

1. はじめに

ProteinDF はガウス型基底関数を用いた密度汎関数法に基づくタンパク質の量子化学計算ソフトウェアである[1]。近年、計算機の高速化や利用できるメモリ・ディスク容量が増大したことにより、ProteinDF によってさらに大きなタンパク質の計算が可能となってきた。ところが大規模なタンパク質の計算を行うにつれ、交換相関ポテンシャルを見積るためのグリッド生成にかかる時間が大幅に増えている。現在 ProteinDF では 306 残基(原子数: 4716)のインスリン 6 量体の全電子計算に成功しているが、全計算時間は Itanium2 1.3GHz 32CPU (166Gflops)の環境で 23 日、その内グリッド生成にかかった計算時間は約 5 日であった[2]。これまでは self-consistent field (SCF)計算前に計算されるグリッド生成の計算時間は比較的問題とならなかったが、超大規模なタンパク質の計算や、特に構造最適化や *ab initio* 分子動力学計算では構造が換わるたびにグリッドを新たに作成する必要があるため重要な問題となるだろう。

タンパク質の全電子計算は、タンパク質の基礎過程や機能の解明に重要な役割を示す。これからますます複雑な計算が要求されることが予想される。その際さらに大規模のタンパク質全電子計算を行う必要性や、タンパク質の構造を変化させながらくり返し全電子計算を行う計算が想定されるため、グリッド生成アルゴリズムの高速化は急務であるといえる。そこで本研究は ProteinDF のグリッド生成ルーチンの高速化を行うことを目的とする。

2. 方法

ProteinDF では、グリッド生成にファジーセル法を採用している[3]。ファジーセル法はボロノイ多面体をベースに各原子の積分領域を分割し、この原子間の積分領域に対し 1 から 0 まで徐々に変化するような重み関数を与える方法である。ProteinDF では図 1 のようなアルゴリズムで重み関数を計算している。

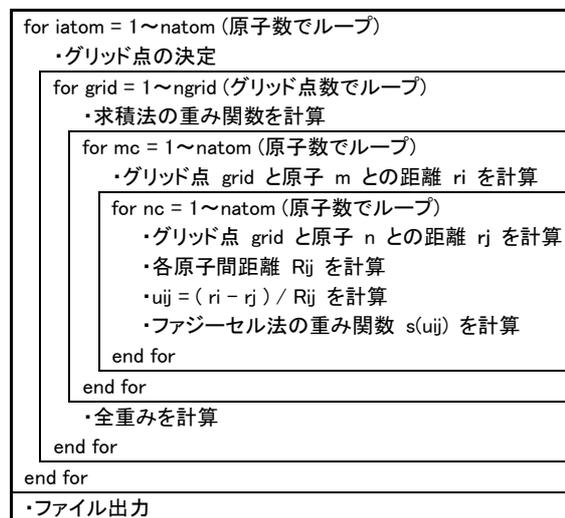


図 1 グリッド生成のアルゴリズム

このアルゴリズムではまず原子数 *natom* 回のループがありその内側に *ngrid* 回、*natom* 回、*natom* 回と、ループが 4 重に存在する。このうち *ngrid* は、原子種に依存するが SG-1 グリッド[4]では高々 5000 程度の定数である。そのため原子数による 3 重のループが、超大規模な計算を考慮する場合に問題となってくる。さて現在のアルゴリズムではメモリを節約する目的もあり、グリッド点 *grid* と原子 *n* 間の距離 *rj* および原子 *m* と *n* 間の距離 *Rij* を一番内側のループで逐次計算している。しかしコンピュータの発展により今ではメモリは以前に比べて多量に利用できるようになっている。

そこでまずこれらの計算をあらかじめ行っておき結果をメモリ上に格納しておくことで、計算回数を減少させ高速化を行う。特に *Rij* をあらかじめ計算することはグリッド点の決定に関係なく行えるため一番外のループ外で実行できる。そのためかなりの高速化が果たせると予想される。

原子間距離をあらかじめ計算するためには、原子数×原子数のハーフ行列分のメモリが確保できればよく、300 残基のタンパク質を計算する場合には約 100 MB、1000 残基では 1.2 GB 程度のメモリが必要である。またグリッド点と原子間距離に関してはそれぞれ 10 MB、30 MB 程度である。

*saisei@fsis.iis.u-tokyo.ac.jp

さらに、巨大な分子を計算する場合、互いが遠方にありファジーセル空間分割に影響を与えない原子の組が存在することが予想される。現状ではこのような原子の組に対しても重み関数を逐次計算するアルゴリズムとなっており無駄が多い。

そこであらかじめ計算不要な原子の組を調べておき最も内側のループの回数を減らすことで、効率的にグリッド生成を行うといった改良を図る。これによって 3 乗の時間依存性が除かれる可能性があり、特に巨大な分子を扱う場合に有効となることが期待できる。

本研究では以上の 2 手法、

①現在のアルゴリズムでメモリを潤沢に使う

②ファジーセル法を効率よく行う

による改良を行い、その効果をみるためのテスト計算として、3 残基のアミノ酸(原子数: 38)、7 残基のアミノ酸(原子数: 101)、14 残基のアミノ酸(原子数: 229)に対するグリッド生成の計算時間を測定し結果を比較した。計算には、ALPHA21264 833MHz(理論ピーク性能 2Gflops)を用いた。

3. 結果

①による高速化の結果を表 1 に示す。①の改良により原子数 38 の 3 残基アミノ酸での計算時間が 39s から 10s、原子数 101 の 7 残基アミノ酸での計算時間が 645s から 185s、原子数 229 の 14 残基アミノ酸での計算時間が 6711s から 1495s となった。

また図 2 では①による高速化前後の、3 残基、7 残基、14 残基アミノ酸でのグリッド生成に要した計算時間を両対数プロットした。予測通り系のサイズの依存性は変化せず(傾きは変化せず)、一様に 4 倍高速となっている。

以上の結果より、①の改良を行うだけで、166G flops の計算機を用いて現時点で約 5 日かかっているインスリン 6 量体のグリッド計算に要する日数が 1 日程度となることが予想される。この改良だけで全体の 22%を占めていたグリッド計算が 5%に圧縮されるため、非常に効果的である。

発表ではさらに②の改良を行った結果を合わせて報告する。

表 1 グリッド生成に要した計算時間
(elapsed time (s))

原子数	従来の ProteinDF	①による高速化後
38	39	10
101	645	185
229	6711	1495

ALPHA21264 833MHz (2Gflops)

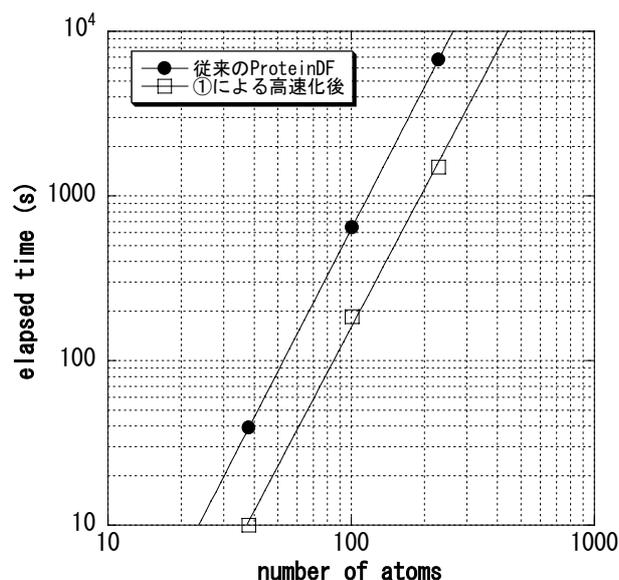


図 2 グリッド生成に要した計算時間の比較

4. 謝辞

本研究は文部科学省 IT プログラム「戦略的基盤ソフトウェアの開発」の支援の下におこなわれました。

5. 参考文献

- [1] F.Sato, Y.Shigemitsu, I.Okazaki, S.Yahiro, M.Fukue, S.Kozuru, H.Kashiwagi, Development of a new density functional program for all-electron calculation of proteins. *Int. J. Quant. Chem.* 63(1997)245-256.
- [2] T.Inaba, F.Sato, P16: All-electron calculation on insulin hexamer by the density functional method. SC2004 Poster-2; 稲葉, 佐藤, 柏木, 密度汎関数法によるインスリン 6 量体の全電子計算. 分子構造総合討論会 2004, 4P126.
- [3] A.D.Becke, A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.*, 88(1988)2547-2553.
- [4] P.M.Gill, B.G.Johnson, J.A.Pople, A standard grid for density functional calculations. *Chem. Phys. Lett.*, 209(1993)506-512.