

## タンパク質の新規サブドメインの提案

—二次構造、モジュールを超える新たな構造単位の定義に向けて—

佐々木 彰吾<sup>†1</sup> 山村 雅幸<sup>†1</sup>

本稿では、タンパク質の階層性を解明することを目的とし、独立したパーツとして利用可能で、ドメインより細かい階層の新たな構造単位を提案する。従来、ドメインより細かい構造単位として「空間的コンパクトさ」を指標とする、モジュールが提唱されてきた。しかし、空間的コンパクトさはパーツの独立性とは直接関係ない。独立したパーツとして利用可能なパーツを探すためには、独立性と直接関係した量を指標にしなければならない。そこで、本研究ではポテンシャルエネルギーを指標として用い、タンパク質のサブドメインを定義し、いくつかのタンパク質で新しいサブドメインの発見を試みた。

### Proposal of New Sub-domain of Protein

SHOGO SASAKI<sup>†1</sup> and MASAYUKI YAMAMURA<sup>†1</sup>

To reveal the hierarchy of a protein structure, we propose a new structural unit that is shorter than a domain. This unit can be treated as an independent part of a protein. As a structural unit which is shorter than a domain, the concept of module has been defined as relatively compact regions of protein. However, the compactness does not relate directly to the independency of a part in a protein. We define a new structural unit adopting the free energy of a protein as index. We also show some new subdomains retrieved by own definition in some well known proteins.

### 1. はじめに

階層性は物事の理解と設計において重要な役割をもっている。人工物の設計の際には、階層性を意識した設計がなされ、これにより他の階層の状態に依存しない柔軟な設計を可能にしている。

タンパク質においては、一次構造、二次構造、三次構造、四次構造と呼ばれる階層が知られている。しかし、これらは階層ごとの独立性を考慮して定義されたものではなく、その導入によって他の階層に依存しない設計を可能にするものではない。例えば、二次構造は単にヘリックス、シート等の特徴的な構造を指すものに過ぎない。

階層の独立性に注目したタンパク質の単位にドメインがある。ドメインは単体でも安定な構造をとり、機能の単位である。単体での安定性から、独立したパーツとして扱われ、ドメインごとのタンパク質組み換えが行われている。ところが、ドメインは独立したパーツとして扱える最小の単位というには大きすぎる。

ドメインより細かい階層の単位としてモジュールが提唱されている。モジュールは空間的にコンパクトなひとつながりのペプチドとして定義され、10～40残基程度の長さをもつ。郷らは、タンパク質中のモジュールの切れ目の位置とエキソン/イントロン境界の位置との相関を示すことで、生物の進化がモジュールの組み換えによって起こること、すなわち、モジュールがタンパク質進化のビルディングブロックであることを示した。<sup>1)2)</sup>

モジュールは空間的なコンパクトな集団として定義され、その結果として単体での独立性や進化の単位としての機能をもつという結果が報告されているが、空間的なコンパクトさと独立性（単体での安定性）は直接関係をもたない。他の階層から独立したパーツを見つけ出すためには、定義の時点から単体での安定性を考慮すべきである。また、単体の安定性に直接関係しているのは物理的相互作用である。さらに、モジュールには、モジュール境界を決定する際の閾値を恣意的に決定する必要があるが、「空間的なコンパクトさ」がタンパク質の物性を直接表す量ではないので、その値に意味を見出すことができない。

本研究では、タンパク質の階層性を解明することを目的とし、独立したパーツとして利用可能なドメインより細かい階層の新たな構造単位を提案する。この構造単位は、タンパク質の安定性を表す自由エネルギーを指標にすることで、サブドメイン決定時の恣意性を排除し、階層性の本質的な理解を目指す。これにより、高精度の立体構造予測やより柔軟な機能設計への応用が期待できる。

<sup>†1</sup> 東京工業大学  
Tokyo Institute of Technology

## 2. 提案するサブドメインの定義

本研究で提案するサブドメインは、平衡状態においてポテンシャル的に他の部分から独立なひとつながりのペプチドとして定義する。タンパク質を二つの集団に分割したとき、そ

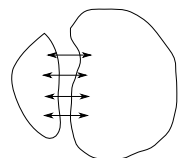


図1 サブドメインのイメージ図

のタンパク質を構成する原子間に働く相互作用は、集団間の相互作用と集団内の相互作用に分けられる(図2)。同一のタンパク質について集団間の相互作用を計算したときその値は分割のしかたにより異なる。集団間の相互作用が小さくなるように分割された断片をサブドメインとして採用する。

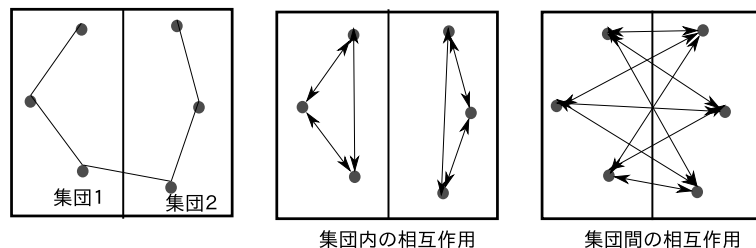


図2 集団内相互作用と集団間相互作用

相互作用の大きさを計算する際のポテンシャル関数は、タンパク質のシミュレーションで一般的に用いられる半経験的なポテンシャル関数を用いた。<sup>3)4)</sup>

本研究の結果の正しさはポテンシャル関数の正しさに依存するが、このポテンシャル関数自体に議論の余地がある。しかし、ポテンシャル関数の設計自体が困難な問題であり、分子シミュレーションにおいて一般的な課題であることから、便宜上これを用いた。

とはいえ、同じポテンシャル関数を用いて計算したもの同士の比較には一定の意味がある

と考える。

$$V = \sum_{\text{bonds}} K_{ij}^R (R_{ij} - R_{ij}^{\text{eq}})^2 + \sum_{\text{angles}} K_{ij}^\theta (\theta_{ijk} - \theta_{ijk}^{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V}{2} [1 + \cos(n\phi_{ijkl} - \gamma_{ijkl})] + \sum_{j=1}^{\text{atoms}} \sum_{j>1}^{\text{atoms}} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{j=1}^{\text{atoms}} \sum_{j>1}^{\text{atoms}} \frac{q_i q_j}{\epsilon R_{ij}}$$

## 3. 計算方法

集団間の相互作用を計算する方法は大きく分けて二つの方法が考えられる。

### 3.1 直接的な方法

ひとつ目の方法は、集団間のポテンシャルを直接計算する方法である。原子間のポテンシャルは、原子の種類、原子間の距離、結合の状態が解れば計算できる。集団間をまたぐ原子ペアのポテンシャルをそれぞれ計算し足し合わせればよい。ただし、結合角・二面角のポテンシャルについては、三つ以上の原子のポテンシャルの和で与えられているので、原子のペアのポテンシャルに分解する必要がある。

$$V_{AB} = \sum_{i \in A}^{\text{atoms}} \sum_{j \in B}^{\text{atoms}} V_{ij}$$

### 3.2 間接的な方法

二つ目の方法は、タンパク質全体のポテンシャルエネルギー、集団単体でのポテンシャルエネルギーから、間接的に集団間のポテンシャルを求める方法である。タンパク質全体のポテンシャルエネルギーと集団単体でのポテンシャルエネルギーは AMBER, GROMACS などの既存ソフトウェアで計算することができる。

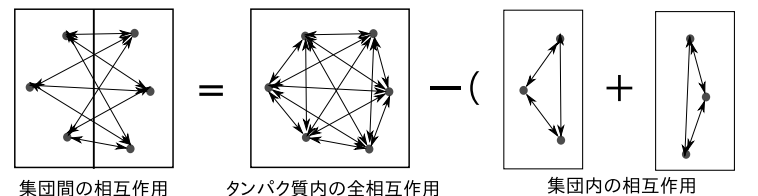


図3 間接的な方法の模式図

### 3.2.1 計算例

上述の間接的な方法の実際の計算結果を示す(図5)。主鎖に二ヶ所の切れ目をいれタンパク質を分割し(図4)、切れ目1の位置を横軸、切れ目2の位置を縦軸とし、それに対応する集団間のポテンシャルを濃淡で表した。切れ目は、主鎖に一残基ごとにいった。すなわちタンパク質の残基数  $N(=312)$  に対し、 $N^2/2$  通りの切れ目の入れ方についてそれぞれ集団間のポテンシャルを計算した。この例において、約20~170残基のポテンシャル的に独立な集団が確認できた。

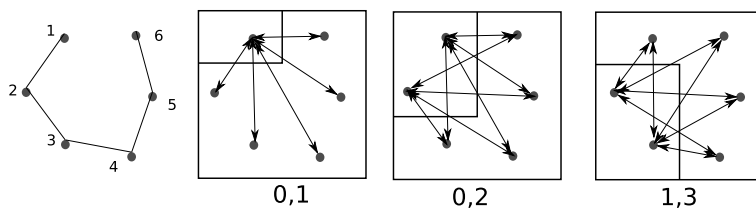


図4 主鎖の分割

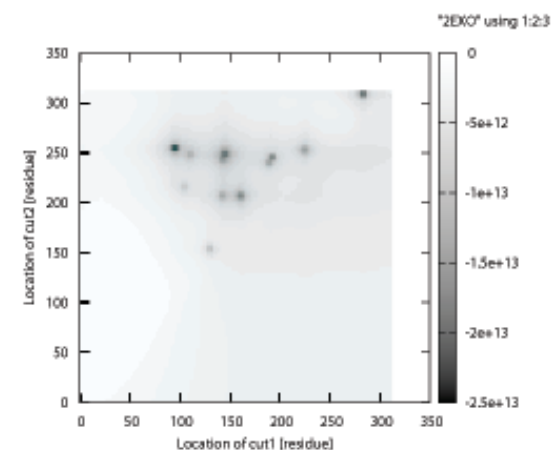


図5 計算結果の例: いくつかのピークが見られるが、例えば(100,250)のピークは100残基目から250残基目までを切り出した部分の独立性が強いことを表している。

## 4. おわりに

本研究では、タンパク質の基本的な構造単位の定義に向け、ポテンシャルエネルギーを指標にしたサブドメインの決定法を提案した。

現状では、まだ予備的な計算が終わった段階で、解決しなければならない課題が多い。「間接的な方法は」既存のソフトウェアで計算することができるが、計算時間が問題である。また、現段階では真空中で計算しているが、将来的には溶媒の影響を考慮する必要がある。今後は、比較的高速に計算できる「直接的な方法」を用いたプログラムを完成させ、様々なタンパク質について網羅的な解析を行い、本研究で提案した単位が構造的・機能的あるいは進化的にどのような意味をもつか検証する。

### 参考文献

- 1) Gō, M., Nosaka, M. Protein architecture and the origin of introns. Cold Spring Harbor Symp. Quant. Biol. Vol.52, pp.915-924 (1987)
- 2) Yoko Sato, Yoshihito Niimura, Kei Yura, Mitiko Gō. Module-intron correlation

and intron sliding in family F/10 xylanase genesl. Gene. Vol.238, Issue 1, pp.93-101 (1999)

- 3) Scott J. Weiner, Peter A. Kollman, David A. Case, U. Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc, 106(3), pp.765-784 (1984)
- 4) Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, Jr., David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc, 117, pp.5179-5197 (1995)
- 5) 名波 剛、山村 雅幸: 郷モジュールに基づくフォールディングシミュレーション . 第29回知能システムシンポジウム, pp.283-288 (2002)