プロジェクト名（タイトル）：第一原理大規模データと機械学習を応用し欲しい物性から分子を設計する

利用者氏名：中田真秀

理研における所属研究室名：柚木計算物性物理研究室

利用者氏名：

1.　本課題の研究の背景、目的

化学おける分子の種類はそれこそ無限にあり、我々にとって重要な分子だけでも億単位はあると思われる。それらの正確なデータの蓄積、分析は化学にとって重要である。さて、量子化学計算は非常に発展し、実験と理論計算は化学にとっての両輪となっている。ただ、量子化学の現状として、分子を決定すれば正確な物性、分子構造など手に入れられるが、ある特定の物性がほしい、ある特定の構造を持った分子を作りたい、となると非常に難しくなる。いわゆる逆問題を解かねばならないからだ。計算化学の逆問題は一般に非常に難しいが、これを解決する一つとして、網羅的な分子データベースの構築、および検索システムの構築、機械学習による補完などのプロジェクト PubChemQC プロジェクト http://nakatamaho.riken.jp/pubchemqc.riken.jp/ [1,2,3]として挑んできた。

また、近年の機械学習の発達により化学データへの需要が高まっている。これに資することができれば幸いである。

2.PubChemQC B3LYP/6-31G*//PM6 データセットの構築と解析 [3]

2.1 2016 年に PubChem から PubChem Compound[4]をダウンロード、これについて、なるべく多くの分子について PM6 で分子構造の最適化を行った。PubChem には約 9100 万分子 SMILES 形式で登録されていた。それらについて網羅的に計算を行い、特に電荷を中性にしたものについて OpenBabel [5]を用い、分子の初期構造を生成、この初期分子構造に基づき、PM6 法で分子の安定構造が求めた[2]。

2.2 PubChemQC B3LYP/6-31G*//PM6 データセットの構築[3]

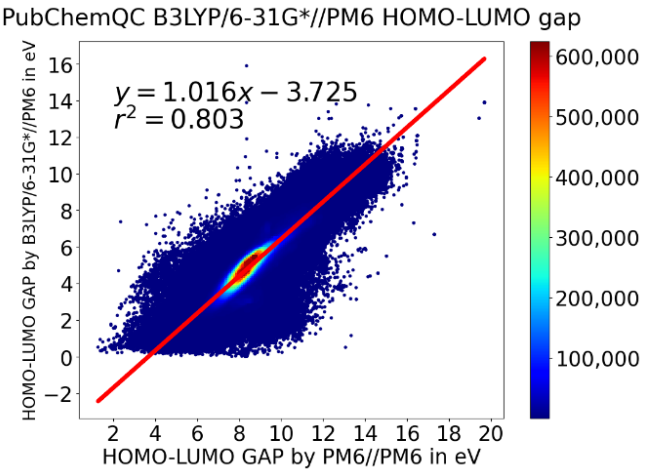
PubChemQC PM6 で求まった 8600 万分子ほぼすべての分子について、GAMESS を用い、B3LYP/6-31G*を用いて電子構造を求めた（H-Kr)。ただし、Rb-Rn については、Def2-SV(P)、また、Stuttgart RSC 1997 ECP も適宜用いた。

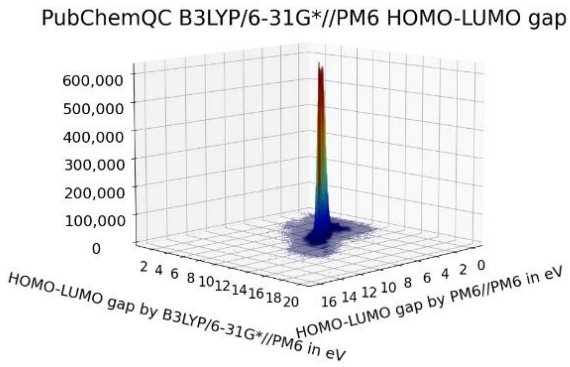
2.3 データの公開について

全データは、CC-BY-4.0 ライセンスの下 https://nakatamaho.riken.jp/pubchemqc.riken.jp/b3lyp_pm6_datasets.html で公開中である。

2.3 HOMO-LUMO gap について

同じ分子構造で、PM6，B3LYP/6-31G*と二通りで電子構造を計算し、特に HOMO-LUMO ギャップを比較した結果を下図にヒートマップおよびヒストグラムを斜め 45 度から見た図を示す。どちらの図もヒストグラムの bin の大きさは 0.1eV で、最大一つの bin に 60 万分子程度入っている。また相関決定係数は 0.8 程度と非常に高い。これは機械学習のモデルの構築で、PM6 のような経験的な量子化学計算法からの補正程度で B3LYP/6-31G*の品質の計算が得られることを強く示唆している。





3. 参考文献

[1] J. Chem. Inf. Model., 2017, 57 6, 1300-1308.

[2] J. Chem. Inf. Model. 2020, 60, 12, 5891-5899.

[3] J. Chem. Inf. Model. 2023, 63, 18, 5734–5754.

[4] Nucleic Acids Res. 2016 Jan 4; 44(Database issue): D1202–D1213.

[5] O'Byole et al., J. Cheminf. 2011, 3:33.

2023 年度　利用報告書

2023 年度　利用研究成果リスト

【雑誌に受理された論文】

“PubChemQC B3LYP/6-31G*//PM6 Data Set: The Electronic Structures of 86 Million Molecules Using B3LYP/6-31G* Calculations” Maho Nakata* and Toshiyuki Maeda, J. Chem. Inf. Model. 2023, 63, 18, 5734-5754

【会議の予稿集】

“3D Molecular Geometry Analysis with 2D Graphs”, Zhao Xu, Yaochen Xie, Youzhi Luo, Xuan Zhang, Xinyi Xu, Meng Liu, Kaleb Dickerson, Cheng Deng, Maho Nakata, Shuiwang Ji, arXiv.2305.13315

【口頭発表】

【ポスター発表】

【その他（著書、プレスリリースなど）】