Chem-Bio Informatics Journal, Vol. 10, pp.24-31 (2010)

Acceleration of monomer self-consistent charge process in fragment molecular orbital method

Takeshi Ishikawa[1,2] and Kazuo Kuwata[1,2,*]

[1]*Division of Prion Research, Center for Emerging Infectious Disease Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan.*

[2]*CREST Project, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*

**E-mail: kuwata@gifu-u.ac.jp*

(Received December 2, 2009; accepted February 15, 2010; published online February 25, 2010)

Abstract

We introduced the dynamic update technique into the monomer self-consistent charge (SCC) process of the fragment molecular orbital (FMO) method to reduce its computational costs. This technique has already been used for solving linear equations in some quantum chemical calculations. After performing test calculations on three typical polyglycines (GLY_{20}, GLY_{40}, and GLY_{60}), we further performed the FMO calculations on the human immunodeficiency virus type 1 protease complexed with lopinavir using the dynamic update technique. These calculations demonstrate that the computational time of the monomer SCC process can be reduced by about one-third. Furthermore, we examined the dependence of the iteration number of the monomer SCC process on parallelization schemes.

Key Words: Fragment molecular orbital method, monomer self-consistent charge process, dynamic update, HIV-1 protease

Area of Interest: Molecular Computing

1. Introduction

In recent years, *ab initio* quantum chemical methods have been applied to large molecules including biomolecular systems. The fragment molecular orbital (FMO) method [1][2][3], which was developed by Kitaura et al., is an efficient approach for such calculations. In this method, target molecules are divided into small fragments and only calculations of the fragments (referred to as monomers) and pairs of the fragments (referred to as dimers) are needed for evaluating total properties. This approach enables us to reduce the computational effort for the quantum chemical calculations of biomolecular systems. However, in the case of large proteins, a high cost is required even if we use the FMO method. Moreover, a recent study reported that the average results from a large number of sampling structures should be used for the examinations of biomolecular systems [4][5]. For such calculations, further enhancements in the computational speed are necessary.

Recently, Ishikawa and Kuwata introduced the resolution of the identity (RI) approximation into the FMO scheme to reduce the computational cost of MP2 [6], the most time-consuming process of FMO calculations. They demonstrated that the RI-MP2 method achieved an approximately 10 fold increase in the speed of the MP2 process in the FMO method. Meanwhile, other processes such as monomer self-consistent charge (monomer SCC) process, calculations with dimer electrostatic approximation, and dimer HF calculations consumed significantly more time. Thus, it is imperative that the computational effort of these processes be reduced.

In this article, we report that computational cost of the monomer SCC procedure can be reduced by the dynamic update technique, that has already been used for solving linear equations in some quantum chemical calculations [7][8]. In the following sections, we discuss the methodological aspects and provide examples of a real biomolecular system. All calculations were performed using our original program PAICS [4].

2. Method

As mentioned above, in the FMO scheme, target systems are divided into small N_f fragments by cutting C-C single bonds and the total properties are evaluated with monomer and dimer calculations. For example, the total energy at HF level can be obtained by the following equation [1][2][3]:

$$E^{HF}_{total} = \sum_{I<J} E^{HF}_{IJ} - \left(N_f - 2\right) \sum_{I} E^{HF}_{I} \ , \tag{1}$$

where E^{HF}_{I} and E^{HF}_{IJ} are the energies of the monomer and the dimer, respectively. These values are calculated with a modified Fock operator [9]:

$$f_X(\boldsymbol{r}) = \tilde{f}_X(\boldsymbol{r}) + \sum_{K \neq X} \{u_K(\boldsymbol{r}) + v_K(\boldsymbol{r})\} + \sum_k B_k |\theta_k\rangle\langle\theta_k| \,, \tag{2}$$

where $\tilde{f}_X(\boldsymbol{r})$ is a conventional Fock operator, $u_K(\boldsymbol{r})$ and $v_K(\boldsymbol{r})$ are the electrostatic potential from *K*-th monomer (nuclear attraction and electron repulsion, respectively), and the third term is the projection operator for cutting C-C single bonds. In Eq. (2), *X* is replaced by *I* and *IJ* in the monomer and dimer calculations, respectively. The electron repulsion potential, $v_K(\boldsymbol{r})$, should be determined from the electron distribution of *K*-th monomer. In the FMO scheme, such electron distributions are obtained by an iterative procedure called monomer SCC process.

In the first iteration of the monomer SCC process, an initial electron distribution for each monomer is calculated in vacuum. In the next iteration, a new electron distribution for each monomer is recomputed including the electrostatic potential from the monomer whose electron distribution has already been obtained from the previous iteration. This iterative process is continued until self-consistency is achieved.

The expression of the total energy in Eq. (1) can be rewritten into the following equation [10]:

$$E^{HF}_{total} = \sum_I E'^{HF}_I + \sum_{I<J} \Delta E^{HF}_{IJ} \,, \tag{3}$$

where E'^{HF}_I is the monomer energy without electrostatic potential from the other monomers and ΔE^{HF}_{IJ} is the inter fragment interaction energy (theoretical details are found in the previous paper [10]). The first term of this equation is used for estimating convergence of the monomer SCC process in our program.

Clearly, the computational time of monomer SCC procedure depends on the number of iterations. In this paper, we introduce the dynamic update technique to reduce the computational time. The essential scheme of the monomer SCC process using the dynamic update is similar to the normal one except for the manner in which the electrostatic potential is determined. In the normal process, electrostatic potential from all other monomers are determined using the electron distribution obtained from the previous iteration. On the other hand, when using the dynamic update, the electrostatic potential from the monomer, whose calculation has been already completed in the present iteration, is determined with the electron distribution obtained in the present iteration. That is, the monomer electron distribution obtained in an iteration is immediately used for the subsequent calculations in this iteration. By this treatment, the number of iterations of the monomer SCC procedure can be reduced for typical biomolecular systems.

Our program is parallelized with the message-passing interface (MPI) [11]. A feature of the dynamic update is that the number of iterations depends on the parallelization scheme used for the FMO calculations. To explain the reasons for this dependence, we considered two types of calculations with different parallelization schemes. One is the non-parallelized calculation and the other is the parallelized calculation with eight CPUs. Figure 1 shows flow charts of the calculations performed in an iteration of the monomer SCC process. In the case of non-parallelized calculation,

for example, fragments 1-8 have already been calculated by CPU1 when the computation on fragment 9 starts. Thus, this calculation is performed with the new electron distributions of fragments 1-8 (the electrostatic potential from other fragments is obtained by the electron distributions of the previous iteration). On the other hand, in the parallelized case, only the calculation of fragment 1 has been completed by CPU1 when the computation on fragment 9 starts. Thus, only the new electron distribution of fragment 1 can be used for the calculation of fragment 9. Consequently, the convergence process of the parallelized calculation is different from that of the non-parallelized calculation. In general, the greater the increase in parallelization numbers, the greater is the increase in iteration numbers. As discussed in the following section, we can control the iteration number of the monomer SCC process by adjusting the parallelization number for the index of the fragment or fragment pair.

3. Results and Discussion

First, we show the effect of the dynamic update with three polyglycines forming a typical α-helix (GLY_{20}, GLY_{40}, and GLY_{60}), whose atomic coordinates were prepared by MD simulations using the AMBER 10 package [12]. In these calculations, each amino acid residue was treated as a single fragment and cc-pVDZ [13] basis set was employed. We used 10^{-5} as a threshold value for estimating convergence of the iterative procedure. The results are summarized in Table 1. As shown in this table, the number of iterations of the monomer SCC process can be reduced by about two-thirds with the dynamic update for the three polyglycines. Because the monomer electron distributions obtained with the dynamic update are slightly different from those with the normal scheme within the threshold value, the final total energies (E^{HF}_{total}) are also slightly different.

We then demonstrated the efficiency of the dynamic update in typical biomolecular systems by performing FMO computations for the human immunodeficiency virus type 1 protease (HIV-1 PR) complexed with lopinavir using cc-pVDZ basis set [13]. This protein, a protease by function, plays a critical role in patients afflicted by Acquired Immune Deficiency Syndrome (AIDS), and lopinavir is a protease inhibitor that is used as a drug to treat patients infected with HIV. Detailed description about the calculated system, e.g., preparation of the atomic coordinates, can be found in a previously published paper [8]. We performed several calculations with different parallel conditions to examine the dependence of the iteration number on them. The results of calculations are summarized in Table 2. All calculations were performed with eight cores and each parallel condition is denoted by values of "np," which is the number of parallelization for fragment or fragment pair index. For example, in the case of np = 2, each monomer or dimer calculation is performed with two cores, and thus, the four individual calculations progress independently at the same time.

The number of iterations of the monomer SCC process with the normal scheme was 52, while those for the dynamic update were 43, 33, 31, and 18 for np = 1, 2, 4, and 8, respectively. As a result, the computational time for the monomer SCC process could be reduced. In particular, in the case of np = 8, the computational time was reduced to about one-third (1619 minutes for the normal

scheme and 556 minutes for the dynamic update).

The number of iterations increased as the np decreased. Thus, the dynamic update should be used together with a larger number of np. However, Table 2 shows that the computational time of dimer calculations increased by employing a larger np. Therefore, an appropriate condition of parallelization is that we use a large np in the monomer SCC calculation and a small np in the dimer calculations. Using such a parallelization condition, we could complete the FMO calculation in 3346 minutes (see Table 2), which was 76.5% of the total time of calculation without the dynamic update (4371 minutes).

4. Summary

In this paper, we introduced the dynamic update technique to reduce the computational effort of the monomer SCC process, which is among the most time-consuming processes of FMO calculations. First, we tested the dynamic update with typical polyglycines forming α-helix (GLY_{20}, GLY_{40}, and GLY_{60}), which showed that the dynamic update could reduce the number of iterations of the monomer SCC process. Further, we performed the FMO computations of the HIV-1 PR complexed with lopinavir to demonstrate the efficiency of the dynamic update for real biomolecular systems. These calculations showed that the iteration number and computational time of the monomer SCC process could be reduced by about one-third. As a result, the total time of the FMO computation was reduced to 76.5% of that without the dynamic update. In case of the monomer SCC calculations with the dynamic update, the iteration number depends on the parallelization condition. This dependence was examined and the most appropriate condition of parallelization was shown.

This study was supported by a Grant-In-Aid for Young Scientist (Start-up) No. 20850020 from the Japan Society for the Promotion of Science as well as by the Program for Promotion of Fundamental Studies in Health Science of the National Institute of Biomedical Innovation. Calculations necessary in this study were performed using the computational resources at Gifu University.

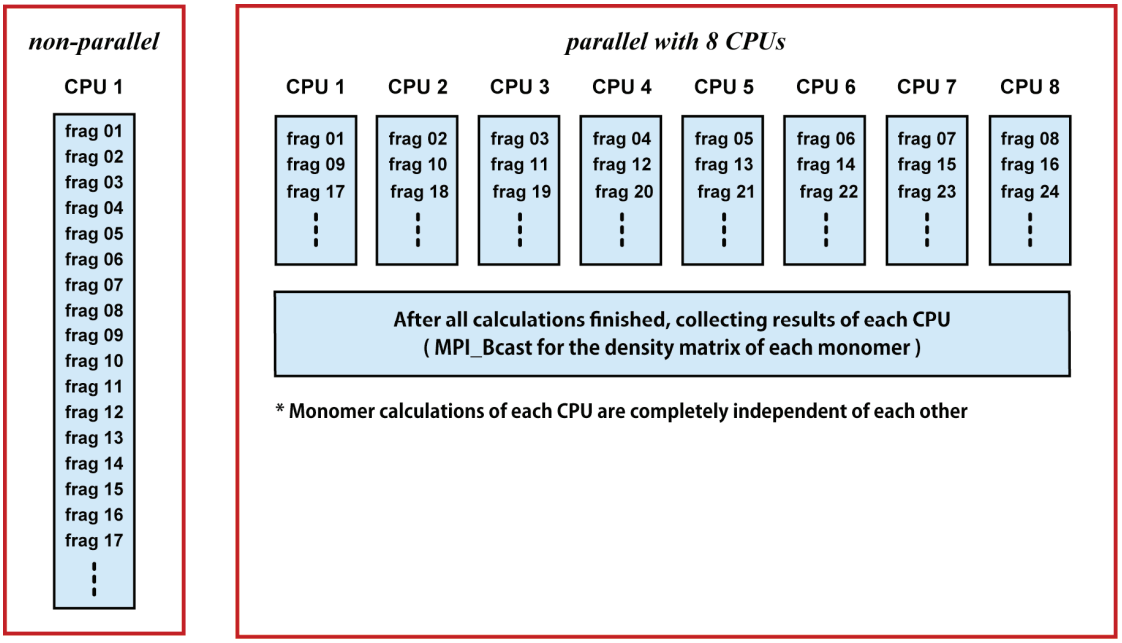

Figure 1. Flow charts of calculations performed in an iteration of monomer SCC process for non-parallelized and parallelized calculations. In the parallelized case, the calculations in each CPU are completely independent of each other until all the monomer calculations have been completed. At the end of each iteration, new electron distributions of the monomers (i.e., new density matrices of the monomers) are shared among all CPUs with an MPI function.

Table 1. Number of iterations of monomer SCC process and total energies of the system for three polyglycines. Total energies are shown in hartree. These calculations were performed with cc-pVDZ.

		iter.	E^{HF}_{total}
GLY_{20}	normal	21	−4212.388080
	dynamic update	14	−4212.388080
GLY_{40}	normal	21	−8349.128939
	dynamic update	14	−8349.128938
GLY_{60}	normal	22	−12485.888805
	dynamic update	14	−12485.888804

Table 2. Number of iterations of monomer SCC process, total energies of the system, and computational times of each process for the HIV-1 PR with lopinavir molecule. Total energies and computational times are shown in hartree and minute, respectively. In these calculations, cc-pVDZ basis set was employed. All calculations were performed with 8 cores (Xeon E5429 and 2.0 GB memory per core). Results of calculations with different parallelization scheme are listed. The "np" represents the number of cores assigned to one fragment or fragment pair calculation (see text).

			time		
	iter.	E^{HF}_{total}	*scc*	*dimer*	*total*
not dynamic update [(a)]	52	−77629.372214	1619	2132	4371
np = 1	43	−77629.372214	1121	2234	3907
np = 2	33	−77629.372215	920	2374	3830
np = 4	31	−77629.372215	857	2478	3889
np = 8	18	−77629.372215	556	2787	3899
np = 8, np = 1 [(b)]	18	−77629.372214	557	2233	3346

[(a)] np = 1 was used.

[(b)] np = 8 and np = 1 were used for monomer SCC and dimer calculations, respectively.

References

[1] K. Kitaura, T. Sawai, T. Asada, T. Nakano, and M. Uebayasi, Pair interaction molecular orbital method: an approximate computational method for molecular interactions, *Chem. Phys. Lett.*, **312**, 319-324 (1999).

[2] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Fragment molecular orbital method: an approximate computational method for large molecules, *Chem. Phys. Lett.*, **313**, 701-706 (1999).

[3] D. G. Fedorov and K. Kitaura, Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method, *J. Phys. Chem. A*, **111**, 6904-6914 (2007).

[4] T. Ishikawa, T. Ishikura, and K. Kuwata, Theoretical Study of the Prion ProteinBased on the Fragment Molecular Orbital Method, *J. Comput. Chem.*, **30**, 2594-2601 (2009).

[5] T. Ishikawa and K. Kuwata, Interaction analysis of the native structure of prion protein with quantum chemical calculations, *J. Chem. Theor. Comp.*, **6**, 538-547(2010).

[6] T. Ishikawa and K. Kuwata, Fragment molecular orbital calculation using theRI-MP2 method, *Chem. Phys. Lett.*, **474**, 195-198 (2009).

[7] G. Rauhut, P. Pulay, and Werner H-J, Integral Transformation with Low-Order Scaling for Large Local Second-Order Møller-Plesset Calculations, *J. Comput. Chem.*, **19**, 1241-1254 (1998).

[8] T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, H. Tokiwa, S. Tanaka, and K. Tanaka, Fragment interaction analysis based on local MP2, *Theor. Chem. Acc.*, **118**, 937-945 (2007).

[9] K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, Fragment molecular orbital method: analytical energy gradients, *Chem. Phys. Lett.*, **336**, 163-170 (2001).

[10] T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Fragment molecular orbital method: use of approximate electrostatic potential, *Chem. Phys. Lett.*, **351**, 475-480 (2002).

[11] Message Passing Interface (MPI) http://www.mpi-forum.org/.

[12] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossv_ary, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, G. Sagui, V. Babin, and P. A. Kollman, AMBER 10, University of California, San Francisco, 2008.

[13] T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.*, **90**, 1007-1023 (1989).