

【はじめに】

本研究グループはタンパク質の性質を電子レベルで理解することを目標に、タンパク質のための密度汎関数法プログラム ProteinDF¹⁾、ならびに、その全電子計算の収束過程を強力にサポートする QCLO (Quasi-Canonical Localized Orbital) 法²⁾の開発を行なっている。QCLO は局在化軌道 (LO) の一種であり、例えばアミノ酸残基といった任意の範囲に局在化しているが、その中ではカノニカル軌道の性質を持つ軌道である。この QCLO をつなげて、タンパク質の高品位な初期値 MO を作成することで、タンパク質の全電子計算をほぼ確実にこなえ、同時に SCF 計算の繰り返し回数を低減できる。この方法によるインスリン単量体の計算は昨年度の本学会で発表した³⁾。ProteinDF を用いて、図1に示すようなインスリン6量体(306 残基、原子数 4716、電子数 18552、基底関数 26766、補助基底関数 48600)の全電子状態計算を行なった。使用した計算機は SGI 社の Altix3700 であり、64 個の Itanium2 プロセッサが NUMalink™(通信速度が 3.2 GB/s)で結合された 256 GB の共有メモリシステムである。本システムのピーク性能は 256 GFLOPS である。この計算は世界最大規模の分子軌道計算である。本発表ではこの計算について報告を行う。

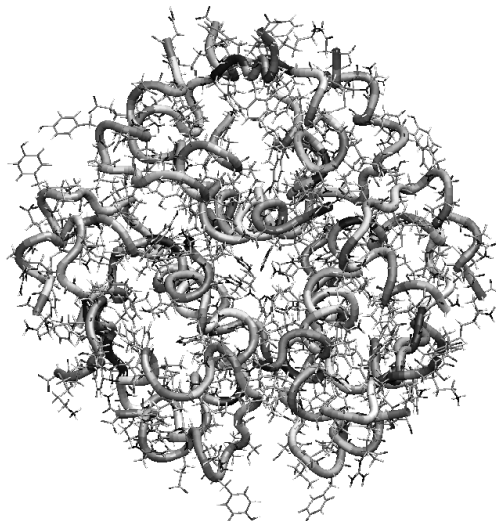


図1 インスリン6量体

【インスリン6量体の計算方法】

インスリン6量体の立体構造データは PDB⁴⁾の 1A10.pdb を用いた。AMBER7⁵⁾を用いて水分子中で 20ps の分子動力学計算を行い安定構造を得た。続いて、2つの Zn^{2+} と全ての水分子を取り除き、アニオン、カチオンなどの荷電アミノ酸近傍に +1、-1 のポイントチャージを置いて系を中性化して、インスリン6量体の計算構造とした。

インスリン6量体の全電子計算は以下に示す5つの計算ステップで達成した。まず、第1ステップでは、アミノ酸1残基ごとの全電子計算を行った。第2ステップでは、第1ステップにて得られたアミノ酸残基の近似電子密度を合成して初期値とし、残基番号 1~3、2~4、3~5 といった重なりを持つ3残基の全電子計算を行った。第3ステップでは、2残基ずつの重なりを持つような8~9残基ペプチドの全電子計算を行った。第4ステップはインスリン単量体の計算で、第3ステップの QCLO を合成して初期値とした。このとき、他の5つのインスリンからの静電効果を考慮するため、第3ステップの結果からマリケン電荷を計算し、それらをポイントチャージとして扱うことにより計算を行っている。最後の第5ステップのインスリン6量体の計算では各インスリン単量体同士は共有結合を持っていないため、第4ステップより得られた近似電子密度を合成して初期値とし、306残基のインスリン6量体の全電子計算を行った。

【ProteinDF の並列化】

26776 軌道の LCAO 行列を倍精度実数型で保存しようとする、約 6GB という大量のメモリ領域が必要となる。そのため、行列要素を各 CPU に分散させて保存させるとともに、様々な演算を並列化させる必要がある。そこで、ProteinDF を並列通信ライブラリ MPI (Message Passing Interface) を用いて並列化した。対角化、行列積、逆行列などの行列演算には、並列版行列演算ライブラリである ScaLAPACK⁶⁾ を用いた。全計算時間の中で大きな割合を占める分子積分は、通信量を極力抑えるために、基底関数の組み合わせをグループ化することで並列化した。それゆえ、ProteinDF では行列要素はすべて分散して保持しており、主要な演算はすべて並列化されているため、タンパク質のような基底関数が膨大となる高分子に対しても分子軌道計算が可能である。

【計算結果】

最終的なインスリン 6 量体全体の全電子計算は 35 回の SCF 計算を要し、SCF 計算 1 回にかかった実時間は 1.3GHz の Itanium2 プロセッサ 32 台を用いて、平均で約 12 時間であった。

図 2 はインスリン 6 量体の全電子計算に対して、4 CPU を基準にした時の ProteinDF の台数効果を示したものである。1 CPU では、計算時間が多くかかることから、4CPU での計算時間を基準とした。図中では分子積分、行列積、対角化、交換相関についても同時に示してある。32CPU にて、70%の並列化効率を得られた。

図 3 はインスリン単量体、インスリン 6 量体などの様々なサイズの分子に対して、基底関数の数と繰り返し計算 1 回の実行時間をプロットしたものである。ProteinDF では、分子積分、行列積、対角化、交換相関の基底関数に対する依存性は、それぞれ 2.2、2.4、2.6、1.9 乗の依存性があり、全体では 2.2 乗の依存性があることがわかった。また、インスリン 6 量体の計算時間の内訳は分子積分、行列積、対角化、交換相関、その他で 53%、24%、12%、6%、5%であった。詳細は当日発表する。

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」において実施された。

参考文献：

- 1) Sato, F. et al.: Calculation of all-electron wavefunction of hemoprotein cytochrome c by density functional theory, Chem. Phys., Lett. 341, pp645-651, 2001
- 2) Kashiwagi, H. et al.: Convergence process with quasi-canonical localized orbital in all-electron SCF calculation on proteins, Molecular Physics, 101, pp81-86, 2003
- 3) 稲葉亨 他: ジスルフィド結合ネットワークタンパク質ヒトインスリンの QCL0 法による全電子計算、分子構造総合討論会 2003 講演論文集 1Pa012
- 4) Protein Data Bank: <http://www.rcsb.org/pdb/>
- 5) AMBER7: <http://www.amber.ucsf.edu/amber/amber.html>
- 6) ScaLAPACK: L. S. Blackford, et. al., ScaLAPACK Users' Guide 1997

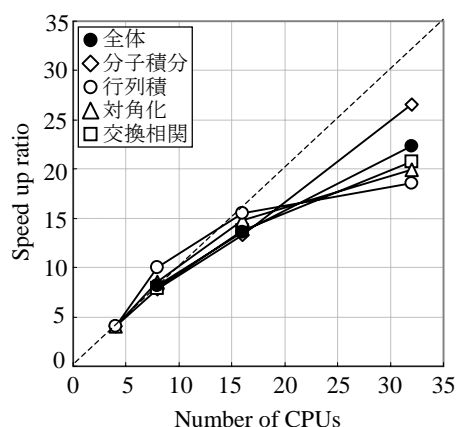


図 2 ProteinDF の台数効果

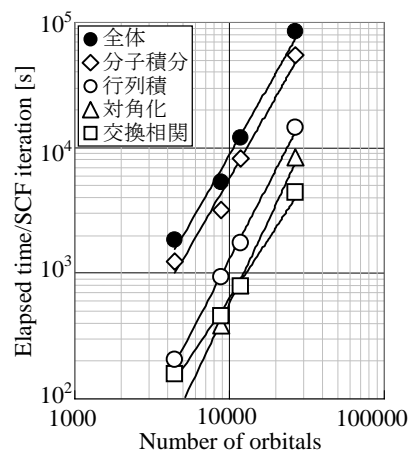


図 3 基底関数の数と計算時間