

## QCLO 法に基づくタンパク質全電子シミュレーションプログラムの開発

(東大生研) ○平野 敏行, 恒川 直樹, 上村(西野) 典子, 佐藤 文俊

【緒言】 QCLO 法[1, 2]は分子の部分系で擬カノニカル局在化軌道(QCLO; Quasi-Canonical Localized Orbital)を段階的に作成・拡大することで、タンパク質などの大規模分子の SCF 計算の初期値を効率良く算出する方法である。大規模分子の全電子量子化学計算において、QCLO 法は SCF 繰り返し計算の良い初期値を与える。しかし、その計算手法は手順が多く煩雑であり、大規模分子ほど手作業による作成は困難になる。1 点計算だけでなく、構造最適化や *ab initio*/DFT MD などのシミュレーションの場合でも、構造変化のたびに QCLO 法を利用することは極めて有効である。本研究では、QCLO 法に基づき *ab initio*/DFT MD などの全電子シミュレーションを半自動的に行う、再利用可能なプログラム・フレームワークを作成し、容易に柔軟なシミュレーションが可能となるワークベンチを作成することを目的とした。

### 【QCLO 法】

QCLO は一種の局在化軌道であり、かつ、例えばアミノ酸残基や数残基のペプチド鎖など、任意の原子グループ単位内で局在化しているが、その単位内ではカノニカル分子軌道(MO)に類似したものとなっている。ここで、局在化している単位をフラグメントと呼ぶ。タンパク質の計算においては、このポリペプチド鎖などサブユニットのフラグメントの QCLO をつなぎ合わせることで、より大きなポリペプチド鎖の高品位な初期値を作成し、段階的に伸長することによってタンパク質の全電子計算に成功している[1]。この時、タンパク質の構造に着目してサブユニットをグループ化すると、効率良く精度の良い初期値が得られる[2]。具体的な計算方法は以下の通りである。

- 手順 1. 計算対象を適当なサブユニット(フレーム分子)に分割する。
- 手順 2. 各フレーム分子の電子状態計算を行い、分子軌道(MO)を求める。
- 手順 3. 得られた MO から局在化軌道(LO)を求める。
- 手順 4. 分割した軌道空間(フラグメント)のうち、QCLO 作成に必要なフラグメントを選び出し、 $C_{LO}$ を作成する。
- 手順 5. フレーム分子全体の Fock(Kohn-Sham) 行列 $F$ を線形変換することにより、フラグメントの Fock(Kohn-Sham) 行列 $F'$ を得る。

$$C_{LO}^\dagger F C_{LO} = F'$$

- 手順 6. これを対角化することにより、QCLO を得る。

$$F' C' = C' E$$

$$C_{QCLO} = C_{LO} C'$$

- 手順 7. 各フレーム分子の $C_{QCLO}$ から計算対象に必要な QCLO を結合し、 $C_{sum}$ を作成する。
- 手順 8. Löwdin の直交化法[3]、または重み付き直交化法[4]により、計算対象分子の初期値 $C$ を得る。

これまで QCLO を求める際に必要であった手順 3 の LO(Localized Orbital)の計算には、Population Localization 法[5]を利用してきた。この LO 計算アルゴリズムの計算量は基底関数の総数  $N$  に対して  $O(N^4)$ になり、大規模分子に適用すると最終的に計算律速となっていた。LO 計

算には幾つかの方法が知られているが、本研究では RLMO 法[6]による LO 計算により計算時間の短縮を試みた。RLMO 法では NBO の生成に類似した方法により作成することができ、 $O(N^3)$  となるため、Population Localization 法に比べ大幅な計算時間の短縮が期待できる。大規模分子における効果的な QCLO 法による初期値作成法について、当日報告する。

#### 【半自動計算プログラムの開発】

流れ作業や定型処理を容易に記述し、ユーザーが望むシミュレーションを円滑に行うために、グルー言語(glue language)の一つとして知られる Python を使用した。全電子計算プログラム ProteinDF をはじめとする外部プログラムの実行では、バッチ処理、MPI 実行環境などマシン環境に即した実行様式を整える必要があり、ユーザー側でカスタマイズできなければならない。本研究では Python を使用してユーザーの実行環境に応じた外部プログラムを呼び出すことによってこの問題を解決した。巨大分子の全電子計算は MPI-OpenMP ハイブリッド並列で行う、複数の低分子の量子化学計算を並行に実行する、別の量子化学計算の結果から得られた原子電荷を新たなモデルの外場として用いるなど、柔軟なシミュレーションを行うことができるようになった。ProteinDF で求めた核座標でのエネルギーの解析的微分から力を算出し、これを ProteinMD へ転送することで、*ab initio*/DFT MD シミュレーションも可能である。QCLO 法ではタンパク質の 2 次構造に基づきサブユニットを形成すると効率良く初期値を作成することができる[2]が、2 次構造判定プログラムと連携することにより、MD シミュレーションにより構造変化が起きても随時適した計算シナリオを作成することもできる。Python が備えるオブジェクト指向言語としての性質を利用し、分子ならびにモデリング操作、各種プログラムの実行をクラス化することによって、一度開発したスクリプトを再利用することが出来る。SCF 繰り返し計算において収束が困難な、高度な技術が必要なヘムタンパク質などの計算条件・シナリオをデータベース化することにより、容易に再実行・応用が可能になると考えられる。今後 GUI とも連携して柔軟なシミュレーションが行える環境を整備していく予定である。

#### 【参考文献】

- [1] H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, *Mol. Phys.*, **1**, 81 (2003).
- [2] N. Nishino-Uemura, T. Hirano, F. Sato, *J. Chem. Phys.*, **127**, 184106 (2007).
- [3] P.O. Löwdin, *J. Chem. Phys.*, **18**, 515 (1950).
- [4] H. Kashiwagi and F. Sasaki, *Int. J. Quant. Chem. Symp.*, **7**, 515 (1973).
- [5] J. Pipek and P.G. Mezey, *J. Chem. Phys.*, **90**, 4916 (1989).
- [6] F.L. Gu, Y. Aoki, J. Korchowiec, A. Imamura, B. Kirtman, *J. Chem. Phys.*, **121**, 10385 (2004).