

# Absolutely Local Occupied and Excited Molecular Orbitals in the Third Order Single Excitation Perturbation Theory for Molecular Interaction

Suehiro Iwata\*

The efficient code to evaluate the molecular interaction energy at the Hartree-Fock level of theory is developed. The method is the third order single excitation perturbation theory based on the locally projected (LP) molecular orbitals (MO). It is shown that with a proper scaling of the molecular orbital integrals the basis set superposition error can be under control even with a small size of basis sets. If the basis set contains augmented diffuse functions, the interaction energy, close to the counterpoise corrected energy, can be directly evaluated without the scaling, as fast as a single supermolecule Hartree-Fock calculation. The key of the method is to use the absolutely local occupied and excited MOs. With these MOs, the amount of the charge-transfer (CT) and the energy contribution of the CT terms are evaluated.

## 1 Introduction

Evaluation of interaction energies between the non-covalent atoms is one of the most important subjects in quantum chemistry. Last two decades the quantum chemical calculations become indispensable tools in the studies of the molecular clusters, in which the non-covalent molecular interaction plays a key role as the cohesive force [1], [2]. Last few years the author has been working for developing a practical but sufficiently accurate molecular orbital theory for the molecular interaction.[3],[4],[5],[6],[7] In particular, the theory is intended to apply for the studies of the molecular clusters and the intra-cluster reactions without using the empirical parameters. Thus, the efficiency is important in the developed method. One of the difficulties in evaluating the weak molecular interaction both with the *ab initio* molecular orbital (MO) theory and with the density functional theory (DFT) is the basis set superposition error (BSSE), which is inherent in the basis set expansions [1], [8]. Both theories are often claimed to be the *first principle* method, or the *ab initio* method. Nevertheless, in both theories, in practice, Schrödinger (or Dirac) equation is solved approximately. The BSSE results from the inconsistency between the approximations for the isolated systems and that for the composite system. In most of the approximation methods, the basis set expansions are used for the one-electron functions (orbitals) and also for the many-electron functions (electron configurations). Liu and McLean defined two inconsistencies in evaluating the interaction between two He atoms, Orbital Basis Inconsistency (OBI) and Configuration Basis Inconsistency (CBI) [9]. This distinction in the inconsistencies is important to avoid or to remove the BSSE. Also it is worth mentioning that the size-consistent [10] and size-extensive [11] many-electron theories do not imply that they are free of CBI. The counterpoise (CP) procedure [12] is mostly used to remove the BSSE. It is in most cases implicitly assumed that the procedure can be applied for the multi-determinant wave functions such as the Møller-Plesset perturbation expansions, but it is not obvious that the CP procedure can remove CBI. Because the closed shell Hartree-Fock (HF) wave function is a single-determinant both for the isolated and composite systems, no CBI sneaks in, and therefore the BSSE is solely the OBI. For the dimer, the OBI in the HF energies can be removed by the CP procedure, because at every step of the procedure the wave function is determined variationally with a given set of one-electron basis functions. However, for the clusters consisting of more than two molecular units, the CP procedure is not unique as discussed by Valiron and Mayer [13], who proposed a more compre-

hensive and time-consuming procedure by extending the method described by White and Davidson [14]. The alternative way to the CP is the extrapolation to the basis set limit [15], and this procedure is possible to be applicable for the correlated wave functions to remove the CBI, although it is a very expensive way. Hobza and his coworkers extensively used the approximate version of this procedure [16].

A simple idea to avoid the OBI is to use the basis sets only local to each molecular unit. Gianinetti and his coworkers derived an equation to determine the molecular orbitals (MOs) under the local expansion constraint, and they called the method "self-consistent field MO for molecular interaction, SCF MO MI" [17], [18]. We reformulated the equations using the projection operators and re-named it "locally projected, LP MO" after the characteristics of the formula to be solved [3]. SCF MO MI is too general for a particular method. Besides, we proved mathematically and also numerically [3] that the original form of SCF MO MI (or LP MO) is a deficient method to study the molecular interaction. It is because the electron delocalization (or charge-transfer) cannot be taken into account in the wave function. Therefore, the binding energy evaluated with this method is inherently underestimated. To introduce the electron delocalization, we developed the perturbation theory using the locally projected excited MOs [4], [5], [6], [7]. Because the LP MOs are canonical for the local Fock matrix but not canonical for the Fock matrix for the full system, the single excitations of the CT type play an important contribution in the perturbation expansion. With the local excited MOs properly orthogonalized to the local occupied MOs, the third and fourth order corrections of the single excitation perturbation theory (LP MO SPT) were examined for some typical hydrogen bonding systems, and the fourth order correction is very small [7]. One of the other important findings is that if the basis sets are not extensive, the proper selection of the local excited MOs is needed. Otherwise, the SPT recovers the BSSE. In the other words, the CT contributions to the binding energy is not separable in the small basis sets. The present work is the extension of those papers.

Khaliullin, Head-Gordon and their coworkers also studied the LP MO (SCF MO MI) [19], [20], [21]. They originally used it to speed-up the SCF calculations for the large ensembles of molecules and then applied it for analyzing the interaction energy. They called the occupied MOs the Absolutely Localized MO (ALMO). To evaluate the CT terms, they had to subtract the BSSE from the 2nd order energy evaluated using all of the local virtual (excited) MOs.

In retrospect, Stoll et al might be the first to derive the variational equations with the local expansion constraint [22]; Their local unit is an atom or a fragment of covalently bonded molecule, and therefore, the energy evaluated with the constrained local wave function was not a good approximation to that obtained without the constraint. Later along this line Sironi and his coworkers developed the extremely localized molecular orbital (ELMO) theory [23], and one of the characteristics is the transferability of ELMO. Very recently to improve the accuracy, they developed the perturbation expansion theory for ELMO[24], obviously influenced by the LP MO SPT calculations. In these works, because the fragment unit is covalent-bonded, a part of local basis sets have to be shared with the neighboring fragment.

In this paper, an efficient third order LP SPT code is reported, and the CT terms in the second and third order SPT are determined in terms of both of the energy and of the electron density. In our scheme no CP correction is required. Besides, the Mulliken population in terms of the second order correction to the electron density is defined.

## 2 Theoretical

A few progress toward our purpose was made after the previous publications [5], [7]. Below, an outline of the locally projected self-consistent (LP SCF) MO theory and of the single excitation perturbation theory (SPT) is only briefly given, while keeping the description self-contained as possible [3], [5], [7], [25].

## 2.1 The direct third order energy

The efficient code to evaluate the third order correction energy  $E^{3SPT}$  directly from the AO (one-electron basis sets) integrals is developed. The correction energy  $E^{3SPT}$  is evaluated with the first order wave function  $\psi^{(1)}$  as

$$E^{3SPT} = \langle \psi^{(1)} | \hat{V} | \psi^{(1)} \rangle \quad (1)$$

For the single excitation PT, the first order wave function is

$$\psi^{(1)} = \sum_i^{occ} \sum_r^{ex} \left| (i \rightarrow r)^S \right\rangle \mathbf{a}_{i,r}^{(1)}, \quad (2)$$

where  $(i \rightarrow r)^S$  stands for the single excitation singlet configuration state function. Below,  $i, j$  and  $k$  are for the occupied MOs, and  $r, s$  and  $t$  are for the excited MOs. The correction energy is

$$E^{3SPT} = \sum_{j,t} \sum_{i,r} \mathbf{a}_{jt}^{(1)} [2 \langle \underline{i}(1)t(2) | r(1)\underline{j}(2) \rangle - \langle \underline{i}(1)t(2) | \underline{j}(1)r(2) \rangle] \mathbf{a}_{ir}^{(1)} \quad (3)$$

where the underbar  $\underline{i}$  implies that in evaluating the integrals, the biorthogonally-transformed orbital  $|\underline{i}\rangle$  has to be used. The occupied MOs are expanded [4] as

$$|i_X\rangle = \sum_{p \in X} |\chi_p\rangle t_{p,i} \equiv \boldsymbol{\chi}_X \mathbf{t}_{X_i} \quad (4)$$

$$|\underline{i}_X\rangle = \sum_Y \sum_{j_Y} |j_Y\rangle u_{j_Y, i_X} \equiv \boldsymbol{\chi} \mathbf{t}_{X_{\underline{i}}} \quad (5)$$

where a row vector  $\boldsymbol{\chi}_X$  is the local basis set vector placed on Molecule  $X$  and a vector  $\boldsymbol{\chi}$  is the full basis set vector. The column vector  $\mathbf{t}_{X_i}$  is the MO coefficient vector. The biorthogonal MO  $|\underline{i}_X\rangle$  has the MO coefficients on the basis sets on the other molecular units through the transformed matrix  $U \equiv \mathcal{S}_o^{-1}$ . Here  $\mathcal{S}_o$  is the MO overlap matrix among the occupied MOs. The third order energy can be written in terms of AOs. By defining the transition density matrix as

$$D_{AB} = \sum_i^{occ} \sum_r^{ex} \mathbf{a}_{ir}^{(1)} t_{A_i} t_{Br} \quad (6)$$

the correction energy is

$$\begin{aligned} E^{3SPT} &= \sum_{A,B} \sum_{C,D} D_{AB} [2 \langle C(1)B(2) | D(1)A(2) \rangle - \langle C(1)B(2) | A(1)D(2) \rangle] D_{CD} \\ &\equiv \sum_{A,B} D_{AB} G_{BA} = \text{trace} [\mathbf{D}\mathbf{G}] \end{aligned} \quad (7)$$

where

$$\begin{aligned} G_{BA} &\equiv \sum_{C,D} [2 \langle C(1)B(2) | D(1)A(2) \rangle - \langle C(1)B(2) | A(1)D(2) \rangle] D_{CD} \\ &\equiv \sum_{C,D} [2 (CD|BA) - (CA|BD)] D_{CD} \end{aligned} \quad (8)$$

Note that both matrices  $\mathbf{D}$  and  $\mathbf{G}$  are not symmetric. By modifying the subroutine for evaluating the two-electron part of Fock matrix, matrix  $\mathbf{G}$  can be computed as fast as the cpu required for constructing a Fock matrix.

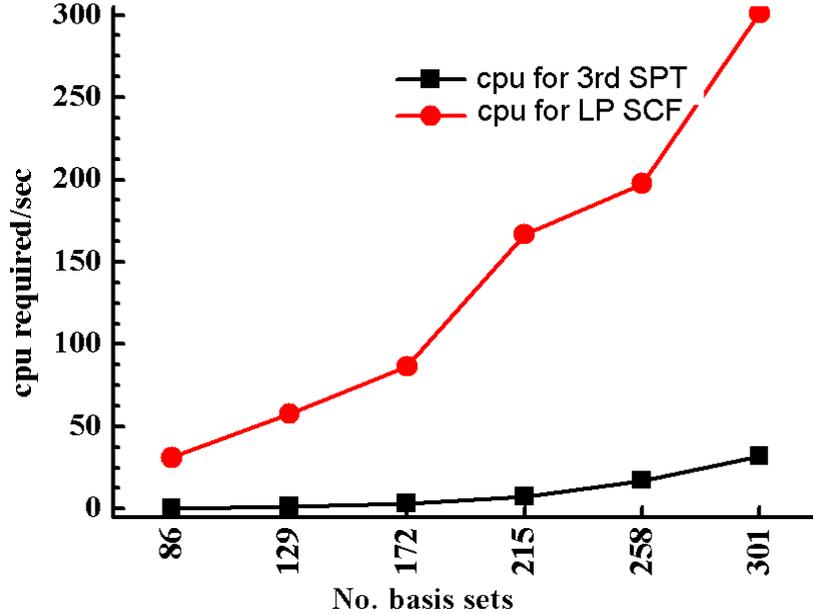


Figure 1: The comparison of cpu required in the LP SCF step and in the second and third order SPT. The computations are for various isomers of water clusters  $(\text{H}_2\text{O})_n$ ,  $n=2 - 7$  with aug-cc-pVDZ. The average is taken over the isomers.

The required cpu for LP SCF (till the convergence) and for SPT ( $E^{2SPT} + E^{3SPT}$ ) is plotted for water clusters in Figure 1. The cpu is an averaged value for several isomers of  $(\text{H}_2\text{O})_n$ . The computer used is Mac Pro OS 10.6.2 (16G memory) with Intel compiler 11.1.080, and BLAS and LAPACK in Intel MKL are used. All two-electron integrals are kept in core during the calculations. Figure 1 shows that independent of the number of basis sets, the required cpu for the 2nd and 3rd order SPT is about 10% of the LP SCF procedure, which is faster than the supermolecule SCF calculations [3], as Khaliullin et al first intended for the method [19]. Also, it can be noticed that the required cpu for LP SCF is almost linear on the size of basis set (the number of water molecules). It is because that the eigenvalue problem to be solved is always 43 (the total number of basis sets for each water molecule), independent on the number of water molecules.

## 2.2 Scaling the off-diagonal elements

The excited (external) MOs in the perturbation expansion should also be expanded locally to each molecular unit. They are determined by solving the matrix eigenvalue problem [4]

$$\tilde{\chi}_X(1 - \hat{\mathbf{P}}_{occ})\chi_X t_{Xr} = \tilde{\chi}_X \chi_X t_{Xr} \eta_{Xr} \quad (9)$$

If the eigenvalue  $\eta_{Xr} = 1$  (in practice,  $> 0.99999$ ), the excited orbital  $|r_X\rangle = \chi_X t_{Xr}$  is local and orthogonal to all of the occupied orbitals; they can be called absolutely local excited orbitals (ALEX MO) [4], [7]. These orbitals can be used directly in the perturbation calculation, but the orbitals  $\chi_X t_{Xs}$ , whose eigenvalue  $\eta_{Xs}$  is smaller than 0.99999, has to be partially delocalized for the perturbation calculation as

$$|s_X\rangle \equiv N_{Xs}(1 - \hat{\mathbf{P}}_{occ})\chi_X t_{Xs} \quad (10)$$

in order to be orthogonal to all of the occupied orbitals. Accordingly the excited MOs in the perturbation expansion are classified by the value of the eigenvalue  $\eta_{Xr}$ .

In the previous works [7], it was demonstrated that the third order corrected binding energy  $E^{3BE}(\eta^{th}) \equiv E^{LPSCF} + E^{2SPT} + E^{3SPT}$  depends on the cut-off value  $\eta^{th}$  in excluding the partially delocalized excited MOs, particularly when the basis set is small. Some of the characteristics are summarized as follows. (1) For large basis sets with diffuse functions such as aug-cc-pVTZ, the corrected binding energy  $E^{3BE}$  is very close to the counterpoise corrected binding energy  $E^{CP-SCF}$ , irrespective of the cut-off value  $\eta^{th}$ . The partially delocalized excited MOs do not contribute to the  $SPT$  binding energy. (2) For small basis sets without diffuse functions such as cc-pVDZ, if all of excited MOs are used (that is,  $\eta^{th} = 0$ ),  $E^{3BE}(0)$  becomes close to the uncorrected SCF binding energy  $E^{uncorr-SCF}$  containing the BSSE. On the other hand, if only ALEx MOs are used in the calculations,  $E^{3BE}(0.99999)$  underestimates the binding energy. (3) For the intermediate size of basis sets such as aug-cc-pVDZ,  $E^{3BE}(0.99999)$  evaluated only with ALEx MOs is slightly smaller than  $E^{CP-SCF}$  (in most cases by 0.3kJ/mol) for the dimers and trimers. But for large sizes of the clusters having many multiple hydrogen bonds, the difference become large. In those cases, if all of the excited orbitals are used,  $E^{3BE}(0)$  overestimates  $E^{CP-SCF}$  by  $1 \sim 2 \text{ kJ/mol}$ . After a few test calculations for a given basis set and for a series of the similar clusters, we can choose an appropriate threshold value  $\eta^{th}$  for the practical studies. A difficulty in using the threshold value to select the excited MOs is that the smooth potential energy curve may not be drawn because the number of excited MOs used is changed at some points.

To avoid this difficulty the scaling of the MO integrals is examined in this work. The linear equation solved for the first order wave function [4],[5] is

$$\sum_{j,s} [(\epsilon_r - \epsilon_i) \delta_{js,ir} - \{[\mathcal{S}_o^{-1}]_{i,j} F_{r,s} - S_{r,s} F_{i,j}\}] a_{js}^{(1)} = \sqrt{2} F_{i,r} \quad (11)$$

If only the interaction term  $F_{i,r}$  between the occupied and excited MOs is scaled as  $f(\eta_r) F_{i,r}$ , the inhomogeneous term of the equation is forced to be small. If the severe scaling factor is used on the excited MO  $r$ ,  $f(\eta_r) F_{i,r}$  becomes nearly zero. But  $\{[\mathcal{S}_o^{-1}]_{i,j} F_{r,s} - S_{r,s} F_{i,j}\}$  is not zero, and the terms may cause the inconsistency, and the linear equation convergence becomes slow. Thus, the MO integrals are scaled as

$$F'_{i,r} = f(\eta_r) F_{i,r} \quad (12)$$

$$F'_{r,s} = f(\eta_r) f(\eta_s) F_{r,s} \quad (13)$$

$$S'_{r,s} = f(\eta_r) f(\eta_s) S_{r,s} \quad (14)$$

The scaling factor tested in the present work is a form

$$f(\eta; \eta_{1/2}, \eta_w) = \frac{1}{2} \left[ 2 - \operatorname{erf} c \left( \frac{\tan \left[ \pi \left( \eta - \frac{1}{2} \right) \right] - \tan \left[ \pi \left( \eta_{1/2} - \frac{1}{2} \right) \right]}{\tan \left[ \pi \left( \eta_{1/2} - \frac{1}{2} \right) \right] - \tan \left[ \pi \left( \eta_w - \frac{1}{2} \right) \right]} \right) \right] \quad (15)$$

The tested sets of parameters  $[\eta_{1/2}, \eta_w]$  in the following figures are  $[0.97, 0.77]$  and  $[0.8, 0.6]$ . The factor  $f(\eta)$  is larger than 0.999 for  $\eta > 0.99$  for the former set and for  $\eta > 0.92$  for the latter set.

### 2.3 The charge-transfer energy with the absolutely local excited molecular orbitals

The  $s + 1$  th order correction energy  $E^{(s+1)SPT}$  is written by the  $s$  th order wave function  $\psi^{(s)}$  as

$$\begin{aligned}
 E^{(s+1)SPT} &= \langle \psi^{(0)} | \hat{V} | \psi^{(s)} \rangle \quad (16) \\
 &= \sum_X^{mol} \left[ \sum_{i,r \in X} \langle \psi^{(0)} | \hat{V} | (i_X \rightarrow r_X)^S \rangle \mathbf{a}_{ir}^{(s)} \right] + \sum_X^{mol} \sum_{Y \neq X}^{mol} \left[ \sum_{j \in X}^{oc} \sum_{t \in Y}^{ex} \langle \psi^{(0)} | \hat{V} | (j_X \rightarrow t_Y)^S \rangle \mathbf{a}_{jt}^{(s)} \right] \\
 &\equiv \sum_X^{mol} E_{LE}^{(s+1)SPT-1}(X) + \sum_X^{mol} \sum_{Y \neq X}^{mol} E_{CT}^{(s+1)SP1}(X \rightarrow Y) \quad (17)
 \end{aligned}$$

where the third line defines the  $s + 1$  order contribution from the local excitation  $E_{LE}^{(s+1)SPT-1}(X)$  and from the charge-transfer excitation  $E_{CT}^{(s+1)SPT-1}(X \rightarrow Y)$ . There is an alternative expression for  $E^{(s+1)SPT}$  for  $s > 1$ . With the  $(2n + 1)$  theorem of the perturbation theory, for instance, the 3rd order correction  $E^{3SPT}$  is

$$E^{3SPT} = \langle \psi^{(1)} | \hat{V} | \psi^{(1)} \rangle \quad (18)$$

$$\begin{aligned}
 &= \sum_X^{mol} \left[ \sum_{i,r \in X} \mathbf{a}_{i,r}^{(1)} \sum_Y^{mol} \sum_Z^{mol} \left\{ \sum_{j \in Y, t \in Z} \langle (i_X \rightarrow r_X)^S | \hat{V} | (j_Y \rightarrow t_Z)^S \rangle \mathbf{a}_{j,t}^{(1)} \right\} \right] \\
 &+ \sum_X^{mol} \sum_W^{mol} \left[ \sum_{i \in X, r \in W} \mathbf{a}_{i,r}^{(1)} \sum_Y^{mol} \sum_Z^{mol} \left\{ \sum_{j \in Y, t \in Z} \langle (i_X \rightarrow r_W)^S | \hat{V} | (j_Y \rightarrow t_Z)^S \rangle \mathbf{a}_{j,t}^{(1)} \right\} \right] \quad (19)
 \end{aligned}$$

$$\equiv \sum_X^{mol} E_{LE}^{3SPT-2}(X) + \sum_X^{mol} \sum_{Y \neq X}^{mol} E_{CT}^{3SPT-2}(X \rightarrow Y) \quad (20)$$

Two definitions for  $E_{LE}^{3SPT}$  and  $E_{CT}^{3SPT}$  numerically result in different values, though the sums (16) and (18) are equal to each other (practically, because of the different conversion for  $\psi^{(2)}$  and  $\psi^{(1)}$ , they are slightly different, in most cases, less than  $1 \text{ kJ mol}^{-1}$ ). In the present code, when  $\psi^{(2)}$  is calculated for  $E^{4SPT}$ , the contributions by both definitions are evaluated. Otherwise, only  $E_{LE}^{3SPT-2}(X)$  and  $E_{CT}^{3SPT-2}(X \rightarrow Y)$  are calculated. In (17), the splitting of the sum to the local excitations and charge-transfer excitations is unique. On the other hand, the definition (18) contains some ambiguity in the terms; the sum for  $E_{LE}^{3SPT-2}(X)$  includes various types of excitations  $| (j_Y \rightarrow t_Z)^S \rangle$  at the kets. In almost all of the numerical test calculations,  $E_{LE}^{3SPT-1}(X)$  is less than  $0.0 \text{ kJ mol}^{-1}$ , but  $E_{LE}^{3SPT-2}(X)$  is not zero as shown later. Note that the CT energy defined in this subsection differs from that defined by Khanliullin et al [21].

### 2.4 The population analysis

In the first paper of our works [3], it was analytically proved under the Mulliken population analysis of the LP SCF wave function that the number of electrons for each molecular unit in the cluster is equal to that of the corresponding monomer and no charge-transfer between the molecular units is allowed. The perturbation expansion of the LP SCF wave function makes correction of this deficiency in the theory. In this subsection, the formulae for the Mulliken population analysis are derived for the first order perturbation wave function. Because the excited MOs are orthogonal to all of the occupied MOs of every molecular unit, the first order correction  $\rho^{(1)}(r)$  to the electron density is zero. To evaluate the 2nd order correction  $\rho^{(2)}(r)$  to the electron density, because of the non-orthogonality

among the occupied MOs and among the excited MOs, a proper normalization of the electron density is required as

$$\begin{aligned} \rho^{(2)}(\mathbf{r}) &= \int \psi^{(1)}(r_2, r_3, \dots, r_N; r) \psi^{(1)}(r_2, r_3, \dots, r_N; r) dr_2 dr_3, \dots, dr_N \\ &\quad - \rho^{(0)}(\mathbf{r}) \int \psi^{(1)}(r_1, r_2, r_3, \dots, r_N) \psi^{(1)}(r_1, r_2, r_3, \dots, r_N) dr_1 dr_2 dr_3, \dots, dr_N \end{aligned} \quad (21)$$

It can be written in terms of basis functions  $\{|\chi_p(\mathbf{r})\rangle\}$  as

$$\rho^{(2)}(\mathbf{r}) = \sum_{p,q} |\chi_p(\mathbf{r})\rangle \left[ \sum_{i,r} \sum_{j,t} \frac{a_{i,r}^{(1)} a_{j,t}^{(1)}}{\sqrt{[\mathcal{S}_o^{-1}]_{ii} [\mathcal{S}_o^{-1}]_{jj}}} \left\{ [\mathcal{S}_o^{-1}]_{i,j} t_{pr} t_{qt} - S_{tr} t_{pj} t_{qi} \right\} \right] \langle \chi_q(\mathbf{r}) | \quad (22)$$

where  $t_{ps}$  ( $s = r$  or  $t$ ) is the coefficient for the excited MOs, and  $t_{q\bar{c}}$  ( $\bar{c} = \bar{a}$  or  $\bar{b}$ ) is defined in (5) for the biorthogonal occupied MOs. It can be proved that the integral  $\int \rho^{(2)}(\mathbf{r}) d\mathbf{r} = \mathbf{0}$ .

The Mulliken population [26] for "molecule  $X$ " in a cluster can be defined as

$$\begin{aligned} &\int \rho^{(2)}(\mathbf{r}) d\mathbf{r} \\ &= \sum_X \sum_{p \in X} \left[ \sum_q^{all} S_{pq}^{AO} \sum_{i,r} \sum_{j,t} \frac{a_{i,r}^{(1)} a_{j,t}^{(1)}}{\sqrt{[\mathcal{S}_o^{-1}]_{ii} [\mathcal{S}_o^{-1}]_{jj}}} \left\{ [\mathcal{S}_o^{-1}]_{i,j} t_{pr} t_{qt} - S_{tr} t_{pj} t_{qi} \right\} \right] \\ &\equiv \sum_X G_X^{2Mull} \end{aligned} \quad (23)$$

By defining the intermediate matrices

$$\mathcal{D}_{p,i}^T \equiv \sum_r^{exc} \frac{a_{ir}^{(1)}}{\sqrt{[\mathcal{S}_o^{-1}]_{ii}}} t_{p,r} \quad (24)$$

$$\mathcal{E}_{p,t}^T \equiv \sum_i^{occ} \frac{a_{i,t}^{(1)}}{\sqrt{[\mathcal{S}_o^{-1}]_{ii}}} t_{p,i} \quad (25)$$

the AO density matrices are calculated as

$$\mathcal{D}_{p,q}^{AO} \equiv \sum_{j,i} \mathcal{D}_{p,i}^T [\mathcal{S}_o^{-1}]_{i,j} \mathcal{D}_{q,j}^T \quad (26)$$

$$\mathcal{E}_{p,q}^{AO} = \sum \mathcal{E}_{p,t}^T S_{tr}^{MO} \mathcal{E}_{p,r}^T \quad (27)$$

Thus, the atomic population is for atom  $K$ ,

$$G_K^{2Mull} = \sum_{p \in K} \sum_q^{all} (S_{p,q}^{AO} \mathcal{D}_{p,q}^{AO} - S_{p,q}^{AO} \mathcal{E}_{p,q}^{AO}) \quad (28)$$

and the molecular population is for molecule  $X$ ,

$$G_X^{2Mull} = \sum_{K \in X} G_K^{2Mull} \quad (29)$$

The first term of (28) is the increase of the electron density, and the second term is the decrease of the electron density; we can define the matrices

$$\begin{aligned}
 G_K^{2Mul} &= \sum_{p \in K} \left[ \sum_L \sum_{q \in L} S_{p,q}^{AO} \mathcal{D}_{p,q}^{AO} - \sum_L \sum_{q \in L} S_{p,q}^{AO} \mathcal{E}_{p,q}^{AO} \right] \\
 &\equiv \sum_{p \in K} \sum_L \left\{ F_{L,p}^{from} - T_{L,p}^{to} \right\} \equiv \sum_L \left\{ F_{L,K}^{from} - T_{L,K}^{to} \right\}
 \end{aligned} \tag{30}$$

With these matrices, the molecular population change  $G_X^{Mul}$  can be defined as

$$\begin{aligned}
 G_X^{Mul} &= \sum_Y \left\{ \sum_{K \in X} \sum_{L \in Y} F_{L,K}^{from} \right\} - \sum_Y \left\{ \sum_{K \in X} \sum_{L \in Y} F_{L,K}^{to} \right\} \\
 &\equiv \sum_Y \left( G_{Y,X}^{from} - G_{Y,X}^{to} \right)
 \end{aligned} \tag{31}$$

### 3 Computational Results

Figures 2 and 3 are the isomer dependence of the hydrogen bond energies (a) for water clusters  $(\text{H}_2\text{O})_n$  and (b) for hydrogen fluoride clusters  $(\text{HF})_n$ . The geometries for  $(\text{H}_2\text{O})_n$  were determined by Ohno et al [27] with B3LYP/6-311+(d,p) and those of  $(\text{HF})_n$  are those reported by Kamiya et al [28] with MP2/cc-pVDZ. The basis sets used are aug-cc-pVDZ in Figure 2 and cc-pVDZ in Figure 3. The ordinate axis is the binding energy per water for  $(\text{H}_2\text{O})_n$  and per hydrogen bond for  $(\text{HF})_n$ . Figure 2 clearly demonstrates that with aug-cc-pVDZ the third order corrected energy  $E^{3BE}$  evaluated with all excited MOs and with a loose scaling parameter set [0.8, 0.6] are close to the CP corrected energy within 0.5 kJ/mol for all of the examined isomers of  $(\text{H}_2\text{O})_n$  and  $(\text{HF})_n$ . Note that the strength of the hydrogen bond is very different among the isomers. To evaluate the CP correction for water septamer  $(\text{H}_2\text{O})_7$  with aug-cc-pVDZ, the SCF calculations have to be performed 8 times with 301 basis sets and 7 times with 43 basis sets. Here the conventional CP procedure, which is an extension of the Boys and Bernardi procedure to the multi-component system, is used. On the other hand, the present LP SCF + 3rd order SPT calculation requires only one LP SCF calculation, and as Figure 1 shows, the SPT calculation is completed less than in 10% of the LP SCF step.

For both isomers, the restriction on the excited MOs to ALEx MOs underestimates the binding energy for large clusters, particularly for those of stronger hydrogen bond systems. This is because in a congested clusters, the enforced condition to be orthogonal to all of the occupied MOs is too strict. The main correction terms in the LP SCF 2nd and 3rd order SPT calculations are the pair type charge-transfer excitations such as molecules  $X \rightarrow Y$ , but to determine the ALEx MO of molecule  $Y$ , the orthogonal condition to the occupied MOs of the molecules other than  $X$  and  $Y$  is imposed. Therefore, the ALEx MO space becomes too small to describe the  $X \rightarrow Y$  charge-transfer interaction.

Figure 3 shows the isomer dependence for a small basis set, cc-pVDZ. As is well known, the BSSE is very large for this size of basis sets even in the SCF level of theory. The difference of the filled square and filled circle points in Figure 3 is BSSE. If all of the excited MOs are used, the calculated  $E^{3BE}$  is close to the uncorrected SCF binding energy. It implies [4], [7] that the charge-transfer terms taken in the 2nd and 3rd order SPT re-introduce the BSSE when the partially delocalized MOs are used in the perturbation expansion. To avoid this difficulty, in the previous paper [7], we applied the cut-off threshold  $\eta^{th}$  for removing some of the partially delocalized excited MOs derived from the vectors of small  $\eta$ . In this work, the scaling of the MO integrals is examined. The plots of both parameter sets

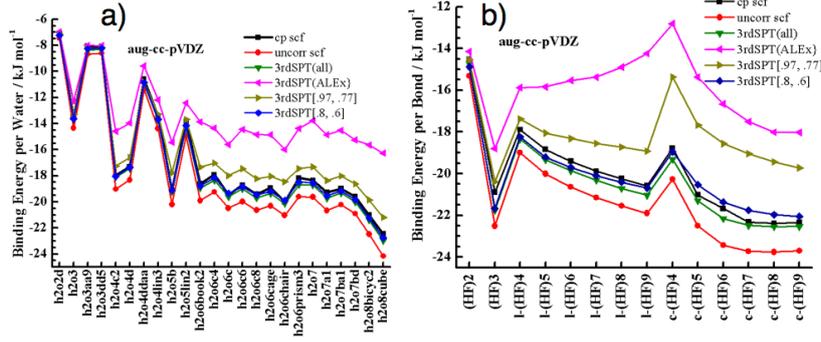


Figure 2: Comparison of the third order LP SPT binding energy  $E^{3BE}$  using the scaled MO integrals with the CP corrected SCF energy (filled square) for various isomers of water (a) and HF (b) clusters. The parameters  $[\eta_{1/2}, \eta_w]$  for scaling is  $[0.98, 0.78]$  (leftward arrow) and  $[0.80, 0.60]$  (rhombus). The other plots are those of the uncorrected SCF (filled circle), those of the third order LP SPT using all excited MOs (downward triangle) and using only ALEx MOs (rightward triangle). The basis set used is aug-cc-pVDZ.

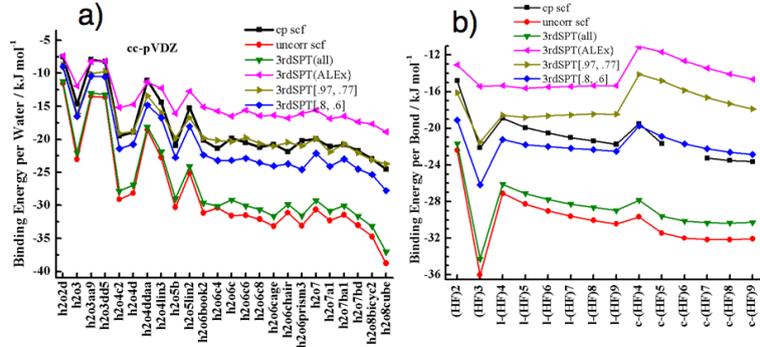


Figure 3: Comparison of the third order LP SPT binding energy using the scaled MO integrals with counterpoise corrected SCF energy for various isomers of water (a) and HF (b) clusters. The symbols are same in Figure 2. The basis set used is cc-pVDZ.

of scaling parameters,  $[0.97, 0.77]$  (leftward triangle) and  $[0.80, 0.60]$  (rhombus), are much closer to those of the CP corrected energy than those of all excited MO and ALEx MO calculations. For water clusters the plots of  $[0.97, 0.77]$  are uniformly close to the CP plots, while for  $(\text{HF})_n$  those are worse than those of  $[0.80, 0.60]$ . Thus, the proper set of scaling parameters is system-dependent and also basis-dependent. Further tuning is not attempted in this study. When the direct MD calculations are carried out with a small basis set, the tuning of the scaling parameters might be necessary.

In Figures 4 and 5, the potential energy curves of  $(\text{H}_2\text{O})_2$  and  $(\text{HF})_2$  are shown. For each O-O and F-F distance, the other geometric parameters are optimized with MP2/aug-cc-pVDZ. Figure 4 clearly demonstrates that with aug-cc-pVDZ basis set the scaling is unnecessary; the curves obtained using all of the excited MOs are close to the CP corrected curves; the error are at most  $0.2 \text{ kJ mol}^{-1}$  both for  $(\text{H}_2\text{O})_2$  and  $(\text{HF})_2$  (note that in Figure 4(b) the curve of the downward triangle is overlapped with the curve of the rhombus). In this basis set, the procedure to determine the excited MOs for the dimers is appropriate and does not introduce the BSSE.

Figure 5 also demonstrates that when all excited MOs are used in the 3rd order SPT, the uncorrected SCF potential energy curves for water and HF dimers are recovered at the full range of the bond length. With this size of the basis sets, the charge-transfer terms cannot be separated from the

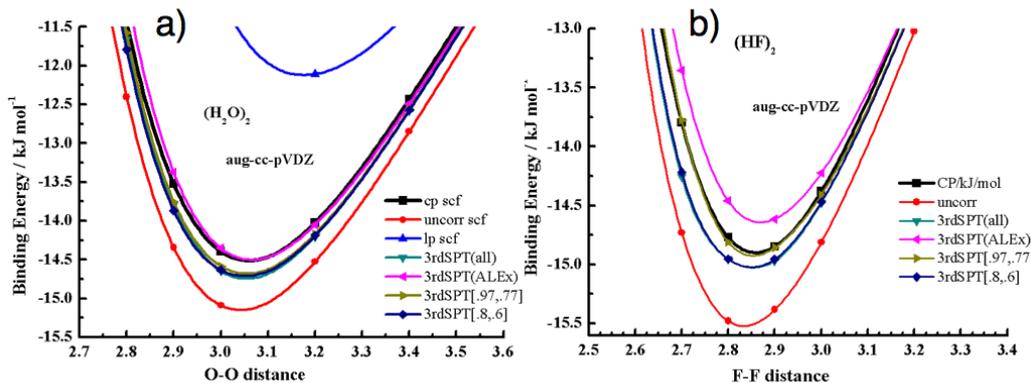


Figure 4: Comparison of the potential energy curve of water dimer (a) and HF dimer (b) calculated by the third order LP SPT using the scaled MO integrals with that of the CP corrected SCF. The basis set is aug-cc-pVDZ. The symbols are same with Figure 2.

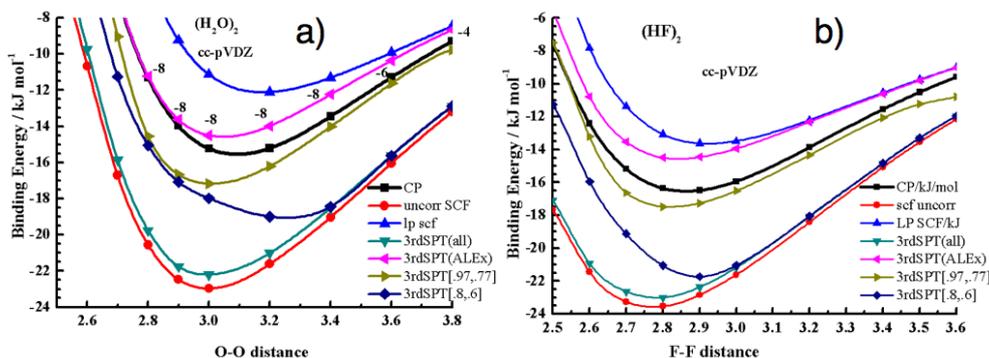


Figure 5: Comparison of the potential energy curve of water dimer (a) and HF dimer (b) calculated by the third order LP SPT using the scaled MO integrals with that of the CP corrected SCF. The basis set is cc-pVDZ. The symbols are same with Figure 2.

BSSE. If the excited MOs are restricted to the ALEx, the systematic underestimation of the binding energy for both dimers are clearly seen, but the amount of the error and the shape of the curves are not terribly bad for the dimer. The numbers given close to the leftward arrows in Figure 5a are the number of the excluded MOs for the ALEx; the number "-8" implies that 8 orbitals from the 22 excited MOs are excluded in the calculations. By scaling the MO integrals, the contribution from the excluded MOs is partly taken into account. For both dimers, the curve evaluated with the set of strict parameters [0.97, 0.77] behaves better than that the loose parameter set [0.8, 0.6]. The curves of the latter eventually coincides with those evaluated with all excited MOs at long bond lengths. Although the loose parameter set gives the better isomer-dependence for the cyclic HF clusters as shown in Figure 3b, in general the strict parameter sets are safe to be used, because their upper bound is the binding energy evaluated with the ALEx MOs.

Figure 6 shows the basis set dependence of the binding energy of  $\text{F}^-(\text{H}_2\text{O})$  and  $\text{Cl}^-(\text{H}_2\text{O})$ . For both cases, the 3rd order energy  $E^{3BE}$  of aug-cc-pVTZ and aug-cc-pVQZ evaluated with the ALEx MO agrees well with the converged CP energy. The 2nd  $E^{2BE}$  and 4th order  $E^{4BE}$  corrections are also given;  $E^{4BE}$  is almost indistinguishable from  $E^{3BE}$  both for  $\text{F}^-(\text{H}_2\text{O})$  and  $\text{Cl}^-(\text{H}_2\text{O})$ . The difference of  $E^{2BE}$  and  $E^{CP}$  persists even at aug-cc-pVQZ, which prompted us to develop the 3rd order SPT.

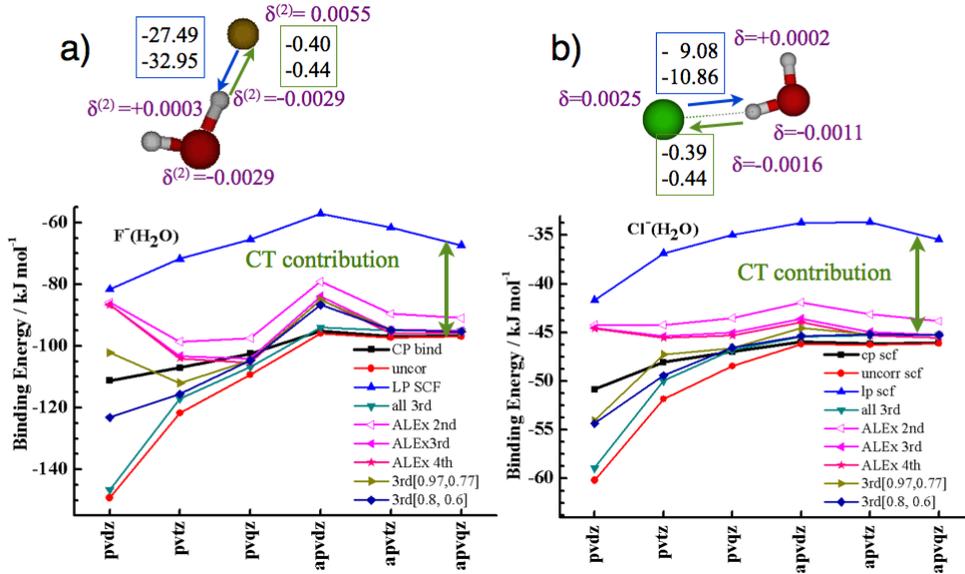


Figure 6: Basis set dependence of the binding energy of  $F^-(H_2O)$  (a) and  $Cl^-(H_2O)$  (b). The symbols are same in Figure 2. The second and fourth order energies evaluated using only ALEX MO are also shown. In the inserted figure, the CT contribution (the upper number in box is for the second order) and the change of Mulliken population are given. They are evaluated with aug-cc-pVTZ using all excited MOs.

The energy evaluated with the LP SCF wave function is also shown (upward triangle) to clearly indicate the charge-transfer energy. Because the LP SCF wave function strictly eliminates the electron delocalization over the molecular units in terms of the basis sets, the difference between this curve and the others is the charge-transfer (CT) energy. As mentioned in section 2.3, the local excitations do not contribute to the binding energy, though they are in the perturbation expansions. The CT energy is a theoretical concept and not an observable; the value is very much dependent on its definition. In the present definition, the reference energy is the energy of the LP SCF wave function which is variationally determined under the strict locality condition of the basis sets. Therefore, as far as the basis sets are placed on the atoms in each molecular unit, the definition of the CT terms is clear and unique. Since the purpose of the present work is not to intend the analysis of the classification of the interaction terms, the comparison with the other methods are not made in this paper. Even with this definition, the CT energy is basis-set-dependent. The difference between the LP SCF and CP SCF energies for  $F^-(H_2O)$  is  $-29.6$  kJ mol<sup>-1</sup> for cc-pVDZ,  $-38.1$  for aug-cc-pVDZ and  $-29.3$  for aug-cc-pVQZ.

In the inserted small box of Figure 6, the 2nd and 3rd order CT contributions (evaluated with aug-cc-pVTZ) are shown. The upper number is the contribution in the 2nd order and the lower one is in a sum of the 2nd and 3rd order. The term from the electron transfer  $F^-$  ( $Cl^-$ ) to  $H_2O$  is  $-32.95$  ( $-10.86$ ) kJ mol<sup>-1</sup>. The back CT contributes to the binding energy only by  $-0.44$  kJ mol<sup>-1</sup> in  $F^-(H_2O)$ . Also the change of the Mulliken atomic and molecular population is given. The amount of the electron transfer in this analysis is unexpectedly small, 0.0055 in  $F^-(H_2O)$  and 0.0025 in  $Cl^-(H_2O)$ .

The change of the Mulliken population and the CT contributions are shown for the cyclic HF clusters in Table 1 and for the linear HF clusters in Table 2. In the cyclic clusters all of HF molecules are equivalent, and so there is no net change of the molecular Mulliken population, but within each molecule, the atomic population is slightly changed; the F atom becomes less negative than in an

	cyc (HF) <sub>4</sub>	cyc (HF) <sub>5</sub>	cyc (HF) <sub>6</sub>	cyc (HF) <sub>7</sub>
$E^{3BE}$ per H bond /kJ mol <sup>-1</sup>	-18.75	-21.19	-22.21	-22.65
$\delta F$	0.0009	0.0011	0.0012	0.0010
LE /kJ mol <sup>-1</sup>	+0.06	+0.07	+0.07	+0.08
CT( $n-1$ ) /kJ mol <sup>-1</sup>	-9.76	-11.47	-11.49	-11.25
CT( $n+1$ ) /kJ mol <sup>-1</sup>	-0.34	-0.38	-0.39	-0.42
CT( $n-2$ ) /kJ mol <sup>-1</sup>	-0.10	-0.15	-0.19	-0.19
CT( $n-3$ ) /kJ mol <sup>-1</sup>		-0.03	0.01	0.00
CT( $n-4$ ) /kJ mol <sup>-1</sup>			-0.04	0.00
CT( $n-5$ ) /kJ mol <sup>-1</sup>				-0.04

Table 1: The change of the Mulliken population on atom F ( $\delta F$ ) and the 3rd contribution of the local excitation (LE) and charge-transfer (CT) terms for the cyclic (HF)<sub>*n*</sub> clusters. The basis set used is aug-cc-pVTZ. All of the excited MOs are used in the perturbation expansion. CT(*m*) stands for CT contribution from the *m* th HF to the *n* th HF

	lin (HF) <sub>4</sub>	lin (HF) <sub>5</sub>	lin (HF) <sub>6</sub>	lin (HF) <sub>7</sub>
$E^{3BE}$ per H bond /kJ mol <sup>-1</sup>	-17.53	-18.62	-19.25	-19.83
$\delta(HF)_1$	+0.0008	+0.0009	+0.0009	+0.0009
$\delta(HF)_n$	-0.0007	-0.0008	-0.0008	-0.0008
CT(1→2) /kJ mol <sup>-1</sup>	-5.36	-5.71	-5.94	-6.14
CT(2→3) /kJ mol <sup>-1</sup>	-6.26	-7.21	-8.01	-8.41
CT(3→4) /kJ mol <sup>-1</sup>	-4.61	-7.20	-8.40	-9.19
CT(4→5) /kJ mol <sup>-1</sup>		-4.87	-7.66	-8.99
CT(5→6) /kJ mol <sup>-1</sup>			-5.05	-7.97
CT(6→7) /kJ mol <sup>-1</sup>				-5.19

Table 2: The change of the Mulliken molecular charge on the end of molecules HF ( $\delta(HF)_1$  and  $\delta(HF)_n$ ) and the 2nd and 3rd order contributions of the charge-transfer terms CT( $n \rightarrow m$ ) for the linear (HF)<sub>*n*</sub> clusters. The numbering is given from the end of the hydrogen acceptor. The basis set used is aug-cc-VTZ. All of the excited MOs are used in the perturbation expansion.

isolated HF molecule; the amount  $\delta F$  of the change is slightly size-dependent. The energy CT( $n-1$ ) stands for the contribution from the CT of the neighboring hydrogen donor HF molecule, and the energy CT( $n+1$ ) stands for the contribution from the CT of the other side of neighboring hydrogen acceptor HF molecule. The energy CT( $n+1$ ) is the "back" CT term. Table 1 shows that the dominant CT term, CT( $n-1$ ), constitutes a substantial part of the hydrogen bond energy; it is nearly equal to a half of the average hydrogen bond energy per bond. The back CT term is small but not negligible. The direct CT contribution from the next neighboring HF molecule is very small, but indirectly the HF molecules other than the neighbors strengthen the hydrogen bond as is seen in the size dependence of CT( $n-1$ ). The change  $\delta F$  in the Mulliken atomic population is tiny and shows a weak correlation with the contribution CT( $n-1$ ).

Table 2 shows the analysis for the chain HF clusters. They are calculated with aug-cc-pVTZ using all of the excited MOs. The numbering of HF molecules in the chain starts at the end of the hydrogen acceptor HF, which is an electron donor as is seen in a positive  $\delta(HF)_1$ . The HF molecule at the other end of the chain is an electron acceptor and  $\delta(HF)_n$  is negative. In these cases, as in F<sup>-</sup>(H<sub>2</sub>O), the amount of the electron transfer in the Mulliken molecular population is very small, and probably it is not suitable to use any model estimation of the change of the electrostatic energy. In Table 2, only the CT terms, CT( $n \rightarrow n+1$ ), of the neighboring hydrogen donor HF molecule is given.

As in the cyclic isomers in Table 1, the longer chain of hydrogen bond network increases the CT terms. The largest CT term in each chain is found at the middle of the chain, which implies that both hydrogen donor and acceptor influences the hydrogen bond strength. Because of this effect, the hydrogen bonds in the cyclic chain shown in Table 1 are stronger than those of the chain isomers. The CT contribution to the hydrogen bond is less than a half of the average bond energy for the chain isomers. The electrostatic and polarization interactions play more important role in the chain isomers than in the cyclic isomers; they are partly cancelled out in the latter isomers.

## 4 Conclusive Remarks

This work shows that at the Hartree-Fock level of the theories the notorious BSSE in estimating weak molecular interaction can be under control even without use of extensive basis sets. The procedure given in section 2.1 can be applied for the studies of properly selected molecular clusters, where the electrostatic, polarization, and charge-transfer terms are dominant and the dispersion term plays a less significant role. The direct Monte Carlo (MC) simulation is the most suitable technique to utilize the method, and the collaborative works are under way. To use the method in the molecular dynamics (MD), the gradient code is needed. It is straightforward for the LP SCF wave function, because the MOs are variationally determined. The gradient codes for the 2nd order SPT might be possible, if all of the excited MOs are used, after some manipulation of the orthogonality condition to the occupied MOs.

To evaluate the dispersion term, the double excitations have to be included in the perturbation expansion. In the earlier work [6], some results are already reported. One of the important findings in the work, the types of the electron configurations in the perturbation expansion have to be carefully selected; if all of the double excitations are included, the calculated binding energy is overestimated. Besides, the developed code based on the configuration state functions was extremely slow, and the test calculations were possible only for the dimers with small basis sets. Now a new code of Slater determinant base, which enables us to evaluate the spin-component scaling (SCS) energy [29], is under development.

## Acknowledgement

The work is partially supported by the Grants-in-Aid for Science Research (No.17550012 and 20550018) of JSPS. The author is appointed as a special professor under the Special Coordination Funds for Promoting Science and Technology from MEXT, Japan in 2004 - 2008. He sincerely thanks Prof. M. Aida of Hiroshima University for giving him an opportunity to return to the research after retirement from Institute for Molecular Science and from National Institute for Academic Degrees and University Evaluation. A part of computations was carried out at the RCCS, Okazaki Research Facilities, National Institutes of Natural Sciences (NINS).

## References

- [1] B. Brutschy and P. Hobza, *Chem. Rev.* 100 (2000) 3861
- [2] I. G. Kaplan, "Intermolecular Interactions: Physical Picture, Computational Methods, and Model Potentials", John Wiley & Sons (2006)
- [3] T. Nagata, O. Takahashi, K. Saito and S. Iwata, *J. Chem. Phys.* 115 (2001) 3553
- [4] T. Nagata and S. Iwata, *J. Chem. Phys.* 120 (2004) 3555
- [5] S. Iwata, *Chem. Phys. Letters*, 431 (2006) 204

- [6] S. Iwata and T. Nagata, *Theor. Chem. Acc.* 117 (2007) 137
- [7] S. Iwata, *J. Phys. Chem. B* 112 (2008) 16104
- [8] P. Hobza, *Annu. Rep. Prog. Chem. Sec. C* 100 (2004) 3
- [9] B. Liu and A. D. McLean, *J. Chem. Phys.* 91 (1989) 2348
- [10] J. A. Pople, *Rev. Mod. Phys.* 71 (1999)
- [11] T. Helgaker, P. Jorgensen, J. Olsen, "Molecular Electronic-Structure Theory", Sec. 4.3, John Wiley & Sons, (2000)
- [12] S. F. Boyes and F. Bernardi, *Mol. Phys.* 19 (1970) 553
- [13] P. Valiron and I. Mayer, *Chem. Phys. Letters*, 275 (1997) 46
- [14] J. C. White and E. R. Davidson, *J. Chem. Phys.* 93 (1990) 8029
- [15] A. J. C. Varandas, *Theor. Chem. Account*, 119 (2008) 511
- [16] P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.* 8 (2006) 1985
- [17] E. Gianinetti, M. Raimondi and E. Tornaghi, *Int. J. Quantum Chem.* 60 (1996) 157
- [18] E. Gianinetti, I. Vandoni, A. Famulari and M. Raimondi, *Adv. Quantum Chem.* 31 (1998) 251
- [19] R. Z. Khaliullin, M. Head-Gordon and A. T. Bell, *J. Chem. Phys.* 124 (2006) 204105
- [20] R. Z. Khaliullin, E. A. Cober, R. L. Lochan, A. T. Bell and M. Head-Gordon, *J. Phys. Chem. A* 111 (2007) 8753
- [21] R. Z. Khaliullin, A. T. Bell and M. Head-Gordon, *J. Chem. Phys.* 128 (2008) 184112
- [22] H. Stoll, G. Wagenblast, and H. Preuss, *Theor. Chim. Acta*, 57 (1980) 169
- [23] A. Fornili, M. Sironi, and M. Raimondi, *J. Mol. Struct. THEOCHEM* 632 (2003) 157
- [24] A. Genoni, K. M. Merz, Jr. and M. Sironi, *J. Chem. Phys.* 129 (2008) 054101
- [25] S. Iwata, *J. Theor. Comp. Chem.* 5 (2006) 819
- [26] R. S. Mulliken, *J. Chem. Phys.* 36 (1962) 3428
- [27] K. Ohno, M. Okimura, N. Akai, Y. Katsumoto, *Phys. Chem. Chem. Phys.* 7 (2005) 3005
- [28] M. Kamiya, S. Hirata and M. Valiev, *J. Chem. Phys.* 128 (2008) 074103
- [29] S. Grimme, *J. Chem. Phys.* 118 (2004) 9095