

# 粗視化モデルを用いた 3 ヘリックスバンドルのフォールディング過程 についての考察

鈴木 陽子<sup>1</sup> 押本 雄輔<sup>2</sup>

## Investigation for Protein Folding Landscape of 3-helix Bundle through $C_{\alpha}$ model simulation

Yoko SUZUKI<sup>1</sup>, Yusuke OSHIMOTO<sup>2</sup>

Based on  $C_{\alpha}$  Gō model, we performed simulations to investigate a free energy landscape of 3-helix bundle protein. We used WHAM (Weighted Histogram Analysis Method) algorithm to obtain thermodynamic variables and free energy landscapes for different temperatures. It has two basins in the landscape corresponding to the folded and unfolded states. Consequently, the folding pathway between the folded and unfolded states is identified. We found that there is no intermediate state in the path.

キーワード：タンパク質フォールディング 1, シミュレーション 2

Keywords : Protein Folding1, Simulation2

### 1. 序

蛋白質は生命が機能を維持する上で様々な機能を担っている分子である。蛋白質は 20 種類のアミノ酸が鎖状に連なっている分子として合成されたのち、折り畳まれて立体構造をとることにより、生体内で機能を果たすことができる。従って蛋白質の生体内での機能を調べたり、その制御法を調べたりするためには、蛋白質の立体構造を解明することが重要である。本研究では、郷モデルを用いたシミュレーションを行うことにより 3 ヘリックスバンドルと呼ばれる蛋白質の折り畳み（フォールディング）過程を調べ、構造の安定性について考察した結果について述べる。

### 2. 目的

#### 2.1 背景

蛋白質は、数十から数千個ものアミノ酸が連なり、数千から数十万個の原子でできる巨大で複雑な分子である。アミノ酸同士はペプチド結合によってつながっており、1 次元配列をなしている。1 次元に連なったアミノ酸は、折り畳まれて 3 次元な構造をとることにより、生体内で機能を発揮することができる。このため、蛋白質の 3 次元立体構造を研究することは、蛋白質自体の機能を知り、それを分子物理学のみならず、分子病理学やバイオテクノロジーの分野へ応用する上で、重要である。

蛋白質の立体構造を理論的に調べる手法として、分子動力学シミュレーションがある。これは蛋白質分子に含まれ

る各原子についての運動方程式を、コンピューターを用いて解くことにより、それぞれの原子の軌跡（トラジェクトリー）を調べ、蛋白質の構造や機能について理解しようとするものである。しかし、分子内におけるすべての原子のトラジェクトリーを追うことは、原子数が多ければ多いほど、膨大な計算資源が必要である。蛋白質を構成する原子数が多いことから、現実的な計算をするには、粗視化を行う必要性がでてくる。

蛋白質の折り畳み（フォールディング）を理論的に考えたとき、天然立体構造の壊れた状態（変性状態）から天然状態に向かって構造を変化させていく際に、蛋白質のとりうる形状について、まったくのランダムサーチで天然状態に達すると仮定すると、フォールディングに要する時間は天文学的な数値になってしまうことが、Levinthal により指摘されている。実際にフォールディングに要する時間はミリ秒から秒のオーダーなので、これは明らかに矛盾している（Levinthal のパラドックス）。

このパラドックスを回避する理論の 1 つにファネル理論がある<sup>(1)</sup>。これは、巻き戻りが進行するにつれて、経路の数が減少し、折り畳まれた状態（天然状態）へ収斂していくというものである。ファネル理論を模式図で表したものが図 1 である。図 1 は、エネルギーを縦軸に、蛋白質の個々の形状に対応した構造空間を横軸にとっており、最小の状態が天然状態に対応し、他の状態よりも際立って深い谷となっていることが、折り畳みがおこるための必要条件になっている。蛋白質は非常に複雑な構造をしているため、エ

1 明星大学理工学部総合理工学科 准教授 専門分野 統計力学・生物物理学

2 明星大学大学院理工学研究科 物理学専攻 博士前期課程 専門分野 統計物理学

エネルギー曲面もでこぼこして多くの極小状態（準安定状態）が存在しているが、蛋白質はこのような極小状態に落ちこんだとしても、熱揺らぎによって極小状態から抜け出し、最終的には、より深い谷になっている天然構造へと巻き戻っていくと考えられている。即ち、蛋白質はこのような構造をとるように、うまく設計されているともいえる。

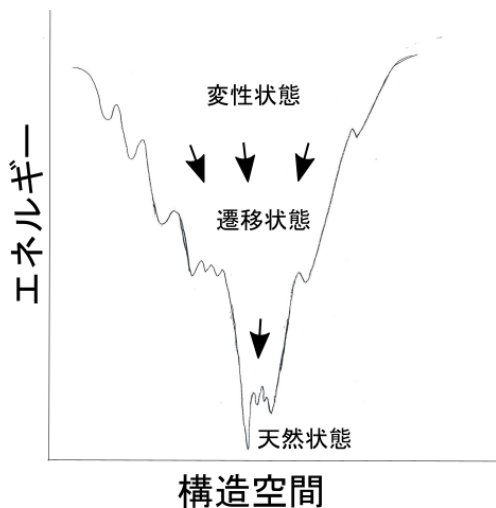


図1 蛋白質の折り畳みの  
エネルギー地形：

縦軸のエネルギーは、蛋白質内の相互作用エネルギーと溶媒による水和自由エネルギーの和になっている。

ファネル理論を具体的に記述するモデルに郷モデルがある<sup>(2,3)</sup>。蛋白質においては、変性状態と天然状態との間の構造転移が、ある種の協同性をもって進むことが知られている。郷モデルはフォールディングが起こる際のこのような共同的ふるまいに着目して作られたものである。

郷モデルについて説明するためには、蛋白質の階層構造について説明する必要がある。蛋白質は、直鎖上に並んだアミノ酸配列からなっているが、これを1次構造と呼ぶ。アミノ酸配列上で近い残基間に働く相互作用（近距離相互作用）により形成される特徴的な構造は2次構造と呼ばれ、らせん状構造（ヘリックス）やプリーツ状の構造（シート）などをとっている。2次構造を形成した複数の部位は、熱力学的安定性をもつように集まって3次元立体構造を

くるが、この構造は3次構造と呼ばれている。また、複数の蛋白質が集まって全体として機能を発現することがあり、この場合の構造は4次構造と呼ばれている。

郷モデルにおいて重要となる構造は、近距離相互作用によって形成される2次構造と遠距離相互作用によって形成される3次構造である。近距離相互作用で形成される2次構造をもった部位同士が安定化するように遠距離相互作用が働く、即ちフォールディングの過程において近距離相互作用も遠距離相互作用も整合して全体を安定化させるような協同性があると考えるモデルが郷モデルである。

郷ポテンシャルは、この特性を具体的に表している。郷ポテンシャルでは、ビーズ鎖を基礎として、フォールドしたとき接触するビーズ間のみにも引力を仮定したものである<sup>(3)</sup>。このポテンシャルが働くビーズ鎖は、接触しやすい近距離の相互作用を最初に形成し、次に遠距離の相互作用を形成して構造を形成する。郷ポテンシャルは、構造の選択性が非常に高い選択性を持ち、余計な構造探索を行わずに、間違った構造を回避して短時間内に構造を形成する。これはファネル構造においては、フォールディングが進むにつれ、際立って自由エネルギーが低い極小状態に収斂していくことに対応している。

本研究では以上のことを踏まえ、 $C_{\alpha}$ 郷モデルを用いて、分子動力学シミュレーションを行った。ここで $C_{\alpha}$ モデルは、蛋白質を構成するアミノ酸（残基）1つ1つをビーズとみなし、蛋白質のアミノ酸配列をそのビーズが1次的に連なった鎖と考えるモデルである。即ち $C_{\alpha}$ 郷モデルは、残基をビーズとみなし、残基間の引力相互作用に郷ポテンシャルを適用したものである<sup>(4)</sup>。

構造の安定性を調べるために、蛋白質の折り畳みの経路として自由エネルギー地形をベースに考え、その経路における残基間のコンタクト（蛋白質が折り畳まれたとき、ある一定の距離の中に残基同士が近づいているとき、その残基対はコンタクトをしていると定義する）がどのくらい安定なものなのかを考察した。

具体的には、3ヘリックスバンドルの構造の安定性について、分子動力学シミュレーションを行った。

## 2・2 3ヘリックスバンドル

計算に用いた3ヘリックスバンドルは、アミノ酸残基55個がつらなる原子数891個の分子であり、図2のように3つの $\alpha$ ヘリックス（螺旋状の構造）が束ねられているような形をなす。残基7から18、27から36、41から55の部位が、それぞれ $\alpha$ ヘリックスにあたる。

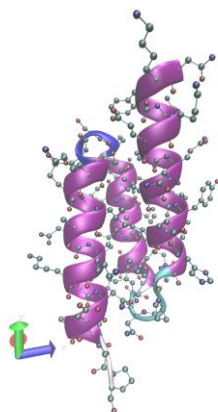


図2 3ヘリックス・バンドルの天然構造：  
X線構造解析による結果(4zdm.pdb)をもとに描  
画したもので、2次構造(局所構造)がわかるよ  
うに、リボン表示を用いて表したもので、ら  
せん構造の部分は、 $\alpha$ -ヘリックス、左上  
の青い部分は $\beta$ -ヘアピン、右下の水色の  
部分はループになっている。

### 3. 手法

#### 3.1 計算方法

本研究では、MDシミュレーション・ソフトウェアパッケージ GROMACS(Groningen Machine for Chemical Simulations)を用いて計算を行った。蛋白質の構造を $C_{\alpha}$ モデルへ粗視化するソフトウェアには SMOG<sup>6)</sup>を用いた。

シミュレーションの設定として時間の刻み幅 $\Delta t$ は 0.0005 ps、時間ステップは $10^7$ ステップ、全シミュレーション時間は 500 ps、ポテンシャルカットオフは 3.0 nm、シミュレーションボックスは 10.6 nm にて行った。

また、温度、圧力一定のもとでシミュレーションを行い、圧力は 1 気圧、温度は、50 K から 200 K まで、5 K ごとでそれぞれシミュレーションを行った。ただし、フォールディング温度の前後 5 K は、1 K 刻みでシミュレーションを行った。

結果の解析は、蛋白質の残基を表すビーズのトラジェクトリー、及び、1000 時間ステップごとに書き出された運動エネルギーやポテンシャルエネルギーについての計算結果のデータを用いて行った。

結果の解析には WHAM (Weighted Histogram Analysis Method)<sup>(6,7,8)</sup>を用いた。この手法を用いることにより、各温度におけるシミュレーション結果をつなげ、種々の熱力学量を温度の関数として表すことができる。また、自由エネルギー地形をコンタクト  $Q$  (フォールディングを行っている経路において、天然状態においてコンタクトをしている残基対のうち、コンタクトがおこっている割合) や RMSD(Root Mean Square Deviation)の関数として表すことも

可能になる。

#### 3.2 WHAM について

一般に、蛋白質のような多自由度系において、エネルギー極小状態(準安定状態)が無数に存在し、このような極小状態を経てエネルギー最小状態(安定状態)になるまでに、要する時間は長い。この場合、シミュレーションでは、計算資源の上で、十分に長い時間が取れず、特に準安定状態のところ状態が落ち込んでしまった場合などは、すべての状態が出現しないことがありうる。このような場合、算出された熱力学量は、統計が不十分で、信頼できないものになってしまう。

このような問題を克服するため、シミュレーションにおいて系が様々なトラジェクトリーをめぐるように、バイアスとなるポテンシャルのセットを用意し、それぞれのセットについて、シミュレーションを行う手法が用いられている。系の自由エネルギーのような熱力学量を考える際、得られたトラジェクトリーから、バイアスの寄与を取り除いて再構成し、任意の状態における熱力学量を算出する手法として用いられているのが WHAM である。

具体的には、バイアスポテンシャルや温度などをセットしたそれぞれの場合について、シミュレーションを実行する。そして、それぞれの反応座標 $\xi$ や系のエネルギー $E$ について度数分布を計算し、そこから、系の状態密度 $\Omega_k(E, \xi)$  ( $k$ は $k$ 番目のセットを意味する)を算出する。本来、状態密度はサンプルの条件によらないはずであるが、実際には、サンプルごとにばらつきがあるので、以下の方法で、 $\Omega_k(E, \xi)$ を最適化する。

まず、実際求めたい $\Omega(E, \xi)$ は、 $\Omega(E, \xi) = \sum_k \omega_k(E, \xi) \Omega_k(E, \xi)$ のように重み因子 $\omega_k(E, \xi)$ を用いて表せるものとする。重み因子の寄与の足し合わせが1となる条件: $\sum_k \omega_k(E, \xi) = 1$ のもとで、 $\Omega(E, \xi)$ の分散が最小になるようにしたとき、確率密度 $P(E, \xi) = \Omega(E, \xi) \exp(-\beta E - \beta V(\xi))$ に対する以下の関係式を得ることができる:

$$P(E, \xi) = \frac{\sum_{k=1}^R g_k^{-1} N_k(E, \xi) \exp(-\beta E - \beta V(\xi))}{\sum_{m=1}^R n_m g_m^{-1} \exp(f_m - \beta_m E - \beta_m V_m(\xi))} \exp(-f_j) = \sum_{E, \xi} P_j(E, \xi)$$

ここで、 $\beta = 1/(k_B T)$  ( $k_B$ はボルツマン定数、 $T$ は系の温度)、 $f = \beta F$  ( $F$ は自由エネルギー)、 $n$ はシミュレーションの総ステップ数、 $N(E, \xi)$ はシミュレーション中に $(E, \xi)$ で指定される状態が現れた回数、 $V(\xi)$ はバイアスポテンシャルを表す。 $\beta_k$ 、 $V_k$ における添え字 $k$ は、 $k$ 番目のセットが設定条件( $\beta_k$ ,  $V_k$ )で与えられていることを表し、 $g_k$ ,  $N_k$ ,  $n_k$ ,  $f_k$ に表れている添え字 $k$ は、 $k$ 番目のセットの下でシミュレーションをした時の $g$ ,  $N$ ,  $n$ ,  $f$ の値を意味する。また、 $g_k$ は $g_k = 1 + 2\tau_k$ で与えられ、 $\tau_k$ はシミュレーションの相関時間を表している。

上記の2つの関係式を自己無撞着的に解くことにより、確率密度 $P(E, \xi)$ を求めることができ、それを用いることによ

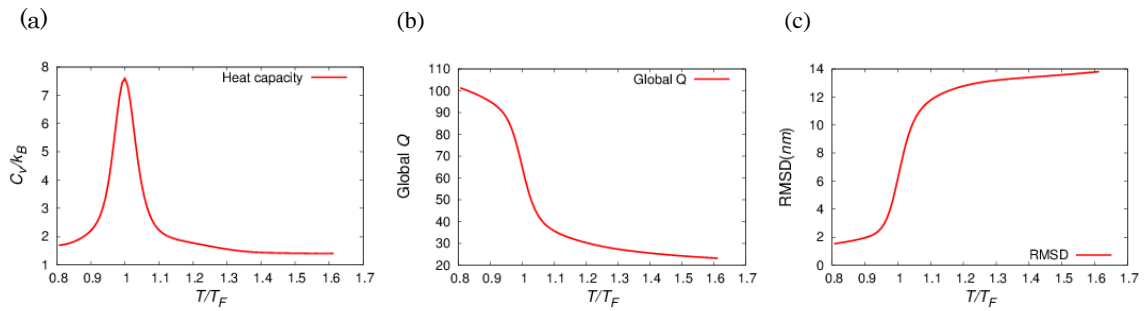


図3 物理量の温度依存性。

(a) 比熱  $C_v$ , (b) コンタクトをしている割合  $Q$ , (c) RMSD (Root Mean Square Deviation)。

比熱のピーク温度をフォールディング温度  $T_F$  とし、横軸は温度  $T$  を  $T_F$  でスケールした値になっている。

って、自由エネルギー  $F$  や比熱  $C_v$  などの熱力学量も得ることができる。

本研究では、郷モデルによりあらかじめトラジェクトリーの選定を行っているため、バイアスポテンシャルを導入することはないが、温度はいくつか選び、それぞれの場合について設定された温度の下でシミュレーションを行っており、それぞれの結果の間をつないで熱力学量についての温度依存性を調べる際に WHAM を用いている。

また、オーダー・パラメータとしてコンタクトしている割合  $Q$  を使い、 $Q$  とエネルギー  $E$  についての度数分布に対して、WHAM のアルゴリズムを用いることにより、確率分布  $P(E, Q)$  を計算している。さらに、得られた  $P(E, Q)$  より、自由エネルギー地形  $F(Q)$  について計算を行っている。

#### 4. 計算結果

図 3(a) は比熱の温度変化を表している。横軸は温度を表しており、フォールディング温度  $T_F$  (アンフォールドし始める温度) で割ったものを表している。ここで、 $T_F$  は、比熱がピークとなる温度をもって決めている。(b) は天然状態においてコンタクトを形成している残基に対して、どのくらいの割合がコンタクトをしているかを示すオーダー・パラメータ  $Q$  の温度変化を示している。温度  $T_F$  を境に  $Q$  の値は 0 から 1 に上がり、 $T_F$  近辺で折り畳み状態と変性状態間の転移が起こっていることを示唆している。この他、エネルギー、エントロピーの温度依存性も調べたが、いずれも  $T_F$  近辺で低い値から高い値に一気に変化しており、転移が起こっていることを示していた。

図 3(c) は、RMSD (Root Mean Square Deviation) の温度依存性を表す。RMSD は X 線構造解析によって得られた結晶構造 (図 2 参照) からのずれに基づいたものである。RMSD も転移温度  $T_F$  を境に一気に大きくなり、低温 (天然状態) のところでは、揺らぎが小さく、高温 (変性状態) では揺らぎが大きくなるのがわかる。

図 4 は温度  $T=1.20T_F$  におけるコンタクト  $Q$  と RMSD に対

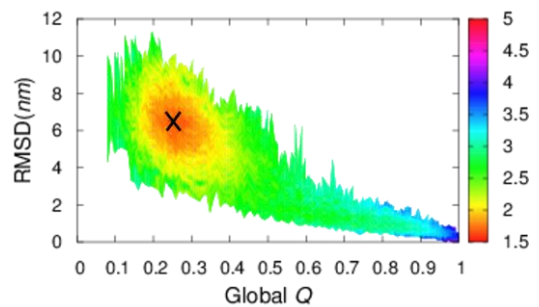


図4 温度  $T=1.20T_F$  におけるコンタクトと RMSD

に対する自由エネルギー地形  $F(Q, \text{RMSD})$  :

横軸はコンタクト  $Q$ 、縦軸は RMSD を表しており、カラーバーは、自由エネルギーの大きさを表している。(  $Q, \text{RMSD}$  ) = (0.25, 6.5) の位置 ( × 印 ) に極小点を持ち、コンタクト数は小さく、RMSD は大きいというアンフォールド状態を示している。

する自由エネルギー地形を表している。横軸はコンタクト  $Q$ 、縦軸は RMSD を表しており、カラーバーによって、自由エネルギーの大きさを表している。このとき、(  $Q, \text{RMSD}$  ) = (0.25, 6.5) の位置に極小点が現れる。この極小点は、コンタクト  $Q$  は小さく、RMSD は大きいという変性状態に対応している。

同様の解析を、 $T=T_F$ 、 $T=0.82T_F$  の温度においても行った。 $T=T_F$  においては (  $Q, \text{RMSD}$  ) = (0.4, 4.2) と (  $Q, \text{RMSD}$  ) = (0.8, 1.0) の 2 点に極小点が現れ、 $T=0.82T_F$  においては (  $Q, \text{RMSD}$  ) = (0.9, 0.4) に極小点が現れた。

$T=T_F$  における 2 つの極小点のうちの 1 つ、(  $Q, \text{RMSD}$  ) = (0.4, 4.2) は、 $T=1.20T_F$  の時と同様に、変性状態に対応しているが、もう 1 つの極小点 (  $Q, \text{RMSD}$  ) = (0.8, 1.0) は、 $Q$  が大きく RMSD が小さいという天然状態に対応している。

また、 $T=0.82T_F$  における極小点は、同様に考えることにより、天然状態に対応していることがわかる。

以上のことから、3 ヘルックスバンドルの転移温度における自由エネルギー地形は、変性状態に対応する極小値と、天然状態に対応する極小値をもち、2 状態遷移であることがわかった。

今後は、フォールディングの経路をさらに詳しく解析することにより、蛋白質のもつ性質や機能を明らかにしていく予定である。

## 参考文献

- (1) J.N. Onuchic, Z. A. Luthey-Schulten, and P. G. Wolynes: "THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective", *Annu. Rev. Phys. Chem.*, 48, pp.545-600(1997)
- (2) N. Gö: "Theoretical Studies of Protein Folding", *Annu. Rev. Biophys. Bioeng.*, 12, pp.183-210(1983).
- (3) Y. Ueda, H. Taketomi, and N. Gö: "Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific interunit", interactions. *Int. J. Pept. Res.* 7, 445-459 (1975).
- (4) C. Clementi, H. Nymeyer, and J. N. Onuchic: "Topological and Energetic Factors: What Determines the Structural Details of Transition State Ensemble and "En-route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins", *J. Mol. Biol.*, 298, pp.937-953 (2000).
- (5) J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford: "SMOG 2: A Versatile Software Package for Generating Structure-Based Models" *PLOS Comp. Bio.* DOI: 10.1371/journal.pcbi.1004794 (2016).
- (6) A. M. Ferrenberg and R. H. Swendsen: "New Monte Carlo technique for studying phase transitions". *Phys. Rev. Lett.*, 61, pp.2635-2638 (1988).
- (7) A. M. Ferrenberg and R. H. Swendsen: "Optimized Monte Carlo data analysis". *Phys. Rev. Lett.*, 63, pp.1195-1198 (1989).
- (8) R. H. Swendsen: "Modern methods of analyzing Monte Carlo computer simulations", *Physica A*, 194, pp.53-62 (1993).