Energy Minimization of Protein Tertiary Structure by Parallel Simulated Annealing using Genetic Crossover

Takeshi Yoshida
Graduate School of Engineering
Doshisha University

Tomoyuki Hiroyasu
Department of Engineering
Doshisha University

Mitsunori Miki
Department of Engineering
Doshisha University
Kyotanabe Kyoto, 610-0321, Japan

Maki Ogura
Graduate School of Engineering
Doshisha University

Yuko Okamoto
Department of Theoretical Studies
Institute for Molecular Science
Okazaki National Research Institutes
Okazaki, Aichi 444-8585, Japan

Abstract

In this paper, Parallel Simulated Annealing using Genetic Crossover (PSA/GAc) is applied to predict protein tertiary structures. The target protein is C-peptide that is consisted of 13 amino acids. The results are compared to those of the former studies. Then it is found that PSA/GAc is an effective method to predict protein tertiary structures.

1 Introduction

Protein is an essential factor of creatures. It is said that the function of the protein is decided by its tertiary structure. At the same time, many medicines are efficacious when the structures of the medicine and the protein are combined tightly. Therefore, it is very important to grasp protein tertiary structure.

Recently, the predictions of protein tertiary structure by computer simulations have been focused because of the easiness and cost of the predictions. One of the methods of the predictions of protein tertiary structure is formulated as a minimization problem on an energy function of protein structures. By this method, the protein structure is determined from only its amino-acid sequence. However, it is very difficult to find an optimum since there are a lot of local minima.

In this study, Parallel Simulated Annealing using Genetic Crossover (PSA/GAc) [Hiroyasu 2000] is applied to predict protein tertiary structures. PSA/GAc is developed by the authors and has a high local and global searching ability. The proteins used in this study is C-peptide that is consisted of 13 amino acids. The results are compared to those of the former studies. Then it is found that PSA/GAc is an effective method to predict protein tertiary structures.

2 Parallel Simulated Annealing using Genetic Crossover

We have developed Parallel Simulated Annealing using Genetic Crossover (PSA/GAc) and PSA/GAc is applied to predict protein tertiary structure in the next chapter.

PSA/GAc is one of parallel SAs and it is also one of hybrid methods. In this section, PSA/GAc is explained.

The concept of PSA/GAc is shown in Figure 1.

In PSA/GAc, there are plural processes and sequential SA is operated in each process. After some steps, the crossover that is usually used in GAs is applied to exchange the information between the solutions. We call this operation genetic crossover. We call a searching point an "individual", the total number of SA searching points the "population size" and annealing steps "number of generations". In optimization problems, the value of the objective function is made smaller or bigger. The objective function is consisted of the fitness in GAs and the energy in SAs. In the genetic crossover, we randomly select two individuals as parents and generate two children by genetic crossover. An individual consists of design variables and the design variables are real numbers. Therefore, the crossover is only performed between the variables. In the operation of genetic crossover, selection is also performed. After the crossover, there are two parents and two children. Then the two individuals that have higher evaluation values are selected. These two individuals become the next searching points. While SA requires high computational costs, particularly in continuous problems, this operation reduces the computational cost. SAs in continuous problems need definitions of the neighborhood and terminal condition. In this paper, Metropolis criterion is used as the acceptance criterion and the exponential annealing $T_{k+1} = 0.93\dot{T}_k$, where k is the generation, is used

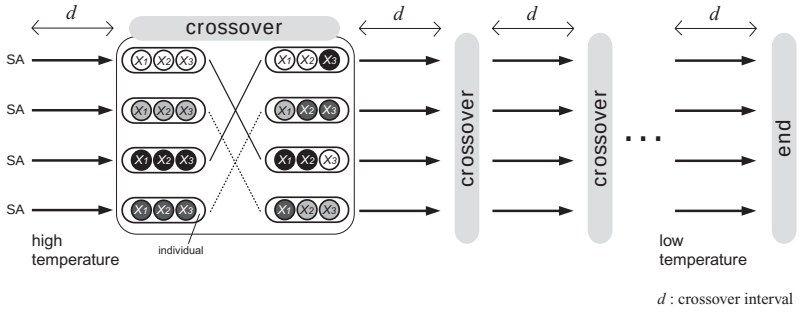

Figure 1: Model of Parallel Simulated Annealing using Genetic Crossover (PSA/GAc)

as the cooling schedule.

3 Energy Minimization of Protein Tertiary Structure

In this section, PSA/GAc is applied to derive protein tertiary structures. Usually, simulated annealing (SA) is applied to find tertiary structures by minimizing the energy of the structure[Okamoto 1991, Okamoto 1993].

3.1 Target Protein Structures

In this study, two types of the protein structures are predicted; thoses are C-peptide that is consisted of 13 amino acids and the fragment of Personal Thyroid Hormone that is consisted of PTH(1-34) amino acids. The sequence of the amino acid of C-peptide is Ly+, Gl-, Thr, Ala, Ala,Ala, Ly+, Phe, Glu, Ar+, Gln, Hi+, Met. Okamoto et al. found the structure whose energy is $-42kcal/mol$ and who has an α-helix which is consisted of 7 amino acids (Helixness$n = 7$) [Okamoto 1991].

3.2 Prediction of Protein Tertialy Structuer by PSA/GAc

In this extended abstract, the result of C-peptide is shown. We have already obtained the result of PTH(1-34). The temperature scheduling and the neighborhood definition are the same as the references [Okamoto 1991, Okamoto 1993]. The used parameter in this simulation is summarized in Table 1.

The results derived by PSA/GAc are compared of those of references [Okamoto 1991, Okamoto 1993].

3.2.1 Results of C-peptide

In this paper, the parameters summarized in Table 1 are used for PSA/GAc to determine the structure of C-peptide. We performed two types of simulations. One of them is a simulation where PSA/GAc are run for10 trials with 4164 MCsweeps and 24 individuals (100,000 MCsweeps totally). The other is a simulation where PSA/GAc are run for 7 trials with 41646 MCsweeps and 24 individuals. In this problem, there are 64 dihedral angles and these are the design variables. In a MCsweep, these 64 design variables are moved one by one. The initial structure is determined randomly.

Table 1: Parameters of PSA/GAc

Parameter	Value
Population size	24
Initial temperature	2.0 (1000K)
Last temperature	0.10 (50K)
Crossover interval	32
Range size	$180° \rightarrow (180 \times 0.3)°$
Num of Processors	6

The energy of the structure that is derived by the simulation of 4164 MCsweeps and 24 individuals is $-46.7kcal/mol$. The dihedral angles of the derived structure is shown in 2.

The derived structure has a α-helix which is consisted of 6 amino acids. The energy of the derived structure is smaller than that of Okamoto et al. and the length of α-helix is shorter.

The result of the simulation of 41646 MCsweeps and 24 indiviulas are shown in Figure 3 and 2.

The energy of the derived structure is $-57.8kcal/mol$. α-helix is constructed, when the values of the dihedral angles (ϕ_i, ψ_i) are in (-60 ± 45, $-50\pm45°$). Therefore, it has a α-helix which consists of 9 amino acids. The energy of the derived structure is much smaller than that of Okamoto et al. Therefore, it can be concluded that PSA/GAc has high searching ability in predicting

Table 2: Main-chain structure of the lowest-energy conformation of C-peptide gained by PSA/GAc (4164MCsweeps)

Residue		ϕ	ψ	ω
1		74	161	180
2		55	44	180
3	*	-72	-43	180
4	*	-76	-39	180
5		-72	112	180
6		-66	94	180
7		-128	167	180
8	*	-71	-22	180
9	*	-67	-38	180
10	*	-61	-41	180
11	*	-74	-44	180
12		-149	108	180
13		-66	158	180

Table 3: Main-chain structure of the lowest-energy conformation of C-peptide gained by PSA/GAc (41646MCsweeps)

Residue		ϕ	ψ	ω
1		18	-55	180
2		-161	162	180
3		-79	76	180
4	*	-72	-27	180
5	*	-68	-44	180
6	*	-63	-40	180
7	*	-65	-47	180
8	*	-63	-37	180
9	*	-68	-42	180
10	*	-64	-40	180
11	*	-71	-35	180
12	*	-65	-43	180
13		-78	96	180

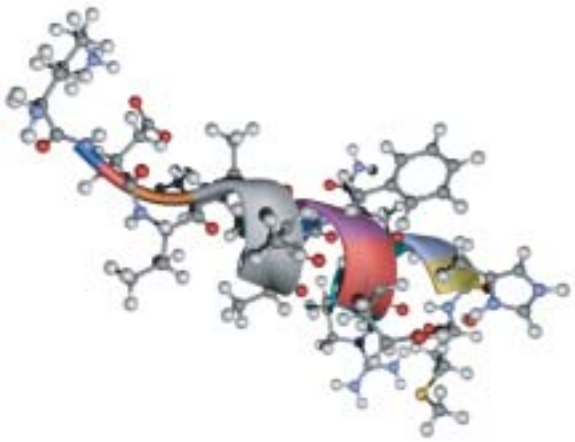

Figure 2: Lowest-energy conformation of C-peptide gained by PSA/GAc (41646MCsweeps)

protein tertiary structure problems.

4 Conclusions

In this study, Parallel Simulated Annealing using Genetic Cross over is applied for predicting protein tertiary structures of C-peptide and PTH. C-peptide is consisted of 13 amino acids and PTH is consisted of 34 amino acids.

From the simulations, the structures who have the smaller energy of the structures derived in the former studies are obtained. Therefore, it is concluded that PSA/GAc is an effective method for predicting protein tertiary structures.

References

[Hiroyasu 2000] T. Hiroyasu and M. Miki and M. Ogura(2000). Parallel Simulated Annealing using Genetic Crossover. Proceedings of the IASTED International Conference PARALLEL AND DISTRIBUTED COMPUTING AND SYSTEMS.139-144

[Okamoto 1991] Yuko Okamoto and Masataka Fukugita and Takashi Nakazawa and Hikaru Kawai (1991). α-Helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A. Protein Engineering.Vol.6 . No.4 .639-647

[Okamoto 1993] Yuko Okamoto and Takeshi Kikuchi and Takashi Nakazawa and Hikaru Kawai(1993). α-Helix structure of parathyroid hormone fragment (1-34) predicted by Monte Carlo simulated annealing. INTERNATIONAL JOURNAL OF PEPTIDE & PROTEIN RESEARCH.Vol.42 .300-303

A Source

Takeshi Yoshida, Tomoyuki Hiroyasu, Mitsunori Miki, Maki Ogura, Yuko Okamoto ,Energy Minimization of Protein Tertiary Structure by Parallel simulated Annealing using Genetic Crossover ,*Genetic and Evolutionary Computation Conference (GECCO-2002)* ,New York City, New York, USA ,GECCO-2002 Workshop Program pp.49-52.