

FMO 計算結果の機械学習による解析

立教大理¹, 東大生研², 産総研³, ナレッジコミュニケーション⁴,
星薬科大薬⁵, 神戸大院情報⁶

○望月祐志^{1,2}, 古明地勇人³, 齊藤天菜¹, 藤本真悠¹, 飯島潤¹, 阿部鷹也¹,
奥脇弘次¹, 土居英男¹, 奥沢明⁴, 牧村健⁴, 中西貴哉⁴, 福澤薫^{2,5}, 田中成典⁶

Machine learning assisted analyses on the results of fragment molecular orbital calculations

○Yuji Mochizuki^{1,2}, Yuto Komeiji³, Sona Saitou¹, Mayu Fujimoto², Jun Iijima¹,
Takaya Abe¹, Kouji Okuwaki¹, Hideo Doi¹, Akira Okusawa⁴, Takeshi Makimura⁴,
Takaya Nakanishi⁴, Kaori Fkuzawa^{2,5}, Shigenori Tanaka⁶

¹*Department of Chemistry, Rikkyo University, Japan*

²*Institute for Industrial Science, The University of Tokyo, Japan*

³*National Institute of Advanced Industrial Science and Technology, Japan*

⁴*Knowledge Communications, Co Ltd.*

⁵*Graduate School of Pharmaceutical Sciences, Hoshi University*

⁶*Graduate School of System Informatics, Kobe University*

【Abstract】 In the fragment molecular orbital (FMO) calculations, the inter-fragment interaction energies (IFIEs) are of the most useful quantity to grasp the nature of interactions in the given system. However, when the number of sample structures grows (e.g. a hundred), the manual analyses for IFIEs should become intractable. Thus, we have been developing the machine learning assisted analyses on IFIEs with the MS Azure environment. Furthermore, Google's TensorFlow as a representative deep learning tool has been used to interpret the IFIE-map which is a two-dimensional visualization of IFIE values. In this presentation, we brief the workflows of these analyses and show several demonstrative examples.

【序】 フラグメント分子軌道(FMO)法[1-3]は、タンパク質のような巨大分子系をそのまま量子論的に扱える手法の一つで、私たちは独自の FMO プログラム ABINIT-MP[3]を開発してきています。FMO 計算からはフラグメント間の相互作用エネルギー(ABINIT-MP 系では IFIE と略記)が得られるため、対象系の解析に好適で、様々な応用例が報告されてきました。これまでのところ、IFIE 解析は PDB に由来する単一構造に基づくものがほとんどですが、水中での構造揺らぎを考慮した統計的な検討も今後は重要となってきます。実際には、AMBER[4]などを用いた古典力場による分子動力学(MD)シミュレーションによって得られる多数の構造サンプルに対し、一連の FMO 計算を行うこととなりますが、応じて生成される IFIE のデータ量は「人

が把握する」にはあまりに膨大になり得ます。そこで私たちは、マイクロソフト(MS) Azure の機械学習[5]を援用して IFIE のデータセットを簡約することを試みています。今回の発表では幾つかの事例について紹介させていただきます。また、別種の試みとして IFIE の可視化画像(IFIE-map)[3]に関する Google の深層学習ツール TensorFlow[6]の使用例にもふれます。なお、解説記事[7]も適宜参考にしていただければ幸いです。

【IFIE データの解析】 これまでに MS Azure の機械学習を適用したタンパク質は、Chignolin(10 残基)、TrpCage(20 残基)、Ubiquitin(76 残基)、Calmodulin(148 残基)です。以下、最初に解析を試みた Chignolin を例に記します。まず、AMBER を使って水和条件下(Na^+ の対イオンを含む)で MD を行い、その後で 100ps 毎に 100 サンプルを液滴状に切り出し、FMO2-MP2/6-31G*レベルで計算し、一連の IFIE セットを得ました。Chignolin では、1-3 以上の有効 IFIE は 36 個になりますが、各 IFIE を目的関数として残基間の重心距離や電荷重心間の距離などを記述子の候補として Azure で提供されるニューラルネットワーク回帰によって解析しました(75 サンプルで学習、25 サンプルで評価)。一例を挙げますと、N 末の Gly(+)と C 末の Gly(-)はクーロン相互作用が支配的ですが、相関係数の上位には距離系のパラメータがリストされ、「物理化学的な直感」に符合した結果が得られました。他のタンパク質でも、荷電-荷電の残基間相互作用はクリアに整理され、予測値-実測値との対応も良好となります。

【IFIE-map の解析】 IFIE-map では、タンパク質の α -ヘリックス、 β -シートの構造は特徴的なパターンとして可視化されます：前者は対角線に沿ったピクセルとして、また後者は斜め 45 度的に並ぶピクセルとして視られます(安定化領域)。そこで、IFIE-map の画像情報のみから α -ヘリックス、 β -シートが対象となるタンパク質に含まれるか否かを深層学習によって判定することを試みました。Google/DeepMind 社 TensorFlow の学習用に 20 種のタンパク質から古典 MD 計算で各 50 サンプルを求め、FMO2-MP2/6-31G レベルで計算を行って IFIE-map のセットを調製しました。学習終了後に、未知のタンパク質の IFIE-map として β -シートのみの画像を TensorFlow に与えると 76%の確度で判定されました。

【謝辞】 本研究開発は、科研費(16H04635)、並びに FS2020(ポスト「京」)プロジェクト-重点課題 6 から支援を受けています。

【参考文献】 [1] “The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems”, (2009, CRC). [2] D. G. Fedorov, T. Nagata, et al., *Phys. Chem. Chem. Phys.* **14** (2012) 7562. [3] S. Tanaka, Y. Mochizuki et al., *Phys. Chem. Chem. Phys.* **16** (2014) 10310. [4] <<http://ambermd.org/>>. [5] <<https://azure.microsoft.com/ja-jp/>>. [6] <<https://www.tensorflow.org/>>. [7] 望月, 奥沢, *計算工学* **22** (2017) 3539.