Finding the minimum energy conformation of protein-like heteropolymers by Greedy Neighborhood Search

Joon Suk Huh

INTRODUCTION

The prediction of a protein’s tertiary structure, given its amino acid sequence, is one of the intriguing problems in biophysics and computational biology. Anfinsen’s thermodynamic hypothesis indicates that the tertiary structure of a protein is the minimum free-energy conformation. Based on Anfinsen’s hypothesis, the protein structure prediction problem can be treated as a global optimization problem of its given energy function. Unfortunately, there is no analytical solution for this problem because of the complexity of the energy function.

Many global optimization methods such as Simulated Annealing (SA), Genetic Algorithm (GA), and Conformational Space Annealing (CSA) have been proposed to deal with such hard optimization problems and successfully applied to various fields of science and engineering problems. A greedy algorithm is any algorithm that always takes the locally optimal solution while finding the global minimum of a function. Compared with above mentioned methods, it often fails to find the global minimum being trapped in a deep local minimum. A newly proposed global optimization method called Greedy Neighborhood Search (GNS) is belonged to the greedy algorithm class. GNS randomly generates the new solution adjacent to the current solution and then replaces the current one with the new one if the new one is better. Also GNS locates all the solutions on the minimum where the first derivative is zero and the second derivative is greater than zero by means of a local minimization algorithm. Because GNS randomly picks candidates on a minimum at each stage, it could escape from a deep local minimum unlike conventional greedy algorithms.

In this paper, the GNS method with a novel conformational sampling method using a spherical distribution called von Mises-Fisher distribution is applied to find the minimum energy conformation of a protein-like heteropolymer model called AB model. The AB model consists of only two types of monomers, Hydrophobic (A) and Hydrophilic (B) monomers analogous to the real proteins. Despite its simplicity, finding the minimum energy conformation of the AB model is very difficult and NP-complete problem. Unlike the traditional Hydrophobic (H), Polar (P) lattice model, only a few works were devoted to search the minimum energy conformations of the AB model polymers. Because there is no rigorous proof to discriminate the true global minimum from many metastable local minima, the question is still open.

MODELS AND METHODS

The AB model

The AB model polymers are modeled as an off-lattice bead-rod chain system in three-dimensional space shown in Fig. 1. The length of a rod connecting two consecutive beads is fixed to 1. The potential energy of an AB model polymer with N monomers is given by

\[ E = \sum_{i=1}^{N-2} E_i(\theta_i) + \sum_{i=1}^{N-2} \sum_{j=i+1}^{N} E_2(r_{ij}, \zeta_i, \zeta_j) \]  (1)

where \( E_i \) is the bond angle energy and \( E_2 \) is the non bonded van der Waals energy. The bond angle energy is given by

\[ E_i(\theta_i) = (1 - \cos \theta_i) / 4 \]  (2)

where \( \theta_i \) is a bond angle defined by three consecutive monomers, \( i-1, i, \) and \( i+1 \). The van der Waals energy is given by

\[ E_2(r_{ij}, \zeta_i, \zeta_j) = 4\left[ r_{ij}^{-12} - C(\zeta_i, \zeta_j) r_{ij}^{-6} \right] \]  (3)

where \( r_{ij} \) is the euclidean distance between \( i \) th monomer and \( j \) th monomer. Each \( \zeta_i \) is either A or B, and \( C(\zeta_i, \zeta_j) \) is +1, +0.5 and -0.5 respectively, for AA, BB, AB pairs, producing strong attraction between AA pairs, weak attraction between BB pairs, and weak repulsion between AB pairs.

A global optimization method

The GNS method is performed on a search space consisting of local minima shown in Fig. 2. GNS consists of following five stages.

1. Random initialization – pick a random conformation from feasible space.
Fig. 1 The bead-rod chain representation of the AB model. Each bead represents a hydrophobic (A) or a hydrophilic (B) monomer.

Fig. 2 A search space consisting of local minima. All sampled conformations are relaxed by local minimization procedure.

2. Candidate generation – generate a trial conformation by a random modification on a current conformation.

3. Energy minimization – using a local minimization algorithm, relax the trial conformation.

4. Replacement – Replace the current conformation with the trial conformation if the trial conformation’s energy is lower than the current one’s.

5. Repeat from Stage 2 to Stage 4 until termination criterion (e.g. maximum iterations) is satisfied.

In this paper, L-BFGS subroutine \(^{12}\) is used to relax the generated conformations.

Fig. 3 A chain system represented by edge vectors.

Fig. 4 Points from the von Mises-Fisher distribution. The concentration parameters are 10 and 100 from left to right. The mean direction is \((0,0,1)^T\).

A conformational sampling method

The bead-rod chain system with \(N\) beads can be represented by \(N-1\) edge vectors connecting two consecutive monomers as shown in Fig. 3.

The conformational sampling method used here generates a trial conformation by sampling these vectors from a spherical distribution called von Mises-Fisher distribution. The von Mises-Fisher distribution is a probability distribution of points or unit vectors on a \(p\)-dimensional hypersphere \(^{13}\). When \(p\) is 2 (on the sphere), the probability density function is given by

\[
p_3(x, \mu, \kappa) = c_3(\kappa) \cdot e^{\kappa \mu^T x},
\]

where

\[
c_3(\kappa) = \frac{\kappa}{4 \pi \sinh \kappa},
\]

where \(\mu\) is the mean direction and \(\kappa\) is the concentration factor. The greater the value of \(\kappa\) is, the higher the concentration of the distribution around the mean direction is, as shown in Fig. 4.

In this paper, the method proposed by wood \(^{14}\) is used to simulate the von Mises-Fisher distribution. A random unit vector \(X\) sampled from the distribution with the mean direction \((0,0,1)^T\) and the concentration parameter \(\kappa\) is given by

\[
X = \sqrt{1-W^2} \cdot V + W \cdot \hat{z},
\]

where

\[
W = \frac{1}{\kappa} \cdot \ln(e^{-x} + \kappa \cdot C_3^{-1}(\kappa) \cdot y_1),
\]

\[
V = \cos \theta \cdot \hat{x} + \sin \theta \cdot \hat{y},
\]

where \(y_1\) and \(y_2\) are random variables from the uniform distribution over the range \([0,1]\). A random unit vector with an arbitrary mean direction is obtained from \(X\) by using following transfer matrices.
where $\theta$ is a polar angle of the mean vector and $\phi$ is an azimuth angle measured from y-axis.

In this paper, trial conformation’s edge vectors are generated from the von Mises-Fisher distribution whose mean directions are current conformation’s edge vectors and the concentration parameter is 10.

**RESULTS AND DISCUSSIONS**

The program is written in FORTRAN90 language. Experiments were conducted for four different sequences whose lengths are 13, 21, 34 and 55 respectively. The full sequences are given as

$S_{13} = ABBABBABAB BAB$,

$S_{21} = BABABBA \circ ABBABBABAB BAB$,

$S_{34} = ABBABBBAB BAB \circ BABABBA \circ BABABBABA BBAB$,

and

$S_{55} = BABABBAB \circ ABBABBABAB BAB \circ ABBABBBAB BAB \circ ABBABBABAB BAB \circ ABBABBABAB BAB$.

Table 1 lists the results obtained by GNS in comparison with those by other methods.

![Table 1](https://example.com/table1.png)

Table 1 mean energies (ME), mean CPU times in second (MT), and the number of times the minimum energy is found out of ten runs (NS).

![Table 2](https://example.com/table2.png)

Table 2 The minimum energies obtained by Greedy Neighborhood Search (GNS) in comparison with those by nPERMis, ELP, STA, and CSA, respectively.
Fig. 5 The minimum energy conformations (right), along with unfolded conformations (left) for (a) S_{13}, (b) S_{21}, (c) S_{34}, and (d) S_{55}.

REFERENCES

1. C. Anfinsen, 1973. Principles that govern the folding of protein chains. *Science*. **81**: 223-230.
2. S. Kirkpatrick, C. D. Gellatt Jr., and M. P. Vecchi, 1983. Optimization by Simulated Annealing. *Science*. **330**: 671-680.
3. N.A. Baricelli, 1957. Symbiogenetic evolution processes realized by artificial methods. *Methods*. **9**: 143-182.
4. J. Lee, H. A. Scheraga, and S. Rackovsky, 1997. New Optimization Method for Conformational Energy Calculations on Polypeptides: Conformational Space Annealing. *J. Comput. Phys.* **18**: 1222.
5. F.H. Stillinger, 1995. Collective aspects of protein folding illustrated by a toy model. *Phys. Rev. E*. **52**: 2872-2877.
6. K.A. Dill, 1985. Theory for the folding and stability of globular proteins. *Biochemistry*. **24**: 1501-1509.
7. H.P. Hsu, V. Mehra, P. Grassberger, 2003. Structure optimization in an off-lattice protein model. *Phys. Rev. E*. **68**: 037703.
8. M. Bachmann, H. Arkin, W. Janke, 2005. Multicanonical study of coarse-grained off-lattice models for folding heteropolymers. *Phys. Rev. E*. **71**: 031906.
9. J. Kim, J.E. Straub, T. Keyes, 2007. Structure optimization and folding mechanisms of off-lattice protein models using statistical temperature molecular dynamics simulation: Statistical temperature annealing. *Phys. Rev. E*. **76**: 011913.
10. S.Y. Kim, S.B. Lee, J. Lee, 2005. Structure optimization by conformational space annealing in an off-lattice protein model. *Phys. Rev. E*. **72**: 011916.
11. J.W. Lee, K.Y. Joo, S.Y. Kim, J. Lee, 2008. Re-examination of Structure Optimization of Off-Lattice Protein AB Models by Conformational Space Annealing. *J. Comput. Chem.* **29**: 2479-2484.
12. J. Nocedalm, 1980. Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*. **35**: 773-782.
13. R.A. Fisher, 1953. Dispersion on a sphere. *Proc. Roy. Soc. London Ser. A*. **217**: 295-305
14. A.T.A. Wood, 1994. Simulation of the von Mises Fisher distribution. *Comm. Statist. Simulation Comput.* **23**: 157-164