1. 序論

1.1 研究背景

グラフィカル ユーザー インターフェイス, アプリケーション

自動的に生成された説明生物学的機能には、タンパク質と他の分子（小型リガンドや高分子など）との結合が関与していることが多い。通常これらの相互作用は、タンパク質の構造に定義された結合部位で起こる。結合部位の位置を知ることは、多くの応用が可能である。結合部位に関して、特に低分子の結合部位は多くの場合、露出した凹ポケット内に位置しており、表面積が増加することで分子内相互作用が最大化される。凹ポケットは、タンパク質のリガンドのない構造に存在しており、このような結合部位は結合ポケットと呼ばれている。しかしながら一方で、リガンドが存在しない通常（アポ構造）状態では閉じているが、結合したとき（ホロ構造）に形成される、または短時間だけ過渡的に開く、隠れたリガンド結合部位が存在することが知られている。このような結合部位はクリプトサイトと呼ばれ（図1）、新たな創薬標的としての応用が期待されている[1]。たんぱく質の結合ポケットを特定するために、多くの計算手法が開発されている。具体的には、以下のような原理に基づいている。

図 1クリプトサイトと通常のポケットの例(アポ構造：2ZB1A, ホロ構造：2NPQA)

1. タンパク質表面の凹み
2. ファンデルワールス項を含むエネルギー関数
3. 既知の結合ポケットとの幾何学的・物理化学的な類似性
4. 異なる特徴を組み合わせて使用する複合的なアプローチ

残念ながら、250Å3以上のポケットを持つと判断されたタンパク質構造は全体の60%程度であり、その多くは薬効のないポケットの可能性があるため、まだ未知の結合ポケットの知見に基づいたリガンド探索ができる可能性があります。結合ポケットとは対照的に、タンパク質のリガンドを含まない構造では、暗号的なサイトは容易に検出できない。例えば、相互作用するタンパク質間の大きくて平坦な界面は、低分子の結合と相まって構造変化を起こすタンパク質界面のいくつかの例が最近報告されているが、薬物を投与することはできないと考えられていました。

現在のところ、これまで発見されているクリプトサイトの多くは、構造生物学解析によって決定されたリガンドと標的タンパク質のホロ構造とアポ構造の比較によって、偶然確認されるものが多い。クリプトサイトを有するタンパク質をアポ構造から予測することができれば、新規標的タンパク質発見が可能になり、新たな創薬研究の展開が期待される。現在、クリプトサイトを誘導する特徴的なフラグメント分子を共溶媒した実験や、分子動力学シミュレーションなどにより、クリプトサイトを予測する手法の開発への取り組みがなされているが、フラグメント分子の汎用性や、大規模なシミュレーションに時間を要するなど課題が多い[1-3]。

1.2 研究目的

本研究では、２つの目標がある。

第１に、アポ構造のタンパク質構造を入力として、Fpocket[4]というポケット検出ソフトウェアを用いてクリプトサイトと表面の凹みに対応する特徴量を可能な限り生成する。なお、作成するデータセット（学習・テスト）については、Peter Cimermancicら[1]の論文のサポート資料（表１と表５）から作成した。その後、クリプトサイトの有無を分類する機械学習モデルを作成することである。

第２に、作成した機械学習モデルを用いて、テストデータに対してクリプトサイトの有無を予測する。予測結果と実際のタンパク質のアポ・ホロペアの表面構造を確認し、予測を誤る場合の表面構造に何か原因がないかどうかを分析する。さらに、作成した機械学習モデル

から因子評価を試み、クリプトサイトを検知するソフトウェア開発の指針を見出すことである。