



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、上村典子、 恒川直樹、西村康幸、山口智美
作成年月日	2009年6月
コード名	ProteinDF
使用言語	C++

◇ProteinDF システム

当グループでは、アミノ酸残基とヘテロ分子で同程度の定量性を持ち、タンパク質をありのまま扱う量子化学計算による解析手段として、密度汎関数 (DF) 法プログラム ProteinDF と、それを中核とする ProteinDF システムを開発してきた[1、2]。ProteinDF は全電子を量子的に扱う、すなわち全カノニカル計算を行うという汎用性から、タンパク質のみに限らずナノ素材のターゲットとなる様々な機能性高分子も取り扱える計算能力があり、ナノスケールものづくりに貢献できる可能性が示唆されていた。本プロジェクトでは、実用的なプリ・ポスト処理機能を追加しつつ、ProteinDF システムをバイオ・ナノ分子特性シミュレーションシステムとして応用・発展すべく、開発・実証計算を行っている。

ProteinDF システムは、次の5つのソフトウェアからなる。1. タンパク質密度汎関数法エンジン ProteinDF とそのソルバ QCLO、2. 力場や量子計算エネルギー勾配を利用した構造最適化や動的構造解析を扱う ProteinMD、3. 配置間相互作用法エンジン ProteinCI、4. タンパク質波動関数データベース ProteinQR、そして5. 統合 GUI の ProteinEditor である。現在もプロジェクトを通して開発中であるが、1のソフトウェアにより300残基超のタンパク質全電子計算が達成でき、2のソフトウェアにより通常の古典 MD や小さなタンパク質の DFT-MD 計算が行える。励起状態の計算が可能な3のソフトウェアは今回非公開だが、設計が終了し、実装・実証計算を行う段階に達した。4のデータベースには100種ほどの全電子計算例が蓄積された。5のソフトウェアにより、100残基規模のタンパク質全電子計算が実行できる統合環境を提供した。今回、統合環境の整備やソルバの性能向上といった様々な工夫により、実用性がさらに増した。ProteinDF システムを用いることによって、容易に最先端の計算を達成し、バイオ・ナノ分子の物性・反応性の理解・応用へ利用できると期待される。ここでは、ProteinDF システムの一部として機能追加・性能向上について紹介する。なお、以上のプログラムは随

時アップデートされており、インターネットからダウンロードして無料で利用することができる[3、4]。

◇ ProteinDF システムの注目すべき新機能・開発項目

タンパク質 DF 計算エンジン (ProteinDF) とポスト処理の高速化：ProteinDF はガウス型基底関数を用いた、RI (Resolution of Identity) 法に基づくタンパク質のための DF 法プログラムである[5]。今回、行列演算ライブラリである LAPACK、ScaLAPACK をランタイムに切り替えることで、1つの実行バイナリで多様な計算機環境とシミュレーションモデルに対応させる改良を施した。また、一般化学分子シミュレーションで標準となっている B3LYP 汎関数の新しい並列化実装に成功し、分散メモリ型並列計算機を使った大規模タンパク質の B3LYP 計算が可能となった。これまでの実証計算の結果から、B3LYP 汎関数は局所密度汎関数よりも計算コストが増加するものの、これまで大規模分子の電子状態計算が抱えていた収束の問題を大幅に解決する傾向にあり、今後主流となる計算手法になると思われる。

全電子カノニカル計算の大きな利点の一つは、これまで培われてきた計算化学の解析手法がそのまま利用できることである。電子状態計算の結果(波動関数)から得られる物理量には、エネルギーやフロンティア軌道である HOMO、LUMO をはじめとする分子軌道、電子密度、静電ポテンシャル、そして IR・Raman、NMR などの分光学的データなど多岐にわたる。このような波動関数から二次的に求められる物理量計算、すなわちポスト処理もまた同様に、シミュレーション規模が大きくなるにつれて計算コストが増加する。例えば 51 残基のインスリンの静電ポテンシャルを求めるためには、Opteron 2.0 GHz を利用した場合、25 時間もの時間を要する。この静電ポテンシャル計算を高速化するため、我々は GPU (Graphics Processing Unit) に注目し、高精度かつ極めて高速な演算手法を開発した[6]。GPU は、元来コンピュータ・グラフィックスをはじめとする画像処理を CPU の代わりに演算する加速装置であり、現在多くの PC やワークステーションには、ほとんど GPU が装着されている。その GPU の極めて高い演算能力が近年着目されており、科学技術計算やシミュレーションなどの汎用的な用途へ広く利用されつつある。我々はこの GPU (Tesla C860) を利用することにより、インスリンの全電子計算に基づく精密な静電ポテンシャル計算をわずか 12 分で達成するコードを開発した。この計算コードには分子積分などの量子化学計算で典型的な律速計算の処理が含まれている。本開発の成功により、他の様々なポスト処理計算にも GPU が利用できる期待が高まった。今後随時、統合環境 GUI である ProteinEditor などから GPU を利用し、デスクサイドでポスト処理ができるようにしたいと考えている。

共通データ・アルゴリズム Bridge クラスの設計と実装：ProteinDF システムは複数のソフトウェアで構成されている。それぞれのソフトウェアは単独動作でも十分強力な機能を有しているが、各ソフトウェアが有機的に連携することで、より多彩なシミュレーションを達成することができる。各プログラムが共通に保有している機能、ならびに連携すべき機能を抽出または新規開発し、クラスライブラリとして利用可能にすることを目的とした。概略図を Figure 1 に示した。それぞれのソフトウェアの橋渡しをするという意味で、Bridge クラスと名付けている。

Bridge クラスライブラリは、データの保持を扱うコンテナクラスと、そのデータに対して様々な処理を行うアルゴリズムの 2 種類のクラスから構成されている。このクラスライブラリは今後 ProteinDF システムの各ソフトウェアに組み込まれ、ソフトウェア間の連携や機能追加が図られる。例えば、

ProteinDF で求めた全電子波動関数から RESP 電荷を求め、このデータを ProteinMD に渡すことで全電子 DF 計算に基づく電荷を用いた古典分子動力学計算が可能となる。また、2次構造判定アルゴリズムは ProteinEditor でタンパク質構造の表示に利用されると共に、QCLO 法に基づく自動計算プログラム (QCLO) は、安全な収束と効率的な計算を達成する構造を考慮したシナリオをユーザに提案することが可能となる。また、Bridge クラスは ProteinCI などの新規モジュールの追加・連携も容易にする。

本クラスライブラリは標準 C++ で記述し、汎用性を持たせた。データのシリアル化には、エンディアンの違いを判別できるデータ構造になっており、環境が変わってもユーザおよびプログラマー共に特別な操作を必要とせず、データの再利用が可能となった。この点は様々な計算機環境からデータを送受信し、データを登録・加工するデータベース (ProteinQR) において、強力な利点になることを強調したい。また、python スクリプトによる簡便なデータハンドリングが可能で、本クラスライブラリをラップした python クラスも用意している。これら C++ クラスライブラリ、python クラスは適宜インターネットなどにより公開する予定である。これにより、ユーザによる ProteinDF システムを利用したツールの独自開発や、自由なシミュレーションの操作が可能になるものと期待される。

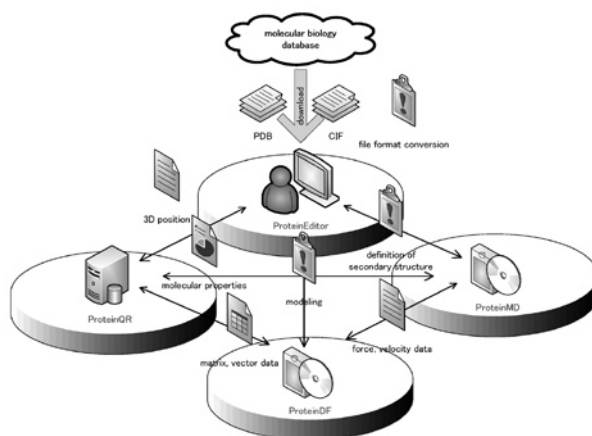


Figure 1. 共通データ・アルゴリズム Bridge クラスの概略図

大規模分子モデリングのための GUI フレームワーク：これまでにタンパク質の密度汎関数法による全電子計算をサポートするまったく新しい GUI、ProteinEditor の開発に成功した。現在では薬剤設計のための構造モデリング機能を強化、新しい物性を表現する手段を開発することによって、より高度な GUI へと成長させている。電子状態計算において、適切に計算構造を決定することは必要不可欠である。これをサポートするソフトウェアは分子モデリングとよばれており、主に低分子を対象にさまざまなアプリケーションが開発され、多くの計算科学者が利用している。タンパク質をはじめとする大規模分子のモデリングにおいても、低分子モデリングで養われた技術は適用可能である。とはいえ、数千から数万にもおよぶ原子座標を手作業で分子モデリングするのは、現実的な作業ではない。特に大規模分子モデリングでは、必然的に超充填複雑構造となることから、許容できないボンドの欠損や原子間の衝突が生じやすく、またこれを見逃しやすいという特徴もある。大規模分子のモデリングには、相応の方法論が必要である。構造編集と構造チェック機能、および編集操作機能を向上させるため、大規模分

子モデリングに特化した GUI、ProteinModeler を新たに設計・試作した。ProteinModeler のフレームワークは、「メインメニュー」、「ツールバー」、「コントロールタブ」、「表示ウィンドウ」、「ログウィンドウ」で構成した (Figure 2)。コントロールタブには、編集時のタンパク質構造データ、およびツールバーで選択されたモデリング機能のパラメータ設定を行うための GUI がタブ形式で表示され、表示ウィンドウには編集時のタンパク質の立体構造を表示、ログウィンドウには操作上のメッセージなどのログが表示される。また、本フレームワークは簡易版インターフェース (I/F) と詳細版 I/F を持たせ、ユーザの要望に広く応える作りとなっている。簡易版 I/F では、一般的な分子モデリングを難しい設定を必要とせず直感的に操作することができる。詳細版 I/F では、詳細な作業が可能な様々な機能を利用できるようになっている。さらにモデリング操作の手順をワークフロー形式のネットワークで表現することにより独自のモデリング操作を作成することができる。ネットワークは保存・再利用することが可能である。この意欲的なモデリングツールは、プロジェクトを通して、前述の Bridge クラスとの連携も図り、大容量のデータ処理と繊細なデータ操作を併せ持つ、大規模モデリングソフトへと発展させる予定である。

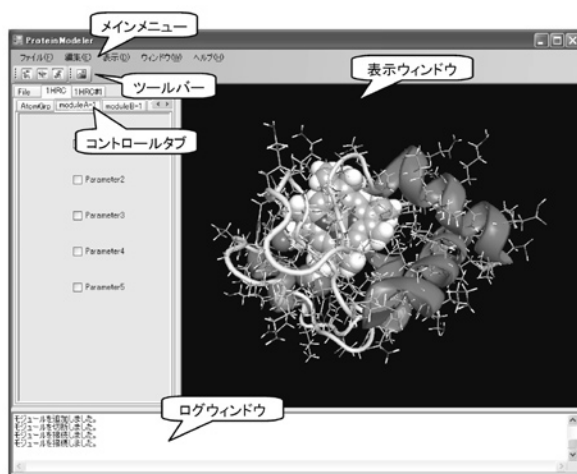


Figure 2. 開発中の ProteinModeler のスナップショット

◇参考文献

- [1] 柏木 浩、佐藤文俊、吉廣 保、稲葉 亨、西川宜孝、小池 聡：“タンパク質量子化学計算—ProteinDF の夢と実現—”, 柏木 浩、佐藤文俊監修、アドバンスソフト (2004).
- [2] F. Sato, T. Yoshihiro, M. Era, H. Kashiwagi : *Chem. Phys. Lett.*, **341**, 645 (2001).
- [3] <http://www.ciss.iis.u-tokyo.ac.jp/>
- [4] 佐藤文俊、中野達也、望月祐志：“プログラムで実践する生体分子量子化学計算”, 森北出版 (2008).
- [5] 柏木 浩、佐藤文俊、恒川直樹、吉廣 保、平野敏行、井原直樹：“タンパク質密度汎関数法”, 森北出版 (2008).
- [6] 湯川英宜、平野敏行、西村康幸、佐藤文俊：“GPU によるタンパク質高精度静電ポテンシャル計算の高速化”, 生産研究, **61**, 103 (2009).