

グロビンフォールドタンパク質における 二次構造ブレイカとしての電荷反転対称性

今井 賢一郎 美宅 成樹*

Secondary Structure Breaker and Charge Symmetry
in Globin Fold Proteins

Kenichiro IMAI and Shigeki MITAKU*

Secondary structure breakers impose strong limitations on the global structure of a protein. Four kinds of secondary structure breakers (proline, glycine, small polar residues and amphiphilic residues) were revealed previously (Imai & Mitaku, 2005, 2007). In addition, the inverse charge symmetry, which drives the formation of hairpin structures, was studied in globin fold proteins, typical all- α type proteins. The inverse charge symmetries were positioned at about two thirds of loops region that do not contain the predicted breakers. Therefore, the charge symmetry is the most plausible mechanism for the fifth type of breaker. The distribution pattern of secondary structure breakers in the remote-homologous globin fold proteins showed that the same fold are realized breaking the secondary structures by various patterns of the breakers.

1. はじめに

タンパク質科学の大きな目標の一つは、タンパク質立体構造形成のプロセスを物理化学的に理解し、コンピュータ上 (*in silico*) においてアミノ酸配列情報から立体構造を再現することである。実際にアミノ酸配列には、タンパク質の立体構造を形成するための情報が含まれているが、タンパク質の立体構造形成プロセスを簡単な物理化学的ルールによって理解できるかどうかは、まだはっきりしていない。もっとも手軽な問題としてタンパク質の二次構造の予測がこれまで精力的に行われてきたが、半世紀を経てその精度は80%を超えることはできていない¹⁾。アミノ酸配列情報のみからタンパク質の立体構造を予測することは非常に難しいのが現状なのである。

我々はアミノ酸配列に含まれるタンパク質の構造の特徴として、秩序構造である二次構造自体ではなく、無秩序構造である二次構造ブレイカに注目した。二次構造ブレイカとは、二次構造を壊し、ループ構造に変えるものであり、二次構造をつなぎ三次構造を形成するうえで重要な構造的特徴となっている。以前より、アミノ酸の出現傾向の解析から、二次構造 (α -ヘリックス, β ストランド) を壊す傾向の強い残基が知られている。^{2), 3), 4), 5)}

ただ、高精度の二次構造ブレイカ予測は行われておらず、我々は二次構造の種類によらず二次構造を壊す傾向の強い4つのタイプ (プロリン, グリシン, 小さい極性残基, 両親媒性残基) のアミノ酸残基クラスタに注目し、予測システムを開発した^{6), 7)}。これらの残基の周辺のアミノ酸配列の物性分布を解析することで、これらの残基が二次構造ブレイカとして働く条件を明確にすることができた。二次構造ブレイカ予測システム SOSUibreaker では、予測した二次構造ブレイカの93%がループ領域に位置しており、ループ領域の80%をカバーしていた。しかし、残りの20%のループ領域については、二次構造ブレイカを予測できておらず、4つのタイプの二次構造ブレイカの他に新たな二次構造ブレイカのみカニズムが存在すると考えられる。

以前、我々は、グロビンフォールドタンパク質において、アミノ酸配列上の電荷の反転対称性 (隣り合う正電荷クラスと負電荷クラス) とヘアピン構造の関係性を調べ、正電荷クラスと負電荷クラスによる静電的な引力がヘアピン構造形成因子であることを指摘した⁸⁾。ヘアピン構造をはじめとする構造の折れ曲がりの形成は、タンパク質の構造に強い制限をもたらすものであり、二次構造の規則性を壊し、これまで明らかにしたものと異なる新しいタイプの二次構造ブレイカのみカニズムとなる可能性がある。本研究では、典型的な球状タンパク質であるグロビンフォールド (all- α 型) に対する二次構造ブレイカ予測と電荷の反転対称性解析を行い、5番

2007年03月14日 受理

*名古屋大学大学院工学研究科マテリアル理工学専攻

目の二次構造ブレイカとしての可能性を検討した。さらに、リモートホモログ（アミノ酸配列の相同性が低いにも関わらず、その立体構造がよく似ている）の関係にあるグロビンフォールドタンパク質において、電荷の反転対称性も含めて、5つのタイプの二次構造ブレイカの出現傾向の比較を行った。二次構造ブレイカは、二次構造をつなぎ三次構造を形成するものとして重要な構造的特徴である。リモートホモログの関係にあるタンパク質間において、出現する二次構造ブレイカを比較することにより、タンパク質の立体構造形成メカニズムについて論議した。

2. 実験方法

2.1 データセット

本研究では、18個のグロビンフォールドタンパク質をデータセットとして用いた。これらは、立体構造分類データベースであるSCOP⁹⁾においてグロビンフォールドに分類されるもので、配列相同性が30%以下のものである。

2.2 二次構造ブレイカ予測

二次構造ブレイカ予測は、大きく二つのステップに分けて行う。(1) 二次構造ブレイカ候補の列挙、(2) 二次構造ブレイカ周辺の配列環境による二次構造ブレイカの判別である。

2.2.1 二次構造ブレイカ候補の列挙

二次構造ブレイカ候補として注目するのは、二次構造の種類によらず二次構造を壊す傾向の強いプロリン、グリシン、側鎖の小さい極性残基、両親媒残基の4つのクラスタである。アミノ酸配列上におけるそれぞれの残基クラスタの位置は、プロリン、グリシン、側鎖の小さい極性残基のアミノ酸配列における位置と両親媒性指標(A index)¹⁰⁾を用いて算出した⁶⁾。プロリンに関しては、そのほとんどが二次構造ブレイカとなっており、判別の必要性がない。しかし、他のものについては、二次構造中にも存在するものもあり、二次構造ブレイカとなるかどうかの判別が必要である。

2.2.2 周辺の配列環境による二次構造ブレイカ判別

二次構造ブレイカの判別に用いるアミノ酸配列の物性パラメータは、二次構造ブレイカ候補周辺のアミノ酸配列の疎水性 $\langle H(j) \rangle$ 、ヘリックス周期性スコア $\langle HPS(j) \rangle$ 、トリプトファン・チロシンの両親媒性指標¹⁰⁾ (両親媒性指標 A' index) $\langle A'(j) \rangle$ 、セリン・スレオニン密度 $\langle ST(j) \rangle$ 、グリシン密度 $\langle G(j) \rangle$ である。 $\langle H(j) \rangle$ は疎水性指標 (KD 指標)¹¹⁾をもとに、 $\langle A'(j) \rangle$ は A' index¹⁰⁾をもとに、 $\langle ST(j) \rangle$ 、 $\langle G(j) \rangle$ は、それぞれの残基のアミノ酸配列における存在をもとに式(1)に従ってj番目の残基の周辺7残基の平均を求めたものである。また、ヘリックス周期

性スコア $\langle HPS(j) \rangle$ は、疎水性指標 (KD 指標)¹¹⁾を用いた以下の式(2)で定義される。

$$\langle X(j) \rangle = \left[\sum_{i=j-3}^{j+3} X(i) \right] / 7 \quad (1)$$

$$\langle HPS(j) \rangle = \max \{ |HP(j)|, |HP(j-1)| \} \quad (2)$$

$$\langle HP(j) \rangle = \left[\frac{H(j+5) - H(j+3) + H(j+1)}{-H(j) + H(j-2) - H(j-4)} \right] / 5$$

二次構造ブレイカとなる残基クラスタとならない残基クラスタの間ではこれら5つのアミノ酸配列の物性パラメータに差がみられる。そこで、周辺のアミノ酸配列の物性の差から環境因子 $\Delta \langle H(I) \rangle$ 、 $\Delta \langle HPS(I) \rangle$ 、 $\Delta \langle A'(I) \rangle$ 、 $\Delta \langle ST(I) \rangle$ 、 $\Delta \langle G(I) \rangle$ を定義し、二次構造ブレイカの判別解析を行い、判別式を得た^{6),7)}。それぞれの判別に有効な環境因子に応じて、グリシンの判別には式(3)を、側鎖の小さい極性残基クラスタの判別には式(4)を、両親媒性残基クラスタの判別には式(5)を用いる。Iは、残基クラスタの位置を表す。

$$Gscore(I) = 2.01\Delta \langle H(I) \rangle + 1.90\Delta \langle HPS(I) \rangle + 5.30\Delta \langle A'(I) \rangle + 32.90\Delta \langle ST(I) \rangle - 0.68 \quad (3)$$

$$SPscore(I) = 2.43\Delta \langle H(I) \rangle + 1.31\Delta \langle HPS(I) \rangle + 1.03\Delta \langle A'(I) \rangle + 178.06\Delta \langle G(I) \rangle \quad (4)$$

$$Ascore(I) = 3.15\Delta \langle H(I) \rangle + 1.36\Delta \langle HPS(I) \rangle + 2.40\Delta \langle A'(I) \rangle + 42.48\Delta \langle ST(I) \rangle - 1.71 \quad (5)$$

予測された二次構造ブレイカは、予測した二次構造ブレイカの93%がループ領域に位置しており、また、ループ領域の80%をカバーしていた。Fig. 1は、これらの予測精度を模式的に表したものである。

70% of predicted breakers (all four types) were actually located in the break region, and 93% were located in the loop region.

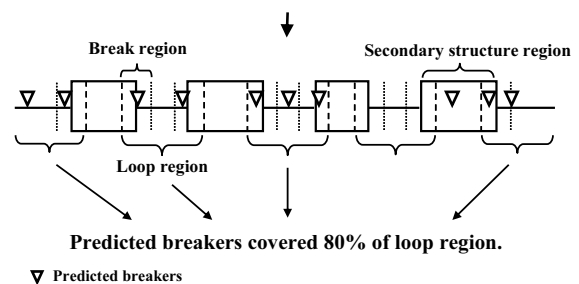


Fig. 1. The diagram illustrating the accuracy of prediction of secondary structure breakers.

2.3 電荷の反転対称性点の解析

符号の異なる電荷は互いに引き合う傾向があるため、アミノ酸配列上において隣り合う正電荷と負電荷クラスターは、ヘアピン構造などの構造の折れ曲がりの原因となる。さらに、構造の折れ曲がりは、タンパク質の構造形成に強い制限を与えるものであり、新しいタイプの二次構造ブレイカとなると考えられる。構造の折れ曲がりの原因となる正電荷と負電荷クラスターの境界は、アミノ酸配列におけるセグメントの電荷の性質の反転が見られるポイント（電荷の反転対称点）として表される。そこで、電荷の反転対称点の解析を行うため、電荷配列の7残基の移動平均である電荷密度 $\langle C(j) \rangle$ の計算を行った。

$$\langle C(j) \rangle = \left[\sum_{i=j-3}^{j+3} C(i) \right] / 7 \quad (6)$$

$C(j)$ の値は、正電荷を持つリジン、アルギニン、ヒスチジン、を +1, 負電荷を持つアスパラギン酸とグルタミン酸を -1 とし、それ以外の残基は 0 とした。アミノ酸配列上の位置 j における電荷の反転対称性の評価は、位置 j を中心に対称的な位置にある電荷値 $\langle C(j-k) \rangle$ と $\langle C(j+k) \rangle$ の差より求めた。もし、位置 j を中心に完全な電荷の反転対称性を示す場合は、 $\langle C(j+k) \rangle$ と $\langle C(j-k) \rangle$ は、符号が異なるがその絶対値は同じ値を示す。このため、 $\langle C(j+k) \rangle$ と $\langle C(j-k) \rangle$ との差を 2 で割った値 $D(j,k)$ は、完全な電荷の反転対称性を示す場合、 $D(j,k) = \langle C(j+k) \rangle$, $D(j,k) = -\langle C(j-k) \rangle$ となる。そこで、 $S(j,m)$ を求め、電荷の対称性を評価した。

$$D(j,k) = \frac{1}{2} (\langle C(j+k) \rangle - \langle C(j-k) \rangle) \quad (7)$$

$$S(j,m) = \frac{1}{2m} \sum_{k=1}^m \left\{ \begin{aligned} & (\langle C(j+k) \rangle - D(j,k))^2 \\ & + (\langle C(j-k) \rangle + D(j,k))^2 \end{aligned} \right\} \quad (8)$$

式中の $2m$ は、対称性を評価するセグメントの長さ (j 番目のアミノ酸を中心とした両脇 m 残基のセグメント) である。電荷の対称性は長いセグメントだけでなく、短いセグメントにも見られるため、評価するセグメントの長さ m を 5 から 15 残基まで変化させ、その最大値を Charge symmetry score (CS score) とした。電荷の反転対称点は、CS score の極小値として表される。

$$\langle CSscore(j) \rangle = \max \{ S(j,m) \}_{m=5-15} \quad (9)$$

3. グロビンフォールドタンパク質における二次構造ブレイカと電荷の反転対称性

3.1 グロビンフォールドタンパク質の二次構造ブレイカ予測と電荷の反転対称性解析

我々は、プロリン、グリシン、側鎖の小さい極性残基、両親媒性残基の 4 つのタイプの二次構造ブレイカ残基に注目し、二次構造ブレイカ予測システムを開発した。二次構造ブレイカ予測を豚回虫由来のヘモグロビンに対して行った結果が Fig. 3 (a) である。図中の上にあるカラーバーは、図中の右にある立体構造に対応した 8 本のヘリックスを、A~E は、ヘリックスを結ぶループ領域を表している。図中のピークは、それぞれ、予測したプロリン、グリシン、側鎖の小さい極性残基、両親媒性残基のクラスターによる二次構造ブレイカの位置を示し、ピークの大さは予測スコアを表している。予測された二次構

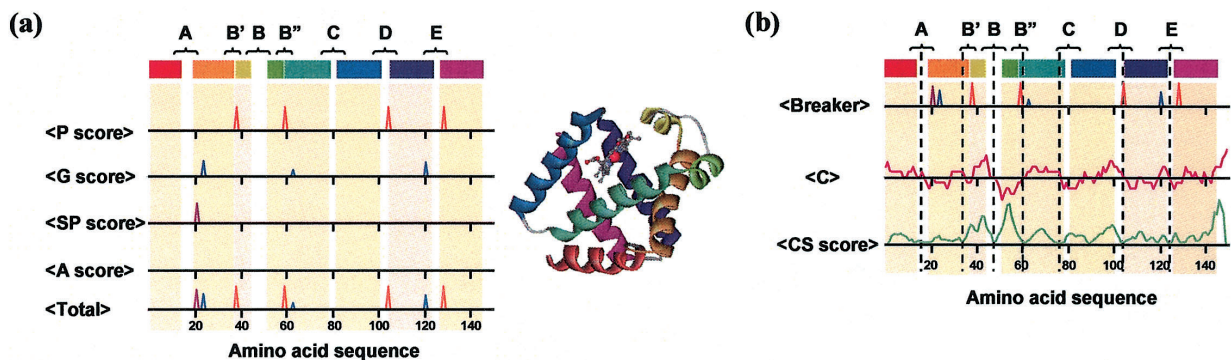


Fig. 2. Results of secondary structure breaker prediction and structure (a) and plots of the charge density and the CS score (b) for pig roundworm hemoglobin (PDB ID: 1ash).

造ブレイカは、二次構造の端付近に現れ、二次構造のブレイカとなっている。しかしながら、ループ領域 B, C は、予測された二次構造ブレイカが周辺に見られない。これらのループ領域には、4 つのタイプの二次構造ブレイカとは異なる二次構造ブレイカのみが存在すると考えられる。

新しい二次構造ブレイカのみをメカニズムとして、我々は、正電荷と負電荷クラスタによる構造の折れ曲がり注目した。Fig. 2 (b) は、豚回虫由来のヘモグロビンの電荷密度 (CD) プロットと CS score である。グロビンフォールドタンパク質では、8 本のヘリックスがループ A~E で折れ曲がることで、球状の立体構造を形成している。CD プロットをみると、折れ曲がりの見られるループ A~E では、正電荷のピークと負電荷のピークが隣り合っており、電荷の反転対称性が見られる。これは、正電荷クラスタと負電荷クラスタの静電的な相互作用がヘアピン構造の原動力となっていることを示唆している。さらに、二次構造ブレイカ予測でとらえることのできないループ領域 B, C においても電荷の反転対称性が見られる。これは、電荷の反転対称性が新しい二次構造ブレイカとなる可能性を示すものである。そこで、グロビンフォールドタンパク質 18 個について、A~E のループ領域ごとに二次構造ブレイカでとらえられないループ領域と電荷の反転対称性の関係性について調べた。アミノ酸配列上における異符号の電荷クラスタによる電荷の反転対称性の特定には、我々の考案した CS score を用いた。Fig. 2(b) では、CS score の極小値と電荷の反転対称点とが良く一致している。このように CS score では、電荷の反転対称性の中心点をその極小値としてとらえることができる。Fig. 3 は、18 個のグロビンフォールドタンパク質の A~E のル

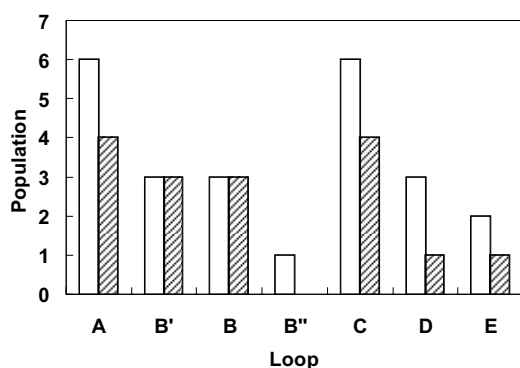


Fig. 3. The populations of loops are not covered by predicted breaker and loops are covered by inverse charge symmetry in these uncovered loops. Open and shade bar shows loops which are not covered by predicted breaker and loops are covered by inverse charge symmetry, respectively.

ープ領域ごとに、二次構造ブレイカでとらえられないループとそのうち電荷の反転対称性を示すループ領域との数を比較したものである。A~E のループ領域それぞれにおいて、予測した二次構造ブレイカでとらえることのできないループ領域がいくつか見られるが、その多くに電荷の反転対称性が見られる。電荷の反転対称性は、これら二次構造ブレイカでとらえることの出来ないループ領域の 3 分の 2 をカバーしていた。これは、電荷の反転対称性が、有力な 5 番目の二次構造ブレイカのみをメカニズムであることを示している。

3.2 グロビンフォールドタンパク質間の二次構造ブレイカ予測と電荷の反転対称性比較

二次構造ブレイカは、二次構造を壊し、ループ構造に変えるといったタンパク質の重要な構造的特徴である。アミノ酸配列の相同性が低いにも関わらず、類似した構造を持つタンパク質間における二次構造ブレイカの比較は、タンパク質の立体構造形成メカニズムを理解するうえで重要な情報となる。

これまで、二次構造ブレイカのみをメカニズムとして、プロリン、グリシン、小さい極性残基、両親媒性残基の 4 つを見出すことができた。そして、今回、5 番目の二次構造ブレイカのみをメカニズムとして、電荷の反転対称性を見出した。そこで、4 つの二次構造ブレイカに電荷の反転対称性を加え、リモートホモログの関係にあるグロビンフォールドタンパク質において、二次構造ブレイカの出現傾向の比較を行った。Fig. 4 は、A~E のループ領域ごとに二次構造ブレイカタイプの出現頻度調べたものである。それぞれの構造上対応するループ領域における各二次構造ブレイカタイプの出現頻度を見てみると、非常に類似した構造を取るタンパク質同士にも関わらず、全てのタンパク質に共通して出現する二次構造ブレイカタイプは存在せず、二次構造ブレイカタイプの出現パターンは全く統一されてはいない。これは、同じ構造パターンを持つタンパク質のループ形成が必ずしも同じメカニズムによってなされているわけではないことを示している。おそらく、タンパク質はいくつかの異なるメカニズムを使い分けてその構造を維持しているのだろう。このようなタンパク質の構造形成メカニズムの多様性がアミノ酸配列と構造との関係を複雑にしている一方で、タンパク質の長い進化の歴史におけるアミノ酸残基の置換に対してもその構造を維持することのできる要因でもありとされる。

4. おわりに

4 つの二次構造ブレイカのみをメカニズムに加え、新たに電荷の反転対称性という新しい二次構造ブレイカのみをメカニズムを見出すことができた。しかし、電荷の反転対称性は二次構造中にも見られる。二次構造中に現れる電荷の反転対称性は、N 端側に負電荷を持ち、C 端側に正電荷

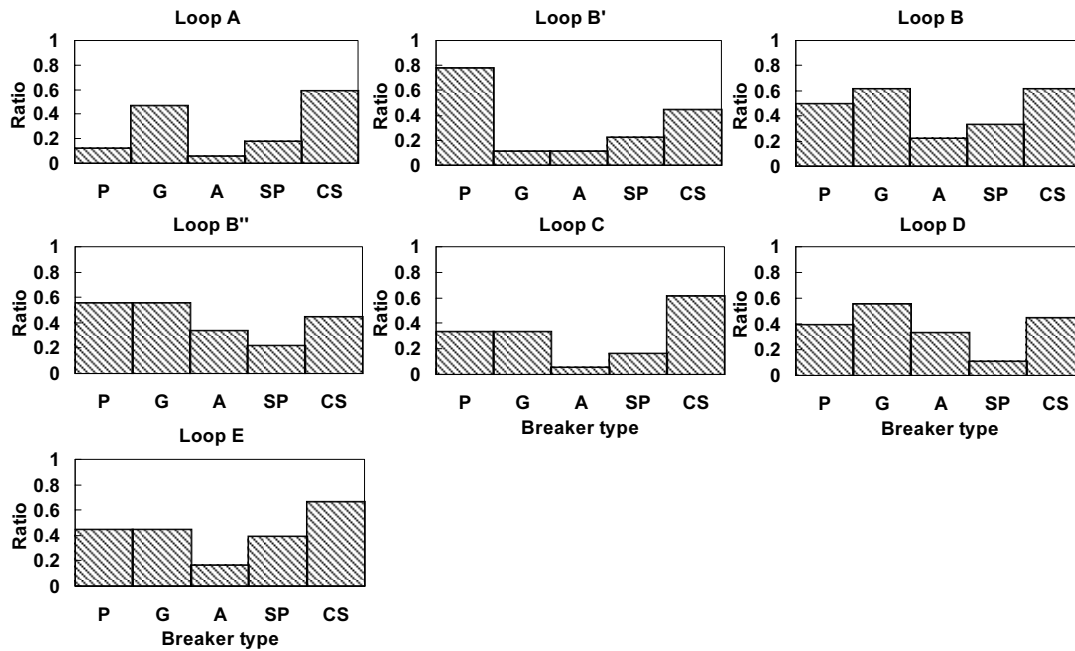


Fig. 4. The appearance ratio of four types of secondary structure breaker and inverse charge symmetry in each loop region. P, G, A, SP and CS show proline breaker, glycine breaker, amphiphilic residues breaker, small polar residues breaker and inverse charge symmetry, respectively.

を持つものが多い。これは、ヘリックス構造の持つ電気双極子と反対の電荷分布であり、ヘリックス構造を安定させるものと考えられる。つまり、電荷分布の反転対称点は二次構造ブレイカ因子とも二次構造形成因子とも働くと考えられる。今後としては、反転対称性が二次構造ブレイカ因子として働く条件、二次構造形成因子となる条件を明確にし、これらを判別することが課題であると考えられる。

また、二次構造ブレイカ予測システム SOSUIbreaker は、以下のウェブサイトで利用できるため、今後の研究に役立てていただけたら幸いです。

http://bp.nuap.nagoya-u.ac.jp/sosui/sosuibreaker/sosuibreaker_submit.html

引用文献

- 1) B. Rost, *Methods. Biochem. Anal.* **44**, 2003, 559-87.
- 2) Levitt, M. *Biochemistry.* **17**, 1978, 4277-4285.
- 3) Chou, P.Y. and Fasman G.D. *Adv Enzymol.* **47**, 1978, 45-148.
- 4) Aurora, R and Rose, G.D. *Pro Sci.* **7**, 1998, 21-38.
- 5) Colloc'h, N and Cohen, F.E. *J Mol Biol.* **221**, 1991, 603-613.
- 6) Imai, K. and Mitaku, S. *BIOPHYSICS*, **1**, 2005, 55-65.
- 7) 今井賢一郎, 美宅成樹, *生物物理*, **47**, 2007, 49-53.
- 8) Imai, K., Asakawa, N., Tsuji, T., Sonoyama, M. and Mitaku, S. *CBIJ*, **5**, 2006, 65-77
- 9) Murzin A. G., Brenner S. E., Hubbard T., Chothia C. *J. Mol. Biol.* **247**, 1995, 536-540.
- 10) Mitaku, S., Hirokawa, T., and Tsuji T. *Bioinformatics.* **18**, 2002, 608-616.
- 11) Kyte, J. and Doolittle, R.F. *J Mol Biol.* **157**, 1982, 105-132.