

# 蛋白質ポケットの druggability 分析のための可視化ツール

理学専攻 情報科学コース 金子 彩香

## 1 概要

薬は人体の主成分である蛋白質に直接作用し、その効果を現わす。蛋白質の薬との親和性の高さを druggability といい、製薬の過程において druggable な蛋白質の発見は重要な課題である。また、近年の研究では、薬成分の多くは蛋白質表面上の窪み（ポケット）に反応することが知られており、druggable なポケットに対する研究が進められている。我々は製薬会社と議論を重ねた結果、ポケットの抽出・分析手法における以下の3つの要望に着目した。

[要望 1] 形状の厳密さよりも計算時間の短縮に重きを置いたポケット抽出手法が欲しい。

[要望 2] 蛋白質にヒットさせる化合物が不特定な状態でも、大雑把に良質そうな形状のポケットを抽出できる手法が欲しい。

[要望 3] ポケットの良質さと各種の特徴量との関係を可視化し、ポケット抽出の判定基準を随時更新しながら利用できる手法が欲しい。

[要望 1] を受け、Nakamura は蛋白質からポケットを高速に抽出する手法 [1] を提案した。本論文では、[1] により抽出されたポケットに対し、[要望 2][要望 3] を満たす2つの druggability 分析のための可視化ツールを提案する。

## 2 提案内容

### 2.1 本研究の全体図

図 1 は本論文で提案する手法のブロック図である。本手法ではデータベースから表面形状とその他の情報を取得し、Nakamura の手法 [1] を用いポケットを抽出する。そしてその抽出結果を可視化ツール 1 と可視化ツール 2 に転送し、可視化ツール 1 ではポケットの形状に基づく druggability 推定による評価結果を表示し、可視化ツール 2 ではポケットにおける特徴量を散布図表示する。これらの可視化結果から得られた知見を、図 1 の点線で示すとおり、ポケット抽出の判定基準として還元することでポケット抽出の改善が期待される。

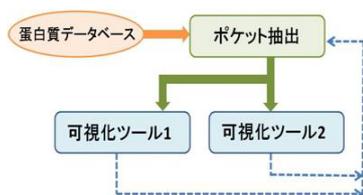


図 1: 本研究の全体図

### 2.2 蛋白質表面形状データの取得

本手法では、蛋白質表面形状データは三角メッシュで構成されていることを前提としており、eF-site[2] よりそのデータを取得している。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 形式で記述している。各頂点には、これら幾何学的情報の他に、疎水性、電位といった化学的情報も与えられている。

### 2.3 ポケットの抽出

本手法では、まず Nakamura の手法 [1] を用い、ポケットを蛋白質表面上から抽出する。処理手順を以下に示す。

1. 蛋白質表面形状にメッシュ単純化を適用
2. 凹凸判定により表面上の窪み部位を抽出
3. メッシュ単純化前の表面形状に窪み部位を投影

これらのステップで抽出された部位をポケットとする。

### 2.4 ツール 1:形状による druggability 推定

#### 2.4.1 形状特徴量の算出

2.3 節によって抽出された各ポケットに対し、形状特徴量を算出する。以下にその処理手順を示す。

1. ポケットの外周を近似する平面  $P_o$  とその中心  $C_{p_o}$  を算出し、 $C_{p_o}$  から法線  $N_{p_o}$  を伸ばす
2.  $P_o$  と  $N_{p_o}$  をそれぞれ  $m$  区間、 $n$  区間に分割する
3. ポケットの表面上に点群を生成する
4. 各点における  $P_o$  からの距離（深さ）と  $N_{p_o}$  からの距離（幅）のヒストグラムを生成する
5. ヒストグラムの各階級の度数を各次元の値とする  $m \times n$  次元ベクトル生成する

この  $m \times n$  次元ベクトルを各ポケットの形状特徴量とする。

#### 2.4.2 druggability 推定値の算出

続いて、あらかじめ druggable であることがわかっているサンプルポケットの形状特徴量群を生成し、ポケットの druggability 推定値を算出する。我々の実装では、PDB に登録されている蛋白質情報内にある非蛋白質原子が密集している部位近辺にあるポケットをサンプルポケットとしている。これは、非蛋白質原子が残っている部分は反応が起きやすいという報告に基づいている。そして、推定値を算出したいポケットとこれら各サンプルポケットの形状特徴量とのコサイン類似度を算出し、その中で値が最大のものをそのポケットの druggability の推定値とみなす。

## 2.5 ツール 2: druggability 基準分析

図 2 は、可視化ツール 2 の表示例である。軸選択領域で、多数のポケットの特徴量の中から散布図の軸に割り当てる 2 項目を選択し、選択された 2 項目についての散布図を画面の左側に表示する。ここで散布図の各プロットはポケットを示しており、赤は druggable、青は undruggable なポケットである。各プロットをクリックすると、右側の画面に対応するポケットの形状が表示される。散布図の軸に割り当てることが可能な各特徴量は以下に示す。

- ポケットの深さ、幅
- 電位の平均・分散
- 疎水性の平均・分散
- 三角形数
- $s$  Å 以上の深さの三角形数
- 電位の値が  $t$  以下 (以上) の三角形数
- 疎水性の値が  $u$  以下 (以上) の三角形数

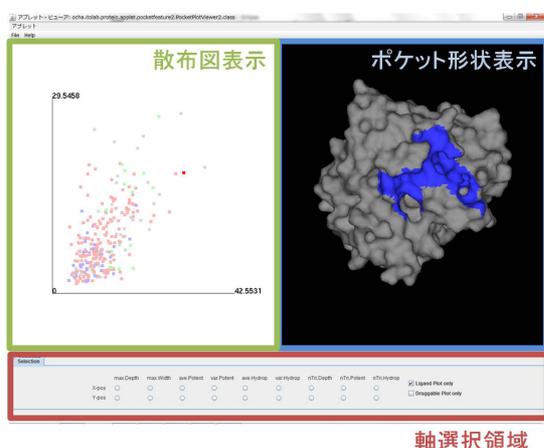


図 2: 可視化ツール 2

## 3 実行結果

### 3.1 ツール 1: 形状による druggability 推定

図 3 は、2.4 節で算出したポケットの druggability 推定値を蛋白質表面上で色表示した結果の一例である。蛋白質は ID:1ke8 は使用している。各ポケットの色は青から赤に向けて推定値が高くなることを意味する。また、蛋白質表面上の微小な球は非蛋白質原子である。結果を観察すると、ポケットが赤く色づけされている部位が 3ヶ所ある。画面右側のポケットは非蛋白質原子が密集しているサンプルポケットの部位であるが、左側の 2つのポケットは非蛋白質原子が密集していないにも関わらず、評価値が高いと算出された。この可視化結果は、新たな druggability の可能性のある部位の発見を示唆している。

### 3.2 ツール 2: 形状による druggability 推定

2.5 節の可視化ツール 2 を使用し、特徴情報と druggability の関係性を観察した結果を示す。使用したポケットは、Halgren らが文献 [3] にて使用した 60 個の蛋白質から抽出した 274 個のポケットである。図 4 は、散布図の x 軸に「電位が -0.1 以下の三角形数」を、y 軸に「疎水性が -0.5 以下の三角形数」を割り当てた結果である。この結果から 2 軸にわりあてた 2 変数には、正の相関性と、y 軸上に値が集中するという、2つの傾向が見られた。また、それらのポケットを調べると、蛋白質 1c14, 1eet, 1e1x などのポケットの多くは y 軸上に存在し、蛋白質 1ptw, 1oyn, 1udu などのポケットの多くは正の相関性上に存在した。1つの蛋白質から抽出されたポケットに両方の傾向が混在する例はなく、蛋白質の種類によってどちらか一方の傾向に属するという推測がたてられた。

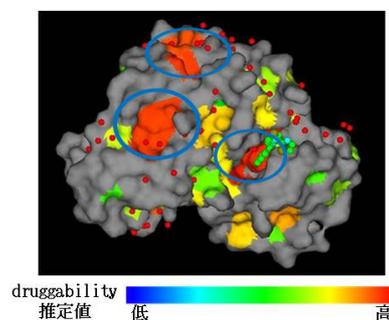


図 3: 可視化ツール 1 の実行結果例

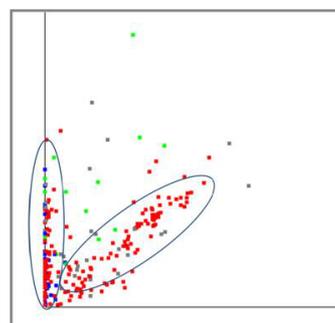


図 4: 可視化ツール 2 の実行結果例

## 4 まとめ

蛋白質ポケットの druggability 分析を目的として、2つの可視化ツールを提案した。今後の展望としては、2.1 節で示したようにこれら可視化ツールで発見した基準をポケット抽出の判定基準として還元することで、本手法をより信頼度の高いものに発展させたい。

## 参考文献

- [1] Y. Nakamura, T. Itoh, Fast Detection and Visualization of Protein Surface Pockets for Drug Design, The 11th Asian Symposium on Visualization, 2011.
- [2] eF-site, <http://ef-site.hgc.jp/ef-site/index.jsp>
- [3] T. A. Halgren, Identifying and Characterizing Binding Sites and Assessing Druggability, Journal of Chemical Information and Modeling, 49(2), 377-389, 2009.