

水の情報エントロピーに注目したタンパク質の リガンド結合部位予測法の開発

佐々木 孝章^{†1} 関 嶋 政 和^{†1,†2}

タンパク質とリガンドの結合は、タンパク質の機能発現や創薬に重要である。タンパク 3000 プロジェクトを始めとするタンパク質の大規模な構造解析により、多くのタンパク質の構造情報が得られてきており、タンパク質とリガンドとの結合を予測し、機能発現や創薬への応用が期待されている。従来、リガンド結合部位は立体構造解析と静電場の解析から求められてきており、その結果、*in silico* による創薬では相補的構造を有する分子を中心に探索されている。本研究では、水分子が自由度を得るとエントロピー的に有利に機能することを利用して、タンパク質周囲の水分子の情報エントロピーに注目し、リガンドの結合部位予測を行うシステムを開発した。

Development of protein ligand binding site prediction focused on water information entropy

TAKA AKI SASAKI^{†1} and MASAKAZU SEKIJIMA^{†1,†2}

Protein and ligand docking is important for expression of protein function and drug development. Protein 3000 project and other structural analysis project determined much protein structures. And this huge structural information is expected to contribute analysis of expression of protein function and drug development. Previously, ligand binding pocket analysis was based on shape complementarity and electrostatic interaction. We developed and evaluated protein ligand binding site prediction system focused on information entropy of water around protein.

^{†1} 東京工業大学 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

1. はじめに

近年、日本では薬の特許切れが相次ぐ一方で、新薬の創出能力の低下が問題視されている。医薬品の開発において、ターゲットとなるタンパク質に結合するリガンドは、天然資源からの単離やハイスループット・スクリーニング (HTS : high-throughput screening) が主に使われてきたが、ハイスループット・スクリーニングは、既知の化合物のみに適用可能な方法であること、極めてコストが大きいこと、偽陽性ヒットの割合の大きいことなどが問題になっている。また、HTS では活性が高いが単位原子当たりの有効性が低い化合物から最適化を始めるために、最終化合物は分子量及び構造の複雑性の増大に伴う物性の悪化が懸念される¹⁾。この問題を解決する手法として、低分子量化合物からリード化合物を作り上げて行くフラグメントベース創薬 (FBDD : fragment based drug design) への取組みが進められている²⁾。FBDD では、活性は低い単位原子当たりの有効性が高い化合物に対して、化合物の伸張や結合等の最適化を行う為に、無駄の少ないコンパクトな最終化合物の作成を可能にする。従来のタンパク質立体構造情報に基づく薬剤設計では、リガンド結合部位の立体構造解析と静電場の解析 (鍵穴構造の解析) から、この部位での相補的構造を有する分子 (鍵) をデータベースから検索し、結合を行う³⁾。具体的には、活性部位の空洞を原子を模した球体で埋め、その球体の中心間の距離を求め、原子間距離が近い化合物を分子構造データベースから検索するという方法である。しかし、この手法では設計されるリガンドは立体構造上の相補性に依存することになり、活性が低い母化合物を拡張することで単位原子当たりの有効性が高まるかは保証されない。つまり、現状ではタンパク質とリガンドが結合した構造 (活性構造) を考慮した最適化を行うことが困難である。本研究では、水分子が自由度を得ると物理的なエントロピー的に有利に機能することを利用して、タンパク質とリガンドが結合した構造を考慮したりガンド設計を可能にするためにリガンド結合部位予測を行うシステムを開発し、その検証を行った。

2. システムの概要

分子動力学法 (MD : molecular dynamics) による水分子の位置に関する情報エントロピーと、ポケット探索サーバによるタンパク質表面の幾何学的な探索を組み合わせ、リガンド結合部位を予測する (図 1)。

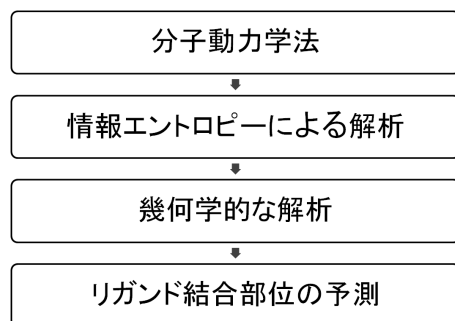


図 1 システムの概要
Fig.1 System overview

2.1 MD シミュレーションによる水分子のダイナミクスの獲得

分子動力学法を用いて、対象となるタンパク質の周囲の水のダイナミクスを獲得する。まず、タンパク質のエネルギー最小化を 1000 ステップおこなう。次に、水分子を対象となるタンパク質の周囲に立方体状に配置する。水分子を緩和させる為に、分子動力学法を用いて 30ps かけて 0K から 300K まで圧力を一定にして温度を上げた後、20ps の等温の MD シミュレーションを実行する。この過程を経た対象となるタンパク質について、常に対象となるタンパク質の構造を固定したまま水分子のみが動くことが可能な 20ns の MD シミュレーションをおこなう。

2.2 水分子の情報エントロピー

MD シミュレーションから得られたトラジェクトリ (軌跡) から PDB ファイルを出力し、解析を行なう。対象となるタンパク質から 10ps 毎の原子の各座標を PDB ファイルとして出力する。次に、対象となるタンパク質と水からなる系を一辺 3.0Å の立方体に細分化する。MD シミュレーションの結果毎に PDB ファイルから水がどの立方体内に存在するかカウントする。各立方体 i において水分子をカウント数を n_i 、全ステップ数を N として、情報エントロピー H を計算する。情報エントロピーの式は以下で表される。

$$H = - \sum_i \frac{n_i}{N} \log \left(\frac{n_i}{N} \right) \quad (1)$$

情報エントロピーが低いという事は、その領域において水分子の動きが少ないという特殊な状況である。そこで、この周辺のタンパク質表面が機能を持っていると考えて、情報エント

ロピー H がある値未満の領域を検出する。

2.3 タンパク質表面の幾何学的なポケット探索

タンパク質表面の幾何学的手法を用いたポケット探索を行なう。本研究では、インターネット上で公開されている探索用のサーバ (Pocket-Finder⁷) を利用する。Pocket-Finder は、Hendlich らの Ligsite⁶) に基づくポケット推定アルゴリズムを利用している。半径 1.6Å のプローブを使って、タンパク質の周りの格子線にそって、0.9Å 毎に探索する。また、立方対角線方向にも探索していく。同じ部分を 7 回探索して少なくとも 5 回、領域の一部として定義された格子点をポケットの領域とする。

2.4 二つの解析の統合

水分子の情報エントロピーの解析と幾何学的なポケット探索の解析の結果を統合する。それぞれの解析結果から、検出された領域が重なる場合に結合部位と予測する。幾何学的な探索を組み合わせる事で、タンパク質表面にある情報エントロピーの低い箇所を排除する一方で、幾何学的手法がはらんでしまう熱揺らぎで生まれる偶然的くぼみを排除する事を可能にする。

3. 実験結果

3.1 実験の条件

本研究では、対象となるタンパク質として、PDBID がそれぞれ 1FKF,3N2U,1UY6 の 3 種類を用いる。周囲に配置する水分子は TIP3P モデルを用いた。分子動力学シミュレーションには、AMBER10 の PMEMD を用いた。長距離の静電相互作用には Particle Mesh Ewald (PME) 法を用いて、カットオフ距離は 10Å に設定した。シミュレーションには、東京工業大学 学術国際情報センター Tsubame2.0 を用いた。情報エントロピーの閾値を 0.1 に設定し、それ未満を結合部位として検出した。

3.2 MD の出力から情報エントロピーの解析

タンパク質の表面上の情報エントロピーが 0.1 未満となる領域の数を図 2 に示す。タンパク質の種類によって領域数が異なり、50~110 個のばらつきがある。0.1 未満となる領域のうち、その殆どは水分子の観測が低頻度となる領域であった。

図 3 は、結合部位を正面にしたタンパク質と情報エントロピーの低い領域を点で示した図である。青い点が、情報エントロピーが 0.1 未満の領域のうち、水分子の観測頻度が非常に高い領域、黄色の点は、情報エントロピーが 0.1 未満の領域の、水分子の観測頻度が低い領域である。また、赤で囲った部分はリガンドの結合部位を示している。

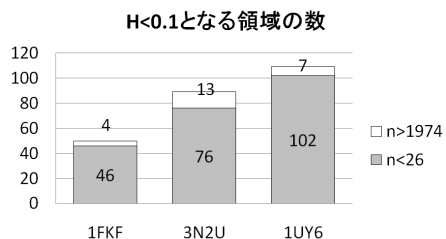


図 2 タンパク質周囲の情報エントロピーの低い領域の個数を示すグラフ。

Fig.2 The graph shows the number of low information entropy area of water around protein.

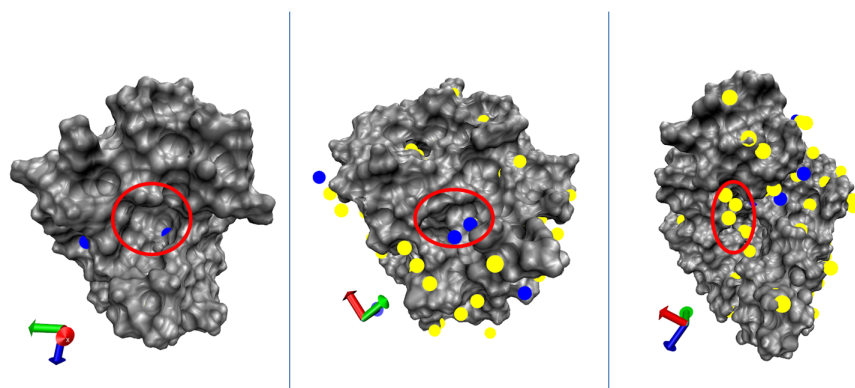


図 3 タンパク質周囲の情報エントロピーの低い領域を青点と黄点で表す。(赤く囲った部分は、リガンド結合部位)。タンパク質は左から 1FKF,3N2U,1UY6。

Fig.3 The blue points and yellow points show the areas of low information entropy around protein (the red ellipses show ligand binding site). The kind of proteins are 1FKF, 3N2U, 1UY6 from left to right.

1FKF では、情報エントロピーの低い領域がリガンド FK506 の結合する箇所とそれ以外の箇所まで計 50 か所発見された (図 3 左)。図 3 左では、青い点の 2 か所が見えているがその他の該当する領域は、裏側の面に存在している。

3N2U でも同様にリガンド D3X の結合部位に情報エントロピーの低い領域が 3 か所発見された (図 3 中央)。3N2U は、情報エントロピーの低い領域が計 89 か所存在しており、図 3 からも分かるように、タンパク質表面のくぼみやそうでないところなど、様々な部分で

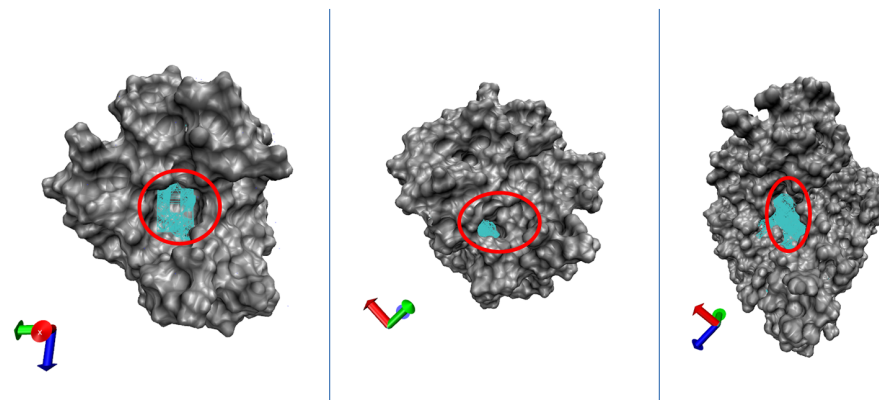


図 4 Pocket-Finder によるポケット探索の結果。タンパク質は左から 1FKF,3N2U,1UY6。

Fig.4 This shows the result of pocket prediction by Pocket-Finder. Order is same as Fig.3 .

情報エントロピーの低い領域が存在している事がわかる。

1UY6 ではリガンド PU3 結合部位に、情報エントロピーの低い領域が現れているが、前の二つとは異なり水分子の観測回数が低い領域が存在している (図 3 右)。情報エントロピーの低い領域の数は、本研究のタンパク質 3 種類の中で最も多く 109 箇所存在している。

3.3 幾何学的な解析

幾何学的な探索として Pocket-Finder による解析の結果を図 4 に示す。それぞれのタンパク質から 10 個ずつポケットの候補とされる領域を出力する。図 4 の水色の部分がポケットと判定された領域である。図 4 は、リガンド結合部位 (赤く囲った部分) を正面とする図である。どのタンパク質についても、リガンド結合部位を判別している事がわかる。図 4 では、リガンド結合部位のみを出力しているように見えるが、実際は正面から見えない側に、他のポケットと判定された領域が存在する。

3.4 2つの解析結果の統合

Pocket-Finder による探索の結果は領域を 10 個のみ出力する。MD から取得した情報エントロピーの解析と幾何学的なポケット探索を統合した結果について、ポケット探索のみの結果と統合した結果の比較を図 6 に示す。2つの解析を統合した結果は、ポケット探索の結果の領域に少なくとも 1 か所、情報エントロピーの低い領域が重なっている領域をカウントしたものである。ポケット探索によるポケットと、水分子の情報エントロピーの低い領

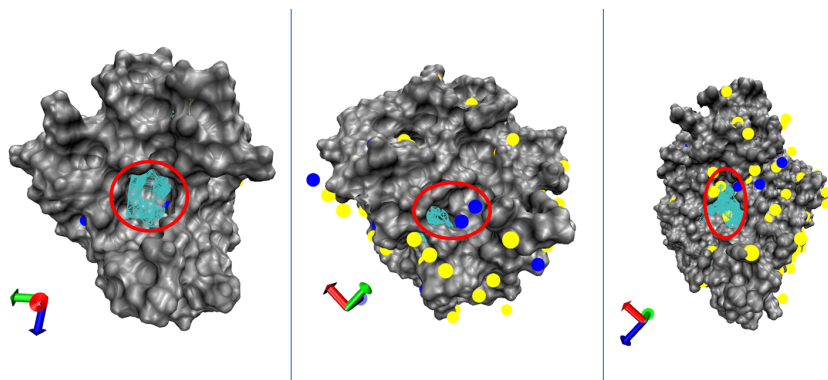


図5 2つの解析を合わせた結果の図. タンパク質は左から 1FKF,3N2U,1UY6.
Fig.5 This figure shows the result joined two analysis.

域が重ならないところがいくつかあり, Pocket-Finder のみを用いた探索より, 該当する領域が削減されている.

1FKF の場合, Pocket-Finder の結果と情報エントロピーの低い領域が重なる領域は, FK506 の結合部位を含め 3 か所となった (図5 左).

3N2U では, 重なった部位は 2 箇所だが, D3X の結合部位とは異なる領域であった. しかし, それぞれの解析結果による領域は D3X の結合部位にあるが, 2つの領域が交わっていない部分がある (図5 中央).

1UY6 では PU3 の結合部位を含めて, 4 箇所発見された (図5 右).

4. ま と め

本研究では, タンパク質のリガンド結合部位を水分子の位置の情報エントロピーから予測するシステムを開発した. タンパク質とリガンドが結合する際, 活性部位に水和している水が外に排出されることによってエネルギー的な安定をはかる⁴⁾. この排出は結合自由エネルギーに物理的なエントロピー的にもエンタルピー的にも寄与する. これは結合前では, 水和水は配向と位置の制約からエントロピー的に不利であり, 疎水性クロージャに水和した水は, 水素結合の補完が形成できないためエンタルピー的に不利となることを利用している.

水分子の情報エントロピーのみの解析では, たくさんの領域が条件に当てはまるが, タンパク質表面の平らな部分にそれらが存在している事も多い. しかしその領域にはリガンドが

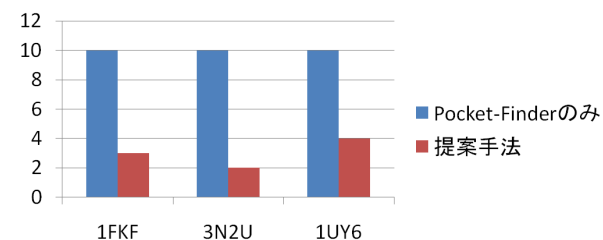


図6 Pocket-Finder と提案手法の領域の数の比較するグラフ.
Fig.6 This graph shows comparison between result of Pocket-Finder and proposal.

結合しない可能性が高く, 偽陽性のヒットが多くなってしまふ. そこで幾何学的なポケット探索を組み合わせる事によって, そのくぼみでは無い部分の領域をカットする事ができる. また逆に, 幾何学的なポケット探索だけではタンパク質の形状相補性のみを考慮するが, それだけでは, 熱揺らぎなどの偶然で生まれたくぼみをカウントする事がある. それを防ぐために, 水分子の情報エントロピーを取り入れる事によって, タンパク質の性質を考慮した探索が可能になる. 二つの手法を組み合わせる事で, 補って結合部位予測を行なう事が可能となった. しかし図7の3N2Uのようにリガンド結合部位の水分子の情報エントロピーの低い領域とポケットの位置が重ならない部分がある. リガンドの分子はある程度の大きさがあるため二つの領域が重ならない場合でもリガンド結合部位を予測ができていている場合があると考えられる. このような領域も含めて発見していく手法を今後考えていく必要がある.

本研究では, タンパク質のリガンド結合部位の予測の検証を行う為に, MD シミュレーションを行い, タンパク質内の水分子の情報エントロピーについて解析を行った. その結果, リガンドが結合可能なポケット様な部位において, 水分子の情報エントロピーが低い事を確認した. さらに幾何学的なポケット探索を組み合わせると, 表面が平らな領域などを除く事によって, リガンド結合部位の予測の精度が向上した.

参 考 文 献

- 1) Oprea, T.I., Davis, A.M., Teague, S.J. and Leeson, P.D.: Is There a Difference between Leads and Drugs? A Historical Perspective, *J. Chem. Inf. Comput. Sci.*, Vol.41, No.5, pp.1308-1315 (2001).
- 2) Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W.: Discovering High-Affinity Ligands for Proteins: SAR by NMR, *Science* 29 November 1996, Vol.274

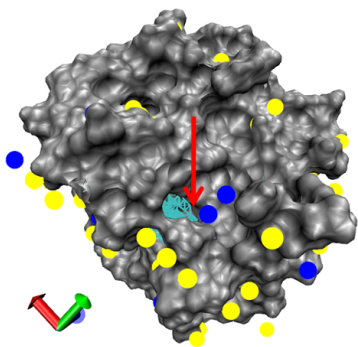


図7 3N2Uの結果. 情報エントロピーの低い領域とポケット探索の領域が重なっていない(赤い矢印).
Fig.7 This shows result of 3N2U.low information entropy area and pocket prediction area do NOT overlap(red arrow).

No.5292 pp.1531–1534 (1996).

- 3) Kuntz, I.D.: Structure-Based Strategies for Drug Design and Discovery *Science* 21 August 1992, Vol.257 No.5073 pp.1078–1082 (1992).
- 4) Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J.: Atomic Structure of FKBP-FK506, an Immunophilin-Immunosuppressant Complex *Science* 10 May 1991, Vol.252 No.5007 pp.839–842 (1991).
- 5) Abel, R., Young, T., Farid, R., Berne, B.J. and Fresner, R.A.: Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding *J. Am. Chem. Soc.*, Vol.130, No.9, pp.2817–2831 (2008).
- 6) Hendlich, M., Rippmann, F., Barnickel, G.: LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins *J Mol Graph Model.*, 15(6):359-63, 389. (1997).
- 7) Pocket-Finder: Pocket Detection <http://www.modelling.leeds.ac.uk/pocketfinder/>