

配列相同性に基づく タンパク質のリガンド結合候補部位の評価

安尾 信明^{1,a)} 関嶋 政和^{1,2}

概要: タンパク質の立体構造情報を用いる structure based drug design (SBDD) において、リガンド結合候補部位におけるアミノ酸の保存性は重要な情報である。本研究では、リガンド結合候補部位におけるアミノ酸の保存性をタンパク質の立体構造上に可視化し、また、リガンド結合候補部位全体としての保存性を定量的に評価するツールを開発した。さらに、その有効性を実際の創薬ターゲットとなるタンパク質を用いて検証した。

キーワード: *in silico* 創薬, SBDD, 配列相同性, リガンド結合候補部位

Evaluation of ligand binding site candidates based on sequence homology

YASUO NOBUAKI^{1,a)} SEKIJIMA MASAKAZU^{1,2}

Abstract: In structure based drug discovery (SBDD) based on three-dimensional protein structure, amino acid conservation of ligand binding site candidates is important. In this study, we made a tool which visualized sequence homology of ligand binding site on protein structure and evaluated whole site quantitatively. We also applied real target of drug discovery to our tool to validate.

Keywords: *in silico* drug discovery, structure based drug design, sequence homology, ligand binding site

1. 序論

タンパク質の立体構造を決定する手法が進歩し、タンパク質立体構造のデータベース件数は大きく増加した [1]。これにより、疾病に関与するタンパク質の立体構造を用いた structure based drug design (SBDD) が可能になった。SBDD におけるシミュレーションでは、多くの場合タンパク質のアミノ酸配列は構造を解析した際の特定の配列となっている。しかし、アミノ酸配列は個体ごとに異なる場合があり、同じタンパク質のアミノ酸配列の中でも、構造

や機能に重要な役割を果たすアミノ酸は変化しにくく、そうでないアミノ酸は変化しやすい傾向にあることが知られている [2]。アミノ酸の保存性は薬剤耐性にも関与する重要な情報であるが、タンパク質全体のアミノ酸の保存度を示す従来の手法 [3] では、リガンド結合候補部位のアミノ酸の保存度を評価する際に問題がある。そこで、本研究では、タンパク質の機能に重要であるリガンド結合候補部位について、アミノ酸配列の保存性をタンパク質の三次元構造上に可視化し、各リガンド結合候補部位について、その保存性を定量的に評価するツールを開発した。また、実際のタンパク質にこれを適用し、その有効性を検証した。

2. 手法

2.1 ツールの概要

本研究で開発したツールによる処理は、図 1 のように行

¹ 東京工業大学 情報理工学専攻 計算工学専攻
Department of Computer Science, Tokyo Institute of Technology

² 東京工業大学 学術国際情報センター
Global Scientific Information and Computing Center, Tokyo Institute of Technology

a) yasuo.n.aa@m.titech.ac.jp

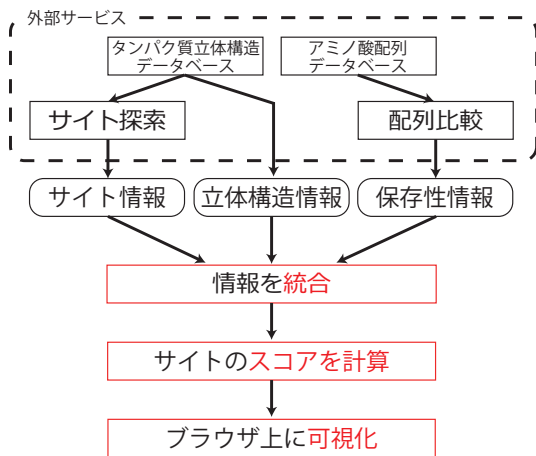


図 1 開発したツールの処理フロー

われる。入力は、タンパク質立体構造情報、リガンド結合候補部位情報、配列相同性情報の三つである。結果は Jmol を使用する HTML ファイルとして出力され、ブラウザで開き操作することができる。

2.2 リガンド結合候補部位全体におけるスコアの計算

リガンド結合候補部位のスコアは、タンパク質全体をマルチプルアラインメントした結果における、各列のアミノ酸の保存性から求めた。

マルチプルアラインメントの i 列目における最大頻度 $p_{max}(i)$ は、 $p_{max}(i) = \max_{x \in A} p_i(x)$ で求められる。ただし、 $p_i(x)$ は i 列目においてアミノ酸 x が現れる確率を表し、 A はアミノ酸の集合を表す。最大頻度は計算が簡単であると同時に、範囲が $0 < p_{max} \leq 1$ であり、可視化の際比較がしやすい利点がある。

本研究におけるリガンド結合候補部位 s のスコア $C(s)$ は、リガンド結合候補部位のアミノ酸の数 $|s|$ を用いて

$$C(s) = \sqrt[|s|]{\prod_{i \in s} p_{max}(i)^2}$$

で定義される。これは、 $(p_{max})^2$ の相乗平均である。このスコアの特徴として、リガンド結合候補部位の大きさに左右されにくいこと、 $0 < C(s) < 1$ が成り立ち、ユーザがスコアの値の大小を理解しやすいことが挙げられる。

3. 実験・考察

本研究では、以下の三つのタンパク質を使用して実験を行い、結晶構造中でリガンドが実際に結合している部位のスコアを計算した。配列データベースは UniprotKB、タンパク質立体構造は PDB から取得した。また、リガンド結合部位探索は Sitemap、マルチプルアラインメントは ClustalW を使用した。使用したデータ及び評価結果のスコアを表 3 に示す。

このツールによる出力の例として、インフルエンザウイルスノイラミニダーゼのタンパク質についての結果を図 2

表 1 実際に使用したデータ

タンパク質	PDB ID	配列数	スコア
ヒト FKBP	1FKB	4	1.0
インフルエンザウイルス ノイラミニダーゼ	2HU4	69	0.81
HIV プロテアーゼ	1SDT	619	0.92

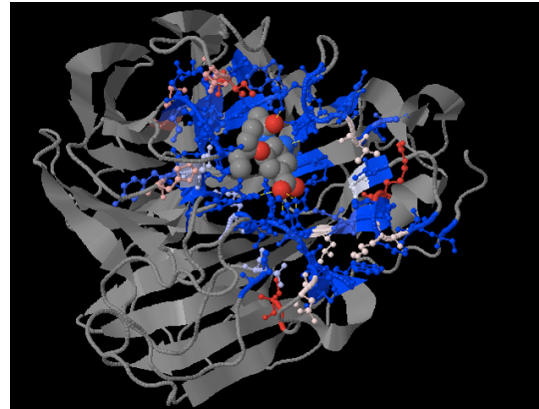


図 2 インフルエンザウイルスノイラミニダーゼに対する出力

に示す。出力結果のアミノ酸は、保存されている割合が高いものから順に青、白、赤と色が変わる。この結果から、ヒトの FKBP タンパク質のリガンド結合部位は、アミノ酸配列が完全に保存されていること、また、ウイルス同士で比較した場合、HIV プロテアーゼのリガンド結合部位は、インフルエンザウイルスノイラミニダーゼのリガンド結合部位よりもよく保存されていることが明らかになった。

4. 結論

本研究では、タンパク質のリガンド結合候補部位について、アミノ酸配列の保存性を三次元構造上に表示することで可視化し、また各薬剤結合候補部位におけるアミノ酸配列の保存性を定量的に評価する指標を開発した。

今後の展開として、アミノ酸の分布などを考えたより良いスコア関数を構築すること、また、開発したツールをオンラインで公開し、誰でもが使用可能なツールにすることが挙げられる。

参考文献

- [1] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, 28: 235-242, 2000.
- [2] Cyrus A. Wilson, J. Kreychman, M. Gerstein, Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores., *J. Mol. Biol.*, 17;297(1):233-49, 2000.
- [3] Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. "ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information.", *Bioinformatics*, 2003. 19(1):163-4.