

課題名(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:

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

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。昨年度から今年度にかけて PubChem[1]に登録されている分子について中性、カチオン、アニオン、スピン反転状態についてそれぞれ PM6 法により分子の構造最適化計算を行い、[2]さらに中性分子について、PM6 で得られた分子構造の上により精密な計算法である B3LYP/6-31G*による電子構造計算を行った。

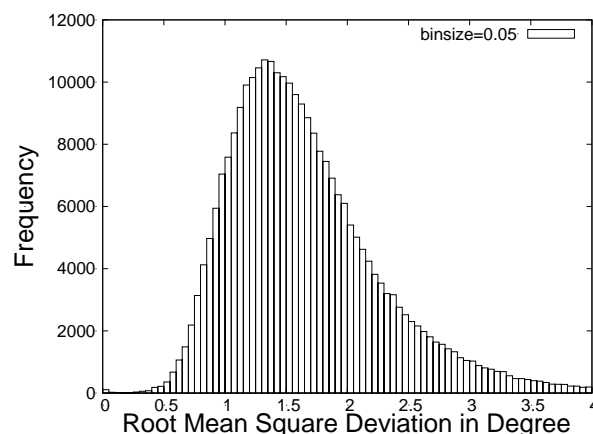
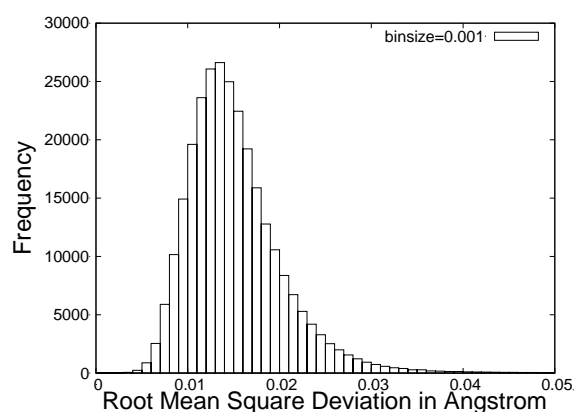
2. 具体的な利用内容、計算結果など

1. PubChem プロジェクトから 2016 年にダウンロードした。この中には全体で 91,679,247 分子存在し、さらに、分子量が 1000 以下かつ中性の 88,886,036 分子が存在した。これらを計算対象とした。
2. Isomeric SMILES 表記から Open Babel [3]を使い分子の初期構造を求めた。
3. Gaussian09 を用い、初期構造から分子の構造最適化計算を PM6 法で行い、計算が成功した分子については、同じ分子のカチオン、アニオン、スピン反転状態の構造最適化計算を行い[4]で公開した。
4. PM6 計算による中性分子の構造最適化は 86,213,135 分子成功した。これらについてより精密な電子構造を求めるため、GAMESS を用いて B3LYP/6-31G*計算を行った。2020/1/7 現在 76,168,195 分子計算が終了している。これは、全体の 88.3%程度であった。

3. 結果

PM6 は経験的方法なので、一般には精度が低い。ただ、分子構造は比較的高精度に求まるとされる。Stewart は 413~5,154 分子について、実験値などと比較し結合距離は 0.025~0.091Å 程度の誤差、244~1,681 分子について同様に実験値などと比較し、結合角は 3.1~7.9 度程度の誤

差があると報告している[5]。我々はすでに B3LYP/6-31G* レベルで構造最適化を行った計算結果を持っている[6]。20 万分子程度で PM6 と B3LYP/6-31G*での結合距離、結合角を比較し、どの程度分子構造が正しく求まっているか比較した。下図は PM6 と、B3LYP/6-31G*計算による結合距離の RMSD、および結合角の RMSD である。平均して結合距離 0.015Å 程度、結合角 1.4 度程度と、確かに PM6 による分子構造、B3LYP/6-31G*による分子構造は十分に近いという結果が得られた(投稿中)[2]。



5. 参考文献

- [1] [Nucleic Acids Res.](#) 2016 Jan 4; 44(Database issue): D1202-D1213.
- [2] Nakata et al., <https://arxiv.org/abs/1904.06046>
- [3] O'Byole et al., [J. Cheminf.](#) 2011, 3:33
- [4] http://pubchemqc.riken.jp/pm6_dataset.html
- [5] *J Mol Model.* 2007 Dec; 13(12): 1173-1213.
- [6] *J. Chem. Inf. Model.*, 2017, 57 (6), pp 1300-1308