

SA/AAN を用いたタンパク質立体構造のエネルギー最小化

Simulated Annealing with Advanced Adaptive Neighborhood for Minimizing the Energy of Protein Tertiary Structures

江上 透

Toru EGAMI

Abstract: Okamoto found that Simulated Annealing (SA) shows an effective method for minimizing the energy of Protein Tertiary Structures. However, detailed examination is not investigated neighborhood range in SA. Well, Simulated Annealing with Advanced Adaptive Neighborhood (SA/AAN) is a SA with an adaptive neighborhood range for maintaining an optimum acceptance ratio, and it shows good performance for typical test problems. This paper deals with minimizing the energy of Protein Tertiary Structures with SA/AAN. This approach, SA/AAN, is more effective than SA in this problem.

1 はじめに

タンパク質の立体構造予測は、近年注目されている問題の一つである。なぜなら、タンパク質は生命現象に直接関わる重要なものであり、構造を解明することは生命現象の仕組みを解明することにもつながるからである。自然に存在するタンパク質の立体構造は、そのタンパク質が持つエネルギーの最小状態に対応している。そのため、アミノ酸配列情報から、エネルギーを最小とするような立体構造を最適化手法を用いて求めることで、立体構造の解析が可能である。

これまで、岡本らは小規模なタンパク質 (Met-enkephalin) の構造解析において、シミュレーテッドアニーリング¹⁾ (Simulated Annealing:SA) の有効性を確かめている²⁾。SA を適用する場合に重要になることは、温度パラメータと近傍の設定方法である。タンパク質の立体構造予測では、構造を決定する二面角と呼ばれる角度を変化させ最適化を行う。よって、連続値である角度の近傍の設計が重要となる。しかしながら、この近傍幅の調節に関しては、詳細な検討は行われていない。

一方、三木らは、最適な受理確率を目標とする適応的近傍を持つシミュレーテッドアニーリング (Simulated Annealing with Advanced Adaptive Neighborhood:SA/AAN) を提案し、代表的な数学関数最小化問題において有効な性能を示した³⁾。

そこで、本研究では、SA/AAN を用いてタンパク質立体構造のエネルギー最小化を行い、その有効性を検討する。

2 Corana の手法 (SA with Adaptive Neighborhood:SA/AN)

2.1 SA/AN のアルゴリズム

SA を用いて探索を行う際、Corana は近傍幅が大き過ぎることにより受理率が低くなり過ぎることや、近傍幅が小さ過ぎることにより受理率が高くなり過ぎることにより、無駄な探索が生じることを指摘した。Corana が提案した SA (SA/AN)⁴⁾ は無駄な探索を生じさせないよう解摂動に用いる近傍幅を受理率が 0.5 になるように近傍を調節するアルゴリズムである。

解摂動は式 (1) で表される一様分布の近傍を考え、現在の各設計変数 x_i から、次状態の各設計変数 x'_i を生成する。

$$x'_i = x_i + rm \quad (1)$$

ここで、 r は $[-1, 1]$ の一様乱数である。また、 m は近傍幅を決定するパラメータである。このパラメータ m をある解の推移毎に変化させることで、適応的な解摂動が行われることを期待する。近傍幅を調節する処理では、式 (1) のパラメータ m を式 (2) のように受理率 p によって変化する関数 $g(p)$ を用いて決定する。また、この $g(p)$ は式 (3) を用いて決定する。

$$m' = m * g(p) \quad (2)$$

$$\left\{ \begin{array}{ll} g(p) = 1 + c \frac{p-0.6}{0.4}, & \text{if } p > p_1 \\ g(p) = \left(1 + c \frac{0.4-p}{0.4}\right)^{-1}, & \text{if } p < p_2 \\ g(p) = 1, & \text{otherwise} \end{array} \right. \quad (3)$$

$$p_1 = 0.6, p_2 = 0.4$$

関数 $g(p)$ をグラフ化したものを Fig. 1 に示す．横軸は受理率 p ，縦軸は近傍幅の拡大率 $g(p)$ を表す．

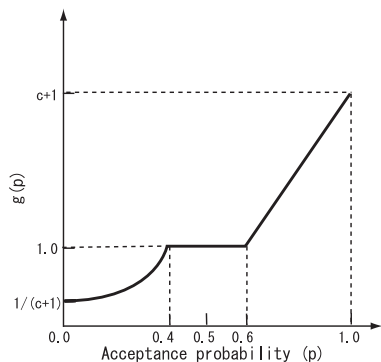


Fig. 1 Corana のアルゴリズム

ここで p は，近傍幅を変更する間隔 N の間に解摂動が受理された回数 n から， $p = n/N$ と計算される．また， c は調節の度合いを決定するパラメータである．

2.2 SA/AN の問題点

SA/AAN を用いることにより，連続関数に SA を適用した場合の近傍設計を自動化できる．しかし，このアルゴリズムでは，一度局所解に陥ると受理率が低下してしまう．この低下を補うために，式 (2) により近傍幅を小さくする．するとますますその局所解から脱出することが困難になってしまう．このため，拡大率自体を適応的にする必要があると考える．

3 最適な受理確率を目標とする適応的近傍を持つシミュレーテッドアニーリング

SA/AN の問題点を解消する手法として，三木らによって最適な受理率を目標とする適応的近傍を持つシミュレーテッドアニーリングが提案されている．SA/AAN のアルゴリズムは，式 (4) に示す階段関数を用いて受理率から近傍幅を決定する．すなわち，受理確率が目標値の上限 p_1 より大きい場合には近傍を H_0 倍し受理率を下げる．一方，目標値の下限 p_2 より小さい場合には近傍を半分に縮小し，受理率を上げる．この時，近傍幅を増加させる拡大率 H_0 を式 (5) のように再帰的に $g(p)$ と定義することにより，受理率が下がりにくい時には，拡大率が十分に大きな値になるようにした．Fig. 2 に関数 $g(p)$ を示す．

$$\begin{cases} m' = m \times g(p) \\ g(p) = H_0(p'), & \text{if } p > p_1 \\ g(p) = 0.5, & \text{if } p < p_2 \\ g(p) = 1.0, & \text{otherwise} \end{cases} \quad (4)$$

$$\begin{cases} H_0(p') = H_0(p') \times H_1, \\ \quad (\text{初期設定: } H_0 = 1.2) \\ H_1 = 1.2, & \text{if } p' > p_1 \\ H_1 = 0.5, & \text{if } p' < p_2 \\ H_1 = 1.0, & \text{otherwise} \end{cases} \quad (5)$$

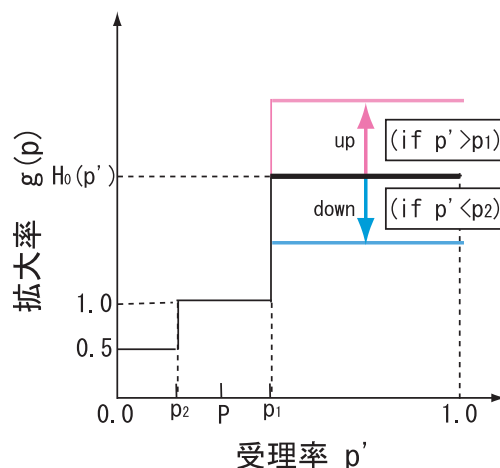


Fig. 2 近傍幅調節の拡大率

ここで p は，近傍幅を変更する間隔 N の間に解摂動が受理された回数 n から， $p = n/N$ と計算される．また，ここで p' は，近傍幅のパラメータ (H_0) を変更する間隔 L の間に解摂動が受理された回数 l から， $p' = l/L$ と計算される．

4 数値実験

本実験では，Met-enkephalin, $(\text{Ala})_{10}$ および C-peptide の 3 つのタンパク質を対象問題とし，SA/AAN を用いてタンパク質立体構造のエネルギー最小化を行った．

Met-enkephalin は Tyr-Gly-Gly-Phe-Met という 5 個のアミノ酸からなり，最小エネルギー $E \leq -11 \text{kcal/mol}$ の領域で最小エネルギー構造となる．

本実験では，岡本らの実験と同様に Met-enkephalin の主鎖における 10 個の二面角と，側鎖における 9 個の二面角をそれぞれ設計変数とし，二面角のとり得る値は $[-180^\circ, 180^\circ]$ の範囲で表現した．つまりこのタンパク質は 19 個の設計変数を持っており，1MCsweep によって 19 回の Metropolis 判定がされるものとした．

$(\text{Ala})_{10}$ は，10 個の Ala というアミノ酸からなり， $E \leq -9.7 \text{kcal/mol}$ の領域で最小エネルギー構造となる．なお，本実験では岡本らの実験と同様に主鎖における 20 個の二面角と，側鎖における 10 個の二面角を設計変数とした．

また，C-peptide は 13 個のアミノ酸残基 Ly+, Gl-, Thr, Ala, Ala, Ala, Ly+, Phe, Glu, Ar+, Gln, Hi+, Met からなり， $E \leq -42 \text{kcal/mol}$ の領域で最小エネルギー構造となる．なお，本実験では岡本らの実験と

同様に主鎖における 26 個の二面角と、側鎖における 38 個の二面角を設計変数とした。

Table 1 に各タンパク質の二面角数 (設計変数の数) および最適解領域を示す。

	Dehdral angles	Optimum region
Met-enkephalin	19	-11kcal/mol or less
(Ala) ₁₀	30	-9.7kcal/mol or less
C-peptide	64	-42kcal/mol or less

実験に用いたパラメータを Table 2 に示す。

パラメータ	Met-enkephalin	(Ala) ₁₀	C-peptide
設計変数	19	30	64
最高温度	2.0	2.0	2.0
最低温度	0.1	0.1	0.1
H_0 (初期拡大率)	1.2	1.2	1.2
H_1 (拡大率)	1.2	1.2	1.2
目標受率率	0.1	0.1	0.1
総 MCsweep 数	100000	100000	100000
H_1 の変化の間隔	5	5	5
$H_0(p)$ の変化の間隔	40	40	40
試行回数	30	30	30

Fig. 3 に、岡本らの SA と SA/AAN を用いた際の近傍の変化の履歴を示す。横軸に MCsweep 数、縦軸に近傍幅を示している。なお、値は Met-enkephalin, (Ala)₁₀ および C-peptide それぞれにおける 30 回試行中のエネルギーが最小となったものを示す。

Fig. 3 に示すように、岡本らの逐次 SA では近傍を 180 度から 54 度まで一定の割合で減少させていたが、SA/AAN を用いることによって問題に依存することなく、受率率によって適応的に近傍幅を調節していることが分かる。

また、Fig. 4 に、岡本らの SA と SA/AAN を用いた際の最適解到達率を示す。

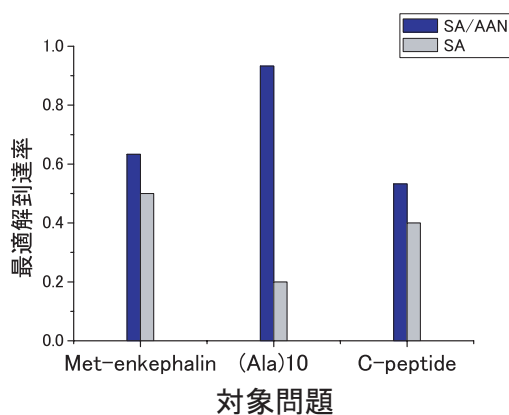


Fig. 4 最適解到達率

Fig. 4 に示すように、3 種のタンパク質すべてにおいて、SA/AAN のほうが高い性能を示した。

また、Fig. 5 に、SA と SA/AAN における Met-enkephalin, (Ala)₁₀ および C-peptide の最良値、中央値および最悪値を示す。

Fig. 5 に示すように、Met-enkephalin, (Ala)₁₀ および C-peptide いずれのタンパク質においても SA/AAN が最も良い性能を示していることが確認できる。

また、Fig. 6 に解収束に要する時間の比較を示す。これは、最適解領域に到達するまでに要する探索数で比較を行う。

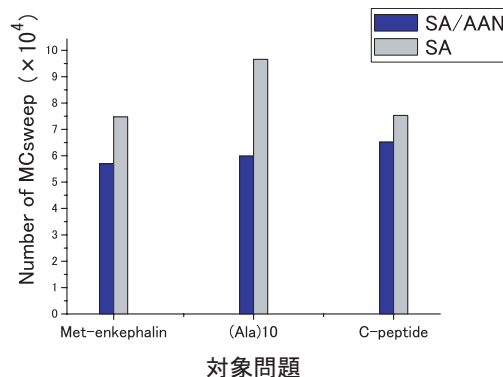


Fig. 6 解収束時間

Fig. 6 に示すように、Met-enkephalin, (Ala)₁₀ および C-peptide いずれのタンパク質においても、解収束に要する時間は SA よりも SA/AAN の方が短いことが分かる。

したがって、Met-enkephalin, (Ala)₁₀ および C-peptide とともに、SA/AAN の方が SA よりも解探索性能に優れていることが分かる。

5 SA/AAN のパラメータの検討

三木らは数学関数最小化問題では、初期拡大率 H_0 は 2 倍、拡大率 H_1 も 2 倍としていた。しかし、タンパク質の近傍は 180 度までであり、拡大率が 2 倍の値を用いてしまうと、近傍が探索途中で一度小さくなったとしてもすぐに大きな値に戻ってしまう。

拡大率を 2 倍にした際の近傍の履歴を、Fig. 7 に示す。

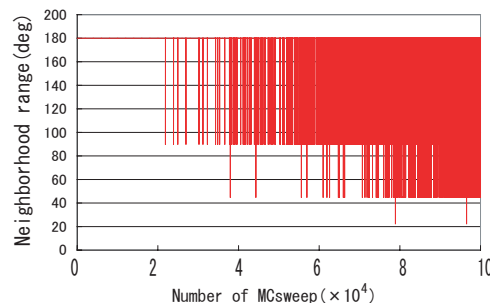


Fig. 7 近傍幅の履歴 (拡大率 2 倍)

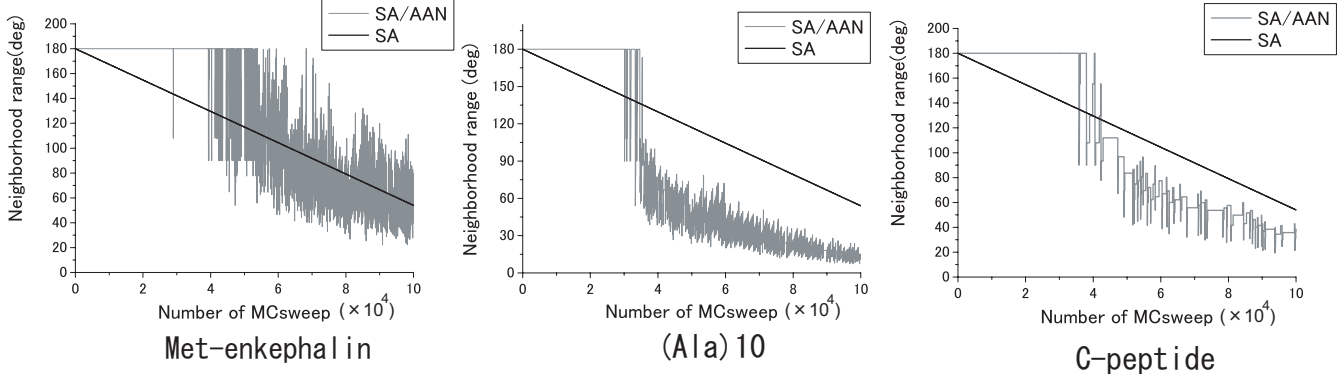


Fig. 3 近傍幅の推移

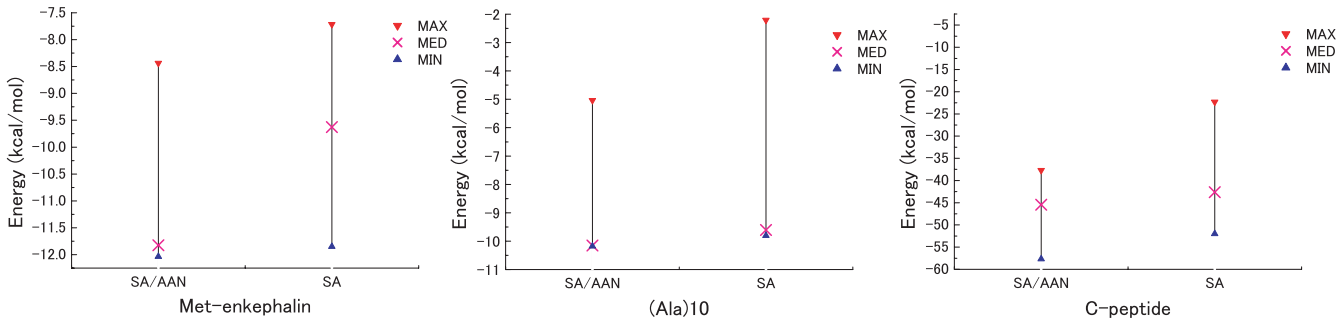


Fig. 5 解探索性能の比較

Fig. 7 に示すように，拡大率が 2 倍の値を用いた場合，探索後半に近傍幅が 180 °となり，有効な探索を行うことができていない。

したがって，今回の実験では初期拡大率 H_0 を 1.2，拡大率 H_1 を 1.2 と小さくした値を用いた。

6 まとめ

本研究では，数学関数最小化問題において有効性が確認されていた，問題に依存せず適応的に近傍を調節するシミュレーテッドアニーリング (SA/AAN) を，連続最適化問題であるタンパク質立体構造のエネルギー最小化問題に適用し，その有効性を検討した。

SA/AAN と岡本らの実験で用いられている逐次 SA の性能を比較するために，SA/AAN を Met-enkephalin, (Ala)₁₀ および C-peptide の 3 種のタンパク質エネルギー最小化に適用した。その結果，いずれのタンパク質においても，SA/AAN は SA よりも高い解探索能力を示すことが明らかとなった。

以上の結果より，SA/AAN はタンパク質のエネルギー最小化計算に対して有効な手法であることが示された。

参考文献

1) Kirkpatrick, S., Gelett Jr. C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, Vol. 220,

No. 4598, pp. 671-680 (1983).

- 2) 岡本 祐幸. モンテカルロシミュレーションで探るタンパク質の折り畳み機構. 物性研究, Vol. 70, No. 6, pp. 719-742, 7 1998.
- 3) 三木 光範, 廣安 知之, 小野 景子: 最適な受理確率を目標とする適応的近傍を持つシミュレーテッドアニーリング, 情報処理学会論文誌 Vol.44, No.1, pp. 1-6(2003) .
- 4) Corana, A., Marchesi, M., Martini, C. and Ridella, S., "Minimizing Multimodal Functions of Continuous Variables with the Simulated Annealing" Algorithm, ACM Trans. on Mathematical Software, Vol. 13, No. 3, pp. 262-280, 1987 .