Theor Chem Acc (2007) 117:541–553
DOI 10.1007/s00214-006-0181-6

REGULAR ARTICLE

Parallelized integral-direct CIS(D) calculations with multilayer fragment molecular orbital scheme

Yuji Mochizuki · Kiyoshi Tanaka · Katsumi Yamashita · Takeshi Ishikawa · Tatsuya Nakano · Shinji Amari · Katsunori Segawa · Tadashi Murase · Hiroaki Tokiwa · Minoru Sakurai

Received: 16 September 2006 / Accepted: 26 October 2006 / Published online: 25 November 2006

Abstract We have developed a parallelized integral-direct code of the perturbative doubles correction for configuration interaction with singles, proposed as CIS(D) by Head-Gordon et al. (Chem Phys Lett 219:21, 1994). The CIS(D) method provides the energy corrections both of the relaxation and differential correlation for the respective CIS excited states. The implementation of CIS(D) is based on our original algorithm for the second-order Møller–Plesset perturbation (MP2) calculations (Mochizuki et al. in Theor Chem Acc 112:442, 2004). There is no need to communicate bulky intermediate data among worker processes of the parallelized execution. This CIS(D) code is then incorporated into a developer version of ABINIT-MP program, in order to improve the overestimation in excitation energies calculated by the CIS method in conjunction with the multilayer fragment molecular orbital scheme (MLFMO-CIS) (Mochizuki et al. in Chem Phys Lett 406:283, 2005). The MLFMO-CIS(D) method is first used in evaluating the lowest $n\pi^*$ excitation energy of the hydrated formaldehyde. The photoactive yellow protein (PYP) is the second target of MLFMO-CIS(D) calculation. Through these applications, it is shown that the CIS(D) correction improves the CIS results favorably.

Keywords Excited states · Fragment molecular orbital · CIS(D) · MP2 · Parallelism · Integral-direct

1 Introduction

A number of biochemical systems, e.g. the photosynthetic center [1], are photoactive through electronic excitations. The central region in photoactive proteins is called as chromophore, and it consists of the photon-absorbing pigment part and some neighbored amino acid residues which should provide the electrostatic or

---

Y. Mochizuki · K. Tanaka
Advancesoft, Center for Collaborative Research,
The University of Tokyo, 4-6-1 Komaba, Meguro-ku,
Tokyo 153-8505, Japan

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

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

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

K. Yamashita · T. Murase
Valway Technology Center, NEC Soft Ltd.,
1-18-7 Shinkiba, Koto-ku, Tokyo 136-8627, Japan

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

M. Sakurai
Center for Biological Resources and Informatics,
Tokyo Institute of Technology, B-62 4259 Nagatsuta-cho,
Midori-ku Yokohama 226-8501, Japan

Present Address:
Y. Mochizuki
Department of Chemistry, Faculty of Science,
Rikkyo University, 3-34-1 Nishi-ikebukuro,
Toshima-ku, Tokyo 171-8501, Japan

hydrogen-bonding interactions with the pigment. Surrounding residues also put some electrostatic influences on the chromophore system probably. These environmental effects should cause the shifts in excitation energies relative to the free pigment, where the positive and negative shifts are usually called as blueshift and redshift, respectively. For example, the opsin shift is an archetype [2]. The similar phenomena have been known for the solvated molecules as the solvatochromism [3]. A sort of hybrid theoretical approach of quantum mechanics (QM) and classical molecular mechanics (MM), denoted as QM/MM, has been frequently used to treat such environmental shifts. Namely, QM or molecular orbital (MO) calculations are applied only to the photoactive molecule and some neighbored areas when needed. Surrounding parts are treated by the MM methods. The ONIOM [4] may be a representative recipe of QM/MM scheme, for example. Although the MM technique makes the inclusion of environmental effects feasible, there exist many prefixed parameters which are empirically adjusted to reproduce the experimental values. The MM description for the hydrogen-bonding, which plays the crucial role for water molecules and amino acids [5], has not yet been well established. The treatment of polarization as a fluctuation from the prefixed values of parameters has been in a similar state. These difficulties lead to an ambiguity in QM/MM calculations. Thus, a full QM treatment would be desirable if possible.

The fragment molecular orbital (FMO) method proposed by Kitaura et al. [6–8] is one of the promising recipes to calculate proteins and solvation clusters in a fully quantum mechanical (QM) manner with reasonable costs of computation. At the Hartree–Fock (HF) level of theory of the FMO scheme, a series of HF calculations is performed for the fragment monomers and dimers under the environmental electrostatic potential (ESP) which is essential to ensure the chemical accuracy. The second-order Møller–Plesset pertubation (MP2) calculation as the simplest correlated method with the size-consistency may be carried out after the respective HF calculations of monomers and dimers. Fedorov et al. [9] have been modifying the GAMESS program [10] for the FMO calculations. Recently, they documented the FMO methodologies and the availability with GAMESS in [11]. Sekino et al. [12] reported some FMO trials with the NWCHEM program [13]. In contrast, the ABINIT-MP program was firstly developed by Nakano et al. [14] as a full-scratch program package for the parallelized FMO-HF calculations of proteins. The fragmentation of bonds in proteins is safely achieved by introducing the special sp^3-hybridized carbon atom called as the bond-detachment atom (BDA) [8,14]. The capability of ABINIT-MP was then enhanced by Mochizuki et al. [15,16] with a parallelized integral-direct implementation of the FMO-MP2 method. Furthermore, Mochizuki et al. [17] implemented the configuration interaction with singles (CIS) [18] in the ABINIT-MP program with an integral-direct parallelism, by accepting the multilayer FMO scheme (MLFMO) [19], where the CIS calculation should be applied to the chromophore region after all the HF calculations of the target system completed. Several illustrative applications performed with ABINIT-MP were documented in [20].

It has been known that the CIS calculations provide semi-quantitative estimates for the low-lying singlet excited states which are predominantly characterized by single excitations, where a typical overestimation in excitation energies is 1–2 eV relative to the experimental values [21]. In fact, such overestimations were obtained by the MLFMO-CIS calculations for the lowest $n\pi^*$ excited state of the hydrated formaldehyde models [17], in comparison with quantitative results reported in the coupled cluster (CC)-based studies [22–24]. The experimental excitation energy of the photoactive yellow protein (PYP), which was first isolated by Meyer from a halophilic bacterium [25], is known as 2.78 eV in blue region of light [26]. However, the MLFMO-CIS value was estimated as high as 4.28 eV. Absolutely, the discrepancies remained in the CIS results of Ref. [17]. The overestimation of CIS energies can be remedied by the second-order perturbative doubles correction (abbreviated as CIS(D)) proposed by Head-Gordon et al. [27]. The CIS(D) method evaluates the energy corrections both from the relaxation and the differential MP2 correlation for CIS excited states. Favorably, the size-consistency is satisfied in the CIS(D) calculations, and its computational cost scales as non-iterative N^5 (where N is the number of basis orbitals) like the MP2 calculation for the ground state [27]. These features of CIS(D) should be promising if the additional correction to the MLFMO-CIS calculations is considered for large scale molecules [17].

In this paper, we report a parallelized integral-direct implementation of the CIS(D) method [27]. Unlike the spin-orbital notation [28] presented originally in Ref. [27], we use a spin-adapted CIS(D) formulation for a separate treatment of the singlet and triplet states as in the previous CIS case [17]. The actual implementation is made through the modification of our parallelized direct MP2 algorithm [15]. No I/O processing and communication of bulky intermediate data is needed during the parallelized execution. In order to augment the MLFMO-CIS ability, the resulting CIS(D) code is then incorporated into a developer version of ABINIT-MP

[14–17,20]. The present work is the first realization that all the integrals associated with CIS(D) calculations are processed ‘on-the-fly’ with parallelism, to author’s knowledge. Remaining of this paper is configured as follows. The formulation of spin-adapted CIS(D) is outlined in Sect. 2, and the corresponding implementation are summarized in Sect. 3. The targets of applications are the same as those in [17]. Namely, the hydrated formaldehyde and the PYP are again employed for the demonstrative calculations, and they are documented in Sects. 4 and 5, respectively. One will see how the CIS(D) correction improves the MLFMO-CIS energies for both examples.

2 Spin-adapted formulation of CIS(D)

Before describing the formulation, we would make some more comments why the CIS(D) method [27] is promising in correcting the results of MLFMO-CIS calculations [17]. The calculations of time-dependent density functional theory (TD-DFT) nowadays become the de facto standard as a quantitative method for excited states [29], where the electron correlation is effectively incorporated in the set of potentials. However, the TD-DFT description could collapse for the charge transfer (CT) excitations with erratically low energies [30]. This situation should cause a difficulty in treating the chromophore consisting of the pigment and neighboring residues (or solvents) since the spurious CT states could appear easily, as exemplified in the TD-DFT treatment of the hydrated s-tetrazine [31]. Although extensive works to overcome such difficulties have been in progress (see [29,32] for example), we would hold a stance that the way of CIS calculation and additional CIS(D) correction is safer or better than that of TD-DFT for the photoactive proteins and solvated systems at the present time.

Head-Gordon et al. [27] discussed that the CIS(D) method can be considered as a second-order approximation to the excited state methods of CC with singles and doubles (CCSD). Oumi et al. [33] examined that the reliability of CIS(D) could degrade in comparison with the CCSD-based reference values if the substantial near-degeneracy of states takes place. The third-order CIS(3) method [34] and also the iterative CIS(D_1) method [35] were thus proposed by Head-Gordon et al. as the improvement over CIS(D). Recently, Hirata [36] developed a partial fourth-order CIS(4_p) method of N^6 scaling, with the help of an automated derivation and implementation system [37] to handle the complicated structure of associated tensor equations. These more sophisticated methods than the original CIS(D) [27] requires much more cost of computation and makes the fully integral-direct implementation with parallelism difficult. This means a potential restriction of the tractable size of molecules. Fortunately, only a few lowest states are of chemical interest for the most of photoactive proteins, and they are typically characterized by the single excitations among a few MOs of the occupied-unoccupied boundary, e.g., the HOMO–LUMO excitation in the PYP [17,26]. It would be expected that the CIS(D) method works well for such cases, as long as the issue of near-degeneracy is not severe [33–35]. Thus, we would adopt the CIS(D) correction [27] to be implemented in a parallelized integral-direct fashion in conjunction with the MLFMO-CIS calculations. Care should be taken in watching the character of target states, of course.

In this paragraph, we briefly summarize the spin-orbital formulation of CIS(D) presented by Head-Gordon et al. [27,35], for later convenience to outline our spin-adapted formulation. The notation of equations is the same as that in Ref. [27]. The indices of *ijk* and *abc* specify the occupied orbitals and unoccupied (or virtual) orbitals, respectively, and the anti-symmetrized two-electron integral (TEI) list [28] is used here: see Head-Gordon’s original paper [27] for more details. A certain CIS state is described as the linear combination of singly excited determinants Φ_i^a from the HF reference determinant Ψ_0,

$$\Psi_{\text{CIS}} = U_1\Psi_0 = \sum_{ia} b_i^a \Phi_i^a, \tag{1}$$

once the eigenvalue equation

$$\begin{aligned}\left\langle \Phi_i^a | H - E^{\text{HF}} | U_1\Phi_0 \right\rangle &= \sum_{jb} \left\langle \Phi_i^a | H - E^{\text{HF}} | \Phi_j^b \right\rangle b_j^b \\ &= \omega b_i^a,\end{aligned} \tag{2}$$

$$\left\langle \Phi_i^a | H - E^{\text{HF}} | \Phi_j^b \right\rangle = \delta_{ij}\delta_{ab}\Delta_i^a + (aj||ib), \tag{3}$$

$$\Delta_i^a = \varepsilon_a - \varepsilon_i, \tag{4}$$

is solved to obtain the excitation energy relative to the HF ground state energy [18]. Namely, U_1 presents all the single excitations with amplitudes b_i^a for the state of interest. The CIS(D) energy correction is then given by the following contractions:

$$\omega^{\text{CIS(D)}} = -\frac{1}{4}\sum_{ijab} \left|u_{ij}^{ab}\right|^2 / (\Delta_{ij}^{ab} - \omega) + \sum_{ia} b_i^a v_i^a, \tag{5}$$

$$u_{ij}^{ab} = \sum_{c}\left[(ab||cj)b_i^c - (ab||ci)b_j^c\right] + \sum_{k}\left[(ka||ij)b_k^b - (kb||ij)b_k^a\right], \tag{6}$$

$$\Delta_{ij}^{ab} = \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j, \tag{7}$$

$$v_i^a = \frac{1}{2}\sum_{jkbc}(jk||bc)\left[b_i^b a_{jk}^{ca} + b_j^a a_{ik}^{cb} + 2b_j^b a_{ik}^{ac}\right], \tag{8}$$

where the coefficients a_{ij}^{ab} are just the MP2 amplitudes due to the perturbation potential V

$$E^{\mathrm{MP2}} = \langle\Phi_0|V|T_2\Phi_0\rangle = \frac{1}{4}\sum_{ijab} a_{ij}^{ab}(ij||ab), \tag{9}$$

$$T_2\Phi_0 = \frac{1}{4}\sum_{ijab} a_{ij}^{ab}\Phi_{ij}^{ab}, \tag{10}$$

$$a_{ij}^{ab} = -\left\langle\Phi_{ij}^{ab}|V|\Phi_0\right\rangle/\Delta_{ij}^{ab} = -(ab||ij)/\Delta_{ij}^{ab}. \tag{11}$$

The operator T_2 generates the double excitations to introduce the electron correlation [28]. The first term in the right-hand side of Eq. (5) should correspond to the relaxation energy which originates from $\langle\Psi_{\mathrm{CIS}}|V|U_2\Psi_0\rangle$, being a negative quantity as can be seen from the definition like MP2. Although the operator U_2 is also of double excitation type, it can be triplet unlike T_2 but like U_1. In contrast, the second term in Eq. (5) should provide the differential MP2 correlation energy given by $\langle\Psi_{\mathrm{CIS}}|V|T_2U_1\Psi_0\rangle_c$ (*connected part*): the *disconnected part* is just the MP2 energy of Eq. (9). This term may be a positive quantity since a hole is created in the occupied space by the single excitation. The computational cost to evaluate Eq. (8) is seemingly of N^6. But, it can be reduced to N^5 through the factorizations [35,38]

$$v_i^a = \sum_{b} w_{ab}b_i^b + \sum_{j} w_{ij}b_j^a + \sum_{kc} w_{kc}a_{ik}^{ac}, \tag{12}$$

$$w_{ab} = \frac{1}{2}\sum_{jkc}(jk||bc)a_{jk}^{ca}, \tag{13}$$

$$w_{ij} = \frac{1}{2}\sum_{kbc}(jk||bc)a_{ik}^{cb}, \tag{14}$$

$$w_{kc} = \sum_{jb}(jk||bc)b_j^b. \tag{15}$$

It is notable that w_{ij} and w_{ab} are independent of both state and spin or reusable once constructed.

Now, we outline the spin-adapted formulation of CIS(D) as below. First, the so-called configuration state functions (CSFs) are introduced to treat the singlet and triplet states separately. They are defined by the products of hole creation operator ξ_i and particle creation operator η_a as

$${}^{(1)}\tilde{\Phi}_i^a = \frac{1}{\sqrt{2}}(\eta_a\xi_i + \eta_{\bar{a}}\xi_{\bar{i}})\Phi_0, \tag{16}$$

$${}^{(3)}\tilde{\Phi}_i^a = (\eta_a\xi_{\bar{i}})\Phi_0, \tag{17}$$

where the bar on indices of i and a means β spin and the tilde on Φ clarifies the spin-adaptation. The corresponding CIS Hamiltonian matrices of singlet and triplet [17] are given, respectively, by

$$\left\langle{}^{(1)}\tilde{\Phi}_i^a|H - E^{\mathrm{HF}}|{}^{(1)}\tilde{\Phi}_j^b\right\rangle = \delta_{ij}\delta_{ab}\Delta_i^a + 2(ia,jb) - (ij,ab), \tag{18}$$

$$\left\langle{}^{(3)}\tilde{\Phi}_i^a|H - E^{\mathrm{HF}}|{}^{(3)}\tilde{\Phi}_j^b\right\rangle = \delta_{ij}\delta_{ab}\Delta_i^a - (ij,ab). \tag{19}$$

The TEI list appeared in Eqs. (18) and (19) are written in the Mulliken convention for spatial MOs [28], and this notation is used hereafter. With the completion of diagonalization for the CIS matrices, the operator U_1 is rewritten in the spin-adapted form as

$${}^{(1)}\tilde{U}_1 = \sum_{ia}{}^{(1)}\tilde{b}_i^a\left[\frac{1}{\sqrt{2}}(\eta_a\xi_i + \eta_{\bar{a}}\xi_{\bar{i}})\right], \tag{20}$$

$${}^{(3)}\tilde{U}_1 = \sum_{ia}{}^{(3)}\tilde{b}_i^a(\eta_a\xi_{\bar{i}}). \tag{21}$$

Next, the spin-adapted form of T_2 operator is defined with five types of doubly excited CSFs as

$$\tilde{T}_2\Phi_0 = \sum_{ia}\tilde{a}_{ii}^{aa}\tilde{\Phi}_{ii}^{aa} + \sum_{i,a>b}\tilde{a}_{ii}^{ab}\tilde{\Phi}_{ii}^{ab} + \sum_{i>j,a}\tilde{a}_{ij}^{aa}\tilde{\Phi}_{ij}^{aa} + \sum_{i>j,a>b}\left[\tilde{a}_{ij}^{ab(\mathrm{SS})}\tilde{\Phi}_{ij}^{ab(\mathrm{SS})} + \tilde{a}_{ij}^{ab(\mathrm{TT})}\tilde{\Phi}_{ij}^{ab(\mathrm{TT})}\right], \tag{22}$$

where the fourth summation has two spin-coupling paths of singlet–singlet (denoted as superscripts of '(SS)') and triplet–triplet ('(TT)') for the $ij \rightarrow ab$ excitation, with respect to the coupling order from 'hole pair' to 'particle pair' [39,40]. Table 1 lists the expressions with the creation operators of holes and particles for the CSFs of $\tilde{T}_2$. The spin-coupling of the creation operators involving $\tilde{\Phi}_{ij}^{ab(\mathrm{SS})}$ and $\tilde{\Phi}_{ij}^{ab(\mathrm{TT})}$ does not directly correspond to the genealogical type [41], although the spin space spanned is totally the same. The corresponding MP2 amplitudes of $\tilde{T}_2$ are included in Table 1, where they were already used to implement the partially renormalized MP2 method [42] in [16] actually. The singlet operator ${}^{(1)}\tilde{U}_2$ has formally the same list of hole-particle operators as $\tilde{T}_2$. But the triplet operator ${}^{(3)}\tilde{U}_2$ has

Table 1 Operator expression of $\tilde{T}_2$ CSFs and corresponding MP2 amplitude given by $[X_1(ia,jb)+X_2(ib,ja)]/\Delta_{ij}^{ab}$

No.	CSF[a]	Operators[b]	X_1	X_2	Δ_{ij}^{ab}
1	$\tilde{\Phi}_{ii}^{aa}$	$\eta_a\eta_{\bar{a}}\xi_i\xi_{\bar{i}}$	-1	0	$2\varepsilon_a-2\varepsilon_i$
2	$\tilde{\Phi}_{ii}^{ab}$	$\frac{1}{\sqrt{2}}(\eta_a\eta_{\bar{b}}-\eta_{\bar{a}}\eta_b)\xi_i\xi_{\bar{i}}$	$-\sqrt{2}$	0	$\varepsilon_a+\varepsilon_b-2\varepsilon_i$
3	$\tilde{\Phi}_{ij}^{aa}$	$\frac{1}{\sqrt{2}}\eta_a\eta_{\bar{a}}(\xi_i\xi_{\bar{j}}-\xi_{\bar{i}}\xi_j)$	$-\sqrt{2}$	0	$2\varepsilon_a-\varepsilon_i-\varepsilon_j$
4	$\tilde{\Phi}_{ij}^{ab\text{(SS)}}$	$\frac{1}{2}(\eta_a\eta_{\bar{b}}-\eta_{\bar{a}}\eta_b)(\xi_i\xi_{\bar{j}}-\xi_{\bar{i}}\xi_j)$	-1	-1	$\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j$
5	$\tilde{\Phi}_{ij}^{ab\text{(TT)}}$	$\frac{1}{\sqrt{3}}[\eta_a\eta_b\xi_i\xi_j+\eta_{\bar{a}}\eta_{\bar{b}}\xi_{\bar{i}}\xi_{\bar{j}}$ $+\frac{1}{2}(\eta_a\eta_{\bar{b}}+\eta_{\bar{a}}\eta_b)(\xi_i\xi_{\bar{j}}+\xi_{\bar{i}}\xi_j)]$	$-\sqrt{3}$	$\sqrt{3}$	$\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j$

[a] Excitation is defined by spatial MOs. There are singlet–singlet and triplet–triplet couplings for $ij \to ab$ excitation, denoted by superscripts as (SS) and (TT), respectively

[b] Acting on Ψ_0 from the left side

Table 2 Operator expression for ${}^{(3)}\tilde{U}_2$ CSFs

No.	CSF[a]	Operators[b]
1	${}^{(3)}\tilde{\Phi}_{ii}^{ab}$	$\eta_a\eta_b\xi_i\xi_{\bar{i}}$
2	${}^{(3)}\tilde{\Phi}_{ij}^{aa}$	$\eta_a\eta_{\bar{a}}\xi_{\bar{i}}\xi_{\bar{j}}$
3	${}^{(3)}\tilde{\Phi}_{ij}^{ab\text{(ST)}}$	$\frac{1}{\sqrt{2}}\eta_a\eta_b(\xi_i\xi_{\bar{j}}-\xi_{\bar{i}}\xi_j)$
4	${}^{(3)}\tilde{\Phi}_{ij}^{ab\text{(TS)}}$	$\frac{1}{\sqrt{2}}(\eta_a\eta_{\bar{b}}-\eta_{\bar{a}}\eta_b)\xi_{\bar{i}}\xi_{\bar{j}}$
5	${}^{(3)}\tilde{\Phi}_{ij}^{ab\text{(TT)}}$	$\frac{1}{2}[(\eta_a\eta_{\bar{b}}+\eta_{\bar{a}}\eta_b)\xi_{\bar{i}}\xi_{\bar{j}}+\eta_a\eta_b(\xi_i\xi_{\bar{j}}+\xi_{\bar{i}}\xi_j)]$

Expressions for ${}^{(1)}\tilde{U}_2$ involving ${}^{(1)}\tilde{\Phi}_{ii}^{aa}$, ${}^{(1)}\tilde{\Phi}_{ii}^{ab}$, ${}^{(1)}\tilde{\Phi}_{ij}^{aa}$, ${}^{(1)}\tilde{\Phi}_{ij}^{ab(SS)}$, and ${}^{(1)}\tilde{\Phi}_{ij}^{ab(TT)}$ are the same as those of $\tilde{T}_2$ (refer to Table 1)

[a] Excitation is defined by spatial MOs. There are singlet–triplet, triplet–singlet, and triplet–triplet couplings for $ij \to ab$ excitation, denoted by superscripts as (ST), (TS), and (TS), respectively

[b] Acting on Ψ_0 from the left side

another list of operators for CSFs [39,40], as shown in Table 2. For the triplet case, there are three spin-coupling paths for the $ij \to ab$ excitation as '(ST)', '(TS)', and '(TT)', and the $ii \to aa$ excitation of closed-shell type disappears. As a whole, the CIS(D) correction for single state may be written in the spin-adapted notation as

$$\omega^{\text{CIS(D)}} = E^{\langle\tilde{U}_2\rangle} + E^{\langle\tilde{T}_2\tilde{U}_1\rangle}, \tag{23}$$

where the first and second terms correspond to the relaxation energy by $\tilde{U}_2$ and the differential MP2 correlation energy by $\tilde{T}_2\tilde{U}_1$, respectively. The gross amount of energy correction is determined by the balance of these two terms.

The relaxation energy for a certain singlet state, $E^{\langle{}^{(1)}\tilde{U}_2\rangle}$, is given as

$$\begin{aligned}E^{\langle{}^{(1)}\tilde{U}_2\rangle} =& -\sum_{ia}|{}^{(1)}\tilde{u}_{ii}^{aa}|^2/(\Delta_{ii}^{aa}-\omega)\\ &-\sum_{i,a>b}|{}^{(1)}\tilde{u}_{ii}^{ab}|^2/(\Delta_{ii}^{ab}-\omega)\\ &-\sum_{i>j,a}|{}^{(1)}\tilde{u}_{ij}^{aa}|^2/(\Delta_{ij}^{aa}-\omega)-\sum_{i>j,a>b}\big[|{}^{(1)}\tilde{u}_{ij}^{ab\text{(SS)}}|^2\\ &+|{}^{(1)}\tilde{u}_{ij}^{ab\text{(TT)}}|^2\big]/(\Delta_{ij}^{ab}-\omega).\end{aligned} \tag{24}$$

The formula for $E^{\langle{}^{(3)}\tilde{U}_2\rangle}$ is similar with this equation, except for the difference in the spin-couplings shown in Table 2. The matrix elements of the energy denominator of Eq. (24) are evaluated by manipulating the hole-particle operators [39,40]. The resultant expressions of integral contraction are somewhat complicated, in comparison with Eq. (6) in the spin-orbital notation [27]. For example, the expressions of ${}^{(1)}\tilde{u}_{ii}^{ab}$ and ${}^{(1)}\tilde{u}_{ij}^{ab\text{(SS)}}$ are written, respectively, as

$$\begin{aligned}{}^{(1)}\tilde{u}_{ii}^{ab} =& -\sum_c[(ia,bc)+(ib,ac)]{}^{(1)}\tilde{b}_i^c\\ &+\sum_k(ia,ik){}^{(1)}\tilde{b}_k^b+\sum_k(ib,ik){}^{(1)}\tilde{b}_k^a,\end{aligned} \tag{25}$$

$$\begin{aligned}{}^{(1)}\tilde{u}_{ij}^{ab\text{(SS)}} =& -\frac{1}{\sqrt{2}}\sum_c[(jb,ac)+(ja,bc)]{}^{(1)}\tilde{b}_i^c\\ &-\frac{1}{\sqrt{2}}\sum_c[(ib,ac)+(ia,bc)]{}^{(1)}\tilde{b}_j^c\\ &+\frac{1}{\sqrt{2}}\sum_k[(ja,ik)+(ia,jk)]{}^{(1)}\tilde{b}_k^b\\ &+\frac{1}{\sqrt{2}}\sum_k[(jb,ik)+(ib,jk)]{}^{(1)}\tilde{b}_k^a.\end{aligned} \tag{26}$$

The separation of intra- and inter-orbital relaxation energies, which are labeled by ii and ij, respectively, may be useful to understand the nature of excited states. Some technique should be employed in the actual implementation to compute $E^{\langle\tilde{U}_2\rangle}$, since the handling of (ia,bc) and (ia,jk) lists could be too demanding if the integral-direct processing is required. The details of technique will be discussed in the next section about the implementation.

The evaluation of $E^{\langle\tilde{T}_2\tilde{U}_1\rangle}$ concerns three intermediate matrices about $\tilde{v}_i^a$, whose expression in the spin-notation [27,35] are given as Eqs. (13), (14) and (15). The matrices of $\tilde{w}_{ab}$ and $\tilde{w}_{ij}$, which are independent of both spin and state, consist of the contributions from five types of CSFs associated with $\tilde{T}_2$. The spin-adapted expressions of integral contraction can be derived through the operator manipulations as for the case of $\tilde{U}_2$. A representative example is the contribution from $\tilde{\Phi}_{ij}^{ab(\mathrm{TT})}$, and it is expressed like

$$\tilde{w}_{ab} \leftarrow \tilde{w}_{ab} + \frac{\sqrt{3}}{2}\sum_{i>j,c}^{(a>c)}[(ib,jc)-(ic,jb)]\tilde{a}_{ij}^{ac(\mathrm{TT})} - \frac{\sqrt{3}}{2}\sum_{i>j,c}^{(c>a)}[(ib,jc)-(ic,jb)]\tilde{a}_{ij}^{ca(\mathrm{TT})}, \tag{27}$$

$$\tilde{w}_{ij} \leftarrow \tilde{w}_{ij} + \frac{\sqrt{3}}{2}\sum_{k,a>b}^{(i>k)}[(ja,kb)-(jb,ka)]\tilde{a}_{ik}^{ab(\mathrm{TT})} - \frac{\sqrt{3}}{2}\sum_{k,a>b}^{(k>i)}[(ja,kb)-(jb,ka)]\tilde{a}_{ki}^{ab(\mathrm{TT})}. \tag{28}$$

The remaining array to be constructed is $\tilde{w}_{kc}$ with spin- and state-dependence. The expressions of integral contraction for the spin singlet and triplet are given, respectively, as

$$^{(1)}\tilde{w}_{kc} = -\sum_{jb}[2(kc,jb)-(kb,jc)]^{(1)}\tilde{b}_j^b, \tag{29}$$

$$^{(3)}\tilde{w}_{kc} = -\sum_{jb}(kb,jc)^{(3)}\tilde{b}_j^b, \tag{30}$$

where the label of state number is omitted for simplicity. The implementation to compute $\tilde{v}_i^a$ will be described in the next section. When the array of $\tilde{v}_i^a$ for all states is available, $E^{\langle\tilde{T}_2\tilde{U}_1\rangle}$ is obtained by the product with the corresponding CIS vector as

$$E^{\langle\tilde{T}_2\tilde{U}_1\rangle} = \sum_{ia}\tilde{b}_i^a\tilde{v}_i^a. \tag{31}$$

The partial summation by fixing an index i provides an orbital-wise differential correlation energy, and this may be useful to see the nature of excited states.

Finally, we address the usage of CIS(D) correction in the MLFMO-CIS context [17]. Prior to the CIS calculation, a series of HF calculations is first performed for the fragment momomers and dimers belonging to both the layer 1 (denoted as L_1) and the layer 2 (L_2). The layer 2 is just the chromophore of which excited states are needed, whereas the layer 1 is the remaining bulky part of the target system. As long as the basis set is common between the layer 1 and layer 2, the corresponding MLFMO–HF energy is evaluated by

$$E^{\mathrm{MLFMO-HF}} = \sum_{X\rangle Y\in L_1,L_2} E^{\mathrm{HF}}_{[XY]} - (N_{\mathrm{Frag.}}-2)\sum_{X\in L_1,L_2} E^{\mathrm{HF}}_{[X]}, \tag{32}$$

where capital X and XY are the fragment momomer and dimer, and $N_{\mathrm{Frag.}}$ means the total number of fragment monomers. When all the HF calculations are complete, the CIS calculation is applied to the layer 2. The CIS(D) calculation is successively performed with the results of CIS. The MLFMO-CIS and MLFMO-CIS(D) total energies are then given, respectively, as

$$E^{\mathrm{MLFMO-CIS}} = E^{\mathrm{MLFMO-HF}} + \omega_{L_2}^{\mathrm{CIS}}, \tag{33}$$

$$E^{\mathrm{MLFMO-CIS(D)}} = E^{\mathrm{MLFMO-CIS}} + \omega_{L_2}^{\mathrm{CIS(D)}}, \tag{34}$$

where $\omega_{L_2}^{\mathrm{CIS}}$ and $\omega_{L_2}^{\mathrm{CIS(D)}}$ are the CIS excitation energy and the CIS(D) energy correction [see Eq. (23)] of the layer 2, respectively. Note that the ESP from the layer 1 onto the layer 2 is fixed at the stage of MLFMO-HF. In other words, the HF charge density of the layer 1 is kept frozen for all the excited states of the layer 2. The layer 2 composition should thus be defined with some care to have a 'margin' around the central pigment. If the size of 'margin' is appropriately set, the polarization effects induced by the excitation of pigment could be taken into account through the CIS(D) calculations.

3 Parallelized direct implementation of CIS(D)

In this section, we outline the implementation of spin-adapted CIS(D) method, based on our parallelized integral-direct MP2 algorithm [15], which can be summarized as below for convenience. With the convention that $pqrs$ specify the atomic orbital (AO) indices, the transformed TEI list needed for the MP2 energy is written as

$$(ia,jb) = \sum_{pqrs} C_{pi}C_{qa}C_{rj}C_{sb}(pq,rs). \tag{35}$$

Actually, this list is generated through the series of quarter transformations

$$(iq,rs) = \sum_p C_{pi}(pq,rs), \tag{36}$$

$$(ia,rs) = \sum_q C_{qa}(iq,rs), \tag{37}$$

$$(ia,js) = \sum_r C_{rj}(ia,rs), \tag{38}$$

$$(ia,jb) = \sum_s C_{sb}(ia,js). \tag{39}$$

The features of our MP2 algorithm [15] are then cited as below.

- There is no I/O processing at the cost of repeated generation of the (pq,rs) list: an 'on-the-fly' algorithm.
- The entire calculation is driven by the i-batch [43], denoted as $[i]$, to adjust the memory requirement to computational resources.
- The standard message-passing interface (MPI) [44] is adopted for parallelization.
- The AO index of s associating with the MO index b is the main parameter of parallelization, where the incompletely transformed list about index s of (ia,jb_s) is the target of MPI_ALL_REDUCE operation.
- There is no communication of bulky intermediate data across worker processes.
- Furthermore, the upper level parallelization is possible by the i-batch in a combinatorial fashion with the lower index s.
- The multiple integral screening is utilized to reduce the operation cost effectively.
- The innermost processing is performed with the standard libraries of the Basic Linear Algebra Subroutines (BLAS) such as DAXPY [45].
- The second-order density matrix can also be directly calculated through the generation of the additional TEI lists of (ia,jk) and (ia,bc) appeared in the occupied-unoccupied response elements.

Reference [16] reported a practical applicability of this algorithm and the resulting code to realize large scale FMO-MP2 calculations with ABINIT-MP.

About the evaluation of $E^{\langle \tilde{T}_2 \tilde{U}_1 \rangle}$, the computation of $\tilde{v}_i^a$ array is the central task. The implementation to construct $\tilde{w}_{ab}$ and $\tilde{w}_{ij}$ is straightforward by modifying the MP2 algorithm/code [15] with the above-summarized features. The computational cost of integral contraction for $\tilde{w}_{ab}$ and $\tilde{w}_{ij}$ is formally of N^5 scaling [recall Eqs. (27) and (28)] as for the transformation itself to prepare the (ia,jb) list. Thus, these contractions are also parallelized with respect to the index i, in the batch loop of MP2 [16]. In contrast, the spin- and state-dependent $\tilde{w}_{kc}$ must be completely constructed before the MP2 processing starts, since this matrix is contracted with the MP2 amplitudes. As suggested by Head-Gordon et al. [35] the Fock-like contraction, which has been employed in AO integral-driven CIS calculations [17,18], is utilized to construct $\tilde{w}_{kc}$. The cost for $\tilde{w}_{kc}$ scales as N^4, and the acceleration through parallelization is obvious [17].

Some technique would be necessary to evaluate $E^{\langle \tilde{U}_2 \rangle}$ if the direct processing of integrals is desired, since both (ia,bc) and (ja,bc) (similarly (ia,jk) and (ja,ik)) are required simultaneously, as shown in Eq. (26). Namely, the simple i-batch processing is not usable. For the evaluation of u_{ij}^{ab} (Eq. (6)), Head-Gordon et al. [35] devised the 'CIS vector-transformed AO-MO coefficients' in their semi-direct implementation of the CIS(D_1) method. We follow this technique and define the modified coefficients

$$\tilde{C}_{sj} = \sum_c C_{sc}\tilde{b}_j^c, \tag{40}$$

$$\tilde{C}_{sb} = \sum_k C_{sk}\tilde{b}_k^b, \tag{41}$$

where the labels of spin and state number are again omitted for simplicity. By using these modified coefficients, the central task in evaluating $E^{\langle \tilde{U}_2 \rangle}$ becomes the preparation of $(ia,j\tilde{b})$, $(ia,\tilde{j}b)$, $(i\tilde{a},jb)$ and $(\tilde{i}a,jb)$. Although there is a similarity to the transformation of (ia,jb) for the MP2 energy, the cost of this preparation is formally $4N_{\text{Vec.}}$ times higher than that for the MP2 case, where $N_{\text{Vec.}}$ is the total number of target vectors depending both on spin and state. The outermost loop of batch, by which the memory fitting is made, can be driven by the canonical ij pair, denoted as $[ij]$. Furthermore, all $N_{\text{Vec.}}$ vectors might be processed at one time, if a sizable memory resource is available. A pretty modification from the original loops for MP2 presented in [15] is thus necessary in the implementation of $E^{\langle \tilde{U}_2 \rangle}$. Figure 1 is the sketch of corresponding loop structure implemented to compute $E^{\langle \tilde{U}_2 \rangle}$ in our spin-adapted CIS(D) code: see Fig. 1 of Ref. [15] for comparison when necessary. The outermost loop of $[ij]$ is parallelizable at the upper level without barrier synchronization. The total cost depends on the number of batches because the generation of (pq,rs) is repeated by just that much. The index s is the parallelization parameter at the lower level. The quarter transformations independently proceed with forming the intermediate integral data in each worker process until calling the MPI_ALL_REDUCE routine [44]. A combinatorial parallelization by the indices of s and ij is possible. The innermost processing is performed by the BLAS libraries of DDOT and DAXPY [45] with the multiple screening, as in the case of regular MP2 calculations. For executions in an environment with small memory resource, we have another version of code in which the outermost loop is driven by $N_{\text{Vec.}}$,

Fig. 1 Outline structure of transformation loops for $E^{\langle\tilde{U}_2\rangle}$

```
Determining the size and times of $ij$-batch depending on memory and $N_{\mathrm{Vec.}}$
Loop over $[ij]$ ! parallelizable at higher level
  Loop over $s$ ! parallelizable at lower level
    Loop over $r$
      Preparing $(pq, rs)$ ! $p \geq q$
      Forming $(iq, rs)$ and $(\tilde{i}q, rs)$ for $N_{\mathrm{Vec.}}$ ! screening, DAXPY
      Forming $(ia, rs)$ and $\{(i\tilde{a}, rs), (\tilde{i}a, rs)\}$ for $N_{\mathrm{Vec.}}$ ! DDOT
      Forming $(ia, js)$ and $\{(ia, \tilde{j}s), (i\tilde{a}, js), (\tilde{i}a, js)\}$ for $N_{\mathrm{Vec.}}$ ! screening, DAXPY
    End of loop over $r$
    Forming $\{(ia, j\tilde{b}_s), (ia, \tilde{j}b_s), (i\tilde{a}, jb_s), (\tilde{i}a, jb_s)\}$ for $N_{\mathrm{Vec.}}$ ! screening, DAXPY
  End of loop over $s$ ! MPI_ALL_REDUCE as barrier
Computing the partial $E^{<\tilde{U}_2>}$ with respect to $[ij]$
End of loop over $[ij]$
```

and the entire computation proceeds vector-by-vector there.

4 Application to hydrated formaldehyde

The developed CIS(D) code with an integral-direct parallelism was incorporated into a developer version of ABINIT-MP program [14–17,20]. We first checked the parallel efficiency of the CIS(D) calculation, by employing the H_2CO+$(H_2O)_{16}$ cluster (abbreviated as FRM+16W) whose geometry was already determined at the HF level in [17] about the CIS calculations. This cluster is a model to mimic the hydrated formaldehyde of which the lowest excited state is characterized by the $n\pi^*$ transition of valence HOMO–LUMO type. The excitation energy itself will be discussed later. The test condition of parallelized CIS(D) calculations was the same as that in [17]. Namely, the 6-31G** basis set [21] was used, where the numbers of basis functions, occupied MOs, and unoccupied MOs were 440, 70 and 352, respectively, under the frozen K-shell restriction. The computational platform was a clustered Intel Dual-Xeon processors (clock rate of 3.06 GHz) with 1000BASE-T ethernet. This resource of computation was of typical in-house class.

We started the efficiency measurement of parallelized CIS(D) calculations on eight processors. Table 3 summarizes the elapsed time (in seconds) and the acceleration relative to the case of eight processors. As addressed in the previous section, our parallelization could be made by both the AO index s and the batch indices of ij or i in a combinatorial fashion. Some possible combinations were thus tried with a symbol like $[P_s \times P_{ij/i}]$ in processor distribution. For the case of eight processors, the parallelization with $[4 \times 2]$ is better than that with $[2 \times 4]$. The parallel efficiency looks good for the processor increase of 8 → 16, where the distribution of $[4 \times 4]$ is recommendable. The efficiency slightly drops for the increase of 16 → 32. With full 64 processors throughout three cases of processor distribution, the job is completed less than an hour. The efficiency of 73% for the total job is acceptable against the double increment of $[4 \times 4] \rightarrow [8 \times 8]$. We consider that the present implementation of parallelized integral-direct CIS(D) calculations is usable with dozens of processors for production runs of photoactive issues, although further tuning of the code would be desirable, especially on the load-balancing.

Hereafter, the lowest $n\pi^*$ excited state of the hydrated formaldehyde is focused on. It has been known that its excitation energy shows the blueshift relative to that of the free formaldehyde as a typical solvatochromism [3]. The estimation of this blueshift has been a challenging target of theoretical calculations. Some studies with correlated wavefunctions are cited as below, as the reference data for our MLFMO-CIS(D) calculations. Martin et al. [46] performed the complete active space self-consistent-field (CASSCF) [47] calculations within the QM/MM framework using classical molecular dynamics (MD) simulation, where the description of water molecules as the solvent was made with a mean-field potential. They obtained the blueshift value of 4.2 kcal/mol (or +0.18 eV) by this simulation. Based on the QM/MM approach of CASSCF and Monte Carlo (MC) configuration sampling, Kawashima et al. [48,49] found the positive relation between the amount of blueshift and the number of water molecules in the first solvation shell and then evaluated a range of 2,540–2,660 cm^{-1} (+0.33 eV). As a highly correlated treatment,

Table 3 Elapsed time [in seconds; timing on Intel Xeon clusters (3.06 GHz clock rate)] and parallel acceleration of CIS(D)/6-31G** calculation for FRM+16W model

Processors[a]	Timing	Acceleration	Relative[b]	Efficiency (%)
8				
[4 × 2]	14168			
[2 × 4]	15800			
16				
[4 × 4]	7613	1.86	[4 × 2]	93
[2 × 8]	7988	1.98	[2 × 4]	99
[8 × 2]	8295	1.71	[4 × 2]	85
32				
[4 × 8]	4666	1.86	[4 × 4]	82
[8 × 4]	4851	1.57	[4 × 4]	78
64				
[8 × 8]	2607	2.92	[4 × 4]	73
[16 × 4]	3005	5.25	[2 × 4]	66
[4 × 16]	3223	4.40	[4 × 2]	55

[a] The way of processor distribution for CIS(D) is shown in a symbol of $[P_s \times P_{ij/i}]$ for the parallelization of s and ij/i, respectively (refer to text)

[b] Acceleration relative to fewer cases of processor distribution

Table 4 Excitation energy (in eV) and blueshift for hydrated formaldehyde models calculated by CIS(D)/6-31G** with and without MLFMO scheme

		CIS		CIS(D)	
MLFMO	Model[a]	Energy	Shift	Energy	Shift
No	FRM[b]	4.75		4.35	
No	FRM+3W	5.01	+0.26	4.43	+0.08
Yes	FRM/16W	5.13	+0.38	4.53	+0.18
Yes	FRM+3W/13W	5.21	+0.46	4.57	+0.22
No	FRM+16W	5.38	+0.63	4.67	+0.32
Yes	FRM+16W/112W	5.38	+0.63	4.66	+0.31

[a] The right side of slash for MLFMO treatemnt presents the number of water molecules in the layer 1 (refer to text)

[b] The geometry was optimized by the HF/6-31++G** procedure in [17]. The reoptimized geometry with HF/6-31G** provides 4.76 eV for CIS and 4.36 eV for CIS(D). The dependence of excitation energies on this basis set effect about the geometry of formaldehyde is found as small as two places of decimals in eV

Kongsted et al. [22,23] utilized the CCSD response [50] calculations combined with a polarizable MM-based dynamics simulation. They estimated the excitation energy of 4.39 eV for the hydrated formaldehyde and the blueshift of 2803±46 cm^{-1} (+0.35 eV). Recently, Hirata et al. [24] presented a conceptually FMO-like pair-interaction method by using the equation-of-motion CCSD (EOM-CCSD) [51] calculation and applied this approximation to several models of the hydrated formaldehyde. With the model employing 81 water molecules, the excitation energy and the blueshift were calculated to be 4.09 eV and 1,360 cm^{-1} (+0.17 eV), respectively. In our previous CIS calculations for several models with and without the MLFMO scheme [17], both the excitation energy and the blueshift for the hydrated formaldehyde were overestimated, in comparison with quantitative results based on the CC treatments [22–24]. The present purpose of application is thus to see that the errors in CIS energies of Ref. [17] are reduced by introducing the contributions from the relaxation and the differential correlation in the CIS(D) calculations.

Table 4 presents the excitation energy obtained by CIS and CIS(D) calculations with and without the MLFMO treatment, where the geometry, the MLFMO modeling and the basis set (6-31G**) were the same as those of Ref. [17] (see this paper for details). One exception was an addition of the FRM/16W model in which only the formaldehyde molecule is set in the layer 2 and 16 water molecules are included in the environmental layer 1. Before discussing the hydrated models, the results of the free formaldehyde are compared with the experimental value of 4.07 eV [49]. The excitation energy is improved from 4.75 eV by CIS to 4.35 eV by CIS(D), but a sizable error is still remained. The geometry was determined at the HF level in [17], and this could be responsible for the discrepancy: see also foot note of Table 4. We thus made a recalculation with the MP2/6-31G** geometry (determined by the Gaussian03 program [21,53]) and obtained 4.09 eV as a better result certainly. On the higher correlation effect for the excitation energies of formaldehyde, the systematic comparisons were already made by Head-Gordon et al. in [27,33,35], showing that the CIS(D) value is comparable to the CCSD-based value for the lowest $n\pi^*$ state. Consequently, we assumed that the issue caused by the HF geometry is transferable for the hydrated case [17] and the evaluation of blueshift is not much affected. In other words, the absolute correspondence in the excitation energy itself is not pursued in the present study.

As can be seen in Table 4, the overestimation of CIS values is corrected by the CIS(D) calculations favorably. It is noteworthy that there is a sizable difference in the reduction of excitation energy from CIS to CIS(D) between FRM and FRM/16W, where the amounts of reduction are 0.40 eV for the former and 0.60 eV for the latter. Namely, more reduction is obtained for FRM/16W incorporating the hydration environment. Only the CIS(D) values are of interest from now on. The blueshift of FRM+3W is as small as +0.08 eV. This suggests that the ‘micro hydration’ with at most three water molecules is insufficient to mimic the hydration of formaldehyde. The results of FRM+3W/13W and FRM/16W models are better. The difference in blueshifts from FRM/16W to FRM+16W implies an importance of the induced polarization or relaxation among the surrounding water mol-

Table 5 Excitation energy (in atomic unit) and decomposition of CIS(D) correction for hydrated formaldehyde models having 16 water molecules

Model	CIS	CIS(D)	Ratio[a]	$E^{\langle\tilde{U}_2\rangle\mathrm{b}}$	$E^{\langle\tilde{T}_2\tilde{U}_1\rangle\mathrm{b}}$	E^{MP2}
FRM	0.1746	0.1597	0.91	−0.1171 (71%)	0.1022 (98%)	−0.3117
FRM/16W	0.1887	0.1666	0.88	−0.1206 (71%)	0.0985 (98%)	−0.3085
FRM+3W/13W	0.1916	0.1678	0.88	−0.1236 (68%)	0.0998 (98%)	−0.9021
FRM+16W	0.1978	0.1717	0.87	−0.1272 (58%)	0.1011 (96%)	−3.5047

[a] Ratio of CIS(D) corrected energy and CIS energy

[b] Value in parenthesis is percentage of contribution from the HOMO

ecules in the layer 2 to which the CIS(D) calculation is applied. Table 5 lists the decomposition of CIS(D) correction for the models having 16 water molecules. A cancellation between the relaxation energy, $E^{\langle\tilde{U}_2\rangle}$, and the differential MP2 correlation energy, $E^{\langle\tilde{T}_2\tilde{U}_1\rangle}$, is obvious for all the cases. The contribution concerning the HOMO leads the relaxation energy, but its amount is decreased against the enlargement of the layer 2. In contrast, the differential correlation energy is kept almost constant and is dominated by the HOMO contribution. This is consistent with the fact that the pair correlation energy is just lost from the HOMO by the excitation.

The FRM+16W model provides +0.32 eV, being very close to +0.31 eV obtained by the largest MLFMO-CIS(D) treatment of FRM+16W/112W. A convergence is here found with 16 water molecules in the layer 2. We would consider that this final value of +0.31 eV is comparable with the reference CC-based values of +0.35 eV by Kongsted et al. [22,23] and of +0.17 eV by Hirata et al. [24], since we have used here only one geometrical configuration for the cluster models. Further investigation combining the MLFMO-CIS(D) method and MD simulations should be necessary if the statistical deviation through configuration sampling is required, as done in [22,23,46,48,49]. Such a combinatorial study would be our future subject.

5 Application to PYP

The PYP system [25] was employed as the second target of application, in order to check again that the CIS(D) correction improves the CIS value suffering from overestimation. This protein contains a total of 125 amino acid residues. The central pigment of PYP is the deprotonated *p*-coumaric acid with the *trans*-form and is linked by a thioester bond with Cys69. Imamoto et al. [26] revealed that the ground state pigment absorbs 446 nm (2.78 eV) blue light and then starts the cycle of photo isomerization. The intense absorption of the ground state in blue should correlate with the vivid yellow. Yoda et al. [54] showed the importance of the positively charged Arg52 as the 'counter ion' to the phenolate part of pigment (negatively charged as $\mathrm{O^-}$-Ph-), at a semi-empirical level of calculations. They also pointed out the presence of hydrogen-bonding network from the coumarate to Glu46, Tyr42 and Thr50 (the farthest). The MLFMO-CIS calculation with 6-31G basis [21] was then performed to estimate the excitation energy, based on the '3PYP' structure archived in the Protein Data Bank (PDB) [55] with some MM refinement for the positions of attached hydrogen atoms. The layer 2 consisted of Arg52, Glu46 and Tyr42, beside the crucial pigment, *p*-coumarate linked with Cys69. The MLFMO-CIS value obtained was as high as 4.28 eV, although the state characterization was successfully made as the HOMO–LUMO transition of π space in the coumarate moiety with a high oscillator strength of 1.52. The overestimation of 1.5 eV from the experimental value of 2.78 eV [26] should be remedied by the CIS(D) correction. The main subject of this section is to confirm this speculation by the actual MLFMO-CIS(D) calculations.

Table 6 shows the MLFMO-CIS excitation energies and the CIS(D) values obtained with the same setting of chromophore and basis set (6-31G) as those in [17]. In this table, the results not only of the lowest singlet state of interest but also of the lowest triplet state is included for comparison. As expected, the overestimated value of 4.28 eV at the CIS level is improved to be 3.29 eV by CIS(D), where the relaxation energy is found to be effective in the gross lowering. A remaining error is 0.5 eV in comparison with the experimental value of 2.78 eV [26], and this discrepancy may be mainly attributed to the issue of geometry being used, as discussed with a free pigment model later. In contrast to the singlet case, the gross correction of CIS(D) slightly raises the excitation energy of the triplet state, which is characterized also by the HOMO–LUMO transition as is the singlet. The differential correlation energy for the triplet is about the same as that for the singlet, whereas the triplet relaxation energy is much smaller than the singlet relaxa-

Table 6 Excitation energy (in eV) and decomposition of CIS(D) correction (in atomic unit) calculated by MLFMO-CIS(D)/6-31G for the lowest state of PYP

	CIS	CIS(D)	Ratio[a]	$E^{\langle \tilde{U}_2 \rangle}$	$E^{\langle \tilde{T}_2 \tilde{U}_1 \rangle}$[b]
Singlet[c]	4.28	3.29	0.77	−0.1430	0.1066
Expt.[d]		2.78			
Triplet[e]	2.47	2.72	1.10	−0.0959	0.1047

[a] Ratio of CIS(D) corrected energy and CIS energy

[b] MP2 correlation energy is −4.4027

[c] Amplitude of the HOMO–LUMO transition is 0.93. Oscillator strength at MLFMO-CIS level is 1.52 [17]

[d] Experimental observation of 2.78 eV by Imamoto et al. [26]

[e] Amplitude of the HOMO–LUMO transition is 0.86

Table 7 Dependence of excitation energy (in eV) on geomtry for pCT$^-$

Basis	Geometry[a]	CIS	CIS(D)
6-31G	HF	4.03	2.95
	MP2	3.87	2.75
6-31G*	HF	4.13	3.08
	MP2	4.01	2.94
Expt.[b]			2.70

[a] Level of geometry optimization

[b] Experimental observation by Nielsen et al. [60]

tion energy. Thus, the loss of correlation energy leads to the gross positive correction. We would note a timing of calculations here. For the lowest singlet state, the entire MLFMO-CIS(D)/6-31G job including the preceding HF and CIS stages took 16 h on 64 Xeon processors, which were employed for the parallel benchmark with the FRM+16W model in the previous section. This demonstrates a practical applicability of our MLFMO-CIS(D) method for realistic photoactive proteins.

In [17], Thr50 was omitted from the layer 2 because of the farthest position from the *p*-coumarate in the hydrogen-bonding network. We made a check calculation of the MLFMO-CIS(D)/6-31G by including Thr50 in the layer 2. The CIS and CIS(D) energies were calculated to be 4.28 and 3.30 eV, respectively. They are essentially the same as those with the original setting, as given in Table 6. The omission of Thr50 in [17] is now justified. To see the effect of a polarization d function on the excitation energy, we performed one more MLFMO-CIS(D) calculation by using 6-31G* basis [21]. The corresponding values at the CIS and CIS(D) levels were obtained to be 4.26 and 3.26 eV, respectively. This means that the balance of cancellation in the CIS(D) correction is not much affected for the present modeling of PYP although the amounts of both relaxation energy and differential correlation energy are increased by the improvement from 6-31G to 6-31G*.

Finally, we discuss the issue of pigment geometry, which might be principally responsible for the remaining discrepancy of 0.5 eV in the MLFMO-CIS(D)/6-31G result relative to the experimental observation of 2.78 eV by Imamoto et al. [26] (see again Table 6). A couple of studies based on the CASSCF second-order perturbation (CASPT2) [56] has been reported for PYP. Molina et al. [57] treated a number of intermediate structures appeared in the photo cycle [26], by using the pigment models replaced Cys69 by –CH_3. They obtained the CASPT2 value of 2.58 eV as the excitation energy of interest. Kawaguchi et al. [58] took care for the geometry preparation of the chromopore in a combination of MD sampling and QM/MM optimization. Furthermore, they used a tricky scheme named as MLSCMO, which is almost the same as the MLFMO scheme [19], in order to incorporate the polarization effect from the surrounding amino acid residues to the central pigment [59]. The final CASPT2 estimate in [58] was 2.88 eV by averaging ten configurations of the chromophore geometry. These two reports [57,58] have exemplified that the theoretical excitation energy has a dependence on the geometry of the PYP chromophore. Although the ‘3PYP’ structure of PDB [55] has been naively used in our calculations of MLFMO-CIS [17] and CIS(D), the assessment of geometry would be desirable, especially for the central pigment moiety or the coumarate. Nielsen et al. [60] measured the absorption spectra of the deprotonated *trans*-thiophenyl-*p*-coumarate, pCT$^-$, which is an experimental model chromophore of PYP, in vacuum. They reported the absorption maximum at 460 nm or 2.70 eV, being slightly lower than the PYP value of 2.78 eV [26]. With this pCT$^-$ molecule, we checked the dependence of CIS(D) energies on geometries. The geometry was optimized at both HF and MP2 levels with 6-31G and 6-31G* basis sets [21,53]. The CIS(D) calculations were performed on the respective geometries. The calculated excitation energies are listed in Table 7. As in the case of the free formaldehyde molecule, the CIS(D) value with the MP2 geometry is in reasonable agreement with the observed value of 2.70 eV [60] relative to the HF-optimized case. The use of polarized 6-31G* yields slightly higher energy than that of 6-31G. The dependence of excitation energies on the geometry just has been shown by using the pCT$^-$ model of PYP. Next, we compared the bond lengths of the phenolate part between the MP2/6-31G geometry of pCT$^-$ and the ‘3PYP’ dataset of PDB [55] and found an average difference of 0.05 Å for the lengths between heavy atoms, where a representative O^--C length was 1.295 Å for the former and 1.345 Å for the latter. Additionally, the higher correlation effect was examined by using the EOM-CCSD module [61,62] available in GAMESS [10]. A simpler model molecule, which was derived from pCT$^-$ by replacing C_6H_5-S- by CH_3-S- at the tail, was employed with the MP2/6-31G geometry optimization [21,53]. The CIS(D) and EOM-CCSD excitation energies with 6-31G were 2.81 and 2.93 eV, respectively. This difference from the higher-order reference value was in a typical range [27,33,35], supporting a validity of CIS(D) calculations for the PYP pigment system. As a whole, we would consider that the naive use of pigment geometry in '3PYP' [55] is mainly responsible for the discrepancy between calculation and experiment [26] found in Table 6. Future investigation with the MLFMO-CIS(D) method should involve the geometry assessment and refinement for at least the chromophore region of PYP.

6 Summary

We have developed a parallelized integral-direct code of the CIS(D) correction [27] with a new spin-adapted formulation for the separate treatment of singlet states and triplet states. The implementation of CIS(D) was made by the modifications of our original MP2 algorithm [15], making an 'on-the-fly' processing of all the integrals possible without communication of bulky data across the worker processes of parallelization. This CIS(D) code was incorporated into a developer version of ABINIT-MP program [14–17,20], in order to correct the excitation energy calculated by the MLFMO-CIS method [17]. With the FRM+16W cluster as a model to mimic the hydrated formaldehyde, the acceleration efficiency of parallel CIS(D) calculations (with 6-31G** basis set [21]) was checked for the lowest $n\pi^*$ state and then shown to be acceptable for production runs. The MLFMO-CIS(D)/6-31G** calculations were performed employing a series of models from FRM/16W to FRM+16W/112W, where the blueshift relative to the free formaldehyde was of interest. The largest FRM+16W/112W model provided a blueshift value of +0.31 eV, being comparable with a reference value of +0.35 eV reported by the CC-based study [22,23]. We furthermore applied the MLFMO-CIS(D) method to PYP [25] whose experimental excitation energy is 2.78 eV [26], as done in [17] at the MLFMO-CIS/6-31G level. The excitation energy was certainly improved from 4.28 eV of the naive CIS to 3.29 eV with the CIS(D) correction. The remaining discrepancy could be mainly attributed to the lack of geometry refinement for the pigment/chromophore, through the supplemental CIS(D) calculations on the pCT$^-$ model pigment of PYP [60]. In summary, our MLFMO-CIS(D) method has opened a new way that a variety of photoactive proteins and solvatochromic systems can be treated in a full QM manner with in-house class resources of the parallel computation.

Acknowledgements Y.M. would thank Prof. Kazuo Kitaura for kind encouragement on the FMO calculations of excited states and also Prof. Shigenori Tanaka for discussion on the photoactive proteins. YM would also be grateful to the anonymous referees for comments to improve the manuscript. The work reported here was supported primarily by the 'Revolutionary Simulation Software for twenty-first Century' (RSS21) project operated by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and partially by the CREST project operated by the Japan Science and Technology Agency (JST). All the calculations were performed on the computational resources at the National Institute of Health Sciences, Kobe University and Rikkyo University.

References

1. Blankenship RE (2002) Molecular mechanisms of photosynthesis. Blackwell, Oxford
2. Lanyi JK (1993) Biochim Biophys Acta 1183:241
3. Reichardt C (1994) Chem Rev 94:2319
4. Svensson M, Humbel S, Froese RJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357
5. Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, New York
6. Kitaura K, Sawai T, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 312:319
7. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 313:701
8. Nakano T, Kaminuma T, Sato T, Akiyama Y, Uebayasi M, Kitaura K (2000) Chem Phys Lett 318:614
9. Fedrov DG, Olson RM, Kitaura K, Gordon MS, Koseki S (2004) J Comp Chem 25:872
10. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) J Comp Chem 14:1347, GAMESS, http://www.msg.ameslab.gov/GAMESS/
11. Fedorov DG, Kitaura K (2006) In: Starikov EB, Lewis JP, Tanaka S (eds) Modern methods for theoretical physical chemistry of biopolymers. Elsevier, Amsterdam, pp 3–38
12. Sekino H, Sengoku Y, Sugiki S, Kurita N (2003) Chem Phys Lett 378:589
13. Kendall RA, Apra E, Bernholdt DE, Bylaska EJ, Dupuis M, Fann GI, Harrison RJ, Ju J, Nichols JA, Nieplocha J, Straatsma TP, Windus TL, Wong AT (2000) Comp Phys Comm 128:260, NWCHEM, http://www.emsl.pnl.gov/docs/nwchem/nwchem.html
14. Nakano T, Kaminuma T, Sato T, Fukuzawa K, Akiyama Y, Uebayasi M, Kitaura K (2002) Chem Phys Lett 351:475, ABINIT-MP
15. Mochizuki Y, Nakano T, Koikegami S, Tanimori S, Abe Y, Nagashima U, Kitaura K (2004) Theor Chem Acc 112:442
16. Mochizuki Y, Koikegami S, Nakano T, Amari S, Kitaura K (2004) Chem Phys Lett 396:473
17. Mochizuki Y, Koikegami S, Amari S, Segawa K, Kitaura K, Nakano T (2005) Chem Phys Lett 406:283
18. Foresman JB, Head-Gordon M, Pople JA, Frisch MJ (1992) J Phys Chem 96:135
19. Fedorov DG, Kitaura K (2005) J Chem Phys 122:054108-1
20. Nakano T, Mochizuki Y, Fukuzawa K, Amari S, Tanaka S (2006) In: Starikov EB, Lewis JP, Tanaka S (eds) Modern

methods for theoretical physical chemistry of biopolymers. Elsevier, Amsterdam, pp 39–52
21. Foresman JB, Frisch A (1996) Exploring chemistry with electrononic structure methods, 2nd edn. Gaussian, Pittsburgh
22. Kongsted J, Osted A, Pedersen TB, Mikkelsen KV, Christiansen O (2004) J Phys Chem A 108:8624
23. Kongsted J, Osted A, Mikkelsen KV, Åstrand P.-O., Christiansen O (2004) J Chem Phys 121:8435
24. Hirata S, Valiev M, Dupuis M, Xantheas SS, Sugiki S, Sekino H (2005) Mole Phys 103:2255
25. Meyer TE (1985) Biochim Biophys Acta 806:175
26. Imamoto Y, Kataoka M, Tokunaga F (1996) Biochem 35:14047
27. Head-Gordon M, Rico RJ, Oumi M, Lee TJ (1994) Chem Phys Lett 219:21
28. Szabo A, Ostlund NS (1982) Modern quantum chemistry. MacMillan, New York
29. Burke K, Werschnik J, Gross EKU (2005) J Chem Phys 123:062206, references therein
30. Dreuw A, Weisman JL, Head-Gordon M (2003) J Chem Phys 119:2943
31. Odelius M, Kirchner B, Hutter J (2004) J Phys Chem A 108:2044
32. Casida ME (2005) J Chem Phys 122:054111
33. Oumi M, Maurice D, Lee TJ, Head-Gordon M (1997) Chem Phys Lett 279:151
34. Head-Gordon M, Lee TJ (1997) In: Bartlett RJ (ed) Recent advances in coupled cluster methods. World Scientific, Singapore, pp 221–253
35. Head-Gordon M, Oumi M, Maurice D (1999) Mole Phys 96:593
36. Hirata S (2005) J Chem Phys 122:094105-1
37. Hirata S (2003) J Phys Chem A 107:9887
38. Pople JA, Head-Gordon M, Raghavachari K (1987) J Chem Phys 87:5968
39. Tanaka K (1972) Dr. Thesis, Department of Chemistry, Faculty of Science, Hokkaido University
40. Tanaka K (1974) Intern J Quant Chem 13:981
41. Pauncz R (2000) The construction of spin eigenfunctions. Kluwer, New York
42. Dykstra CE, Davidson ER (2000) Intern J Quant Chem 78:226
43. Head-Gordon M, Pople JA, Frisch MJ (1988) Chem Phys Lett 153:503
44. Message Passing Interface (MPI), http://www.mpi-forum.org/
45. Basic linear algebra subroutines (BLAS), http://www.netlib.org/blas/
46. Martin ME, Sánchez ML, Olivares del Valle FJ, Aguilar MA (2000) J Chem Phys 113:6308
47. Roos BO, Taylor PR, Siegbahn PEM (1980) Chem Phys 48:157
48. Kawashima K, Dupuis M, Hirao K (2002) J Chem Phys 117:248
49. Dupuis M, Kawashima Y, Hirao K (2002) J Chem Phys 117:1256
50. Koch H, Jϕrgensen P (1990) J Chem Phys 93:3333
51. Stanton JF, Bartlett RJ (1993) J Chem Phys 98:7029
52. Robin MB (1985) Higher excited states of polyatomic molecules. Academic, New York
53. Gaussian 03 Revision B 0.4, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, BaroneV, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda, R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PM W, Johnson B, Chen W, Wong MW, Gonzalez C, (2003) Pople JA, Gaussian, Pittsburgh, http://www.gaussian.com/
54. Yoda M, Houjou H, Inoue Y, Sakurai M (2001) J Phys Chem B 105:9887
55. Protein Data Bank (PDB), http://www.rcsb.org/
56. Andersson K, Malmqvist PÅ, Roos BO, Sadlej AJ, Wolinski K (1990) 94:5483
57. Molina V, Merchan M (2001) Proc Natl Acad Sci USA 98:4299
58. Kawaguchi K, Yamato T (2006) Chem Phys Lett 430:386
59. Matsuura A, Sato H, Houjou H, Saito S, Hayashi H, Sakurai M (2006) J Comp Chem 27:1623
60. Nielsen IB, Boyé-Péronne S, El Ghazaly MOA, Kristensen MB, Nielsen SB, Andersen LH (2005) Biophys J 89:2597
61. Kowalski K, Piecuch P (2001) J Chem Phys 115:2966
62. Piecuch P, Kucharski SA, Kowalski K, Musial M (2002) Comp Phys Comm 149:71