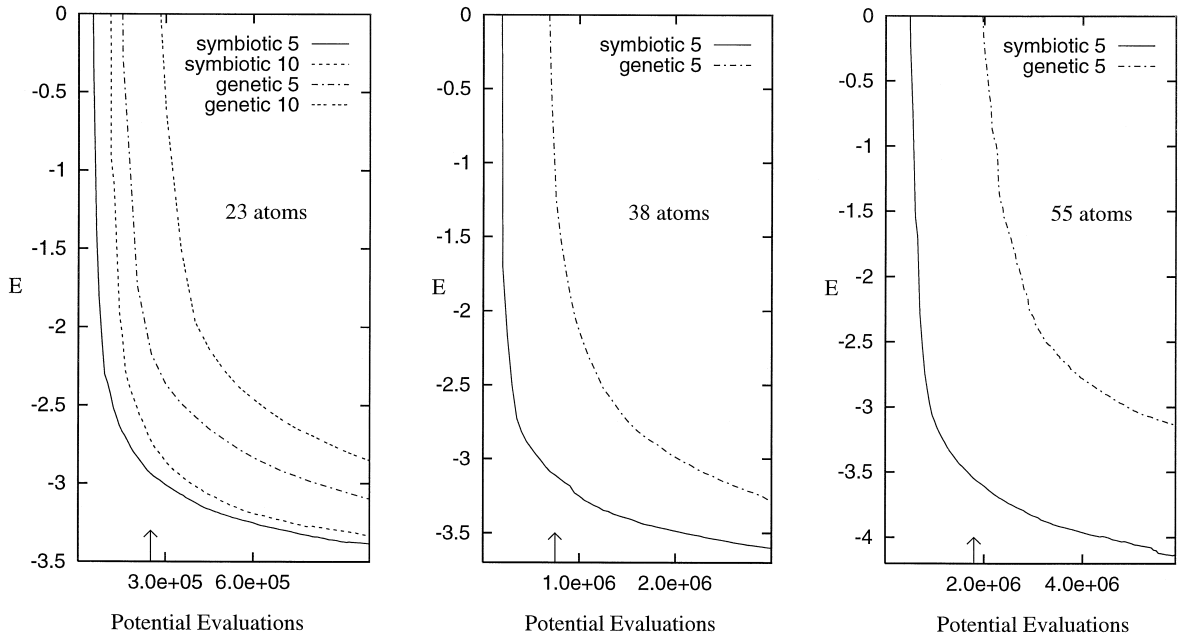

Fig. 1. Comparison of the best energy found with the symbiotic and genetic algorithms as a function of the number of potential evaluations (Eq. (1)) for three different sized LJ clusters. The curves are averages of 20 separate runs for each algorithm. For the cluster size of 23 atoms results for population sizes of 5 and 10 individuals are given.

previous generation. One of these four suffered a mutation of one bit. The best solution was always passed to the subsequent generation without change [17]. Significant variations of this reproduction scheme produced only very small changes in the relative efficiency of the two algorithms.

It was determined to be most efficient to stop the symbiotic algorithm and begin local minimization at $4.0\times10^3, 1.5\times10^5, 2.5\times10^5$, 7.5×10^5, and 1.8×10^6 potential evaluations (arrows Fig. 1) for the cluster sizes of 6, 18, 23, 38 and 55 atoms respectively. These numbers are the result of an evaluation of the average CPU time required by the symbiotic algorithm to arrive at a particular energy, the time required to minimize locally (to $\Delta E/E = 10^{-8}$) from this energy with the conjugate gradient technique, and the probability of entrapment in high energy local minima. [2] Requiring the genetic algorithm to reach the same best energy values, Fig. 1 then indicates that the symbiotic approach is approximately 2.8, 3.4 and 5.5 times more efficient in terms of potential evaluations than the standard genetic approach at finding low energy candidates suitable

Table 1

Size	Symbiotic		Hybrid Genetic
	each run [a]	global minimum [b]	global minimum [c]
6	0.03	0.36	0.90
18	0.66	42	240
23	1.16	122	155 [d]
38	4.73	189200	
55	12.65	337333	

[a] Average CPU time [s] (Alpha Station 500, 266 MHz) for each individual run (including local optimization). [b] Average global minimum encounter times [s]. [c] Global minimum encounter times of Gregurick et al. [22] divided by a factor of 7 (see text). [d] This result was obtained for global optimization starting from a configuration seed of the fully optimized 22 atom cluster [22]. All our results are for completely random starting configurations.

[2] This conclusion, to stop the global optimization at a relatively high energy and to follow it with a fast conjugate gradient local minimization, rather than to continue the global optimization to much lower energies, was also arrived at for the configuration of LJ clusters using simulated annealing as the global optimization technique [5].

for local optimization for the 23, 38 and 55 atom cluster respectively.

Table 1 compares the global minimum mean encounter time for the symbiotic algorithm with the results of Gregurick et al. [22] for a hybrid genetic algorithm employing local minimization of each individual before selection (similar to the potential energy surface transformation described by Wales and Doye [9]). The results of Gregurick et al. have been divided by a factor of 7 to account for the differences in computer hardware [3]. Average CPU times for each individual run are also listed in the table. The symbiotic algorithm finds the global minima in less time than the hybrid genetic approach, at least up to cluster sizes of 23 atoms for which CPU times have been published [4].

The symbiotic algorithm not only outperforms the hybrid genetic algorithm in the global minimum search but also provides the energy distribution of the local minima and encounter statistics, giving an indication of their attraction basin widths. Fig. 2 shows the energy distribution of the lowest energy stationary points found within an energy of 0.17 of the ground state, after local conjugate gradient minimization of 80000 separate runs of the symbiotic algorithm. The plots include about 10000 distinct minima for cluster sizes greater than 18 atoms. It was confirmed that these distributions were similar to those obtained with the genetic algorithm after minimizing locally from the the same energy values. The density of configurations (per unit energy) is plotted in the insets of Fig. 2 as a function of the cluster energy. This information is normally not available from most implementations of the genetic algorithm (which fixate on finding the global minimum) but is of considerable value in mapping out the potential energy surface and thereby elucidating the thermodynamic behavior of the cluster [28,29]. In

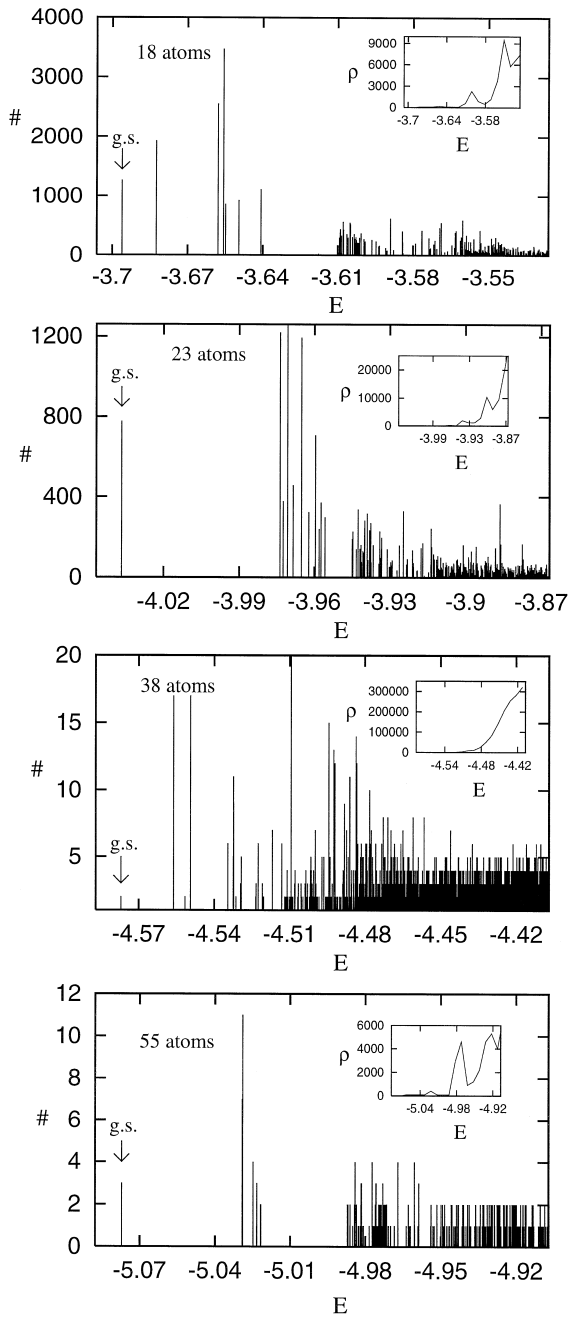

Fig. 2. Distribution in energy of the stationary points, within an energy interval of 0.17 of the ground state, found in 80000 separate runs of the symbiotic algorithm after local minimization. The ground states are marked with an arrow. The insets give the configuration density as a function of energy.

[3] Bench marks available via www.sissa.it/furio/mdbnch.html.

[4] Niesse and Mayne [26] have reported improved performance with the hybrid genetic algorithm by using two point crossover (both our and Gregurick et al. crossover is limited to the usual one point) and base−10 real variables and operators instead of binary. However, due to ambiguous reporting of CPU times, it is impossible to make a quantitative comparison here.

Table 2
Energies and the number of times (in parenthesis) encountered in 80000 runs (81624 runs for the 23 atom cluster) of the 10 lowest energy stationary points found. Saddle points are marked with a '∗', all others are stable minima.

Isomer	6	18	23	38	55
1	−2.118677 (7425)	−3.696163 (1267)	−4.036716 (778)	−4.577064 (2)	−5.077245 (3)
2	−2.050488 (71758)	−3.682476 (1935)	−3.973725 (1221)	−4.559273 (1)	−5.029169 (7)
3	−2.013172 (9) ∗	−3.657906 (2557)	−3.972523 (379)	−4.556166 (17)	−5.029149 (6)
4	−1.938385 (10) ∗	−3.655614 (3473)	−3.970773 (1261)	−4.551543 (2)	−5.029032 (11)
5	−1.889006 (50) ∗	−3.654815 (861)	−3.968442 (457)	−4.549414 (17)	−5.024851 (4)
6		−3.649542 (923)	−3.965148 (1194)	−4.534616 (6)	−5.023399 (1)
7		−3.640785 (1114)	−3.962312 (325)	−4.532567 (1)	−5.023373 (2)
8		−3.617695 (23)	−3.959557 (707)	−4.532498 (5)	−5.023352 (3)
9		−3.609328 (300)	−3.958110 (241)	−4.532387 (11)	−5.021806 (2)
10		−3.608971 (392)	−3.957323 (372)	−4.532281 (9)	−4.987240 (2)

Table 2, the energies and encounter statistics of the 10 lowest energy stationary points are listed.

4. Discussion and conclusions

Ecological or physical symbiosis between distinct organisms, once adapted on simpler fitness landscapes, played an essential role in the evolutive history of all complex biological organisms [30]. Symbiosis aids in the development of new organisms adapted to more complex niches for which mutation and crossover alone would take prohibitively long.

In the symbiotic optimization of the cluster, fitness is based on the local structure and the problem has effectively been reduced from one of $3 \times n$ variables to n problems of approximately $3 \times m$ variables each $(m \ll n)$. The fitness landscapes (potential energy surfaces) are considerably simpler and the crossover operation is optimized, acting only locally within the cell. Connectivity between cells is afforded by their overlapping nature, the symbiotic events, and the local optimization of the entire cluster. For large Lennard-Jones clusters, the results of this Letter indicate that the symbiotic approach is substantially more efficient than the standard genetic algorithm in finding low energy stable structures, and that this advantage increases with cluster size.

The symbiotic algorithm, in the configuration presented here, provides an unbiased, complete and accurate energy distribution of the low energy minima, and does not just fixate on localization of the global minimum. Nevertheless, it outperforms the hybrid genetic technique in finding the global minimum. Mean global minimum encounter times could be significantly reduced by; starting from seed configurations [22,26], minimizing locally only the lowest energy global results [9], preventing premature convergence to low energy local minima by augmenting the diversity of solutions with either a larger initial population [31] or by introducing niche interaction [2]. Techniques which would improve the overall performance of the algorithm are; relaxing the precision on the energy, hybrid local optimization of the entire cluster before selection at the symbiotic events, multi-point crossover [26,31], base-10 real variable encoding and operators, instead of binary ones [26,13], and removal of the six degrees of freedom related to the rotational and translational invariance of the cluster.

As the range of the potential increases at constant cluster size, or the size of the cluster decreases at constant potential range, the local tight coupling of variables through the fitness function is lost and to maintain optimal efficiency the ratio of the cell radii to the radius of the cluster would have to increase correspondingly. Under these conditions the imperative for symbiosis is eventually lost and the symbiotic approach merges into the genetic approach. However, such clusters are more amenable to optimization since extending the range of the potential adiabatically removes local minima, and, few atom clusters have few minima.

The symbiotic algorithm may be regarded as the genetic analogue of the divide-and-conquer method of Yang [32] in quantum density functional theory.

In this approach the system is partitioned into subsystems in physical space and the electron density determined locally for each subsystem. The computational advantage lies in the fact that the construction and diagonalization of the global Hamiltonian matrix is not required. The coupling between each subsystem can, in many cases, be shown to be minimal.

Although we stressed the advantages of a semiempirical first approach to finding the low energy structures of large clusters, there is no fundamental limitation, other than that of CPU time, in coupling the symbiotic algorithm to ab initio or density functional calculations. If the interest is in determining stable cluster geometries rather than in specific dynamical trajectories, this approach would significantly improve on the popular Car-Parinello technique [33] employing molecular dynamics.

We have also applied a similar symbiotic approach to clusters bound through an *n*-body, tight-binding, Gupta potential used in modeling transition and noble metal clusters [34]. Similar improvements in the efficiency were obtained and new global minima and low energy isomers were discovered [35]. The symbiotic algorithm should also be applicable to the more general case of a complex many-variable problem where, although the variables are independent, there is a strong coupling of variable groups through the fitness function.

Acknowledgements

The author would like to thank I.L. Garzon, R. Santamaria and J. Heras for useful comments on the manuscript. Financial support of CRAY-UNAM grant SC-003095, DGAPA-UNAM grant IN101297, and CONACYT grant 25083E is gratefully acknowledged.

References

[1] R.P. Andres, T. Bein, M. Dorogi, S. Feng, J.I. Henderson, C.P. Kubiak, W. Mahoney, R.G. Osifchin, R. Reifenberger, Science 272 (1996) 1323.
[2] D.B. McGarrah, R.S. Judson, J. Comp. Chem. 14 (1993) 1385.
[3] I.G. Kaplan, I.L. Garzon, R. Santamaria, B.S. Vaisberg, O. Novaro, J. Mol. Struct. Theochem 398-399 (1997) 333.
[4] F.E. Harris, in: Computational Methods for Large Molecules and Localized States in Solids, eds. F. Herman, A.D. McLean, R.K. Nesbet (Plenum Press, New York, 1973) p. 81.
[5] L.T. Wille, Chem. Phys. Lett. 133 (1987) 405.
[6] J. Farges, M.F. de Feraudy, B. Raoult, G. Torchet, J. Chem. Phys. 84 (1986) 3491.
[7] I.L. Garzon, M. Avalos Borja, Phys. Rev. B40 (1989) 4749.
[8] M.R. Hoare, Advan. Chem. Phys. 40 (1979) 49.
[9] D.J. Wales, J.P.K. Doye, J. Phys. Chem. A 101 (1997) 5111.
[10] R.S. Judson, M.E. Colvin, J.C. Meza, A. Huffer, D. Gutierrez, Int. J. Quant. Chem. 44 (1992) 277.
[11] B. Hartke, J. Phys. Chem. 97 (1993) 9973.
[12] Y. Zeiri, Phys. Rev. E 51 (1995) R2769.
[13] Y. Zeiri, Comput. Phys. Commun. 103 (1997) 28.
[14] J. Holland, Adaptation in Natural and Artificial Systems (University of Michigan Press, Ann Arbor, 1975).
[15] J.T. Alander, An Indexed Bibliography of Genetic Algorithms in Chemistry and Physics (University of Vaasa, Finland, 1996) available via anonymous ftp: site ftp.uwasa.fi, directory cs/report94-1, file gaCHEMPHYSbib.ps.Z
[16] D.M. Deaven, N. Tit, J.R. Morris, K.M. Ho, Chem. Phys. Lett. 256 (1996) 195.
[17] K. Michaelian, Am. J. Phys. 66 (1998) 231.
[18] P. Sutton, S. Boyden, Am. J. Phys. 62 (1994) 549.
[19] R.S. Judson, E.P. Jaeger, A.M. Treasurywala, M.L. Peterson, J. Comput. Chem. 14 (1993) 1407.
[20] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning (Addison-Wesley, Massachusetts, 1989).
[21] D.M. Deaven, K.M. Ho, Phys. Rev. Lett. 75 (1995) 288.
[22] S.K. Gregurick, M.H. Alexander, B. Hartke, J. Chem. Phys. 104 (1996) 2684.
[23] J. Farges, M.F. de Feraudy, B. Raoult, G. Torchet, Surf. Sci. 156 (1985) 370.
[24] J.P.K. Doye, D.J. Wales, Phys. Rev. Lett. 80 (1998) 1357.
[25] J.A. Northby, J. Chem. Phys. 87 (1987) 6166.
[26] J.A. Niesse, H.R. Mayne, J. Chem. Phys. 105 (1996) 4700.
[27] G. Chang, M. Lewis, Acta. Cryst. D50 (1994) 667.
[28] D.J. Wales, Science 271 (1996) 925.
[29] K.D. Ball, R.S. Berry, R.E. Kunz, F.Y. Li, A. Proykova, D.J. Wales, Science 271 (1996) 966.
[30] L. Marguilis, D. Sagan, What is Life (Simon and Shuster, New York, 1995).
[31] J. Mestres, G.E. Scuseria, J. Comput. Chem. 16 (1995) 729.
[32] W. Yang, Phys. Rev. Lett. 66 (1991) 1438.
[33] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.
[34] I.L. Garzon, A. Posada-Amarillas, Phys. Rev. B54 (1996) 11796.
[35] K. Michaelian, I.L. Garzon, N. Rendon, Energy Distribution of Minima in the Potential Energy Surface of Metal Clusters, Bulletin of the American Physical Society, March Meeting, Los Angeles, CA, March 16–20, 1998, vol. 43, No. 1, p. 363.