

蛋白質ポケット特徴量からの druggability 予測のための可視化手法

三好 真紀子 (指導教員：伊藤 貴之)

1 はじめに

薬は、生体の主成分である蛋白質に反応することで効くことができる。蛋白質は、20種類のアミノ酸が多数連結して出来た高分子化合物である。アミノ酸が連結した後、複雑に折りたたまれることで特有な立体構造をとっている。蛋白質の表面は凸凹とした形状をしており、表面の中でもある程度の深さ大きさの凹みをポケットと呼ぶ。薬が効くためには、薬成分がこのポケットに入りとどまる必要がある。

薬と蛋白質ポケットの親和性の高さを druggability と呼ぶ。親和性が高いポケットを druggable なポケット、そうではないポケットを undruggable なポケットと呼ぶ。創薬において、druggable なポケットを発見することが非常に重要視されている。本研究は、蛋白質の構成要素であるアミノ酸に着目した手法を提案する。蛋白質とアミノ酸との親和性の高さには傾向があると発表されている [1]。そこで、蛋白質ポケットに特化してアミノ酸との親和性の高さを調査し、ポケットの druggability とアミノ酸との関連性を可視化する。またポケットとアミノ酸の関係から、ポケットの druggability を予測する手法を提案する。

2 関連研究

Yamaguchi らは、蛋白質にすでに低分子が結合している部位に対して、どのアミノ酸と親和性が高いのかを統計をとり分析をしている [1]。Het-PDB というデータベースを実装し、蛋白質の低分子結合部位に対してアミノ酸に関する統計をとることで、知られていなかった低分子とアミノ酸の相互作用を発見している。このことから、蛋白質ポケットとアミノ酸の関係を調べることは有意義であると考えられる。そこで、本論文では蛋白質ポケットとアミノ酸の相互作用を、蛋白質ポケットとアミノ酸間の距離分布を分析することで調べている。

3 提案手法

3.1 蛋白質表面形状データの取得

本研究では蛋白質表面形状が三角メッシュで近似されていることを前提としている。eF-site は蛋白質の分子表面に関するデータベースである。eF-site では PDB (Protein DataBank) に登録されている蛋白質を対象としている。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形メッシュを XML 形式で記述されている。

3.2 ポケット抽出

本手法では、まずポケットを蛋白質表面から抽出する。本研究の先行研究である Nakamura [2] の手法を用いてポケットを抽出する。蛋白質表面形状データから以下の手順を踏んでポケットを高速に抽出する。

1. 三角メッシュ単純化により、表面の細かい凹凸を除去して平滑化された粗い形状を形成する。
2. 平滑化された表面形状から、凹部形状を抽出する

ことで、ペプチドサイズの凹部を抽出する。

3. 抽出された凹部を、単純化前の三角メッシュに投影する。
4. 投影された凹部形状の不要な部位を除去する。

3.3 ポケットとアミノ酸間の距離算出

次に、蛋白質を構成する 20 種類のアミノ酸とポケット間のユークリッド距離を算出する。本研究では二種類の距離を算出する。ひとつは、ポケットの入り口 1 点から蛋白質を構成するアミノ酸までの距離。もうひとつは、ポケットの最深部から蛋白質を構成するアミノ酸までの距離である。入り口を構成する点群のうち 1 点または最深部の 1 点と、アミノ酸を構成している原子間のユークリッド距離を算出する。

1 つの蛋白質には同じアミノ酸が複数存在することも多い。距離算出の際には、入り口または最深部からアミノ酸までの最短の距離を各アミノ酸とポケット間の距離として定義する。

3.4 可視化分析

ポケットの入り口・最深部付近に特定のアミノ酸が 2 つ存在した場合、ポケットの druggability に影響するのか調べる。金子らが提案した可視化ツール [3] を用いる。このツールは、散布図を用いて蛋白質ポケットに関する特徴量とポケットの druggability の関係を調べることができる。本研究では、この可視化ツールに距離のデータを適用し散布図の二軸にポケットと任意のアミノ酸間の距離を割り当て、druggability とアミノ酸の関係を可視化する。

次に散布図で可視化した結果、190 組を一度に把握するためにマトリクス図で可視化する。マトリクス図の縦横軸ともに同じ順番でアミノ酸を割り当て、1 つの領域の色が二つのアミノ酸が近くにあるポケットの druggability を表す。各領域の色は、2 つのアミノ酸との距離が 5 未満の時の druggable なポケット数を全ポケット数で割った値 ds に依存する。 ds が 0.85 未満の時に青色、0.85 以上の時に赤色で色付けされ $|ds - 0.85|$ と全ポケット数の値が大きい程彩度が大きくなる。

しかしポケットの druggability に関係するのはアミノ酸 2 つだけとは限らない。そこで、20 種類のアミノ酸の各々からポケットまでの距離を 20 次元ベクトルとしてポケット周辺のアミノ酸分布を比較するため、自己組織化マップの一つである GeodesicSOM [4] に同様のデータを適用する。自己組織化マップは、多次元情報データの類似度をマップ上の近さで表現することができる。

3.5 druggability 予測

蛋白質ポケットの druggability 予測を行う。GeodesicSOM [4] に距離データを適用した結果から、druggability が未知の蛋白質ポケットの druggability の予測する。以降 druggability が未知である蛋白質を入力蛋白質と呼ぶ。druggability が既知である蛋

白質ポケットについて 3.3 節で述べた距離を計算し、GeodesicSOM に適用する。また、入力蛋白質に対してもポケットとアミノ酸間の距離を計算する。そして GeodesicSOM の結果マップに表示されるニューロンの中で最も似ている距離の値を持つ重荷ベクトルをユークリッド距離をそれぞれ算出することで探索する。また同様に、入力蛋白質と周りのニューロンとの違いもユークリッド距離で算出する。算出されたユークリッド距離 D から、平均と標準偏差を計算する。算出されたユークリッド距離の内、平均と標準偏差の和以上の値をもったものがあれば、その値を持ったニューロンのラベルを予測の候補に入れるのをやめる。そして選ばれたニューロンのラベルから、そのラベルの druggability ごとに $V = \sum \frac{1}{d}$ を算出する。ここで算出された値のうち、最も大きい値が算出された druggability を入力蛋白質の druggability として予測する。

4 実行結果

本手法を実行するにあたり、Halgren[5] が文献にてテストデータとして使用した蛋白質を使用した。文献中にこれらの蛋白質の druggability 情報が紹介されており、druggable, undruggable もしくはどちらとも言えない difficult という情報が得られた。60 個の蛋白質からポケットを抽出し可視化を行った。

可視化ツール [3] を用いてポケット最深部と任意の 2 つのアミノ酸間の距離を可視化した結果が図 1 である。ポケット付近にアラニンとチロシンがあると druggable なポケットであることが多いと示された。

ポケット最深部とアミノ酸間の距離分布をマトリクス図で可視化した結果が図 2 である。図 2 を見ると一部に赤く色付けられた領域が集まっていることがわかった。つまりそれらの領域に割り当てているアミノ酸が近くにあるポケットが druggable なポケットであることが多いことがわかった。また割り当てられているアミノ酸を見るとセリンやヒスチジンなどの親水性アミノ酸とバリンやイソロイシンなどの疎水性アミノ酸であることがわかった。つまり親水性と疎水性のアミノ酸の組み合わせが近くにあるポケットは druggable なポケットであることが多いということがわかった。

druggability 予測についても同様のデータを扱った。はじめにデータを実際に予測を行うデータを全体の 40%、GeodesicSOM[4] に適用するデータを全体の 60% になるようにランダムに分けた。データの蛋白質 druggability はすべて既知であるわけだが、予測する段階では蛋白質 druggability は未知であると仮定した。予測を行った結果、druggability 予測が的中した割合は、81.6% であった。

5 まとめと今後の課題

本研究では、蛋白質ポケットの druggability 分析の一環としてポケットとアミノ酸間の距離分布を可視化し結果を示した。その結果として、特定のアミノ酸の距離が変化すると蛋白質ポケットの druggability が変化してくることが示唆された。また可視化結果から蛋白質ポケットの druggability を推測する手法を提案した。

今後は距離の計算手法を検討し、また予測手法の改善を行いたい。また他のデータにも提案手法を適用し

さらにアミノ酸とポケットの関係を調査していきたい。

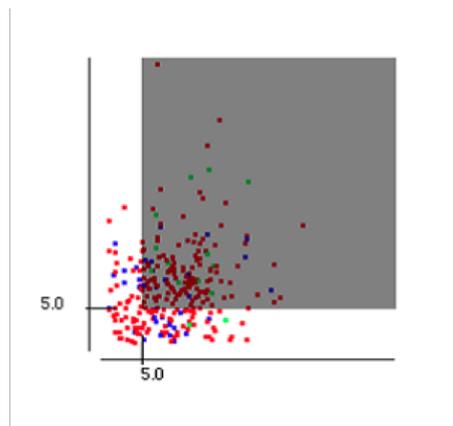


図 1: ポケットとアラニン間の距離を横軸、ポケットとチロシン間の距離を縦軸に割り当てた可視化結果

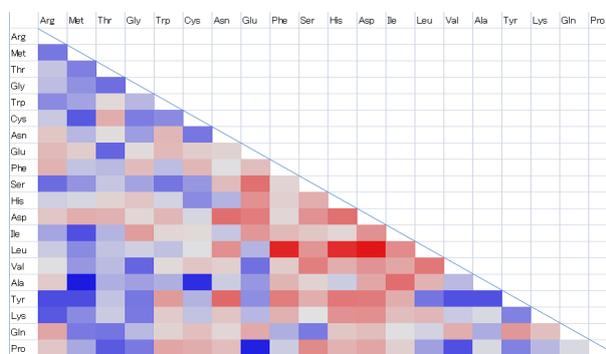


図 2: ポケットとアミノ酸間の距離を可視化した結果

参考文献

- [1] Yamaguchi, A., Iida, K., Matsui, N., Tomoda, S., Yura, K. and Go, M.: Het-PDB Navi.: A Database for Protein-Small Molecule Interactions, *Journal of biochemistry*, Vol. 135, No. 1, pp. 79-84 (2004).
- [2] Nakamura, Y. and Itoh, T.: Fast Detection and Visualization of Protein Surface Pockets for Drug Design, in *The 11th Asian Symposium on Visualization* (2011).
- [3] 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, *芸術科学会論文誌*, Vol. 12, No. 3, pp. 104-113 (2013).
- [4] Wu, Y. and Takatsuka, M.: Spherical Self-Organizing Map Using Efficient Indexed Geodesic Data Structure, *Neural Networks*, Vol. 19, No. 6, pp. 900-910 (2006).
- [5] Halgren, T. A.: Identifying and characterizing binding sites and assessing druggability, *Journal of chemical information and modeling*, Vol. 49, No. 2, pp. 377-389 (2009).