

4Pa035

ProteinDF をベースにしたタンパク質構造最適化プログラムの開発

(東大生研[†]・富士総研[‡]・アドバンスソフト^{*})

○井原直樹[†]・金城友之[†]・恒川直樹[†]・岡田一俊[‡]・西川宜孝[‡]・上野哲哉[†]・佐藤文俊[†]・柏木浩^{†*}

【背景】ポストゲノム時代に突入し、タンパク質の立体構造および、その機能の解析がさらに注目を浴びている。実験研究では、X線構造解析・NMR・電子顕微鏡などを用いて原子レベルの分解能でタンパク質の立体構造が次々と決定されている。一方、理論研究においてもコンピュータシミュレーションによる立体構造予測が精力的になされているが、そのシミュレーションの多くはデータベース型の力場をもとにした古典シミュレーションであり、これらの力場が提供するエネルギー精度は十分なものではない。自然界のタンパク質は生体内の環境下で機能している。温度 300K の生体溶液中におけるタンパク質の構造や機能の変化は、0.1eV 程度のわずかな自由エネルギー変化によっておこる。このことは、タンパク質のエネルギー計算において許される誤差も 0.1eV 以下が望ましいことを意味する。しかしながら、既存のデータベース型の力場では、この精度を保証していない。これらの力場では、同じ種類のアミノ酸に同じ力場を使用しているが、実際にはタンパク質内のアミノ酸毎にこれらのパラメータが異なる。その端的な例を表 1 に示す。これは、密度汎関数法によるシトクロム *c* の全電子計算によって得られたアラニンの Mulliken 電荷である。アラニンは全部で 6 つあり、各アラニンの電荷分布は当然異なっている。その量をそれぞれの電荷の和で比較してみると、最大で 0.1 以上である。この電荷の違いから、単一力場による方法では 10Å 離れた先においても誤差が約 0.1eV にも達することを示している。電荷に限らず、その他のエネルギー成分においても、ほぼ同様の懸念が生じる。タンパク質の全電子状態計算は、このような問題点を解消する最良の方法の一つである。

表 1： シトクロム *c* の全電子計算によるアラニン(ALA)の Mulliken 電荷

	ALA15	ALA43	ALA51	ALA83	ALA96	ALA101
N	-0.4820	-0.5236	-0.4759	-0.5203	-0.4889	-0.4735
C α	-0.0941	-0.1405	-0.1075	-0.0903	-0.0955	-0.1010
C	0.2254	0.2771	0.2057	0.2198	0.2409	0.2497
O	-0.3562	-0.3706	-0.3363	-0.3673	-0.3673	-0.3887
C β	-0.6150	-0.5116	-0.5652	-0.5912	-0.6137	-0.6123
H	0.4107	0.4350	0.3329	0.3799	0.4018	0.4016
H α	0.2457	0.1971	0.2642	0.2405	0.2448	0.2400
H β 1	0.2590	0.2132	0.2573	0.2478	0.2275	0.2453
H β 2	0.2134	0.1791	0.2411	0.2101	0.2755	0.2404
H β 3	0.2457	0.2625	0.2179	0.2294	0.2514	0.2227
合計	0.0526	0.0177	0.0342	-0.0416	0.0765	0.0241

【目的】発表者らは、タンパク質のような巨大な系を扱うことができ、精度の良い計算が望める方法として、現在の計算機事情では密度汎関数法が適当であると考え、密度汎関数法によるタンパク質の全電子計算を実現するソフトウェア ProteinDF^[1]を開発している。本研究の目的は、こ

の ProteinDF から得られる電子状態を反映させた力場を取り込み、タンパク質の立体構造を高精度で求めるための構造最適化プログラムの開発を行うことである。

【方法】近年、電子の運動は量子化学的に、核の運動は古典力学的に取り扱い、これらをカップルさせることによって分子動力学計算 (MD) を行う方法が開発され、一般に *ab initio* MD と呼ばれている。本研究では、電子状態の計算には ProteinDF を採用し、核の運動には新たに MD プログラムを開発することにより、タンパク質の *ab initio* MD および *ab initio* 構造最適化を実現している。ProteinDF は大容量メモリの安全管理やプログラムの拡張性を考慮に入れて、オブジェクト指向言語 C++ によってコーディングされている。そこで、本研究の MD プログラムも C++ でコーディングを行った。これによる最も大きな利点は、ProteinDF のデータと MD プログラムのデータのやり取りは、そのデータ構造をあらわに取り扱う必要がないオブジェクトを介して行うことができるため、容易かつ安全に結合することができることである。

本 MD プログラム構築にもオブジェクト指向の考え方を導入した。本プログラムの制御オブジェクトから、シミュレーションに必要な種々のデータを保持する MD オブジェクトと Verocity Verlet 法や擬 Newton 法といった様々な計算オブジェクトが生成される。これらの計算オブジェクトは、MD オブジェクト内で抽象化されたインテグレータオブジェクトと連結することにより、さまざまな処理を行う。このような実装法により、MD シミュレーションのカーネルに手を加えることなく、今後も様々な力場やアルゴリズムを追加することができる。また、複数の MD オブジェクトを生成することにより、レプリカ交換法によるシミュレーションが可能となる。

タンパク質のようなパラメータの多い分子における構造最適化の研究は、古典系においても大変困難なフロンティアである。本研究では、レプリカ交換法^[2,3]、マルチカノニカル法^[4,5]といった大局的な構造最適化法と共に、擬 Newton 法のような局所構造最適化法を採用する。しかし、現在のところ量子化学計算には膨大な計算時間を要するため、タンパク質の構造最適化においては、活性中心近傍における部分構造最適化が現実的であると考え、これが取り扱いやすいように実装を工夫している。その他、計算速度を上げる手法として、空間分割法や擬カノニカル局在化軌道法に基づいた *ab initio* MD や *ab initio* 構造最適化の実装も予定している。詳細は当日報告する。

【謝辞】

本研究は、文部科学省 IT プログラム「戦略的基盤ソフトウェアの開発」において実施された。(関連各位に謝意を表する。)

【参考文献】

- [1] F. Sato, Y. Shigemitsu, I. Okazaki, S. Yahiro, M. Fukue, S. Kozuru and H. Kashiwagi, *Int. J. Quantum Chem.* **63**, 245 (1997)
- [2] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996)
- [3] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999)
- [4] N. Nakajima, A. Kidera and H. Nakamura, *J. Phys. Chem. B* **101**, 817 (1997)
- [5] U. Hansmann, Y. Okamoto and F. Eisenmenger, *Chem. Phys. Lett.* **259**, 321 (1996)