

課題名(タイトル):

第一原理大規模データ及び機械学習に依る有機イメージセンサーの開発

利用者氏名: ○中田真秀(1)、島崎智実(2)

理研における所属研究室名:

(1)情報システム本部 研究開発部門 計算工学応用開発ユニット

(2)AICS 量子系分子科学研究チーム(旧所属)

1. 本課題の研究の背景、目的

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、および検索システムの構築、機械学習による補完で挑んでいる。今年度は引き続き計算を行いつつ、データベース作成、web サービスの構築を目標とした。

2. 具体的な利用内容、計算方法

前年度に引き続き、GAMESS[2]および Firefly[3]を用いて Pubchem[1]に登録されている分子の計算を行っている。

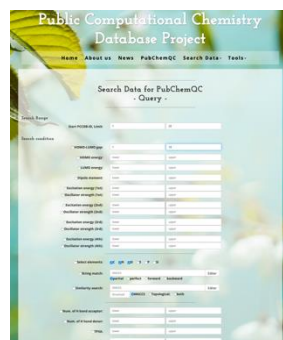
1. Pubchem[1]のデータのダウンロードを行う。
2. Isomeric SMILES 表記から OpenBABEL[4]を使い分子の初期座標を求めた。
3. GAMESS を用い、経験的手法である PM3 を用い、構造最適化を行った。
4. この構造を用い、STO-6G 基底、Hartree-Fock 法で構造最適化を行った。
5. さらに、Firefly[3]を用い、6-31G(d)/B3LYP で構造最適化を行った。
6. 収束パラメータの若干の違いがあるため再び GAMESS で構造最適化を行った。
7. 最後に収束した構造からインプットファイルを作成しなおし、GAMESS で構造最適化を行った。これはインプットファイルに最適化された構造を入力し、その構造が最適化されているかをチェックするためである。
8. 最適化された構造を用い 6-31G+(d)/TD-B3LYP で励起状態を 10 個求めた。
9. 分子量が 500 以上のもの、およびいくつかの分子の混合系については計算を除外している。

データベース構築については、計算結果のアーカイブを展

開、データを cclib で処理し、json 形式に変換、postgresql に読み込んだ。この初期データベースを 2000 分子ずつ区切って HOMO-LUMO ギャップ/HOMO/LUMO/dipole moment/励起エネルギーなどで検索できるようにした。さらに、それに対する web interface を作った。

3. 結果

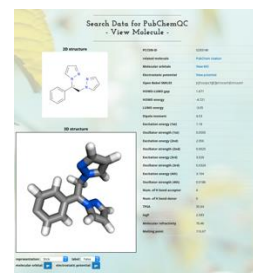
<http://pccdb.org/> にその結果がある。まずは、検索を行ってみる。http://pccdb.org/search_pubchemqc/query



からは ID を 2000 個ずつ(負荷を避けるため)入力可能である。その下に HOMO-LUMO ギャップ、HOMO エネルギー…と検索キーが続く。今回試しで 1000 分子、HOMO-LUMO が gap 1-2ev の分子を検索すると、以下のように 611 分子得られた(左下図)



これを一分子ずつ詳しく表示させたものが下図となる



4. 今後の計画・展望

分子種が 400 万程度と少ないため、利用の幅が小さい。計算の品質を下げつつも増強すること、機械学習による物性値予測(現在進行中)を行っている。これらからイメージセンサーの開発につなげたい。

5. 参考文献

Maho Nakata and Tomomi Shimazaki, "PubChemQC Project: a Large-Scale First-Principles Electronic Structure Database for Data-driven Chemistry", J. Chem. Inf. Model., 2017, 57 (6), pp 1300-1308.

平成 30 年度 利用研究成果リスト

【雑誌に受理された論文】

(平成 29 年度)

Maho Nakata and Tomomi Shimazaki, "PubChemQC Project: a Large-Scale First-Principles Electronic Structure Database for Data-driven Chemistry", J. Chem. Inf. Model., 2017, 57 (6), pp 1300-1308.

【その他(著書、プレスリリースなど)】

<http://pccdb.org/> の立ち上げ、メンテナンス