

# タンパク質相互作用面におけるアミノ酸残基の絡まり度と保存度との関係

中川 真理子(指導教員:由良 敬)

## 1 はじめに

生物は細胞を基本単位として構成されている。細胞内には、分子の集合体である大小の様々な細胞小器官が存在する。これら細胞小器官を構成する分子の多くは、タンパク質である。タンパク質は、化学反応の触媒や小型の分子やイオンの運搬など、様々なはたらきをしている。タンパク質の化学的な実体は、20種類のアミノ酸が枝分かれすることなく1本の鎖状につながった高分子である。それぞれのアミノ酸は隣接するアミノ酸とペプチド結合でつながっている。ペプチド結合は、1つのアミノ酸のカルボキシル基と、もう1つのアミノ酸のアミノ基から水1分子に相当する原子が脱離して形成される。アミノ酸から水1分子に相当する原子がとれた構造はアミノ酸残基とよばれる。タンパク質は多くのペプチド結合できあがっているため、ポリペプチド鎖とよぶことも出来る。タンパク質はその種類によってアミノ酸の並びが決まっており、それをアミノ酸配列とよぶ。

異なる生物では、類似のアミノ酸配列をもつタンパク質が、同じ機能を担っていることがほとんどである。これは共通祖先がもっていたタンパク質のアミノ酸配列が、生物進化の過程でアミノ酸残基の置換や挿入、欠失を伴って変化した結果と考えられている。タンパク質の機能にさほど影響をおよぼさない変化は、次世代に受け継がれていくが、タンパク質の機能に多大な影響をおよぼすようなアミノ酸配列の変化が起こってしまうと、そのようなタンパク質をもつ個体は生きることができないため、次の世代に受け継がれることはない。したがって、アミノ酸配列のうち、タンパク質の機能にとって重要な部分は進化の過程で保存される。

タンパク質は、他のタンパク質と相互作用することで、細胞小器官を構築し、細胞内で様々なはたらきを実現している。これらのタンパク質は、多種多様な分子で混雑した細胞内でも、ごく短い時間のうちに特異的な相互作用相手を見出し相互作用するに至るが、その機構は明らかにされていない。複数のタンパク質が相互作用して構築されるタンパク質複合体の結晶構造は、X線回析の手法で解析することにより明らかにされており、そのデータは Protein Data Bank (PDB) [1] に登録されている。そこで、できあがった構造のデータから、タンパク質が高速でかつ特異的に複合体を形成するために必要な特徴を見出すことができるかを検討した。この結果に基づき、細胞内でタンパク質がどのようにして相互作用するに至るのかを推定することを試みた。

我々の研究室では、タンパク質多量体の相互作用面に、両タンパク質が比較的平らな構造を提供して相互作用を成立させている場合と、お互いのポリペ

プチド鎖が絡まり合ったように見える構造(絡まり構造)を形成している場合の2通りがあることを報告した。そして、これら2種類の相互作用面をタンパク質複合体構造から自動的に同定する手法(2.1参照)を開発した。タンパク質の相互作用面の絡まり構造部位は、タンパク質が単体で存在する際には、定まった構造を取らない天然変性領域である傾向があることも見出した。このことは、タンパク質が相互作用に至る過程において、絡まり構造部位が溶媒中でふらふらしながら相互作用相手を探す、フライキャスティング現象[2,3]が起きていることを示唆する。我々は、絡まり構造がこの現象の痕跡であると推測した[4]。

本研究では、この絡まり構造が、進化的にどのようなになっているのかを検討した。

## 2 手法

### 2.1 絡まり度(Surroundedness Factor)の計算

相互作用面の絡まり構造部位では、ある鎖に属するアミノ酸残基が他の鎖に突っ込んでいる/絡まっている。突っ込んでいる/絡まっているアミノ酸残基は、他の鎖のアミノ酸残基に囲まれている。どのアミノ酸残基がどの程度他の鎖に突っ込んでいるか/絡まっているか、どの程度他の鎖に属するアミノ酸に囲まれているかは、以下のように計算できる。相互作用面におけるアミノ酸配列の*i*番目の残基の  $C\alpha$  原子(アミノ基とカルボキシル基の両方が結合している炭素原子)から、固定長半径 $R\text{\AA}$ 以内に位置する水素原子以外(X線結晶構造解析では、散乱断面積が小さい水素原子は見えない)の原子で、同じポリペプチド鎖に属する原子の数を $N_s$ とし、異なるポリペプチド鎖に属する原子の数を $N_d$ とする。*i*番目のアミノ酸残基とその前後*L*個のアミノ酸残基について、 $N_d/N_s$ を計算し、その値のアミノ酸配列に沿った平均値を*i*番目のアミノ酸残基の Surroundedness Factor (SF) と定義する(式(1))。

$$SF(R, i) = \frac{1}{2L+1} \sum_{k=-L}^L \frac{N_{d_{i+k}}}{N_{s_{i+k}}} \quad (1)$$

本研究では、先行研究にならって、 $R = 12$ 、 $L = 2$ を用い、SFの値が1.25以上ならば、そのアミノ酸残基は絡まっているアミノ酸残基であるとした。

### 2.2 保存度の計算

共通祖先由来のタンパク質群のアミノ酸配列をアラインメントし、各位置のアミノ酸残基の保存度を測定する。指標として「アミノ酸保存度」と「ギャップの割合」を用いた。

アラインメントの位置*x*におけるアミノ酸保存度  $AAS_x$  は以下のように定義する。まず、位置*x*に現れ

るアミノ酸残基種*i*の出現頻度 $f_{x,i}$

$$f_{x,i} = \frac{n_{x,i}}{\sum_{j=1}^{20} n_{x,j}} \quad (i = 1, \dots, 20) \quad (2)$$

を求める。ただし、 $n_{x,i}$ は位置*x*に現れるアミノ酸残基種*i*の個数。これを全てのアミノ酸残基種に対して計算し、シャノンエントロピー[5]を用いて $AASC_x$ を以下のように求める。

$$AASC_x = 4.32 - (-\sum_{i=1}^{20} f_{x,i} \log f_{x,i}) \quad (3)$$

ただし、 $4.32 \approx \log 20$ であり、これはアミノ酸残基種の出現に関するシャノンエントロピーの最大値である。したがって、 $AASC_x$ の値が高いほど保存されていることを意味する。

アラインメントの位置*x*におけるギャップの割合は、

$$Gap\_Frequency_x = \frac{n_{x,gap}}{N} \quad (4)$$

と定義する。ただし、 $n_{x,gap}$ はアラインメントの位置*x*におけるギャップの個数で、 $N$ はアラインメントされたアミノ酸配列の本数を表す。

### 3 結果と考察

772個のタンパク質立体構造のデータにおいて、全アミノ酸残基の絡まり度、アミノ酸保存度、ギャップの割合を計算した。

絡まっている ( $SF \geq 1.25$ ) アミノ酸残基のアミノ酸保存度の分布と絡まっていない ( $0 < SF < 1.25$ ) アミノ酸残基のアミノ酸保存度の分布を比較した。図1は、絡まっているアミノ酸残基のヒストグラム①の各棒の縦軸の値から、絡まっていないアミノ酸残基のヒストグラム②の各棒の縦軸の値を引いた差の棒グラフである。各棒で示される差について、10,000回のリサンプリングによるブートストラップ検定を行った結果、絡まっているアミノ酸残基は全体に比べてアミノ酸保存度が低い傾向にあることが分かった。

また、絡まっているアミノ酸残基のギャップあり・なしの分布と、絡まっていないアミノ酸残基のギャップあり・なしの分布を比較した。図2は、アミノ酸保存度の場合と同様にして両分布の差をとった結果を示す棒グラフである。10,000回のリサンプリングによるブートストラップ検定を行った結果、絡まっているアミノ酸残基は全体に比べてギャップが存在する傾向にあることが分かった。

### 4 考察と今後の課題

タンパク質の相互作用面において、絡まっている部分はそうでない部分と比べて保存度が低いことが分かった。このことは、タンパク質の進化において、機能に重要な部位が保存されることに反するよう見える。一般に、天然変性領域は保存度が低いので[6]、このことが天然変性領域の傾向がある絡まり構造部位の保存度に影響しているのではないかと考えられる。天然変性領域では、短い領域が重要な機能を担うと考えられており[7]、絡まり構造の中にもこのような短い領域がある可能性がある。その場合、短い

領域が相互作用相手と相互作用に至るための引き金になっている可能性もある。このことを明らかにするには、絡まり構造の中に保存度が高い領域を見出し、分子動力学計算によって、その部分が相互作用に至る過程をシミュレーションする必要がある。

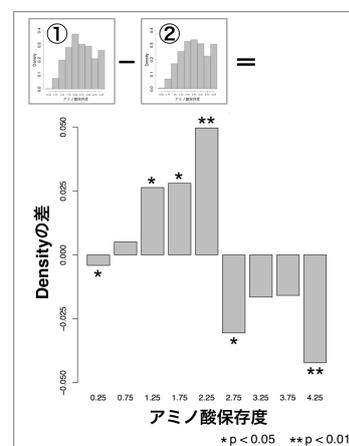


図1: アミノ酸の保存度についての比較結果。①は絡まっているアミノ酸残基について、横軸にアミノ酸保存度、縦軸に Densityをとったヒストグラム。②は絡まっていないアミノ酸残基について、①と同じように描いたヒストグラム。

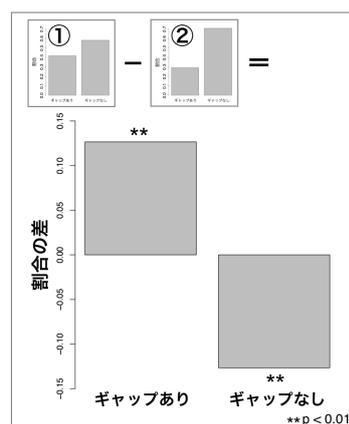


図2: ギャップの割合についての比較結果。①は絡まっているアミノ酸残基において、「ギャップあり」の割合と「ギャップなし」の割合を描いた棒グラフ。②は絡まっていないアミノ酸残基において、①と同じように描いた棒グラフ。

### 参考文献

- [1] The RCSB Protein Data Bank, <http://www.rcsb.org/pdb/home/home.do>
- [2] Levy, Y. *et al.* (2005) A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.*, **346**, 1121-1145.
- [3] Shoemaker, B. A. *et al.* (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc. Natl. Acad. Sci. USA*, **97**, 8868-8873.
- [4] Yura, K. and Hayward, S. (2009) The interwinding nature of protein-protein interfaces and its implication for protein complex formation. *Bioinformatics*, **25**(23), 3108-3113.
- [5] Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27**(3), 379-423.
- [6] Brown, C. J. *et al.* (2011) Evolution and disorder. *Curr. Opin. Struct. Biol.*, **21**(3), 441-446.
- [7] Francesca, D. *et al.* (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* **13**, 6580-6603.