Development of the new quantum chemical calculation system for proteins based on the density functional method program ProteinDF

F. Sato, N. Ihara, T. Ueno, S. Tahara, N. Tsunekawa, Y. Nishimura, N. Nishino, T. Hirano, T. Yoshihiro, Y. Nishikawa*, T. Inaba*, Y. Nagamine*, T. Nishimura*, S. Koike*, H. Koike*, H. Kashiwagi

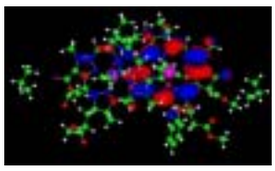
Institute of Industrial Science, University of Tokyo, *AdvanceSoft Corporation

We are developing a system for quantum chemical calculation of proteins based on the gaussian-type orbitals density functional method program ProteinDF. It can treat the whole protein as a molecule and calculate the all-electron taking into account the electron correlation effect. This system will become a valuable tool for the study of biotechnology in the post-genomic age. The system consists of 5 sub-systems: (1) an integrated GUI, (2) automatic simulation, (3) *ab initio* molecular dynamics and geometry optimization, (4) a module for supercomputers, and (5) a protein wavefunction database.

Why are quantum chemical calculations of proteins required?

Proteins are huge molecules consisted of many amino acids, and about 1/3 amount of them contain metal ions. In addition, proteins function within a biological environment (~300K). This means that the analysis of protein function requires at least 0.1eV energy accuracy. However, the most of conventional simulations of protein are classical ones using database force fields, which assign the same parameters to all the same amino acids in proteins, so that they may not provide sufficient accuracy.

Ionization potentials of	Fe(II) ion	~700 *kcal/mol*
	heme	~100
	hemoprotein	< 10



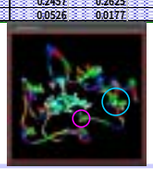
The 3D graphics of HOMO of special pair model.

ProteinDF is

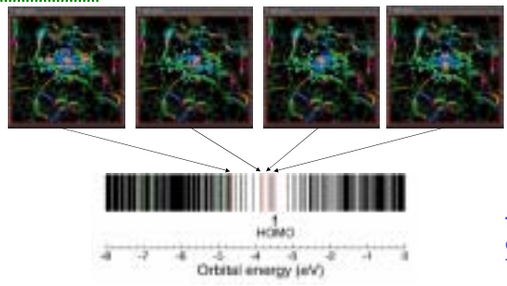
- A gaussian-type orbitals density functional method program which can attain all-electron canonical wavefunction of proteins,
- Suited for huge molecular systems, especially for metal-containing proteins,
- Coded by the object-oriented language C++.

Mulliken charge of ALAs in cytochrome *c* by ProteinDF.
The left column is the RESP charge of ALA in AMBER99.

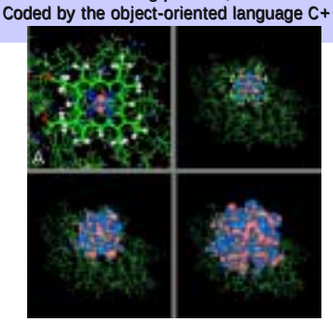
	ALA15	ALA43	ALA51	ALA83	ALA96	ALA101	ALA AMBER99
N	-0.4820	-0.5236	-0.4759	-0.5203	-0.4889	-0.4735	-0.4157
Cα	-0.0941	-0.1405	-0.1075	-0.0903	-0.0955	-0.1010	0.0337
C	0.2254	0.2771	0.2057	0.2198	0.2409	0.2497	0.5973
O	-0.3562	-0.3706	-0.3363	-0.3673	-0.3673	-0.3887	-0.5679
Cβ	-0.6150	-0.5116	-0.5652	-0.5912	-0.6137	-0.6123	-0.1825
H	0.4107	0.4350	0.3329	0.3799	0.4018	0.4016	0.2719
Hα	0.2457	0.1971	0.2642	0.2405	0.2448	0.2400	0.0823
Hβ1	0.2590	0.2132	0.2573	0.2478	0.2275	0.2453	0.0603
Hβ2	0.2134	0.1791	0.2411	0.2101	0.2755	0.2404	0.0603
Hβ3	0.2457	0.2625	0.2179	0.2294	0.2514	0.2227	0.0603
Total	0.0526	0.0177	0.0342	-0.0416	0.0765	0.0241	0.0000



Positions of ALAs in cytochrome *c*.



Distribution of Kohn-Sham orbital energy of cytochrome *c*.

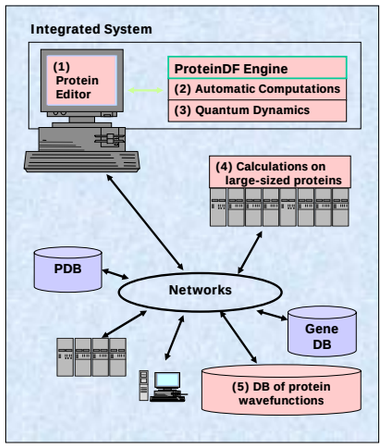


The 3D graphics of the 3293-th MO (HOMO) of d^6-low-spin ferrocytochrome *c*.
The isosurface values are ±0.05 (above left), ±0.005 (above right), ±0.0005 (below left) and ±0.00005 (below right), respectively. The scale of (A) is magnified twice.

This is the first time all-electron calculation of metal-containing protein (9,600 orbitals) in 2000.

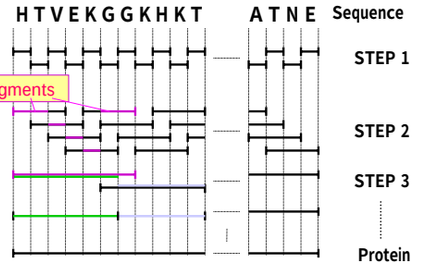
ProteinDF system

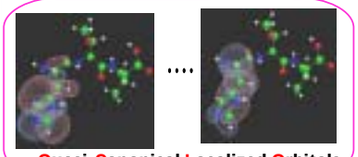
We are developing a new quantum chemical software for proteins based on ProteinDF. It will support the semi-automatic computations on fixed coordinate all-electron calculation, *ab initio* molecular dynamics and geometry optimization of proteins. We prepare the integrated environment, through which we can easily operate all of simulations. We are also constructing the database of the protein wavefunctions.



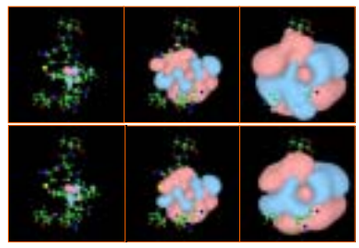
Topics I; Automatic Computation & QCLO

QCLO is localized in a fragment, but it is also the canonical MO of the fragment. Computational process with QCLOs (right) improves the SCF convergence.





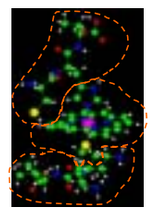
Quasi-Canonical Localized Orbitals



Application of QCLO for the active center including heme.

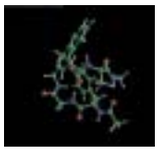
1 fragment
= usual canonical MO

3 fragments QCLO
(right) definition of fragments



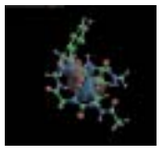
The 3D graphics of the d_{xz} orbitals in active center model.
From the left, the isosurface values are ±0.05, ±0.005 and ±0.0005, respectively.

Topics II; Comparison with classical method



AMBER99

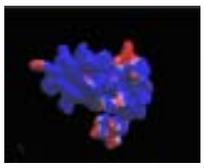
Φ_2 = -165.2, Ψ_2 = 55.2
Φ_3 = 82.3, Ψ_3 = -41.9



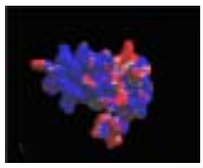
ProteinDF

Φ_2 = -161.4, Ψ_2 = 37.1
Φ_3 = 103.7, Ψ_3 = -30.9

Optimized Geometry of TYR-GLY-GLY-PHE-ALA peptide.



AMBER99

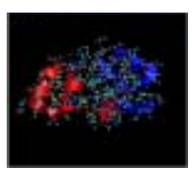


ProteinDF

Electrostatic potential of Insulin (51 residues, 4,439 orbitals).

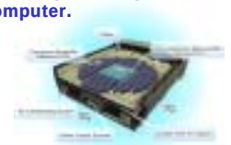
Topics III; Computational time & Future view

Computational time of Insulin dimer.
•102 residues (8,878 orbitals)
•Altix3700 (256GFlops)
55 min./SCF
(1.5 days for convergence)



Future Dream;
Electron transfer from cytochrome *c* to oxidase.
Application of QCLO and DFT/CI. The right figure is the virtual composition of F.Sato et al. *Chem.Phys.Lett.* **341**, 645-651 (2001) and T.Tsukihara et al. *Science* **272**, 1136-1144 (1996).

Predicted elapse time of 1SCF for 100,000 orbitals proteins (1,000 residues) by 40TFlops computer.



Calculation Step \ Molecule	100,000 orbitals estimated elapse time (sec)
Total Energy Calculation	1,646
Generation of Kohn-Sham Matrix	1,540
Fitting of Electron Density	1,270
Fitting of Exchange Correlation Potential	76
Matrix Diagonalization	8,804
Matrix Multiplication	1,926
Total	15,262

Frontier Simulation Software for Industrial Science, 2004
http://www.fsis.iis.u-tokyo.ac.jp