

次世代量子化学計算グループ; ProteinDF システム

佐藤 文俊, 井原 直樹, 上野 哲哉, 田原 才静, 恒川 直樹, 西村 康幸, 西野 典子, 平野 敏行, 吉廣 保,
西川 宜孝*, 稲葉 亨*, 長峰 康雄*, 西村 民男*, 小池 聡*, 小池 秀輝*, 柏木 浩
東京大学生産技術研究所, *アドバンスソフト(株)



当グループはガウス型基底関数を用いた密度汎関数法プログラムProteinDFをエンジンとした次世代の量子化学計算システムを開発しています。ProteinDFはタンパク質を丸のまま扱って、電子相関効果を取り込みながら精密な全電子の計算を行うことができます。このシステムはポストゲノム時代におけるバイオ研究開発の有用な道具となるでしょう。本システムは5つのサブシステムから構成されています: (1) 統括グラフィカル・ユーザ・インターフェース、(2) 自動計算法、(3) 量子動力学法および構造最適化法、(4) 大規模計算機用エンジン、そして (5) タンパク質波動関数データベースです。

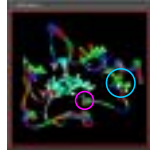
なぜタンパク質の量子化学計算が必要なのでしょうか?

タンパク質はたくさんのアミノ酸残基がつながった巨大分子で、そのうち約 1/3 は金属を持っています。そのうえ、タンパク質は体内という低い温度 (~300K) で働きますから、タンパク質の機能を解析するには少なくとも数千キロカロリーの精度で、巨大分子の性質を見積もる必要があります。タンパク質のシミュレーションは未だに古典力学に基づく方法が主流です。しかし、アミノ酸残基ごとに固定された力場をパラメータに用いる古典法では十分な精度で計算ができません。

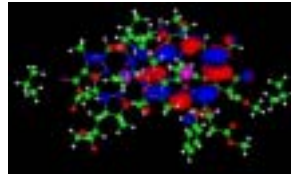
イオン化ポテンシャル	Fe(II) イオン	~700 kcal/mol
	ヘム	~100
	ヘムタンパク質	< 10

ProteinDFで計算されたシクロム c におけるALAのMulliken電荷。
右はAMBER99におけるALAのRESP電荷。

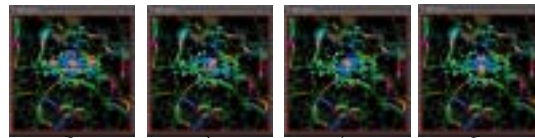
	ALA7E	ALA43	ALA61	ALA63	ALA65	ALA101	ALA/AMBER99
N	-0.4820	-0.5236	-0.4759	-0.5203	-0.4889	-0.4735	-0.4157
O	-0.0941	-0.1405	-0.1075	-0.0903	-0.0865	-0.1010	-0.0239
C	-0.2254	0.2771	-0.2671	0.2198	-0.2408	0.2497	0.5973
O	-0.3662	-0.3706	-0.3963	-0.3873	-0.3673	-0.3867	-0.5679
Cl	-0.6158	-0.5116	-0.5952	-0.5912	-0.6137	-0.6123	-0.1825
H	0.4107	0.4350	0.3329	0.3799	0.4018	0.4018	0.2718
Hx	0.2457	0.1971	0.2642	0.2405	0.2448	0.2400	0.0829
Hβ1	0.2590	0.2132	0.2573	0.2478	0.2275	0.2453	0.0603
Hβ2	0.2104	0.1791	0.2411	0.2101	0.2255	0.2404	0.0603
Hβ3	0.2457	0.2625	0.2179	0.2294	0.2514	0.2227	0.0603
合計	-0.0526	0.0177	-0.0342	-0.0416	0.0786	0.0241	0.0000



シクロム c における ALA の位置。



光合成反応中心タンパク質スペシャルペアのHOMO.

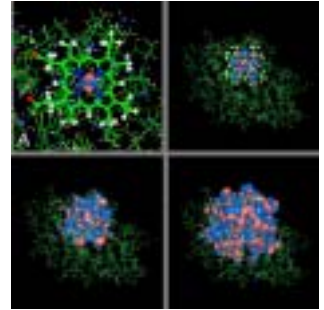


シクロム c の Kohn-Sham 軌道エネルギー分布.

この金属を持ったタンパク質の全電子計算は (9,600 軌道) は 2000 年に世界で初めて達成されました。

ProteinDF とは

- タンパク質の全電子カノニカル波動関数計算ができるガウス型基底関数を用いた密度汎関数法プログラムです。
- 大規模な分子系、特に金属を持ったタンパク質の計算に威力を発揮します。
- オブジェクト指向言語C++によるプログラムです。



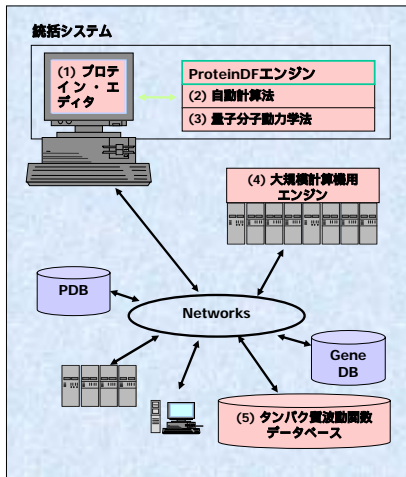
フェロシクロム c の HOMO (3293番目の分子軌道). 等値面の値は ±0.05 (左上A), ±0.005 (右上), ±0.0005 (左下), ±0.00005 (右下). (A)はスケールを倍にして描いている。

ProteinDF システム

当グループはProteinDFをベースにタンパク質のための新しい量子化学計算ソフトを開発しています。

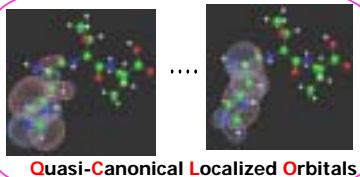
一点計算だけでなく、量子動力学計算や構造最適化計算もサポートし、これら難しい計算がほぼ自動的に達成できる仕組みを組み込んでいます。

さらに、使い勝手を考慮した統括的な環境を開発しており、計算したタンパク質の波動関数のデータベースも提供します。

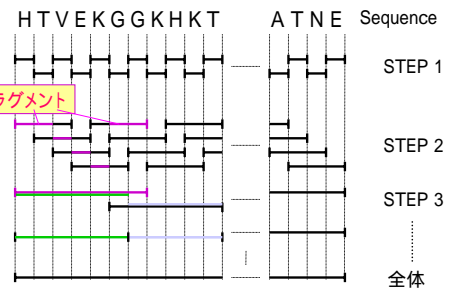


トピックス I; 自動計算と QCLO

QCLO はフラグメントに局在化していますが、その中では通常のカノニカル軌道と同じ性質を持つ特殊な軌道です。QCLO をタンパク質の全電子計算に応用すると、難しいシミュレーションがスムーズに達成できます。



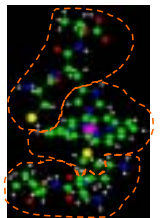
Quasi-Canonical Localized Orbitals



ヘムを持つ活性中心における QCLO の応用.

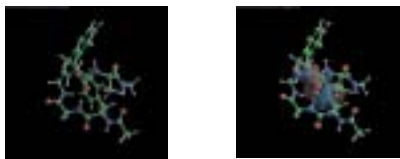
1 フラグメント = 通常のカノニカル軌道

3 フラグメントの QCLO 右図はフラグメントの定義

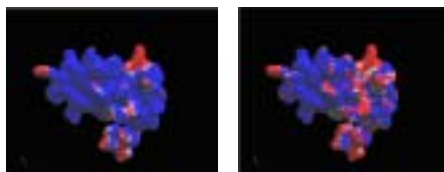


トピックス II;

ProteinDF と古典力学法との結果比較



AMBER99 $\Phi_2 = -165.2$, $\Psi_2 = 55.2$ $\Phi_3 = 82.3$, $\Psi_3 = -41.9$
ProteinDF $\Phi_2 = -161.4$, $\Psi_2 = 37.1$ $\Phi_3 = 103.7$, $\Psi_3 = -30.9$
5残基ペプチド(TYR-GLY-GLY-PHE-ALA)の最適化構造の比較.



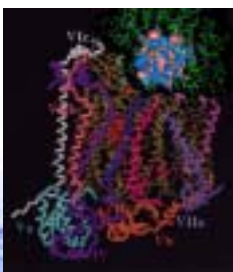
AMBER99 ProteinDF
インスリン (51 残基, 4,439 軌道) の静電ポテンシャルの比較.

トピックス III; 計算時間の飛躍的な短縮と未来展望

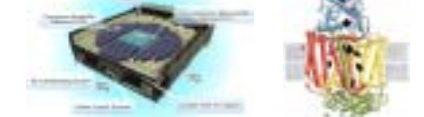
インスリン2量体の計算時間.
•102 残基 (8,878 軌道)
•Altix3700 (256GFLOPS)
55 分/ 回転 (最終結果まで 1.5 日)

ちなみに...
2000年に達成したシクロム c の計算は2ヶ月以上.

近未来の夢:
シクロム c とオキシダーゼ間の長距離電子移動解析.
密度汎関数法と配置間相互作用法を QCLO を使って組み合わせたら...
右図は仮想的な合成図です.
F.Sato et al. *Chem.Phys.Lett.* **341**, 645-651 (2001) and T.Tsukihara et al. *Science* **272**, 1136-1144 (1996).



40TFLOPSのスーパーコンピュータにおける1,000 残基タンパク質 (100,000 軌道) 全電子計算1回転分の時間予測.



計算ルーチン	分子	100,000 軌道
全エネルギー計算		1,646
Kohn-Sham 行列生成		1,540
電子密度フィッティング		1,270
交換相関ポテンシャルフィッティング		76
行列対角化		8,804
行列積		1,926
合計		15,262