

シミュレーテッドアニーリングを用いた タンパク質立体構造エネルギー最小化における温度パラメータの検討

廣安 知之[†], 三木 光範[†], 米田 真純^{††}, 岡本 祐幸[‡]

シミュレーテッドアニーリング (Simulated Annealing : SA) では温度の移り変わりを示す温度スケジュールによって得られる解の精度は大きく異なり, 温度スケジュールを決定するパラメータについて様々な研究が行われている. しかし, SA をタンパク質立体構造エネルギー最小化問題に適用した際の温度スケジュールの設定について検討を行った研究は少ない. そこで, 本研究では SA をタンパク質エネルギー最小化問題に適用した際の温度パラメータについて検討を行う. その結果, 従来の温度スケジュールよりも良好な結果を得ることを示した.

The Examination of the Temperature Range in the Simulated Annealing for Minimizing the Energy of Protein Tertiary Structure

Tomoyuki HIROYASU[†], Mitsunori MIKI[†], Masumi YONEDA^{††}, and Yuko OKAMOTO[‡]

In Simulated Annealing (SA), a temperature range gives great influence on an accuracy of a solution. So far, many researches about the determination of the temperature schedule have been shown. However, in the field of minimizing the energy of the protein tertiary structures by SA, not so much researches on the determination of the temperature schedule have been done. This paper shows that the SA which uses the temperature schedule determined by important temperature is more effective than the SA with conventional temperature schedule.

1 はじめに

タンパク質は生物の生命現象を直接担っている重要な物質であり, タンパク質の生化学的機能や特性を決定づけているのが立体構造であるといわれている. 自然に存在するタンパク質の立体構造は, 系の自由エネルギーの最小状態に対応しており, 系の正しいエネルギー関数とアミノ酸配列情報から最小化問題としてタンパク質の立体構造予測が可能であると考えられる. これまでのアミノ酸配列情報からのタンパク質立体構造予測においては, ヒューリスティック法の 1 つであるシミュレーテッドアニーリング (Simulated Annealing : SA) がよく使用されてきた¹⁾.

Kirkpatrickらによって提案された SA は, 局所探索を行いながら解が改善方向へも移動する可能性を残すことで, 局所解に陥ることを防ぎ, 与えられた初期状態から最終的には最適状態に行きつくことが期待されている.

SA の最大の特徴である改善方向への遷移は, 温度と呼ばれるパラメータと状態遷移に伴うエネルギー変位量から確率的に行われる. そのため, SA 試行中の温度の移り変わりを示す温度スケジュールによって得られる解の精度は大きく異なり, 温度スケジュールを決定するパラメータについて様々な研究が行われている^{2, 3, 4)}.

SA の温度スケジュールに関しては様々な研究が行われているが, SA をタンパク質立体構造エネルギー最小化問題に適用した際の温度スケジュールを設定するパラメータについて検討を行った研究は少ない. そこで本研究では, Met-enkephalin を対象問題とし SA をタンパク質立体構造エネルギー最小化問題に適用した際の温度パラメータについて検討を行う. SA において, 特定の一定温度で解探索することにより良好な解探索能力を示すことが報告されており^{2, 3, 4)}, この温度で解の探索を行うことが重要と考えられている. しかし, タンパク質エネルギー最小化問題を対象とした場合には, この温度の存在は確認されていない. そこで, タンパク質エネルギー最小化問題においても, 良好な解探索能力を示す温度領域 (重要温度領域) が存在するかどうかの検証を行う. また, この温度領域とタンパク質が折り畳まれるときに必要な 2 段階の温度との比較を行う. そして, 重要温度領域をもと

[†] 同志社大学工学部
Knowledge Engineering Dept., Doshisha University

^{††} 同志社大学大学院
Graduate School of Engineering, Doshisha University

[‡] 岡崎国立共同研究機構分子科学研究所
Department of Theoretical Studies, Institute for
Molecular Science

にした, SA の温度パラメータの検討を行い, タンパク質エネルギー最小化の効率化について検証する.

2 シミュレ - テッドアニーリング

SA のアルゴリズムでは, 現状態から次状態を生成する「生成処理」, 現状態から次状態に遷移するかどうかを判定する「受理判定」, 現在の温度から次の温度を生成する「クーリング」の 3 つの過程が重要となる⁵⁾.

生成処理では, 現状態から次状態を生起する確率分布として一様分布や正規分布が用いられる.

受理判定では, 通常は式 (1) に示す Metropolis 基準⁶⁾ を用いて, 次状態へ遷移するか否かの判定を行う.

$$A(E, E', T) = \begin{cases} 1 & \text{if } \Delta E < 0 \\ \exp(-\frac{\Delta E}{T}) & \text{otherwise} \end{cases} \quad (1)$$

この値は温度 T によって変化するため, 温度 T はエネルギーが増大する方向への遷移確率に重大な影響を与えるパラメータである.

生成処理, 受理判定をある程度繰り返した後にクーリングを行う. クーリングにおいて, 最適解の漸近収束性を保証するためには式 (2) に示す対数型アニーリング以上に急速に冷やしてはならない⁵⁾.

$$T_k = \frac{T_1}{\log k} \quad (k \geq 2) \quad (2)$$

しかし, 対数型アニーリングでは収束があまりにも遅く膨大な時間を要するため, 通常は式 (3) に示す指数型アニーリングを用いることが多い.

$$T_{k+1} = \gamma T_k (0.8 \leq \gamma < 1) \quad (3)$$

SA 開始時の温度の決定方法については様々な研究が行われている^{4, 7, 8)}. これらの研究では SA の初期温度は, 解の遷移で局所解に陥らないような高い値が設定されており, いずれの研究でも初期温度の決定には多くの予備実験を必要としている.

3 SA における重要温度領域とタンパク質エネルギー最小化における転移温度

3.1 重要温度領域

Connolly²⁾ と Mark^ら^{3, 4)} の研究において, 特定の一定温度で解探索することにより, 良好な解探索能力を示すことが報告されている. SA において, 特定の温度で解の探索を行う事が重要なことから, 本論文ではこの温度のことを重要温度領域と呼ぶ. Mark^らの研究では, 重要温度領域による一定温度 SA を巡回セールスマン問題, 二次配置割り当て問題およびグラフ分割問題に適用し, 従

来の温度スケジュールを用いた逐次 SA より高い解探索能力を示すことを確認した.

Connolly, Mark^らは, 様々な温度で一定温度 SA を何度も試行し, 各温度での解の品質 (式 (4)), 最適解発見率 (式 (5)), 実行時間 (式 (6)) から重要温度領域を決定している.

$$\mathbb{E}[\min\{f(X_1), \dots, f(X_N)\}] \quad (4)$$

$$P(\tau(T_{opt}) \leq N) = \max_{T \in [0, \infty]} P(\tau(T) \leq N) \quad (5)$$

$$\mathbb{E}[\tau(T_{opt})] = \min_{T \in [0, \infty]} \mathbb{E}[\tau(T)] \quad (6)$$

式 (4) に示す $f(X_t)$ は, t 番目の遷移で得られた解 X_t のエネルギーを示す. この式では, 予備実験で決定する評価計算回数 N の一定温度 SA を行い, 各温度で得られた最良エネルギーの中から最良のものを得た温度を重要温度領域としている.

式 (5) では, 一定温度 SA で最適解に到達する確率を全ての温度で比較し, その中から最大値を取る時の温度を重要温度領域としている.

式 (6) では, 一定温度 SA で最適解に到達するまでの時間を全ての温度で比較し, その中で最小の時間を取った温度を重要温度領域としている. また, 一定温度 SA で最適解に到達しない場合は, 最適解から誤差 ϵ の解 x を得るまでに要した時間から, 重要温度領域を決定する.

Mark^らが行った式 (4), 式 (5), 式 (6) から決定した重要温度領域の比較では, どの方法においてもほぼ近似した温度が重要温度領域と考察された. しかし, 評価計算回数 N や誤差 ϵ によって決定する重要温度領域の値が異なり, この重要温度領域を決定するために膨大な計算が必要であることも示されている⁴⁾.

このように組合せ最適化問題に SA を適用した研究においては, 特定の温度領域でのアニーリングが SA の解探索能力に大きく影響することがわかっている.

3.2 転移温度

SA を用いたタンパク質立体構造エネルギー最小化において, タンパク質が折り畳まれるときには最低 2 段階を経ると考えられている⁹⁾. 第 1 段階はランダムコイル状態からコンパクトなグロビュール状態に移行する段階であり, 第 2 段階はコンパクトな構造の中から自然の立体構造 (最小エネルギー構造) に最終的に折り畳まれる段階である. そして, これら 2 つの転移に対応して, それぞれ転移温度が存在する. そのため, SA においてこれらの温度での探索が必要不可欠であると考えられる.

4 数値実験

4.1 対象問題

本論文では Met-enkephalin を対象とし、温度パラメータチューニングによるタンパク質立体構造エネルギー最小化の効率化を行う。

Met-enkephalin は 5 個のアミノ酸残基からなり、主鎖に 10 個、側鎖に 9 個の二面角を持つ。本研究では、各設計変数において順に生成、受理判定を行うこととし、以後この処理を 1Monte Carlo sweep(MCsweep) と呼ぶ。つまり、Met-enkephalin においては 1MCsweep によって 19 回の Metropolis 判定が課される。

Met-enkephalin では、ECEPP/2 エネルギー関数¹⁰⁾に基づいた気相中において、 $E \leq -11\text{kcal/mol}$ の領域で最小エネルギー構造を形成することが確認されている¹¹⁾。

4.2 SA における重要温度領域

タンパク質立体構造エネルギー最小化問題における重要温度領域の存在を確認するために、一定温度でアニーリングを行う SA (一定温度での Metropolis 判定) を様々な温度でタンパク質エネルギー最小化に適用した。実験に用いたパラメータを表 1 に示す。

一定温度での SA に用いるそれぞれの温度は、最高温度 (Maximum Temperature) と最低温度 (Minimum Temperature) の間を等比的に温度数 32 で分割した値を割り当てた。なお、最高温度と最低温度は岡本らが実験で用いたパラメータと同じ値を用いた。

表 1: Parameters of SA with a constant temperature

Parameter	Value
Maximum temperature	2.0
Minimum temperature	0.1
Range size	$180^\circ (180 \times 0.3)^\circ$
Number of temperatures	32
Number of MCsweeps	100000
Trials	50

32 温度で一定温度 SA を行い、各試行で得られた最小エネルギーの平均値を図 1 に示す。図の横軸は温度、縦軸はエネルギーを示す。

図 1 に示すように、特定の温度範囲による一定温度 SA が良好な結果を示しており、重要温度領域が存在することが確認できる。

図 1 から各タンパク質の重要温度領域を検証する。3.1 節に示す重要温度領域の決定方法のうち、本論文では Connolly による式 (4)²⁾ を参考にした。

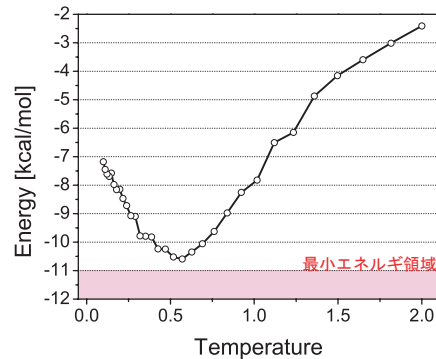


図 1: An important temperature range in a SA with a constant temperature

複数の温度で複数回実行した一定温度 SA の中で、最も良好な平均値を得た温度を重要温度とし、その解と最も不良な平均値 (今回の結果では温度 2.0 での平均値) との差の 15% 以内の解を得た温度領域を重要温度領域と決定する。今回の実験では Met-enkephalin の重要温度は 0.56、重要温度領域は 0.3 ~ 0.8 であった。

4.3 タンパク質エネルギー最小化における転移温度

先ほど求めた重要温度領域と転移温度の比較を行う。岡本は論文⁹⁾において、Met-enkephalin の 2 つの転移温度は「 $0.46 \pm 0.6, 0.56 \pm 0.4$ 」であると述べている。Met-enkephalin においてはこれら 2 つの転移温度が誤差の範囲内でつながっており、その範囲は 0.4 ~ 0.6 である。我々が求めた Met-enkephalin の重要温度領域は 0.3 ~ 0.8 であり、岡本らの定義する転移温度の範囲を包含している。

4.4 温度パラメータの検討

4.2 節の実験で求めた重要温度領域はタンパク質が折り畳まれる段階に必要な転移温度に対応している。そこで、重要温度領域の値をもとにチューニングした温度パラメータを設定し、解探索の効率化を図る。

Met-enkephalin の重要温度領域は 0.3 ~ 0.8 であるため、最高温度を 0.8、最低温度を 0.3 とした温度パラメータと従来から用いていた最高温度 2.0、最低温度 0.1 の温度パラメータでの結果と比較を行う。実験に用いたパラメータを表 2 に示す。4.2 節の実験における一定温度 SA の最高温度と最低温度と区別するために、最高温度を Initial Temperature、最低温度を Last Temperature とする。

比較結果を図 2、図 3 に示す。図 2 ではエネルギー履歴の比較を行っており、横軸は MCsweep 数、縦軸はエネルギー、そして凡例は「最高温度 - 最低温度」を示している。また、図 3 では、最小エネ

表 2: Parameters

Parameter	Value	
Initial temperature	2.0	0.8
Last temperature	0.1	0.3
Range size	180 ° (180 × 0.3) °	
Number of MCSweeps	100000	
Trials	50	

ルギ領域到達率 (Success Rate) と各試行の最小エネルギーの中央値での比較を行い、横軸は「最高温度 - 最低温度」、縦軸は最小エネルギー領域到達率とエネルギーを示している。

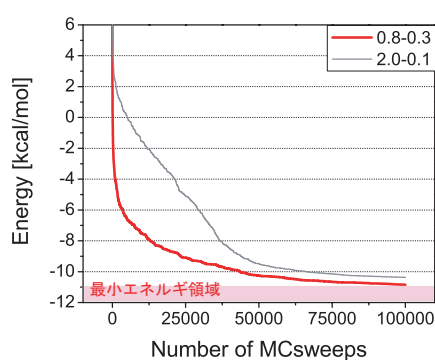


図 2: Result of SA changing temperature schedule (Energy transition)

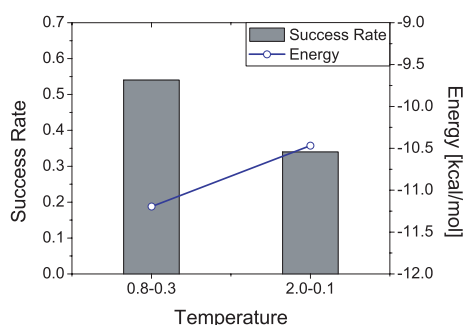


図 3: Result of SA changing temperature schedule (Success Rate, Energy)

図 2 より、重要温度領域を考慮した温度パラメータ設定では従来から用いていたものと比較して、探索序盤から有効な解探索が行われ、探索終盤も局所解に陥らずに探索が進んでいる。また、図 3 より、最小エネルギー領域到達率、エネルギー値ともに良好な結果を得ており、最小エネルギー領域到達率においては 20 %もの向上が見られる。

これらのことから、重要温度領域を考慮した温度スケジュールによって探索序盤から有効な解探索を行い、効率的な解探索が行えることが示された。

5 結論

本論文では、SA をタンパク質立体構造エネルギー最小化問題に適用した際の温度スケジュールを設定するパラメータについて検討を行った。SA において解の探索に重要な温度パラメータについて検討を行い、タンパク質エネルギー最小化問題における重要温度領域の存在を確認した。また、その温度領域とタンパク質の折り畳みに必要不可欠である転移温度との比較を行い、重要温度領域と転移温度が対応していることを確認した。そして、重要温度領域を考慮した温度パラメータを設定することによって、効率的な解探索が行えることを示した。

参考文献

- 1) Hikaru Kawai, Takeshi Kikuchi, and Yuko Okamoto. A prediction of tertiary structures of peptide by the Monte Carlo simulated annealing method. *Protein Engineering*, Vol. 3, No. 2, pp. 85–94, 1989.
- 2) David T.CONNOLLY. An improved annealing scheme for the qap. *European Journal of Operational Research*, Vol. 46, pp. 93–100, 1990.
- 3) Mark Fielding Harry Cohn. Simulated annealing: Searching for an optimal temperature schedule. *SIAM J. Optim.*, Vol. 9, pp. 779–802, 1999.
- 4) Mark Fielding. Simulated annealing with an optimal fixed temperature. *SIAM J.*, Vol. 11, No. 2, pp. 289–307, 2000.
- 5) 中野良平 Bruce E Rosen. シミュレーテッドアニーリング -基礎と最新技術-. 人工知能学会誌, Vol. 9, No. 3, 1994.
- 6) Jr. M.P.Vecchi S.Kirkpatrick, C.D.Gelatt. Optimization by simulated annealing. *Science*, Vol. 220, No. 4598, pp. 671–680, 1983.
- 7) Steave R. White. Concepts of scale in simulated annealing. *Proceeding IEEE Intl. Conf. Comp. Des. (ICCD)*, pp. 646–651, 1984.
- 8) L.A. McGeoch D.S. Johnson, C.R. Aragon and C.Shevon. Optimization by simulated annealing : An experimental evaluation, part2, graph coloring and number partitioning. *Operations Research*, Vol. 39, pp. 378–406, 1991.
- 9) 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, 1998.
- 10) F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga. *J. Phys. Chem.*, Vol. 79, pp. 2361–2381, 1975.
- 11) Yuko OKAMOTO, Takeshi KIKUCHI, and Hikaru KAWAII. Prediction of low-energy structure of met-enkephalin by monte carlo simulated annealing. *CHEMISTRY LETTERS*, pp. 1275–1278, 1992.