

蛋白質ポケットの druggability 分析のための可視化ツール

金子彩香 中村友香理 伊藤貴之

お茶の水女子大学大学院 人間文化創成科学研究科

Visualization for Druggability Analysis of Protein Pockets

Ayaka Kaneko Yukari Nakamura Takayuki Itoh

Graduate School of Humanities and Sciences, Ochanomizu University

{ayaka, sincere, itot} @ itolab.is.ocha.ac.jp

アブストラクト

薬は人体の重要な構成成分である蛋白質に作用する。薬との親和性の高さを druggability といい、親和性の高い蛋白質の発見は製薬過程における重要な課題である。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、我々も druggable なポケットに対する抽出・分析の研究を進めてきた。本論文では、蛋白質表面形状に対し、ポケットを抽出する手法、並びに、ポケットの druggability 分析のための 2 種類の可視化ツールを提案する。ポケット抽出手法では、三角メッシュで構成された蛋白質表面上に対し、メッシュ単純化により微小凹凸を平滑化し、その表面上から凹部をポケットとして抽出する。druggability 分析のための 1 つ目の可視化ツールでは、形状の観点から各ポケットの druggability 推定値を算出し可視化する。2 つ目の可視化ツールでは、各ポケットの特徴量を 2 軸に割り当てた散布図とポケット形状を 1 画面上に表示する。本論文では 2 つの可視化ツールを 60 種類の蛋白質に適用し、適切と思われる druggability 推定値が可視化されたことや、ポケットの電位や疎水性といった化学的特性と druggability との関係性を散布図で可視化されたことを示す。

キーワード 蛋白質, ポケット, druggability, 抽出, 分析

Abstract

Research of protein is pivotal to the drug discovery, since most of the drugs act upon proteins inside human body. Drugs act when they are close to the concave portions, so called “pockets”, of the protein surfaces. Therefore, detection and analysis of the pockets are also important for the drug discovery. This paper presents a fast method with pocket extraction and evaluation technique for the protein surfaces. When the protein surfaces are provided as triangular meshes, the method first applies mesh simplification to smoothing small geometric features. It then detects concave portions as pockets from the simplified triangular meshes. The method then evaluates the pockets from the following viewpoints: geometry and chemistry. This paper introduces our case study which applied the presented technique to 60 proteins, and successfully visualized appropriate druggability estimation and correlation between chemical properties and druggability.

Keyword protein, pocket, druggability, extraction, analysis

1. はじめに

蛋白質は20種類のアミノ酸で構成される高分子化合物であり、その種類・機能は極めて多種多様である。また、蛋白質は生命体を構成する細胞の主成分であり、生命維持の重要な役割を担うことから、様々な分野で注目されている。近年では計算機技術の発達に伴い、蛋白質を計算機によって分析する学術分野が確立され、活発に議論が進んでいる。

薬品開発の分野においても蛋白質の研究は非常に重要視されている。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。主に実験によって導かれる蛋白質と薬との親和性の高さを *druggability* といい、*druggability* が高いと評価される蛋白質をしばしば「*druggable* である」という。創薬のターゲットを特定する過程において、*druggable* な蛋白質・*druggable* な蛋白質部位の発見は重要な課題である。創薬過程で行われる計算機を用いた蛋白質分析技術として、構造既知の蛋白質を鋳型として構造未知の蛋白質構造を予測するモデリングや、蛋白質や化合物の構造情報に基づくドッキングシミュレーションにより創薬ターゲットに対し活性を持つ化合物を選別するパーチャルスクリーニングなどがあげられ、これらの工程に対して今日までに様々な手法、ツール、データベース等が考案されてきた。その中の一つとして、蛋白質ポケットについての研究分野がある。ポケットとは、蛋白質表面に存在する窪み形状の部位の総称である。近年の研究により、薬成分の多くはこのポケットに入り込み反応することが報告されている。このことから、蛋白質ポケットの抽出・分析は非常に重要であると考えられ、これらについてもこれまでに多くの手法が報告されてきた。

我々は製薬会社および計算生物学の専門家と議論を重ねた結果、蛋白質ポケットの抽出・分析において以下のような要望に着目した。

[要望 1] 従来のポケット抽出・分析手法は、計算時間が膨大であるという問題点があった。また、蛋白質は分子運動により形が常に変化する。そこで、形状の厳密さは重要視せず計算時間の短縮に重きを置いた手法を確立させたい。

[要望 2] ポケット形状の良質さは本来なら、蛋白質にヒットさせる化合物の形状が特定されてから決定されるものであるが、現実には化合物が特定されない状態でポケットを抽出する場合もある。そこで、そのような場合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] ポケットの良質さの基準は曖昧であり、それぞれの手法が主観によって各種の特徴量（幅、深さ、アミノ酸配列など）を用いている。そこで、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。

これらの要望を受け本論文では、蛋白質表面上からポケットを高速に抽出し、そのポケットを形状に基づいて分析する可視

化ツール1、さらには各ポケットが持つ特徴量と *druggability* の関係性の分析を行う可視化ツール2を提案する。

本手法では、まず、三角メッシュで構成された蛋白質表面上をメッシュ単純化により平滑化し、ポケットとなる表面上の窪みを抽出する。メッシュ単純化を適用することで、蛋白質の微細な形状の特徴を排除し、本質的な形状の特徴のみからポケットを抽出することができる。しかも、頂点や三角形の数を低減したメッシュでの抽出処理によって、計算量の増加を抑えている。以上から上述の[要望 1]を満たすと考えられる。

可視化ツール1では、抽出されたポケットに対し、形状に基づいて *druggability* を推定し、その結果を可視化する。本ツールでは抽出したポケットの形状に対して、既に *druggable* と判断されている蛋白質から抽出されたサンプルポケットの形状との類似度から、*druggability* の推定値を算出する。このツールにより、ポケットにヒットする化合物を特定することなく、形状的な観点から薬との親和性が高いと推定されるポケットがどのような形状をしているかを視覚的に観察できる。以上から上述の[要望 2]を満たすと考えられる。また抽出されたポケットに対して、このツールを用いて形状的な良質さを判定しておけば、例えば創薬過程のドッキングシミュレーションにおいて、蛋白質表面全体をシミュレートするのではなく、良質な形状を有するポケットの周辺のみを対象を絞ってシミュレーションを進めることができるため、作業工程の短縮が期待される。

可視化ツール2では、各ポケットの特徴量を2軸に割り当てた2次元の散布図を表示し、さらに散布図上のマウス操作で特定したポケットの形状を同時に表示する。本ツールを用いることで、特徴量と *druggability* の関係性を直感的に分析し、ポケット抽出の判定基準の検討に貢献できると考える。以上から上述の[要望 3]を満たすと考えられる。現在、特徴量がどのような数値基準を満たすポケットが *druggable* なポケットであるかについて統一した基準はない。また、*druggable* なポケットとそうでないポケットを特徴量で判別分析することは非常に難しく、研究組織ごとに用途や方針が異なることなどからも、基準分析を全自動化することは簡単ではないと考える。そこで本ツールでは、分析支援のために可視化を用いることで、各ユーザーの主観的判断に基づく基準分析を支援できると考える。

2. 関連研究

2.1 蛋白質の *druggability* に関する研究

蛋白質の *druggability* に関する研究は近年急速に発展しており、そのサーベイ論文もいくつか発表されている。その代表例としてPerotらは、*druggable* な蛋白質ポケット・結合部位の抽出や類似度比較評価、その結果公開のデータベースをサーベイした論文を発表している[1]。ここでサーベイされているポケット抽出手法は、主に幾何学的観点から抽出する手法と化学的構造から抽出する手法の大きく2種類に分けられる。本手法は幾何学的観点をを用いた手法に分類される。幾何学的観点をを用いた手法には、格子を用いて蛋白質分子との衝突判定により窪んだ形状を探索するPocketDepth[2]や、蛋白質表面に小さな球を転がし、それが

入り込むか否かで窪みを探索するGHECOM[3]などがあげられる。また、結合部位の類似度比較の研究も、異なる2つの蛋白質が幾何学的かつ化学的な類似性を共有していた、という事例から注目を集めている。その代表的な手法として、蛋白質構造に幾何学的ハッシュ法を適用し類似性の高い部位を予測するMultiBind[4]や、蛋白質分子表面形状の各頂点に対しクリーク探索法を適用するeF-seek[5]などがあげられる。

2.2 蛋白質表面形状分析手法 PROTEIN

蛋白質表面にポリゴン処理を施すという観点から、Nishiyamaらは蛋白質表面形状の局所凹凸形状の抽出・分類・観察のための手法およびユーザインターフェースPROTEIN (Partial Relief Observation Technique and Interface) [6]を提案している。この手法は、蛋白質間の反応は分子表面形状に大きく依存するという報告から分子表面形状を用いており、また、他の化合物と合致する基本部分形状となる局所凹凸形状に着目している。まず、蛋白質の分子表面形状から局所凹凸形状部位を抽出し、各部位の形状特徴量を算出する。そして、その算出結果をクラスタリングし可視化する。また、局所凹凸形状の隣接関係にも着目し、類似部分形状を抽出するために分子表面形状をグラフ化し、生成されたグラフ経路の比較観察を行っている。

局所的な形状把握は、他の化合物との結合の際に非常に重要な情報となる。しかし、その局所部分が実際に他の化合物と反応するかどうかは、さらに広域な凹部形状、つまりはポケットの把握が必要となる。そこで本手法では、この局所凹凸形状分類の考え方をポケット形状に特化させ、ポケットの抽出・分析手法に拡張している。

2.3 3次元モデルの形状類似度比較手法

本手法では、可視化ツール1のポケットのdruggability推定値を決定する過程において、ポケットの形状特徴量を算出し、形状類似度比較を行う。3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、それらは、モデルの幾何学的形状から特徴量を求める手法、モデルの位相情報を用いる手法、モデルの投影後の外観に基づく手法の三種類に分類できる。特にモデルの幾何学的形状から特徴量を求める手法が多く、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法[7]、それをさらに拡張させたAAD法[8]がある。これらの手法は、モデルの位置や向きを正規化を必要とせずにモデルの形状特徴量を算出することができる。その他にも、モデルの慣性主軸に沿った包絡線を用いたPS法[9]などもある。また、位相情報を用いる手法にはHilagaらの手法[10]、見掛けに基づいた手法にはLFD法[11]などが上げられる。

本手法はポケット形状の特徴量算出の際、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。そこで本手法では、D2法やPS法、また前節で紹介したPROTEINで用いられている特徴量算出方法を参考にした形状特徴量算出手法により、多次元ベクトルでポケット形状を表現している。

3. 本手法の概要

本章では提案手法の概要を述べる。図1は本論文で提案する手法のブロック図を示したものである。本手法ではデータベースから表面形状その他の情報を取得し、そこからまず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。また、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。

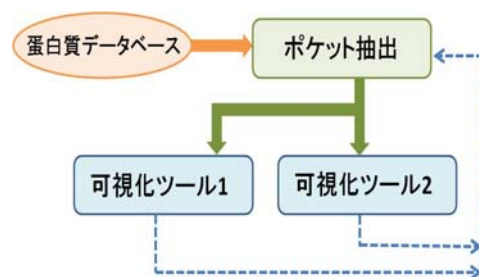


図1. 提案手法の概要

本手法では、主に以下2つの蛋白質データベースから収集されたデータを使用する。抽出・分析で用いる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[12]よりそのデータを取得している。また、可視化ツール2で用いる疎水性・電位の情報はeF-siteより取得しており、可視化ツール1で用いる非蛋白質原子の情報はPDB[13]より取得している。

eF-site

eF-siteでは、PDBに登録されている蛋白質のうち数十万種類について、コノリ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。各頂点には、頂点座標、法線ベクトル、最大曲率、最小曲率などの幾何学的情報の他に、各頂点には、疎水性、電位といった化学的情報も与えられている。

PDB (Protein DataBank)

PDBとは生体高分子の立体構造座標に関する世界的なデータベースである。PDBには、蛋白質のアミノ酸配列や原子構造情報はもちろんのこと、蛋白質には直接関係ないが蛋白質に付随している非蛋白質原子の情報なども含んでいる。PDBに蓄積されている構造データは、X線結晶解析法、NMR法（核磁気共鳴法）、単粒子電子顕微鏡観測法などにより実験的に決定されたデータであり、理論的な予測（蛋白質構造予測）で推定されたデータは蓄積していない。

4. ポケット抽出

本手法では、まず、ポケットを蛋白質表面上から抽出する。本手法のポケット抽出は、eF-siteで取得した蛋白質表面形状データに対し、メッシュ単純化、凹凸判定、凹部投影、の3つの処理を行う。以下、各処理の詳細を示す。また、蛋白質1ezqを適用したこれら一連の処理の流れを図2に示す。

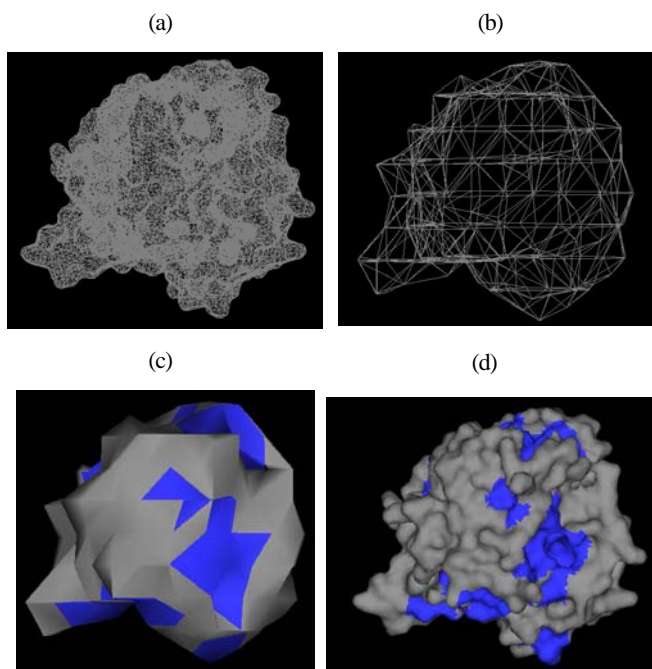


図2. 蛋白質 (ID :1ezq) の表面形状データとポケット抽出の実行例

(a)蛋白質表面形状データ (b)メッシュ単純化
(c)凹部形状の抽出 (d)もとの三角メッシュに投影

4.1 メッシュ単純化

メッシュ単純化は、まず蛋白質表面上を包括するような3次元格子を生成する。次にその各格子点にスカラ値を付与し、最後にスカラ値を元に格子上に新たな分子表面を形成する頂点を決定する。以下にそれら各処理の詳細を示し、全体の流れを図3に示す。メッシュ単純化を適用することにより蛋白質表面上は微小な凹凸形状は除外され、大きな特徴的凹凸形状のみを把握することが可能となる。

4.1.1 3次元格子の生成

まず、蛋白質表面形状を包括するような3次元格子を生成する。蛋白質表面を構成する頂点座標の最大値と最小値を算出し、その最大値から最小値までの間に等間隔に格子点を生成していく。この時、格子点および格子線の間隔を調整することで、蛋白質の度合いを変えることが出来る。

4.1.2 スカラ値の付与

次に、生成した各格子点にスカラ値を付与する。各格子点において、蛋白質表面形状を構成する各格子点との距離を算出する。その距離の最小値をその格子点のスカラ値とする。各格子

点において、一番距離の小さい蛋白質表面の頂点を始点とし、格子点を終点とするベクトルを求め、それらのベクトルに対し、法線ベクトルとの内積を求める。この時、蛋白質表面を陰関数曲面に見立て、内積が負の値、つまり格子点が蛋白質表面内部に含まれると判断された場合、格子点に付与されたスカラ値を負の値に変換する。こうして、蛋白質分子表面内部には負、外部には正のスカラ値が付与される。

4.1.3 等値面の生成

付与されたスカラ値をもとに、格子辺上に新たな蛋白質表面を形成する頂点を決定する。格子辺の両端にあたる2格子点のスカラ値をそれぞれ a および b とする。蛋白質内部の格子点には負の値、外部の格子点には正の値が付与してあるため、蛋白質表面は内部点と外部点にはさまれた格子辺上にあると仮定される。そこで、 $a \times b < 0$ となる格子辺を $|a|:|b|$ に内分する点を求め、これを等値面を構成する頂点とする。このような頂点をいくつか抽出した後、マーチングキューブ法を適用して頂点を連結することで、格子内部に多角形を生成する。以上の処理を各格子に適用することで等値面を生成する。

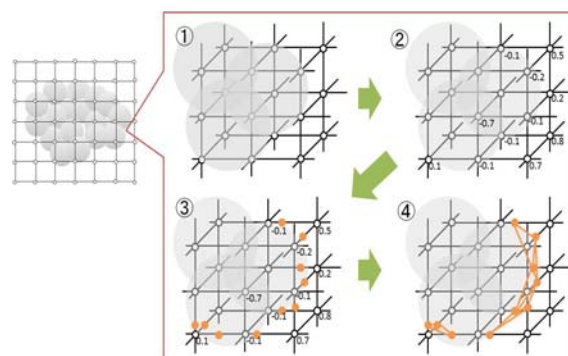


図3. メッシュ単純化の処理手順

①3次元格子の生成②スカラ値の付与
③④等値面の生成

4.2 凹凸判定

次に、前節の処理により平滑化された蛋白質表面形状に対し、ポケットとなりうるような窪み形状の部位(凹部)を抽出する。まず、平滑化された蛋白質表面形状を構成する各頂点 v_i について、位置を (x_i, y_i, z_i) 、法線ベクトルを $(n_{x_i}, n_{y_i}, n_{z_i})$ とする。この時、 v_i における接平面は以下の式(1)において $t = 0$ を満たす面である。

$$t = n_{x_i}(x - x_i) + n_{y_i}(y - y_i) + n_{z_i}(z - z_i) \quad \cdots (1)$$

ここで本手法では v_i と稜線を共有する隣接頂点 (x, y, z) に対して、式(1)により判別値 t を算出する。本手法では、すべての隣接頂点に対して t が負であるならば、 v_i は「凹」の属性を付与する。この凹の属性を与えられた部位がポケット候補の部位となる。それ以外の頂点においても同判別値 t を算出し、すべての隣接頂点に対して t が正ならば、その頂点は「凸」とし、それ以外は「その他」属性を付与する (図4参照)。

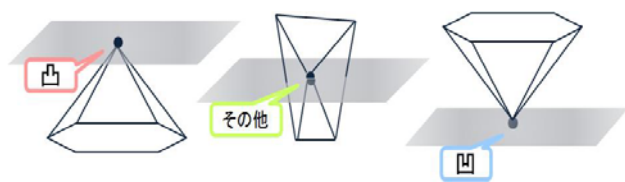


図 4. 凹凸判定

(左)凸の属性(中央)その他の属性(右)凹の属性

続いて、三角形にも同様に属性を付与する。平滑化された蛋白質表面を構成する各三角形において、その三角形を構成する3つの頂点のうち、1つでも「凹」の属性を持つ頂点があれば、その三角形の属性は「凹」とする。「凹」の属性を持つ部位がポケット候補の部位となる。

4.3 凹部投影

次に、前節で凹の属性を付与されたポケット候補の部位を単純化前の三角メッシュではどの部位に対応していたか特定する。単純化後の三角メッシュの各三角形の重心を G_1 とし、単純化前の三角メッシュの各三角形の重心を G_0 とする。ここで、蛋白質全体の中心点 O から G_0 および G_1 に結んだ線分の内積 p を式(2)により算出する。

$$p = \frac{(G_0 - O) \cdot (G_1 - O)}{\sqrt{(G_0 - O)^2} \sqrt{(G_1 - O)^2}} \quad \dots (2)$$

この値が1に近いとき、両線分の蛋白質重心からの方向はほぼ等しいことになる。そこで単純化後の各三角形 T_1 について、単純化前の三角形 T_0 との間に p の値を算出し、その値が1に非常に近い三角形を列挙する。そして、列挙された三角形 T_0 の中で G_1G_0 間の距離が最も近い三角形を選出し、この三角形に単純化後の三角形と同じ凹の属性を付与する。以上の処理により単純化前の三角メッシュ上のいくつかの三角形に凹の属性が付与される。この属性を持つ三角形の集合で構成される部位を、抽出された「ポケット」と定義する (図5参照)。

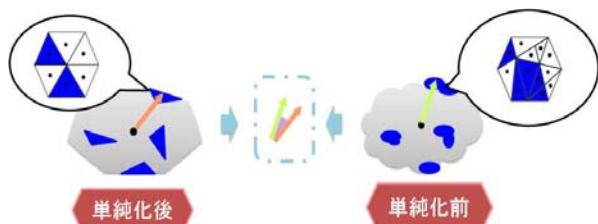


図 5. 凹部投影

5. ツール 1 : 形状からの druggability 推定

1つ目の可視化ツールは、蛋白質ポケットの形状特徴に基づいたdruggability推定値を可視化する。本ツールでは前処理として、あらかじめdruggableであることがわかっている既知のポケットをサンプルポケットと位置づけ、これらの形状特徴量をデータベース化する。そして、4章で提案した手法で抽出した各ポケットに対して形状特徴量を算出する。これらの形状特徴量の類似

度を、抽出された各ポケットのdruggability推定値とみなし、その結果を可視化する。この可視化により、蛋白質表面上のどのような形状のポケットがdruggableであると推定されるかを、視覚的に把握することが容易になる。

5.1 ポケットの形状特徴量算出

4章の処理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。本手法の特徴量は、薬成分の入り込みやすさに影響を及ぼす「深さ」「口の広さ」に基づいた多次元ベクトルとなっている。以下にその処理手順を示す。

1. ポケットの外周面を近似する平面 P_0 を算出し、その平面の中心 C_{p_0} からの平面 P_0 の法線ベクトル N_{p_0} を伸ばす
2. P_0 と N_{p_0} を C_{p_0} から等間隔にそれぞれ m 区間、 n 区間に分割する
3. ポケットを構成する三角メッシュ上に等密度に点群を生成する
4. 各点における P_0 からの距離(深さ)と N_{p_0} からの距離(幅)のヒストグラムを生成し、各階級の度数を各次元の値とする $m \times n$ 次元ベクトルを形状特徴量とする

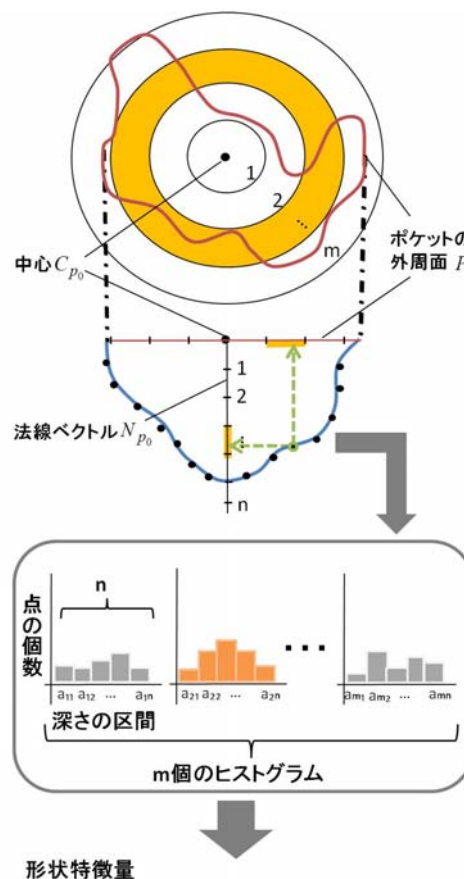


図 6. 形状特徴量算出の処理手順

図6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。なお、手順2における各区間の間隔と区間数は、現段階の実装では、経験的に、区間

の間隔は1Å, 区間数は $m = 16, n = 16$ とした。

5.2 ポケットの druggability 推定値算出

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, データベースから検索する。そして当該ポケットの形状特徴量と各サンプルポケットの形状特徴量との類似度の最大値を求める。この値を, 当該ポケットの druggability の推定値とみなす。以下サンプルポケット群生成と類似度比較についての詳細を示す。

5.2.1 サンプルポケット群の生成

サンプルポケットの収集には druggability を判定する実験が必要であるが, 我々自身でその実験を実施するのは困難である。そこで現時点の我々の実装では, 専門家からアドバイスを受けた妥当性のある代替手段として, PDB に登録されている蛋白質情報から druggable と推測されるポケット群を収集している。PDB に登録されている蛋白質の中には, 蛋白質自身の構成には直接関係ない「非蛋白質原子」と呼ばれる原子の情報が記載されているものがあり, HETATM というレコードで記録されている。これら非蛋白質原子は主に蛋白質を結晶化する過程で表面上と結合し残留したものであり, 非蛋白質原子が残っている部位は反応が起きやすいといわれている。そこで我々の実装ではこの前提に基づき, 非蛋白質原子の集中部位周辺にあるポケットを druggable なポケットと仮定し, それらをサンプルポケットとして収集する。そして収集した多数のサンプルポケットに対して, 前節と同じ処理で形状特徴量を求め, これをデータベース化する (図7参照)。

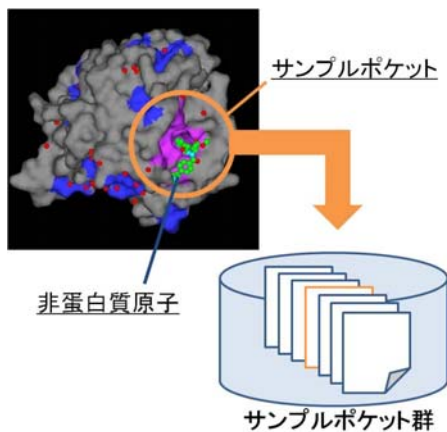


図7. サンプルポケット群の生成

5.2.2 ポケットの形状類似度比較

続いて, サンプルポケットとの形状類似度比較により, ポケットの druggability 推定値を算出する。推定値を算出したいポケットの形状特徴量を A , 収集された k 個のサンプルポケット群の各形状特徴量を B_1, B_2, \dots, B_k とすると, ポケット A に対し, サンプルポケット B_1, B_2, \dots, B_k 各々とのコサイン類似度 S_1, S_2, \dots, S_k を式(3)より求める。

$$S_i = \frac{A \cdot B_i}{\sqrt{|A|^2} \sqrt{|B_i|^2}} \quad \dots (3)$$

そして, S_1, S_2, \dots, S_k の中で値が最大のものをポケット A の druggability 推定値とする。この推定値を色表示することで, 形状に基づくポケットの評価結果を可視化する (図8参照)。

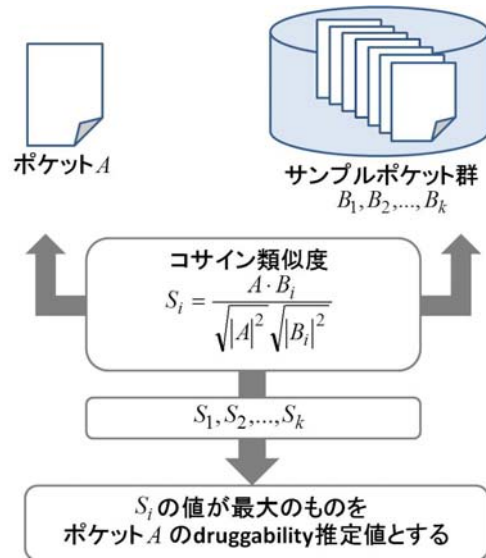


図8. druggability 推定値の算出手順

6. ツール2 : druggability 基準分析

2つ目の可視化ツールでは, ポケットの druggable 判定の有効な基準を発見することを目的として, 各ポケットが持つ特徴量と druggability の関係性を散布図として可視化する。

図9は, 可視化ツール2の操作画面である。操作画面は3つの領域に分かれており, 左側は散布図表示領域, 右側はポケットの形状表示領域となっている。また, 下側の領域は, 特徴量選択領域となっており, 多数のポケットの特徴量の中から散布図の2軸に割り当てる2項目を選択すると, 選択された2項目についての散布図が散布図表示領域に表示される。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 なポケット, 青は undruggable なポケット, 緑は druggable でも undruggable でもない中間のポケットである。各プロットをクリックすると, 右側の形状表示領域に対応するポケットの形状が表示される。このように, 様々な特徴量を割り当てた散布図およびポケット形状を同時に表示・観察できることから, 本可視化ツールは各特徴量と druggability 間の複雑な相関関係の発見の手助けとして期待できる。

現段階で我々が実装している特徴量を以下の(1)~(8)に示す。なお, 以下(6)から(8)にて使用する変数 s, t, u はユーザの必要に応じ値を調節することが可能である。

- (1)深さ: 各ポケット表面を構成する各頂点から平面 P_0 に向かって法線ベクトル N_{p_0} と平行なベクトルをそれぞれ伸ばし, その中で最も長いベクトルを持つ点を, そのポケットの最深部とし, その長さをポケットの深さと定義する。
- (2)幅: 各ポケットの平面 P_0 上に点群を生成し, また, ポケッ

ト外周上の各頂点を P_0 上に投影する。 P_0 上の各点に対し、外周の頂点との距離が最小なものを算出し、その中で最も距離の値が大きいものをそのポケットの幅の値と定義する。

(3)電位の平均・分散：電位の情報はeF-site上にて三角メッシュの各頂点に付与されている。これを利用し、ポケット全体の電位の平均値・分散値を算出する。

(4)疎水性の平均・分散：電位と同様にeF-site上の情報を利用し、ポケット全体の疎水性の平均値・分散値を算出する。

(5)三角形数：各ポケットの形状を構成している三角形の総数。ポケットの表面積に比例する値として扱う。

(6) $s \text{ \AA}$ 以上の深さの三角形数：ポケットを構成する三角形のうち、平面 P_0 から $s \text{ \AA}$ 以下の深さにある三角形の総数。ポケット中の一定の深さを有する部位の面積に比例する値として扱う。

(7)電位の値が t 以下 (以上) の三角形数：ポケットを構成する三角形のうち、3頂点における電位の平均値が t 以下または t 以上の三角形の総数。電位の値が特定の条件を満たす部位の面積に比例する値として扱う。

(8)疎水性の値が u 以下 (以上) の三角形数：ポケットを構成する三角形のうち、3頂点における疎水性の値が u 以下または u 以上の三角形の総数。疎水性の値が特定の条件を満たす部位の面積に比例する値として扱う。

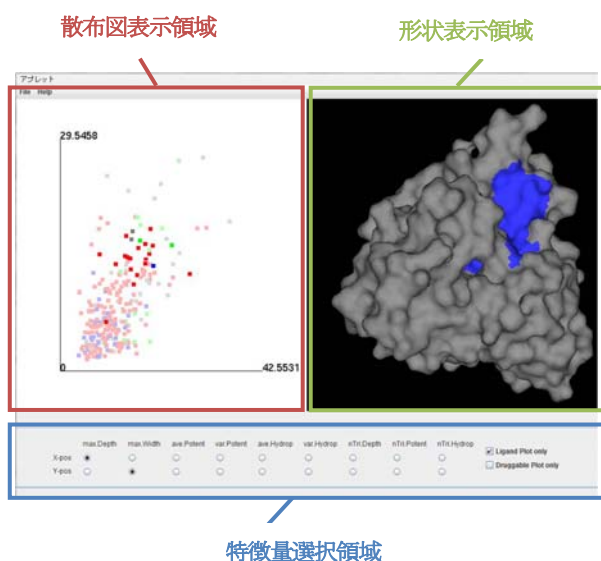


図 9.可視化ツール 2 の操作画面

これらの特徴量を実装している背景は以下のとおりである。eF-site にて蛋白質表面上の各頂点に登録されている情報は、幾何的情報の他に、化学的情報として電位と疎水性がある。我々の実装では表面形状に関する入力情報は以上に限定されることから、幾何的情報から算出される(1)(2)(5)(6)、化学的情報から算出される(3)(4)(7)(8)を、我々自身の経験に基いて特徴量に採用した。また(1)~(5)の特徴量はポケット全体に関する特徴量として採用し、(6)~(8)はポケット表面上の特定の条件を満たす部位の面積に関する特徴量となっている。なお eF-site から入手できる表面形状は均一な大きさの三角形の集合で表現されている

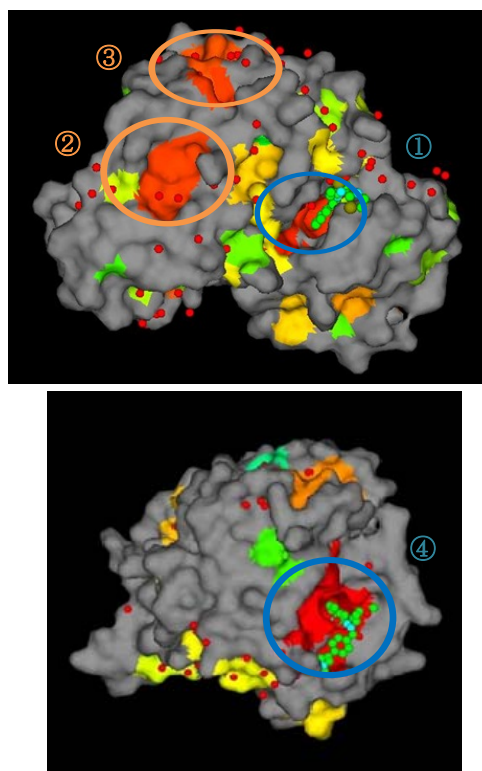
ことから、現時点での我々の実装では面積を三角形の個数に置き換えて扱っている。

7. 実行結果

7.1 ツール 1：形状からの druggability 推定

図 10 は、5 章の処理で各ポケットの druggability 推定値を算出し、蛋白質表面上で色表示した実行結果の一例である。図 10(上)が蛋白質 ID:1ke8 に適用した実行結果であり、図 10(下)が蛋白質 ID:1ezq に適用した実行結果である。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。オレンジまたは赤のポケットは、サンプルポケット群のどれか一つと非常に形状が似ていることを意味する。また、蛋白質表面上の微小な球は非蛋白質原子である。これら結果を見ると、双方とも非蛋白質原子が密集している部位付近のポケット (図 10(上)の①と(下)の④) が赤く色づけされており、正しく推定値の計算が行われていることが確認できる。

さらに図 10(上)の 1ke8 による実行結果を観察すると、ポケットがオレンジまたは赤に色づけされている部位が 3 ヶ所ある。その内①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、②と③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、推定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。



druggability 推定値 低 高

図 10. 可視化ツール 1 による実行結果

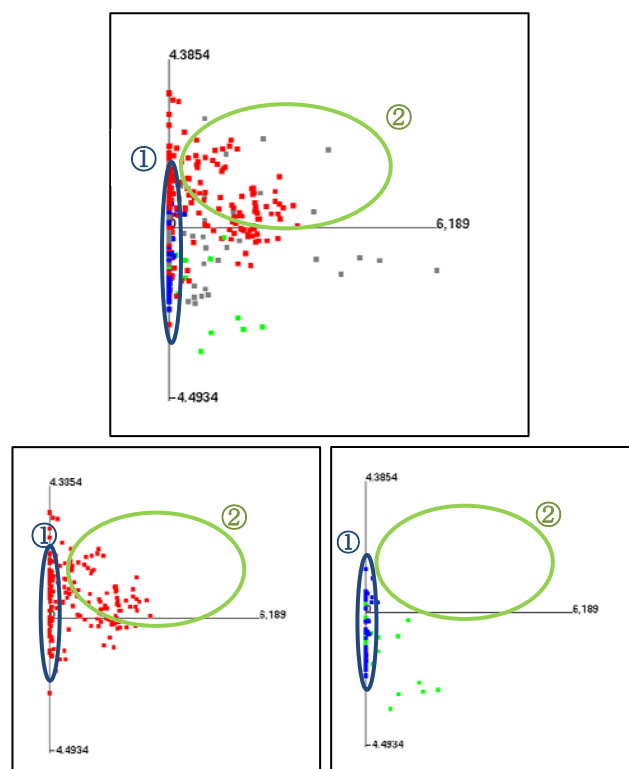
(上)蛋白質 ID : 1ke8 に適用した結果
(下)蛋白質 ID : 1ezq に適用した結果

7.2 ツール2: druggability 基準分析

6章の可視化ツールを使用し、特徴量と druggability の関係性を観察した。中でも散布図上で顕著な関係性を観察できた特徴量について、結果を以下に示す。使用したポケットは、Halgren らが文献[14]にてテストデータとして使用した 60 個の蛋白質から抽出した 274 個のポケットである。

観察結果1

図 11 は、散布図の x 軸に「電位が-0.1 以下の三角形数」を、y 軸に「疎水性の平均」を割り当てた結果である。この結果を観察すると、まず、undruggable なポケットが y 軸上、つまりは、電位が-0.1 以下の三角形数が 0 である部分 (図 11 の①) に集中していることが観察できる。また、druggable なポケットは散布図全体に散らばっているが、druggable 以外のポケットは、y 軸上を除いた y の値が正の領域、つまり、電位が-0.1 以下の三角形数と疎水性の平均が共に正である領域 (図 11 の②) には存在していないことが観察できる。これらのことから、この 2 軸において、druggable に関する相関性は観察できなかったが、undruggable なポケットに対しては、そのポケットの表面は電位が-0.1 以下の値を示す部位がないことが多いという推測が立てられた。また、druggable 以外のポケットで電位が-0.1 以下の値を示す部位がある場合、そのポケットは疎水性の平均の正にはなりにくいという傾向も推測出来た。



電位が-0.1 以下の三角形数 (x 軸) 疎水性の分散 (y 軸)

図 11. 散布図の観察結果(1)

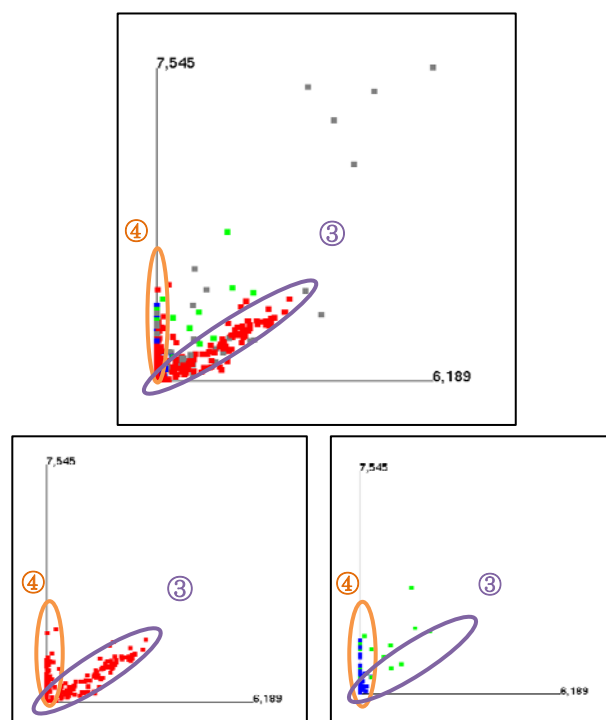
(上) すべてのポケットを表示

(左下) druggable のみ表示 (右下) druggable 以外表示

観察結果2

図 12 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。この結果を観察すると、この 2 軸に割り当てた 2 変数には、正の相関 (図 12 の③) と y 軸上 (図 12 の④) に値が集中するという 2 つの傾向を観察することができた。また、それら傾向を構成するポケットを調べると、蛋白質 lptw, loyn, ludu, などから抽出されたポケットの多くは正の相関上に存在し、蛋白質 lc14, leet, le1x などから抽出されたポケットの多くは y 軸上に存在した。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かった。以上のことから、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立てられた。

また、undruggable なポケットは y 軸上に多く観察でき、正の相関上には undruggable なポケットは観察できなかった。この結果から、例えば、薬品開発の過程において、undruggable なポケットの混在を許してでもできるだけ多くの druggable なポケットを抽出したい場合は、正の相関と y 軸の 2 つの相関を抽出条件として追加し、逆に、undruggable の混在は許さず druggable なポケットだけを厳選して抽出したい場合は、正の相関のみを抽出条件として追加する、というように発見した傾向を目的に合わせて活用することが期待できる。



電位が-0.1 以下の三角形数 (x 軸)
疎水性が-0.5 以下の三角形数 (y 軸)

図 12. 散布図の観察結果(2)

(上) すべてのポケットを表示

(左下) druggable のみ表示 (右下) druggable 以外表示

8. まとめと今後の展望

本論文では、蛋白質ポケット抽出、また、その druggability 分析を目的とし、ポケットの抽出手法と2つの可視化ツールを提案した。抽出手法では、三角メッシュで構成された蛋白質表面形状に対し、メッシュ単純化を適用することで、抽出の高速化を図った。1つ目の可視化ツールでは、ポケット形状に着目し、サンプルポケットとの形状類似度比較を行うことでポケットの druggability を推定して可視化した。2つ目の可視化ツールでは、ポケットの特徴量と druggability の関係性を観察するための散布図と形状の表示機能を搭載した。さらに、それら2つのツールによる実行結果を示し、そこから得られる知見について議論した。

今後の展望として図1にも示した通り、可視化ツール1,2で発見した基準をポケット抽出の際の条件に追加することで、本手法・ツールをより信頼度の高いものに発展させたいと考えている。また、可視化ツール1と可視化ツール2を統合した一つの可視化ツールとして実装することも可能であることから、その有効性についても議論したい。その上で、提案手法による可視化結果の妥当性を再検証したい。さらには、本手法・ツールによる結果と化学的な実験データを比較検証することも視野に入れている。

参考文献

- [1] 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*, 15(15-16), 656-667, 2010.
- [2] Y. Kalidas, N. Chandra, PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins, *Journal of Structural Biology*, 161, 31-42, 2008.
- [3] T. Kawabata, N. Go, Detection of pockets on protein surfaces using small and large probe spheres to find putative ligands Binding Sites, *PROTEINS: Structure, Function, and Bioinformatic*, 68, 516-529, 2007
- [4] A. Shulman-Peleg, et al, MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Research*, 36(2), W260-264, 2008.
- [5] K. Kinoshita, et al., Identification of protein functions from a molecular surface database, eF-site. *Journal of Structural and Functional Genomics*,9-22, 2002.
- [6] 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*, 37(3), 181-188, 2008.
- [7] R. Osada, T. Funkhouser, B. Chazelle, D. Dlbkin, Matching 3D Models with Shape Distributions, *Proc. Int'l Conf. on Shape Modeling and Application 2001*, 154-166, 2001.
- [8] R. Ohbuchi, T. Minamitani, T. Takei, Shape-similarity search of

3Dmodels by using enhanced shape functions, *International Journal of Computer Applications in Technology(IJCAT)*, 23(3/4/5), 70-85, 2005.

[9] 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*, 56(10), 48-56, 2002.

[10] M. Hilaga, Y. Shinagawa, T. Kohmura, T. Kunii, Topology matching for fully automatic similarity estimation of 3D Shape, *ACM SIGGRAPH 2001*, 203-212, 2001.

[11] D. Y. Chen, X. P. Tian, Y. T. Shen, M. Ouhyoung, On visual similarity based 3D model retrieval, *Computer Graphics Forum*, 22(3), 223-232, 2003.

[12] <http://ef-site.hgc.jp/eF-site/index.jsp>

[13] <http://www.rcsb.org/pdb/home/home.do>

[14] T. A. Halgren, Identifying and characterizing binding sites and assessing druggability, *Journal of Chemical Information and Modeling*, 49(2), 377-389, 2009.

[15]金子彩香, 中村友香理, 伊藤貴之, 蛋白質ポケットの druggability 分析のための可視化ツール, *NICOGRAPH 2012*, pp. 1-8, 2012.

金子 彩香



2011年お茶の水女子大学理学部情報科学科卒業。2013年お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース博士前期課程修了。

中村 友香理

2010年お茶の水女子大学理学部情報科学科卒業。2012年お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース博士前期課程修了。

伊藤 貴之



1990年早稲田大学理工学部電子通信学科卒業。1992年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年日本アイ・ビー・エム(株)入社。1997年博士(工学)。2000年米国カーネギーメロン大学客員研究員。2003年から2005年まで京都大学大学院情報学研究科COE研究員(客員助教授相当)。2005年お茶の水女子大学理学部情報科学科助教授。2011年同大学教授。ACM, IEEE Computer Society, 情報処理学会, 芸術科学会, 画像電子学会, 可視化情報学会, 他会員。