

# 機械学習を用いた暗号的結合部位を持つタンパク質の網羅的予測

大規模知識発見分野 熊田 匡仁

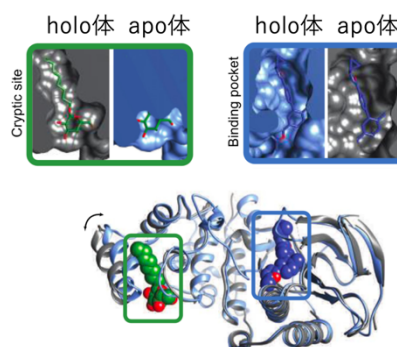
## 1. 背景

創薬は、タンパク質表面上の薬効のあるポケットにリガンドが結合するとタンパク質が機能を失活する鍵と鍵穴の原理に基づいている。これまで研究者はリガンドフリーの構造でポケットの存在するタンパク質を創薬ターゲットとし、薬効の有無について研究してきた。

しかし、多くのタンパク質は、リガンドフリー構造では容易に検出できない低分子結合ポケットを持っていることが知られている。つまりは、アポ構造ではポケットを形成していないが、ホロ構造ではポケットを形成する暗号的結合部位と呼ばれる部位を持つタンパク質が多く存在する[1]。

暗号的結合部位を理解し、正確に特定することで創薬ターゲットを拡大することができる[3]。

これまでクリプティックサイトは、フラグメントベースのリガンド探索によって実験的に、また、長時間の分子動力学シミュレーションとフラグメントドッキングによって計算的に特定されてきた[2]。



Peter Cimermancic et.al, PubMed 2016

## 2. 研究目的

本研究ではアポ構造のタンパク質について、薬効がありポケットになり得る表面の凹みと薬効のない表面の凹みを分類する機械学習モデルを作成する。分析対象タンパク質データセットに対し、作成した機械学習モデルを適用することで、表面上の凹みから暗号的結合部位の予測と因子評価を試みる。

## 3. 手法

1. クリプティックサイトを持つ構造的に定義されたアポのデータセットを論文[1]に従って構築する。
2. fpocket(<http://fpocket.sourceforge.net>)というタンパク質表面上のポケット検出ソフトウェアを用いて構築したデータセットに対し、特徴量を作成する。
3. fpocket では、各タンパク質について暗号的結合部位になり得る凹みとその他の凹みを共に検出するため、pymol を用いて目視で暗号的結合部位の凹みかどうかを確認し、ラベリングした。

4. 3.までで構築したデータセットを学習データとし、決定木ベースのモデルを用いて暗号的結合部位の凹みか否かの分類するモデルを作成する。
5. 機械学習モデルの分類に関して、特徴量の各要素からどの要素が寄与しているかを分析。

## 結果

—

## 参考文献

- [1] Peter Cimermancic et al. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol.* 2016 Feb 22
- [2] Kimura SR et.al. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J Chem Inf Model.* 2017 Jun 26
- [3] Antonija Kuzmanic et.al. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations, *Chem. Res.* 2020, 53, 3, 654–661 Publication Date: March 5, 2020