7. 結論

本研究ではアポ構造のタンパク質構造を入力として、クリプトサイトの有無を分類するSVMを用いた機械学習モデルを作成した。テストデータについて F1値で71.0%の性能を達成した。機械学習モデルが正答した場合と誤答した場合について、タンパク質のポケット周辺の表面構造を確認した。現状としてモデルが予測を誤答する場合はアポ構造においてクリプトサイトになりうる凹みが浅く、その他の凹みと判定を誤ったと考えられる。モデルが学習において重要と判断した特徴量を可視化した。その結果、特徴量の内、Hydrophobicity score、Alpha sphere density、Polarity scoreが重要特徴量の上位であることがわかった。今後は、以上得られた知見をもとに、まずはFpocket[4] をカスタマイズし、クリプトサイト検出精度の向上に取り組みたい。さらに、そこで得られた知見をもとに、最終的には独自のクリプトサイト検出ソフトウェアの開発に試みたい。