1189

Structure Prediction of Benzene Clusters Using a Genetic Algorithm

W. J. Pullan

Department of Mathematics and Computing, Central Queensland University,
Rockhampton, Queensland, Australia 4702

This paper describes a real coded, parallel genetic algorithm implemented to find global minimum energy structures of clusters of benzene (C_6H_6) molecules. Starting from randomly generated structures, the genetic algorithm was able to find minimum energy structures for clusters of two to fifteen benzene molecules.

1. INTRODUCTION

While considerable effort has been dedicated to the study of global optimization methods applicable to the structural optimization of atomic clusters, relatively less effort has been applied to the optimization of molecular clusters. Global optimization of molecular clusters is inherently more difficult than that of atomic clusters due to the need to determine the three angular coordinates specifying the orientation of each molecule (other than the first). In addition, for a flexible molecule, there may be intramolecular interactions which determine the shape of the molecule and which require additional parameters for their specification. Also, as molecules generally consist of more than a single type of atom, the intramolecular and intermolecular force fields contain a number of terms which are likely to result in a more complex optimization problem.

Whenever molecules are close to each other, three types of interactions occur. The long-range Coulombic interaction arises from the molecular electric potential, usually viewed as originating from net charges on the atoms, and may be either an attractive or repulsive interaction. The intermediate-range dispersion interaction due to instantaneous induced dipole interactions is always attractive, while the short-range exchange repulsion interaction from electron−electron interactions increases rapidly at close range and is responsible for the "size" of the atom. This combination of short-, intermediate-, and long-range interactions leads to a complicated intermolecular potential energy surface which includes many local minima.

This paper extends a genetic algorithm optimization method which has been successfully applied to clusters of carbon atoms,[3] pure atomic clusters,[4] mixed atomic clusters,[8] and water clusters[7] to clusters of benzene molecules. Section 2 describes the force fields used to model the interactions between benzene molecules and also overviews previous global optimization studies of clusters of benzene molecules. Section 3 describes the optimization method used in this study, while the results obtained are described and discussed in section 4. Possible future research directions and a summary are presented in section 5.

2. BENZENE CLUSTERS

For this study, a rigid planar benzene (C_6H_6) molecule was assumed, with atomic coordinates shown in Table 1[11] and defined relative to a coordinate system whose origin is
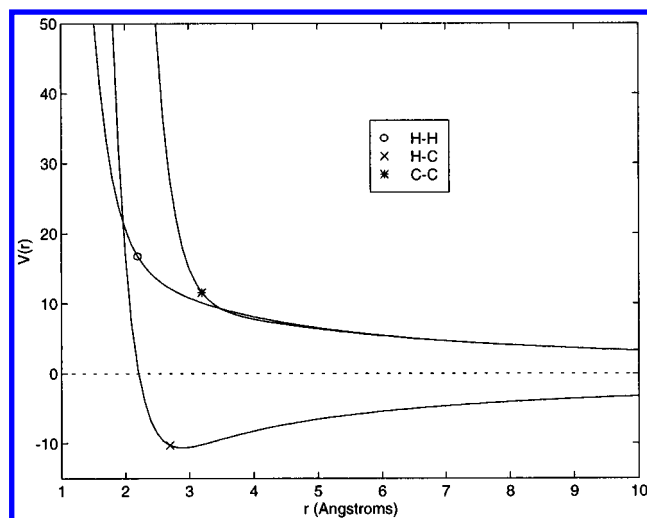

Figure 1. Benzene intermolecular atomic pair potentials for the H−H, H−C, and C−C interactions.

at the center of mass of the benzene molecule and the *xy*-plane defined by the plane of the benzene molecule.

The total potential energy for a benzene molecular cluster is

$$V = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \upsilon(r_{ij})$$

with the three components of the intermolecular force field modeled by the following atom-to-atom nonbonded potential energy[11]

$$\upsilon(r_{ij}) = B \exp(-Cr_{ij}) - Ar_{ij}^{-6} + q_i q_j r_{ij}^{-1}$$

where, for atoms *i* and *j*, $V(r_{ij})$ is the potential energy of interaction and r_{ij} is the interatomic distance. The dispersion energy coefficient is *A*, while *B* and *C* are the exchange repulsion coefficients with q_i and q_j being the net atomic charges for atoms *i* and *j*. The numeric values associated with these coefficients are dependent on the types of atoms involved in the interaction and are shown in Table 2 for the three possible interactions (C−C, H−C, H−H). Figure 1 shows the form of $V(r)$ for these three atom-to-atom interactions.

From Figure 1, the pair potential able to contribute the lowest energy is clearly that due to the H−C interaction. However, to obtain a low-energy cluster, the H−H and C−C distances also need to be relatively large. This can be

---

Table 1. Benzene Molecule Cartesian Coordinates (Å)[a]

atom	x	y	z	atom	x	y	z
1 (C)	1.3970	0.0000	0.0000	7 (H)	2.4240	0.0000	0.0000
2 (C)	−1.3970	0.0000	0.0000	8 (H)	−2.4240	0.0000	0.0000
3 (C)	−.6985	1.2098	0.0000	9 (H)	1.2120	2.0992	0.0000
4 (C)	−0.6985	−1.2098	0.0000	10 (H)	−1.2120	−2.0992	0.0000
5 (C)	−0.6985	1.2098	0.0000	11 (H)	−1.2120	2.0992	0.0000
6 (C)	0.6985	−1.2098	0.0000	12 (H)	1.2120	−2.0992	0.0000

[a] Note that the hydrogen coordinates correspond to foreshortened C−H bond distances of 1.027 Å rather than nuclear−nuclear distances of 1.097 Å.

Table 2. Benzene Molecules Nonbonded Potential Parameters[a]

param	H−H	H−C	C−C
A, kJ mol^{-1} Å^6	136	573	2414
B, kJ mol^{-1}	11677	65485	367250
C, Å^{-1}	3.74	3.67	3.60
$q(H)$	0.153 e		
$q(C)$		−0.153e	

[a] Note that the A, B, and C parameters are geometric means.
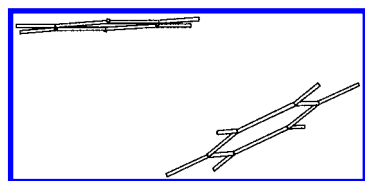
Figure 2. Low-energy configuration for the benzene dimer.

achieved by rotating one molecule to give the edge-to-plane orientation shown in Figure 2 which allows some close H−C interaction (minimum of 3.2 Å) while also maintaining H−H and C−C distances relatively large (minimum of 3.8 Å).

In previous optimization studies of benzene clusters, Williams[11] investigated benzene clusters using a Newton−Raphson local optimization method. With assumption of a center of symmetry (where possible) and with specially chosen initial configurations, local energy minima were found for benzene clusters containing 2, 3, 5, 7, 9, 11, 13, and 15 molecules. Subsequently van der Waal[10] was able to find lower minima for benzene clusters containing between 2 and 7 molecules using a method which started from configurations derived from either a 13 molecule icosahedral cluster by removal of the appropriate number of molecules or by removal of molecules from optimized clusters so obtained.

Using randomly generated initial configurations, Williams,[12] with an off-ridge eigenvector minimization with an annealing technique, found energy minima for benzene clusters of two to four molecules. This technique used a combination of simulated annealing and Newton−Raphson local optimization but, when the Hessian was no longer positive definite and the Newton−Raphson method was not guaranteed to find a minimum energy, adopted a technique of performing a linear search in the direction of the eigenvector corresponding to the largest negative eigenvalue of the Hessian. This was found to be an effective technique for quickly moving to a region where the Hessian became positive definite and the Newton−Raphson local optimizer was again reliable. Subsequently, a binary-coded genetic algorithm (GA) implemented by Xiao and Williams,[14] with randomly generated starting configurations, was able to find the currently accepted global energy minima for benzene clusters of two to four molecules.

3. OPTIMIZATION METHOD

The optimization method used in this study to optimize benzene clusters was a parallel real-coded genetic algorithm. GAs are based on the process of natural selection where the proportion of "fitter" individuals in a population tends to increase at each generation.[5] Standard GAs operate on a binary encoded version of the parameters of a problem and use genetic operators such as proportional selection (based on a fitness factor), random mutation, and crossover to generate a new population from an existing population.

Theoretical analysis and experimental results[5] show that standard (or binary encoded) GAs are efficient when applied to problems whose parameters can be encoded as short, low-order schema which are relevant to the underlying problem and relatively unrelated to schema over other fixed positions.[5] However, the benzene cluster optimization problem meets neither of these criteria because of the following:

1. Molecular positions and orientations are real valued (floating point) parameters. Using a binary coded GA requires that these positions be discretized so they can be coded as binary integers. To obtain an accuracy of 3 digits after the decimal place, for a 10 molecule benzene cluster, requires a relatively large binary solution vector of length 702 bits (13 binary bits to hold each of the 6 parameters that define the position and orientation of 9 molecules). This in turn generates a very large search space of approximately 10^{211} elements.

2. In evaluating the potential energy (and hence the fitness factor) of a benzene cluster, the position of each molecule is only meaningful in the context of the positions of all other atoms. That is, schema over different positions are not unrelated, so standard GAs can be expected to perform poorly when applied to the molecular cluster problem.[5]

As the theoretical basis for GAs was originally restricted to standard GAs, the alternative of using GAs whose parameters were encoded as real numbers has been relatively infrequently chosen. However, it has been shown recently[1] that the "implicit parallelism"[5] of standard GAs does not depend on binary encoding, and there is a growing body of evidence that GAs with real parameters have acceptable performance[2,13] provided suitable genetic operators are utilized.

With regard to molecular structure problems, this was first demonstrated by Deaven and Ho,[3] where they addressed the first problem listed above by the use of a local (deterministic) optimizer which had the effect of mapping a range of parameter values (those corresponding to energies within a "catchment basin" on the potential energy surface) to a particular set of parameter values (that corresponding to the local minimum of the catchment basin). In effect the local optimizer performs a natural discretization of the parameter

STRUCTURE PREDICTION OF BENZENE CLUSTERS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 6, 1997 **1191**

values where the interval of discretization is directly related to the size of a catchment basin. Deaven and Ho addressed the second problem identified above by the use of innovative geometric genetic crossover operators operating in the problem (phenotype) space rather than the more normal genetic crossover which operates in the parameter (genotype) space. For molecular structure problems in particular, these geometric crossover operators tend to preserve the structural relationships that have developed within individuals and provide the building blocks necessary to create individuals with higher fitness factors.

Direct evidence of the effectiveness of this type of GA, as compared to the standard GA, is provided by results obtained Gregurick et al.[6] for optimizing pure atomic clusters as compared to those obtained by Deaven et al.[4] Gregurick et al.[6], using randomly generated initial clusters with a standard GA, successfully optimized all clusters in the range $N = 2...20$. By contrast Deaven et al.,[4] using randomly generated initial clusters, were able to obtain the currently accepted global minima for all clusters in the range $N = 2...100$.

The GA used to obtain the results presented in this paper contains many of the features of the Deavon et al.[3,4] GA and is now described.

3.1. Genetic Encoding. For benzene clusters, molecules were treated as rigid bodies and, for other than the first molecule which defined the coordinate system, were encoded using six real-valued parameters. With reference to an external coordinate system, three parameters specified the center of mass of the molecule, and the three remaining parameters specified the molecular orientation using Eulerian angles. This encoding requires $6N - 6$ parameters to fully specify an N molecule benzene cluster.

3.2. Genetic Crossover. The genetic crossover operator for generating new clusters translated each parent cluster so that the center of mass of the cluster was positioned at the origin of the coordinate system and then the clusters were randomly rotated around each axis of the coordinate system. Molecules located within a hemisphere, quadrant, or octant of each parent cluster were then used to create two new child clusters. For the hemisphere variant, child clusters were created using opposing hemispheres (above and below the *xy*-plane) from each of the parent clusters, which were initially translated directly away from the *xy*-plane and then moved together, in small incremental steps, with the cluster energy calculated at each step. Starting from the minimum energy configuration thus found, a BFGS local optimization of the new child cluster was performed. Child clusters constructed from quadrant and octant volume elements were created using an analogous procedure.

3.3. Genetic Mutation. The following equally likely mutation operations were applied to all child clusters with probability p_m:

(a) A combination of BFGS local minimizations and APSE[9] searches were used to perform a random search within the close structural vicinity of a cluster.

(b) A randomly selected molecule was removed, a new molecule added at a randomly selected position and orientation, and a subsequent BFGS local optimization performed.

(c) The molecule which had the highest binding energy was removed, a new molecule added at a randomly selected position and orientation, and a subsequent BFGS local optimization performed.

Chart 1

Master	Slave(s)
$g = 0$	*Start*:
For p=1..NP	**Receive**(Command, I_1, I_2)
Send(New, p)	If (Command = New)
For i = NP+1..NR	**Create**(I1)
Receive(Command, I_1, p)	**Send**(Add, I_1, Master)
Send(New, p)	Go to Start
Next:	If (Command = Generate)
$g = g + 1$	**Crossover**(I_1, I_2, I_3, I_4)
Select	**Mutate**(I_3, I_4)
For i=1..NI	**Send**(Add, I_3, I_4, Master)
For j=1..i	Go to Start
Receive(Command, I_1, I_2, p)	If (Command = Terminate)
Send(Generate, P_g[i], P_g[j], p)	Stop
if *finished*	
For p=1..NP	
Send(Terminate, p)	
Stop	
else	
go to *Next*	

(d) All Euler angles for a randomly chosen molecule were "stepped", in small increments, through their allowed range. The minimum energy configuration thus found was subsequently locally optimized using a BFGS local optimizer.

3.4. GA Parameters. The population size was always set at 12 with a probability of mutation (p_m) of 0.1, and all possible crossover pairings were performed. No individual was added to the genetic pool if its energy was within 0.05 of that of an individual already in the genetic pool.

3.5. Parallel Implementation. To obtain the computational power needed to optimize benzene clusters, the GA was implemented in a parallel computing environment. The relevant parallel aspects of the algorithm are now presented using the following definitions: g, generation number; p, slave processor number; NP, number of slave processors; NR, number of initial randomly generated individuals; I_i an individual; P_g, current population for generation g; $P_g[i]$, an individual of P_g; NI, number of individuals in $P_{*}g$. Thus, the parallel GA is implemented as given in Chart 1. Details of the elements of the algorithm in Chart 1 are as follows:

(a) **Send** transmits a command and optional data to the nominated processor.

(b) **Receive** receives a command and optional data and, in the case of the master processor, adds all received individuals to a temporary genetic pool T_g.

(c) **Create** creates a new molecular cluster by randomly placing each molecule within a cube of side 30 Å. These clusters are then locally optimized using a BFGS optimizer.

(d) **Select** chooses clusters form P_{g-1} and T_{g-1} for inclusion in P_g. Selection is based on cluster energy with the proviso that only one cluster will be chosen with a given energy (plus or minus a small Δ). At the completion of **Select**, T_g is cleared.

(e) **Crossover** applied the genetic crossover operators described in subsection 3.2 to clusters from P_g to generate new clusters for the subsequent **Mutate** operator.
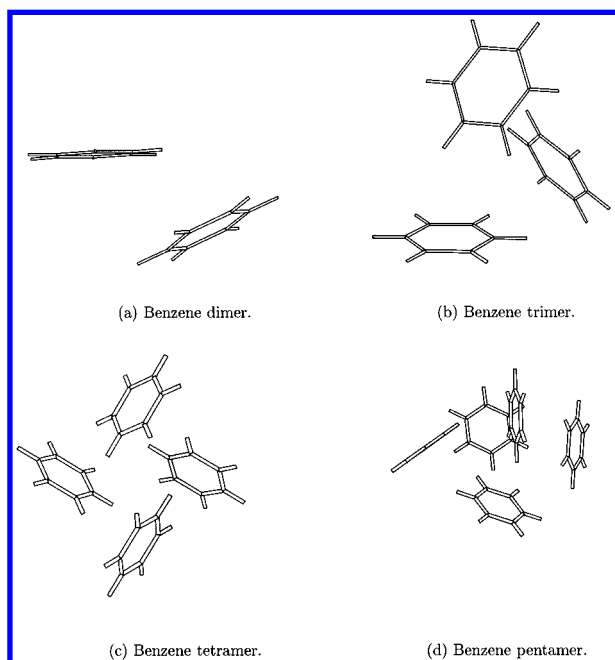
(f) **Mutate** performs BFGS local minimization, with a performance related early termination option, on all new clusters. In addition, with probability p_m, one of the equally likely mutation operators described in subsection 3.3 was applied.

An important efficiency feature of the parallelization algorithm is that there are no synchronization points ensuring that at no point will have a slave processor enter a wait state.

Table 3. Optimized Potential Energies for $N = 2...15$ Benzene Molecular Clusters[a]

	energy (kJ mol^{-1})	
N	refs 11 and 10	this study
2	−11.0	−11.0
3	−32.1	−32.1
4	−55.6	−55.6
5	−79.1	−79.1
6	−106.1	−106.4
7	−134.1	−134.1
8		−161.4
9	−161.0	−190.8
10		−221.5
11	−237.0	−251.8
12		−286.3
13	−325.3	−324.6
14		−347.5
15	−372.3	−371.3

[a] The energy values shown for this study were those attained at the completion of 50 generations.

(a) Benzene hexamer.   (b) Benzene septamer.

(c) Benzene octamer.   (d) Benzene nonamer.

Figure 4. Structures found during optimization of benzene hexamer, septamer, octamer, and nonamer.
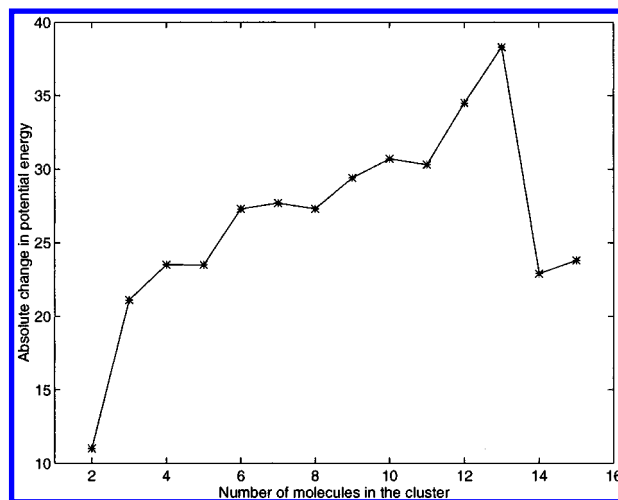
Figure 5. Absolute change in the minimum potential energy as the number of molecules in the benzene cluster increases.

(a) Benzene dimer.   (b) Benzene trimer.

(c) Benzene tetramer.   (d) Benzene pentamer.

Figure 3. Structures found during optimization of benzene dimer, trimer, tetramer, and pentamer clusters.

4. RESULTS

Table 3 contains all previously found optimal potential energy values in addition to those obtained by this GA after 50 generations had elapsed for each cluster configuration. As an indication of the computational requirements of the GA, to optimize a 10 molecule benzene cluster takes approximately 30 min on a 500 MHz DEC. Alpha workstation.

The optimized structures found for benzene clusters in the range $N = 2, ..., 8$ are shown in Figures 3 and 4. The predominant feature of these benzene cluster configurations is the tendency toward close C⋯H approaches due to the Coulombic interaction between the positively charged hydrogen atoms and the negatively charged carbon atoms. This effect favors an edge-to-face structure, where the hydrogen atoms at the edge of one molecule are pointed toward the face of the other molecule. This gives rise to large dihedral angles between the molecular planes and is characteristic of the "herringbone" packing relationship found in crystalline benzene and many other crystals.[11] In more detail, the optimized structures for benzene clusters are as follows: Benzene dimer (Figure 3a) (a dihedral angle of 26.3° and an intermolecular center-to-center distance of 4.69 Å); benzene trimer (Figure 3b) (all dihedral angles are 61.1°, and all intermolecular center-to-center distances are 4.98 Å); benzene tetramer (Figure 3c) (this is a tetrahedral configuration with dihedral angles of 77.5° (for adjacent pairs) and 55.3° (for diametrically opposing pairs); the respective intermolecular distances are 5.04 and 5.76 Å); benzene pentamer (Figure 3d) (structurally this cluster can be viewed as two fused tetrahedra sharing a common face; intermolecular distances within the tetrahedral cluster range from 5.0 to 5.7 Å and dihedral angles range from 32 to 95°); benzene hexamer (Figure 4a) (this structure is a flattened octahedron where the average intermolecular distance in the central square is 5.8 Å); benzene septamer (Figure 4b), benzene octamer (Figure 4c), and benzene nonamer (Figure 4d) (the basic structure of all these benzene clusters is a partial icosahedral shell surrounding a central molecule).

Figure 5 shows the absolute change in the minimum potential energy as the cluster size increases. The absolute

change in minimum potential energy increases as the partial icosahedral shell around the central molecule is constructed until, at $N = 13$, the shell is completed. Subsequent molecules are positioned in an outer icosahedral shell with a correspondingly smaller absolute change in cluster potential energy.

Given that, using random initial configurations, the largest benzene cluster previously optimized was $N = 4$ by Xiao and Williams;[14] the GA performed well by successfully finding all currently accepted global minima for $N = 2, ...7$, and also obtained values for $N = 8, ..., 15$. In fact, with the exception of $N = 13$ and 15, the results obtained by this GA either matched or improved all previous results. In contrast to most previous optimization studies, this GA started from completely random initial configurations and used no heuristic information during the optimization process.

5. CONCLUSION

This paper extended a genetic algorithm optimization method which has been successfully applied to clusters of carbon atoms,[3] pure atomic clusters,[4] mixed atomic clusters,[8] and water clusters[7] to clusters of benzene molecules. With regard to previous global optimization studies of these clusters, the GA reproduced existing results and produced some new lower potential energy minima.

While the GA was able to successfully optimize benzene clusters, further improvements can be made in both the computational efficiency and the optimization method. Currently almost all of the processor time is used by the BFGS local optimizer. Improvements in efficiency could be implemented using partial optimization techniques and early stopping of some local optimizations when sufficient gains were not being achieved.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Antonisse, J. A. New Interpretation of Schema Notation That Overturns Binary Encoding Constraint. *Proceedings of the Third International Conference on Genetic Algorithms*; George Mason University, Fairfax VA; Morgan Kaufmann: 1989.

(2) Janikow, C. Z.; Michalewicz, Z. In An Experimental Comparison of Binary and Floating Point Representations in Genetic Algorithms. *Proceedings of the Fourth Conference on Genetic Algorithms*, San Mateo, CA; Morgan Kaufman: 1991.

(3) Deaven, D. M.; Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett*. **1995**, *75*, 288−291.

(4) Deaven, D. M.; Tit, N.; Morris, J. R.; Ho, K. M. Structural Optimization of Lennard-Jones Clusters by a Genetic Algorithm. *Phys. Chem. Lett*. **1996**, *256*, 195.

(5) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley Publishing Co. Inc.: Reading, MA, 1989.

(6) Gregurick, S. K.; Alexander, M. H.; Hardtke, B. Global Geometry Optimization of $(Ar)_n$ and $B(Ar)_n$ Clusters using a Modified Genetic Algorithm. *J. Chem. Phys*. **1996**, *104*, 2684−2691.

(7) Pullan, W. J. Global Optimization Applied to Molecular Architecture. Ph.D. Thesis, Department of Mathematics and Computing, Central Queensland University, **1996**.

(8) Pullan, W. J. Energy Minimization of Mixed Argon-Xenon Microclusters Using a Genetic Algorithm *J. Comput. Chem*. **1997**, *18*, 8.

(9) Pullan, W. J. A Direct Search Method Applied to a Molecular Structure Problem. *Aust. Comput. J*. **1996**, *29*, 4.

(10) van der Waal, B. W. Computed Structure of Small Benzene Clusters. *Chem. Phys. Lett*. **1986**, *123*, 69−72.

(11) Williams, D. E. Calculated Energy and Conformation of Clusters of Benzene Molecules and Their Relationship to Crystalline Benzene. *Acta Crystallogr*. **1980**, *36*, 715−723.

(12) Williams, D. E. OREMWA prediction of the structure of benzene clusters: Transition from subsidiary to global energy minima *Chem. Phys. Lett*. **1992**, *192*, 538−543.

(13) Wong, K. P.; Wong, Y. W. Floating-Point Coding Method for Genetic Algorithms. *Proceedings of ANZIIS*-93 Perth, Western Australia; 1993.

(14) Xiao, Y.; Williams, D. E. Genetic Algorithm: A New Approach to the Prediction of the Structure of Molecular Clusters. *Chem. Phys. Lett*. **1993**, *215*, 17−24.

CI970283N