

タンパク質のアミノ酸残基間の相互作用エネルギーに関する機械学習

Application of machine learning to interaction analysis of residues in proteins

立教大理¹, 神戸大院シス情², 産総研³, 星薬科大⁴, 東大生研⁵

○畑田 峻¹, 八幡 研一郎¹, 奥脇 弘次¹, 田中 成典², 古明地 勇人³, 福澤 薫⁴, 望月 祐志^{1,5},

Rikkyo Univ.¹, Kobe Univ.², AIST.³, Hoshi Univ.⁴, Univ. Tokyo.⁵,

○Ryo Hatada¹, Kenichiro Yahata¹, Koji Okuwaki¹, Shigenori Tanaka²,

Yuto Komeiji³, Kaori Fukuzawa⁴, Yuji Mochizuki^{1,5}

E-mail: fullmoon@rikkyo.ac.jp

【序】フラグメント分子軌道(FMO)法は、タンパク質のような巨大分子系を量子論的に扱える手法であり、この解析から得られるフラグメント間相互作用エネルギー(IFIE)はタンパク質のアミノ酸残基間の相互作用解析に有用である。しかし、タンパク質の構造揺らぎを考慮する場合、構造サンプリングを増やして解析を行うため、計算コストや出力される IFIE データが劇的に増加する。また、膨大なデータの中から重要な情報の抽出を行う必要がある。そこで先行研究[1,2]では、機械学習を用いて IFIE 値と物理量の相関解析や IFIE 値の相関解析と予測を試みた。しかし、非荷電アミノ酸残基ペアの解析精度が不十分である等の課題があった。そこで本研究では、記述子セットの改良、および新たな解析手法を用いて解析精度の向上を試みた。

【計算・解析スキーム】先行研究[1,2]と同様、Trp-Cage や Ubiquitin 等のタンパク質を対象とした。先ず AMBER[3]を用いて水溶液中条件で MD 計算を行い、100 個の構造サンプルのデータを得た。次に ABINIT-MP[4]を用いて FMO-MP2/6-31G*レベルで構造毎に各残基ペアの IFIE を算出した。先行研究では、IFIE 値や FMO 計算から得られる電荷や距離情報等の物理量を 1 つにまとめて解析していたが、今回、GAUSSIAN [5]で算出した双極子モーメントと分極率の値を説明変数として追加した。また、データセット内で、アミノ酸個々で記述されていた説明変数をペア化した。その際、先行研究で用いていたニューラルネットワーク回帰に加え、Scikit-learn [6] によってサポートベクター回帰(SVR)等の複数の回帰手法を用いた。

【結果】ここでは、Trp-Cage の SVR の結果を Figure 1 に示す。約 95% の高い予測精度が得られており、IFIE に対して重要な特徴量を確認したところ電荷変数が重要であることがわかった。

【謝辞】本研究は、文科省ポスト「京」(富岳)重点課題 6、立教 SFR から支援を受けている。

【文献】[1] 望月他, 2016 年 第 77 回応用物理学会秋季学術講演会, 14p-P18-7. [2] 望月他, *J. Comp. Chem. Jpn.* **16**, 119 (2017). [3] <<https://ambermd.org/>>. [4] Tanaka et al., *Phys. Chem. Chem. Phys.* **16**, 10310 (2014). [5] <<https://gaussian.com/>>. [6] <<https://scikit-learn.org/stable/>>.

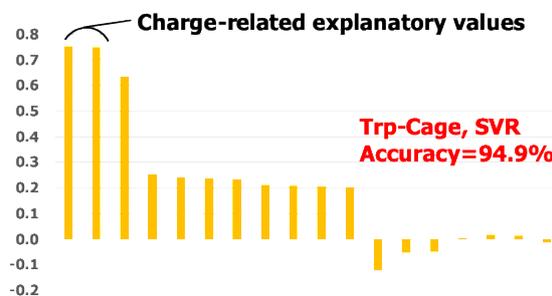


Fig 1. Correlation coefficient of explanatory variables and accuracy