

QCLO 法の並列処理 - タンパク質の全電子自動計算のために -

上野哲哉¹ 吉廣 保¹ 住田豊州² 柏木 浩^{1,3,4} 村松伸哉⁴ 西川宜孝⁴ 佐藤文俊¹

東大生産技術研¹ 富士通九州² 九工大³ アドバンスソフト⁴

【はじめに】

タンパク質の機能を理解・解明するため、発表者らは密度汎関数法を用いたタンパク質全電子計算プログラム ProteinDF を開発している^[1]。タンパク質のような巨大で複雑な系の SCF の繰り返し計算においては、計算を非常に良い初期値から始めないと発散してしまうか正しい解に収束しない。この良い初期値を作成する方法として、局在化軌道 (Localized Orbital:LO) を応用した擬局在化軌道 (Quasi-Canonical LO:QCLO) 法を開発し^[2]、これを用いて初期データの作成から全電子計算までのタンパク質半自動計算プログラムを開発した。ProteinDF は分散並列処理環境に対応している^[3]が、QCLO 法を用いた半自動計算は分散並列処理環境に対応していない。発表者は、大規模なタンパク質計算に QCLO 法に対応させるため、並列化を行った。

【QCLO 法】

発表者らは、タンパク質全電子計算の初期値を求める方法として、タンパク質を構成するペプチド鎖を計算単位として、それを次第に大きくしながら計算を進める近似電子密度合成法を開発した。しかし、この方法は補助基底関数の展開に一様性がないためかなり大きな誤差を生み出し、収束を悪化させていた。そこで、これを改善する方法として QCLO 法を開発した^[2]。QCLO は一種の局在化軌道で、この方法によって求められた分子軌道はアミノ酸残基などのフラグメントと呼ばれる単位で局在化しているが、フラグメントのカノニカル分子軌道(MO)に似た軌道である。この軌道をつなぎ合わせて次のペプチド鎖分子の初期値とすることで、誤差を数十分の一に減少させることに成功した^[2]。

【QCLO 法アルゴリズム】

計算対象となるペプチド鎖 (フレーム分子と呼

ぶ)を順次伸張していく。これをステップと呼ぶ。各ステップでの計算は以下のとおり (図1)。

STEP1:

タンパク質をアミノ酸残基ごとに分割し、各アミノ酸の MO を計算。

STEP2:

- (1) STEP1 のアミノ酸残基の近似電子密度を合成し、これを初期値として 3 残基の MO を計算。
- (2) LO を計算 (Population Localization 法^[4])。
- (3) フラグメント分子の QCLO を計算。
- (4) フレーム分子の基底に展開。

STEP3:

- (1) STEP2 で計算した QCLO を結合し、直交化 (Löwdin の直交化^[5])。
- (2) フレーム分子の Kohn-Sham 行列を計算。
- (3) フラグメント分子の QCLO を計算。
- (4) フレーム分子の基底に展開。

STEP4 以降:

STEP3 と同様の計算をフラグメント分子 = タンパク質となるまで繰り返す。

タンパク質の全電子計算の流れは図1のようになる。この計算の流れをあらかじめ指定することで、タンパク質全電子計算が半自動的に行われる。

【QCLO 法の並列化】

QCLO 法の並列化は、計算対象によって、いくつかの異なる方法が考えられる。1つはフレーム分子の計算を同時に行う方法である。各フレーム分子の計算は完全に独立に計算できる。そのため、ProteinDF を複数同時に実行することができれば並列実行が可能となる。この方法を並行化と呼ぶ。次の方法は、各フレーム分子内の計算を並列に行う方法である。これには2つのアプローチがあり、1つは各フラグメント分子を同時に計算する方法 (並列化1) で、もう1つは個々のフラグメント分子の計算を分割して計算する方法 (並列化2) である。

Parallel Control of Quasi-Canonical Localized Orbital Method -For Automatic All-Electron Calculations on Proteins-

¹ Tetsuya Ueno, Tamotsu Yoshihiro, Fumitoshi Sato, Institute of Industrial Science, University of Tokyo

² Toyokuni Sumita, Fujitsu Kyusyu System Engineering

³ Hiroshi Kashiwagi, Kyusyu Institute of Technology

⁴ Shinya Muramatsu, Nobutaka Nishikawa, AdvanceSoft

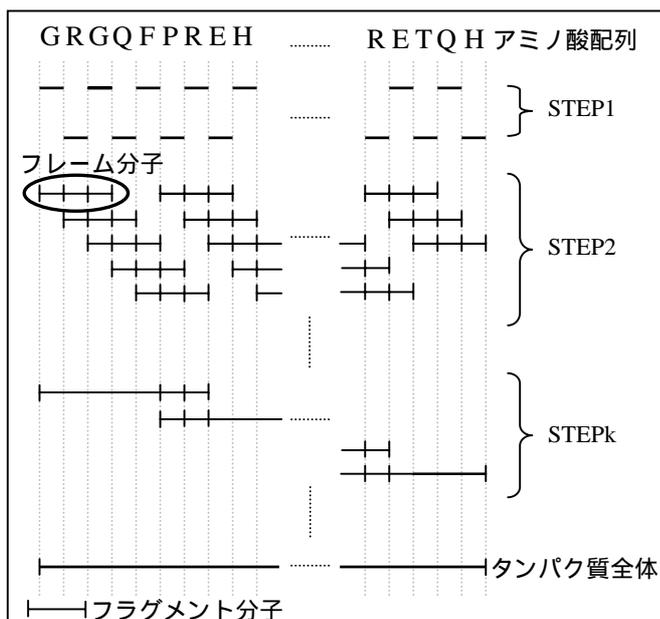


図1 QCLEOを用いたタンパク質全電子計算の流れ

【並行化】

計算対象となるフレーム分子が異なるだけで、各プロセッサでの処理は同じであり、また、処理中に通信も発生しない。計算量がなるべく均等になるようにするため、計算の終了したプロセッサに次の計算を割り振る動的なタスク分配を行った。

【並列化1】

STEP3における(1)、(2)、(4)の計算はフレーム分子全体にわたる計算であり ProteinDF の機能により並列処理可能である^[3]。ここでは(3)のフラグメント分子の QCLEO 計算を複数プロセッサで同時に行うことで並列化を行った。負荷分散は並行化と同様の動的タスク分配にて実現した。

【並列化2】

STEP3の(3)における個々のフラグメント分子の QCLEO 計算内部の並列化を行った。QCLEO の計算は基本的に行列の積と対角化によって実行される。行列積及び対角化の並列化については、ProteinDF において用いられている方法^[3]を応用した。

QCLEO 法の計算では、ステップが後半になるほどフレーム分子やフラグメント分子のサイズが大きくなる。そのため、全ステップを通じて最適となる方法は無く、各ステップにおいて最適な並列計算法を用いる必要がある。STEP1、STEP2 のような多数の小さなフレーム分子の計算の場合には並行計算を利用し、最終付近のステップでの大き

なフラグメント分子の計算には並列化2を用いるといった具合である。そのため、発表者は、並行化及び並列化1、並列化2に加え、並列化1と並列化2の中間にあたる階層並列化も行った。

【階層並列化】

同一フレーム分子内のフラグメント分子のサイズに大きなばらつきがある場合(図1 STEPk 参照)は、並列化1を用いたフラグメント分子の計算の動的タスク分配では計算量を均等化できない。また、並列化2を用いた場合、小さなフラグメント分子を多数のプロセッサで並列に計算してしまうため、通信量との兼ね合いから十分な効率が得られない。そこで、これらの中間を補うものとしてこれらを組み合わせた並列処理を行う。処理形態としては、プロセッサを幾つかのグループに分け、グループ単位で1つのフラグメント分子の計算を並列に実行するものである(図2)。同モデルを階層化マスタ・スレーブモデルと名付けた。各グループ内のスレーブ台数は、フラグメント分子のサイズをもとに計算時間が最短となるよう適化した。

各並列処理の計算結果については当日会場にて発表する。

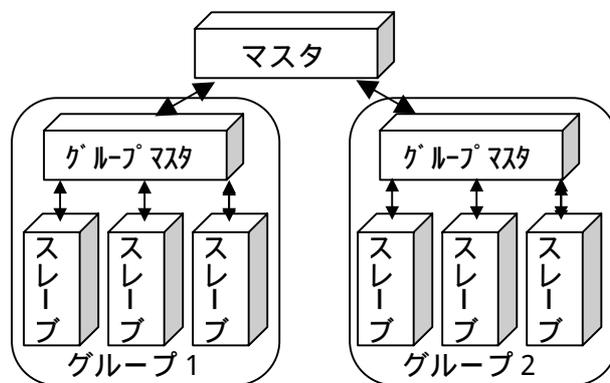


図2 階層化マスタ・スレーブモデル

【謝辞】

本研究は文部科学省リサーチ・レボリューション計画(RR2002)、ITプログラム「戦略的基盤ソフトウェアの開発」の支援の下に行われました。

【参考文献】

- [1] F. Sato, Y. Shigemitsu, S. Yahiro, M. Fukue, S. Kozuru and H. Kashiwagi, *Int. J. Quant. Chem.* **63**, 245-256(1997)
- [2] H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro and F. Sato, *J. Mol. Phys.* in press.
- [3] T. Yoshihiro, F. Sato and H. Kashiwagi, *Chem. Phys. Lett.* **346**, 313-321(2001)
- [4] J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916(1989)
- [5] P. O. Lowdin, *J. Chem. Phys.* **18**, 515-520(1950)