On the Use of Direct Search Methods for the Molecular Conformation Problem

J. C. Meza and M.L. Martinez
Scientific Computing Department
Sandia National Laboratories
Livermore, CA 94551-0969
meza@ca.sandia.gov

ABSTRACT

An important area of research in computational biochemistry is the design of molecules for specific applications. The design of these molecules, which depends on the accurate determination of their three-dimensional structure, can be formulated as a global optimization problem. In this study, we present results from the application of a new conformation searching method based on direct search methods. We compare these results to some earlier results using genetic algorithms and simulated annealing.

Keywords: global optimization, molecular conformation, nonlinear programming.

Published in Journal of Computational Chemistry, Vol. 15, No. 6, 627-632, 1994

1. Introduction

An important area of research in computational biochemistry is the design of molecules for specific applications. Examples of these types of applications occur in the development of enzymes for the removal of toxic wastes, the development of new catalysts for material processing, and the design of new anti-cancer agents. The design of these drugs depends on the accurate determination of the structure of biological macro-molecules. This problem, known as the molecular conformation problem, consists of finding the configuration of a molecule that yields its lowest free energy.

Under the assumption that the native structure of a molecule corresponds to a conformation for which the energy is at or near the global minimum, the molecular conformation problem can be formulated as an optimization problem. In general, one can decompose the search for a global minimum into two phases. In the first phase, we are interested in generating conformations that will be used as starting guesses for a second energy minimization phase. Using standard optimization techniques, each starting structure can be minimized to yield a final low-energy configuration. Unfortunately, because the total energy of a molecule depends on atom-atom interactions, the number of possible low-energy configurations can grow exponentially with the number of atoms and has been estimated by Hoare to be on the order of $O(e^{N^2})$ for an N-atom molecule [1].

There have been many attempts at developing efficient methods for the solution of this optimization problem. For a review of conformational searching methods see [2, 3]. Other recent attempts at conformational searching involve parallel stochastic methods as in [4, 5], and direct optimization methods as in [6, 7]. To address the multiple minimum problem, several people have attempted to use new search algorithms based on genetic algorithms [8] and simulated annealing [9]. Genetic algorithms (GA) [10] belong to a class of stochastic optimization methods based on analogies to natural selection strategies from evolution. Simulated annealing (SA) [11] is another stochastic optimization method that is based on a thermodynamic process called annealing. Both GA and SA have the property that they allow uphill directions and the iterates will therefore not necessarily decrease monotonically. This property can be exploited to allow the search methods to jump out of local wells by moving uphill. However, both of these approaches can be time consuming and depend heavily on certain carefully chosen parameters.

As has been previously pointed out [3] the generation of the starting geometries will have the greatest effect on the final configurations found. The purpose of this study is to suggest an alternative methodology to genetic algorithms and simulated annealing for conformational searching. The new approach is based on the use of a particular instance of a class of methods known as pattern search methods. The new approach which uses a method known as the parallel direct search method [12] is robust, depends on only one parameter, and is easily parallelized.

2. Numerical Methods

2.1 Model Problem

For the purpose of this study, it was helpful to develop a model problem for which we could easily estimate the global minimum and which could easily be scaled while still retaining the same functional characteristics. We used a two-dimensional polymer consisting of N atoms connected by rigid rods of unit length. The function used to describe the energy of this system is given by a pairwise additive function consisting of Lennard-Jones potentials, that is,

$$V(\Theta[r]) = \sum_{i,j>i}^{N} \left[\left(\frac{\alpha}{r_{ij}} \right)^{12} - 2 \left(\frac{\alpha}{r_{ij}} \right)^{6} \right], \tag{1}$$

where r_{ij} is the distance between the atoms x_i and x_j, Θ is the bond angle between any three consecutive atoms, and α is a constant. In this study, we have conducted the conformational search in internal coordinates which has the advantage of implicitly enforcing the bond length constraints. If we set $\alpha = 1$, the minimum for a given pair of atoms will correspond to a unit distance, and the global minimum corresponds to closed hexagonal packs with unit spacing. This test problem has two nice features: 1) the dimension of the problem can be easily adjusted, and 2) the minimum energy can be easily deduced for any given dimension.

For this model problem, the parameter space can be visualized as having two easily distinguished regions. The first region is characterized by high energies and corresponds to configurations that are knotted. The second region is characterized by energies less than zero and corresponds to un-knotted configurations. The probability that a random configuration will have at least one knot (and therefore fall into the first region) is a function of the number of atoms in a molecule with that probability approaching one for even moderately sized molecules. As an example, for a molecule with 61 atoms, of 10,000 randomly generated conformations, only one conformation had an energy less than 0.0.

2.2 Direct Search Methods

Direct search methods belong to a class of optimization methods that do not compute derivatives. Examples of direct search methods are the Nelder-Mead Simplex method [13], Hooke and Jeeves' pattern search [14], the box method [15], and Dennis and Torczon's parallel direct search algorithm (PDS). The PDS algorithm can be viewed as an intelligent adaptive grid search algorithm employing a multi-sided simplex.

Starting from an initial simplex S_o, the function value at each of the vertices in S_o is computed and the vertex corresponding to the lowest function value, v_o, is determined. Using the underlying grid structure, the simplex S_o is rotated 180° about v_o and the function values at the vertices of this rotation simplex, S_r, are compared against v_o. If one of the vertices in the simplex S_r has a function value less than the function value corresponding to v_o, then an expansion step to form a new simplex, S_e, is attempted in which the size of S_r is expanded by some multiple, usually 2. The function values at the vertices of S_e are compared against

the lowest function value found in S_r. If a lower function value is encountered, then S_e is accepted as the starting simplex for the next iteration; otherwise S_r is accepted for the next iteration. If no function value lower than the one corresponding to v_o is found in S_r, then a contraction simplex is created by reducing the size of S_o by some multiple, usually 1/2, and is accepted for the next iteration.

Because PDS only uses function comparisons it is easy to implement and use. Since the rotation, expansion, and contraction steps are all well-determined it is possible to determine ahead of time a set of grid points corresponding to the vertices of the simplices constructed from various combinations of rotations, expansions, and contractions. Given this set of grid points, called a search scheme, the PDS algorithm can compute the function values at all of these vertices in parallel and take the vertex corresponding to the lowest function value. An interesting consequence of this approach is that the PDS algorithm can jump out of local wells by using a large enough search scheme size. By varying the size of the search scheme one can therefore use the PDS algorithm as a means of efficiently generating conformations in a manner similar to GA and SA.

It is also worthwhile to contrast PDS with grid search methods. In a grid search method the grids are generated by starting with a fixed molecule and systematically varying one of the parameters. This method works well for small molecules but becomes computationally prohibitive for larger molecules. The grid in PDS however is adaptive and will automatically change in response to the contours of the energy surface.

3. Numerical Results

In an earlier study [8], it was noted that both genetic algorithms and simulated annealing outperform random search for large molecules. However, both approaches were computationally expensive and depended heavily on carefully chosen certain parameters. In this study, we performed numerical experiments to compare direct search methods against both genetic algorithms and simulated annealing. The PDS code we used is a slight modification to a code developed by Torczon and obtained from the Center for Research on Parallel Computation at Rice University. For the energy minimization phase we used several well-known optimization methods, including a conjugate gradient method, a limited memory BFGS (LBFGS) method, and a quasi-Newton method, SUMSL, developed by D. Gay. Details of these methods can be found in [16, 17, 18]. All numerical tests were run on SGI workstations using IEEE double precision arithmetic with a machine precision, $\mu \approx 1.1 \cdot 10^{-16}$.

Using the model problem (1) we tested various combinations of methods for molecules with different numbers of atoms. We studied molecules of 19, 37, and 61 atoms because the minima correspond to configurations that are hexagonal closed packs of radii 2, 3, and 4. Each test case consisted of running a set of trials starting from a set of randomly chosen starting points. To provide consistency across all of the test cases a set of 1000 configurations was generated from a uniform random distribution and stored in a file. This file was then used as the set of initial guesses for all of the test cases.

Because of the different methods used there were three different stopping criteria used.

Table I: Comparison of different PDS search size schemes (SSS) for 61 atom case.

Method	Trials	Minimum	% Negative
PDS(118)	200	-92.63	0.5
PDS(295)	200	-115.95	45.0
PDS(590)	200	-113.94	78.0
PDS(1180)	100	-115.02	83.0
PDS(2360)	100	-117.47	85.0

The PDS method uses a step tolerance, that is, the method terminates whenever the following condition is met:

$$\frac{1}{\Delta}\left(\max_{i,j}||v_i - v_j||_2\right) \leq \text{XTOL},$$

where $\Delta = \max(1, ||v_0||_2)$, and v_0 is the initial guess. The conjugate gradient method from [19] used a convergence criteria based on the function values:

$$2\frac{|f_{k+1} - f_k|}{|f_{k+1}| + |f_k| + \epsilon} \leq \text{FTOL},$$

where ϵ is equal to the machine precision. For the LBFGS method, the convergence test used consists of a test on the gradient, that is,

$$||g_k|| \leq \text{GTOL}\left(\max(1, ||x_k||)\right).$$

Unless otherwise indicated, all of our tests used the following tolerances: XTOL $= 10^{-3}$, FTOL $= 10^{-5}$, and GTOL $= 10^{-5}$.

3.1 Search Scheme Size

The first question we wished to address is whether the size of the search scheme has an effect on the geometries generated by PDS. Clearly the larger the scheme size the larger the sampling of the parameter space, at the expense of each iteration being more costly. In the first set of test cases, we chose the simplex shape to be regular, allowing PDS to construct the entire initial simplex from the initial random configuration. We considered search scheme sizes (SSS) of $2m$, $5m$, $10m$, $20m$, and $40m$ where $m = N - 2$ for a molecule of N-atoms and ran 100 trials for each test case. Table I contains the results for the 61 atom case.

As Table I shows, all of the test cases yielded configurations with negative energies. For the first three tests, we ran 200 trials while the last two test cases used 100 trials. Each trial typically converged in about 50 iterations and took about 1-3 minutes of CPU time per trial. The fourth column of Table I also displays the percentage of all trials that yielded a negative energy. In a sense, this is an indication of how robust the method is. We note that

Table II: Comparison of PDS versus GA and SA for 19, 37, and 61 atom test cases.

Method		Minimum		
		19 atoms	37 atoms	61 atoms
GA	CG	-44.3	-97.3	-166.6
SA	CG	-44.2	-94.8	-164.4
PDS	CG	-45.3	-95.3	-142.6
Best Known		-45.3	-98.3	-170.5

with a search scheme size of $10m$ approximately 80% of the trials had configurations with negative energies. As the search scheme size increases we get a higher percentage of trials with negative energies but at the cost of substantially more CPU time per trial. Figure 1 contains a representative configuration for the 61 atom test case computed by PDS using a search scheme size of 2360 points.

3.2 Energy Minimization

The next step in our numerical experiments was to test the combination of PDS with an energy minimization phase. For the purposes of a direct comparison with GA and SA, we ran a test case using PDS with the same CG routine used in [8]. The lowest energies found are given in Table II and compared against the best known global minimum energy.

For the 19 and 37 atom test cases, the results for all three methods are similar in terms of the final energies. In the 61-atom case, the various methods start to exhibit greater differences. In this case, the lowest energies computed by GA with CG and SA with CG are much closer to the best known global minimum than PDS with CG. Table II however does indicate that PDS can be used in a similar manner to GA or SA to generate good starting conformations for the energy minimization phase.

The next set of tests combine PDS with gradient methods. All of these tests use PDS with a value of SSS of $40m = 2360$. From tests using random search we can predict that it would not be useful to do an energy minimization starting from a configuration that has a high energy. We therefore modified our algorithm to set a user-defined energy tolerance such that if the value returned from the PDS algorithm is higher than this energy tolerance, the minimization phase is skipped.

We also note that PDS was modified so that it would restart after it had converged using the best vertex as its initial guess for the restart. To allow for greater flexibility, we allowed the user to set the maximum number of restarts. In these tests we set the number of restarts equal to 5.

The results of these tests are displayed in Table III. The first observation we make is that PDS is not effective as an energy minimization algorithm. Even with restarts, PDS does not

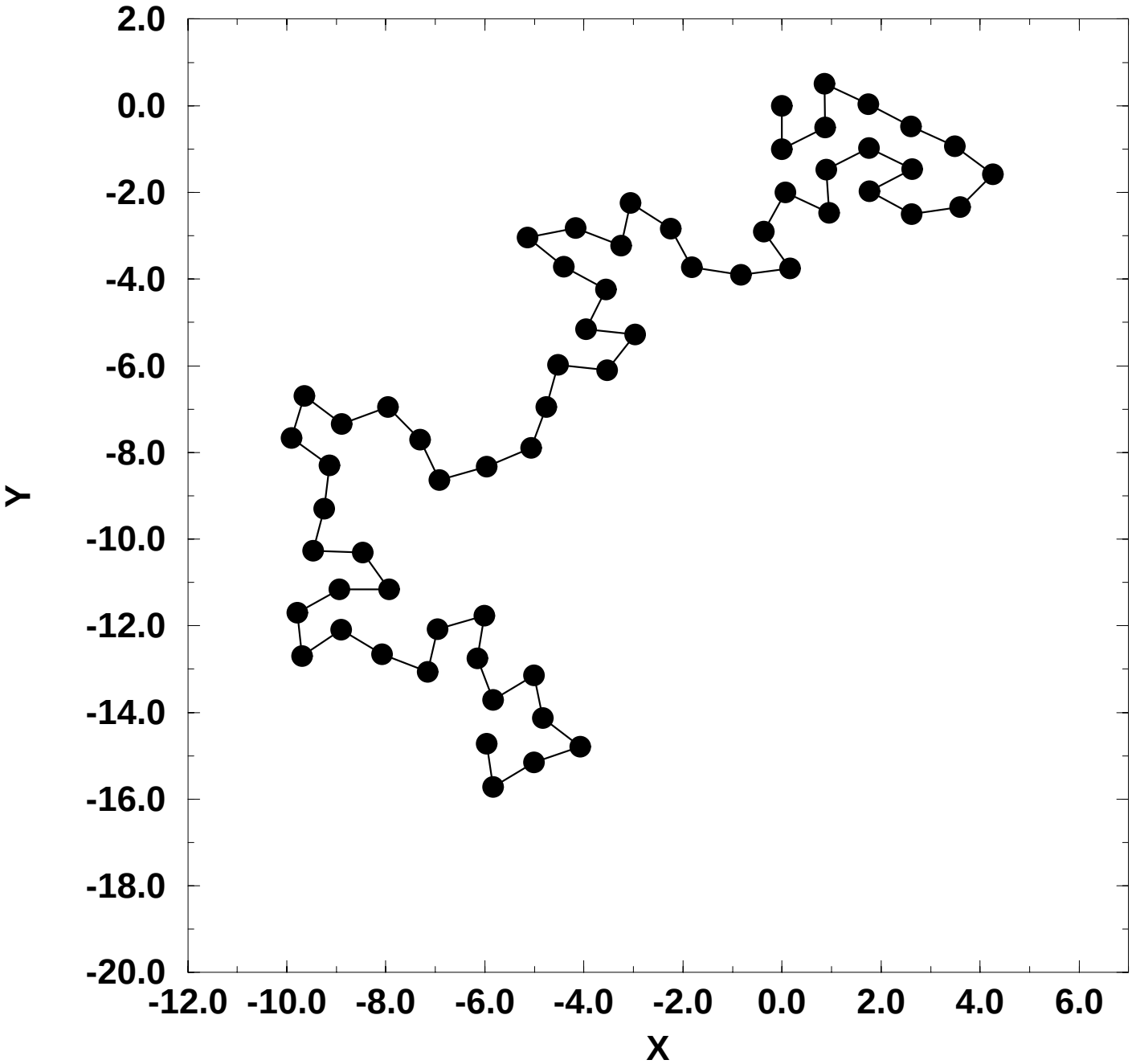

Figure 1: Conformation computed for a 61 atom test case using PDS(2360).

Table III: Minimum energies found using PDS for 61 atom test case.

Method		Energy
PDS(2360)	PDS(2360)	-117.5
PDS(2360)	CG	-142.6
PDS(2360)	LBFGS	-167.2
PDS(2360)	SMSNO	-166.0
PDS(2360)	SUMSL	-163.6

get a solution near a minimum. Theoretically PDS should not converge to a non-minimizer so this behavior seemed unusual at first. In fact, if the search scheme size is increased at the same time that a restart occurs, then PDS will start to progress towards a minimum again, although at an extremely slow rate.

For the gradient based methods, the results are all similar in terms of the final energy. The combination of PDS with LBFGS yielded the lowest energy overall, but not substantially better than the other methods. The difference between SUMSL and SMSNO lies in the availability of first derivatives. The SMSNO method uses finite differences to compute gradients, whereas SUMSL uses analytic gradients. For these test problems, we used ADIFOR to generate analytic derivatives [20]. Figure 2 contains a representative configuration computed by using a combination of PDS(2360) and SUMSL.

3.3 Distribution of Minima

Although our main goal is to find the global minimum it is also important to find any local minima that are close to the global minimum. One way of depicting this result is to generate a distribution of the local minima found for each method. Figure 3 contains the distributions for the test cases corresponding to the 61 atom molecule. The three curves using PDS(2360) all compare favorably with the distribution corresponding to the test results using GA. In fact, all four curves generate distributions with at least 80% of the final configurations having negative energies. Simulated annealing (SA) is the only method that does not do well on this problem. We also note that of these methods, the combination of PDS(2360) with SUMSL yields the best distribution with approximately 80% of the minima having energies below -144.0.

4. Conclusions

We can make several conclusions from the numerical results. In [8], we studied GA and SA with the conclusion that GA could be used for conformational searching. In this study, we showed that PDS can also be used for conformational searching and performs at least as

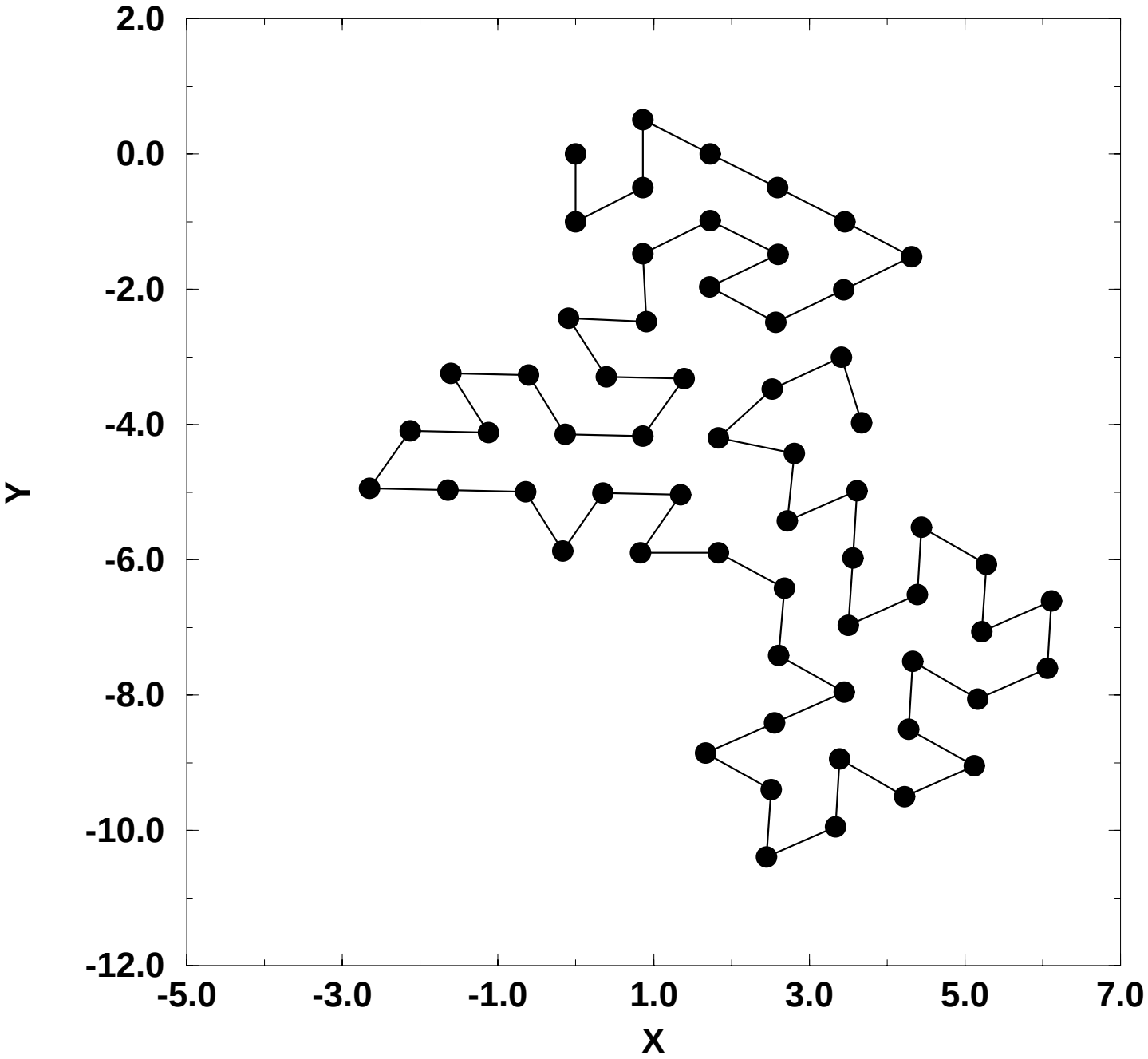

Figure 2: Final configuration computed using PDS(2360) + SUMSL for 61 atom test case.

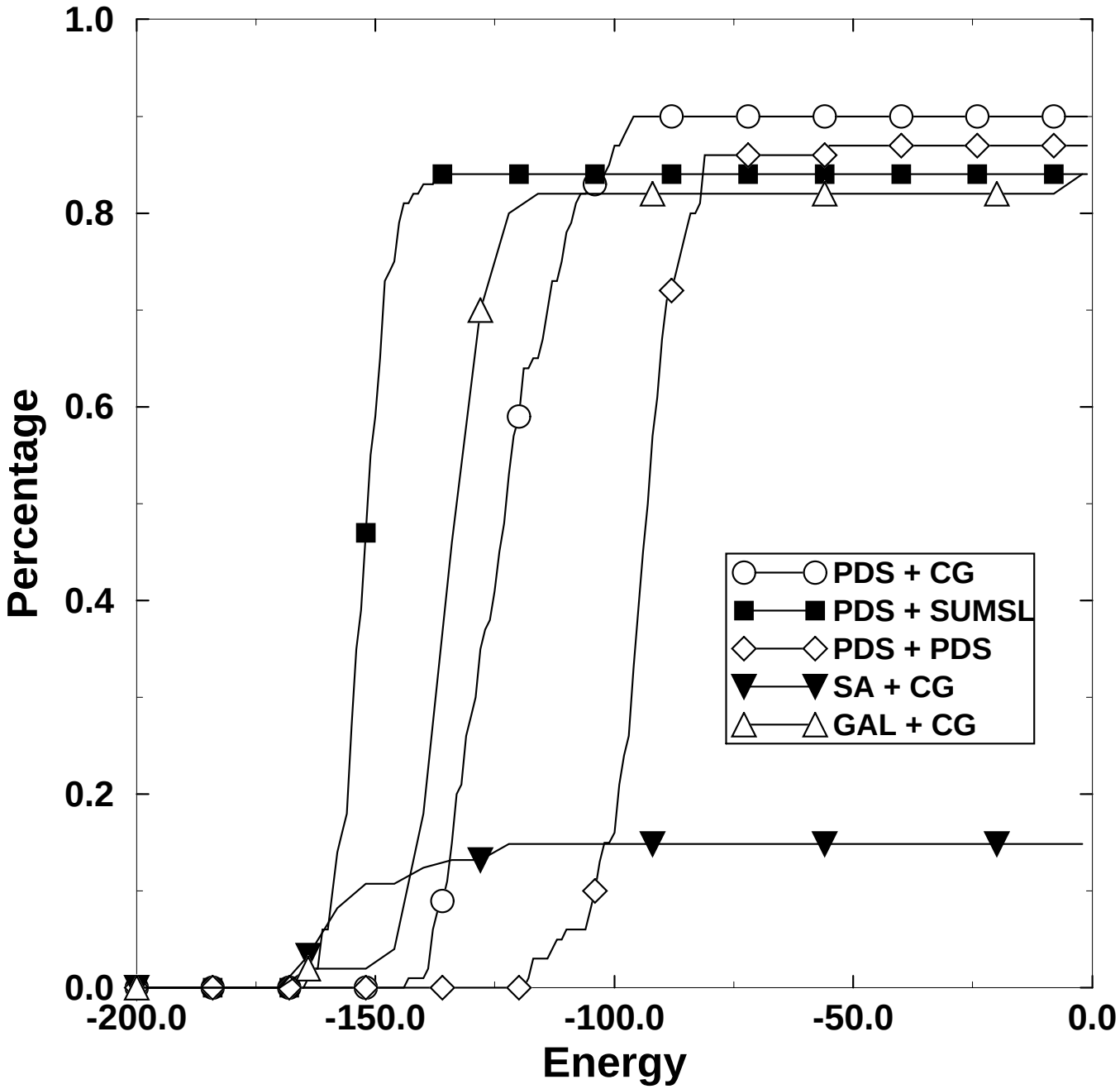

Figure 3: Distribution of minima for 61 atom test case.

well as GA and substantially better than SA for larger molecules. One of the advantages of the PDS algorithm is that the user only has to decide on the value of one parameter. This parameter, the search scheme size, has some effect on the convergence rate, but typically the parameter can be chosen to be some small multiple of the dimension of the problem. The PDS algorithm can also be easily parallelized allowing for even greater computational efficiency.

We feel that these results are suggestive of the power of these new optimization methods for conformational searching for large molecules. Further research will concentrate on new experiments using more realistic molecules. In particular, it would be interesting to do a direct comparison with the methods used in [3] for computing the conformations of cycloheptadecane.

REFERENCES

[1] M.R. Hoare. Structure and dynamics of simple microclusters. *Adv. Chem. Phys.*, 40:49–135, 1979.

[2] A.E. Howard and P.A. Kollman. An analysis of current methodologies for conformational searching of complex molecules. *J. Med. Chem.*, 31:1669–1675, 1988.

[3] Martin Saunders, K.N. Houk, Yun-Dong Wu, W. Clark Still, Mark Lipton, George Chang, and Wayne C. Guida. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J. Am. Chem. Soc.*, 112:1419–1427, 1990.

[4] R.H. Byrd, E. Eskow, R.B. Schnabel, and S.L. Smith. Parallel global optimization: Numerical methods, dynamic scheduling methods, and applications to molecular configuration. Technical Report CU-CS-553-91, University of Colorado at Boulder, 1991.

[5] R.H. Byrd, E. Eskow, and R.B. Schnabel. A new large-scale global optimization method and its application to Lennard-Jones problems. Technical Report CU-CS-630-92, University of Colorado at Boulder, 1992.

[6] R.S. Maier, J.B. Rosen, and G.L. Xue. Discrete-continuous algorithm for molecular energy minimization. Technical Report 92-031, AHPCRC, 1992.

[7] G.L. Xue. Improvement on the Northby algorithm for molecular conformation: Better solutions. Technical Report 92-055, University of Minnesota, 1992.

[8] R.S Judson, M.E. Colvin, J.C. Meza, A. Huffer, and D. Gutierrez. Do intelligent configuration search techniques outperform random search for large molecules? *International Journal of Quantum Chemistry*, 44:277–290, 1992.

[9] S.R. Wilson and W. Cui. Applications of simulated annealing to peptides. *Biopolymers*, 29:225–235, 1990.

[10] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, 1989.

[11] S. Kirkpatrick, C.C. Gelatt, and M.P Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[12] J.E. Dennis and V.J. Torczon. Direct search methods on parallel machines. *SIAM J. Optimization*, 1(4):448–474, 1991.

[13] J. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7:308–313, 1965.

[14] R. Hooke and T.A. Jeeves. Direct search solution of numerical and statistical problems. *J. Assoc. Comp. Mach.*, 8:212–229, 1961.

[15] G.E.P. Box and K.B. Wilson. On the experimental attainment of optimum conditions. *J. Royal Statistical Society, Series B*, 13:1–45, 1951.

[16] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, New Jersey, 1983.

[17] D.C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.

[18] Tamar Schlick. *Optimization Methods in Computational Chemistry*, chapter 1, pages 1–71. Reviews in Computational Chemistry. VCH Publishers, 1992. (ed. Kenny B. Lipkowitz and Donald B. Boyd).

[19] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes*. Cambridge University Press, Cambridge, 1987.

[20] C. Bischof, A. Carle, G. Corliss, A. Greiewank, and P. Hovland. ADIFOR–generating derivative codes from Fortan programs. *Scientific Programming*, 1(1):11–29, 1992.