

蛋白質の量子化学反応解析システムの開発

アドバンスソフト株式会社 ○小池 秀耀

Development of a quantum chemical analysis system for proteins
Hideaki Koike, AdvanceSoft Corporation

Abstract:

We have developed a quantum chemical analysis system for proteins. It supports quantum chemical calculation of a proteins, all-electron and/or structural optimization calculations, and the integrated environment through which we can easily and intuitively operate the simulation.

1. はじめに

ポストゲノム時代において、タンパク質の機能・構造解析は重要なテーマである。量子化学に基づくタンパク質の機能・構造解析を実現することを目的として、密度汎関数法に基づく分子軌道法プログラムProteinDFを核とした、スーパーコンピュータ・ネットワーク上で分散・並列処理が可能な大規模タンパク質の量子化学計算システムを開発した。

2. 研究開発項目とその成果概要

2.1 研究開発項目

量子化学に基づくタンパク質の機能・構造解析を実現することを目的として、密度汎関数法に基づく分子軌道法プログラムProteinDFを核とした、スーパーコンピュータ・ネットワーク上で分散・並列処理が可能な大規模タンパク質の量子化学計算システムを開発した。本研究開発項目を以下に示す。

- 最適機種選択ネットワークコンピューティング機能
- タンパク質全電子自動計算用初期値作成シナリオ・エディタ
- 高速構造最適化計算システム
- PDB、MDとのインターフェース

2.2 蛋白質の量子化学反応解析システムの概要

蛋白質の量子化学反応解析システムは、密度汎関数法に基づく分子軌道法プログラムProteinDF、タンパク質全電子自動計算の初期値作成のための半自動化プログラム、構造最適化プログラム、PDB、MDインターフェース、ユーザ・インターフェースから構成される。蛋白質の量子化学反応解析システムのシステム構成を図2に示す。

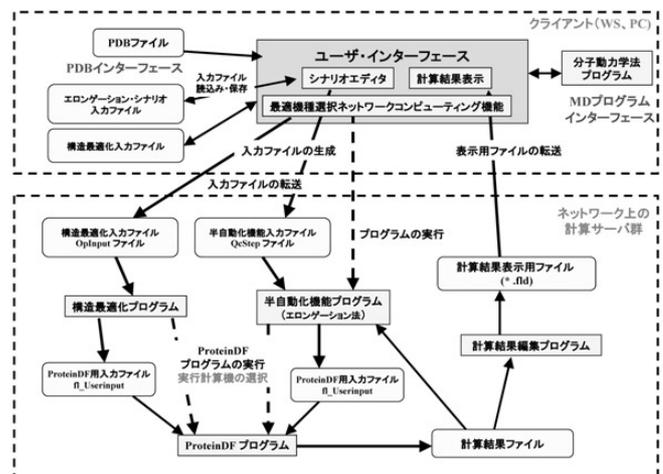


図1 蛋白質の量子化学反応解析システムのシステム構成

2.3 タンパク質全電子自動計算用初期値作成シナリオ・エディタ

タンパク質全電子自動計算を行うためには、適当な初期データを作成する必要がある。擬カノニカル局在軌道法は、小規模なアミノ酸残基の計算結果から、段階的に結合させながら、大規模なタンパク質全電子計算のための初期データを作成していく手法である。計算対象のタンパク質をフラグメント分割し、各フラグメントの擬カノニカル局在軌道を結合して初期値を作成する。シナリオ・エディタは、擬カノニカル局在軌道法に基づいて、シナリオの作成・編集、および半自動化プログラムおよび分子軌道法プログラムProteinDFを用いて初期値を作成、全電子計算を行うシステムである。ProteinDFシステム GUIとシナリオ・エディタを図2に示す。



図2 ProteinDFシステム GUIとシナリオ・エディタ

シトクロムcの活性中心の近傍モデルである6残基モデルを計算事例として、シナリオ・エディタを用いて、全電子計算の自動計算を行った。6残基Minimalモデルは、鉄とポルフィリン、17-19残基、79-81残基から構成される分子モデルである。ユーザは、計算シナリオを記述した入力データを作成することにより、ProteinDFシステムが一連の全電子計算の処理を自動的に実行する。シトクロムc 6残基モデルの計算シナリオを図3に示す。6残基モデルは、鉄とポルフィリンの分子モデルから始めて、鉄とポルフィ

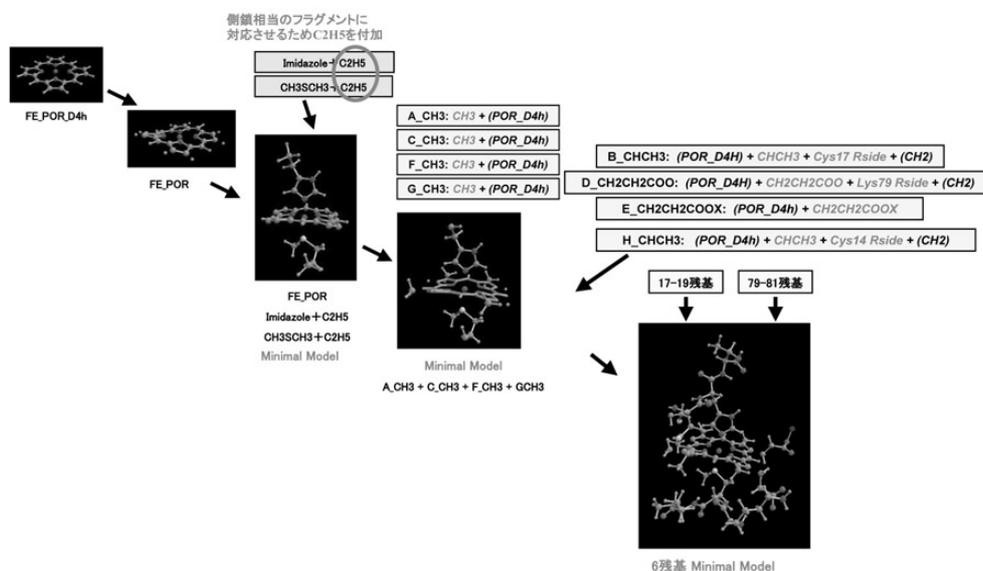


図3 6残基モデルの計算シナリオ

リンと18番残基と79番残基の一部の分子モデル、6残基モデルと計算している。図4に、6残基モデルの計算結果であるHOMOとLUMOの等値面 (± 0.005) を示す。

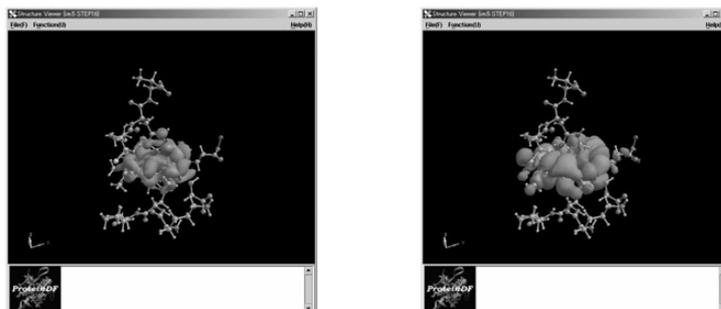


図4 6残基モデルの全電子計算結果 (HOMO, LUMOの等値面 (± 0.005))

2.4 構造最適化計算プログラム

タンパク質の構造解析を目的に、構造最適化プログラムの開発を行った。構造最適化プログラムはエネルギー微分計算と最適化計算から構成される。エネルギー微分計算は、分子軌道法プログラムProteinDFの積分計算部を拡張した。基底関数および補助基底関数は、s,p,d軌道まで対応している。積分計算はObara-Saikaの手法、最適化手法は共役勾配法 (CG法) 等を用いている。構造最適化プログラムのプログラム構造を図5に示す。構造最適化プログラムは、高速化のために並列化を行っている。

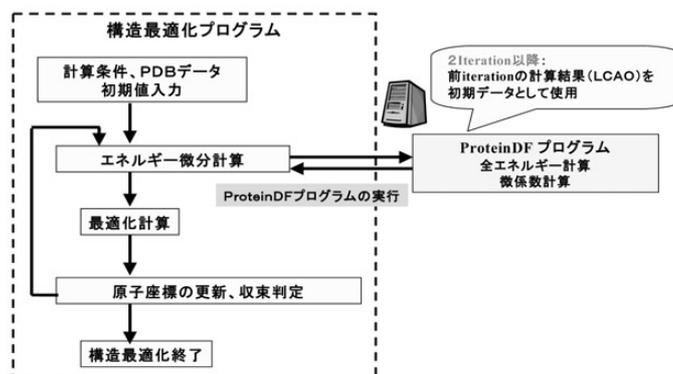
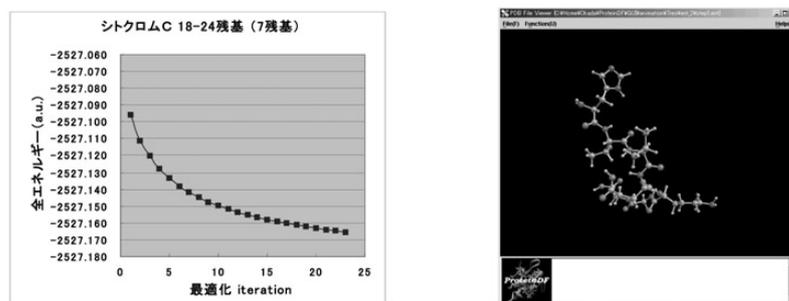


図5 構造最適化システムのプログラム構造

構造最適化計算の事例として、シトクロムc ペプチド鎖に対し構造最適化計算を行った。計算事例は、シトクロムc ペプチド鎖 7残基 (18-24番残基、原子数101、軌道数559) である。シトクロムc ペプチド鎖 7残基の構造最適化の計算結果を図7に示す。



(a) 全エネルギー

(b) 構造最適化後の分子構造

図6 シトクロムc ペプチド鎖7残基の構造最適化計算結果

2.5 PDB, MDとのインターフェース

タンパク質の解析を行うためにはネットワーク上に分散するPDB（プロテイン・データバンク）などのデータベースへのアクセスやMD（分子動力学）などの結果の参照が不可欠である。PDB、MDインターフェースは、PDBデータに対し、ProteinDFシステム用の入力データの分子構造を生成する。PDBデータに対し、溶媒（水分子）を付加し、分子動力学法プログラムを用いて、エネルギー最小化計算を行い、構造最適化を行う。次に、溶媒（水分子）を取り除くことにより、タンパク質を抽出し、タンパク質の中酸化、表面処理を行う。図7に、PDB、MDインターフェースの概要を示す。

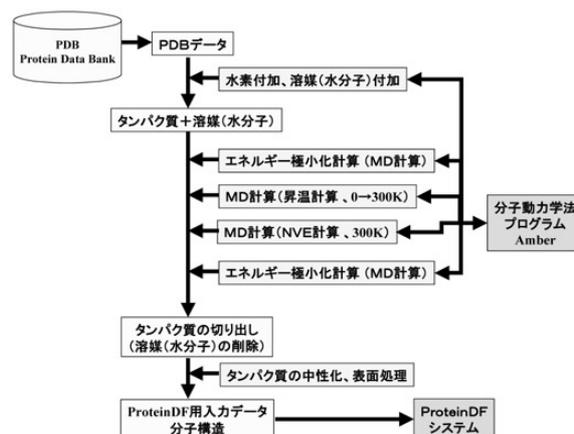


図7 PDB、MDインターフェースの概要

2.6 最適機種選択ネットワークコンピューティング機能

巨大な量子化学計算にはネットワーク上の分散並列処理が必要不可欠である。特に、タンパク質全電子自動計算では、小規模なアミノ酸残基の計算結果から、段階的に結合させながら、大規模なタンパク質全電子計算のための初期データを作成しているため、ProteinDFプログラムの複数実行が必要となり、並列分散処理を行うことにより、計算処理の大幅な高速化が図れる。タンパク質の自動計算において、計算内容を判断し、ネットワーク上に分散する計算資源を自動選択して、並列分散処理を行う機能を開発した。計算フレーム分子のサイズに応じた並列処理が可能である。自動計算法における並列分散処理の概要を図8に示す。

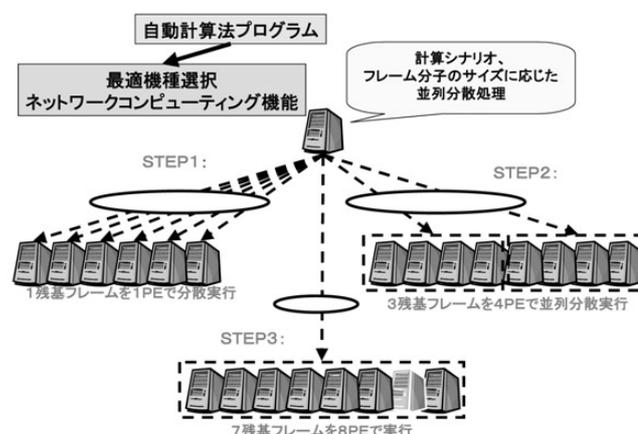


図8 自動計算法における並列分散処理の概要

3. ネットワークの活用について

ネットワーク上に分散した計算資源、並列計算機上において、システムの動作実験を行った。

4. まとめ

蛋白質の量子化学反応解析システムの開発として、タンパク質全電子自動計算用初期値作成シナリオ・エディタ、構造最適化計算プログラム、最適機種選択ネットワークコンピューティング機能、PDBインターフェースの開発を行い、蛋白質の量子化学反応解析システム全体のプロトタイプを開発した。シクロロムcの一部を例題として、システム全体および各機能の有効性を確認した。

5. 研究開発実施体制

代表研究者 アドバンスソフト株式会社 小池 秀耀

研究分担

東京大学 生産技術研究所 計算科学技術連携研究センター 佐藤 文俊

アドバンスソフト 株式会社 技術第2部 小池 聡、長峰 康雄

研究協力

東京大学 生産技術研究所 / アドバンスソフト 柏木 浩、 株式会社 富士総合研究所 西川 宜孝