

蛋白質表面形状分析手法 PROTEIN を用いたポケット形状評価

金子彩香 (指導教員：伊藤貴之)

1. 概要

人体の主成分である蛋白質は、薬品開発において非常に注目されている物質であり、多くの薬が蛋白質に作用することから、各蛋白質が薬とどの程度の反応性を持つかを把握することが重要となっている。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、その場所で一定時間に渡って化学反応を起こす。このような窪みを我々はポケットと呼ぶ。これらのことから我々は、蛋白質表面形状分析プロジェクト PROTEIN の一環として、蛋白質の薬との反応性における評価値算出手法の確立を目指す。その前段階として、本研究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) プロジェクトと称して、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およびその結果の可視化[1]に取り組んだ。また我々は、蛋白質表面形状からのポケット高速抽出[2]に取り組んだ。本研究では、蛋白質表面形状からのポケット抽出に[2]の手法を適用する。

2.2 蛋白質の結合部位の類似度評価値算出手法

文献[3]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。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、提案手法に文献[3]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 3D モデルの類似度比較手法

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、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 法[4]、モデルの慣性主軸に沿った包絡線を用いた PS 法[5]などがある。提案手法では、D2 法や PS 法を参考にした特徴量算出方法を適用している。

3. 提案手法

3.1 蛋白質表面形状

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。本手法の実装は、eF-site[6] 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に分子表面形状を算出し、データベース化して公開しており、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。

3.2 ポケット抽出

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。本手法ではポケット抽出手段

として[2]の手法を用いる。以下にその処理手順を示す。

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、微小凹凸を平滑化した表面形状を得る。
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1に示す。

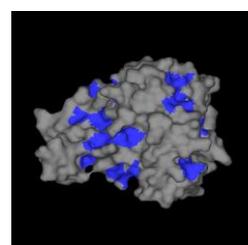


図1 ポケット抽出の例

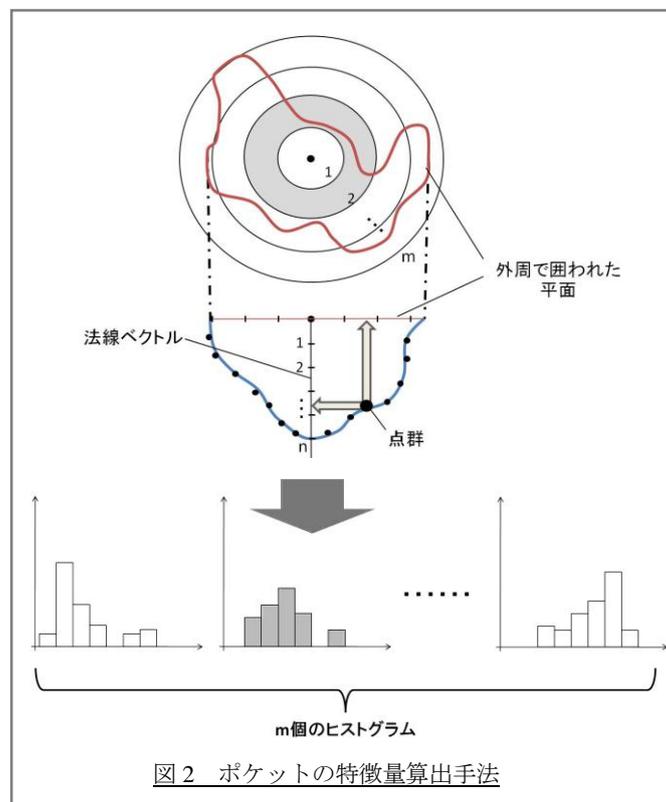


図2 ポケットの特徴量算出手法

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2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。

1. 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、それぞれ分割する。
2. ポケットの外周で囲われた平面とその中心を算出する。
3. ポケットを構成する三角メッシュ上にできるだけ均等に点群を生成する。
4. 各点に対して、その点を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。

平面上の m 区間において、各々の区間に含まれている点の深さ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、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。

3.4.1 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、薬との反応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の多くは蛋白質と反応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、薬との反応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3に示す。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。

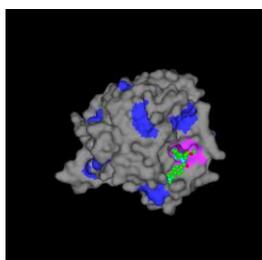


図3 サンプルポケットの例

3.4.2 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。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。

$$S_k = \frac{A \cdot B_k}{\sqrt{|A|^2} \sqrt{|B_k|^2}} \quad (1)$$

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4に示す。ここで色は、青から赤に向けて評価値が高くなることを示す。

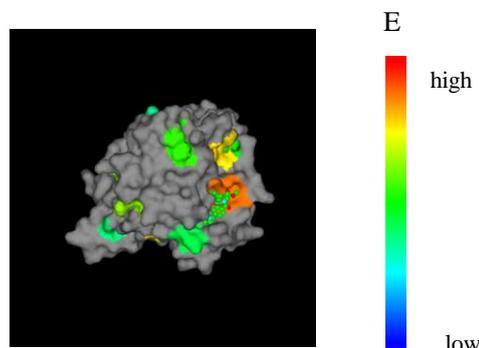


図4 実行結果

5. まとめ

本報告では、蛋白質に薬との関係性における評価値算出手法を確立するための前段階として、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。今後は、以下の課題に取り組んでいきたい。

- 多数の蛋白質に対して、本手法を適用し検証する。
- 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を評価値の算出に適用する。
- 薬との反応性の低さを表すルールを定義し、そのルールとの適合度判定を評価値算出の処理過程に組み込む。
- 蛋白質表面の幾何学的情報の他に、化学的情報も用いて評価値を算出する。

参考文献

- [1] 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.
- [2] 中村, 伊藤, 蛋白質表面形状分析手法 PROTEIN を用いたポケット分析, 第2回データ工学と情報マネジメントに関するフォーラム(DEIM 2010).
- [3] Stephanie Perot, Olivier Sperandio, Maria A. Miteva, Anne-Claude Camproux, Bruno 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.
- [4] 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.
- [5] 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.
- [6] eF-site, <http://ef-site.hgc.jp/ef-site/index.jsp>