

# タンパク質の折り畳みでの エネルギー地形上の確率流ネットワーク

松下勝義<sup>1,2</sup>, 杉原秀理<sup>1,3</sup>, 菊池誠<sup>1,3,4</sup>, 能川知昭<sup>5</sup>, 佐々木志剛<sup>6</sup>

<sup>1</sup> 大阪大学サイバーメディアセンター, <sup>2</sup> 大阪大学蛋白研, <sup>3</sup> 大阪大学理学研究科,  
<sup>4</sup> 大阪大学生命機能研究科, <sup>5</sup> 東邦大学医学部, <sup>6</sup> 東北大学工学部

## 概要

タンパク質が複雑な天然構造へ折りたまれる機構は生物物理学における長年の謎である。その理解の鍵は折りたたみのエネルギー地形にある。そのため、タンパク質の構造情報からエネルギー地形を再現することは重要な課題である。しかしながら、エネルギー地形をこれまでに蓄積されたタンパク質構造から研究する手法は十分に洗練されていない。我々は折りたたみのエネルギー地形の知見を得る手法として、エネルギー状態ネットワーク上の確率流を見積もるモンテカルロシミュレーションを利用した手法を開発した。そして、この手法をエネルギー地形がファネル型になると想定されるタンパク質に応用し折りたたみのファネル地形に由来する確率流ネットワークを観測した。

## Probability flow on the protein folding energy landscape.

Katsuyoshi Matsushita<sup>1,2</sup>, Hidetoshi Sugihara<sup>1,3</sup>, Macoto Kikuchi<sup>1,3,4</sup>,  
Tomoaki Nogawa<sup>5</sup>, Munetaka Sasaki<sup>6</sup>

<sup>1</sup> Cybermedia Center, <sup>2</sup> Institute for Protein Research,

<sup>3</sup> Graduate School of Science, <sup>4</sup> Graduate School of Frontier Biosciences; Osaka University,

<sup>5</sup> Faculty of Medicine, Toho University, <sup>6</sup> Graduate School of Engineering, Tohoku University

## Abstract

The mechanism of the protein folding into the complex native structure has been a major mystery in biophysics for a long time. The key to understanding the mechanism is in the folding energy landscape on the protein structure space. Thus the reconstruction of the energy landscape from structure informations of protein is one of important issues. However, the method for investigating the energy landscape from the accumulated protein structure data so far is not sufficiently sophisticated. We develop a method using a Monte Carlo simulation to estimate probability flows on energy state networks from protein structures in order to gain insight into their folding energy landscapes. We apply it to a protein for which the folding funnel is expected and observe a probability flow network reflecting the folding funnel.

## 1 はじめに

タンパク質の構造はその生物学的機能を定める [1]. そのためタンパク質の天然構造への一意構造形成, すなわち折りたたみ, は生物物理学における重要な研究

テーマである。構造形成は 80 年代から 90 年代にかけて唱えられた Go の整合性原理 [2] や Bryngelson-Wolynes の競合極小化原理 [3] により大枠の理解がなされた。これらの原理はタンパク質の相互作用の

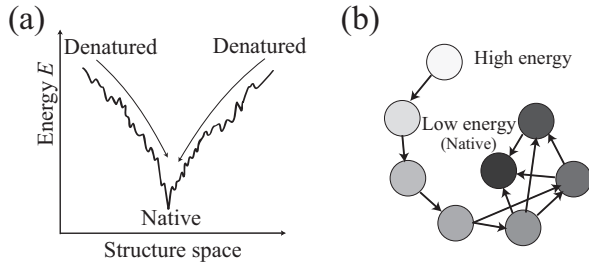


図 1: (a) 折りたたみのファネル地形. (b) エネルギー状態とその間の確率流. ノードはエネルギー状態を表し色が濃いほどエネルギーは低い. 各ノードのエネルギー差は均等であると仮定している. 矢印はエネルギー状態間の確率の流れ.

整合性, もしくは競合のなさがタンパク質折りたたみの原因であることを主張する. そしてその結果として構造空間上のエネルギー地形は図 1(a) にあるようなファネル (漏斗) 状の構造を形成するとされる [4]. ファネル構造の底はタンパク質の天然構造に対応しており, ファネルの形状がその天然構造へのタンパク質の折りたたみを誘導するのである.

この仮説は現在もっとも受け入れられているものであるが, その正しさを確認したという研究はほとんどない. あったとしても格子タンパクで折りたためる配列でそのエネルギー形状を観測したという主張はあるものの, ファネル形状が観測されるのが天然構造の周りの摂動領域のみだけであった [4]. また最近では分子動力学法を用いた計算が行われているものの, ファネル形状が最初から仮定されそのパラメータをフィッティングしたのみでファネルの存在の証とはなっていない [5]. 特にタンパク質の観測データにより蓄積された構造データを生かした研究はなく, 簡単に多くの構造に関してエネルギー地形形状の特徴を抽出する手法開発はファネル地形仮説の確認のための重要課題である.

本研究ではファネル地形確認のための手法開発の第一歩として, ファネル地形を持つと想定されるタンパク質のエネルギー状態が作るネットワーク上での確率流を調べる. もしエネルギー地形がファネル的であるならば, そのエネルギー状態ネットワーク上の確率流は図 1(b) のようにファネルの導く通りに流れるであろう. 高エネルギー状態ではエネルギー地形の傾きがなだらかである. そのため, エネルギーが高い状態からは少しエネルギーの低い状態に構造の存在確率の流れが起きると予想される. 一方, エネルギーの低い状態では強いエネルギー地形の勾配

を反映し, より低い状態への流れが起きるのである. このような確率流はエネルギー状態空間上のマルコフ過程で表現される. 我々はそのマルコフ過程をタンパク質構造データを基にした模型上のモンテカルロシミュレーションから構成する. その結果我々はファネル地形由来の確率流の成すネットワーク構造を得ることができた.

## 2 模型と手法

タンパク質はアミノ酸が一次的にならんだ鎖から成っている. タンパク質の構造形成の過程は天然構造におけるアミノ酸残基の間の結合のみを考慮するだけで理解されることが経験的に知られている [2]. そのような天然構造を基にしたタンパク質模型の一つに Wako-Saitô-Muñoz-Eaton 模型 [6, 7, 8, 9] がある. この模型はアミノ酸残基内の離散的な立体角自由度をスピン自由度  $m_i$  で表す. ここで  $i$  はサイトの番号であり  $m_i$  は 0 か 1 をとる.  $m_i = 1$  は天然構造における立体角状態を表し,  $m_i = 0$  はその他の状態をとっていることを表す.

天然構造におけるアミノ酸残基ペア間の集合  $\langle i, j \rangle$  に対して, 次のように Hamiltonian が定義される.

$$\mathcal{H}(m_i) = -\epsilon \sum_{\langle i, j \rangle} \prod_{k=i}^j m_k. \quad (1)$$

一方非天然構造の集合のエントロピーを表現するため次のエントロピー項を導入する,

$$\mathcal{S}(\{m_i\}) = \sum_i (1 - m_i). \quad (2)$$

そして状態  $\{m_i\}$  に対する平衡分布は  $P(\{m_i\}) \simeq \exp[-\beta\mathcal{H}(\{m_i\}) + \mathcal{S}(\{m_i\})]$  で与えられる.

さて, 本研究ではこの模型に対する確率流を計算する. エネルギー状態間の確率流を与えるためには微視的ダイナミクスを与える必要がある. 平衡分布  $P(\{m_i\})$  を与える自然なダイナミクスの一つは Glauber (熱浴) ダイナミクスである. このダイナミクスでは微視的遷移確率を遷移元の状態  $\{m_i\}$  と遷移先の状態  $\{m'_i\}$  から  $W(\{m'_i\}|\{m_i\}) = P(\{m'_i\}) / (P(\{m_i\}) + P(\{m'_i\}))$  と与え, 遷移先はランダムにサイト  $i$  を選択し,  $m_i$  をフリップさせて生成された状態とする.

この Glauber ダイナミクスは微視的なダイナミクスであり, 我々が計算するエネルギー状態間の確率流を与えるためには微視的状态空間を巨視的なエネル

ギー状態空間に疎視化しなければならない。その際重要になるのは遷移元の巨視的状态 ( $E, S$ ) から遷移先の状態 ( $E', S$ ) への遷移経路の密度  $\rho(E', S'|E, S)$  である。ここで  $E$  と  $S$  は式 (1) と (2) で与えられるエネルギーとエントロピーである。この密度がわかれば、巨視的な遷移確率は

$$W(E'|E) = \sum_{S, S'} \rho(E', S'|E, S) W(E', S'|E, S) \quad (3)$$

の形に与えることができる。ここで  $W(E', S'|E, S)$  は先に述べた Glauber ダイナミクスの微視的遷移確率である。

$\rho(E', S'|E, S)$  は先に示した Glauber ダイナミクスのうち、 $m_i$  をフリップさせるサイトをランダムに選ぶ部分で決まる量である。従ってランダムサンプリングのモンテカルロシミュレーションで遷移のヒストグラム  $h(E', S'|E, S)$  を作り、それを遷移元の状態 ( $E, S$ ) を固定して規格化することで  $\rho(E', S'|E, S)$  を得ることができる。本研究ではマルチカノニカルサンプリングを用いてこれを実行し巨視的遷移確率  $W(E'|E)$  を得る [10]。この遷移確率が計算したいエネルギー状態間の確率流を与える。

今回は構造形成のエネルギー地形が生み出すエネルギー緩和過程の確率流を捕らえる。そのために、すべての巨視的状态  $E$  に対して、それぞれ初期分布  $P_E(E') = \delta_{E, E'}$  を定義し、それらの分布すべての集合からの緩和を考える。この確率流は各エネルギー状態からのエネルギー地形由来のエネルギー緩和の様子を描き出してくれると期待できる。今回は複雑さを避けるため初期分布の集合からの 1 スピンフリップでのエネルギー緩和に対する確率流を計算する。

### 3 結果

ここまで説明した手法をよくベンチマークとして構造が計算される図 2(a) のアルブミン結合 GA モジュールに適用し、タンパク質折りたたみのファネル地形を反映した確率流が得られるかを確かめる。天然構造は三つのヘリックスバンドルからなる。現状のところ他に安定な構造は知られていないためこの構造を安定化する相互作用のみを考慮する。今回はこの構造での残基間結合を残基内の  $C\alpha$  原子間距離が  $6.5\text{\AA}$  以下の場合に成立しているとみなし、WSME 模型を構成した。その結果得られた  $W(E'|E)$  を元に図 2(b) にエネルギー状態間の確率流を表すネットワークを示した。ここで温度は生理温度における天

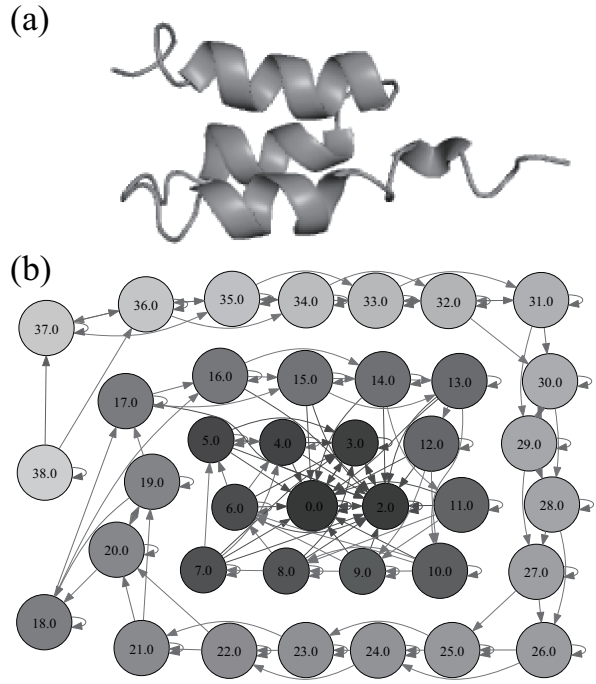


図 2: (a) アルブミン結合 GA モジュールの天然構造 (PDB:1prb). (b) アルブミン結合 GA モジュールの天然構造を元に得たエネルギー状態間の確率流. ノードはエネルギー状態を表し色が濃いほどエネルギー  $E$  が低く、数字は  $E/\epsilon$  を表す. 矢印はエネルギー状態間の確率の流れであり、色が濃いほど確率流が大きい. 見易さのため遷移確率が小さいエッジは表示していない (本文参照). ノードの空間配置には意味は無いがエッジで繋がるノードは近くに位置する.

然構造と変性状態の双安定状態を再現できる折りたたみ温度  $T_f = 0.64\epsilon$  に設定した。

各ノードはエネルギー状態を表し、色の濃さがエネルギーの低さに対応している。ノード上に与えられた数値は  $E/\epsilon$  である。エネルギー状態の間の矢印は  $W(E'|E)$  から自己ループを除いて再規格化した遷移確率をもとに遷移確率が高ければ濃い矢印、遷移確率が低ければ薄い矢印で遷移を表した。矢印の数が多くなりネットワーク構造の本質を得るため 20%/1step 以上の再規格化した遷移確率を持つ場合のみを図には示してある。自己ループは同じエネルギー状態への遷移かもしれないが遷移そのものがおきなかった確率である。この自己ループに関しては他の遷移とは違い、そのまま  $W(E|E)$  の値を元に矢印の色を決めている。この場合も濃いほど自己遷移確率が高く、20%/1step 以上の場合のみ矢印を書いている。

図 1(b) と良く似ているエネルギー状態からすぐ下

のエネルギー状態への遷移が高エネルギー状態で見られる。また、低エネルギー状態になるとより大きなエネルギー減少に対応する矢印が観測される。アルブミン結合 GA モジュールの場合、タンパク質のエネルギーは天然構造と相関が強く、エネルギーはそのまま天然構造の秩序変数とみなすことができる。従って、ここで述べたネットワーク構造の特徴が図 1(a)に見られるようにファネル地形が構造が天然構造に近づくほどエネルギー変化が急になっていることを反映していると考えられる。

## 4 結び

本研究ではタンパク質のエネルギー地形の情報を得るためにエネルギー状態の成すネットワーク上で確率流を求める手法を考案した。そして、それをアルブミン結合 GA モジュールに適用して、実際ファネル地形に由来するエネルギー状態間の確率流ネットワークの振る舞いを確認した。この手法による計算は分子動力学シミュレーションによる大規模計算より低コストに実行でき、今後、タンパク質構造データベースからのエネルギー地形再構成によるファネル仮説検証の有効な手段を与えると期待される。

まだこの手法は開発されたばかりで今後の課題は山積している。まず、この手法を用いたより定量的な解析手法が確立できていない。エネルギー地形の性質をさらに深く知るための物理量、もしくは図 2(b)に示されたネットワーク構造の特徴量を見つけることが今後の課題として残されている。さらに、ファネル地形にならないと予測される“相互作用の競合に由来する高エネルギー構造”を考慮しなければならない場合 [11, 12, 13] やタンパク質間の結合のエネルギー地形にも応用し [14]、非自明なエネルギー地形の情報を得るための手法としてのこの方法の有効性を確認する必要もあるだろう。

## 参考文献

- [1] E. Fischer, Ber. Dt. Chem. Ges. **27** (1894) 2985 .
- [2] N. Gō, Annu. Rev. Biophys. Bioeng. **12** (1983) 183.
- [3] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84** (1987) 7524.

- [4] P. E. Leopold *et al.*, Proc. Natl. Acad. Sci. USA **89** (1992) 872.
- [5] J. Wang *et al.*, Proc. Natl. Acad. Sci. USA **109** (2012) 15763.
- [6] H. Wako and N. Saitō, J. Phys. Soc. Jpn. **44** (1978) 1931.
- [7] H. Wako and N. Saitō, J. Phys. Soc. Jpn. **44** (1978) 1939.
- [8] V. Muñoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. USA **95** (1998), 5872.
- [9] V. Muñoz and W. A. Eaton, Proc. Natl. Acad. Sci. USA **96** (1999) 11311.
- [10] T. Nogawa *et al.*, Phys. Rev. E **86** (2012) 41133.
- [11] K. Matsushita and M. Kikuchi, J. Chem. Phys. **138**, (2013) 105101.
- [12] K. Matsushita and M. Kikuchi, Interdiscip. Info. Sci. **19** (2013) 29.
- [13] N. C. Shirai and M. Kikuchi, arXiv:1310.0889.
- [14] K. Matsushita and M. Kikuchi, J. Phys.: Conf. Series **454** (2013) 012034.