

## FMO データを用いた機械学習モデルの開発 Development of machine learning models using FMO data

加藤 幸一郎

Koichiro Kato

九州大学 大学院工学研究院 応用化学部門

Department of Applied Chemistry, Graduate School of Engineering, Kyushu University

生体関連分子の立体構造研究や、薬物分子設計における相互作用研究では、タンパク質などを計算対象とした分子シミュレーションの利用が一般的となっている。これらのシミュレーションにおいて分子のエネルギーや各原子に働く力は、物理化学的描像に基づくポテンシャル関数で表現され、パラメータセットと合わせて古典分子力場と呼ばれる。分子シミュレーションの精度は使用する古典分子力場によって決定されるため、古典分子力場の更なる精度向上が期待されている。しかしながら、タンパク質の生体高分子と医薬低分子化合物等との相互作用は非常に複雑であり、例えばハロゲン結合等の相互作用は単純なポテンシャル関数形では表現することは困難である。精度向上の意味では、計算対象となる分子系のダイナミクスを全て量子化学(QM)的に計算するという作業も考えられるが、たとえ「富岳」をもってしても計算コストの面で非現実的である。創薬をはじめとした生体関連分子研究にかかる時間・コストの更なる削減、コロナウィルスの様なパンデミックに対応するための迅速な研究開発など、更なる分子シミュレーション活用が必須になっていく事は明白であるものの、そのためには古典分子力場の計算コストで量子化学(QM)レベルの計算精度の実現に向けた従来力場の枠組みに囚われない大転換、すなわちゲームチェンジが不可欠である。このゲームチェンジの鍵になると我々が注目し研究開発に着手しているのが機械学習力場である。これはQM計算で作成したデータをArtificial Neural Network(ANN)により学習することで、低コストで高精度な力場を構築するというコンセプトの手法[1]である。類似手法も合わせて無機結晶材料や低分子化合物へと適用範囲の広がりを見せているが、巨大な生体関連分子で機械学習力場を構築した事例は未だ報告されていない。これは高品質なタンパク系の学習データセットの構築が困難であることが一因であると考えられる。

我々は、この高品質なタンパク系の学習データセットの構築をフラグメント分子軌道法(FMO法)[2, 3]の活用により克服し、機械学習力場の構築に向けた検討を進めている。FMO法により生み出される唯一無二のデータを用いた機械学習モデルの構築は、力場開発に留まらず非常に広範なポテンシャルを秘めていると我々は考えている。本講演では、FMOデータを用いた機械学習による原子電荷予測や相互作用予測、さらには機械学習力場構築に向けた状況を紹介する。

[1] Behler, J., Parrinello M., Phys. Rev. Lett. 98, 146401 (2007)

[2] Fedorov, D. G., Nagata, T., Kitaura, K., Phys. Chem. Chem. Phys. 14, 7562-7577 (2012).

[3] Recent Advances of the Fragment Molecular Orbital Method: Enhanced Performance and Applicability. ed. by Mochizuki Y, Tanaka S, Fukuzawa K, Springer, (2021).