# 1L05

## 第一原理計算による分子の物理化学データベース構築

○中田真秀

理化学研究所 情報基盤センター

maho@riken.jp

### 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」ということにつきる。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。

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う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 を用い、6-31G(d)/B3LYP を用いて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 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、その構造が本当に正しいかをチェックするためである。
8. また、さらに同じ構造を用い 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

2014/4/21 現在、67000 個の分子のデータを全て http://pubchemqc.riken.jp で公開している。Java がインストールされていれば、構造を Jmol[5]を用いて web ブラウザから見ることが可能である。理研 RICC などを利用し、一日 1000 分子程度の構造最適化を行っている。また、励起状態の計算結果については近日公開中である。

[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/