

## 相互作用類似尺度を利用したタンパク質／リガンド結合予測法の開発

広川貴次

産業技術総合研究所 生命情報科学研究センター 分子設計チーム

計算機的手法による従来の SBDD では、厳密な立体構造情報が必要な上に、効果的な構造探索や相互作用エネルギー値の精度が求められている。一方で、標的タンパク質として重要なファミリーの一つである GPCR などは、結晶構造が皆無に近いという現状があり、構造予測法により構築した立体構造の精度でどれほど信頼性のある SBDD が可能であるか課題が残っている。このような背景の中、私たちはドッキング計算後のエネルギー評価値のみではなく、タンパク質／リガンドの結合様式を複数のリガンド間で相互作用類似尺度により比較解析することで直感的かつ柔軟なリガンド結合を予測する手法を開発した。本発表では幾つかの GPCR に対する阻害剤の結合予測事例を紹介する。

### Molecular modeling of the receptor-ligand complex using the similarity of molecular interaction profile

Takatsugu Hirokawa

Molecular Modeling and Design Team, Computational Biology Research Center

National Institute of Advanced Industrial Science and Technology

The difficulty with receptor-ligand docking is in part due to the dependence of structural resolution of target proteins. We developed a molecular modeling technique for the prediction of receptor-ligand binding models using comparative ligand-binding analysis (CoLBA) that takes into account not only interaction energy, but also similarity of interaction profiles among the ligands. The advantage of CoLBA is to realize intuitive and flexible screening from docking results when protein structures with low resolution (or theoretical models) are targeted. As the results of ligand binding prediction to G-protein coupled receptor targets, the key residues for binding using CoLBA are consistent with the observation of mutagenesis studies.

#### 1. 背景と目的

特定のリガンドが結合するタンパク質の立体構造情報が得られた場合、活性部位周辺の形状を「鍵穴」と見立て、「鍵」となるリガンドの結合モデルを原子レベルで推定し、その分子基盤を新規リガンド探索に利用する Structure-Based Drug Design (以下

SBDD。本研究では、計算機的手法による SBDD を前提とする) が効果的な創薬アプローチの一つとして利用されている。近年では、構造ゲノミクスやタンパク質立体構造予測技術の発展に伴い標的タンパク質の立体構造情報は急増しており、その結果、SBDD に対する期待も大きくなっている。しかし現状

では、標的タンパク質の情報量に比べると新薬の誕生が思うように加速していない。その原因の一つに SBDD の基盤技術であるドッキング計算が十分に活用されていない点が挙げられる。特に標的タンパク質の立体構造が X線や NMR のような実測されたものではなく立体構造予測法により得られた立体構造モデルを利用する場合は、深刻な問題となる。具体的には、ドッキング計算では標的タンパク質上のリガンドの配座・配向が十分に探索されているが、これを結合エネルギーのような評価スコアだけでは正解の結合モデルを同定できず、候補となる数十の結合モデルからしばしば人的介入によって任意の結合モデルを選び出す作業が必要になり、これが SBDD の律速段階になっている。

一方でバイオインフォマティクスの分野では、META サーバー<sup>1)</sup>に代表されるように単一の評価尺度では予測が困難な問題に対して複数の解析結果を利用したコンセンサス評価で負の過剰予測を減らすこと成功している。そこで本研究では、通常人的介入により行われるドッキング計算結果の補正過程をコンセンサス手法により自動化することを試みた。ただし、本研究のコンセンサス手法では、異なるドッキング計算手法間の比較ではなく、単一のドッキング計算手法で同じ作用機序を持つとされている複数のリガンドのドッキング計算を実施し、候補として得られた複数のポーズをリガンド間で比較することに着目した。著者らは、これを Comparative Ligand Binding Analysis (以下、CoLBA) 法と提案した。実際に、標的タンパク質であるヒスタミン H1 受容体 (GPCR ファミリー) を実施例として、既知拮抗リガンドとのドッキング計算および CoLBA 法による解析を行った結果、アミノ

酸残基置換実験等で推定されている作用機序に合致した拮抗リガンドの結合モデルをマニュアルドッキングや人的介入なしに予測することができた。

## 2. 方法

### 2-1. CoLBA 法

CoLBA 法のプロトコルは、次の通りとなる：①標的タンパク質に対し、同一の作用機序（作用・拮抗）が想定される複数のリガンドを既存のドッキング計算アルゴリズム等で個別にドッキングした後、②各リガンドについてドッキングスコア順に上位複数の候補ポーズを保存、③候補ポーズをリガンド間で交差比較し相互作用類似尺度および既に得られているドッキングスコアをエネルギー順位スコアに変換した平均値を計算、④相互作用類似尺度とエネルギー順位スコアからなる散布図解析により統計的に有意と表現される候補ポーズの組み合わせを選択し、各リガンドの標的タンパク質に対する結合モデルとする。

#### 2-1-1. 相互作用類似尺度

2つのリガンド A, B についてドッキング計算から得られたそれぞれ任意の候補ポーズ  $a, b$  間の相互作用類似尺度  $S_{ab}$  は、リガンド-タンパク質間の水素結合数 HB と疎水性接触数 HC の類似性  $s_{ab}^{HB,HC}$  により計算される。

$$s_{ab}^{HB,HC} = \frac{\sum_{i=1}^L P(i)_a^{HB,HC} P(i)_b^{HB,HC}}{\sum_{i=1}^L \{P(i)_a^{HB,HC}\}^2 + \sum_{i=1}^L \{P(i)_b^{HB,HC}\}^2 - \sum_{i=1}^L P(i)_a^{HB,HC} P(i)_b^{HB,HC}} \quad (1)$$

ここで、 $s_{ab}^{HB,HC}$  は、Tanimoto 係数で表現される相互作用類似性スコアで、水素結合数類似性  $s_{ab}^{HB}$  と疎水性接触数類似性  $s_{ab}^{HC}$  は、個別に

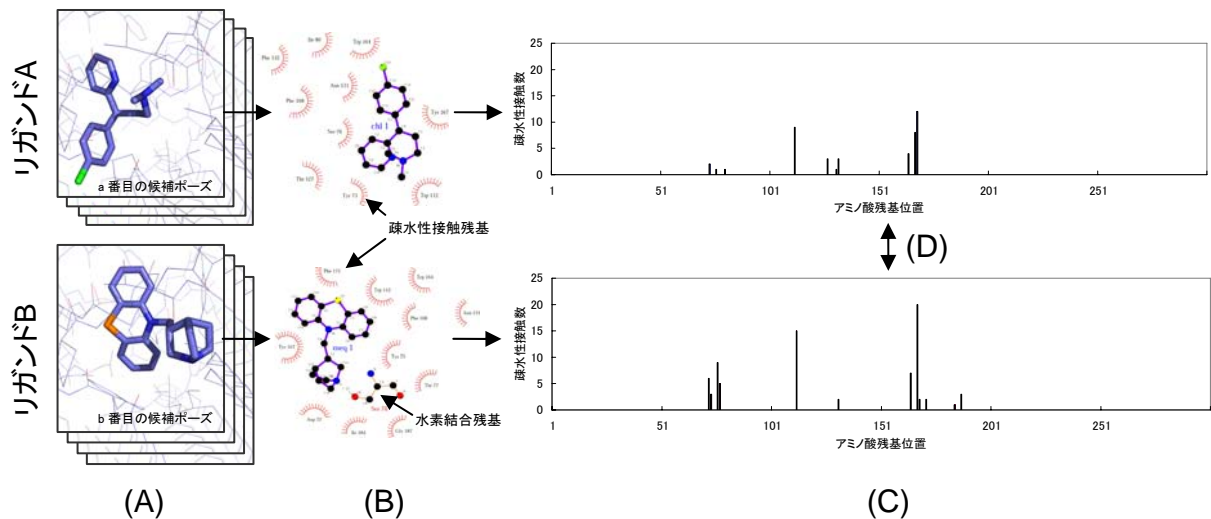


図 1 CoLBA 法における相互作用類似尺度の計算：(A) ドッキング計算により 2 つのリガンドの候補ポーズ (PDB 座標) を生成する。例としてリガンド A の候補ポーズ a とリガンド B の候補ポーズ b を比較対象とする。(B) リガンドとタンパク質間の疎水性接触数と水素結合数を LIGPLOT により測定し、疎水性接触数や水素結合数に関与しているアミノ酸残基番号を列挙する。(C) (B) により列挙した残基について、疎水性接触数や水素結合数をそのままプロファイルとしてタンパク質配列に対して表現する。関与していない他の残基については、0 とする。(D) プロファイル間を Tanimoto 係数で比較する。

計算される。 $P_{a,b}^{HB,HC}$  は、水素結合数や疎水性接触数をそれに関与するタンパク質残基の  $L$  次元のベクトルとして表現 (図 1) されるものである。 $L$  は、タンパク質残基長となる。水素結合数および疎水性接触数は、LIGPLOT プログラム<sup>2)</sup>により同定した。

最終的に相互作用類似尺度  $S_{ab}$  は、 $s_{ab}^{HB}$ 、 $s_{ab}^{HC}$  に重み  $w$  と水素結合残基包括率  $C$  を考慮した、次式で計算される。

$$S_{ab} = w \cdot C \cdot s_{ab}^{HBond} + (1-w) s_{ab}^{HC} \quad (2)$$

$$C = \frac{a_{ab}}{N_{A,B}^{max}} \quad (3)$$

$N_{A,B}^{max}$  は、リガンド A、B の全ての候補ポーズの中の水素結合関与残基数の最大値で、 $a_{ab}$  は、候補ポーズ  $a$ 、 $b$  で共通に存在する水素結合関与残基数である。

### 2-1-2. エネルギー順位スコア

比較するリガンド A、B は、それぞれ分子量や原子組成が異なることから、ドッキング計算スコアをそのまま利用せず、それぞれの

リガンドについてドッキング計算で 1 位のドッキングスコアからの差を任意の候補ポーズ  $a$ 、 $b$  の差分スコアとして定義した (1 位の候補ポーズは、0 となる)。

$$\Delta E_{a,b} = E_{A,B}^{lowest} - E_{a,b} \quad (4)$$

そしてリガンド A、B の任意の候補ポーズ  $a$ 、 $b$  間のエネルギー順位スコア  $ES_{ab}$  は、それぞれの差分スコアの平均で評価した。

$$ES_{ab} = \frac{|\Delta E_a - \Delta E_b|}{2} \quad (5)$$

### 2-2. 散布図解析

リガンド A、B の各候補ポーズをリガンド間で総当りの交差比較し、 $S_{ab}$  および  $ES_{ab}$  の値を計算後、それぞれ標準偏差で規格化し散布図を作成する。各々の候補ポーズペアの特性は、散布図の原点を通る直線  $ES_{ab} = S_{ab}$  からの垂線の距離 (Z-score) で評価する。

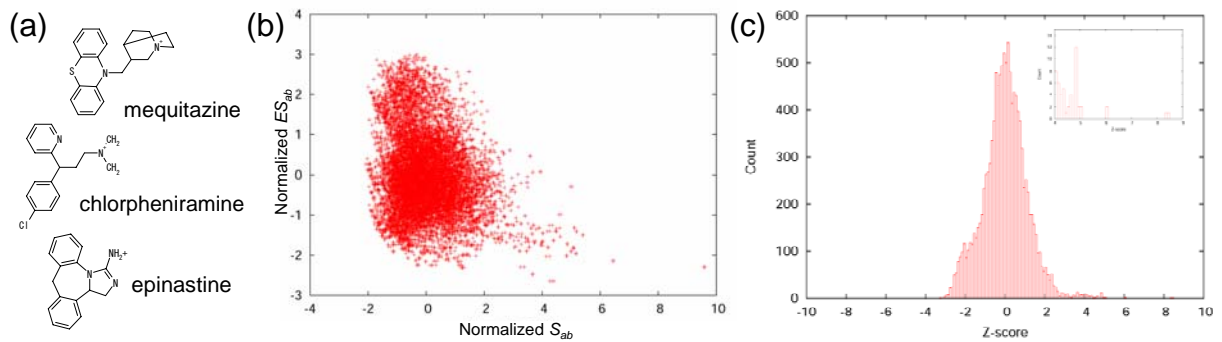


図 2 3つの拮抗リガンドのCoLBA法による解析結果：実施例として用いたヒスタミンH1受容体の既知拮抗リガンド (a)。CoLBA法の散布図結果 (b) と Z-score 測定結果のヒストグラム (c)

$$Z\text{-score} = \text{sign}(S_{ab}, ES_{ab}) \times \sqrt{\left(\frac{S_{ab} - ES_{ab}}{2}\right)^2 + \left(\frac{ES_{ab} - S_{ab}}{2}\right)^2} \quad (6)$$

$$\text{sign}(S_{ab}, ES_{ab}) = \begin{cases} -1: & ES_{ab} > S_{ab} \\ 1: & \text{otherwise} \end{cases} \quad (7)$$

分布から大きく右下に位置するプロット (ドッキングスコアでも上位で、かつ、リガンド間でタンパク質に対する相互作用様式が似ている) が理想的な候補ポーズペアとなる。

### 2-3. タンパク質立体構造モデリングとリガンドドッキング計算

CoLBA法の実施例として用いたヒスタミンH1受容体は、構造未知の標的タンパク質であるため、類縁タンパク質である視覚ロドプシン (PDB: 1F88\_A) を鋳型に比較モデリング法により構築した。ヒスタミンH1受容体と視覚ロドプシンのアラインメントは、GPCRDB<sup>3)</sup>より、Class A Rhodopsin-likeのアミン系サブファミリーから40個の代表配列を加えたマルチプルアラインメント (ClustalW1.8) で実施した。膜貫通ヘリックスを繋ぐループ構造は、モデリングの対象外とした。側鎖構造は、Ponder-Richardsの

ライブラリから得られるものを初期構造とし、その後、Simulated Annealing (AMBER8 / parm96) により側鎖構造サンプリングを行った。最終構造は、サンプリング構造から、Lipid Compatibility Score<sup>4)</sup>を指標に選定した。

リガンドは、ヒスタミンH1受容体の既知拮抗リガンドのクロルフェニラミン、エピナスチン、メキタジンを対象に (図 2a)、ヒスタミンH1受容体のリガンド結合部位上で、それぞれのリガンドの側鎖構造の自由度を考慮したドッキング計算を行い、結合エネルギーの安定なものから順に25程度を各リガンドの候補ポーズとした。リガンドの構築とドッキング計算には、CCG社のMOEパッケージを利用した。

### 3. 結果と考察

ヒスタミンH1受容体の構造予測を行った後、クロルフェニラミン、エピナスチン、メキタジンのドッキング計算を行った結果、それぞれ25、25、19の候補構造が得られた。これらの候補構造について、11875通りの組み合わせでCoLBA法による解析を行った。3種のリガンドの比較では、相互作用類似尺度、エネルギー順位スコアは、3通りのペア比較の平均により算出した。図 2bは、結果を散布図により表現したものである。この散

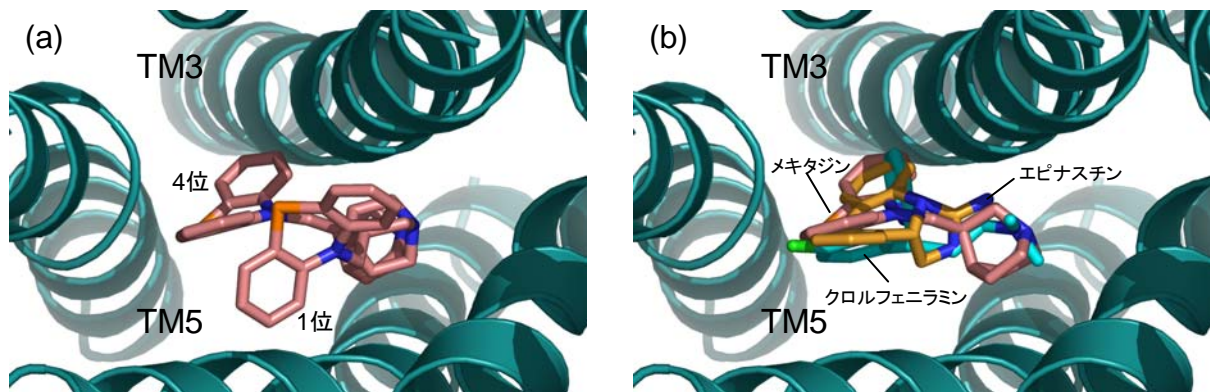


図 3 ヒスタミン H1 受容体 (リボン表示) へのリガンド結合の分子グラフィックス: ドッキング計算によって評価されたメキタジンの 1 位構造と 4 位構造の比較 (a)。4 位構造は、CoLBA 法によって選定されたものである。CoLBA 法によって選定された Z-score が最大の 3 つのリガンド候補ポーズの重ね合わせ (b)。クロルフェニラミン、エピナスチン、メキタジンは、ドッキング計算ではそれぞれ 1 位、2 位、4 位の候補ポーズ。

布図からわかるように、分布から右下に離れたところにある組み合わせが存在していることがわかる。これは、クロルフェニラミン (1 位)、エピナスチン (2 位)、メキタジン (4 位) の候補ポーズの組み合わせによるものである。Z-score は、8.41 と有意のものであることが確認できた (図 2c)。図 3a は、メキタジンを例に、ドッキング計算の結合エネルギーで一位と評価された候補ポーズと CoLBA 法により選定された 4 位の候補ポーズの重ね合わせである。図 3b は、CoLBA 法で選択された 3 つのリガンドの候補ポーズの重ね合わせである。各リガンドには、カチオン性窒素が含まれているが、それがヒスタミン H1 受容体の特異的な位置 (3 番目の膜貫通ヘリックス (TM3)) に共通して配向し、疎水性の官能基は、TM5 に対してほぼ同一の空間に存在していることがわかる。このようなファルマコフォアと呼ばれる機能に関係する官能基の空間的特性から、ドッキング計算のみで決定される最安定構造よりも CoLBA 法により選定された候補ポーズが作用機序を説明する上でも適切なことがわかる。詳細は割愛するが、ヒスタミン H1 受容体について、このファルマコフォアが情報伝達を拮抗するモデルとして理想的であるこ

とが、アミノ酸残基の変異実験等からも確認されている。

CoLBA 法によるドッキング計算結果の補正的処理は、非常に単純なプロトコルであるが、創薬現場の専門家から見て直感的に良い結果が得られやすい (ここでは紹介できなが、他の実施例でも良好な結果が得られている)。ドッキング計算結果をコンセンサスにより選び出すアプローチは、既に創薬現場で行われているが、従来の方法は、候補ポーズ間のリガンド構造に着目した類似尺度でコンセンサス評価を行っている。この場合、結果がリガンドの構造類似性の信頼性に依存することになり、一見、リガンド間の構造骨格や官能基組成に共通性がない場合は困難となる。これに対し、CoLBA 法は、候補ポーズ間の類似性をリガンド自身の構造特性 (配置・配向) ではなく、相互作用情報のプロファイルとしてタンパク質配列に一度変換することで、情報科学的処理による比較計算を容易にし、リガンド同士の類似性に依存しない比較法を実現した。

もう一つの長所としては、今回のように標的タンパク質が実測された構造ではなく解像度の低下が伴う計算機モデリングでも適用範囲が広がる点が挙げられる。一般のドッ

キング計算の結合エネルギー計算は原子座標にとっても敏感で、タンパク質側鎖の向きが少しでも変わっただけで数十 kcal/mol の変動が生じる。CoLBA 法では、ドッキング計算の結合エネルギーだけではなく、水素結合数や疎水性接触数といった粗視化した情報に基づいた相互作用類似尺度を評価軸に取り入れたことで、標的タンパク質構造精度に対して柔軟な手法であると考察できる。

ヒスタミン H1 受容体をはじめ GPCR などの標的タンパク質についても解像度の高い結晶構造が得られ、ドッキング計算でも十分な構造探索と厳密な結合エネルギー評価法が開発されれば、補正技術は将来不必要になるであろう。そう考えると CoLBA 法は、一時的な橋渡しの役割に過ぎないのかもしれない。しかし、一刻でも早く新薬創製が望まれる創薬分野において、この手法が創薬現場の専門家の人的介入に代わる基盤技術となり、新薬が少しでも早く世の中に誕生することに貢献できるのであれば幸いに思っている。

#### 4. 参考文献

- [1] Bujnicki, J.M. *et al.* (2001) Structure prediction meta server. *Bioinformatics*, **17**, 750-751.
- [2] Wallace, A.C. *et al.* (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, **8**, 127-134.
- [3] <http://www.gpcr.org/7tm/>
- [4] Dastmalchi, S. *et al.* (2001) Modeling of the structural features of integral-membrane proteins reverse-environment prediction of integral membrane protein structure (REPIMPS). *Protein Sci.*, **10**, 1529-1538.