

## 蛋白質のノットによる分類

## Classification of protein by knot

○野口勝<sup>1</sup>, 山中雅則<sup>2</sup>\*Masaru Noguchi<sup>1</sup>, Masanori Yamanaka<sup>2</sup>

Abstract: This study classifies proteins listed on Protein Data Bank by topological genus which calculated from secondary structure with main chain and hydrogen bonds of protein. For a structure drawing, we applied two methods one is named 1page-representation at this study which using only upper plane on main chain, the other is named 2page-representation which using upper and lower plane. Finally we compare both results of genus and discuss which one is natural method as classification of protein.

## 1. はじめに

蛋白質立体構造の解析と研究は、近年精力的に進められている。構造データベースとしてはPDB[1],をはじめとして、CATH[2], SCOP[3]などがあげられる。これらのデータベースは経験的手法により、蛋白質の立体構造の分類をするものである。これらの手法以外にも高分子の数学的なトポロジーを用いて構造を分類する方法の研究も盛んである。RNAに関するpseudoknotによる位相幾何学的分類においては、高分子から骨格を抽出し、2次元トーラスへ埋め込み、その種数によって分類が行われている。本研究では、蛋白質の骨格となる主鎖と水素結合のデータから得られるFigure1のような2次元のグラフを2次元トーラスに埋め込み、そのトーラスの種数によって分類する。また、この手法を適用し、蛋白質を分類する場合、種数が不自然に大きくなる傾向がある。本研究では、この問題の解決のために新たな2次元グラフの描写の手法を考案し、種数を算出すると共に、どちらが自然な分類の手法かを検討したので報告する。

## 2. 分類

本研究では 2 種類のグラフの描写を試みた。一方は 1page-representation である。この手法では Figure1 のように水素結合を主鎖の上半面のみで描く。もう一方は 2page-representation である。この手法では Figure5 のように主鎖を境に上下の半面に水素結合を分配する。

## 2-1. 1page-representation を用いた分類

初めに、蛋白質の 3 次元構造から蛋白質の骨格となる主鎖と水素結合のデータを抽出し、Figure1 のように蛋白質の N 末端から C 末端に至る主鎖を直線で描き、主鎖の上半面に水素結合をアーチ状に描く。Figure1(c) では主鎖を黒、水素結合を青のアーチで描いた。

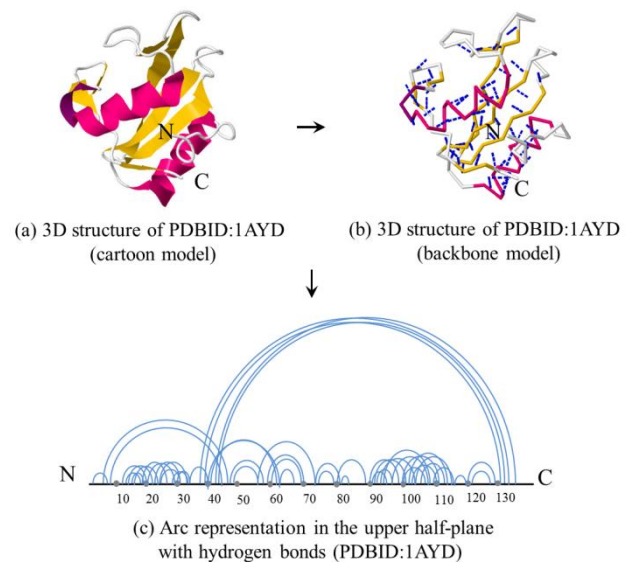


Figure 1. Example of a pseudoknot structure is drawn (PDBID:1AYD)

得られたグラフを Figure2 のように既約分解し、分解したそれぞれのグラフの種数を算出する。

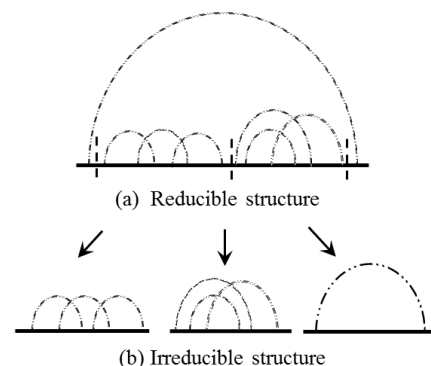


Figure 2. The process of irreducible decomposition

蛋白質中に存在する水素結合の数を  $H$ 、主鎖と水素結合が形成するループ構造の数を  $L$  としたとき、グラフ理論から種数  $g$  は(1)式で与えられる。

$$g = \frac{H - L}{2} \quad (1)$$

次に、分解した Figure2(b)の各グラフの種数の算出を

Figure3 に示す. ここではループ構造を赤で示した.

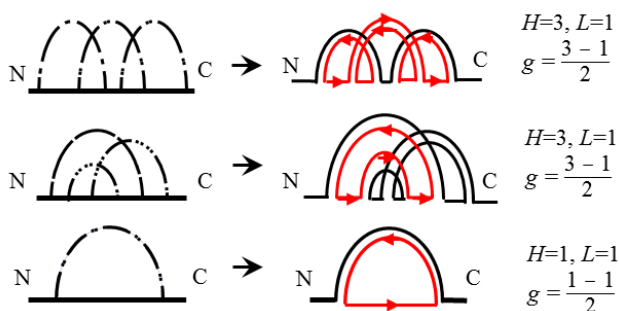


Figure 3. Example of calculation of genus

Figure3 の 3 つのグラフは算出した種数を持つ 2 次元トーラスに埋め込まれる. その過程を Figure4 に示す. 埋め込みでは, 2 次元のグラフの N 末端から C 末端に至る主鎖を Figure4(b) のように円形に変換する. その後, Figure4(c) のように点線で描いた水素結合が互いに交差しないようトーラスに埋め込まれる.

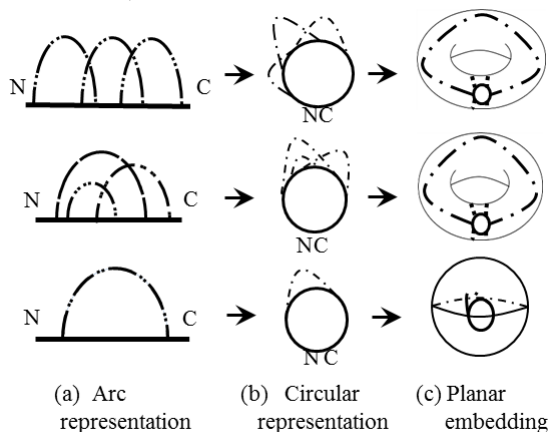


Figure 4. Process of embedding on torus

### 2-2. 2page-representation を用いた分類

ここでは既約分解したそれぞれのグラフの水素結合を Figure5 のように直線で描いた主鎖の上下に分配する.

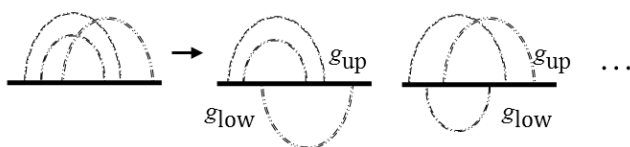


Figure 5. Arc representation both in the upper and lower half-planes with hydrogen bonds

割り振られたグラフのパターンの総数は  $2^H$  通りとなる.  $2^H$  通り存在する上側のグラフの種数  $g_{up}$  と, 下側のグラフの種数  $g_{low}$  を (1) 式から算出し, (2) 式からその合計値  $g_{2page}$  を算出する. 最後に,  $g_{2page}$  の中で最小となる数値  $g_{min}$  を種数  $g$  とする (3) 式.

$$g_{2page} = g_{up} + g_{low} \quad (2)$$

$$g = g_{min} \quad (3)$$

### 3. 結果

現時点で分類した既約分解後のグラフの総数は 860 個となった. それぞれのグラフの水素結合の数と種数の関係をプロットする. 赤は 1page-representation を用いた種数の結果である. 青は 2page-representation を用いた種数の結果である.

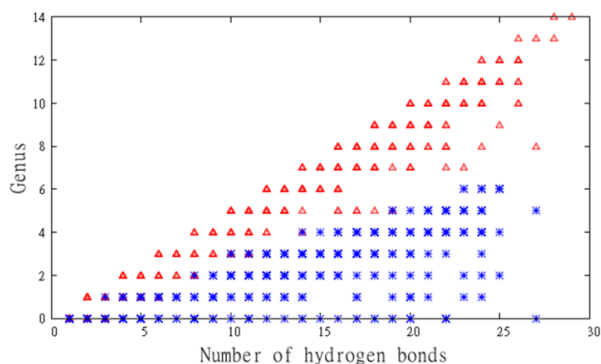


Figure 6. Relation between genus and the number of hydrogen bonds

直観的に平面的な逆平行  $\beta$  シートにおいて 1page-representation を用いた分類を行うと種数が大きくなり不自然である. しかし, 2page-representation による分類では種数が低くなり, 種数の過剰な上昇を避けられたと考えられる. この点から, 2page-representation による分類は 1page-representation を用いた分類に比べると自然な種数が得られていると考えられる. 種数の値が高い領域においては Figure7(a) のような  $\alpha$  ヘリックスなど水素結合が互いに交差し合った結合網を形成するものが多く, 種数の低い領域では Figure7(b) のような可約な構造を有する蛋白質が多い. 中間層では Figure7(a) と Figure7(b) のような構造を共有した蛋白質が多いと考えられる.

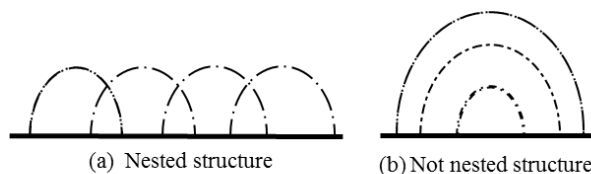


Figure 7. Nested structure and not nested structure

### 4. 今後の課題

蛋白質の網羅的分類のため, これまでに分類を行っていない蛋白質の分類を行う.

### 5. 参考文献

- [1] PDB: <http://www.pdb.org>
- [2] CATH: <http://www.cathdb.info/latest/index.html>
- [3] SCOP: <http://scop.mrc-lmb.cam.ac.uk/scop>