

部分構造最適化に分子動力学法を取り入れたタンパク質立体構造予測の提案

吉田 昌太

1 はじめに

近年、タンパク質の立体構造予測が注目されている。タンパク質はアミノ酸が複数連なって構成される物質で、自然界ではある決まった構造に折りたたまれた状態で存在している。この構造をタンパク質の立体構造と呼び、立体構造と機能は密接に関わっていると言われていいる。そのため、立体構造を解明することは病理の解明や新薬の開発につながることを期待されている。また、タンパク質は最もエネルギーの低い安定した構造に折りたたまれるため、エネルギー最小化問題と捉えることができる。本研究室では、「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 (Parallel Simulated Annealing with Genetic Crossover : PSA/GAc) を用いて立体構造予測を行う。本報告では、現在の解析手法に新たに分子動力学法を組み合わせ、その有効性を検証する。

2 タンパク質立体構造予測

2.1 エネルギー関数

本研究室では、目的関数として、タンパク質の系をモデル化したエネルギー関数を用いる。しかし、タンパク質のエネルギー関数は非常に複雑で、大域的にいくつかの、局所的に無数の極小値を持つと考えられている。そのため、分子動力学計算プログラムパッケージである TINKER¹⁾ に名古屋大学の岡本先生が手を加えたものをエネルギー関数として使用する。設計変数は、原子間の回転角である二面角を用いる。

2.2 予測結果の評価基準

タンパク質の立体構造予測を行う際に評価する基準は 2 つある。1 つはエネルギー値、もう 1 つは構造の形である。構造が既知のタンパク質ならば、シミュレーション結果がどの程度その構造に似ているかが重要となる。2 つの構造の差異を定量化するために用いられる量が RMSD (Root Mean Square Deviation) である。RMSD は 2 つの分子構造を重ね合わせて、対応する各原子のずれの二乗を平均したものの平方根で定義される。式 (1) に RMSD の求め方を示す。

$$RMSD(A, B) = \sqrt{\frac{1}{N} \sum_{i=1}^N (a_i - b_i)^2} \quad (1)$$

RMSD の単位は Å で、値が小さいほど 2 つの構造がよく似ていることになる。

3 分子動力学法

3.1 概要

分子動力学法 (Molecular Dynamics method) は、2 体あるいはそれ以上の原子間ポテンシャルの下に、古典力学におけるニュートン方程式を解いて、系の静的、動的安定構造や、動的過程を解析する手法のことである。分子動力学法が扱える系の規模としては、最大で数億原子からなる系の計算例がある。

しかしながら分子間の相互作用は古典的力場関数では精度よく記述できない場合が多く、第一原理 (量子力学) に基づいた巨大分子計算手法の実現が必要であり、この古典力学の限界を解決するために考案された手法が第一原理分子動力学法である。以後、第一原理分子動力学のことを MD と呼ぶ。しかし、MD では最大で数千個の原子数しか扱えず、格段に減るとい問題点もある。Fig. 1 に分子動力学法の流れを示す。

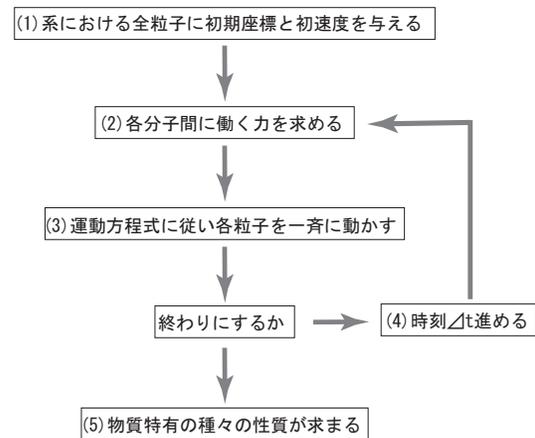


Fig. 1 分子動力学法の流れ (出典：自作)

3.2 分子動力学法における温度パラメータの定義

分子動力学法にも温度という概念が存在する。しかし、本研究室では MD を使用することが初めてのため、温度パラメータの調整が難しい。そこで Fig. 2, Fig. 3 に示す、名古屋大学の岡本先生が定義したパラメータを基に検討を行う。

Fig. 2, Fig. 3 に示すように、350K ではどの力場においても時間が進むにつれて構造も変化しているが、300K ではほとんど変化が見ることができない。そのため 300K 以下ではタンパク質の立体構造はほとんど変化しないと考えられるため、この値を基準として温度をあげて検討する必要があると考えられる。

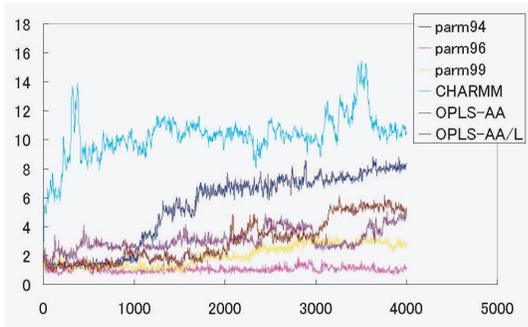


Fig. 2 温度 350K における探索履歴
(出典：参考文献 2 より引用)

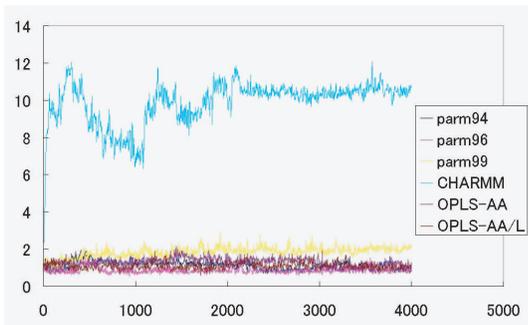


Fig. 3 温度 300K における探索履歴
(出典：参考文献 2 より引用)

3.3 分子動力学法における温度パラメータの検討

3.2 節で挙げたように 300K を基準として温度パラメータの検討を行う。以下に示す max 温度から min 温度へ変化させる。本報告では、初期状態から MD を 100 万回計算させたものと、PSA/GAc で最適化を行い、その後さらに 100 万回 MD を実行させたものを検討する。

max	1500	1200	1000	800	500	100	50	10
min	300	300	300	300	300	0	0	0

Table 1 temperature

このように、min 温度を 0K に設定したものは、PSA/GAc で最適化を行った値よりも低くなっていることがわかる。これにより、温度が低い状態だとエネルギーがどんどん下がること、また、PSA/GAc では完全にエネルギー最小化を行えていないことが明らかになった。

次に各温度における立体構造の RMSD 値を示す。Fig. 5 を見てわかるように、PSA/GAc 後に MD を実行した構造のほうが全体的に RMSD 値がよくなっており、PSA/GAc と MD を組み合わせることで精度の良い結果が出るのがわかった。

温度 (K)	Energy(kcal/mol)
1500-300	-796.442
1200-300	-805.598
1000-300	-823.356
800-300	-834.188
500-300	-830.838
500-0	-1023.73
100-0	-989.445
50-0	-983.096

Table 2 初期状態から MD を実行した構造のエネルギー値

温度 (K)	Energy(kcal/mol)
PSA/GAc 後のエネルギー	-905.208
1500-300	-782.802
1200-300	-797.27
1000-300	-802.206
800-300	-815.324
500-300	-796.864
500-0	-992.99
100-0	-980.133
50-0	-980.353

Table 3 PSA/GAc 後に MD を実行した構造のエネルギー値

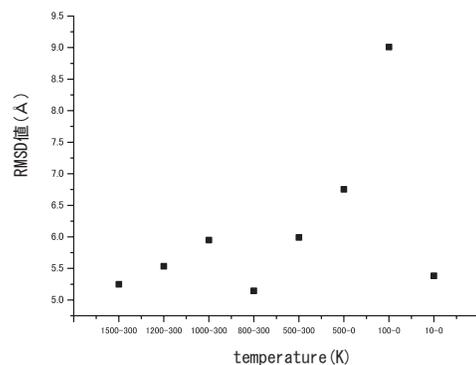


Fig. 4 初期状態から MD を実行した構造の RMSD 値
(出典：自作)

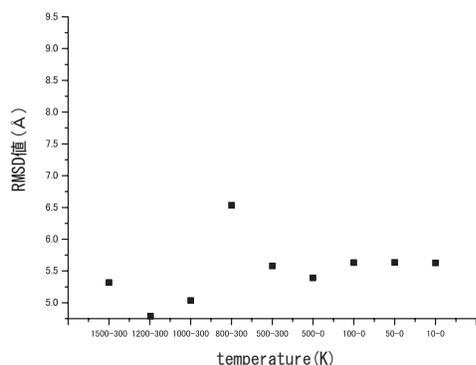


Fig. 5 PSA/GAc 後に MD を実行した構造の RMSD 値 (出典：自作)

4 部分構造最適化に分子動力学法を取り入れたタンパク質立体構造予測の提案

4.1 部分構造に分割する有効性

現段階で PSA/GAc で解くことができるタンパク質は非常に小規模なものに限られている。例えば、13 残基からなる C-peptide などは、エネルギー最小化で良い結果が得られている。しかし、Protein-A や Protein-G などのタンパク質においては、いまだ良い結果が得られていない。しかし、全体としては天然構造に似ていなくても、部分的な構造は非常に良く似ている場合がある。つまり、ある部分構造ではうまく予測できるが、ある部分構造の予測はうまくいかないということである。その例を Fig. 6 と、Fig. 7 に示す。この部分構造は Protein-A の各部分を最適化した結果である。Fig. 6 は 16~30 残基、Fig. 7 は 39~54 残基である。また、左に Native 構造、右に実験で得た最小エネルギー構造を示す。

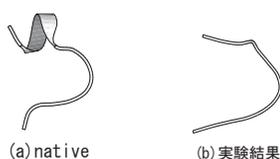


Fig. 6 Protein-A における 16~30 残基 (出典：自作)



Fig. 7 Protein-A における 39~54 残基 (出典：自作)

以上のように、ターンの部分も α ヘリックスもうまく予測できていることがわかる。そのため部分構造最適化の組み合わせによるタンパク質の立体構造予測は有効性があると考えられる。

4.2 分子動力学法を取り入れた提案手法の概要

4.1 節に示したように、部分構造を最適化し組み合わせることの有効性は示された。しかし、その部分構造も nativa と全く同じというわけではない。そこで章で示した分子動力学を導入することでより精度の良い結果が得られるのではないかと考えた。そこで分子動力学を導入した新たな手法を考案した。それを Fig. 8 に示す。

また Fig. 8 にあるように、部分構造が示されても、現在は一つのタンパク質として組み直すことが容易ではない。そこで、ユーザの好きなように部分構造を組み合わせることができたり、ユーザが最適化させたい部分だけを最適化させ残りの構造は固定させる自動化プログラムを作成した。

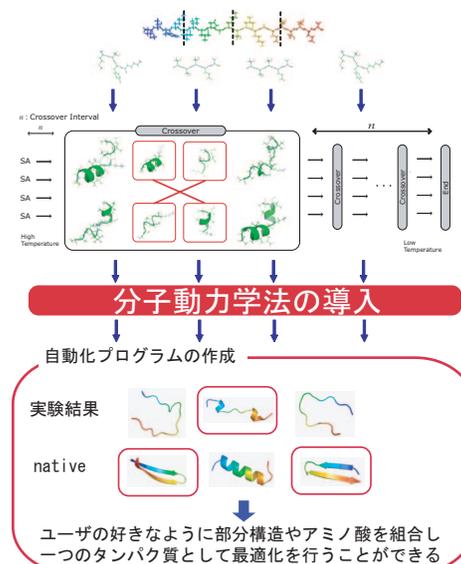


Fig. 8 分子動力学法を取り入れた提案手法 (出典：自作)

5 まとめ

本報告では、分子動力学法の有効性の検証と部分構造最適化に分子動力学法を導入した新たなタンパク質の立体構造予測の提案を行った。実験により、以前よりは精度が向上したが、まだ native と同じ構造にはならないため、分子動力学の使用法やパラメータの検討がより必要だと考えられる。そして、今後はこれらの分析を元に精度を向上させ大規模なタンパク質の構造予測を完成させたい。

参考文献

- 1) TINKER Home Page .
<http://dasher.wustl.edu/tinker/>
- 2) 計算結果報告 20060118 榮慶丈著