UC-411

SAND93-8233
Unlimited Release
Printed May, 1993

A Direct Search Method for the Molecular Conformation Problem

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

ABSTRACT

An important area of research in computational biochemistry is the design of molecules for specific applications. The design of these molecules depends on the accurate determination of their three-dimensional structure or conformation. Under the assumption that molecules will settle into a configuration for which their energy is at a minimum, this design problem can be formulated as a global optimization problem. The solution of the molecular conformation problem can then be obtained, at least in principle, through any number of optimization algorithms. Unfortunately, it can easily be shown that there exist a large number of local minima for most molecules which makes this an extremely difficult problem for any standard optimization method.

In this study, we present results for a direct search method applied to a molecular conformation problem. We compare the new method against genetic algorithms and simulated annealing. The major result of this study is that the direct search method when used in combination with standard enery minimization algorithms can find a large number of low enery conformations on a test problem.

Contents

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. Unfortunately, since 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 [7].

The purpose of this study is to consider the use of a new optimization algorithm for searching conformation space. For a recent review of conformational searching see [**?**, **?**]. Other recent attempts at conformational searching involve parallel stochastic methods as in [3, 2], and direct optimization methods as in [11, 14].

We will decompose the search for a global minimum into two distinct phases: a global search phase and a local energy minimization phase. In the global search phase or global phase, we are interested in generating a distribution of good initial guesses for the second phase. In the energy minimization pahse we assume that the initial guess provided by the global phase is inside a region of attraction for some local minima that is close to the global minimum.

2. Numerical Methods

We first describe the model problem used and some of its characteristics. As in [8], 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 described by the bond angles between adjacent pairs of atoms. In this way, the nonlinear constraints are easily enforced at the expense of a coordinate transformation each time a function evaluation is needed. 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 test 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 low energies 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. For a molecule with 61 atoms, for example, the probability is already greater than 99.9999% that it will have at least one knot. Additionally, the energies for two configurations in the two regions can differ by $15 - 16$ orders of magnitude.

These points are worth noting because previous numerical studies showed that it was difficult for optimization algorithms to reach low energy conformations from random high-energy configurations. For the purposes of this paper, we will categorize optimization methods as either stochastic or deterministic. Stochastic methods incorporate a random element in their search strategies. Deterministic methods are those that do not incorporate a random element into their behavior. The rest of this section is devoted to briefly describing these methods.

2.1 Stochastic Methods

Genetic algorithms(GA) [6] are optimization methods based on analogies to natural selection strategies from evolution. The idea is to encode the problem parameters as a binary bit string called a gene. A population of these binary genes are randomly generated and then ranked according to a measure of fitness which in most cases is taken to be the value of the objective function. A new generation of genes is then created using two concepts called crossover and mutation. In crossover, two parent genes are chosen and two new genes are produced by combining a subset of the genes from one parent gene with a subset from the other parent. Mutation represents the changing of a single bit of a binary gene.

Simulated annealing [9] (SA) is another stochastic optimization method and is based on a thermodynamic process called annealing. It is known that whenever a liquid is allowed to cool slowly, it will settle into its lowest-energy state or global minimum. This physical process can be mimicked in an optimization algorithm by allowing the algorithm to accept steps that not only decrease the function value, but with some probability will also accept steps that increase the function value. The probability of accepting a step that increases the objective function value depends on another function of a parameter known as the temperature. This function is designed such that for high values of the temperature the probability of accepting a step that increases the objective function value is large and for low temperatures the probability is small.

Both GA and SA have the property that they allow uphill directions and the iterates will therefore not necessarily decrease monotonically. This property is sometimes used to justify the assertion that GA and SA are global optimization methods because they can jump out of local wells by moving uphill. The use of genetic algorithms for conformational searching has been studied in [8] and the use of simulated annealing has been studied in [**?**].

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, pattern search, and the box method. Dennis and Torczon's parallel direct search (PDS) algorithm is an extension of the Nelder-Mead simplex method that uses the concept of searching in several directions simultaneously. Starting from an initial simplex, S_0, the function value at each of the vertices of S_0 is computed and the vertex corresponding to the lowest function value, v_0, is determined. The simplex S_0 is rotated 180° about v_0 and the function values at the vertices of this rotation simplex, S_r, are compared against v_0. If one of the vertices in S_r has a function value lower than v_0, then the simplex, S_r, is expanded to form the simplex S_e. 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 at the time that S_r is constructed no function value lower than v_0 is found in S_r, then a contraction simplex, S_c is created by reducing the size of S_0 by some multiple, and the contracted simplex is accepted as the new simplex for the next iteration.

Because PDS only uses function comparisons it is relatively easy to implement and use. Since the rotation, expansion, and contraction steps are all well-determined it is also possible to determine ahead of time a set of grid points corresponding to the vertices of 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. One approach therefore is to take the scheme size to be equal to the number of processors available. Another approach, which we have followed, is to vary the size of the search scheme in an attempt to sample the parameter space for the global minimum. Used in this manner, PDS has the property that it can jump over barriers in what would amount to uphill directions. The PDS algorithm can also be shown to converge under some mild assumptions [13].

2.3 Gradient Methods

In the case that derivatives are available analytically or can be easily computed, it is useful to take advantage of this higher order information during the energy minimization phase. In this study we used a conjugate gradient method from [12], a limited memory BFGS method developed by Nocedal, and a quasi-Newton method with a BFGS update formula developed by Gay. The last two methods in particular possess good local convergence properties. Details of these methods can be found in [5, 10, **?**].

3. Numerical Results

In [8] it was noted that a combination of algorithms yielded better results than a single method. In this study, we explore these ideas in more detail. As such, the tests include several optimization algorithms described in the previous section, as well as combinations of these

algorithms. The PDS code used is a slight modification to a code developed by Torczon and obtained from the Center for Research in Parallel Computations at Rice University. The CG code used was taken from [12]. The limited memory BFGS (LBFGS) method was provided by Nocedal [10]. The quasi-Newton method, SUMSL, was developed by D. Gay. 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. 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 [12] 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 PDS tests

The first question that arises is whether the size of the search scheme has an effect on the minimum found. Clearly the larger the scheme size the larger the sampling of the parameter space. The disadvantage is that each iteration will be more costly. In the next 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 1 contains the results for the 61 atom case.

As Table 1 shows, all of the test cases yielded configurations with negative energies. The third column of Table 1 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 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 function evaluations per iteration. Figure 3 contains a representative configuration for the 61 atom test case computed by PDS using a search scheme size of 2360 points.

3.2 Hybrid runs

The previous study [8] indicated that both GA and SA yielded good approximations to the global minimum. 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, we ran a test case using PDS with the same CG routine used in [8]. The lowest energies found are given in Table 2 and compared against the best known global minimum energy.

For the 19 and 37 atom test cases, the results 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 2 also indicates that PDS can be used in a similar manner to GA or SA, that is as the global phase of a two-phase algorithm.

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 the results using random search we can predict that it would not be useful to start a local minimization from a configuration that has a large energy value. We therefore modified our algorithm to set a user-defined energy tolerance such that if the value returned from the global phase (in this case PDS) is higher than this energy tolerance, the local phase is skipped.

We also note that PDS was modified for the local phase 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 3. The first observation we make is that PDS is not as effective in the local phase as any of the gradient based methods. Even with restarts, PDS does not 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+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 [1].

Figure 4 contains a representative configuration computed by a hybrid method using PDS(2360) for the global phase and SUMSL in the local phase.

Table 1: Comparison of different PDS search size schemes (SSS) for 61 atom case.

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

Table 2: Comparison of PDS versus GA and SA for 19, 37, and 61 atom test cases.

Method		Minimum		
Global	Local	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

Table 3: Minimum energies found using PDS as the global phase method for 61 atom case.

Hybrid Method		
Global	Local	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

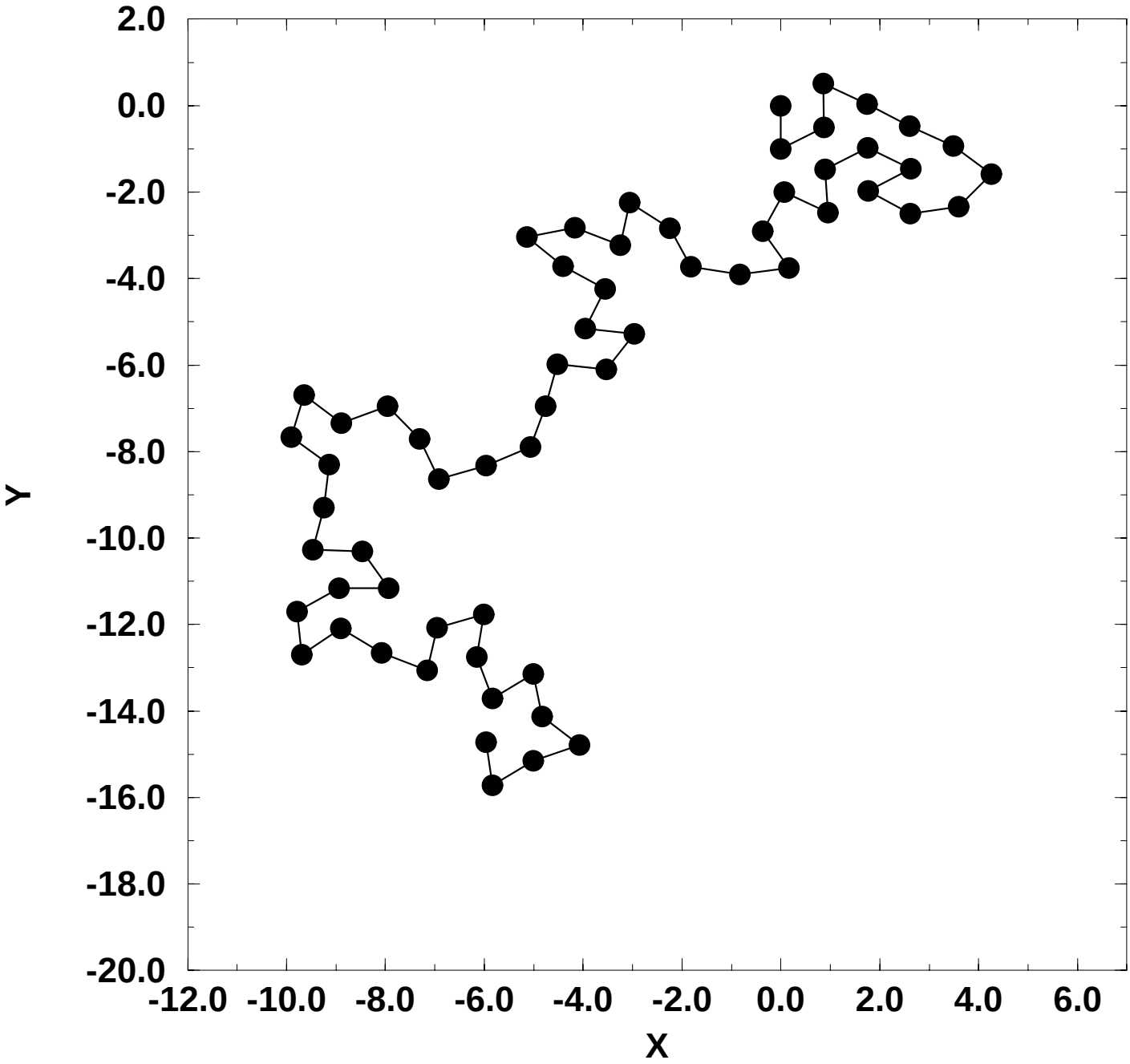

Figure 1: Mininum energy conformation computed for a 61 atom test case using PDS(2360).

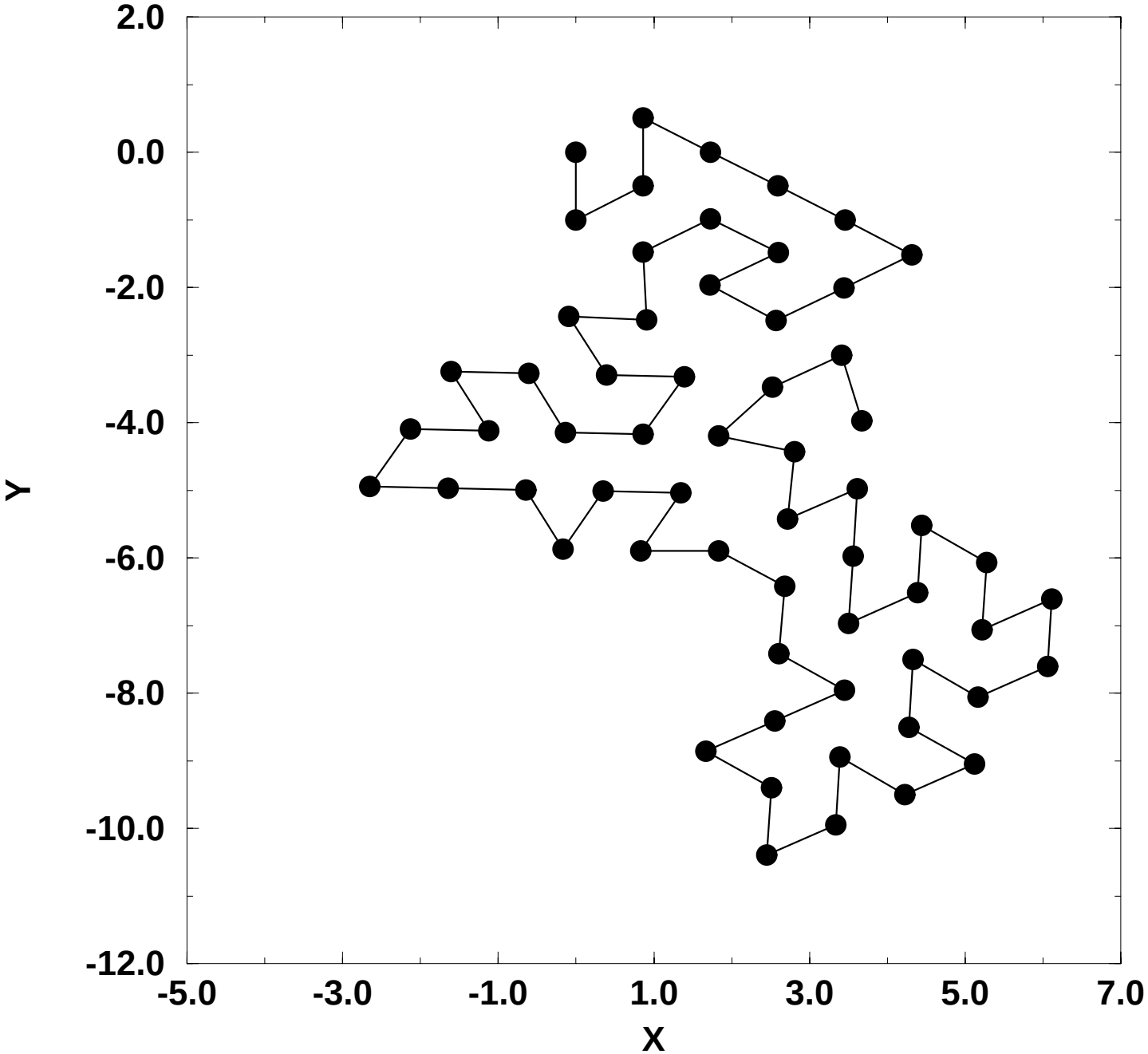

Figure 2: Final configuration after a hybrid run using PDS(2360) + SUMSL for N=61 atom test case.

3.3 Distribution of Minima

Although our main goal is to find the global minimum it is also important to find local minima that might be 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 5 contains the distributions for the test cases corresponding to the 61 atom molecule. The three curves using PDS(2360) as the global phase all compare favorably with the distribution corresponding to the test results using GA as the global phase. 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. Methods that did not use derivative information such as GA, SA, and PDS could all make the transition into the low energy region but once inside this region they exhibited a very slow convergence rate. Hybrid methods that incorporate the properties of several methods tend to work well. In [8], we studied GA and SA with the conclusion that GA served well as the first pass of a multi-pass algorithm. In this paper, we showed that PDS can also be used in a similar manner. When PDS is used in the global phase it can usually succeed in generating good starting conformations for the energy minimization phase. The effect of the search scheme size has some effect on the success ratio, but after a certain search scheme size is reached there are marginal returns. The PDS algorithm can also be shown to converge under some mild conditions.

Further research: 1) 3d problem perhaps cycloheptadecane 2) combination with molecular mechanics 3) parallel versions

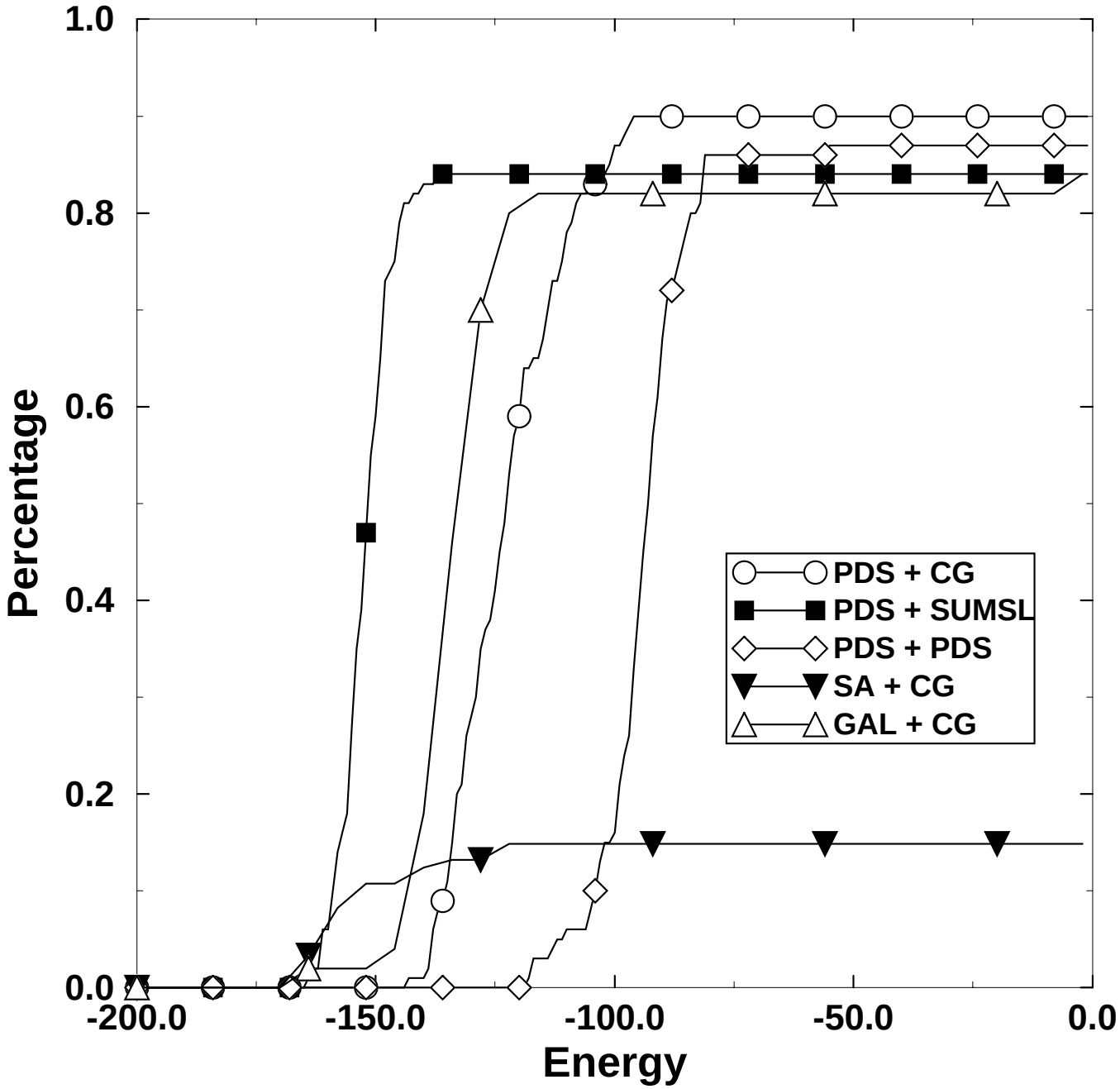

Figure 3: Distribution of minima for N=61 atom test case.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] A. Dekkers and E. Aarts. Global optimization and simulated annealing. *Mathematical Programming*, 50:367–393, 1991.

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

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

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

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

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

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

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

[13] V.J. Torczon. On convergence of multi-directional search algorithm. *SIAM J. Optimization*, 1(1):123–145, 1991.

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