

タンパク質立体構造予測における温度パラメータチューニング

The importance of the temperature range in the Simulated Annealing applied to the Protein Tertiary Structure Prediction

米田 真純

Masumi YONEDA

Abstract: The Protein Tertiary Structure Prediction is one of the optimization problems. This paper examined a temperature schedule in the Simulated Annealing applied to the Protein Tertiary Structure Prediction. Consequently, it turns out that a best temperature schedule changes with target protein.

1 はじめに

タンパク質立体構造予測は、最適化問題の 1 つである。タンパク質は生命現象に直接関わる重要なものであるため、構造を解明することは生命の仕組みを説明することにもつながる。タンパク質の立体構造はエネルギーの最小状態に対応しており、エネルギーが最小となるような構造を最適化手法を用いて求めることで、構造予測が可能である。これまで、タンパク質の構造予測においては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⁾。

SA は、物質を高温から徐々に冷却し欠陥の少ない低エネルギーの結晶を得るといった焼きなましの過程をシミュレートした手法である²⁾。SA では改悪方向への解遷移を確率的に受理することで、最適状態に遷移していくことが期待されている²⁾。しかし、この遷移は温度パラメータに依存しており、そのパラメータのチューニングには多くの予備実験が必要となる。

これまで知的システムデザイン研究室タンパク質班では、岡崎国立共同研究機構分子科学研究所の岡本らが実験で用いた最高温度と最低温度をすべてのタンパク質に適用しており、最適パラメータの検証を行っていなかった。そこで、本研究では温度パラメータのチューニングを行い、解精度について検証する。

2 タンパク質立体構造予測における重要温度領域

組合せ最適化問題に SA を適用した研究において、特定の温度領域でのアニーリングが SA の解探索性能に大きく影響することが分かっている。しかし SA をタンパク質立体構造予測に適用する場合において、この温度領域 (重要温度領域) の存在は確認されていない。そこで、一定温度でアニーリングを行う SA をタンパク質立体構造予測に適用し、重要温度領域の存在を確認する。

実験では、最高温度 2.0 と最低温度 0.01 の間を等比的

に 32 分割し、一定温度で探索を行う。各温度の総探索数は 6000MCsweep とした。対象問題を Met-enkephalin, (Ala)₁₀, (Val)₁₀ と C-peptide としたときの実験結果を Fig. 1 に示す。Fig. 1 は 50 回試行の平均値であり、横軸は温度、縦軸はエネルギーを示している。なお、ECEPP/2 エネルギー関数¹⁾に基づいた気相中において、Met-enkephalin は -11.0kcal/mol 以下、(Ala)₁₀ は -9.7kcal/mol 以下、(Val)₁₀ は 1.8kcal/mol 以下、C-peptide は -42.0kcal/mol 以下の領域で最小エネルギー構造をとることが確認されている。

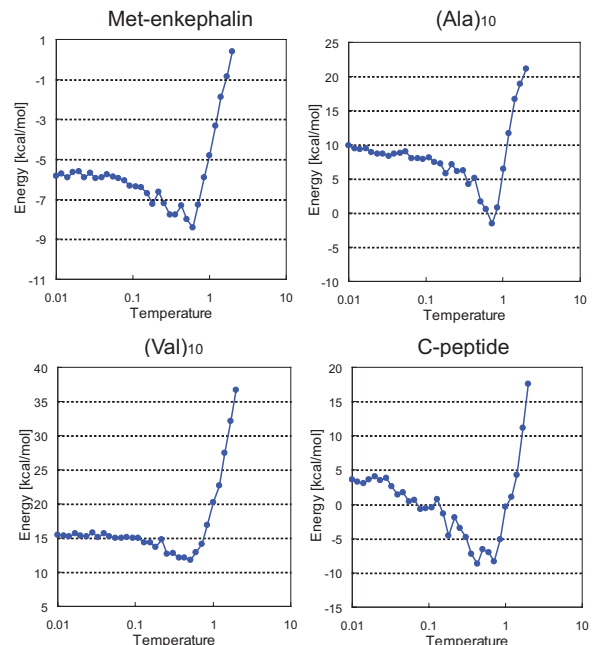


Fig. 1 タンパク質立体構造予測における重要温度領域

Fig. 1 より、Met-enkephalin においては温度 0.7 付近、(Ala)₁₀ においては温度 0.9 付近、(Val)₁₀ においては温度 0.7 付近、さらに C-peptide においては温度 0.8 付近で他の温度よりも精度の良い解が得られており、タンパク質立体構造予測において重要温度領域の存在が確認できる。しかし、特定の温度のみの探索では最適解を発見することが困難であることがわかる。

3 タンパク質立体構造予測における温度パラメータチューニング

前節において一定温度 SA をタンパク質立体構造予測に適用し、重要温度領域の存在を確認した。そこで、この温度領域を考慮した温度スケジュールと従来のパラメータを用いた温度スケジュールの比較を行う。従来のパラメータを Table 1 に示す。計算時間の都合上、C-peptide を対象問題とした場合には総探索数を 6000MCsweep とした。

パラメータ	値
最高温度	2.0
最低温度	0.1
近傍	180 ° 54 °
総 MCsweep 数	100000(C-peptide のみ 6000)
試行回数	50

3.1 最高温度パラメータに関する検討

本節では重要温度領域を考慮した最高温度の設定を行う。実験では、最高温度のみを調節し、最高温度以外のパラメータについては Table 1 と同値にした SA を行った。実験結果を Fig. 2 に示す。Fig. 2 において、C-peptide のグラフは 50 回試行の平均値、その他は 50 回試行の最適解発見率である。横軸は最高温度を示している。

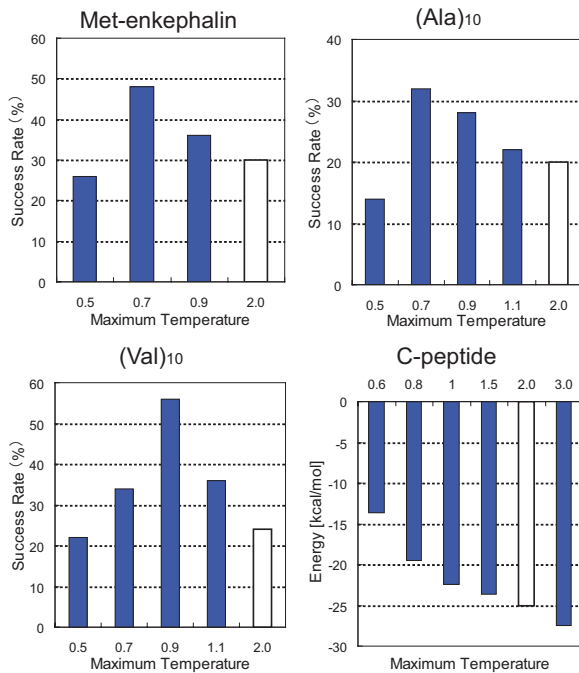


Fig. 2 最高温度パラメータに関する検討

Fig. 2 より、Met-enkephalin、(Ala)₁₀ と (Val)₁₀ については最高温度を重要温度付近まで下げることで最適解発見率が向上した。しかし、C-peptide については、最高温度を従来の 2.0 から 3.0 へと上げることで低いエネルギー値が得られた。

3.2 最低温度パラメータに関する検討

本節では重要温度領域を考慮した最低温度の設定を行う。実験では、最低温度のみを調節し、最低温度以外のパラメータについては Table 1 と同値にした SA を行った。実験結果を Fig. 3 に示す。Fig. 3 において、C-peptide のグラフは 50 回試行の平均値、その他は 50 回試行の最適解発見率である。横軸は最高温度を示している。

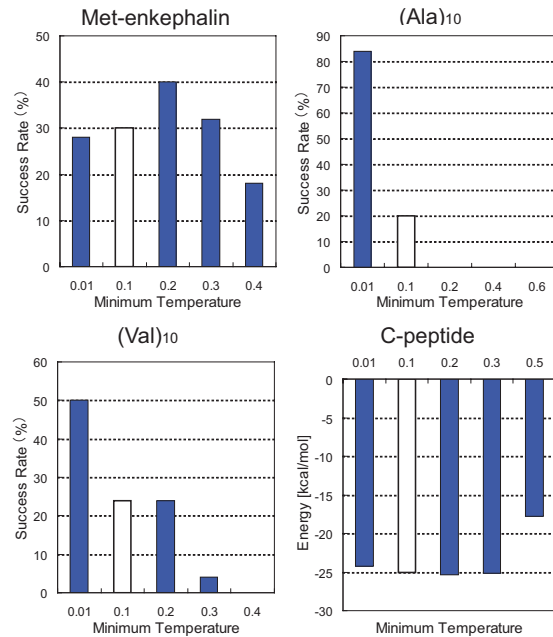


Fig. 3 最低温度パラメータに関する検討

Fig. 3 より、Met-enkephalin においては最低温度を重要温度付近とすることにより最適解発見率が向上した。しかし、(Ala)₁₀ と (Val)₁₀ については最低温度を従来の 0.1 よりも低い 0.01 にすると、最適解発見率が大幅に向上した。また、C-peptide については最低温度が 0.5 以外の結果では、ほぼ同等のエネルギー値であった。

4 結論

本研究では、SA を用いたタンパク質立体構造予測における重要温度領域の検証および、重要温度領域を考慮した温度パラメータの設定を行った。

実験結果より、対象としたタンパク質において重要温度領域が存在する、一定温度のみの探索では最適解を発見することは困難である、そして、各タンパク質で最適な温度スケジュールが存在するということが示された。

今後は SA の探索で、より良好なタンパク質立体構造を得るために最適パラメータの設定、もしくは最適な温度スケジュールを自動的に設定する必要がある。

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. 1989.
- 2) Bruce E. Rosen, 中野 良平. 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. 1994.