

論文の内容の要旨

論文題目 密度汎関数法を用いた並列カノニカル分子軌道法による
タンパク質電子状態解析に関する研究

氏 名 稲葉 亨

タンパク質は生体内のあらゆる場所に存在し、生命活動を支える最も重要な物質である。タンパク質の機能は、酵素、生体構造、エネルギー源の貯蔵・輸送、情報伝達、免疫などであり、実に多種多様である。工学的にもこれらの機能を直接応用したデバイスなどが考案されつつあり、多くの注目を集めている。このように複雑な機能、性質を持つタンパク質を電子レベルで解明することは非常に重要である。タンパク質は、化学的に見ると、炭素、酸素、窒素、水素、硫黄といった原子からなるアミノ酸が連結した数百〜数十万の原子を持つ巨大分子である。

カノニカル電子状態計算は分子規模に対して原理的には4乗の依存性がある。大規模分子の場合、非常に多くの計算時間が必要となり、効果的なアルゴリズムの開発や並列計算機を利用した並列処理が必須の技術となる。また、各分子軌道の係数を表す **Linear Combination of Atomic Orbital (LCAO)** 行列など各種行列要素を保存するために多くのメモリ容量も必要となり、メモリ管理も重要となる。一方、大規模分子の電子状態計算の **Self-Consistent Field (SCF)** イタレーションを無事に収束させるためには、初期値が非常に重要となる。初期値の精度が悪いと **SCF** イタレーションが収束せず、振動もしくは発散してしまう。大規模分子の電子状態計算では、初期値に対して特に注意を払う必要がある。

このように、タンパク質の電子状態解析は非常に困難となることから、タンパク質レベルの分子規模の電子状態解析を行うことができるグループは世界的にも数グループに限られている。そこで、本研究では、タンパク質などの大規模分子の電子状態を実用的な時間と計算機で解析可能とすることを目的に、高効率の並列計算を行うことができる密度汎関数法をベースとした電子状態解析手法を開発した。さらに、カノニカル全電子計算では世界最大となる300残基レベル(5000原子レベル)のタンパク質の電子状態を解析した。

本解析手法では、タンパク質などの大規模分子の電子状態解析で有利となる **Resolution of Identity** 法に基づく密度汎関数法を採用している。分子積分に関しては、

イタレーションのたびに再計算を行うダイレクトSCF法、大規模分子において分子積分の数を大幅に減らすことができるカットオフ法、さらに、収束に近づくにつれて分子積分の数をさらに減らすことができるアップデート法を使用している。これらの手法を適用することにより、本解析手法では、Hartree-Fock法とほぼ同じ計算量で相関相互作用の効果を取り入れたタンパク質の電子状態解析を行うことができる。

大規模分子の電子状態計算において行列サイズが大きくなる課題に対して、本研究ではすべての行列を並列計算機の各メモリに分散させて保持させた。これにより、本解析の実行限界は、並列計算機の各メモリ容量に制限されず、総メモリ容量に制限されることになる。

演算量の増大に伴い、並列解析が必須になるが、本解析において計算負荷の高い、分子積分、交換相関項の数値積分、行列演算はすべて並列化を行った。行列演算に関しては汎用の行列演算ライブラリであるScaLAPACKを用いて並列化を行った。分子積分、交換相関項の数値積分に対しては独自のアルゴリズムにて並列化を行った。分子積分については大規模行列を扱いながら、並列化を行う必要があるが、一時的な部分行列を用いた分子積分法を新たに開発した。分子積分、交換相関項の数値積分はネットワークスピードの遅い並列計算機(例えば、ネットワークスピードがGbit/s)でも十分な並列化効率を得られる。一方、行列演算に関しては、ネットワークスピードの遅い並列計算機では並列化効率が下がり、全体の並列化効率を大きく低下させる。

計算時間に大きく依存するSCFイタレーションの回数を低減することを目的に、従来のSimple mixing法に加えて、新たな収束加速方法として、DIIS法、Anderson法を追加した。これらの手法を用いて、インスリン単量体、2量体、インターロイキンの全電子計算を実施した結果、DIIS法、Anderson法の収束回数は、Simple mixing法のイタレーション回数と比較すると、平均で57%、44%程度のイタレーション回数で収束することがわかり、大幅に電子状態解析の計算時間を短縮することができる。

一方、タンパク質の電子状態計算の収束性を向上させるためには、精度の高い初期値を作成することが重要である。本研究では、タンパク質の電子状態解析を一度で解析せずに、小さな分子に分割し、徐々に分子サイズを大きくしながら、最終的にはタンパク質全体の解析を行う手順としている。各分子の全電子計算の初期値として、1つ前のステップのQuasi-Canonical Localized Orbital (QCL0)という局在化軌道を結合した波動関数は精度の高い初期値となる。QCL0はある領域に局在化しているが、その領域内ではカノニカル軌道のような性質をもつ局在化軌道である。QCL0を結合した波動関数を初期値として、複雑なトポロジー構造を持っているインスリン単量体(51残基、基底関数の数4439、補助基底関数の数8060)に対して全電子計算を行ったところ、ほぼ試行錯誤なしで達成することができた。この時の初期値と収束値の差は全エネルギーが0.006 a. u.、マリケン電荷が0.08 a. u. となり、QCL0による初期値作成方法が大変有効であることを示した。

これらの解析手法を用いて、306残基のタンパク質であるインスリン6量体の電子状態の計算（残基数、原子数、電子数、基底関数の数、補助基底関数の数は、それぞれ、306、4728、18552、26790、48684）を行った。初期値は、各インスリン単量体の波動関数(LCAO行列)を組み合わせたものとした。収束加速法にはAnderson法を用いて、17回のイタレーションで収束させることができた。初期値作成には24台の3.2GHz Xeonクラスター(理論ピーク性能：153.6Gflops, ネットワークスピード：1Gbit/s)を使用し、計算時間は、26時間42分であった。インスリン6量体の計算は64台の1.3GHz Itanium 2クラスター(理論ピーク性能：332.8Gflops, ネットワークスピード：3.2GB/s)を使用し、計算時間は64時間58分であった。並列計算機の理論ピーク性能を考慮入れると、初期値作成時間は全体の計算時間のわずか16%であり、大きな割合ではないことがわかる。インスリン6量体の電子状態計算にて並列化効率は82.5%, 並列化率は99.6%と、高い並列化性能を得ることができた。本研究による解析手法により、300残基レベルのタンパク質の全電子計算が現実的な並列計算機を用いて、かつ実用的な時間で計算することができることを実証できた。

インスリン6量体の電子状態計算から得られた分子表面の静電ポテンシャルと分子力場の固定電荷から得られた分子表面の静電ポテンシャルを比較した結果、分子表面では大きな違いがないが、分子内部、特に単量体どうしの界面で大きな違いが生じることがわかった。タンパク質-タンパク質間の相互作用などを計算するときは電子移動を考慮した電子状態計算が必要であることを示唆している。

本解析によりインスリン6量体の解離による電荷変動が大きいアミノ酸を特定した。6量体から2量体の解離では6量体の中心付近のアミノ酸、2量体から単量体の解離では2量体界面のアミノ酸が、解離により電荷の変化が大きいことが示された。

PCクラスターなどの安価で高性能な計算機システムが増えている現状を考えれば、上記解析結果が示すように、タンパク質の電子状態計算はますます短時間でできるようになると予想される。

本研究で開発した解析手法は様々なタンパク質の電子状態を計算するのに有効なツールであり、バイオ分野におけるタンパク質-タンパク質間相互作用、タンパク質-基質間相互作用を評価することができる理論的医薬品開発ツールの一つとなる。また、密度汎関数法をベースとした分子軌道法プログラムであるため、ナノ材料分野における様々な材料開発のツールともなり、応用範囲、適用範囲は非常に多岐にわたると考えられる。