FMO-MO Method as an Initial Guess Generation for SCF Calculation: Case of (−)-Epicatechin Gallate

Katsuhiro Tamura,[1,2] Toshio Watanabe,[3]
Takayoshi Ishimoto,[3] Hiroaki Umeda,[3]
Yuichi Inadomi,[4] and Umpei Nagashima*[1,3]

[1]Department of Chemistry, University of Tsukuba,
1-1-1 Tennodai, Tsukuba 305-8571

[2]Shizuoka Industrial Research Institute of Shizuoka
Prefecture, 2078 Makigaya, Aoi-ku, Shizuoka 421-1298

[3]Research Institute for Computational Sciences,
National Institute of Advanced Industrial Science
and Technology, CREST-JST, 1-1-1 Umezono,
Tsukuba 305-8568

[4]Research Institute for Information Technology,
Kyushu University, 3-8-33-710 Momochihama,
Sawara-ku, Fukuoka 814-0001

Received July 30, 2007; E-mail: u.nagashima@aist.go.jp

FMO-MO (Molecular orbitals (MOs) of the fragment MO (FMO) method) for (−)-epicatechin gallate were calculated to examine the dependency of the accuracy on the fragmentation patterns and on the basis set, and the results were compared to those obtained using conventional MO calculations. Results also show that the FMO-MO can be used as the initial guess in self-consistent field calculations.

The fragment molecular orbital (FMO) method[1–5] has been proposed as a method for calculating the electronic structure of large systems, such as proteins,[6] DNA,[7–9] and molecular clusters. The FMO method divides a molecule into fragments, and then, the molecular orbitals (MOs) of the fragments and the fragment pairs are calculated to obtain the total energy and other properties of the molecule. The MOs for an entire molecule can also be obtained using the FMO method,[10] although more calculation time is required.

The conventional MO method for large systems requires a good initial guess, which is sufficiently near the self consistent field (SCF) solution, for SCF convergence. A rough initial guess requires a large number of SCF iterations and, especially for large systems, has an extremely tenacious divergence problem. A quasi-canonical localized orbitals method[11] (QCLO), in which the initial guess is determined from the subsystem MO calculations, has been developed and applied to all-electron density functional calculations for insulin.[12] Although the electron density calculated using the FMO-MO method reportedly can be used as a good initial guess for SCF calculation,[13] detailed analysis of the dependency of the SCF convergence on the accuracy of FMO-MO calculations has not been done yet.

In the FMO method, selection of the fragmentation position of a molecule and the distribution pattern of two electrons at a breaking bond is crucial for performing highly accurate electronic structure calculations. The accuracy of the FMO method has only been confirmed for fragmentation at a sp^3 C–C bond. In particular, the accuracy for fragmentation at the Cα–C* bond of a protein and enzyme and at the C4′–C5′ bond of deoxyribose in DNA has been extensively documented, because applications of the FMO method have mainly focused on proteins and DNA. Recently, Ishikawa et al. have reported calculations on bond-breaking at a Si atom using the FMO method and have shown the accuracy of the fragmentation at Si bonds.[14] We have previously reported the fragmentation pattern dependence of the total energy and atomic charge difference of (−)-epicatechin gallate ((−)-ECg)[15] and have found that an optimal pattern exists for individual molecules, for which results comparable to conventional MO calculations can be achieved. Also, the difference in the potential energy surfaces between the FMO method and the conventional ab initio MO method has been investigated using the optimal fragmentation pattern of (−)-ECg,[16] and the FMO method has been found to be applicable to quantum molecular dynamics (QMD) calculations using the optimal fragmentation pattern.[17]

In this study, FMO and FMO-MO calculations with four basis sets were done using three fragmentation patterns of (−)-ECg to examine the dependency of these calculations on the fragmentation pattern and on the basis set. Then, SCF calculations were done using the FMO-MO density as an initial guess to examine the convergence rate.

Computational Method

The FMO method divides a molecule into fragments, and then, the MOs of fragments and fragment pairs are calculated to obtain the fragment and the fragment pair energies and electron densities (denoted as $\{E_I\}$, $\{E_{IJ}\}$, $\{\mathbf{D}_I\}$, and $\{\mathbf{D}_{IJ}\}$, respectively). The FMO energy, E^{FMO}, and density, $\mathbf{D}^{\mathrm{FMO}}$, are obtained by using the following equations:

$$E^{\mathrm{FMO}} = -(n-2)\sum_I E_I + \sum_I \sum_{J<I} E_{IJ}, \tag{1}$$

$$\mathbf{D}^{\mathrm{FMO}} = -(n-2)\sum_I \mathbf{D}_I + \sum_I \sum_{J<I} \mathbf{D}_{IJ}. \tag{2}$$

The Fock matrix of the entire molecule can be calculated from $\mathbf{D}^{\mathrm{FMO}}$, and the MOs in the FMO method (FMO-MO) and the electron density in FMO-MO, $\mathbf{D}^{\mathrm{FMO\text{-}MO}}$, are calculated using a generalized diagonalization of this Fock matrix.[10] The energy calculated from $\mathbf{D}^{\mathrm{FMO\text{-}MO}}$, hereafter denoted $E(\mathbf{D}^{\mathrm{FMO\text{-}MO}})$, reportedly is a more accurate approximation of the energy and density than E^{FMO} and $E(\mathbf{D}^{\mathrm{FMO}})$, although the computational cost to generate a Fock matrix is higher than that used in the FMO method.[10]

To examine the suitability of $\mathbf{D}^{\mathrm{FMO\text{-}MO}}$ as an initial guess in SCF calculations, the SCF calculation was started from an initial guess of $\mathbf{D}^{\mathrm{FMO\text{-}MO}}$. The energy calculated by the first SCF iteration is equal to $E(\mathbf{D}^{\mathrm{FMO\text{-}MO}})$. We defined the energies that reached the SCF convergent threshold of 1.0×10^{-10} and 1.0×10^{-7} as the exact energy and the convergent energy, respectively. The energy difference (ΔE) was calculated from the exact energy as a standard. For SCF iterative calculation,

Published on the web February 12, 2008; doi:10.1246/bcsj.81.254

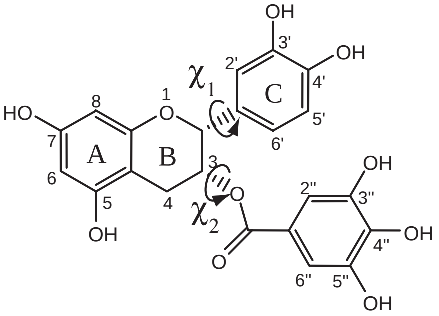

Fig. 1. Molecular structure of (−)-ECg with $(\chi_1, \chi_2) = (0, 0)$.

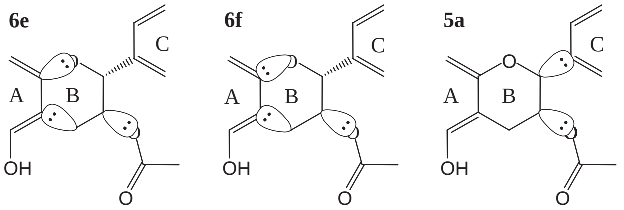

Fig. 2. Molecular fragmentation patterns and bonding-electron distribution.

acceleration procedures are commonly used.[18,19] In our study, however, we did not use any acceleration procedure, because our goal was to analyze the differences due to different initial guesses of $\mathbf{D}^{\text{FMO-MO}}$.

The geometry of the (−)-ECg used in this work was the geometry used in a previous study to determine the FMO fragmentation patterns at the HF/STO-3G.[15] Figure 1 shows the molecular structure of (−)-ECg. We selected three fragment patterns for the (−)-ECg calculation (see Fig. 2), because these fragmentation patterns accurately represented the total energy and electronic population obtained using conventional MO at the HF/STO-3G. To examine the basis set dependency, we used four basis sets: the STO-3G,[20] which has been used to determine the fragmentation in Ref. 14, and the 6-31G, 6-31G*, and 6-31G**. All parameters for fragmentations at the O–C and sp^2-hybridized C–X bonds were the same as those used for fragmentation at the Cα–C* bond.

All FMO calculations were performed using ABINIT-MP,[21] and the FMO-MO calculations and SCF calculations were performed using a program[22] written by Inadomi et al.

Results and Discussion

Table 1 lists the FMO and FMO-MO calculation results. E^{FMO} with the **6f** and **6e** fragmentation patterns using the STO-3G basis set were the most accurate values ($\Delta E^{\text{FMO}} =$ 0.87 and 0.94 kcal mol^{-1}, respectively) and are similar to the values reported in Ref. 14. For the series of 6-31G basis sets (6-31G, 6-31G*, and 6-31G**), ΔE^{FMO} was five times larger than that of the STO-3G basis set. The relative order of ΔE^{FMO} calculated using the same basis set had the same trend: **6f** ≈ **6e** < **5a**. These results indicate that the optimal fragmentation pattern for the STO-3G basis set is not the optimal pattern for the other basis sets, although the relative accuracy of the fragmentation patterns within the same basis set has similar trends.

Table 1. Energy Differences of FMO, $\mathbf{D}^{\text{FMO}}$, and $\mathbf{D}^{\text{FMO-MO}}$ (in kcal mol^{-1}) and the Number of Iterations for Convergence of Each Model Calculation

		ΔE^{FMO}	$\Delta E(\mathbf{D}^{\text{FMO}})$	$\Delta E(\mathbf{D}^{\text{FMO-MO}})$	$N_{\text{iteration}}$[a)]
STO-3G	**5a**	1.769	3.318	0.005	3
	6e	0.940	1.748	0.004	4
	6f	0.873	3.071	0.006	4
6-31G	**5a**	7.761	6.503	0.097	13
	6e	5.018	3.141	0.053	7
	6f	4.281	2.338	0.052	10
6-31G*	**5a**	7.698	6.144	0.113	8
	6e	5.135	3.328	0.055	7
	6f	4.558	2.719	0.055	9
6-31G**	**5a**	7.344	5.811	0.117	9
	6e	4.839	3.129	0.054	7
	6f	4.247	2.484	0.054	10

a) Criteria for convergence was 10^{-7} au.

For a given basis set and fragmentation pattern, $E(\mathbf{D}^{\text{FMO-MO}})$ was more accurate than E^{FMO}, whereas the accuracy of $E(\mathbf{D}^{\text{FMO}})$ was on the same order as E^{FMO}. The accuracy of $E(\mathbf{D}^{\text{FMO-MO}})$ using the STO-3G basis set was high: <0.006 kcal mol^{-1}. Furthermore, the accuracy of E^{FMO} strongly correlated with the accuracy of $E(\mathbf{D}^{\text{FMO-MO}})$. Ratios of $\Delta E(\mathbf{D}^{\text{FMO-MO}})$ to ΔE^{FMO} were 1:137–383 for the STO-3G basis set and 1:63–95 for the series of 6-31G basis sets. The FMO calculations based on these fragmentation patterns using the STO-3G basis set well reproduced not only the energies but also electron populations,[15] and therefore, the FMO calculation should be suitable for use in the FMO-MO calculation. The lower flexibility of the STO-3G basis set compared with that of the series of 6-31G basis sets might be another reason for the accuracy of the STO-3G results in the FMO-MO calculation.

Figure 3 shows ΔE during SCF iteration starting from $\mathbf{D}^{\text{FMO-MO}}$. The SCF calculations using the STO-3G basis set quickly converged within 4 SCF iterations. This rapid convergence is due to the accuracy of $\mathbf{D}^{\text{FMO-MO}}$ as initial guess. All SCF calculations with **6e** using the series of the 6-31G basis sets converged within 7 SCF iterations. These results confirm the effectiveness of $\mathbf{D}^{\text{FMO-MO}}$ as a good initial guess in SCF calculations.

In the SCF calculations for each basis set, except STO-3G, the fragmentation pattern of **6e** converged more rapidly compared with those of **5a** and **6f**. The number of iterations strongly depended on the fragmentation pattern, and there was no correlation between the number of iterations and accuracy of either E^{FMO}, $E(\mathbf{D}^{\text{FMO}})$, or $E(\mathbf{D}^{\text{FMO-MO}})$. The SCF calculations on the fragmentation patterns of **5a** and **6f** using the 6-31G basis set converged more slowly compared with the calculations using the more flexible basis sets 6-31G* and 6-31G**. These slower calculations might be accelerated by using a convergence acceleration procedure, such as the direct inversion in the iterative subspace (DIIS).[18]

DIIS accelerates the convergence near the convergence region, whereas it often needs another acceleration procedure in regions far from the convergence.[19] The initial guess of $\mathbf{D}^{\text{FMO-MO}}$ provides the SCF calculation with a starting point near the convergence point. Therefore, convergence of the SCF calculations might be accelerated by the DIIS procedure, although we did not examine these calculations in this study.

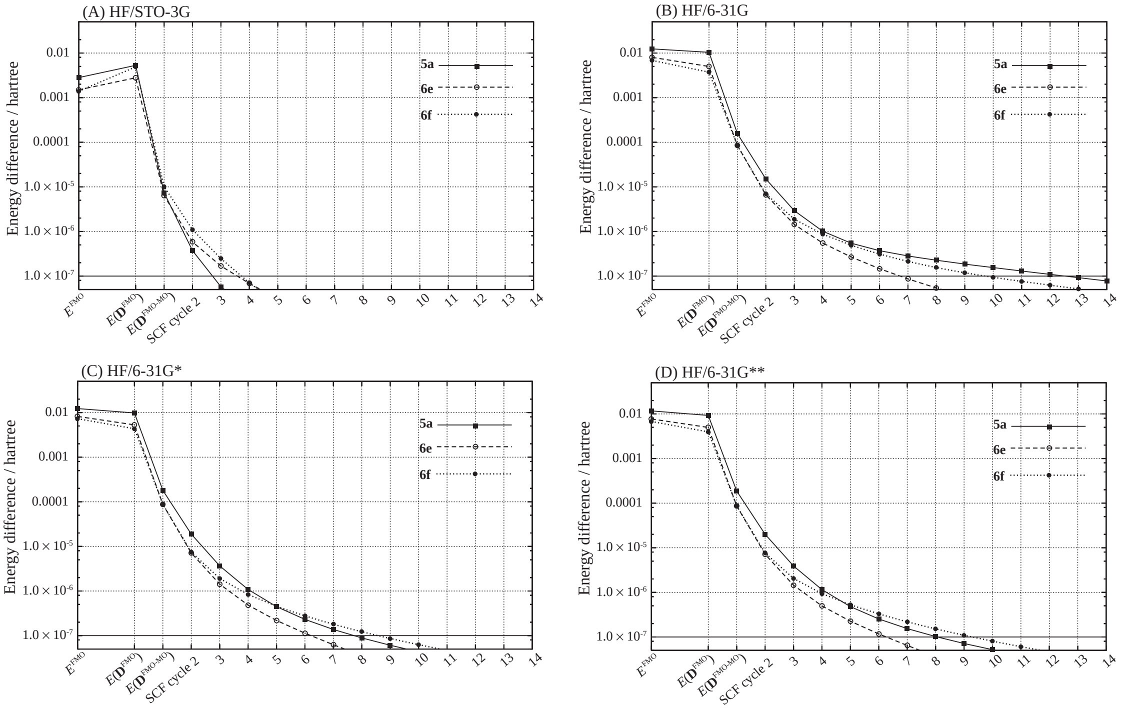

Fig. 3. Energy differences during SCF iteration: (A) HF/STO-3G, (B) HF/6-31G, (C) HF/6-31G*, and (D) HF/6-31G**.

Conclusion

FMO and FMO-MO calculations with four basis sets using three fragmentation patterns of (−)-ECg were carried out. The optimal fragmentation pattern for the STO-3G basis set did not yield an accurate FMO energy using the series of 6-31G basis sets, indicating that the FMO fragmentation pattern depends on the basis set. $\Delta E(\mathbf{D}^{\text{FMO-MO}})$ yielded a difference as small as 1% of the ΔE^{FMO}. The SCF calculations with an initial guess of $\mathbf{D}^{\text{FMO-MO}}$ showed rapid convergence, especially when the STO-3G basis set was used. These results demonstrate the importance of an optimized fragmentation pattern in each basis set, and the effectiveness of using the FMO-MO method to accurately calculate the energy and density, which is then used as a good initial guess for SCF calculations.

This work was supported by the research project, "Development of MO calculation system for large molecule on Grid" of CREST, Japan Science and Technology Agency.

References

1 K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701.

2 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2000**, *318*, 614.

3 K. Kitaura, S. Sugiki, T. Nakano, Y. Komeiji, M. Uebayasi, *Chem. Phys. Lett.* **2001**, *336*, 163.

4 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2002**, *351*, 475.

5 D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.

6 K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura, T. Nakano, *J. Phys. Chem. B* **2006**, *110*, 16102.

7 S. Sugiki, M. Matsuoka, R. Usuki, Y. Sengoku, N. Kurita, H. Sekino, S. Tanaka, *J. Theor. Comput. Chem.* **2005**, *4*, 183.

8 K. Fukuzawa, Y. Komeiji, Y. Mochizuki, A. Kato, T. Nakano, S. Tanaka, *J. Comput. Chem.* **2006**, *27*, 948.

9 T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, U. Nagashima, *J. Phys. Chem. B* **2007**, *111*, 9621.

10 Y. Inadomi, T. Nakano, K. Kitaura, U. Nagashima, *Chem. Phys. Lett.* **2002**, *364*, 139.

11 H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, *Mol. Phys.* **2003**, *101*, 81.

12 T. Inaba, S. Tahara, N. Nishikawa, H. Kashiwagi, F. Sato, *J. Comput. Chem.* **2005**, *26*, 987.

13 Y. Inadomi, H. Umeda, T. Watanabe, U. Nagashima, T. Ishimoto, American Chemical Society National Meeting 232, Atlanta, GA, March, **2006**.

14 T. Ishikawa, Y. Mochizuki, K. Imamura, T. Nakano, H. Mori, H. Tokiwa, K. Tanaka, E. Miyoshi, S. Tanaka, *Chem. Phys. Lett.* **2006**, *430*, 361.

15 K. Tamura, Y. Inadomi, U. Nagashima, *Bull. Chem. Soc. Jpn.* **2007**, *80*, 721.

16 K. Tamura, T. Watanabe, T. Ishimoto, U. Nagashima, *Bull. Chem. Soc. Jpn.* **2007**, *80*, 1939.

17 K. Tamura, T. Watanabe, T. Ishimoto, U. Nagashima, *Bull. Chem. Soc. Jpn.* **2008**, *81*, 110.

18 P. Pulay, *J. Comput. Chem.* **1982**, *3*, 556.

19 K. N. Kudin, G. E. Scuseria, E. Cancès, *J. Chem. Phys.* **2002**, *116*, 8255.

20 W. J. Hehre, R. F. Stewart, J. A. Pople, *J. Chem. Phys.* **1969**, *51*, 2657.

21 ABINIT-MP, NIHS, available from http://moldb.nihs.go.jp/abinitmp.

22 Y. Inadomi, T. Nakano, K. Kitaura, U. Nagashima, HPC and Networking, 9th International Conference, HPCN Europe 2001, Amsterdam, The Netherlands, June, **2001**.