

課題名 (タイトル) :

第一原理計算による分子の物理化学データベース構築

利用者氏名 : 中田真秀

所属 : 情報基盤センター

はじめに

化学の研究は「ある物性、機能を持った分子を設計したい」ということにつける。逆に個々の分子を特定すれば、量子化学計算を行うことで、実験結果と同等または凌駕するような計算結果が得られる。量子化学計算は精度は高いが、化学を主導しているか、といえは否であろう。

この研究では、量子化学計算を使いデータベースを作成し、そのデータから、完全なる *in silico* での新規機能分子の創出、提案を行うシステム構築を目指している。元来、新規物質を作るといのは人間の経験や勘で行われるのだが、コンピュータでも近年、大量のデータ(ビッグデータ)をもとにした機械学習の発達により、IBM ワトソンのように非常に高いレベルの人間の思考が再現されつつある。そこで、我々はまず、分子の正確なデータの蓄積を行う。幸い、構造式レベルでは、PubChem プロジェクト (<https://pubchem.ncbi.nlm.nih.gov/>) [1] に 4000 万分子ほど登録されており、さらにこれは日々増えてきている。これらの分子の網羅的な量子化学計算により、実験値とほぼ同等な分子構造、物理化学的性質の基礎的データの蓄積を行う。それをを用い、得られたビッグデータの解析を通して新規物質の開発、開発支援システムの構築の研究を行うのが最終的な目標である。

計算手法

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。

計算結果

2016/2 現在 3,066,709 個分子が計算されている。これは昨年に比べて、ほぼ 2 倍程度増加した(昨年度同時期は 1,532,781)。さらに昨年度は計算、公開できていなかった励起状態の計算も 2,676,326 分子についてもデータを公開できた。構造は JSmol[5]を用い、何も設定しない状態でも web ブラウザから見る事が可能である。一日数千分子程度の電子状態計算を行えたことになる。

- [1] [Bolton E, Wang Y, Thiessen PA, Bryant SH, Chapter 12 IN Annual Reports in Computational Chemistry, Volume 4, American Chemical Society, Washington, DC, 2008 Apr.](#) [2] [M.S.Gordon, M.W.Schmidt pp. 1167-1189, in "Theory and Applications of Computational Chemistry: the first forty years" C.E.Dykstra, G.Frenking, K.S.Kim, G.E.Scuseria \(editors\), Elsevier, Amsterdam, 2005.](#) [3] Alex A. Granovsky, Firefly version 8.0 [4] N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. J. Cheminf. (2011), 3, 33. [5] <http://www.jmol.org/>

平成 27 年度 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

なし

【国際会議などの予稿集、proceeding】

The PubChemQC project: A large chemical database from the first principle calculations, [Nakata Maho](#).
AIP Conf. Proc. 1702, 090058 (2015).

【国際会議、学会などでの口頭発表】

Maho Nakata, Advanced Center for Computing and Communication, RIKEN,
“The PubChemQC Project: a large chemical database from the first principle calculations”, Kobe
workshop for material design on strongly correlated electrons in molecules and materials, 2016/2/17
<http://www.aics.riken.jp/labs/cms/workshop/201602/index.html>

【その他（プレスリリース、学術会議以外の一般向けの講演など）】

なし