

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
米田 真純

1 タンパク質の立体構造エネルギー最小化

1.1 序論

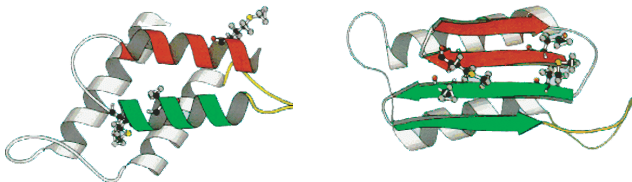
タンパク質は生物の様々な生命現象を担っている物質の一つである。タンパク質の機能は、構造と密接に関わっており、立体構造の解明は新薬の開発や病理の解明につながると考えられている。しかし、現在では多くのタンパク質において、その立体構造が未知であるため、立体構造の解析と、立体構造と機能の対応については世界中で盛んに研究されている。

1.2 タンパク質の立体構造予測の有用性

自然のタンパク質は複数個のアミノ酸が連なった鎖状の生体高分子で、特定の立体構造に折り畳まれた状態で存在している。タンパク質はアミノ酸配列が特定の立体構造をとる時のみ、その生化学的機能を発揮する。

タンパク質の立体構造を解明することは機能の解明にもつながり、新薬の設計、特定の機能を持った人工タンパク質の設計、タンパク質の誤った折り畳みに起因する病理の研究につながると考えられている。

生体内において、正常でない立体構造をもつタンパク質は本来の機能を発揮できないため、病気の原因などになっている。誤った折り畳みによる症例としては、アルツハイマー病やプリオン病が挙げられる。プリオンタンパク質の折り畳みの例を Fig. 1 に示す。



正常なプリオンタンパク 感染症のあるプリオンタンパク
Fig. 1 折り畳みの異なるプリオンタンパク質

1.3 タンパク質の立体構造予測の手法

これまでのタンパク質の構造研究は、実際のタンパク質を用いる、核磁気共鳴法や X 線結晶構造解析といった実験生物学的なアプローチがほとんどであった。しかし、実験環境が解析結果に影響することや、純粋なタンパク質が大量に必要であるなどの問題点が存在する。一方、立体構造予測方法としてはコンピュータシミュレーションによるタンパク質の立体構造予測がある。

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いることが明らかとなっている。このため、系の正しいエネルギー関数が与えら

れば、アミノ酸配列情報からコンピュータシミュレーションにより最小化問題 (最適化問題) としてタンパク質の立体構造が予測可能であると考えられる。

しかし、タンパク質はアミノ酸原子間の結合を中心として自由に回転できるため、タンパク質の折り畳まれ方は原理的には膨大な数にのぼることが予想される。このことから、タンパク質のエネルギー関数には Fig. 2 のように局所解が無数にあると考えられ、局所解に捕捉されにくい最適化手法を用いることが必要となる。

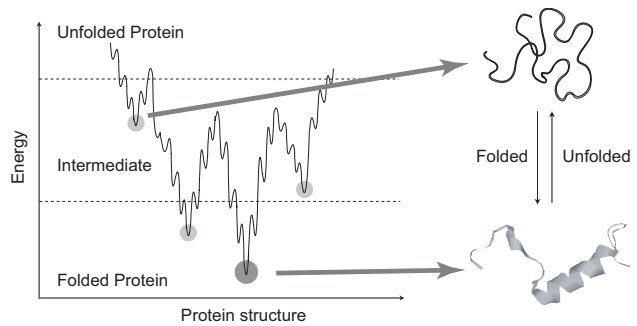


Fig. 2 タンパク質のエネルギー関数と立体構造の関係

これまで、我々の研究グループではシミュレーテッドアニーリング (SA) を用いてタンパク質エネルギー最小化を行ってきた。そして、遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) と呼ばれる手法を提案しており、この手法を用いていくつかのタンパク質に対してエネルギー最小化を行っている。また去年から、SA だけでなく GA によるアプローチも行っている。

1.4 PSA/GAc によるタンパク質エネルギー最小化

タンパク質グループでは、これまでに小規模なタンパク質の立体構造エネルギー最小化を行ってきた。Fig. 3 はヒト副甲状腺ホルモンのフラグメントの立体構造結果である。

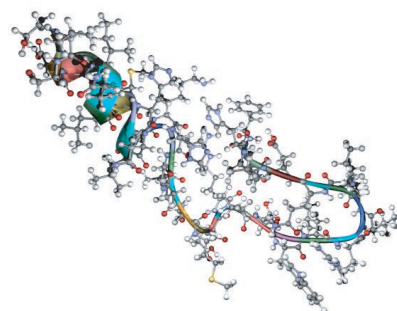


Fig. 3 PTH(1-34) の立体構造予測結果

2 SA を用いたタンパク質立体構造エネルギー最小化

2.1 序論

SA 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によって、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¹⁾。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、対象問題に適した温度スケジュールを決定しなくてはならない²⁾。

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検討を行い、温度パラメータと解精度の関係について検証する。

2.2 タンパク質立体構造エネルギー最小化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¹⁾。しかし SA をタンパク質のエネルギー最小化問題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 をタンパク質のエネルギー最小化問題に適用し、重要温度領域の存在を確認する。

Met-enkephalin, (Ala)₁₀ と C-peptide の 3 つのタンパク質を対象問題としたときの実験結果を Fig. 4 に示す。Fig. 4 は各試行で得られた最小エネルギーの平均値であり、横軸は温度、縦軸はエネルギーを示している。なお、本研究では、Met-enkephalin は -11.0kcal/mol 以下、(Ala)₁₀ は -9.7kcal/mol 以下、C-peptide は -42.0kcal/mol 以下を最適解とした。

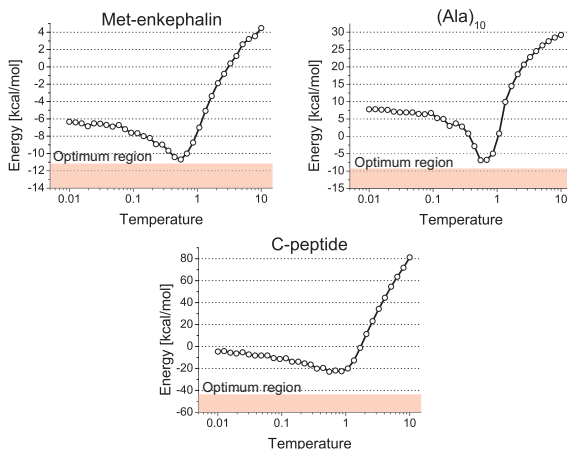


Fig. 4 一定温度 SA による重要温度領域の検証

Fig. 4 より、Met-enkephalin においては温度 0.2 ~ 0.7, (Ala)₁₀ においては温度 0.4 ~ 0.8, C-peptide においては温度 0.3 ~ 1.1 の範囲で他の温度よりも精度の良い解が得られており、タンパク質のエネルギー最小化において重要温度領域の存在が確認できる。しかし、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ことがわかる。

2.3 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

本節では重要温度領域を考慮した最高温度および最低温度の調節を行う。実験では、岡本らが実験で用いた温度スケジュール（最高温度 2.0, 最低温度 0.1）の一方を固定し、他方を調節した。

検討を行った中で、最も良好な結果を示した試行の温度スケジュール（最高温度 最低温度）および最適解領域到達率を Table 1 に示す。Table 1 の Success Rate の括弧中は、岡本らが実験で用いた温度スケジュールとの比較結果を示している。

Table 1 温度パラメータの検討結果

Problem	Temperature Schedule		Success Rate
Met-enkephalin	0.7	0.1	48 % (+18 %)
(Ala) ₁₀	2.0	0.01	84 % (+64 %)
C-peptide	2.0	0.1	40 %

Met-enkephalin については、ほぼ重要温度領域のみを探索する温度スケジュールが最も良好な結果を示した。

(Ala)₁₀ については、岡本らが用いた温度 0.1 よりも下げた、温度 0.01 を最低温度とした試行が良好な結果を示した。このことより、(Ala)₁₀ を対象とした場合は 0.1 ~ 0.01 の温度領域での探索が必要であると考えられる。

C-peptide を対象とした場合には、岡本らが実験で用いた、最高温度を 2.0, 最低温度を 0.1 とした試行が良好な結果を示し、岡本らのパラメータ設定は妥当であるといえる。

2.4 結論

本研究では、SA 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は存在するが、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ことを確認した。そして、重要温度領域を考慮した温度スケジュールを設定することにより、良好な結果を得ることができた。

参考文献

- 1) 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.
- 2) Harry Cohn, Mark Fielding. Simulated Annealing: Searching for an optimal temperature schedule. SIAM J. Optim. 1999.