

2P114

タンパク質全電子計算における Schwegler-Challacombe フォック交換項計算加速の検証
(東大生研¹, 東大院工²) ○阿部 敏彦¹, 平野 敏行¹, 石川 健太郎², 佐藤 文俊¹

【はじめに】

密度汎関数法による量子化学計算において、SCF 計算を安全に、高速に収束へと導く技術は、HOMO-LUMO ギャップが狭いタンパク質のような巨大分子では特に重要である。密度汎関数法の交換相関汎関数にはさまざまな汎関数が提案されており、中でも B3LYP^[1]などのハイブリッド汎関数はその計算精度の高さから、現在標準的に用いられている。ハイブリッド汎関数は Fock の交換項を含んでおり、計算には 4 中心 2 電子積分が必要である。その演算量は基底関数の総数を N とすると $O(N^4)$ となるので、巨大分子の計算では計算精度を保ちながら演算量を大量に削減する方法が必須である。Fock の交換項は、その対称性から RI 法のような恒等分解法で計算サイズ依存性を下げることができないため、Fock の交換項の物理的性質に着目した独特の高速計算法が提案されている。

本研究では、演算量を理論上 $O(N)$ まで削減させることができる Schwegler と Challacombe の方法^[2]に着目した。ただしこの方法は計算精度を完全には保証せず、また計算する系に応じて自明でない複数のパラメータを適切に設定する必要がある。本研究では Schwegler-Challacombe の Fock 交換項計算加速法について検証を行った。カットオフ閾値や交換項の距離依存性を近似する定数など、設定が必要なパラメータについて SCF 計算の収束に与える影響を調べ、タンパク質の全電子計算に適したパラメータ設定の検討を行った。

【Schwegler-Challacombe のフォック交換項計算加速法】

Fock の交換項は、密度行列 P と 4 中心 2 電子積分を用いて以下のように表わされる。

$$K_{pr} = \sum_{qs} P_{qs} \langle pq | rs \rangle$$

これは、以下のように書き換えられる。

$$K_{pr} = \iint \rho_{pr}(\mathbf{r}, \mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \rho(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}'$$

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{qs} P_{qs} \rho_{qs}(\mathbf{r}, \mathbf{r}'), \quad \rho_{pr}(\mathbf{r}, \mathbf{r}') = g_p^*(\mathbf{r}) g_r(\mathbf{r}')$$

ここで $g_p(\mathbf{r})$ は基底関数である。Schwegler-Challacombe 法では、これを

$$\tilde{K}_{pr} = \iint \rho_{pr}(\mathbf{r}, \mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|^l} d\mathbf{r} d\mathbf{r}'$$

により近似する。ただし l は系に依存するスケールパラメータであり、通常は 1.0 とする。 \tilde{K}_{pr} が閾値 ϵ_1 よりも小さい場合に K_{pr} の計算を省略(カットオフ)することにより、 $O(N^4)$ であった 4 中心 2 電子積分を、絶縁体などバンドギャップが存在する物質に対しては、理論上 $O(N)$ まで削減可能であるとしている。

【方法】

Lys-Gly-Arg の 3 残基を用いて SCF 計算を行った。計算には ProteinDF を用いた。収束判定には、全エネルギー変化の絶対値が 1.0×10^{-5} a.u. 以下で、かつ電子密度行列要素の変化の絶対値の最大値が 1.0×10^{-4} 以下の条件を用いた。まずパラメータ設定として、 $\epsilon_1 = 1.0 \times 10^{-10}$ および $l = 1.0$ として実験を行ったところ、SCF 計算は収束しなかった。そこで、SCF 計算を収束させるための条件について、

① ϵ_1 を変えた場合 ② l を変えた場合 ③ 計算法の修正について検討した。

本研究で用いた計算法の修正は以下のとおりである。Schwegler-Challacombe の方法では、4 中心 2 電子積分値の大きいものまで捨ててしまうものの、理論によればこれに対応する電子密度行列要素が十分に小さくなり、Fock の交換項の要素には影響を与えないと考えられていた。ところが、上記の系を用いて詳細に検討したところ、SCF 計算を収束させるうえで重要なものまでカットしていることが明らかとなった。

そのようなカットをしている諸条件を検討したところ、重なり行列の値と関連があることがわかった。そこで、Schwegler-Challacombe 法に新たに重なり行列 S の値に基づく条件を加えた修正法を提案した。具体的には、 $\tilde{K}_{pr} < \epsilon_1$ かつ $S_{pq} < \epsilon_S$ かつ $S_{rs} < \epsilon_S$ の場合に、 $\langle pq | rs \rangle$ の計算を省略する。閾値 ϵ_S を定めるため、収束しなかった最初の実験でカットオフされている 4 中心積分に対応する、 S の分布を用いた。なるべくカットオフ数を減らさないように、 S の大部分よりも閾値を大きくするために、 S の平均+標準偏差 $\times 2$ を用い、 $\epsilon_S = 0.05$ とした。

【実験結果および考察】

SCF 計算の最大繰り返し回数を 100 回とし、その中で収束するかどうか、また収束までに要した CPU 時間の計測を行った。使用した CPU は、Intel (R) Pentium (R) D CPU 3GHz である。

1. ϵ_1 を変えた場合： $\epsilon_1 = 1.0 \times 10^{-11}, 1.0 \times 10^{-12}, 1.0 \times 10^{-13}, 1.0 \times 10^{-14}$ と変えたところ、 $\epsilon_1 = 1.0 \times 10^{-14}$ で収束した。
2. l を変えた場合： $l = 1.1, 1.2, 1.5$ と変えたところ、 $l = 1.5$ で収束した。
3. 提案法(計算法の修正)： $\epsilon_1 = 1.0 \times 10^{-10}, l = 1.0$ で収束した。

上記のうち、収束したものはすべて SCF 計算 36 回で計算が収束した。4 中心積分の総数 3859610 のうちカットオフされた数と、SCF 計算 1 回分の時間を、計算時間の小さい順に並べたのが下の表である。比較のため、カットオフなしの場合を基準として、差を右側に記した。また、収束しなかった最初の実験についても、参考として示した。

Schwegler-Challacombe カットオフ条件	カットオフ数	計算時間(秒)	基準からの差(秒)
$\epsilon_1 = 1.0 \times 10^{-10}, l = 1.0$ (収束せず)	2154210	316	- 10
$\epsilon_1 = 1.0 \times 10^{-10}, l = 1.0$ 提案法	1286231	319	- 7
$\epsilon_1 = 1.0 \times 10^{-14}, l = 1.0$	308298	326	0
$\epsilon_1 = 1.0 \times 10^{-10}, l = 1.5$	1972	326	0
Schwegler-Challacombe カットオフなし(基準)	0	326	0

実験結果によれば、 l の値を大きくする、あるいは ϵ_1 を小さくすることで SCF 計算を収束させることができたが、計算時間短縮の効果は小さい。本実験系は 3 残基と小さいものの、タンパク質では、計算時間短縮効果は大変重要である。オリジナル法では 4 中心積分のうち必要なものとそうでないものを適切に選別するためには条件をきつくするしかないが、提案法では重なり積分に基づく閾値の追加により、選別が効果的に行われているものと考えられる。

本研究は「イノベーション基盤シミュレーションソフトウェアの研究開発」により行われた。

参考文献

- [1] A. D. Becke: *J. Chem. Phys.*, **98**-7, 5648-5652 (1993).
- [2] E. Schwegler and M. Challacombe: *J. Chem. Phys.*, **105**-7, 2726-2734 (1996).