

Protein: 蛋白質分子表面の局所凹凸形状分類の一手法

西山 慧子

(指導教員：伊藤貴之)

1. はじめに

蛋白質は、生体反応の触媒である酵素、臓器等の構成成分、機能を調節するホルモンなどの原料となる物質であり、また生体内の反応において大変重要な働きを担う。蛋白質の機能は表面形状に大きく依存し、表面の局所凹凸形状の特徴との関係が深いと言われている。このような蛋白質の分子表面の局所凹凸形状を分析することにより、他の分子との相互作用や機能、他の既知の蛋白質との関係などがわかるようになる。

本論文では、蛋白質表面形状データベース(eF-site)に記録されている分子表面情報に基づいた、分子表面の局所凹凸形状を観察するための手法およびインタフェースとして、Protein (Partial Relief Observation TEchnique and INterface) を提案する。蛋白質の表面微細形状を分析する手法は以前にもいくつか報告されている[1][2]が、計算量が膨大になることが多い。この問題を解決するために、蛋白質の局所凹凸形状を単位とした特徴分析が合理的であると判断した。

2. システムの概要

今回実装したシステムは、2工程より構成される。

工程1：各局所凹凸形状の分類・観察。

- 蛋白質の分子表面から、局所凹凸形状を抽出。
- 三次元モデルの類似度判定手法を応用して、局所凹凸形状の特徴量を算出し、局所凹凸形状間の類似度を算出。
- 局所凹凸形状をクラスタリング。

工程2：複数の局所凹凸形状よりなる部分形状の抽出・観察。

- 三角メッシュの単純化によるグラフ生成。
- グラフの経路比較により、類似部分形状を抽出。
- クラスタリング結果、抽出結果の可視化。

図1に、本論文において構築したシステムの処理の流れを示す。

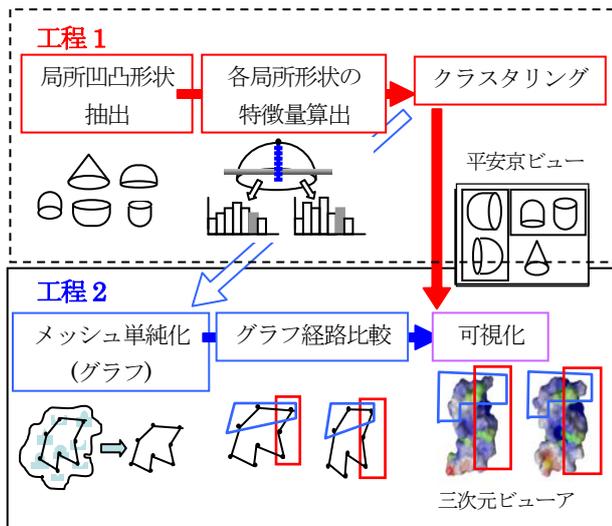


図1：システムの処理の流れ。

2.1 各局所凹凸形状の分類・観察

局所凹凸形状の抽出処理において、まず各頂点に対して、法

線ベクトルを用い接平面を定める。このとき、全ての近接頂点が接平面の下部に位置するならば、凸部に属すると判断する。逆に全てが上部に位置するならば、凹部に属すると判断する。以上の判断結果から、各頂点に「凸」「凹」「該当なし」のいずれかの属性を与える。続いて、同一属性を有する隣接頂点を、再帰的に集めることで、1個の局所凹凸を構成し、その頂点群に固有のラベル（通し番号）を与える。

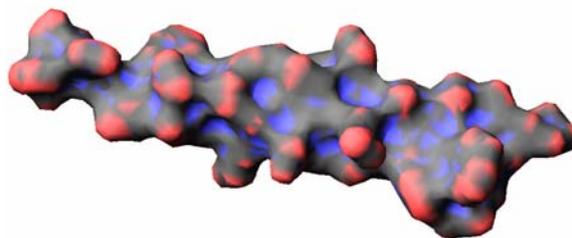


図2：ラベリング結果。赤は凸部、青は凹部を示す。

特徴量算出の処理では、前節までの処理によって抽出された局所凹凸形状の表面上に、点群をランダムに生成する。そして局所凹凸形状の主軸から各点への距離を算出し、その平均と分散のヒストグラムを特徴量として算出する。この手法は文献[3]に提案されているPS法に類似している。

クラスタリング処理においては、まず局所凹凸形状間の類似度を、それらの特徴量のユークリッド距離を用いて算出する。続いてK-means法を適用し、局所凸部形状と局所凹部形状を混在させてクラスタリングする。

2.2 複数の局所凹凸形状よりなる部分形状の抽出・観察

隣接する複数の局所凹凸形状の隣接関係の類似度を算出することで、異なる蛋白質間に対して、やや大きめの局所類似性を算出することがあげられる。我々は、以下の処理手順により類似部分形状の抽出を、実現する。

メッシュの単純化の処理では、各々のクラスタにラベル（通し番号）を与え、同クラスタに属する局所凹凸形状に同番号を与える。局所凹凸形状の頂上頂点または底面頂点だけを残すように三角ポリゴンを単純化し、これを局所凹凸形状グラフとして扱う。本手法は、頂上および底面頂点を消去しないことを除き、文献[4]で提案されているポリゴン単純化と類似している。

局所凹凸形状グラフの経路比較による類似部分の抽出処理では、構成する三頂点のクラスタ番号が等しい三角形群に、固有のラベル（通し番号）を与える。同ラベルを持つ各三角形群の接続関係を比較することで、類似経路を持つ類似部分形状を抽出する。

2.3 可視化インタフェース

本手法では、クラスタリング結果を視覚的に探索するためのインタフェースとして、大規模階層型データ可視化手法「平安

京ビュー」を用いる。ここで本手法では、局所凹凸形状のクラスタリング結果から、以下のように階層型データを構築する。

- 局所凹凸形状をクラスタリング結果に従って分類。
- 複数の蛋白質から抽出した局所凹凸形状が各クラスター内に混在している場合には、蛋白質ごとに再分類。
- 局所凹形状と局所凸形状に応じて、再分類。

また本手法では蛋白質分子表面の3次元ビューアを提供し、図2に示すような凹部、凸部を色分け表示する機能の他に、特定のクラスターに属する局所凹凸だけを表示する機能や、2.2章で述べた処理より抽出した類似部分形状を表示する機能を搭載する。

「平安京ビュー」の画面上で興味深いクラスターを見つけ、そのクラスター番号を3次元ビューア上で入力すると、そのクラスターに属する局所凹凸形状だけを抽出して表示する、または、抽出した類似部分形状を表示する、というような機能を実装できる。

3. 実行結果

本章の実験では、表面形状が非常に類似している、蛋白質ID(1yee,1yec)を使用した。図3は、両蛋白質のクラスタリング結果を、平安京ビューにて表示した結果画像である。

図3の結果画像より、各クラスターに、ほぼ同数の局所凹凸形状が分類されていることが観察される。この結果より、2種の蛋白質は、形状の似た局所凹凸形状を、ほぼ同数持ち合わせていることが示される。また第三階層に注目すると、下部に位置する局所凹形状のクラスターは、局所凸形状クラスターに比べ、属するノード数が、多いことが観察される。この結果より、両蛋白質は、反応発生の際、受動体となる傾向が強いことが観察された。また、多くのノードが、中間値を示すオレンジで表現されているが、点線の丸で囲ったクラスターにおいては、複数の色のノードが観察された。この両クラスターは、1yecの局所凹形状クラスターであった。このように、本手法を用いることにより、両蛋白質の有する異なる形状、属性部分を容易に観察できる。また、両蛋白質が異なる反応を有する場合、その反応の要因となる形状群の絞込みに効果的である。

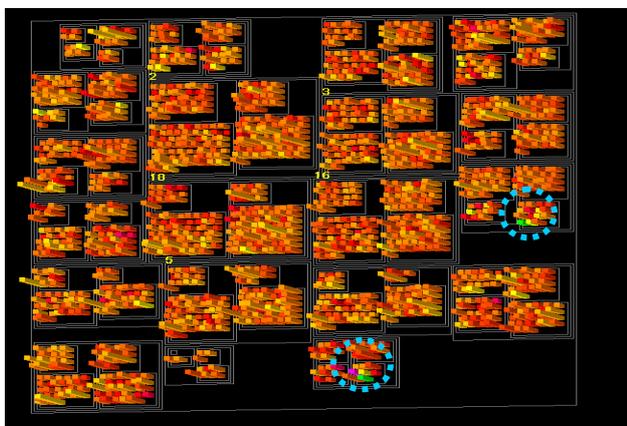


図3: 温度を高さで、電位を色で表現した可視化例。

図4は、2.2章の処理より得られた、類似部分形状の結果画像である。上図は、両形状の局所凹凸形状グラフを示しており、両部分形状は4つの局所凹凸形状より構成されており、互いに(0.2.2.0)のクラスターに属する形状を有することが観察できる。

また下図では、局所凸形状を赤で、局所凹形状を青で示している。この図より、各々の有する形状属性(凸属性、凹属性)も類似していることが観察される。抽出した類似部分形状は、両蛋白質の角形状の根元部分に位置し、反応の受動体となる窪み形状部分であった。このように本手法を用いることにより、機能発現において重要な役割を担う部分形状の抽出が容易になる。

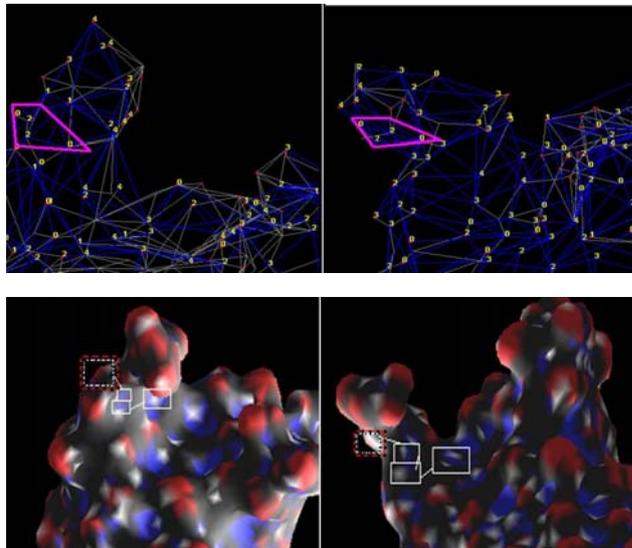


図4: (上)1yee(左)と1yec(右)の局所凹凸形状グラフ。
(下)1yee(左)と1yec(右)の表面形状。

4. まとめ・今後の課題

本論文では、蛋白質の分子表面から局所凹凸形状を抽出し、各々の特徴量を算出し、これに基づいて局所凹凸形状間の類似度を算出し、これらを分類する手法を提案した。また、分類結果を用い、局所凹凸形状グラフを作成し、大きめの類似部分形状を抽出する手法を提案した。今後の課題として、計算時間の削減を試みたい。また経路比較において、頂点位置情報を考慮したい。

謝辞

本報告にあたって、蛋白質表面形状の構造類似度算出に関して、東京大学の木下賢吾准教授より、また形状属性判定に関して、お茶の水女子大学の萩田真理子准教授より、貴重な情報を賜ったことを感謝いたします。

参考文献

- [1] 清水, Nripendra L. S., 大川., 類似表面結合処理の分割による蛋白質分子表面モチーフ抽出の並列化方式, 情報処理学会論文誌: 数理モデル化と応用(TOM), Vol. 47, No. SIG 1(TOM14), pp.120-129, 2006.2.
- [2] Via, A., Ferre, F., Brannetti, B., Helmer-Citterich, M., Protein surface similarities: a survey of methods to describe and compare protein surfaces, Cell MolLife Sci., Vol. 57, No. 13-14, pp. 1970-1977, 2000.
- [3] 小田切, 衣嶋, 武井, 大淵, 慣性モーメントの包絡線を用いた3次元モデルの形状類似検索, 映像情報メディア学会誌, Vol. 56, No. 10, pp. 1589-1597, 2002.
- [4] Hugues, H., Progressive meshes, Computer Graphics (SIGGRAPH 96), ISBN:0-89791-746-4, pp. 99-108, 1996.