

蛋白質表面形状分析手法 PROTEIN を用いたポケット形状評価

A Pocket Shape Evaluation Technique Applying Protein Surface Technique

金子 彩香*
Ayaka KANEKO

伊藤 貴之*
Takayuki ITO

* お茶の水女子大学大学院 人間文化創生科学研究科
Graduate School of Humanities and Sciences, Ochanomizu University

Protein is a main component of our body. It is important to determine the affinity of proteins with drugs (so called “druggability”) in the medicine development. It is well-known that the drugs go in hollows (so called “pocket”) of protein surfaces to act. As a part of protein surface analysis project “PROTEIN”, we attempt to evaluate druggability of the proteins. As a preliminary step, this paper presents a pocket shape evaluation technique. Supposing that the protein surfaces are provided as triangular meshes, the technique first extracts the pockets from the protein surfaces and calculates feature values of pocket shapes. It then compares the feature values with feature values of sample-pockets which has high affinity with the drugs.

1. 概要

蛋白質は細胞の主要成分であり、細胞の働きを推し進める、調節するなどといった人体において非常に重要な役割を担っている。細胞内で働く蛋白質は何万種類と存在し、それぞれが独自の構造体と機能を持っている。また、薬の多くはその成分が蛋白質に作用することから、蛋白質は薬品開発においても注目されており、各蛋白質がそれぞれ薬とどの程度の親和性を持つかを把握することが重要となっている。本報告では、蛋白質と薬との親和性を *druggability* と呼ぶ。また、薬が蛋白質に作用する際、薬成分は蛋白質表面上のある程度大きくて深い窪みに入り込み、その場所に一定時間留まり相互作用することが多いとされている。このような蛋白質表面上の窪みを、本報告ではポケットと呼ぶ。

以上の背景から我々は、蛋白質表面形状分析プロジェクト PROTEIN の一環として、各蛋白質が薬とどの程度の親和性を有するかを判定するための蛋白質の *druggability* 評価手法の確立を目指している。その前段階として、本報告では蛋白質表面上の各ポケットの形状評価値を算出する手法を提案する。

2. 関連研究

2.1 蛋白質表面形状分析手法 PROTEIN

我々は、PROTEIN (Partial Relief Observation TEchnique and INterface) と称して、蛋白質表面形状分析に取り組んでいる。その一環として我々は、蛋白質表面の局所凹凸形状分類手法 [Nishiyama 08] や、蛋白質表面形状からのポケット抽出手法 [Nakamura 11] を発表している。このうち [Nishiyama 08] では、蛋白質の反応性が部分形状に密接に関係していることに着目し局所凹凸形状の分類およびその可視化を試みた。また [Nakamura 11] では、蛋白質表面上のポケット高速抽出を試みた。本報告では、蛋白質表面形状からのポケット抽出に [Nakamura 11] の手法を適用する。

2.2 蛋白質の類似度評価値算出手法

[Perot 10] は、様々な蛋白質の結合部位の類似度評価値算出手法と、その結果公開のデータベースをサーベイしている。それらの中で、本手法と同じく蛋白質の表面形状に着目した手法の一例として、eF-site の三角メッシュデータを用いて各頂点の物性付きベクトルに対しクリーク探索法を適用する eF-seek [Kinoshita 02] がある。

蛋白質ポケットの反応性評価にあたり、現時点で我々は、蛋白質表面形状の幾何学的情報のみから評価値を算出している。しかし、より信頼性の高い評価値を算出するためには、幾何学的情報以外の情報も参照する必要がある。そこで将来的には、提案手法に [Perot 10] で紹介されているような手法を併用し、化学的情報も加えた評価値算出手法の確立を視野に入れている。

2.3 三次元モデルの類似度比較

三次元モデルの類似度比較においても著名な手法が数多く存在し、それらの手法は、特徴量として注目する構造が幾何学的構造のものと、位相的構造のものに分けることができる。幾何学的構造のものはさらに、頂点に注目したもの、面に注目したものの、ボックスに注目したものに分けられる。代表的な手法として、モデルの表面上に生成された点群に対して 2 点間の Euclid 距離のヒストグラムを用いた D2 [Osada 01]、D2 をさらに拡張した AAD [Ohbuchi 03]、モデルの慣性主軸に沿った包絡線を用いた PS [小田切 02]、などがあげられる。我々はポケット形状の特徴量算出に際して、ポケットの多くはお椀や円錐のような形状をしており、その入り口から底に向かった主軸に沿って特徴量を算出するのがよい、と直感的に考えた。そこで提案手法では、D2 や PS を参考にした特徴量算出方法を適用し、これらの手法と同様に多次元ベクトルでポケット形状を表現している。

3. 提案手法

本手法の処理の流れを図 1 に示す。本手法では、三角メッシュで近似表現された蛋白質表面形状から抽出した各ポケットに対して形状特徴量を算出し、薬との親和性が高いとされるサンプルポケットの形状特徴量との類似度比較により評価値を算出する。

連絡先: 金子 彩香, お茶の水女子大学大学院 人間文化創生科学研究科, 〒112-8610 東京都文京区大塚 2-1-1,
Tel: 03-5978-5399, E-mail: ayaka@itolab.is.ocha.ac.jp

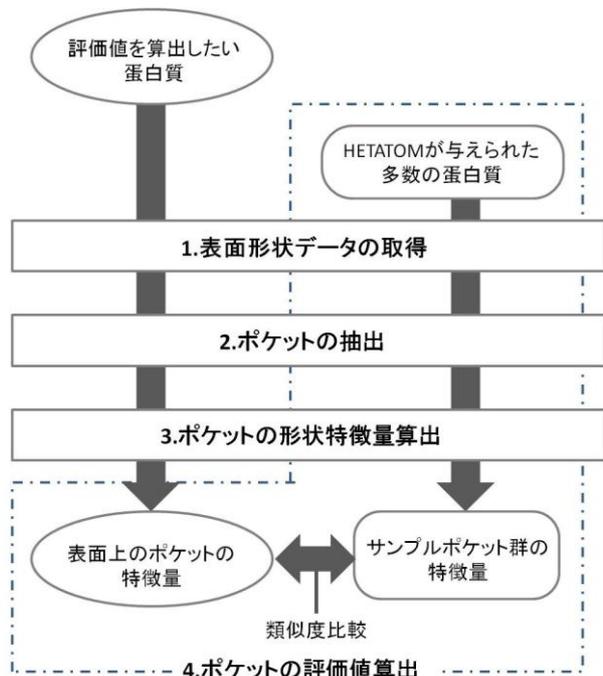


図1 本手法の流れ

3.1 蛋白質表面形状データの取得

提案手法では、蛋白質表面形状は三角メッシュで近似されていることを前提とする。我々による本手法の実装は、eF-site に登録されている分子表面形状を利用している。eF-site では、PDB(Protein DataBank) に登録されている蛋白質を対象として、コロニー曲面抽出技術を適用して分子表面形状を算出し、データベース化して公開している。eF-site では、頂点、2 頂点を両端とする稜線、3 稜線によって囲われた三角形領域、の集合で構成される三角メッシュを、XML 形式で記述して提供している。また、このような幾何学的情報の他に、各頂点には、疎水性、温度、電位といった化学的情報も同時に与えられている。

3.2 ポケットの抽出

本手法では、まず、評価値を算出したい蛋白質のポケットを蛋白質表面上から抽出する。ポケットを抽出する手段として [Nakamura 11] の手法を用いる。この手法は、以下の手順で構成されている。

1. 蛋白質表面を構成する三角メッシュに対してメッシュ単純化を適用し、微小な凹凸を平滑化した概略的な表面形状を得る。
2. 単純化された三角メッシュ上の各頂点に対して、接平面を算出し、隣接頂点がすべて接平面の内側なら当該頂点に凹部の属性を与える。
3. 凹部頂点群を構成する三角形群を抽出する。
4. 抽出された凹部三角形群を単純化前の三角メッシュに投影する。

これらの処理により単純化前の三角メッシュ上に得られた凹部がポケットとなる。抽出されたポケットを青く塗った蛋白質表面の例を、図 2 に示す。

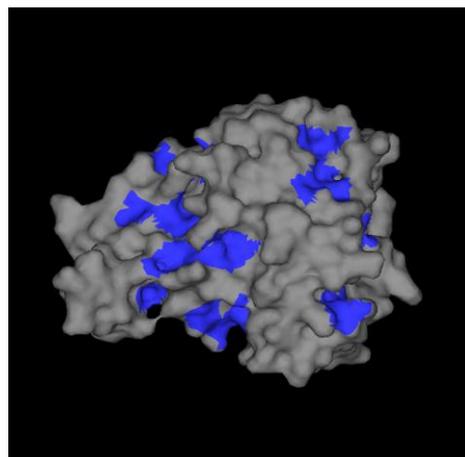


図2 ポケット抽出の例

3.3 ポケットの形状特徴量算出

続いて、前節の処理によって抽出されたポケットの形状特徴量を算出する。本手法の特徴量は、薬成分の入り込み易さに影響を及ぼすとされるポケットの深さと口の広さにもとづいた多次元ベクトルである。以下にその処理手順を示す。また、図 3 はポケット形状の一例を簡略化し、上と横の双方から見た図を用いて処理手順を示したものである。

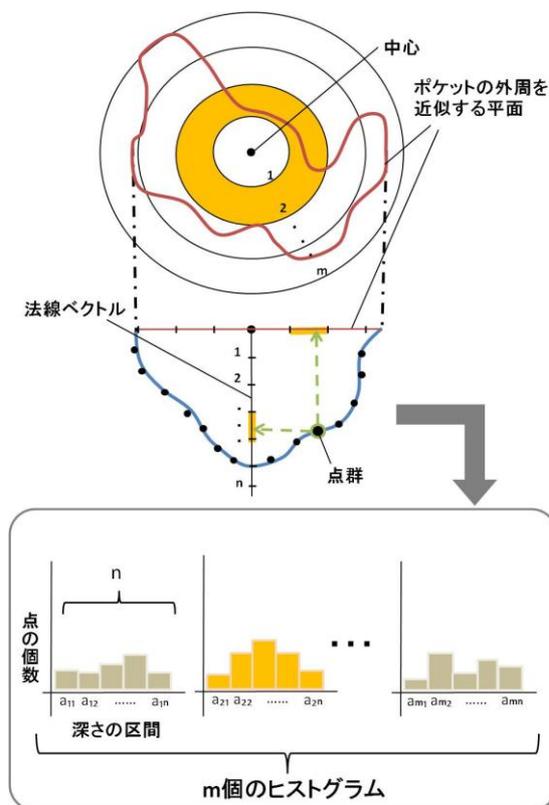


図3 ポケットの特徴量算出手法

1. ポケットの外周面を近似する平面とその平面の中心を算出する。
2. 平面を中心からの距離で m 区間に、また、平面の中心からの法線ベクトルを平面からの深さで n 区間に、それぞれ等間隔に分割する。
3. ポケットを構成する三角メッシュ上にできるだけ等密度に点群を生成する。
4. 各点に対して、その点が中心からの距離と平面からの距離がそれぞれの区間に属しているかを求める。
5. 平面上の各区間に対し、各々の深さの区間に含まれている点の個数についてのヒストグラムを生成し、各階級の度数を特徴量の各次元の値とする。

なお、手順 2 において、各区間の間隔は経験的に 1 \AA とした。また、手順 4 において、中心からの距離が m 区間目を超えてしまった点はすべて m 区間目に属するとする。深さについても同様の処理をする。以上の処理手順より $m \times n$ 次元の形状特徴量が得られる。

3.4 ポケットの評価値算出

続いて本手法では、薬との反応性が高いと推測されたサンプルポケット群を用意し、これらについて形状特徴量を算出する。そして、与えられたポケットの特徴量との類似度比較により、当該ポケットの評価値を算出する。

(1) サンプルポケット群の生成

現時点での我々の実装では、PDB に登録されている蛋白質から、薬との親和性が高いと推測されるポケット群を収集する。PDB に登録されている蛋白質の中には、主に結晶化の過程で残留した、蛋白質の構成分子以外の分子(HETATM)の情報が記載されているものが多い。HETATM の多くは蛋白質と相互作用し易い部位に位置していることが知られている。このことから我々は、HETATM の集中部位周辺のポケットを、薬との親和性が高いポケットと仮定し、サンプルポケットとする。

我々の実装では、HETATM が与えられた多数の蛋白質からサンプルポケットを抽出し、これらの形状特徴量を前節と同じ処理で求め、それをサンプルポケット群の特徴量 B_1, B_2, \dots, B_k とする。抽出されたサンプルポケットを蛋白質表面上に表示した例を図 4 に示す。赤、緑、青の微小な球が HETATM を示し、紫で塗られたポケットがサンプルポケットである。なお HETATM の色は赤が酸素、緑は炭素、シアンは窒素を示す。

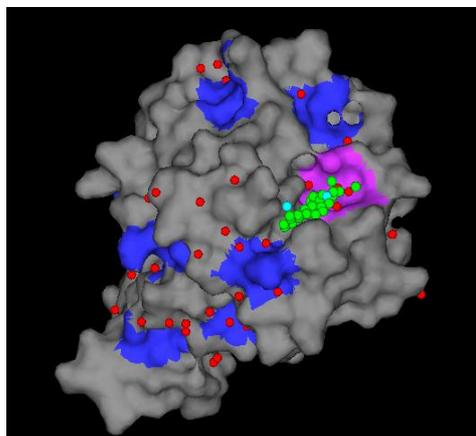


図 4 サンプルポケットの例

(2) 形状特徴量の類似度比較

続いて、評価値を算出したいポケットと各サンプルポケットを類似度比較する。評価値を算出したいポケットの特徴量 A と、サンプルポケット群の特徴量 B_1, B_2, \dots, B_k の各々に対して、式(1)を適用して類似度 S_1, S_2, \dots, S_k を求める。

$$S_i = \frac{A \cdot B_i}{\sqrt{|A|^2} \sqrt{|B_i|^2}} \dots (1)$$

そして S_1, S_2, \dots, S_k の中で最大値のもの、つまり A に最も類似する B との類似度の値を、 A の評価値 E とする。

4. 実行結果

前節までの処理で算出した各ポケットの評価値を色付けで蛋白質表面形状に表示した実行結果を図 5 に示す。図 5 では、1ezq と 1glf の 2 つの蛋白質における実行結果を示しており、ポケットの色は青から赤に向けて評価値が高くなる。

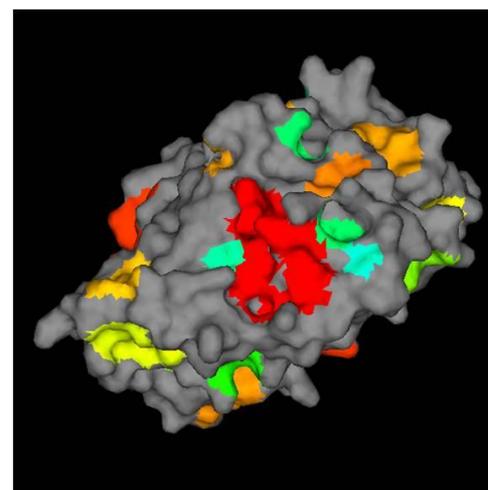
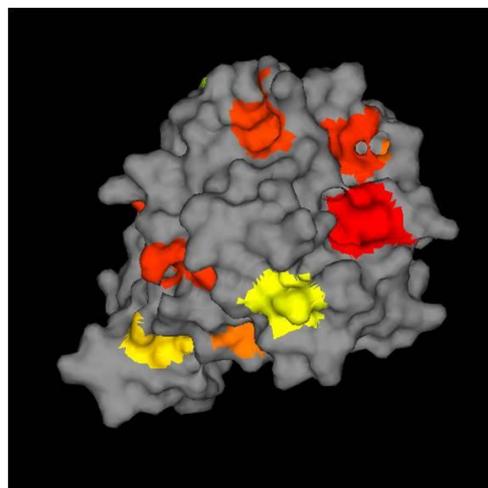


図 5 ポケットの評価値算出結果の例

5. まとめ

本報告では、蛋白質表面上の各ポケットの形状評価値を算出する手法を提案した。この形状評価値算出手法は、蛋白質の druggability 評価手法を開発するための前段階と位置づけられる。また、本研究は製薬関係の専門家と共同で進めているものであり、今後製薬の現場で実用化するために以下の点を課題としている。

1. 多数の蛋白質に対する本手法の適用と検証。
2. 特徴量算出方法に関しての検討と改善。
3. 形状評価方法に関しての検討と改善。
4. 化学的特性の考慮
5. 本研究の最終目的である蛋白質本体の druggability 評価値算出への拡張。

現在、主に 1, 2 の課題に取り組んでいる。課題 1 については現在、文献[Halgren 2009]がテストデータとして適用している約 60 個の蛋白質に本手法を適用し、結果を検証している。この 60 個の蛋白質にはすべて HETATM 情報が記載されているので、それらの周囲にあるポケットをサンプルポケット群として生成し、それに類似して見える形状のポケットに対して高評価を与えているかを検証する。また課題 2 の特徴量算出方法に関する改善点として、特徴量の次元の高さに伴う悪影響の回避が考えられる。そこで現在、特徴量算出手法の処理過程に主成分分析を用いた次元削減を加え、次元の呪いの回避を試みている。

課題 3 に関しては、ポケットの形状評価を回帰分析問題として扱いサンプルポケットの特徴量の学習させる、形状的欠点(=薬との親和性の低さ)を表すルールを定義し、それとの適合度判定を行う、などの処理を評価値算出の過程に組み込み、評価値の信頼性向上を図ることを考えている。課題 4 に関しては、eF-site が提供する分子表面形状データには疎水性と電位が記録されていることから、これらを考慮して druggability 値を補正することを考えている。

参考文献

- [Nishiyama 08] K. Nishiyama, T. Itoh; PROTEIN: A Visual Interface for Classification of Partial Reliefs of Protein Molecular Surfaces, The Institute of Image Electronics Engineering of Japan, Vol. 37, No. 3, pp. 181-188, 2008.
- [Nakamura 11] Y. Nakamura, T. Itoh; Fast Detection and Visualization of Protein Surface Pockets for Drug Design, *The 11th Asian Symposium on Visualization*, 2011.
- [Perot 10] S. Perot, O. Sperandio, M. A. Miteva, A.-C. Camproux, B. O. Villoutreix; Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery, *Drug Discovery Today* Vol. 15, Issues 15-16, pp. 656-667, 2010.
- [eF-site] K. Kinoshita; eF-site,
<http://ef-site.hgc.jp/eF-site/index.jsp>
- [Kinoshita 02] K. Kinoshita, et al.; Identification of protein functions from a molecular surface database, eF-site. *J. Struct. Funct. Genom.* Vol. 2, pp. 9-22, 2002.
- [Osada 01] R. Osada, T. Funkhouser, B. Chazelle, D. Dlbkin; Matching 3D Models with Shape Distributions, *Proc. Int'l Conf. on Shape Modeling and Application 2001*, pp. 154-166, 2001.
- [Ohbuchi 03] R. Ohbuchi, T. Minamitani, T. Takei; Shape Similarity Search of 3D Models by using Enhanced Shape

Functions, *Proc. Theory and Practice in Computer Graphics*, 2003.

[小田切 02] 小田切智, 衣嶋昌俊, 武井毅, 大淵竜太郎, 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. 48-56, 2002.

[Halgren 2009] T. A. Halgren, Identifying and Characterizing Binding Sites and Assessing Druggability, *Journal of Chemical Information Modeling*, Vol. 49, pp. 377-389, 2009.