Theor Chem Account (2007) 118:937–945
DOI 10.1007/s00214-007-0374-7

REGULAR ARTICLE

Fragment interaction analysis based on local MP2

Takeshi Ishikawa · Yuji Mochizuki · Shinji Amari · Tatsuya Nakano · Hiroaki Tokiwa · Shigenori Tanaka · Kiyoshi Tanaka

Received: 11 December 2006 / Accepted: 15 June 2007 / Published online: 19 July 2007

Abstract We have developed a fragment interaction analysis based on local MP2 (FILM) in the context of the fragment molecular orbital (FMO) scheme. The primary purpose of this work is to provide a tool for analyzing inter-fragment interaction associated with dispersion interactions in a large molecule such as protein and DNA. Our implementation of local MP2 (LMP2) is based on the algorithm developed by Pulay and Werner. A potential of FILM was demonstrated using the human immunodeficiency virus type 1 protease (HIV-1 PR) complexed with lopinavir (LPV). The total energy, binding affinity, and inter-fragment interaction energy (IFIE) by the FMO method using LMP2 were compared with those obtained by canonical MP2 and the site-specific information in dispersion interaction was obtained. It turned out that the FILM is a useful tool for analyzing the dispersion interaction between an amino acid residue and a specific site of a ligand.

Keywords Local MP2 · Fragment molecular orbital method · FMO · FILM · Interaction analysis

T. Ishikawa (✉) · Y. Mochizuki · H. Tokiwa
Department of Chemistry, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
e-mail: ishi@tvs.rikkyo.ne.jp

T. Ishikawa · Y. Mochizuki · T. Nakano · S. Tanaka
CREST Project, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

S. Amari · K. Tanaka
Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

T. Nakano
Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

S. Tanaka
Graduate School of Science and Technology, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Hyogo 657-8501, Japan

K. Tanaka
AdvanceSoft Corporation, Center for Collaborative Research, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

1 Introduction

In recent years, ab initio quantum mechanical calculations have been applied to large molecules such as proteins and nucleic acids. Various approaches have been proposed for avoiding the high cost in calculating a large molecule with quantum mechanical method. The fragment molecular orbital (FMO) method, which has been developed by Kitaura et al. [1–5], is one of the most efficient approaches for quantum mechanical investigation of large molecules. In the FMO method at the Hartree–Fock (HF) level, a target molecule is divided into small fragments and calculations are performed only for each fragment and pair of fragments using a modified Hamiltonian including the environmental electrostatic potential (ESP). It has been shown that this approach reduces computational cost extremely with keeping the chemical accuracy. The FMO scheme has been extended to incorporate the electron correlation in terms of the second-order Møller–Pleset perturbation theory (MP2) [6–8] and coupled-cluster (CC) theory [9]. Additionally, the methods of configuration interaction singles (CIS) [10] and its perturbative doubles (CIS(D)) [11] and multiconfigurational self-consistent-field theory (MCSCF) [12] are combined with the FMO method in the framework of multilayer formulation [13]. A large biomolecular system which includes heavy metal atoms, such as the hydrated cisplatin-DNA complex,

has also been calculated by introducing the model core potential (MCP) into the FMO scheme [14].

A valuable feature of the FMO method is that the interfragment interaction energy (IFIE) is clearly defined in the expression of the total energy [15]. Some studies are reported in which the IFIE is used to analyze the interaction between an amino acid residue and a ligand. Ode et al. [16] performed calculations of interaction energies between human immunodeficiency virus type 1 protease (HIV-1 PR) and its inhibitors. In their study, both of the wild type and L90M mutant were calculated by the FMO scheme for investigation of the mechanism of drug-resistance. Amari et al. [17] developed a visualized cluster analysis of protein–ligand interaction (VISCANA) using the IFIE. Their idea leads to a systematic analysis of the interactions between a specific protein and various ligands. Mochizuki et al. [18] developed a modified version of configuration analysis for fragment interaction (CAFI). The CAFI provides the visualized information of charge transfer in hydrogen bonding interaction.

The IFIE analysis could provide the energetic information between fragments in the region of chemical interest, e.g., pharmacophore. Although the IFIE information is certainly useful in understanding the nature of ligand–protein interaction, one may expect to obtain more site-specific information by which the importance of each fundamental substitution is revealed. The CAFI is one of such tools, but the analysis is performed essentially at the HF level. It has been known that the dispersion or van der Waals interaction is crucial in the protein–ligand interaction. The MP2 calculation introduces the van der Waals interaction energy into the FMO scheme appropriately. If site-specific information of van der Waals interaction is available, this could be very useful as a complementary tool to the CAFI. Thus we consider that the local MP2 (LMP2) method plays a role as such a tool.

For about 30 years, various local correlation methods using localized molecular orbitals have been developed by some groups including Pulay and Werner [19–22]. Among such approaches, the LMP2 method [23–29] requires the lowest computational cost. In LMP2, the total correlation energy is described as the sum of the pair correlation energies based on localized molecular orbitals. By combining LMP2 with the FMO method, we would be able to obtain the IFIE which is decomposed into the pair correlation energies among localized orbitals. Another advantage of LMP2 is to reduce the computational cost to formally linear scaling with molecular size by utilizing two individual approximations, i.e., the restriction of virtual space and the selection of correlated orbital pair. Additionally, by multistep screening [30] and the multipole approximation [31,32] in the procedure of integral transformation, the computational efforts have been reduced significantly. The restriction of virtual space yields reduction of the basis set superposition error (BSSE) from interaction energy in calculations of supermolecular systems [22,33].

In the present paper, we will report the implementation of LMP2 into the context of the FMO method in a developer version of the ABINIT-MP [15]. At the beginning of this paper, we should emphasize that the primary purpose of this work is not to save the cost of MP2 calculations but to develop a tool of fragment interaction analysis based on LMP2 (abbreviated as FILM). As mentioned above, we consider that introduction of LMP2 into the FMO scheme would provide a useful tool to analyze interaction in large molecules including biomolecular systems. In the next section, we will describe the FMO method and our implementation of LMP2. The parallel performance of our program is given in Sect. 3. In the following section, we will demonstrate analyses on the HIV-1 PR with inhibitor and show performance of the FILM.

2 Method and implementation

2.1 Brief description of FMO method and IFIE

In the FMO scheme, a molecule is divided into N_f small fragments by cutting C−C single bonds [2,3]. In this paper, we use terminology of "monomer" and "dimer" which refer to a fragment and a pair of fragments, respectively.

In the case of the HF level of theory (FMO-HF), only energies of all monomers and dimers are calculated by solving the Fock equation with modifications of the ESP from surrounding fragments. Further special attention should be paid to the electron occupation on the orbitals contributed to the single bonds at the both edges. The two electrons belonging to these bonds are assigned to either of fragments. Modified Fock operators for monomer and dimer are shown below [34]:

$$\tilde{f}^I(1) = f^I(1) + \sum_{K\neq I}\left\{u^K(1) + v^K(1)\right\} + \sum_k B_k|\theta_k\rangle\langle\theta_k|, \tag{1}$$

$$\tilde{f}^{IJ}(1) = f^{IJ}(1) + \sum_{K\neq I,J}\left\{u^K(1) + v^K(1)\right\} + \sum_k B_k|\theta_k\rangle\langle\theta_k|, \tag{2}$$

where $f^I(1)$ and $f^{IJ}(1)$ are the conventional Fock operators of monomer and dimer. $u^K(1)$ and $v^K(1)$ are one electron potentials, where the former is the nuclear attraction from the Kth monomer and the latter is the repulsion given by the electronic distribution in the Kth monomer. The third terms are introduced for the sake of projecting out the orbitals θ_k which should not be occupied in each fragment calculation [2,3]. The parameter B_k is a sufficiently large positive value and is usually set at about 10^6 to 10^8. The electronic distributions in $v^k(1)$ are updated by iterative calculations

until self-consistency is achieved. The monomer and dimer energies are denoted by $E^{\mathrm{HF}}{}_I$ and $E^{\mathrm{HF}}{}_{IJ}$, respectively. The total energy is obtained by the following equation [1]:

$$E^{\mathrm{HF}}{}_{\mathrm{total}} = \sum_{I>J} E^{\mathrm{HF}}{}_{IJ} - (N_f - 2) \sum_I E^{\mathrm{HF}}{}_I. \tag{3}$$

Equation (3) is rewritten by Eq. (4) using $\Delta E^{\mathrm{HF}}{}_{IJ}$ which is the IFIE in FMO-HF as

$$E^{\mathrm{HF}}{}_{\mathrm{total}} = \sum_{I>J} \Delta E^{\mathrm{HF}}{}_{IJ} + \sum_I E'^{\,\mathrm{HF}}{}_I, \tag{4}$$

where $E'^{\,\mathrm{HF}}_I$ is the monomer energy without electrostatic potentials produced by other monomers. The detail of the formulation of the IFIE was described by Nakano et al. [15].

In the MP2 level of theory (FMO-MP2), the total energy is calculated using

$$E^{\mathrm{MP2}}{}_{\mathrm{total}} = \sum_{I>J} E^{\mathrm{MP2}}{}_{IJ} - (N_f - 2) \sum_I E^{\mathrm{MP2}}{}_I, \tag{5}$$

where $E^{\mathrm{MP2}}{}_I$ and $E^{\mathrm{MP2}}{}_{IJ}$ are MP2 energies of monomer and dimer, respectively [6–8]. The definition of the IFIE in FMO-MP2 is given by

$$\Delta E^{\mathrm{MP2}}{}_{IJ} = \Delta E^{\mathrm{HF}}{}_{IJ} + \Delta E^{\mathrm{corr}}{}_{IJ}, \tag{6}$$

where $\Delta E^{\mathrm{corr}}{}_{IJ}$ is the correction on the IFIE by including the electron correlation and obtained by the following equation:

$$\Delta E^{\mathrm{corr}}{}_{IJ} = E^{\mathrm{corr}}{}_{IJ} - E^{\mathrm{corr}}{}_I - E^{\mathrm{corr}}{}_J. \tag{7}$$

The quantities of $E^{\mathrm{corr}}{}_I$ and $E^{\mathrm{corr}}{}_{IJ}$ are electron correlation energies by MP2 calculations of monomer and dimer, respectively.

2.2 Incorporation of local MP2 into FMO scheme

Our implementation of LMP2 was made according to the formulation and algorithm developed by Pulay and Werner [22–26,30]. In LMP2 method, the virtual space is limited to the modified atomic orbitals (AOs) spatially close to each local occupied orbital. The limited virtual space and modified AO are called as "domain" and "projected AO", respectively. In the conventional LMP2, the projected AOs are defined by excluding occupied orbitals $|i\rangle$ from a normal atomic orbital $|\mu\rangle$ [25]:

$$|p\rangle = \left(1 - \sum_i |i\rangle\langle i|\right)|\mu\rangle. \tag{8}$$

In the case of LMP2 combined with the FMO method, the projected AOs should be constructed by excluding not only occupied orbitals $|i\rangle$ but also $|\theta_k\rangle$ appearing in Eqs. (1) and (2). We obtained the definition of projected AO instead of Eq. (8) as

$$|p\rangle = \left(1 - \sum_{\alpha\beta \in |i\rangle,|\theta_k\rangle} |\alpha\rangle S^{-1}{}_{\alpha\beta}\langle\beta|\right)|\mu\rangle, \tag{9}$$

where $S_{\alpha\beta}$ is overlap between $|\alpha\rangle$ and $|\beta\rangle$. Occupied orbitals are orthogonalized one another. In contrary, $|i\rangle$ and $|\theta_k\rangle$ are not orthogonal and $|\theta_k\rangle$ are not orthogonal one another as well.

A method determining a domain has been proposed by Boughton and Pulay [24]. Although we employed their method in our implementation, two thresholds were introduced for determination of the domain. These are denoted by **Th$_1$** and **DTh$_1$**. The former is used regularly, except for the inter-fragment orbital pairs in dimer calculation whose domain is determined by using with **DTh$_1$**. Introduction of **DTh$_1$** is intended to retain more accurate pair correlation energies for inter-fragment orbital pairs without intractable computational cost (see Sect. 4.2).

In LMP2 method, only selected orbital pairs are considered. A criterion for the selection of the orbital pair has been proposed by Hampel et al. [22]. In their criterion, the minimum distance between atoms contained in the domain of each orbital is used for the selection of orbital pairs. This criterion was adopted and a threshold value is denoted by **Th$_2$**.

The algorithm of the integral transformation for LMP2 was described by Saebø et al. [30] and the prescreening using the Cauchy–Schwarz inequality has been proposed by Werner et al. [26]. Their algorithms were used and here we denote this screening threshold as **Th$_4$**. Another approximation for reducing the intermediate buffer of the transformation was proposed by Rauhut et al. [25]. Introducing their approximation, the number of the first half transformed integrals can be reduced. In the present work, this threshold is denoted by **Th$_3$**. The five thresholds used in our FMO-based LMP2 are summarized in Table 1. Setting of these threshold values will be discussed in Sect. 4.2.

Table 1 Summary of the five thresholds introduced in our implementation of LMP2 into the FMO scheme

Threshold	Effect
Th$_1$	Determination of the domain regularly
DTh$_1$	Determination of the domain only for pairs of inter-fragment orbitals in dimer calculations
Th$_2$	Selection of pairs of orbitals
Th$_3$	Reducing the number of the first half transformed integrals
Th$_4$	Screening in the first half transformation

In the FMO scheme with LMP2 (FMO-LMP2), the total energy is evaluated as

$$\tilde{E}^{\mathrm{LMP2}}_{\mathrm{total}} = \sum_{I>J} \tilde{E}^{\mathrm{LMP2}}_{IJ} - (N_f - 2) \sum_{I} \tilde{E}^{\mathrm{LMP2}}_{I}. \quad (10)$$

We distinguish the energy calculated by FMO-LMP2 from that by FMO-MP2 by using a sign of tilde. Equation (10) are simply obtained from Eq. (5) by replacing $E^{\mathrm{MP2}}{}_{IJ}$ and $E^{\mathrm{MP2}}{}_{I}$ by $\tilde{E}^{\mathrm{LMP2}}_{IJ}$ and $\tilde{E}^{\mathrm{LMP2}}_{I}$, respectively. On the other hand, the definition of the IFIE in FMO-LMP2, which is denoted by $\Delta\tilde{E}^{\mathrm{LMP2}}_{IJ}$, is given by Eqs. (11) and (12):

$$\Delta\tilde{E}^{\mathrm{LMP2}}_{IJ} = \Delta E^{\mathrm{HF}}{}_{IJ} + \Delta\tilde{E}^{\mathrm{corr}}_{IJ}, \quad (11)$$

$$\Delta\tilde{E}^{\mathrm{corr}}_{IJ} = \sum_{i'\geq j'} \epsilon_{i'j'}, \quad (12)$$

where $\epsilon_{i'j'}$ in Eq. (12) are pair correlation energies obtained from dimer calculation. Because of high locality of occupied orbitals in LMP2, they can be assigned to either fragment. The summation of orbital pair $i'j'$ in Eq. (12) is then limited within the span of only inter-fragment orbital pairs. As clearly shown in Eqs. (11) and (12), we are able to decompose the IFIE originated from the dispersion or van der Waals interaction into contributions of each orbital pair of inter-fragment. These two equations represent the essence of the FILM.

3 Tests of program and parallel performance

We would demonstrate parallel performance of the program before discussing how the FILM is useful to analyze the interactions in large biomolecular systems. For simplicity, we show the performance on a model system of polyglycine peptide with 10 residues (Gly_{10}) without the FMO scheme. The 6-31G** [35] basis set was employed and the number of basis functions was 775. The total correlation energy obtained by the canonical MP2 was −6.3405 hartree, and that by the LMP2 method was −6.1996 hartree. The loss of the correlation energy by LMP2 was only 2.2 %, and this difference falls in a typical range for a double-zeta basis set [26].

In the present program, two time-consuming procedures were parallelized with message-passing interface (MPI) [36]. The transformation of two-electron integrals was parallelized under the quadruple loop for shells of contracted basis functions. The simplified loop structure was illustrated by Saebø et al. [30]. Another parallelization was done at the procedure of making the residuum matrices [25] which are calculated for pairs of localized orbitals in solving the linear equation. We summarizes the acceleration and efficiency in Table 2. These data were taken using a linux-based cluster of 5 nodes Intel Dual-Xeon (dual core, 2.8 GHz clock rate). The efficiency of the parallelization in the procedure of integral transformation is better than that of solving the linear equation. Since a *dynamic* update procedure [25] was adopted in solving the linear equation, the number of iterations increased with many CPUs. The ratio of the total computational time with 10 CPUs was 0.15 to that with single CPU and its acceleration was 6.6. In the FMO method, parallelization is done for each monomer or dimer [6,37], and the parallel acceleration of our program is expected to be practical.

Table 2 The ratios of the computational time[a] and accelerations on a model system of Gly_{10} in LMP2 without the FMO method ($\mathbf{Th_1} = 0.02$, $\mathbf{Th_2} = 8.0$ Å, $\mathbf{Th_3} = 0.01$, $\mathbf{Th_4} = 10^{-8}$)

Processor	Transformation	Linear equation	LMP2 total
1	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
2	0.52 (1.92)	0.73 (1.37)	0.55 (1.82)
4	0.28 (3.57)	0.43 (2.33)	0.30 (3.33)
5	0.23 (4.35)	0.39 (2.56)	0.25 (4.00)
10	0.15 (6.67)	0.22 (4.55)	0.15 (6.67)

[a] 5 nodes Intel Dual-Xeon (dual core, 2.8 GHz clock rate) were used

4 Demonstration of interaction analysis for HIV-1 PR with LPV

4.1 Target system and details of calculations

In this section, we demonstrate the performance of the FILM by applying it to the HIV-1 PR with lopinavir (LPV), which is illustrated in Fig. 1. The HIV-1 PR protein is composed with two identical polypeptides, each of which consists of 99 amino acid residues. The LPV molecule is one of inhibitors of the HIV-1 PR. The X-ray crystal structure of a complex was reported by Stoll et al. [38] (Protein Data Bank [39] code: 1MUI). This structure was used as the initial coordinate set. The terminating hydrogen atoms were attached and one water molecule, which was considered to be crucial for the ligand–protein binding [40], was added. Energy minimization was carried out using the force-field of MMFF94 [41] for the hydrogen atoms in the ligand binding pocket and using the force-field of AMBER99 [42] for the remaining hydrogen atoms. We used the structure obtained above throughout this paper.

In our calculations, each amino acid residue of the HIV-1 PR or a water molecule was treated as a single fragment. The LPV molecule was divided into four fragments denoted by LPV_1, LPV_2, LPV_3, and LPV_4. In Fig. 2, the chemical structure of LPV is illustrated together with the fragmentation. Totally 203 fragments were considered in our calculations; 198 amino acid residues, a water molecule, and four fragments of LPV. The interaction energy between LPV and a certain amino acid residue is given by summing up the IFIEs with respect to four fragments of LPV. Among 198 amino

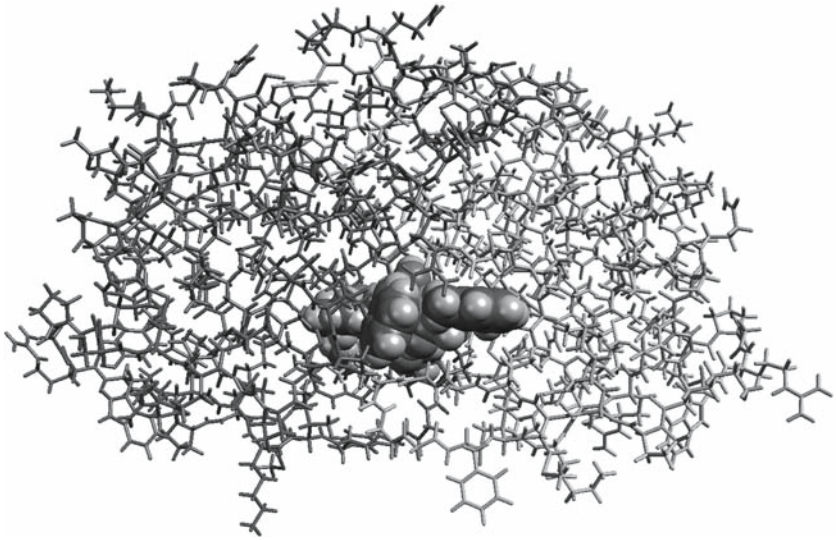

Fig. 1 Graphic representation of the HIV-1 PR complexed with lopinavir (LPV). Two identical protein chains are drawn with *dark and light sticks*, and LPV is illustrated in ball representation

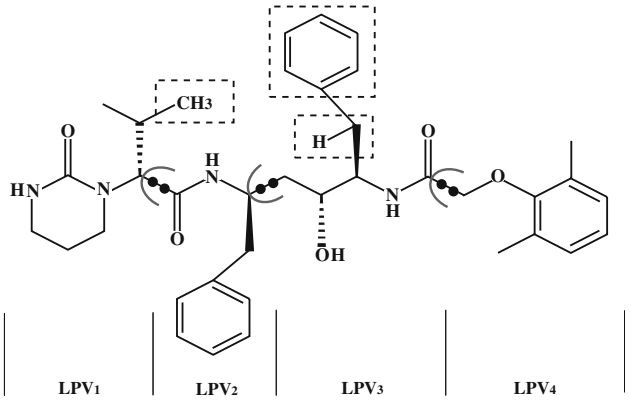

Fig. 2 The chemical structure of LPV and the manner of division into four fragments. The *squares with dotted lines* show three sites which are referred for the demonstration of the FILM in Sect. 4 (see text)

acid residues, 18 residues are located within 5.0 bohr from LPV. We selected 14 of them for the object of the present analysis. Those are $Asp25_B$, $Asp29_B$, $Asp30_B$, $Ile47_B$, $Ile50_B$, $Pro81_B$, $Ile84_B$, $Asp25_A$, $Asp29_A$, $Asp30_A$, $Ile47_A$, $Ile50_A$, $Pro81_A$, and $Ile84_A$.

Both the Pipek-Mezey [43] and the Boys [44] localization schemes were implemented in our program. We adopted the former by which the σ–π separation can be retained. Calculations of FMO-MP2 were carried out for comparison.

4.2 Tests for threshold values in LMP2

We checked the dependence of the results on the five thresholds, which have been introduced for improving the efficiency of the FMO-based LMP2 calculation (see Table 1). A dimer of $I84_B$ and LPV_1 was selected as an example for the test calculations, because the dispersion interaction between $I84_B$ and LPV was found to be crucial, as shown later. The 6-31G* basis set [35] was employed and the total number of basis functions of this dimer was 352. For the check of accuracy, energies were monitored in five decimal places.

First, we examined three thresholds of $\mathbf{Th_1}$, $\mathbf{DTh_1}$ and $\mathbf{Th_2}$, which were used to determine the domain and orbital pair. These thresholds influence the correlation energy and the IFIE directly. The results are summarized in Table 3. The LMP2 correlation energy, namely $\tilde{E}^{\mathrm{LMP2}}{}_{IJ} - E^{\mathrm{HF}}{}_{IJ}$, is shown in row (A) and the correction on the IFIE with FMO-LMP2, which is denoted by $\Delta\tilde{E}^{\mathrm{corr}}{}_{IJ}$ in Eq. (12), is shown at row (B). In the case of FMO-MP2 as the reference, the correlation energy is −2.54953 hartree and correction on the IFIE is −0.00346 hartree. On the other hand, FMO-LMP2 calculation with $\mathbf{Th_1} = 0.02$, $\mathbf{DTh_1} = 0.02$, and $\mathbf{Th_2} = 8.0$ Å provides −2.49210 hartree (97.75 %) and −0.00227 hartree (65.65 %), respectively.

When the three thresholds are tightened, the IFIE obtained with FMO-LMP2 gets close to that with FMO-MP2. For example, in the case of $\mathbf{Th_1} = 0.001$, $\mathbf{DTh_2} = 0.001$, and $\mathbf{Th_2} = 10.0$ Å, the correction on the IFIE become −0.00299 hartree, which recovers 86.42% of that with FMO-MP2. However, the memory requirement is increased by 12.5 times, and the CPU time is increased by 2.7 times. On the other hand, in the case that only $\mathbf{DTh_1}$ is set to be smaller, i.e., $\mathbf{Th_1} = 0.02$, $\mathbf{DTh_1} = 0.001$, and $\mathbf{Th_2} = 8.0$ Å, the IFIE obtained with LMP2 is −0.00295, which is 85.26% of that with FMO-MP2. In this setting, the memory and the CPU time are required about 4.1 times and 1.3 times of those with the first thresholds, respectively. By exploiting the threshold of $\mathbf{DTh_1}$, we can obtain the more accurate IFIEs with small memory requirement and little CPU time. This computational advantage is expected to be crucial in the case of the FMO calculation including large fragments. It should be noted that the BSSE reduced by FMO-LMP2 is reintroduced by using

Table 3 (A) and (B) denote LMP2 correlation energies and corrections on the IFIE by FMO-LMP2/6-31G*, respectively, using some threshold values

$\mathbf{DTh_1}$		$\mathbf{Th_2} = 8.0$ Å			$\mathbf{Th_2} = 10.0$ Å		
		$\mathbf{Th_1}$			$\mathbf{Th_1}$		
		0.020	0.010	0.001	0.020	0.010	0.001
0.020	(A)	−2.49210	–	–	−2.49212	–	–
	(B)	−0.00227	–	–	−0.00229	–	–
0.010	(A)	−2.49220	−2.51404	–	−2.49223	−2.51406	–
	(B)	−0.00236	−0.00238	–	−0.00239	−0.00239	–
0.001	(A)	−2.49303	−2.51487	−2.54703	−2.49306	−2.51489	−2.54703
	(B)	−0.00295	−0.00297	−0.00299	−0.00298	−0.00299	−0.00299
FMO-MP2	(A)	−2.54953					
	(B)	−0.00346					

Other threshold values are set as $\mathbf{Th_3} = 0.01$ and $\mathbf{Th_4} = 10^{-8}$

a small value for $\mathbf{DTh_1}$. We will discuss about the BSSE at the Sect. 4.4 in more detail. In the following, we will show the results using the setting of $\mathbf{Th_1} = 0.02$, $\mathbf{DTh_1} = 0.001$, and $\mathbf{Th_2} = 8.0$ Å.

Finally, we should test the remaining two thresholds, $\mathbf{Th_3}$ and $\mathbf{Th_4}$, which are used in the procedure of the integral transformation. After several trials on these values, we have chosen 0.01 and 10^{-8} for $\mathbf{Th_3}$ and $\mathbf{Th_4}$, respectively. In the case of this test, our setting leads to errors less than 0.00002 hartree for the correlation energy and the correction on the IFIE.

4.3 Total energy and binding affinity

Table 4 summarizes the total energies of the HIV-1 PR complexed with LPV by the FMO method employing the 6-31G* basis set. Recall that the total energy in FMO-HF, FMO-MP2, and FMO-LMP2 are defined by Eqs. (3), (5), and (10), respectively. Total correlation energy obtained by FMO-LMP2 is −220.0701 hartree which is 97.1% of that obtained by FMO-MP2. This loss in the correlation energy by FMO-LMP2 is of the same order as that found in the conventional LMP2 (see Sect. 3).

The binding energy between the HIV-1 PR and LPV was calculated by the three methods. The definitions of the binding energy in FMO-HF and FMO-MP2 are given by

$$E^{\mathrm{HF}}{}_{\mathrm{BA}} = E^{\mathrm{HF\ (Complex)}}_{\mathrm{total}} - E^{\mathrm{HF\ (HIV-1\ PR)}}_{\mathrm{total}} - E^{\mathrm{HF\ (LPV)}}_{\mathrm{total}}, \tag{13}$$

$$E^{\mathrm{MP2}}{}_{\mathrm{BA}} = E^{\mathrm{MP2\ (Complex)}}_{\mathrm{total}} - E^{\mathrm{MP2\ (HIV-1\ PR)}}_{\mathrm{total}} - E^{\mathrm{MP2\ (LPV)}}_{\mathrm{total}}, \tag{14}$$

where $E^{\mathrm{(Complex)}}_{\mathrm{total}}$, $E^{\mathrm{(HIV-1\ PR)}}_{\mathrm{total}}$, and $E^{\mathrm{(LPV)}}_{\mathrm{total}}$ indicate the total energies of the HIV-1 PR complexed with LPV, HIV-1 PR, and isolated LPV, respectively. The binding energy by FMO-LMP2 is given by

$$\tilde{E}^{\mathrm{LMP2}}{}_{\mathrm{BA}} = E^{\mathrm{HF}}{}_{\mathrm{BA}} + \sum_{I\ \in\ \mathrm{LPV}} \sum_{J\ \in\ \mathrm{PR}} \Delta\tilde{E}^{\mathrm{corr}}{}_{IJ}, \tag{15}$$

where $\Delta\tilde{E}^{\mathrm{corr}}{}_{IJ}$ is defined by Eq. (12). The difference between the binding affinities obtained by FMO-MP2 and FMO-LMP2 is −13.7 kcal/mol. We consider that the main origin of this difference is the reduction of the BSSE by using LMP2, and this subject will be discussed in the next subsection. The binding energies with FMO-MP2 or FMO-LMP2 increase by about three times from the HF value. This indicates the importance of the electron correlation in the studies of the interaction between HIV-1 PR and LPV.

Table 4 The total energies of HIV-1 PR complexed with LPV and the binding energies of LPV by the FMO-HF, FMO-MP2, and FMO-LMP2 employing 6-31G*

	Total energy	Correlation energy	Binding energy	
FMO-HF	−77621.8730	–	−0.0807	(−50.6)
FMO-MP2	−77848.4591	−226.5860	−0.2510	(−157.5)
FMO-LMP2	−77841.9431	−220.0701	−0.2292	(−143.8)

Each energy is shown in hartree and in kcal/mol in parenthesis

4.4 Interaction energies between LPV and amino acid residues

In Table 5, we summarize interaction energies between LPV and selected 14 amino acid residues obtained by FMO-HF, FMO-MP2, and FMO-LMP2 with the 6-31G* basis set. These numerical data are graphically shown in Fig. 3. Since the definition of the IFIE in FMO-LMP2 (Eqs. (11) and (12)) is different from that in FMO-MP2 (Eqs. (6) and (7)), $\Delta E^{\mathrm{MP2}}{}_{IJ}$ and $\Delta\tilde{E}^{\mathrm{LMP2}}{}_{IJ}$ are not necessarily equal to each other. However, in a chemical sense, two values should be comparable. As clearly shown in Fig. 3, the order of the

Table 5 The results of calculations for interaction energies between LPV and selected 14 amino acid residues in HIV-1 PR employing 6-31G*

	FMO-HF	FMO-MP2	FMO-LMP2	(FMO-LMP2)−(FMO-MP2)
$D25_B$	−0.0463 (−29.1)	−0.0538 (−33.8)	−0.0512 (−32.1)	+0.0026 (+1.6)
$D29_B$	−0.0141 (−8.8)	−0.0270 (−16.9)	−0.0273 (−17.1)	−0.0003 (−0.2)
$D30_B$	−0.0120 (−7.5)	−0.0148 (−9.3)	−0.0144 (−9.0)	+0.0004 (+0.2)
$I47_B$	−0.0003 (−0.2)	−0.0053 (−3.3)	−0.0045 (−2.8)	+0.0009 (+0.5)
$I50_B$	−0.0125 (−7.8)	−0.0231 (−14.5)	−0.0209 (−13.1)	+0.0022 (+1.4)
$P81_B$	+0.0033 (+2.1)	−0.0029 (−1.8)	−0.0024 (−1.5)	+0.0005 (+0.3)
$I84_B$	+0.0040 (+2.5)	−0.0043 (−2.7)	−0.0029 (−1.8)	+0.0014 (+0.9)
$D25_A$	−0.0395 (−24.8)	−0.0480 (−30.2)	−0.0478 (−30.0)	+0.0003 (+0.2)
$D29_A$	−0.0040 (−2.5)	−0.0118 (−7.4)	−0.0102 (−6.4)	+0.0016 (+1.0)
$D30_A$	−0.0041 (−2.6)	−0.0127 (−8.0)	−0.0111 (−7.0)	+0.0016 (+1.0)
$I47_A$	−0.0020 (−1.3)	−0.0058 (−3.6)	−0.0053 (−3.3)	+0.0005 (+0.3)
$I50_A$	−0.0081 (−5.1)	−0.0187 (−11.7)	−0.0169 (−10.6)	+0.0017 (+1.1)
$P81_A$	+0.0042 (+2.6)	−0.0022 (−1.4)	−0.0019 (−1.2)	+0.0003 (+0.2)
$I84_A$	−0.0007 (−0.4)	−0.0045 (−2.8)	−0.0038 (−2.4)	+0.0008 (+0.5)

Each energy is shown in hartree and kcal/mol in parenthesis. These values are graphically expressed in Fig. 3

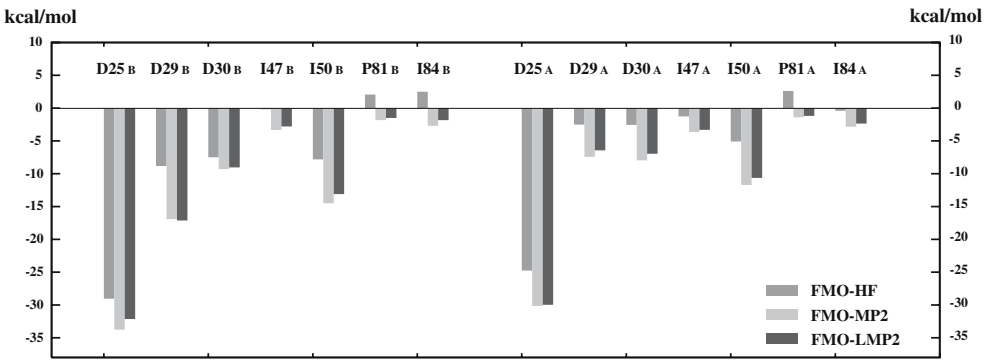

Fig. 3 Interaction energies between LPV and selected 14 amino acid residues in HIV-1 PR by FMO-HF, FMO-MP2, and FMO-LMP2 using 6-31G* (also see Table 5)

corrections on the IFIE obtained from the two methods is qualitatively similar for every amino acid residues. The differences of the IFIEs between by FMO-MP2 and FMO-LMP2 range from −0.2 to +1.6 kcal/mol (Table 5).

As mentioned in Sect. 1, it has been known that LMP2 reduces the BSSE from the interaction energy [22,33]. We consider that the large portion of the difference in the IFIEs between by FMO-MP2 and FMO-LMP2 is attributed to the reduction of BSSE. In order to confirm this issue, we performed simple test calculations for a typical complex model composed with benzene and methanol. This model was selected by the assumption that a polar situation in biomolecular systems was mimicked by insertion of the hydroxyl group. The configuration of this model is shown in Fig. 4. We performed conventional MP2 calculations to obtain the BSSE arising from the electron correlation using the counterpoise method [45], because it is difficult to estimate the amount of BSSE in the context of the FMO scheme. By comparison with the difference in the IFIE by FMO-MP2 and FMO-LMP2, we are able to estimate the reduction of BSSE.

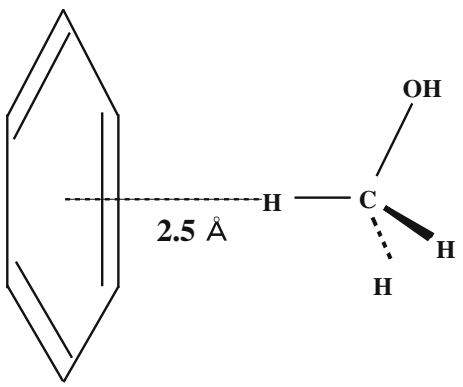

Fig. 4 Configuration of a complex modelcomposed with C_6H_6 and CH_3OH. The distance between two molecules, i.e., 2.5 Å, is close to the minimum point on the potential curve obtained byMP2/6-31G**

Table 6 The IFIEs between C_6H_6 and CH_3OH in a typical model employing 6-31G** (**DTh1** = 0.001)

	IFIE	(FMO-MP2)–(FMO-LMP2)
FMO-MP2	−0.0043 (−2.7)	–
FMO-LMP2	−0.0032 (−2.0)	−0.0011 (−0.7)
BSSE[a]	−0.0014 (−0.9)	

The BSSE arising from electron correlation was calculated by the counterpoise method in conventional MP2 calculations. Each energy is shown in hartree and kcal/mol in parenthesis
[a] The BSSE at the correlated level : $E(BSSE)^{MP2} - E(BSSE)^{HF}$

The results of these calculations are summarized in Table 6. The difference of the IFIEs between by FMO-MP2 and FMO-LMP2 (−0.7 kcal/mol) is comparable to the BSSE (−0.9 kcal/mol). It is expected that the difference is caused mainly from the reduction of BSSE by FMO-LMP2.

In the case of FMO-HF, one can obtain the IFIE originated from the electrostatic interaction mainly by the averaged electronic distribution. For example, $D25_B$ and $D25_A$ have large interaction energies with LPV, which are −29.1 and −24.8 kcal/mol, respectively. Both amino acid residues have a charged $-COO^-$ which is directed to −OH of LPV (the location of $D25_B$, $D25_A$ and LPV are illustrated in Fig. 5a). These geometrical configurations are consistent with the large IFIEs at HF level of theory.

On the other hand, dispersion or van der Waals interaction can be described by MP2 level of theory. $P81_B$, $I84_B$, and $P81_A$ have repulsive interactions with LPV using FMO-HF/6-31G*, which are +2.1, +2.5, and +2.6 kcal/mol, respectively. However, interaction energies obtained by FMO-MP2 or FMO-LMP2 become attractive; corresponding values are −1.8, −2.7, and −1.4 kcal/mol with FMO-MP2, and are −1.5, −1.8, and −1.2 kcal/mol with FMO-LMP2. As shown in the geometrical configurations of these amino acid residues, the stabilizations are originated from the typical dispersion interaction of CH/CH type or CH/π type [46] (the location of $P81_B$, $I84_B$, $P81_A$, and LPV are shown in Fig. 5b). It is well known that these interactions are very important in many biomolecular systems.

The FILM can decompose the IFIEs originated from dispersion interactions into the pair correlation energies of localized orbitals as shown in Eq. (12). By exploiting this feature, one obtains the site-specific information in dispersion interaction. An illustrative example is given using LPV and $I84_B$ in the following. We picked up the three sites of LPV, which are shown in Fig. 2, $-CH_3$, $-C_6H_5$, and −CH. These sites are spatially close and directed to $I84_B$. The accumulation of the pair correlation energies between orbitals located at each site and the all orbitals belonging to $I84_B$ provides the information of the importance of each site of LPV for the binding with $I84_B$. With a detailed analysis, it was found that the $-CH_3$, $-C_6H_5$, and −CH undertook 29.6, 29.2, and 11.8% of the total stabilization originated from dispersion interaction, respectively.

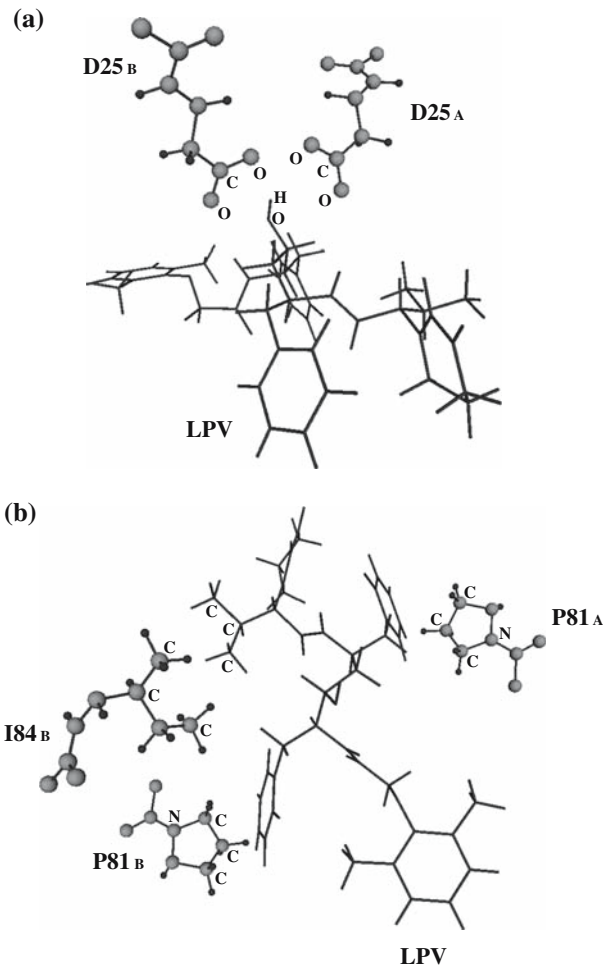

Fig. 5 The geometrical configurations of $D25_B$, $D25_A$, and LPV in **a** and $P81_B$, $I84_B$, $I84_A$, and LPV in **b**

In the present work, the 6-31G* basis set have been used mainly. However, it is known that this type of basis sets are not sufficient to describe the dispersion interaction quantitatively. We consider that larger basis sets are needed in order to obtain a reliable information about the interaction between fragments and this is the subject of our future works.

5 Summary

In this work, a fragment interaction analysis based on local MP2 (FILM) was developed in the FMO scheme [1,2] as a new analysis tool for large molecules including biomolecular systems. We expect that the FILM can provide a useful information of dispersion or van der Waals interactions which are decomposed into the pair correlation energies of localized molecular orbitals.

In combining LMP2 with the FMO method, the conventional definition of the projected AO [25] was modified (see Eq. (9)). For distinction from the regular threshold for determination of the domain [24], another threshold (**DTh$_1$**) was introduced. This is employed only for the domain of pairs of inter-fragment orbitals in a dimer calculation. By exploiting this threshold, the more accurate IFIE would be obtained with tractable computational cost. In FMO-LMP2, the definition of the IFIE was shown in Eqs. (11) and (12), unlike that in FMO-MP2 [6–8] (see Eqs. (6) and (7)). The parallel efficiency of our program was checked for a model system of Gly_{10} without the FMO scheme.

After determining a reasonable set of five thresholds, we calculated the total energy, binding affinity, and IFIEs in the HIV-1 PR complexed with LPV using the FMO-LMP2 as well as the FMO-HF and FMO-MP2. The results showed that these values obtained by FMO-LMP2 were comparable to those by FMO-MP2. Finally, the FILM was applied to the interaction between LPV and $I84_B$ which is one of the amino acid residues of HIV-1 PR. A site-specific information in dispersion interaction could be obtained by partial summation of the pair correlation energies.

It was found that the FILM could provide valuable information, that is to say, the relative importance of the specific sites in the protein–ligand interaction. We conclude that the FILM is a useful tool for fundamental researches on biomolecular systems and drug designs.

Acknowledgments The authors would thank Dr. Kaori Fukuzawa for the discussion on the importance of dispersion interaction in biomolecular systems and also Prof. Kazuo Kitaura for general encouragement. We would be grateful to anonymous referees for the valuable comments to improve the manuscript. The work reported here was supported primarily by the CREST project operated by the Japan Science and Technology Agency (JST) and partially supported by the Revolutionary Simulation Software for 21st Century (RSS21) project operated by Ministry of Education, Culture, Sports, Science and Technology (MEXT). All the calculations were performed on the parallel computing resources at Rikkyo University and Kobe University.

References

1. Kitaura K, Sawai T, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 312:319
2. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 313:701
3. Nakano T, Kaminuma T, Sato T, Akiyama Y, Uebayasi M, Kitaura K (2000) Chem Phys Lett 318:614
4. Fedorov DG, Kitaura K (2004) Chem Phys Lett 389:129
5. Fedorov DG, Kitaura K (2004) J Chem Phys 120:6832
6. Mochizuki Y, Nakano T, Koikegami S, Tanimori S, Abe Y, Nagashima U, Kitaura K (2004) Theo Chem Acc 112:442
7. Mochizuki Y, Koikegami S, Nakano T, Amari S, Kitaura K (2004) Chem Phys Lett 396:473
8. Fedorov DG, Kitaura K (2004) J Chem Phys 121:2483
9. Fedorov DG, Kitaura K (2005) J Chem Phys 123:134103
10. Mochizuki Y, Koikegami S, Amari S, Segawa K, Kitaura K, Nakano T (2005) Chem Phys Lett 406:283
11. Mochizuki Y, Tanaka K, Yamashita K, Ishikawa T, Nakano T, Amari S, Segawa K, Murase T, Tokiwa H, Sakurai M (2007) Theo Chem Acc (in press)
12. Fedorov DG, Kitaura K (2005) J Chem Phys 112:054108
13. Fedorov DG, Kitaura K (2005) J Phys Chem 109:2638
14. Ishikawa T, Mochizuki Y, Nakano T, Amari S, Mori H, Honda H, Fujita T, Tokiwa H, Tanaka S, Komeiji Y, Fukuzawa K, Tanaka K, Miyoshi E (2006) Chem Phys Lett 427:159
15. Nakano T, Kaminuma T, Sato T, Fukuzawa K, Akiyama Y, Uebayasi M, Kitaura K (2002) Chem Phys Lett 351:475
16. Ode H, Neya S, Hata M, Sugiura W, Hoshino T (2006) J Am Chem Soc 128:7887
17. Amari S, Aizawa M, Zhang J, Fukuzawa K, Mochizuki Y, Iwasawa Y, Nakata K, Chuman H, Nakano T (2006) J Chem Inf Model 46:221
18. Mochizuki Y, Fukuzawa K, Kato A, Tanaka S, Kitaura K, Nakano T (2005) Chem Phys Lett 410:247
19. Pulay P (1983) Chem Phys Lett 100:151
20. Pulay P, Saebø S, Meyer W (1984) J Chem Phys 81:1901
21. Saebø S, Tong W, Pulay P (1993) J Chem Phys 98:2170
22. Hampel C, Werner H-J (1996) J Chem Phys 1996(104):6286
23. Pulay P, Saebø S (1986) Theo Chim Acta 69:357
24. Boughton JW, Pulay P (1993) J Comp Chem 14:736
25. Rauhut G, Pulay P, Werner H-J (1998) J Comp Chem 19:1241
26. Schütz M, Hetzer G, Werner H-J (1999) J Chem Phys 111:5691
27. Subotnik JE, Head-Gordon M (2005) J Chem Phys 122:034109
28. Scuseria GE, Ayala PY (1999) J Chem Phys 111:8330
29. Werner H-J, Manby FR, Knowles P (2003) J Chem Phys 118:8149
30. Saebø S, Pulay P (2001) J Chem Phys 115:3975
31. Hetzer G, Pulay P, Werner H-J (1998) Chem Phys Lett 290:143
32. Hetzer G, Schütz M, Stoll H, Werner H-J (2000) J Chem Phys 113:9443
33. Schütz M, Rauhut G, Werner H-J (1998) J Phys Chem A 102:5997
34. Kitaura K, Sugiki S-I, Nakano T, Komeiji Y, Uebayasi M (2001) Chem Phys Lett 336:163
35. Foresman JB, Frisch A (1996) Exploring chemistry with electronic structure methods. 2nd edn. Gaussian, Pittsburgh
36. Message Passing Interface (MPI) http://www.mpi-forum.org/
37. Fedorov GD, Olson RM, Kitaura K, Gordon MS, Koseki S (2004) J Comp Chem 25:872
38. Stoll V, Qin W, Stewart KD, Jakob C, Park C, Walter K, Simmer RL, Helfrich RH, Bussiere D, Kao J, Kempf D, Sham HL, Norbeck WD (2002) Bioorg Med Chem 10:2803
39. Protein Data Bank (PDB) http://www.pdb.org/
40. Kagan RM, Shenderovich MD, Heseltine PNR, Ramnarayan K (2006) Protein Sci 14:1870
41. Halgren T (1999) J Am Chem Soc 112:4710
42. Wang J, Cieplak P, Kollman PA (2000) J Comp Chem 21:1049
43. Pipek J, Mezey G (1998) J Chem Phys 90:4916
44. Boys SF (1966) In Quantum Theory of Atoms, Molecule and the Solid State, Löwdin, PO, Ed.. Academic Press, New York
45. Boys SF, Bernardi F (1970) Mol Phys 19:533
46. Nishio M, Hirota M, Umezawa Y (1998) The CH/π interaction. Wiley–VCH, New York