

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 29 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2013～2014

課題番号：25730174

研究課題名(和文) タンパク質の自動量子化学計算法と電子状態DBの開発

研究課題名(英文) Study of Auto Quantum Chemical Calculation Method and Electron Structure Database for Proteins

研究代表者

平野 敏行 (Hirano, Toshiyuki)

東京大学・生産技術研究所・助教

研究者番号：60451887

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：タンパク質の電子状態計算結果は、タンパク質の物性・反応性を理解する上で有益な情報を与える。タンパク質のような巨大分子の電子状態計算において、分子モデリング・電子状態計算ともに工程が複雑で達成が難しく、試行錯誤が必要であった。本研究では、分子モデリングに必要な処理を自動的に行うプログラムを作成し、また電子状態計算処理をほぼ自動的に効率良く行う計算方法を確立した。作成したプログラムはインターネット上に公開した。

研究成果の概要(英文)：The electronic structure of proteins gives useful information in understanding its physical and reactive properties. In the electronic structure calculation of large molecules such as proteins, the modeling and single point calculation processes are complicated, so it is difficult to achieve. In this study, the programs that approximately automatically perform molecular modeling and the calculation of electronic structure have been developed. The developed programs have shown on the internet.

研究分野：量子化学

キーワード：量子化学 タンパク質 電子状態計算

1. 研究開始当初の背景

量子化学計算は、シュレーディンガー方程式に基づき物質の電子状態を明らかにするため、物質の物性・反応性を理解する上で極めて強力なツールである。申請者は、タンパク質をターゲットとした量子化学計算プログラム ProteinDF を開発している [T.Hirano, et. al., JSST, 2009]。ProteinDF はタンパク質をまるごと扱った、カノニカル分子軌道計算を行えることが特長であり、反応に関与するフロンティア軌道(分子軌道)、エネルギー準位、静電ポテンシャルなど様々な物理量が求められる。

一般に、タンパク質の機能解析や機能強化設計といった研究において、変異体やファミリー、フォールドなど、複数の関連したタンパク質の間で、様々な情報を比較検討することが常套手段である。本研究では、タンパク質の電子状態も生物情報の一つとして位置づけ、PDB と同様に蓄積・共有化し、バイオインフォマティクス技術と融合することで、新たな知見が得られると考えられる。そのためには、PDB に登録されている構造の量子化学計算を行い、結果をデータベース(DB)に登録する仕組みが求められる。これらが自動的に実行されるか、少なくともタンパク質構造の登録者・作成者自らが簡単に実行できる環境を整備する必要がある。計算コストの高いタンパク質の量子化学計算は短時間で計算出来るようになりつつあるものの、計算に必要な知識・技術は極めて専門性が高く、達成は困難である。

本研究では、タンパク質のモデリング・分子軌道計算など量子化学計算に関連するプロセスの自動化に関する基礎研究を行い、タンパク質の電子状態とバイオインフォマティクスの融合や応用に展開するための基礎研究を行う。近い将来、タンパク質の量子化学計算とバイオインフォマティクスが融合し、このことがタンパク質の機能の解明と応用に繋がれると期待している。本研究は、その基盤となる技術の開発を目指す。

2. 研究の目的

タンパク質量子化学計算の自動化に関する基礎研究を行い、タンパク質の電子状態とバイオインフォマティクスの融合や応用に展開するための基盤研究を行う。以下の項目を明らかにすることを目的とする。

(1) 量子化学計算の前処理として必要な、タンパク質モデリングの自動化に関する研究を行う。PDB などから取得したタンパク質構造から、欠損部位、水素原子を付加し、構造緩和を行い、異常接近原子対や二面角のチェックをする、一連のモデリング作業を自動化する。

(2) トラブルのない安定したタンパク質量子化学計算の自動解法に関する研究を行う。タンパク質の量子化学計算の実施にあたり、SCF 繰り返し計算に必要な初期値を用意する必要がある。良い初期値を求める手法の1つとして、擬カノニカル局在化軌道(QCLO)法 [T.Hirano, et. al., *J. Chem. Phys.*, 127, 184106] を採用し、効率的に初期値を算出する。

3. 研究の方法

(1) タンパク質モデリングの自動化に関する研究

PDB に登録されているタンパク質構造データの多くは、X線結晶構造解析とNMR解析によるものである。水素原子の付加や欠損部位の処理、結晶水やヘテロ原子に対する処置、溶媒の取り扱いなど、量子化学計算に至るまでに必要な前処理・モデリングは多段階に及んでいる。厳密なモデリングが必要な理由は、1原子たりとも矛盾のある分子構造を許さない量子化学計算の要請によるものである。

タンパク質は原子数が多い。異常接近原子対や構造の歪みに対するチェックを、すべて人手で行うには限界がある。本研究では、PDB 構造データから量子化学計算に至るまでのモデリング手順をプログラミングし、インタラクティブ性の無い、自動バッチ処理が可能なモデリング技術の確立を目指す。

(2) タンパク質量子化学計算の自動化・効率化に関する研究

大量のタンパク質量子化学計算をスムーズに行うために、構造に問題が無ければほぼ100%失敗せずに安定して解けるタンパク質量子化学計算の自動化を行う。

安定して量子化学計算の収束解が得られるためには、適切な初期電子密度が重要である。低分子の計算で用いられる Hückel 法や Harris の汎関数をそのままタンパク質量子化学計算に適用した場合、初期電子密度と収束解との差が大きく、用いることが出来ないことを確認している。本研究ではタンパク質量子化学計算の初期値の作成に QCLO 法 [H. Kashiwagi, et. al., *Mol. Phys.*, 101 81 (2003)] を用いる。QCLO 法は、まず1残基ごとの量子化学計算を行い、その結果から3残基の初期値を作成、3残基の結果から7残基の初期値を作成・・・と順次計算領域を拡大し、結果として分子全体の良好な初期値を作成する方法である(図1)。

この QCLO 法を発展させ、イオン結合、ジスルフィド結合、2次構造を考慮に入れることで更に収束効果が高くなる計算手法を開発している [T.Hirano, et. al., *J. Chem. Phys.*, 127, 184106]。本研究では、ヘテロ原子を含むタンパク質量子化学計算の自動

化に着手し、ほとんどのタンパク質の自動計算が行えるシステムを構築する。

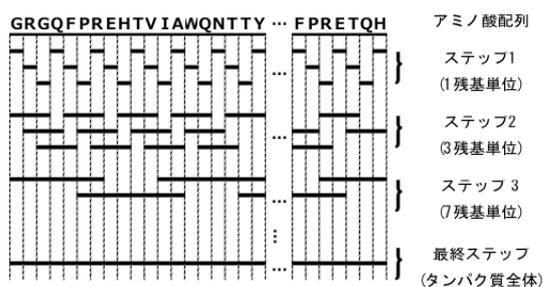


図 1. QCLO 法の計算シナリオ

4. 研究成果

(1) 自動モデリング法

PDB に登録されている結晶構造の多くは水素原子が欠損しており、また NMR 構造の多くは構造的歪みが随所に見られる。低分子の量子化学計算では、量子化学的に構造最適化を行い、その構造における電子状態を議論するが、タンパク質のような大規模分子の場合、1 点計算でも大変な作業であり、量子化学的な構造最適化は困難である。そこで本研究では、分子力学(MM)・分子動力学(MD)法を用いて計算構造のモデリングを自動的に行う手法を開発した。本手法では対象のタンパク質に対し、1) 水素付加 2) 構造緩和 3) 荷電アミノ酸に対する対イオンの付加 4) 再度構造緩和 の手順を段階的に行った。分子動力学プログラムおよび水素付加・力場パラメータの付与は、タンパク質の MD プログラムとして実証例の多い Amber プログラムを用いた。各ステップの処理内容は Python スクリプトで記述しており、障害などにより処理途中で停止した場合でも、途中から再開できるように工夫した。

力場が用意されていないヘテロ分子に対しては、自動的に力場パラメータを用意し、適用するようにした。対象となるヘテロ分子の構造を抜き出し、AM1 計算を行い、その計算結果から GAFF 力場を出力、その力場を割り当てられるようになっている。

分子動力学プログラム実行の前処理として、それぞれのタンパク質構造に関する情報を割り当てる必要がある。まず、接近した荷電アミノ酸残基に対し、イオン対を構成するかどうかの判断処理を実装した。荷電アミノ酸残基側鎖におけるアミノ基やカルボキシル基などの重心からの距離に対し、閾値(4.0 Å)よりも短い場合は、それらの荷電アミノ酸残基をイオン対として判断した。イオン対を構成しない荷電アミノ酸残基は、対イオンとして Na⁺または Cl⁻を近傍に配置した。これらの対イオンは、量子化学計算時にはダミー原子として処理される。また、接近した 2 つのシステイン残基はジスルフィド結合を形成するとして判断、結合処理を行っている。

溶媒モデルは Generalized Born/Surface Area (GB/SA)モデルを使用した。露な溶媒分子を使用しないことで、計算コストを下げることが出来た。

こうして開発したプログラムは、後述の QCLO 法プログラム QCLObot に同梱している。

(2) タンパク質量子化学計算の自動化・効率化に関する研究

① 自由なフラグメント分割が可能な QCLO 法の開発

QCLO 法において、量子化学計算を行う計算単位をフレーム分子、QCLO を構成する単位をフラグメントとよぶ。フレーム分子は 1 つ以上のフラグメントから構成される。QCLO 法における計算手順を計算シナリオとよび、重なり部分を持たせながら伸長する独特な方法を採用している。

ペプチド結合のような規則正しい結合様式において、その QCLO フラグメントの分割をコンピュータでパターン化して処理することは比較的容易である。しかし、ヘテロ分子はパターン化が難しく、フラグメント分割の自動化は困難である。本研究では、まずフラグメント分割を手で行うことを前提に、そのフラグメント分割からフレーム分子の構築、そしてフレーム分子の量子化学計算結果から各フラグメントの QCLO 作成までを簡単な操作で達成するプログラムの作成を行った。

プログラミング言語として、Python を採用した。Python は世界的に広く使われているスクリプト言語であり、プログラムの改変が容易である。オブジェクト指向プログラミングが可能であり、また誰が書いても同じような処理になるため、可読性が高い特徴がある。したがって、本研究で開発されたプログラムはメンテナンスが容易である特徴を自動的に有する。Python はグルー言語として、C/C++などの開発言語で作られたプログラムとの接続性も高く、ProteinDFなどの量子化学計算プログラムとの親和性が高いことも Python を採用する理由の 1 つとなった。本研究で開発したプログラム QCLObot はオープンソース (GPL v3)として、インターネット上に公開した(<http://proteindf.github.io/>)。

本プログラムを用いて QCLO 法による計算シナリオを記述する場合は、まずフラグメントを定義し、そのフラグメントを束ねてフレーム分子を構築する。操作としては、フラグメントオブジェクトは辞書型オブジェクトであるフレーム分子オブジェクトに代入することで表現できるため、簡単に操作できる。作成したフレーム分子オブジェクトは、1 点計算メンバ関数により、自動的に計算環境を整備し、1 点計算を行う。基底関数や交換相関汎関数の種類など、量子化学計算に必要なパラメータは、フレ

ーム分子オブジェクトのプロパティとして
 予め設定できる。計算完了後は、フレーム
 分子を構成するフラグメントを新たなフレ
 ーム分子のフラグメントとして登録するこ
 とで、自動的に QCLO を計算し、初期値と
 して利用することができる。計算シナリオ
 のプログラミングでは、フレーム分子を構
 成するフラグメントの取り扱いにだけ注意
 すれば良く、QCLO の作成など煩雑な処理
 は自動的に行われるように工夫されている。

また本プログラムの特徴として、冪等性
 を重視した。良好な初期値を得るために、
 計算シナリオを途中で組み直すことは多い。
 その度に再計算するのは時間の無駄である
 とともに、余計なエラーを生みやすい。本
 プログラムでは、同じ操作を何度行っても
 同じ結果が得られるように工夫されている。
 また、処理済みの項目は再度処理されない
 ように工夫されている。

②スコア関数の改良

QCLObot プログラムによって、フラグ
 メントを自由にプログラミングが可能にな
 ったが、それだけではフレーム分子とフラ
 グメントの取り扱いが容易になっただけで、
 正しい初期値は得られない。フラグメント
 の QCLO を求める際に、局在化軌道(LO)
 計算とその LO を拾い上げる操作(pick up)
 が必要である。本研究では、pick up の操
 作を改良し、フラグメント分割が自由に行
 えるように工夫した。

これまでのフラグメント分割では、分割
 に起因する誤差を小さくするために、単結
 合の結合間で分割する方法を採用していた。
 ペプチド結合の場合は、主鎖の C-alpha 炭
 素がそれに対応する。LO 計算後の pick up
 操作では、各 LO がどの原子に主に所属す
 るかを population 計算により判断し、その
 原子が所属するフラグメントに LO を割り
 当てていた。

本研究では、原子ではなくフラグメント
 の population を計算し、population が大
 きいフラグメントに LO を割り当てるよう
 にした。現在の LO プログラムは、Mulliken
 Population を最大にするように局在化する
 Pipek-Mezey 法を用いているが、コレス
 キー分解法や RLMO 法などフラグメント
 毎に局在化できる手法を採用して高速化が
 図れると期待できる。

③新しい QCLO 法の実証計算

ヘテロ分子の一例として、フラビンアデ
 ニンジヌクレオチド(FAD)を新しい QCLO
 法で計算した。

フラビンアデニンジヌクレオチドを 6 つ
 のフラグメントに分割し、リン酸基を重なり
 領域として持たせた(図 2)。本研究の
 QCLO 法によって初期値を得ることに成
 功し、良好な収束解を得ることができた。
 一方、Harris の汎関数による初期値を使用

した場合は収束解が得られなかった。本研
 究で開発した QCLO 法プログラムがヘテ
 ロ分子でも有効に機能することが示された。

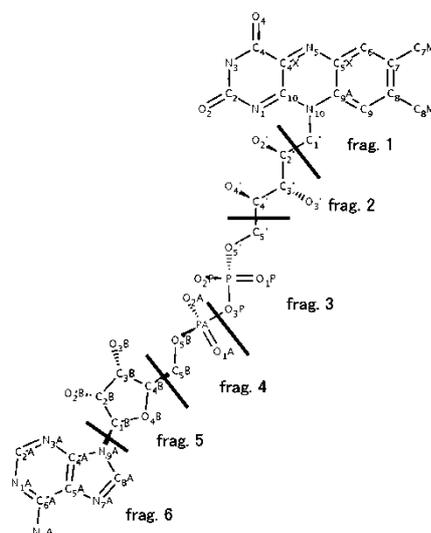


図 2. FAD のフラグメント分割

③第三世代密度汎関数計算法の開発

QCLO 法では、対象となるタンパク質だ
 けでなく、計算シナリオ中のフレーム分子
 の量子化学計算も行う必要があるため、1
 点計算をなるべく速く行う必要がある。現
 在主流の分散メモリ型並列計算機で効率的
 に計算するために、第三世代密度汎関数計
 算法を開発、提案した[T.Hirano, et. al.,
 PCCP, 16, 14496]。

第三世代密度汎関数計算法では、コレス
 キー分解法とグリッドフリー法に基づき、
 計算律速となる分子積分を SCF 繰り返し
 計算の前に 1 度だけ行い、SCF 繰り返し計
 算中では行列演算のみを行い電子状態を計
 算するものである。従来グリッド法による
 数値積分で求めた交換相関項も、グリッド
 フリー法を適用することによって行列演算
 に置換することができた。分散メモリ型並
 列計算機において、巨大な大域行列を各計
 算ノードに分散保持し、計算機に最適化さ
 れた ScaLAPACK ライブラリにて行列演
 算することができるので、タスク分散も容
 易に高速化が図れる。

以上の研究成果により、PDB から得られ
 たタンパク質構造情報から計算モデルを作
 成し、そのモデル構造に対する 1 点計算が
 一通り可能となった。量子化学計算に耐え
 るタンパク質のモデリング方法は類が無く、
 本研究で開発したモデリング方法はこれ
 を解決する一つの方法として利用すること
 ができる。また、自由なフラグメント分
 割可能な QCLO 法プログラムを開発し、大
 規模量子化学計算に必要な初期値を容易に
 作成できるようになった。第 3 世代密度汎
 関数法の開発により、量子化学計算が PC
 クラスタなどの分散メモリ型並列計算機に

においても効率良く達成できるようになった。規則性のあるペプチド鎖だけでなく、ヘテロ分子にも対応できるようになったことで、取り扱えるタンパク質の種類が一気に増加した。しかし、依然としてヘテロ分子の計算シナリオは試行錯誤が必要であり、今後自動的に最適な計算シナリオを作成する仕組みを開発する必要があるだろう。

5. 主な発表論文等

[雑誌論文] (計 1 件)

① T.Hirano, F.Sato, “A third-generation density-functional-theory-based method for calculating canonical molecular orbitals of large molecules.”, *Phys. Chem. Chem. Phys.*, 査読有, **16**, 14496-14503 (2014).

DOI: 10.1039/c3cp55514c

[学会発表] (計 6 件)

① 平野敏行, 佐藤文俊, “自由なフラグメント分割可能な QCLO プログラムの開発”, 分子科学討論会, 大阪大学(大阪), 2015.

② J. Yoshida, T. Hirano, F. Sato, “Study on QCLO method for canonical molecular orbital calculation of cofactor-containing protein”, CBI 2014 annual meeting, タワーホール船堀(東京), 2014.

③ 平野敏行, 佐藤文俊, “グリッドフリー密度汎関数計算法の進展”, 分子科学討論会, 広島大学(東広島), 2014.

④ 吉田洵也, 平野敏行, 佐藤文俊, “酸化型グルコースオキシダーゼの電子状態計算”, 理論化学討論会, 名古屋大学(名古屋), 2014.

⑤ 平野敏行, 佐藤文俊, “次世代スーパーコンピュータに向けたタンパク質の電子状態シミュレーション”, 東京大学生命科学シンポジウム 2014, 東京大学(東京), 2014.

⑥ 平野敏行, 樋口恒, 佐藤文俊, “コレスキ分解を用いたグリッドフリー密度汎関数計算の研究”, 第 7 回分子科学討論会, 京都テルサ(京都), 2013.

[その他]

ホームページ等

① <https://proteindf.github.io/proteindf/proteindf/>

② <https://proteindf.github.io/proteindf/qclobot/>

③ <http://satolab.iis.u-tokyo.ac.jp/>

6. 研究組織

(1) 研究代表者

平野 敏行 (HIRANO Toshiyuki)

東京大学 生産技術研究所 助教

研究者番号: 60451887