



分類	量子化学計算
キーワード	タンパク質、量子化学計算、密度汎関数法
開発者	佐藤文俊、平野敏行、阿部敏彦、下向智美、谷村景貴、恒川直樹、松田潤一
作成年月日	2013年3月
コード名	ProteinDF
使用言語	C++

◇ ProteinDF System 2012 概要

当グループは、密度汎関数法でタンパク質分子の全電子カノニカル波動関数を計算する ProteinDF システムを開発している[1、2]。計算法の汎用性から、タンパク質に限らず様々な巨大分子の電子状態を解析できるため、ナノスケールものづくり設計に貢献できる。本公開ソフトウェアは、2012年度最終版である。本ソフトウェアはダウンロードして無料で利用することができる[3]。

◇ 注目すべき新機能・開発項目

分散メモリ型並列計算機に対応した超並列密度汎関数計算エンジンの改良：1000 残基クラスタンパク質のカノニカル分子軌道計算の場合、1つあたりの行列のサイズは約 80GB にも及ぶ。行列積を計算するだけでもその3つ分、すなわち 240GB のメモリを必要とする。ProteinDF では大規模分子のカノニカル分子軌道計算を限られたメモリ上で遂行するために、必要最低限の大規模行列のみをメモリ上に保持し、それ以外の行列はディスク領域に退避させる方法を採用している。ディスクを用いる方法は非効率的であるが、計算ノードの搭載メモリ量が限られている現在において、大規模分子の量子化学計算を行う次善の策といえる。スーパーコンピュータ「京」をはじめとする現在主流のスーパーコンピュータでは、分散メモリ型並列計算機を採用している。このとき、計算ノードより大きな行列を扱う場合は分散保持するしかない。個別の行列要素へのアクセスが必要な分子積分計算では、ノード間通信速度がボトルネックになり非効率である。そこで、行列要素の参照が必要な分子積分計算を極力排除し、チューニングされている行列積により高速計算を達成するコレスキー分解法を ProteinDF に実装した。本手法では4中心積分 ($pq|rs$) を要素とする行列 V をコレスキーベクトル L とその共役転置行列の積であらわす。

$$V_{pq,rs} = (pq | rs) \approx \sum_{K=1}^M L_{K,pq} L_{K,rs}$$

この L を用いることによって、Kohn-Sham 行列におけるクーロン項 J ならびに Fock 交換項 K を線形演算によって求めることができる。

$$J_{pq} = \sum_{rs} (pq | rs) P_{rs} \approx \sum_{rs} \sum_I L_{I,pq} L_{I,rs} P_{rs}$$

$$K_{pq} = \frac{1}{2} \sum_{rs} (pr | qs) P_{rs} \approx \sum_i \sum_I X_{I,pi} X_{I,qi}$$

$$(X_{I,pi} = \sum_r L_{I,pr} Q_{ri}, P_{pq} = \sum_J Q_{p,J} Q_{q,J})$$

本手法により、カノニカル分子軌道計算を不得手とする分散メモリ型計算機においても効率良く大規模分子の電子状態を求めることが可能になった。

bridge 共通データ・アルゴリズムクラスの開発：様々なタイプのシミュレーションを円滑に実行できるように、ProteinDF をはじめとする各プログラムで利用するデータ構造・アルゴリズムをコンポーネント化し、データ構造・アルゴリズムの共通化を行った。データ構造はエンディアンの異なる計算機にも対応する MessagePack を元に設計し、アルゴリズムは C++ のほか、グルー言語として知られる Python で開発した。これにより、2 次構造に基づいて QCLO 計算シナリオを自動的に作成する、ProteinDF 計算で得られた Mulliken 電荷に応じて原子を色付けするといったプレ・ポスト処理を Python スクリプトで記述することができるようになる。本クラスによって、ファイル形式や計算機の違いを気にすることなく、分子のモデリングやポスト処理に集中できる環境が整った。オープンソース・ソフトウェアとして公開し、将来にわたり機能追加・改良していく予定である。

部分構造最適化機能：エネルギー勾配から構造最適化計算を行う機能を実装した。分子の性質や反応を解析するためには、ポテンシャルエネルギー曲面の情報が不可欠である。系の自由度が極めて多いタンパク質全ての構造においてこれを求めることは事実上不可能であるが、活性中心回りの構造において得られれば十分実用的である。すでに最急降下法、共役勾配法、BFGS 法を実装していたが、これらに加えて、収束性を大幅に改善する DIIS (direct inversion in the iterative subspace) 法を実装した。DIIS 法は残差ベクトルを最小化する方法で、商用ソフトウェアにも実装されている信頼性の高いアルゴリズムである。energy-DIIS (EDIIS) 法、augmented Roothaan-Hall energy-DIIS (ADIIS) 法、linear-expansion shooting technique (LIST)、遷移状態を探索する機能も実装中であり、随時公開する予定である。これらの収束加速法以外にも、大規模構造最適化計算を効果的に実行するための工夫として、タンパク質の骨格だけを固定する機能、 α 炭素を固定する機能なども実装した。

NMR 計算機能：核磁気共鳴 (NMR) 分光は、分子構造や電子配置などの詳細情報が得られるため、生体分子のみならず、ナノ分子の観測手段として用いられる重要な方法の一つである。そのため理論計算による結果解析や予測の必要性が高まっている。タンパク質の分子軌道計算により得られた波動関数からの化学シフト計算を実現することを目的として、スクラッチからプログラムの Hartree-Fock (HF) 法レベルで開発を行った。ゲージ依存性の問題を解決する方法として、Gauge Invariant Atomic Orbitals (GIAO) 法を採用した。GIAO 法では原子ごとにベクトルポテンシャルを定めるゲージの原点

を取ることで、基底関数のゲージ依存性を取り除くため、高精度の計算が可能である。磁気遮蔽定数の反磁性項および常磁性項の計算法を開発し、常磁性項の計算で必要となる CPHF 方程式の安定解法を新たに考案した[4]。得られた反磁性項と常磁性項から、最終的に核磁気遮蔽テンソルが求められる。分子積分の大規模化と高速化を行い、大規模分子への適用を可能とした。

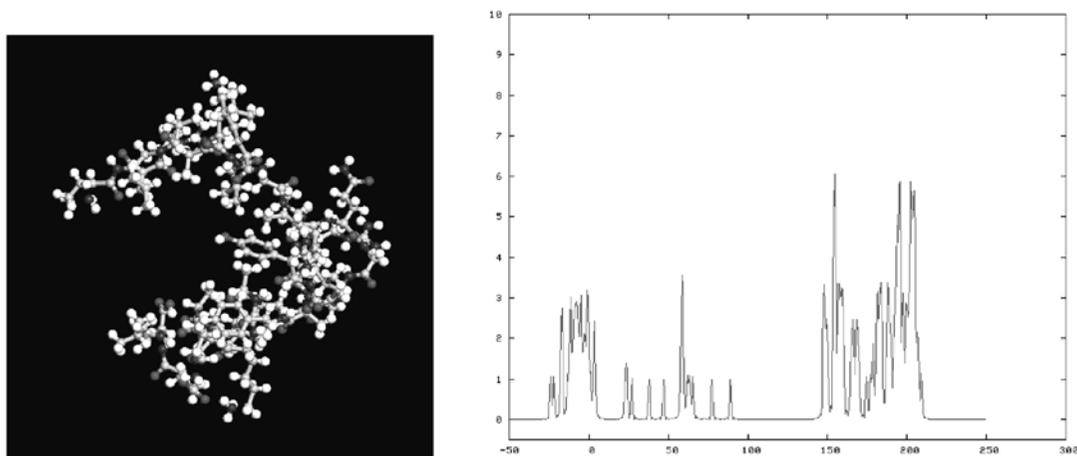


Fig 1 : Molecular structure of 28-residue peptide (left) and calculated ^{13}C FT-NMR spectrum (right)

統括 GUI : HPCI の構築を迎え、タンパク質の網羅的な分子軌道計算が現実的なものとなった。大規模計算データ群から情報を得るには、可視化などの工夫が必要である。統括 GUI、ProteinEditor では、タンパク質の電子構造解析法において新たな解析手法の提案を行い、それらを実現する可視化ツールを開発した[5]。分子軌道と軌道エネルギーから、実験で得られる様々な物理量を量的に見積もることができるのが分子軌道法の魅力であるが、なぜ物理量がそのような値に見積もられたのかを解析するためには、分子軌道一つ一つに立ち戻らなければならない。ここで、必ずしも HOMO や LUMO だけが重要であるとは限らない。解析には、総合的な判断が必要となる。タンパク質ともなればその解析は大変困難な作業となる。また逆に、タンパク質をどのように改変したら、望みの物理量が得られるようになるのかという設計指針を得るのも至難の業である。そこで、分子軌道の空間分布と分子軌道エネルギーの分布との相関を、タンパク質の構造と関連付けて解析する方法を提案した。具体的には、特定のアミノ酸残基や α ヘリックスといった着目する構造モチーフを決め、それが各分子軌道に占める割合(電子数相当)が閾値以上となる分子軌道を網羅的にピックアップし、局所状態密度(LDOS)として表示・解析するという方法である。本方法の有効性を調べたところ、興味深い相関が存在することが明らかとなった。タンパク質の電子構造原理の理解や新規タンパク質の分子設計に役立つだろう。

◇参考文献

- [1] 佐藤文俊、恒川直樹、吉廣 保、平野敏行、井原直樹、柏木 浩：“タンパク質密度汎関数法”、柏木浩監修、森北出版(2008)
- [2] 上村(西野)典子、佐藤文俊、恒川直樹、西村康幸、平野敏行、吉廣 保、甘利真司、加藤昭史、

小林将人、田中成典、中野達也、福澤 薫、望月祐志：“プログラムで実践する 生体分子量子化学計算 –ProteinDF/ABINIT-MP の基礎と応用”、佐藤文俊、中野達也、望月祐志編、森北出版 (2008)

[3] <http://www.ciss.iis.u-tokyo.ac.jp/>

[4] T. Abe, Y. Sekine, F. Sato : “Solving a coupled perturbed equation by the residual cutting method”, Chem. Phys. Lett., 557, 176 (2013)

[5] 下向智美、千葉貢治、松田潤一、平野敏行、佐藤文俊：“タンパク質の全電子波動関数と状態密度解析”、可視化情報学会、32、127 (2012)

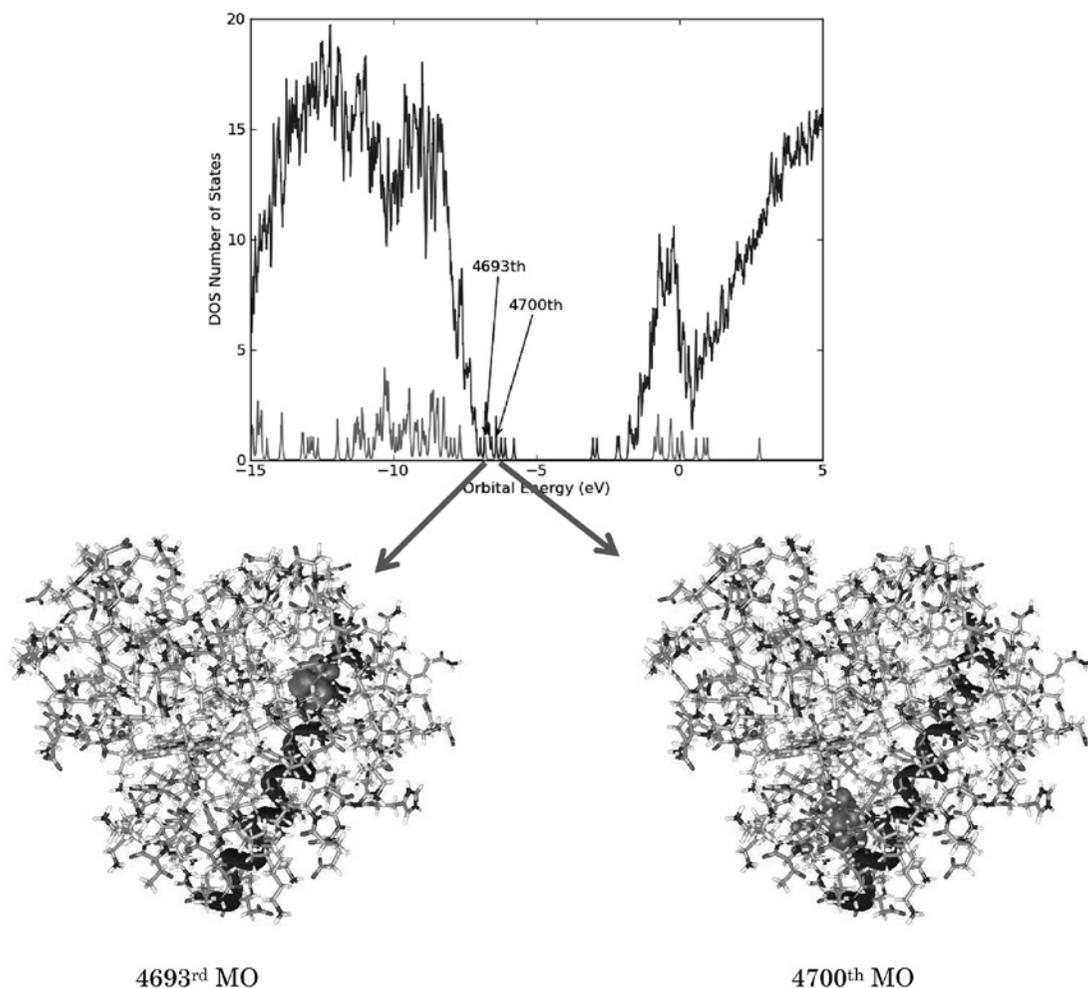


Fig 2 : Analysis of B3LYP all-electron MO calculation on carbonmonoxy-myoglobin

(執筆責任者：佐藤文俊)