

Powerful Simulated-Annealing Algorithm Locates Global Minimum of Protein-Folding Potentials from Multiple Starting Conformations

Mark E. Snow

University of Michigan, Scientific Computation Group, 535 W. William St., Ann Arbor, Michigan 48103-4943

Received 29 July 1991; accepted 15 November 1991

Protein-folding potentials, designed with the explicit goal that the global energy minimum correspond to crystallographically observed conformations of protein molecules, may offer great promise toward calculating native protein structures. Achieving this promise, however, depends on finding an effective means of dealing with the multiple-minimum problem inherent in such potentials. In this study, a protein-folding-potential test system has been developed that exhibits the properties of general protein-folding potentials yet has a unique well-defined global energy minimum corresponding to the crystallographically determined conformation of the test molecule. A simulated-annealing algorithm is developed that locates the global minimum of this potential in four of eight test runs from random starting conformations. Exploration of the energy-conformation surface of the potential indicates that it contains the numerous local minima typical of protein-folding potentials and that the global minimum is not easily located by conventional minimization procedures. When the annealing algorithm is applied to a previously developed actual folding potential to analyze the conformation of avian pancreatic polypeptide, a new conformer is located that is lower in energy than any conformer located in previous studies using a variety of minimization techniques.

INTRODUCTION

Protein-folding potentials are potential-energy functions that have been specifically designed with the goal that the global minimum of the potential function correspond to the native conformations of protein molecules.^{1,2} They differ from commonly used molecular-mechanics potentials in that their construction and parameterization is mathematically rather than chemically justified, i.e., there is no specific attempt for the potential to reproduce local geometry of small molecules correctly or to have any particular chemical interpretation aside from the goal that the global minimum of the function correspond to observed conformations of protein molecules. While such potentials cannot be used to ask the variety of questions that molecular-mechanics potentials can, they offer numerous advantages toward the end of calculating native protein structures. That molecular-mechanics potentials give energies of the order of or even lower than the native for nonnative protein conformations³⁻⁶ makes them less than ideal for this purpose. A further advantage of protein-folding potentials is that the functional form can be chosen to be easily and efficiently computed.

One difficulty that protein-folding potentials

have in common with conventional molecular-mechanics potentials is a multiple-minimum problem. While the molecular representation and functional form of folding potentials make the multiple minimum problem less severe than in molecular-mechanics potentials, the problem is still drastic enough that it is not computationally feasible to locate all minima for a protein or to know when the global minimum has been found.

For this study, a test system has been created with a known unique global minimum. To accomplish this, a small (21 amino acid) system was chosen and an excessive number of adjustable parameters were allowed in parameterizing the potential. The energy-conformation surface of the test system was studied to verify that the system exhibits all the properties of protein-folding potentials in general. With this test system in place, it was then possible to explore the feasibility of simulated annealing to address the global optimization of protein-folding potentials.

Simulated annealing has not previously been used with protein-folding potentials, but has shown itself in general to be a very powerful approach to the global optimization of functions in many variables.⁷ The method has recently found considerable application to problems arising from

the complexity of the energy-conformation surface of molecules using conventional molecular-mechanics potentials.⁸⁻¹⁰

METHODS

The Test Problem

While developing and testing the annealing algorithm, it was desirable to choose a problem, potential, and parameterization in which there was no question that the potential has a unique global minimum corresponding to the crystallographically observed conformation. The problem should be small enough that extensive computation was feasible, yet large enough that the conformation space retain the overall structure and extensive multiple-minimum problem found in very large molecules. The 21 amino-acid fragment of bovine pancreatic trypsin inhibitor (BPTI) containing residues 16-36 meets these criteria and was chosen as a test system. The structure of BPTI has been determined crystallographically.^{11,12} Residues 16-36 form an antiparallel β -sheet that has been identified as a relatively rigid structural element by normal-mode analysis.¹³

The Potential and Parameterization

As developed previously,¹ residues are modeled as single points centered on the α -carbon. The form of the potential is

$$E = \sum_{ij} e(d_{ij}, \varepsilon, \rho)$$

where $e(d_{ij}, \varepsilon, \rho) = \varepsilon[5(\rho/d_{ij})^{12} - 6(\rho/d_{ij})^{10}]$ and d_{ij} is the distance between atoms i and j . The ε and ρ terms are broken into groups and are different for nearest neighbor interactions, for $(i, i + 2)$ interactions (where ε and ρ depend on the class of residue $i + 1$), for $(i, i + 3)$ interactions (where ε and ρ depend on the class of residues $i + 1$ and $i + 2$), and for long-range interactions (where ε and ρ depend on the class of residues i and j). In general, the residue classes are chosen so as to keep down the number of parameters while maintaining the desired properties of the potential.^{1,2} In this case, however, each residue was assigned to its own class to ensure that the potential could be parameterized with an unambiguous global minimum corresponding to the crystallographically determined structure.

With each residue belonging to its own class, and only a single structure to model, it is trivial to parameterize the potential to have a single global minimum corresponding to the observed conformation. If the ρ s are chosen to agree with the values ob-

served in the crystal structure, then any choice of the ε s where all values are nonzero and not infinite will give the desired result. For this study, all nearest neighbor interactions had an ε of 200 (arbitrary) units, $(i, i + 2)$ interactions had an ε of 50 units, $(i, i + 3)$ interactions had an ε of 20 units, and long-range interactions had an ε of 5 units.

For the generation of starting conformations and for simulated annealing, independent variables were taken to be the "bond" angle formed by three consecutive α -carbons and the dihedral angle formed by four consecutive α -carbons. At this stage, $C_\alpha-C_\alpha$ bond lengths were held fixed at 3.8 Å. The conjugate-gradients minimization, however, was performed in Cartesian space. Conversion from the angular representation to the Cartesian representation was performed by placing the first carbon at the origin, the second on the $-x$ -axis, and subsequent atoms in accordance with the bond and dihedral angles, with the third atom lying in the xy -plane.

Generation of Starting Conformations

Random starting conformations in the angular representation were generated by choosing bond angles at random between $+90$ and $+180^\circ$ and choosing dihedral angles at random between -180 and $+180^\circ$.

The Annealing Algorithm

The performance of annealing algorithms is highly dependent on appropriate choices of the independent variables, step size, initial temperature, and cooling schedule.⁷ Each of these variables was explored in the development of the algorithm used in this study. Dihedral and bond angles were found to be far more natural and efficient variables for the simulated annealing than Cartesian coordinates, although Cartesian coordinates are preferable for the final conjugate-gradients minimization at the end of an annealing run (see below). A crucial variable is the step size. Simulated-annealing implementations for computational chemistry potentials (see, e.g., refs. 8-10) traditionally choose some range of possible step sizes from which steps are chosen at random, this range of step sizes being kept fixed as the temperature is gradually lowered over the course of the annealing run. I have found it much more effective to gradually lower *both* the temperature and the step-size range over the course of the run. If this is done so as to keep the acceptance rate of trial conformations in the 50% range, the algorithm appears to sample the space much more efficiently. This approach is based on an idea derived from the work of Haines¹⁴ and has

been explored by the author on a number of test problems.¹⁵

In outline, the algorithm used here is as follows: The simulation proceeds according to the Metropolis et al. algorithm¹⁶ with a move consisting of a small step in one of the bond or dihedral angles, an evaluation of the energy of the new conformation, and a decision, based on the Boltzmann criterion, as to whether to keep the conformation. The initial step-size range and temperature are inputs to the program. An iteration consists of several sweeps through all the angles in the molecule, the number of sweeps per iteration, n , also being an adjustable parameter. If n is too small, trivial local minima are located and the procedure does not do much better than conventional minimization algorithms. With large enough n , only very good minima are located. At the bottom of the iteration loop, the acceptance rate of trial conformations is calculated and the temperature and step-size range may be adjusted so as to keep the acceptance rate in the 40–60% range. This is accomplished by multiplying the temperature by a factor a if the acceptance rate is less than 0.4 and by a factor b if the acceptance rate is greater than 0.6. a must be greater than 1.0, but should be no greater than 2.0 for the algorithm to work well. Likewise, b should be less than 1.0, but no less than 0.5. If, at the end of an iteration, the acceptance rate is in the allowed range, a check is made to determine if the best-observed energy has dropped since the last iteration. If the energy is dropping, both the temperature and step-size range remain unchanged for the next iteration. If the energy has not dropped, the temperature and step-size range are both decreased by multipliers c and d , respectively. These multipliers should be less than 1.0 but greater than 0.5, and, as a practical matter, the drop in the step-size range should not be as great as that in the temperature. The ability of the temperature to self-adjust independent of the step-size range allows correct acceptance rates to be achieved, even with poor choices of c and d , but at a cost in performance. I have found $c = 0.6$ and $d = 0.75$ to be good numbers. Iterations continue until either an iteration limit is reached, a lower limit on the step size is reached, or a lower limit on the temperature is reached. Note that if an inappropriate initial temperature is chosen it will quickly half or double during the initial iterations of the program until the acceptance rate is brought into the 50% range. Thus, reasonable results will still be obtained, but at a slightly increased cost in CPU time. After a few runs with a given potential, it becomes apparent what an appropriate initial temperature is, and this parameter can be reset.

At some point during an annealing run, the temperature and step size become small enough that

the procedure is within the radius of convergence of a given local minimum. At this point, simulated annealing is no longer a cost-effective strategy and the program switches to a conjugate-gradients minimizer¹⁸ in Cartesian coordinate space. The use of Cartesian coordinates at this point in the procedure makes it possible to achieve the true minimum. This cannot be achieved in an angle-based scheme where bond lengths are held fixed.

RESULTS

The Energy-Conformation Space

If a minimization algorithm designed with this test system is to be useful in general, it is essential that the test system exhibit the characteristics that have been previously observed for standard folding potential with large molecules. Such systems exhibit numerous local minima, and have the property that low-energy local minima resemble the global minimum except for rotation about a few bonds such that near-optimal distances are maintained for most atom pairs, i.e., local structure tends to be correct, but regions may be rotated with respect to each other.^{1,17}

To explore the characteristics of the energy-conformation space of this test system, 50 random starting conformations were generated as described in Methods. Twenty-five of these were energy minimized using conjugate-gradient minimization¹⁸ and 25 were subjected to a simulated annealing run in Cartesian coordinate space, followed by conjugate-gradients minimization. The 50 minimized structures were analyzed using a previously described cluster analysis method.⁶ In this analysis, a distance matrix is constructed for each of the structures; then, a new comparative matrix is computed by comparing (pairwise) all of the 50 individual distance matrices. This analysis revealed that the 50 starting conformations refined to 37 distinct local minima. The global minimum was located and was reached from 4 of the 50 starting conformations. The energy of the local minima ranged from -5650.94 to -6073.52 units (the global minimum). The runs in which Cartesian space simulated annealing was used did no better than those in which simple minimization was used. A very gradual annealing schedule run long enough would have probably done better, but this was not pursued since annealing in angle space was producing superior results in much shorter times (see below). It is interesting to note that the local minima fall into two distinct populations. Thirty-seven of the minimized structures, representing 29 local minima, have minimized energies between -5650.94 and -5694.26 . The other 13 minimized structures, representing 8 local minima,

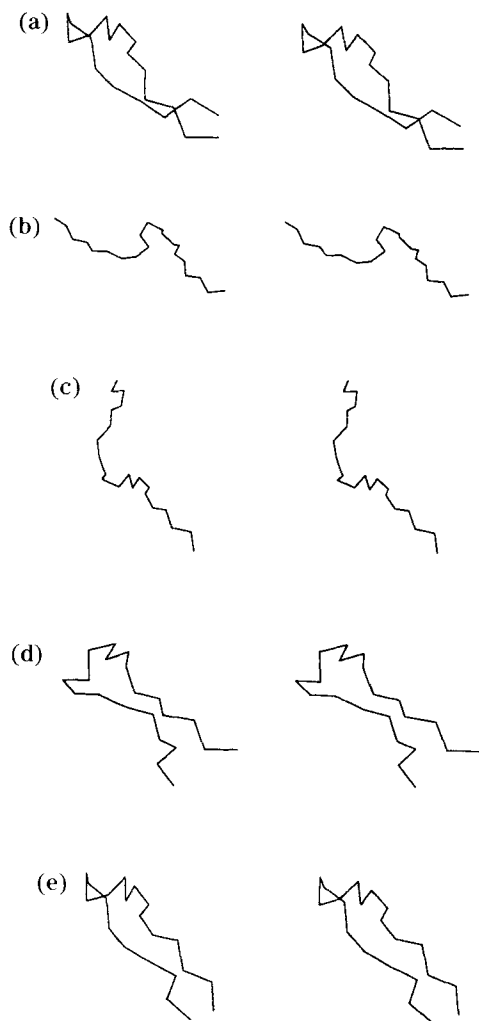


Figure 1. Stereo drawings of some of the local minima located during the exploration of the energy-conformation space. (a) Global energy minimum (energy = -6073.52). This is the conformation observed in the crystal structure. (b,c) have energies of -5683.55 and -5677.01 , respectively. These minima are in the high-energy subpopulation of local minima and do not form a β -sheet. (d,e) have energies of -5913.71 and -5999.47 , respectively. These minima are in the low-energy subpopulation of local minima. They do form a β -sheet, but do not have a correct turn. Conformation (d) is the worst (highest energy) structure in the low-energy subpopulation. It also has the poorest turn.

have energies between -5913.71 and -6073.52 (the global minimum). There were no structures found with energies between -5694.26 and -5913.71 . Further, when the conformations are examined visually it is apparent that the structures in the second population have formed a correct β -structure, but may show a slight difference from the global minimum, primarily in the β -turn. The structures in the first population, however, do not have the two strands of the sheet folded correctly against each other. The global and a number of the local minima are shown in Figure 1. These results indicate that the energy-conformation space of the

test problem does indeed have the properties of general energy-conformation spaces of protein-folding potentials for large molecules.

Simulated Annealing

An angle-space simulated-annealing run was attempted on this system with an initial step size of 15° for both bond and dihedral angles, an initial "temperature" of 1.6, and 50 sweeps per iteration. The annealing run reached the iteration limit (50), then switched to conjugate-gradients minimization. This run converged to the global minimum and took 604 s of CPU time on a SunSparcstation IPC.

To ascertain whether finding the global minimum in a single simulated annealing run was a fluke, several more runs were attempted from different random starting conformations. In a total of eight runs, the global minimum was located four times. The other four runs found very low energy minima with energies ranging from -5970.52 to -6030.02 . The five minima located by this procedure all fall in the second population of local minima and all exhibit good β -structure (Figs. 2 and 3).

Results with an Actual Folding Potential

The unique well-defined global minimum of the test problem was essential for the development and evaluation of the annealing algorithm. It is important, however, to also test the algorithm on an actual folding potential. The folding potential developed by Crippen and Snow¹ has the same functional form as the test function used in this study except that the $(i, i + 1)$ interactions have the form $E = \epsilon(d_{i,i+1}^2 - \rho^2)^2$,² where d and ρ are the observed and optimal interresidue distance. This potential was parameterized by solving a nonlinear program that attempts to force previously located minima that are not near the native conformation to have higher energies than the near-native minimum.¹ The energy-conformation space of the potential has been studied extensively^{1,19} for avian pancreatic polypeptide.²⁰ Since the global minimum of this system is not known, achieving the global minimum cannot be used as a measure of the annealing algorithm in this case. It is possible, however, to compare results achieved with the annealing algorithm to results using several other minimization algorithms that have been implemented with the potential.

Twenty-five annealing runs were attempted with $n = 500$, $a = 1.6$, $b = 0.75$, $c = 0.6$, $d = 0.75$, and an initial temperature of 30. Trial runs with smaller n did not produce good results, indicating that this is indeed a more complex potential than the test function. The lowest energy conformation

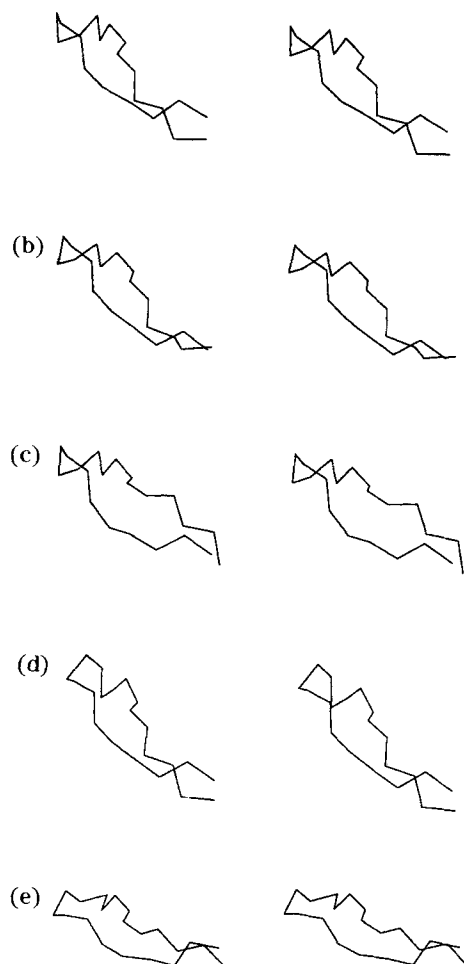


Figure 2. Stereo drawings of the five conformations located by simulated annealing. (a) Global energy minimum (energy = -6073.52). This conformation was reached from four of the eight random starting points. (b,c,d,e) have energies of -6030.02 , -5996.22 , -5984.78 , and -5970.52 , respectively. All of these conformations lie in the low-energy subpopulation of local minima. All form good β -sheets and closely resemble the native conformation.

located had an energy of -400.01 units at a rms distance from the native conformation of 5.49 \AA . Five more runs were attempted with n increased to 3000 and other parameters unchanged. The lowest energy conformation located had an energy of -407.33 units at an rms of 5.35 \AA vs. the native structure. This conformation has lower energy than any conformer previously located for this potential. For comparison, in the original paper,¹ of 737 conformations generated using the embed algorithm,²¹ then conventionally minimized, the lowest energy conformation located had an energy of -406.9 units at an rms from the native of 1.84 \AA . Further, various attempts at energy embedding from higher dimensional spaces²² never achieved a minimum lower than -399.6 units at an rms of 7.52 \AA .¹ In a separate study,¹⁹ the lowest energy

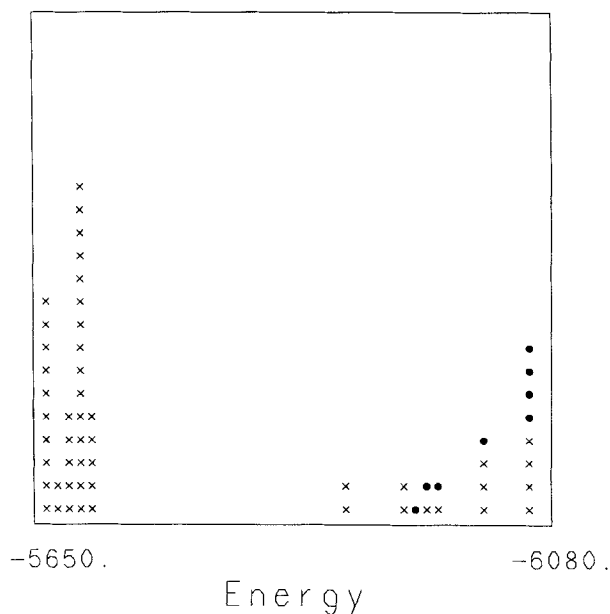


Figure 3. Distribution of local minima. The 50 minima located during the exploration of the energy-conformation space are marked X and the 8 minima located by angle-space simulated annealing are marked by filled octagons. The plot is divided into bins having a width of 10 energy units each. The cluster of structures represented by the Xs on the left side of the plot do not have the β -strands folded correctly. Note that the eight simulated-annealing structures all lie on the right side of the plot with low energies.

conformer located by rotational energy embedding had an energy -402.5 at an rms of 2.05 \AA .

DISCUSSION

The test system described here, by having a unique well-defined global minimum corresponding to the crystallographically determined conformation, has provided an excellent opportunity to develop and explore simulated annealing algorithms and their applicability to protein-folding potentials.

That the annealing algorithm located the global energy minimum from four of eight random starting conformations is extremely encouraging. Perhaps most encouraging is the fact that, with an actual folding potential, the algorithm located a lower energy conformation for avian pancreatic polypeptide than had previously been found using hundreds of minimizations from well-chosen starting points, using energy embedding, or using rotational energy embedding.

In the future, annealing algorithms should enhance the feasibility of finding global and near-global minima of folding potentials and may make it possible to extend the use of folding potentials to larger protein molecules.

The author is grateful to Dr. Gordon Crippen for valuable discussions regarding this work.

References

1. G.M. Crippen and M.E. Snow, *Biopolymers*, **29**, 1479 (1990).
2. P. Seetharamulu and G.M. Crippen, *J. Math. Chem.*, **6**, 91 (1991).
3. J. Novotny, R. Bruccoleri, and M. Karplus, *J. Mol. Biol.*, **177**, 787 (1984).
4. S. Bryant and L.M. Amzel, *Int. J. Peptide Prot. Res.*, **29**, 46 (1987).
5. J. Novotny, A.A. Rashin, and R.E. Bruccoleri, *Prot. Struct., Funct., Genet.*, **4**, 19 (1988).
6. M.E. Snow and G.M. Crippen, *Int. J. Peptide Prot. Res.*, **38**, 161 (1991).
7. R.H.J.M. Otten and L.P.P.P. van Ginneken, *The Annealing Algorithm*, Kluwer Academic Press, Boston, 1989.
8. A. Nayeem, J. Vila, and H.A. Scheraga, *J. Comp. Chem.*, **12**, 594 (1991).
9. S.R. Wilson, W. Cui, J.W. Moskowitz, and K.E. Schmidt, *J. Comp. Chem.*, **12**, 342 (1991).
10. L.B. Morales, R. Garduno-Juarez, and D. Romero, *J. Biomol. Struct. Dynamics*, **8**, 721 (1991).
11. A. Wlodawer, J. Deisenhoffer, and R. Huber, *J. Mol. Biol.*, **193**, 145 (1987).
12. J. Deisenhoffer and W. Steigemann, *Acta Crystallogr. B*, **31**, 238 (1975).
13. T. Nishikawa and N. Go, *Prot. Struct., Funct., Genet.*, **2**, 308 (1987).
14. L.M. Haines, *Technometrics*, **29**, 439 (1987).
15. M.E. Snow and S.B. Crary, *The Michigan Academician*, in press.
16. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
17. M.E. Snow and G.M. Crippen, unpublished results.
18. R. Fletcher and C.M. Reeves, *Computer J.*, **7**, 149 (1964).
19. G.M. Crippen and T.F. Havel, *J. Chem. Inf. Comp. Sci.*, **30**, 222 (1990).
20. I. Glover, I. Haneef, J. Pitts, S. Wood, D. Moss, I. Tickle, and T. Blundell, *Biopolymers*, **22**, 293 (1983).
21. G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, Wiley, New York, 1988.
22. G.M. Crippen and P.K. Ponnuswamy, *J. Comp. Chem.*, **8**, 972 (1987).