

## 2E17

ハイブリッド汎関数を用いたタンパク質全電子密度汎関数計算

(東大・生研<sup>\*</sup>, 日立・機械<sup>\*\*</sup>, 東大・情報基盤<sup>\*\*\*</sup>) ○平野 敏行<sup>\*</sup>、稲葉 亨<sup>\*\*</sup>、佐藤 文俊<sup>\*\*\*</sup>

【序】当グループでは、タンパク質の全電子カノニカル波動関数を計算する密度汎関数法プログラム ProteinDF を開発している[1]。基底関数にはガウス型関数を、分子積分の高速化には RI(Resolution of Identity)法[2]を採用した。この ProteinDF を用いて、局所密度近似(LDA)汎関数 SVWN[3]で、現在世界最大の全電子計算である、306 残基、26,790 軌道のインスリントタンパク質 6 量体の全電子計算に成功した[4]。一方、近年の低分子密度汎関数計算では、非局所密度近似補正(GGA)や Hartree-Fock(HF)法における交換項を組み合わせたハイブリッド交換相関汎関数が良い結果を与えることが検証されてきた。なかでも B3LYP [5]は、数多くの計算結果と実験結果との比較が行われ、高精度の結果を示した多くの計算例が報告されている。低分子の密度汎関数計算において B3LYP は成功を収めているが、大規模タンパク質の全電子計算の例はほとんど無く、評価も十分に行われていない。そこで、B3LYP ハイブリッド汎関数を ProteinDF に実装し、タンパク質電子状態計算における計算量や結果の評価・比較を行った。

【理論・方法】ハイブリッド汎関数を用いたタンパク質全電子計算を困難にしていた主な要因として、①大規模分子における交換相関項の数値計算に加えて、計算コストの高い Fock の交換項の分子積分が必要であること、②大規模分子の分子積分・数値積分と、数万次元の大規模行列演算に耐え得る計算機資源とソフトウェアが必要であること、が考えられる。

①Fock の交換項は、次の 4 中心積分で計算される。

$$\langle pq|rs\rangle \equiv \iint g_p(\mathbf{r})g_q(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}g_r(\mathbf{r}')g_s(\mathbf{r}')d\mathbf{r}d\mathbf{r}'$$

ここで  $g_p$  は基底関数である。この積分は、クーロン項の計算でなじみのカットオフ法や FMM 法以外にも、交換項の性質を利用して効率良く求める独自の手法[6]が提案されているが、本研究では、クーロン項の計算とあわせて分子積分を統一的に制御できる利点に着目し、Fock の交換項の計算にも RI 法を採用した。RI 法では補助基底関数  $g_a^\rho$  を用いて、以下の 3 中心積分

$$\langle pq|\alpha\rangle = \iint g_p(\mathbf{r})g_q(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}g_a^\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}'$$

により、一般に 2 電子積分は次のように求めることができる[7]。

$$\langle pq|rs\rangle = \langle pq|v\rangle V_v^{-1}\langle t|rs\rangle$$

ここで  $V^{-1}$  は補助基底関数ベースの電子反発積分  $\langle\alpha|\beta\rangle$  の逆行列である。Fock の交換項  $\mathbf{K}_{pq}$  は、

$V^{-1/2}$  から作成された  $\langle pq|v\rangle$  を利用して次のように求められる。

$$\mathbf{K}_{pq} = \sum_{vrs} \sum_{\sigma} \langle pr|v\rangle \langle v|qs\rangle \mathbf{C}_{r\sigma} \mathbf{C}_{s\sigma}$$

$$\langle pq|v\rangle = \sum_t V_{vt}^{-1/2} \langle pq|t\rangle$$

この積分は  $O(nN^2N')$  ( $N$  は軌道数、 $n$  は補助基底関数の数、 $N'$  は占有軌道の数) の計算量を持つものの、カットオフ法などを併用すれば、さらに高速化することができる。さらにこの手法の利点は、既に ProteinDF に実装されている、高速な 3 中心積分ルーチンとカットオフテーブルが利用可能となることである。本研究では、この手法によって得られた Fock の交換項を用いて、ハイブリッド交換相関汎関数を ProteinDF に実装した。

②計算律速となる分子積分や交換相関項の数値積分や、数万次元の行列演算を行うために、SPMD (single program multiple data) 方式の並列化を行った。分子積分や数値積分は、独自のアルゴリズムにより高度に並列化されている。数万次元にも及ぶ大規模行列演算には、膨大な容量のメモリ(主記憶装置)とその管理が必要となる。そこで ProteinDF では、巨大な行列をプロセッサ毎に分散保持する、分散メモリモデルを採用した[8]。1 つの巨大な行列は、図 1 のように各プロセッサに分散される。行列積・逆行列・対角化といった行列演算には、並列版線形計算ライブラリ ScaLAPACK を用いた。行列要素を分散保持した結果、ProteinDF の計算可能サイズは、理論上 CPU あたりのメモリ容量からクラスタ全体の総メモリ容量あたりに依存することとなり、PC クラスタから SMP システムまで様々な計算機環境で効率よく動作する密度汎関数プログラムとなった。

本発表では、B3LYP ハイブリッド汎関数の ProteinDF への実装と、これを用いたタンパク質密度汎関数計算の結果について報告する。

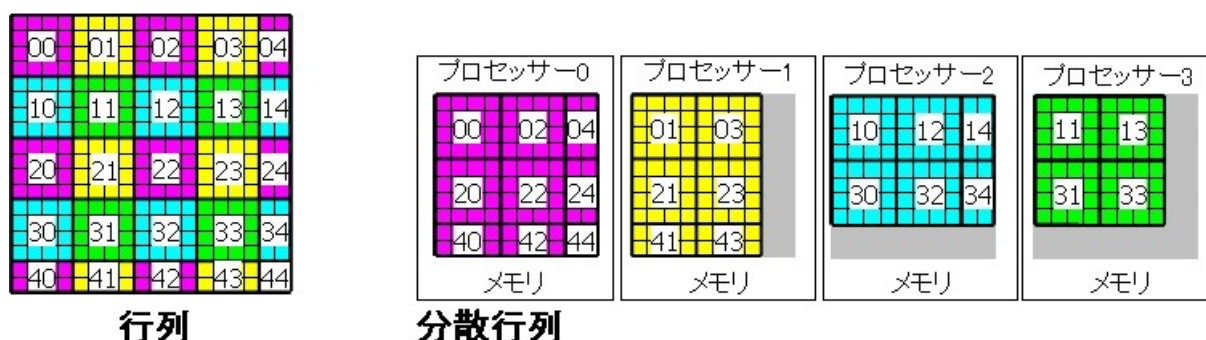


図 1. 分散メモリを用いた行列要素の保持例。ScaLAPACK では左図の密行列を右図のように各 CPU で分散保持する。

本研究は文部科学省次世代 IT 基盤構築のための研究開発プログラム「革新的シミュレーションソフトウェアの研究開発」の支援の下に行われた。計算には、統計数理研究所の統計科学スーパーコンピュータシステムを利用した(2007-ISM-CRP-1012)。

[1] F. Sato, T. Yoshihiro, M. Era and H. Kashiwagi, *Chem. Phys. Lett.*, **341**, 645 (2001).  
 [2] J. Andzelm and E. Wimmer, *J. Chem. Phys.*, **96**, 1280 (1992).  
 [3] S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, **58**, 1200 (1980).  
 [4] T. Inaba, N. Tsunekawa, T. Hirano, T. Yoshihiro, H. Kashiwagi and F. Sato, *Chem. Phys. Lett.*, **434**, 331 (2007).  
 [5] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, **98**, 11623 (1994).  
 [6] E. Schwegler and M. Challacombe, *Theor. Chem. Acc.*, **103**, 344 (2000).  
 [7] O. Vahtras and J. Almlöf, *Chem. Phys. Lett.*, **213**, 514 (1993).  
 [8] T. Inaba, F. Sato, *J. Comput. Chem.*, **28**, 984 (2007).