

タンパク質立体構造中のリガンド結合候補部位 における配列保存性評価手法の開発

安尾 信明[†] 関嶋 政和[‡]

東京工業大学 情報理工学研究科 計算工学専攻^{†,‡} 東京工業大学 学術国際情報センター[‡]

1. 序論

近年、NMR や X 線結晶構造解析等のタンパク質の立体構造を決定する手法が進歩すると共に、それと共に、タンパク質立体構造のデータベース数は大きく増加した¹。疾病に関与するタンパク質の立体構造を用いる structure-based drug design (SBDD) は効率的な薬剤候補化合物の探索手法と考えられるようになってきている。SBDD で使用される分子動力学やドッキング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立体構造が解析された際の特定の株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、特に変異の速度が速いウイルスの場合はその差が顕著である。個体間のアミノ酸配列の保存性を調べる手法は多数存在するが、アミノ酸配列は一次元の情報であり、三次元構造における位置関係情報が欠落している。

リガンド結合候補部位は一般にその他の部位に比べ保存性が高いため、タンパク質全体のアミノ酸の保存度を示す従来の手法^{2,3}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その他の部位の影響を受けてしまう。そのため、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、また、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手法を開発した。同時に、実際の創薬ターゲットとなっている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 手法の概要

本研究での処理は、図1のように行われる。

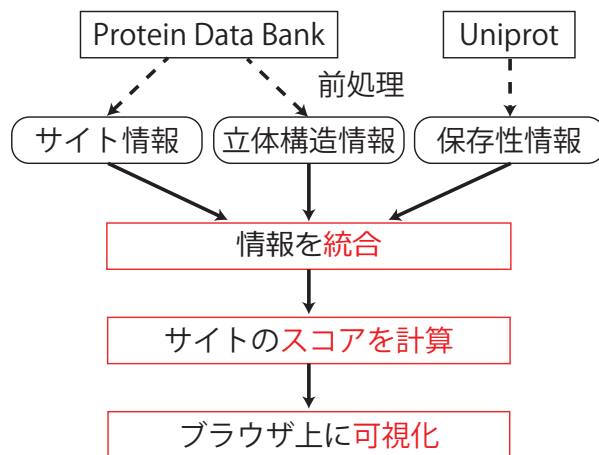


図1. 全体のフロー

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 pymol の入力ファイルとして出力され、ユーザはブラウザ上で操作するか、pymol インターフェース上で操作するかを選択することができる。

2.2 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ている。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。最大頻度は配列の本数やアミノ酸の種類によらず他のサイトの残基などとも比較が可能である利点があり、値は $0 < p_{\max}(i) \leq 1$ である。

また、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、 $(p_{\max}(i))^2$ 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、 $0 < C(s) \leq 1$ が成り立ち、他のサイトと比較しやすいことが挙げられる。

Evaluation method of ligand binding site based on sequence conservation

Nobuaki Yasuo[†], Masakazu Sekijima^{†,‡}

[†]Department of computer science,
Tokyo institute of technology

[‡]Global scientific information and computing center,
Tokyo institute of technology

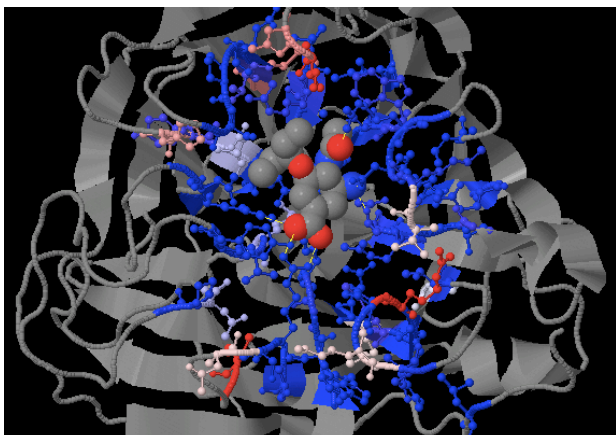


図 2. Jmol による表示

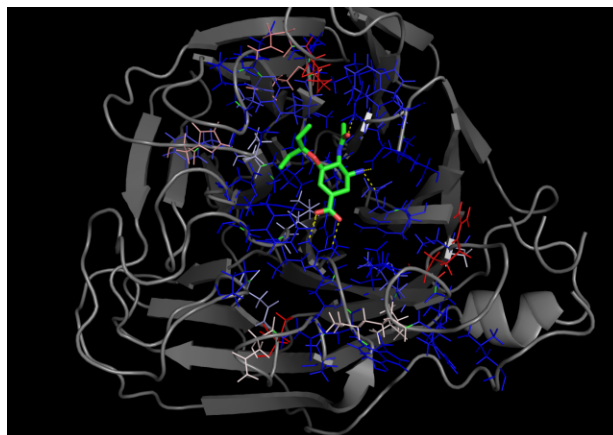


図 3. pymol による表示

2.3 使用する配列

本研究で使用する配列はユーザが選択して入力するものとした。これは、ユーザが想定するターゲットによって使用したい配列の範囲が異なり、同種内の亜型同士の類似箇所や相違点を見る場合や近縁の種間での類似性を見る場合などが考えられるからである。

3. 実験・考察

本研究では、実際の抗ウイルス薬の創薬ターゲットタンパク質を用いて、可視化やスコアについての実験を行なった。使用したターゲットを表 1 に示す。今回の実験での配列は、Uniprot データベース中で類似した配列が取り除かれているデータベースである Uniref90 から、BLAST を用いて、標的以外の配列が現れるまで E-value 順に取得した。BLAST のクエリは結晶構造の配列を元にし、配列の本数は、ターゲット順に 69 本、920 本、619 本となった。

可視化の結果のうち、実験 No.1 であるインフルエンザウイルス A 型のノイラミニダーゼのものを図 2, 図 3 に示す。図 2 は Jmol プラグインによるブラウザ上の表示、図 3 は pymol インターフェースによる表示である。出力結果において、リガンド結合候補部位に含まれているアミノ酸は、保存度 $p_{\max}(i)$ が高いものから順に青、白、赤と色が変わる。

この結果から、このタンパク質の結合サイトの
表 1. 使用したターゲット

No.	ウイルス	タンパク質	PDB ID
1	インフルエンザウイルス A 型	ノイラミニダーゼ	4M4Q
2	ヒト免疫不全ウイルス (HIV)	リバーstrans スクリプターゼ	1JLB
3	ヒト免疫不全ウイルス (HIV)	プロテアーゼ	1SDT

うち、リガンドと直接相互作用する部位はよく保存されている一方、変化のしやすいアミノ酸も存在することがわかる。また、このタンパク質のリガンドは抗インフルエンザ薬で知られるオセルタミビル (タミフル) であるが、既知の薬剤耐性をもたらす変異である H274Y が検出できており⁴、図 2 中央左の保存性がやや低くピンク色になっているアミノ酸が該当する。

保存性スコアは、ターゲット順に 0.81, 0.83, 0.92 となっており、3 ターゲットの中では HIV プロテアーゼのリガンド結合候補部位が比較的よく保存されていることがわかる。

4. 結論

本研究の手法により、タンパク質のリガンド結合候補部位において、アミノ酸配列の保存性を三次元構造上に可視化することができ、また部位全体の保存性を定量的に評価することが可能になった。今後の展開として、アミノ酸の性質を考慮した類似性を用いること、結合候補部位のスコアの良さを定量的に評価することなどが考えられる。

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] Vyas J, Gryk MR, Schiller MR. "VENN, a tool for titrating sequence conservation onto protein structures," *Nucleic Acids Research* 37(18):e124 (2009).
- [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*. 19(1):163-4 (2003).
- [4] Ives JA1, Carr JA, Mendel DB, Tai CY, Lambkin R, Kelly L, Oxford JS, Hayden FG, Roberts NA. "The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo." *Antiviral Research* 55, 307-317 (2002).