Density-Matrix Based Scheme of Basis Selection for Linear Combination of Fragment Molecular Orbitals

Okiyama, Yoshio

Mochizuki, Yuji

Yamanaka, Masanori

Tanaka, Shigenori

Density-Matrix Based Scheme of Basis Selection for Linear Combination of Fragment Molecular Orbitals

Yoshio Okiyama[1]*, Yuji Mochizuki[2], Masanori Yamanaka[3], and Shigenori Tanaka[4]†

[1]*Division of Medicinal Safety Science, National Institute of Health Sciences, Kawasaki, Kanagawa 210-9501, Japan*

[2]*Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, Toshima, Tokyo 171-8501, Japan*

[3]*Department of Physics, College of Science and Technology, Nihon University, Chiyoda, Tokyo 101-8308, Japan*

[4]*Graduate School of System Informatics, Kobe University, Kobe 657-8501, Japan*

A dependable description of the frontier orbitals around the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is essential for the theoretical elucidation of exciton and charge-transfer dynamics in molecular systems. In this study, an efficient scheme for constructing molecular orbitals (MOs) of large molecular systems is proposed within the framework of the fragment molecular orbital (FMO) method. In contrast to the previously proposed energy-based methodology, called the linear combination of molecular orbitals of FMO (FMO-LCMO), the present scheme employs the density matrices of monomer and dimer MOs of fragments to select the monomer MOs as the basis set for constructing the Hamiltonian matrix of the total system. The accuracy and computational cost of the proposed scheme are assessed in comparison with those of the conventional energy-based methods for three model systems of DNA stacked base pairs, pseudo-glycine pentamer, and water clusters, thus demonstrating its efficiency in terms of data compression.

1. Introduction

A number of fragment-based methods have been proposed to conduct quantum chemical calculations of large molecular systems, including proteins, nucleic acids, and nanomaterials.[1] One of the most successful examples of such attempts is the fragment molecular orbital (FMO) method,[2,3] which applies fragmented molecular orbital (MO) calculations in the embedding electrostatic field of the entire system. The FMO method has been used primarily for calculations of the energies and gradients in geometry optimizations and molecular dynamics,[4,5] and because the method is an energy-based fragmentation approach, obtaining the MOs and their energies for whole molecular systems is not easy. The simplest way to obtain the MOs and their energies in an FMO scheme is to take the union of fragment (monomer) MOs, which corresponds to the one-body expansion, FMO1. Pair corrections can then be added to the FMO-MO method,[6] in which FMO provides an initial density for a couple of iterations in regular MO schemes, or in the linear combination of MOs of fragments (FMO-LCMO) method,[7] where one can obtain a subset of orbitals and energies around the Fermi level or the frontier orbitals.[8–14] In the present study, we focus on this FMO-LCMO method. Relevant descriptions of frontier orbitals are essential for understanding charge transfer and exciton dynamics in large molecular systems.[15,16]

The FMO-LCMO method was originally proposed to obtain canonical MOs of large systems within the framework of the FMO scheme[7] (see Sect. 2 for details). The FMO method first decomposes the total system into fragments and calculates the MOs of each fragment self-consistently under the electrostatic field of the other fragments. Dimer, trimer,[17,18] or tetramer[19] calculations are then carried out to account for the charge transfer and exchange interactions over multiple fragments. With the FMO-LCMO method, to remove the excess energy, the diagonal elements of the Hamiltonian matrix of the total system are computed from the results of these fragment calculations, whereas the off-diagonal elements are constructed from the dimer or trimer matrices projected to the monomer MO space.[7,8] The total Hamiltonian matrix is then diagonalized to determine the canonical MOs of the entire system. It is thus supposed that the Hamiltonian matrix before diagonalization provides appropriate diagonal site energies and off-diagonal transfer integrals, including the exchange and charge-transfer effects.[9–13] Although the FMO-LCMO method has been reported to attain a high accuracy at the FMO3 level,[8] inaccuracies in the virtual orbitals have been considerable for some systems at the FMO2 level.[20] The reason for this difficulty is likely to be the lack of higher-order exchange effects in the embedding potential. The accuracy of the FMO-LCMO method can then be assessed in terms of the orbital energies and density of states for a representative set of materials compared with full MO results without fragmentation.

The FMO-LCMO method is regarded as a computational scheme that can efficiently describe the MOs of large-scale molecular systems by using fragment monomer orbitals as the bases. In general, there is arbitrariness in the selection of the monomer MO bases; because interest is primarily in the canonical orbitals near the highest occupied molecular orbital (HOMO)/lowest unoccupied molecular orbital (LUMO), the HOMO/LUMO vicinities of the monomer MOs are uniformly selected as the bases. Therefore, our question is the following: Is there a basis selection scheme that most efficiently reproduces the canonical orbitals over this arbitrariness?

Thus, the current problem that should be resolved is how to optimize the selection of relevant fragment MOs for constructing the Hamiltonian matrix of the total system such that the physicochemical behaviors of the frontier orbitals are appropriately described with a relatively lower computational cost. Herein, we consider that one can construct a density matrix obtained by projecting a monomer MO onto a dimer MO. Each matrix element consists of an overlap integral between them, representing the contribution of a monomer
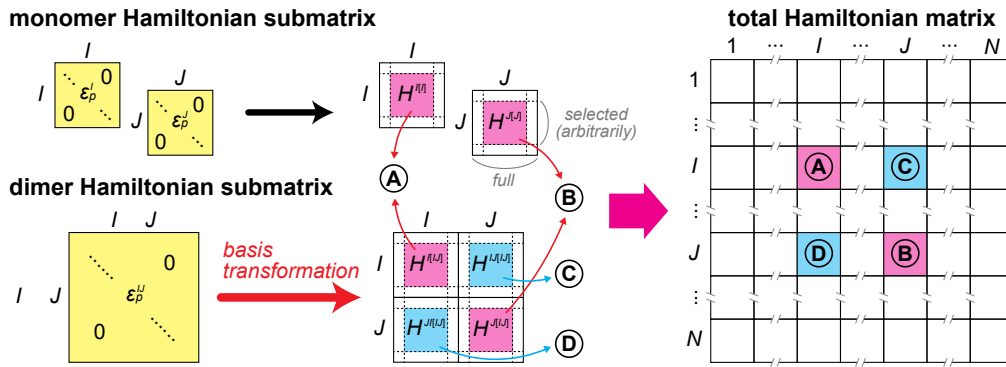
Fig. 1. (Color online) Construction of the total Hamiltonian matrix in the FMO2-LCMO method for a system containing N fragments.

MO to make up a dimer MO. To realize this concept, we diagonalize the density matrix and select the monomer MOs according to their eigenvalues and eigenvector components.

One possible solution is thus the diagonalization of the density matrix (referred to as TFDM below) between the monomer orbitals and the higher-order polymer orbitals. From the eigenvalues (occupation numbers) and eigenvectors (where their constituents are monomer orbitals), the monomer orbitals important for constituting the polymer orbitals are determined. That is, the idea here is that the monomer orbitals are superposed onto the higher-order polymer orbitals, and the monomer orbitals having a large overlap are essential for reproducing the higher-order polymer orbitals. In this study, we propose a new calculation scheme that extracts the effective basis space by efficiently compressing the unoptimized basis space, such as HOMO/LUMO neighborhoods that have been employed thus far.

In the following Sect. 2, we illustrate the basic theory and associated computational scheme of the present proposal. In Sect. 3, the test calculations for three molecular systems are shown to assess the usefulness of the present method. Some concluding remarks are given in Sect. 4.

2. Theory

In this section, we first review the fundamental formulation of the FMO method and the conventional monomer MO-based FMO-LCMO method. We next introduce a novel density-matrix-based scheme and related considerations. Furthermore, we show the computational levels and settings used in this study.

2.1 FMO-LCMO scheme

The FMO method[2–5] defines the total energy of each subsystem X ($X = I, IJ, IJK, \ldots$) as E_X, including the environmental electrostatic potential from other fragments, where I, J, K represent fragment indices and their combinations refer to the dimer and higher-order fragments. Once each monomer energy within the environmental potential is self-consistently obtained, the total energy of the entire system can be expressed in a many-body expansion manner with a correction energy of ΔE_X:

$$E^{\mathrm{FMO}} = \sum_I E_I + \sum_{I>J} \Delta E_{IJ} + \sum_{I>J>K} \Delta E_{IJK} + \cdots. \quad (1)$$

In the pairwise approximation of FMO (FMO2), dimer energies are required in the correction formula

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J, \quad (2)$$

which indicates the difference in energy from the monomers.

In the FMO-LCMO method, the total Hamiltonian matrix of the entire system is analogically constructed in the monomer MO representation[7]

$$H^{\mathrm{FMO}} = H^{(1)} + \Delta H^{(2)} + \Delta H^{(3)} + \cdots \quad (3)$$

by adding the ζ-th order corrections $\Delta H^{(\zeta)}$ to the first-order Hamiltonian matrix $H^{(1)}$. Each Hamiltonian matrix consists of its submatrices of each subsystem X having fragment MOs $|\phi^X\rangle$ and associated energies ε^X with their indices of p, q, r, s as follows:

$$
\begin{aligned}
H_{pq}^{Y[X]} &= \langle \phi_p^I | \hat{H}^X | \phi_q^J \rangle \\
&= \sum_r \sum_s \langle \phi_p^I | \phi_r^X \rangle H_{rs}^X \langle \phi_s^X | \phi_q^J \rangle \\
&= \sum_r \varepsilon_r^X \langle \phi_p^I | \phi_r^X \rangle \langle \phi_r^X | \phi_q^J \rangle, \quad (4)
\end{aligned}
$$

where Y represents a block of the Hamiltonian matrix based on the monomer MOs, taking only a diagonal block ($Y = I$, $I = J$) or off-diagonal block ($Y = IJ$, $I \neq J$). The superscript $Y[X]$ denotes the contribution of the Hamiltonian components from a subsystem X to a block Y of the entire system. When Y is equal to X, $H_{pq}^{Y[X]}$ is simply expressed as

$$H_{pq}^{Y[X]} = \varepsilon_p^X \delta_{pq}, \quad (5)$$

where δ_{pq} is Kronecker's delta. The matrix element for the first-order Hamiltonian is then represented as

$$H_{pq}^{(1)} = \sum_I H_{pq}^{I[I]}, \quad (6)$$

and that for the second-order correction is also given as

$$
\Delta H_{pq}^{(2)} =
\begin{cases}
\sum_{J \neq I} (H_{pq}^{I[IJ]} - H_{pq}^{I[I]}) & \text{(diagonal block)} \\
H_{pq}^{IJ[IJ]} & \text{(off-diagonal block)}
\end{cases}, \quad (7)
$$

as required for the pairwise approximation of FMO-LCMO (FMO2-LCMO). Figure 1 depicts the contribution of matrix elements above to the total Hamiltonian matrix. After the construction of the total Hamiltonian matrix, solving the generalized eigenvalue problem for the matrix yields delocalized LCMOs with one-electron orbital energies and the corresponding coefficients in the monomer MO representation. Because there exists arbitrariness in spanning the basis space of the monomer MOs, their HOMO/LUMO

vicinities are uniformly selected as the bases in a conventional scheme. In the present research, as shown below, we confine ourselves to the consideration of FMO2-LCMO as a first step in the formulation.

2.2 Basis selection based on density matrix

Herein, we express the MOs of the fragment monomer and dimer as $|\phi_p\rangle$ and $|\psi_t\rangle$, respectively. A matrix element of the *pseudo* density matrix **D** for each monomer, ρ_{pq}, is defined as follows:

$$\rho_{pq} = \sum_t \langle\phi_p|\psi_t\rangle\langle\psi_t|\phi_q\rangle, \tag{8}$$

where the target range of t for dimers can be taken arbitrarily and may be near the frontier orbitals of the dimers. To emphasize the inter-relationship between the fragment monomer and dimer MOs, we hereafter call this **D** the trans-fragment density matrix (TFDM).[21] Upon diagonalizing **D**, we obtain the matrix **L**, which satisfies the following equation with the identity matrix **I**:

$$\mathbf{L}^{-1}\mathbf{D}\mathbf{L} = \vec{\Lambda}\mathbf{I}, \tag{9}$$

where **L** constitutes eigenvectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_j, \ldots$ associated with eigenvalues $\Lambda_1, \Lambda_2, \ldots, \Lambda_j, \ldots$, respectively. The eigenvalues correspond to the occupation numbers, and the associated eigenvectors consist of the coefficients for the MOs of each monomer. The most efficient way to compress the LCMO space based on the monomer MOs is to select the monomer MOs that contribute significantly to the eigenvalues and eigenvector components. For this selection, we introduce the following *contribution index* for the i-th monomer MO:

$$c_i := \sum_j^N \Lambda_j^\ell |a_{ij}|^m > c^{\min}, \tag{10}$$

where N is the number of eigenvalues, ℓ is the power of the eigenvalues Λ_j, and m is the power of the absolute value of the eigenvector components a_{ij}. The monomer MO is thus adopted as the basis when c_i is larger than the threshold $c^{\min}$. Once the essential monomer MOs have been extracted for each dimer, their aggregation is considered as the basis set for the LCMO. We call such an adaptive selection based on TFDM a DM scheme, and distinguish it from the conventional scheme based on a predetermined range of MO energy levels, which herein is called the EL scheme. More specifically, the DM scheme first targets the dimer MOs and then selects the monomer MO bases according to the fitness for the dimer MOs, whereas the EL scheme directly targets the monomer MOs used as the bases as is. In the present study of the DM scheme, we fix the parameter values as $\ell = 1$ and $m = 2$. For $c^{\min}$, we investigated its effect within the range of 0.01, 0.05, 0.1, 0.2, and 0.5.

We make some supplementary remarks on the present scheme. Let us consider a density operator in terms of the following wave functions $|\Psi_\nu\rangle$:

$$\hat{W} = \sum_\nu w_\nu |\Psi_\nu\rangle\langle\Psi_\nu|, \tag{11}$$

with the coefficients w_ν. As a special case for this, we consider a projection operator as

$$\hat{P} = \sum_{\nu \in target} |\Psi_\nu\rangle\langle\Psi_\nu|, \tag{12}$$

where $w_\nu = 1$ when ν belongs to the target states and is $w_\nu = 0$ otherwise. If we consider a density matrix with the basis set of fragment monomer MOs as

$$\rho_{pq} = \langle\phi_p|\hat{P}|\phi_q\rangle = \sum_{\nu \in target} \langle\phi_p|\Psi_\nu\rangle\langle\Psi_\nu|\phi_q\rangle, \tag{13}$$

with "*target*" being the target (e.g., frontier) orbitals of the dimer MOs, it recovers the density matrix of Eq. (8). Furthermore, if ν in Eq. (12) forms a complete system, we find

$$1 = \sum_\nu |\Psi_\nu\rangle\langle\Psi_\nu|, \tag{14}$$

and Eq. (13) becomes

$$\rho_{pq} = \delta_{pq}, \tag{15}$$

thus implying that the selection of the target states is essential. Whereas the earlier FMO-LCMO scheme corresponds to an approximation that the monomer MOs around the HOMO and LUMO are selected as the target states, the present scheme focuses on the inclusion of inter-fragment orbitals associated with the charge transfer with the exchange effects. This idea thus complies with the essence of the density matrix renormalization group (DMRG)[22–24] concerning the inclusion of boundary effects upon the expansion of the system size.

In addition, the present renormalization procedure is intimately associated with the singular value decomposition (SVD),[25–28] which is represented as follows:

$$\langle\phi_p|\psi_r\rangle = \sum_\alpha U_{p\alpha}\lambda_\alpha V_{r\alpha}, \tag{16}$$

where $U_{p\alpha}$, $V_{r\alpha}$, and λ_α refer to the left-hand matrix, right-hand matrix, and eigenvalue of the SVD, respectively. If Eq. (16) is substituted into Eq. (8), we find

$$\rho_{pq} = \sum_r \sum_\alpha \sum_\beta U_{p\alpha}\lambda_\alpha V_{r\alpha}V_{r\beta}\lambda_\beta U_{q\beta}, \tag{17}$$

and hence

$$\mathrm{Tr}\,\rho_{pq} = \sum_{p,q,r} \sum_{\alpha,\beta} \delta_{pq}U_{p\alpha}U_{q\beta}V_{r\alpha}V_{r\beta}\lambda_\alpha\lambda_\beta. \tag{18}$$

If we assume the monomer MOs $\{\phi_p\}$ form a complete system and use

$$\sum_p U_{p\alpha}U_{p\beta} = \delta_{\alpha\beta}, \tag{19}$$

we are led to

$$\mathrm{Tr}\,\rho_{pq} = \sum_\alpha \sum_r \lambda_\alpha^2 V_{r\alpha}^2. \tag{20}$$

When the states r form a complete system, we find that $\sum_r V_{r\alpha}^2 = 1$ and λ_α^2 refer to the eigenvalues of the density matrix ρ_{pq}. The idempotence of the density matrix is retained when the basis set forms a complete system, but not when some truncations are introduced.

2.3 Computational details

In this study, we employ the Hartree–Fock (HF)

Y. Okiyama et al.

approximation with the 6-31G* basis set for the FMO2-LCMO analysis. Although more accurate descriptions based on the density functional theory (DFT) are feasible, we suppose that the prototypical HF-level analysis will provide essential insights into the development of the FMO-LCMO methodology. We used the developer version of ABINIT-MP[5,29] for the following analysis.

In the following, we compare the results of the FMO-LCMO scheme obtained by the earlier EL-based method and the present DM-based method with those of the canonical orbitals obtained using a conventional MO method. When employing all monomer MOs as bases of FMO-LCMO, the scheme is called *full* FMO-LCMO, which indicates the ultimate performance of the FMO-LCMO method. The resultant orbital of the full FMO-LCMO should be distinguished from the *canonical* orbital. Regarding the choice of target orbitals ("window"), we consider three schemes: 1) full, 2) HOMO$-n$:LUMO$+n$, and 3) LOMO:LUMO$+n$. For the HOMO$-n$:LUMO$+n$ scheme, we account for HOMO$-n$,…, HOMO-1, HOMO, LUMO, LUMO$+1$,…, LUMO$+n$. For the LOMO:LUMO$+n$ scheme, we consider all occupied orbitals, where LOMO is the lowest occupied molecular orbital. To distinguish between localized monomer MOs and delocalized MOs such as LCMOs that linearly combine the localized MOs, or canonical MOs, we use the superscript "(d)" for the latter MOs, as MO$^{(d)}$. The redundancy of the bases derived from the bond detached atom (BDA)[5] is removed by the diagonalization of overlap integrals when we do not employ the minimal basis set.[8]

3. Results and Discussion

In this section, we demonstrate the results of the DM-based LCMO scheme using stacked base pairs of deoxyribonucleic acid (DNA), pseudo-glycine pentamer, and water clusters as test systems in comparison with those of the conventional EL-based scheme.

3.1 Stacked base pairs of DNA

First, we consider the stacked Watson–Crick (WC) base pairs of DNA with four types of AA, AT, CG, and GA.[30] Figure 2 shows the structure of CG stacked base pairs, where each base is a single fragment without BDA. We carried out the DM (HOMO$-n$:LUMO$+n$) calculations in comparison with the EL scheme. Figure 3 illustrates the comparison results concerning the distribution of selected monomer MOs in the case of the CG stacking, where the corresponding orbital energies of each monomer are shown in Fig. S1 in the Supplemental Material section.[31] For the EL scheme, only
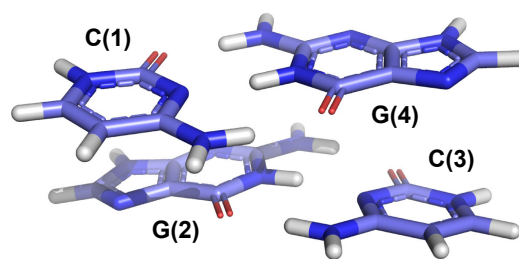

Fig. 2. (Color online) Structure of CG stacked base pairs. Nucleobase codes are written with the fragment number in parentheses.
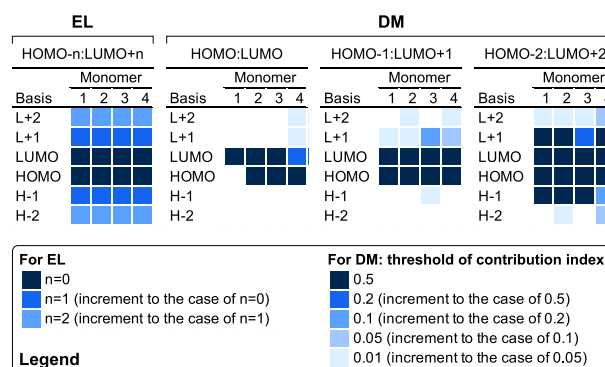

Fig. 3. (Color online) Monomer MO distribution as selected bases in CG stacked base pairs, determined by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO$-n$:LUMO$+n$ ($n = 0, 1, 2$). The basis labels of MO energy levels below the HOMO and above the LUMO are abbreviated as H$-k$ and L$+k$ ($k = 1, 2, \ldots$), respectively. For EL, only the frontier MOs (HOMO:LUMO) of all monomers are uniformly selected as the bases in the case of $n = 0$ (the bases are depicted in a color labeled as $n = 0$), one monomer MOs located just at the lower and upper energy positions of the above bases are added in the case of $n = 1$ (the increment bases for $n = 0$ are depicted in a color labeled as $n = 1$), and further ones are similarly added in the case of $n = 2$ (the increment bases for $n = 1$ are depicted in a color labeled as $n = 2$). For DM, the selected bases vary depending on the overlapped dimer orbitals and the threshold value of the contribution index. When some bases selected with the threshold of 0.5 are depicted in color and labeled as 0.5, other monomer MOs are added stepwise to the bases according to the reduction of the threshold of 0.2, 0.1, 0.05, and 0.01 (the increment bases for 0.5, 0.2, 0.1, and 0.05 are depicted in colors labeled as 0.2, 0.1, 0.05, and 0.01, respectively).

the frontier MOs (HOMO:LUMO) of all monomers are uniformly selected as the bases in the case of $n = 0$, and each monomer MO located just at the lower and upper energy positions of the above bases are added in the case of $n = 1$, and further monomer MOs are added in the case of $n = 2$; the total numbers of selected bases are 8, 16, and 24, respectively. For the DM scheme, the selected bases vary

Table I. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO$-n$:LUMO$+n$ ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for CG stacked base pairs.

	Canonical	Full	EL			DM ($c^{\min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	610	610	8	16	24	7	9	16
LUMO+2	0.173	0.173	0.176	0.174	0.113	0.176	0.174	0.174
LUMO+1	0.118	0.117	0.117	0.117	0.043	0.117	0.117	0.117
LUMO	0.113	0.113	0.113	0.113	0.037	0.113	0.113	0.113
HOMO	-0.263	-0.262	-0.263	-0.263	-0.263	-0.263	-0.263	-0.263
HOMO-1	-0.266	-0.266	-0.266	-0.266	-0.266	-0.266	-0.266	-0.266
HOMO-2	-0.339	-0.340	-0.341	-0.341	-0.341	-0.341	-0.341	-0.341
HOMO/LUMO gap	0.375	0.375	0.376	0.376	0.299	0.376	0.376	0.376

depending on the overlapped dimer orbitals and the threshold value of the contribution index. For the basis selected with the threshold of 0.5, other monomer MOs are added stepwise to the bases according to the reduction of the threshold of 0.2, 0.1, 0.05, and 0.01, and the total numbers of selected bases with the threshold of 0.1 are 3, 6, and 7 for the targeted dimer MO windows ($n = 0$, 1, and 2), respectively. As seen in this comparison, the present DM scheme adaptively selects the monomer MO bases by fitting to the designated dimer orbitals with the energy width n and given threshold parameters, whereas the EL scheme homogeneously selects the monomer MO bases around HOMO/LUMO according to the designated value of n for the monomer MO energy levels. Thus, in the present DM scheme, the essential monomer MOs are efficiently selected according to the threshold parameters such that the number of monomer MO bases is adaptively changed. We then compare the resultant FMO-LCMO orbitals with the canonical orbitals in Table I. In contrast to the full 610 bases, both the EL and DM schemes well reproduce the orbital energies around HOMO$^{(d)}$/LUMO$^{(d)}$ with only tens of monomer MO bases. However, unnatural orbitals appear at LUMO$^{(d)}$ and LUMO+1$^{(d)}$ in the case of $n = 2$ of the EL scheme, leading to a reduction in the HOMO$^{(d)}$/LUMO$^{(d)}$ gap. In principle, such unnatural orbitals must be shifted to higher energy levels owing to their orthogonality with lower-energy-level occupied orbitals that should canonically exist. In the FMO-LCMO scheme, which drastically limits the basis space of the monomers, the occupied orbitals in depth are often omitted, resulting in such energy level discrepancies. This difficulty can be attributed to the fact that the required orthogonality to the occupied orbitals is not attained in the extended basis sets beyond the minimal set.[8]

To cope with this orthogonality problem, we further conducted a DM (LOMO:LUMO+n) analysis in comparison with the corresponding EL calculation, where all occupied orbitals down to the LOMO and the unoccupied orbitals near the LUMO are considered. The distribution of the monomer MOs selected as bases is shown in Fig. 4, and the corresponding orbital energies of the monomer MOs are shown in Fig. S1 in the Supplemental Material section,[31] where all occupied orbitals are employed as bases in both the EL and DM schemes. This comparison shows that the unoccupied orbitals of the monomers selected in the DM scheme are distributed discontinuously and broadly according to the criteria of the target width and contribution threshold. Figure 5 then illustrates the dependence of LCMO energy levels near the LUMO$^{(d)}$ on the threshold of the contribution index in the DM scheme. Thus, we observe that the energy levels of LUMO$^{(d)}$ and LUMO+1$^{(d)}$ converge well within the range of 0.01–0.5 of the threshold parameter. In this case, for example, the LUMO of fragment 4 does not contribute to the LUMO$^{(d)}$ of LCMO in the case of an $n = 0$ target fitting. However, a target value of $n = 1$ or higher and a threshold value of 0.1 or less are required to reproduce the LUMO+2$^{(d)}$ orbital energy of LCMO. Table II shows a comparison of the LCMO energy levels by EL and DM methods with those of the canonical orbitals in the case of targeting LOMO:LUMO+n for CG stacked base pairs. Thus, we find the resolution of the orthogonality difficulty, leading to an accurate reproduction of the canonical orbitals in both
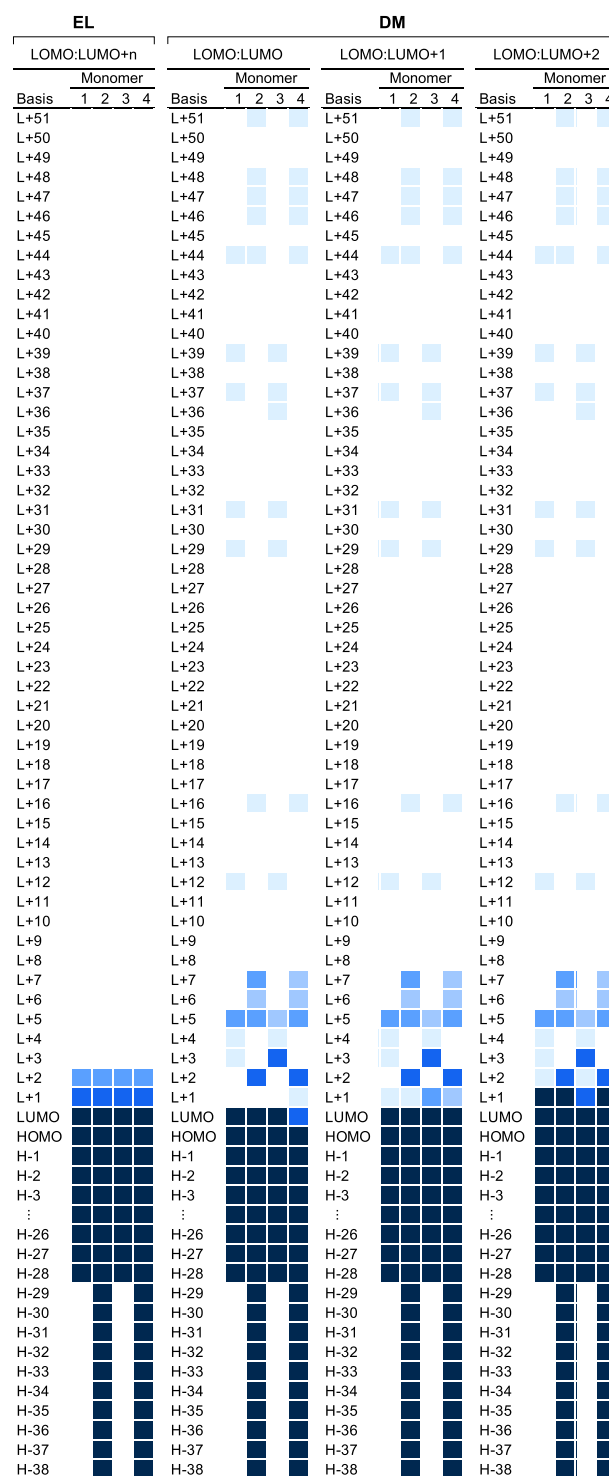

Fig. 4. (Color online) Monomer MO distribution as selected bases in CG stacked base pairs, determined through EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$). The color coding follows the legend in Fig. 3. The basis labels of monomer MO energy levels below the HOMO and above the LUMO are abbreviated as H−k and L+k ($k = 1, 2, \ldots$), respectively.

the EL and DM schemes, whereas the required numbers of monomer bases are substantially increased up to 140–150.

Furthermore, we carried out analogous analyses for AA-, AT-, and GA-stacked WC base pairs, the structures of which are depicted in Figs. S2, S3, and S4, respectively, in the Supplemental Material section.[31] The dependence of the FMO-LCMO energy levels near the LUMO$^{(d)}$ on the
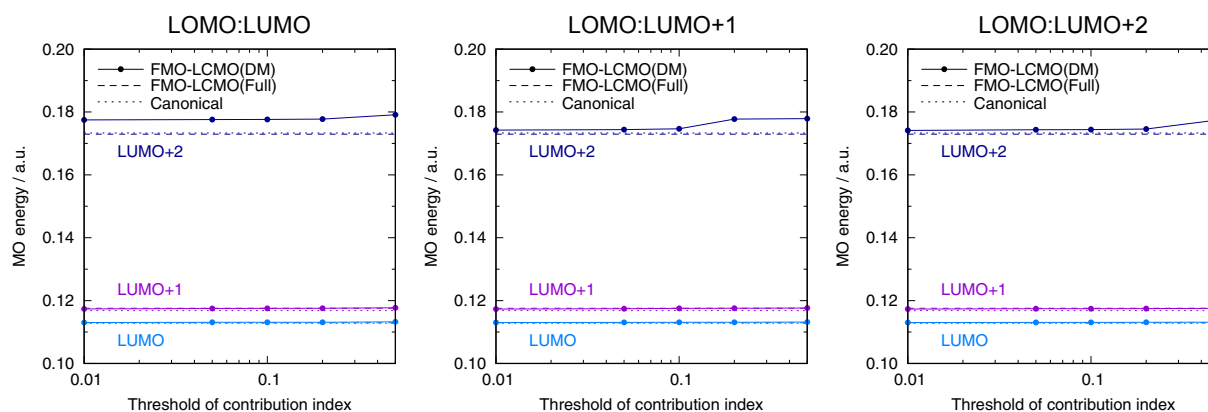
Fig. 5. (Color online) Dependence of delocalized MO energy levels near LUMO on the threshold of contribution index in DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for CG stacked base pairs.

Table II. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for CG stacked base pairs.

	Canonical	Full	EL			DM ($c^{\min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	610	610	140	144	148	147	148	151
LUMO+2	0.173	0.173	0.178	0.175	0.175	0.178	0.175	0.174
LUMO+1	0.118	0.117	0.118	0.118	0.117	0.118	0.118	0.117
LUMO	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113
HOMO	−0.263	−0.262	−0.262	−0.262	−0.262	−0.262	−0.262	−0.262
HOMO−1	−0.266	−0.266	−0.265	−0.265	−0.265	−0.265	−0.265	−0.265
HOMO−2	−0.339	−0.340	−0.340	−0.340	−0.340	−0.340	−0.340	−0.340
HOMO/LUMO gap	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375

threshold parameter $c^{\min}$ in the DM-based LOMO:LUMO+n target scheme are illustrated in Figs. S5–S7 in the Supplemental Material section[31] in comparison with the full FMO-LCMO and canonical MO results. Tables S1–S6 in the Supplemental Material section[31] show comparisons of the orbital energy levels between the EM/DM based FMO-LCMO results and the canonical MO results in the cases of HOMO−n:LUMO+n and LOMO:LUMO+n target schemes. Taken together, the results for AA, AT, and GA stacked pairs are similar to those for the CG case above but indicate that the orthogonality problem may occur because of the increase in the number of MO bases even in the DM and EL schemes. That is, whereas the HOMO−n:LUMO+n target scheme seems to work well in the DM case for the stacked base pairs of DNA, we should focus on the appearance of unnatural orbitals because of the orthogonality difficulty when the number of monomer MO bases increases.

3.2 Pseudo-glycine pentamer

Next, we consider a pseudo-glycine pentamer molecule with the BDA, where the N-terminal glycine residue is modified from $NH_2CH_2CO–$ to $HCONHCH_2CO–$, and the C-terminal glycine residue is modified from $–NHCH_2COOH$ to $–NHCH_3$.[7] Figure 6 shows its structure and fragmentation. In this system, it is remarkable that the orthogonality difficulty owing to a reduction of occupied orbitals in the target vividly appears, as shown in Table III. In other words, when we employ the HOMO−n:LUMO+n target scheme, large deviations in the LCMO orbitals from the canonical orbitals are observed on the LUMO$^{(d)}$ side in both the EL and
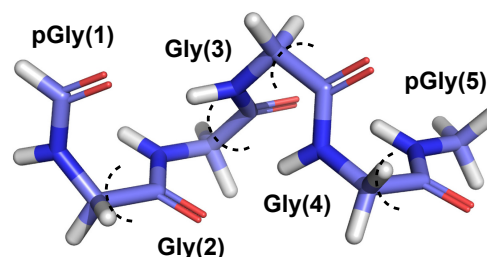

Fig. 6. (Color online) Structure and fragmentation of pseudo-glycine pentamer. The glycine residues at the N- and C-termini are modified as $HCONHCH_2CO–$ and $–NHCH_3$, respectively, and thus called "pseudo" glycine residues. This peptide is divided into five fragments at the sites of each Cα atom. Residue codes are written with the fragment number in parentheses.

DM schemes. It should also be noted that there are small errors in the orbital energies at the HOMO$^{(d)}$ (−0.012 a.u.) and LUMO$^{(d)}$ (−0.007 a.u.) levels, even in the full LCMO scheme.

Hence, we need the LOMO:LUMO+n target scheme to reproduce the canonical orbitals. Figure 7 shows the distribution of the monomer MOs selected as bases, and the corresponding orbital energies are shown in Fig. S8 in the Supplemental Material section.[31] The distribution is non-trivial and discontinuous, which is analogous to the DNA base pair case. For example, LUMO+36 in monomer fragment 5 selected for $c^{\min} = 0.05$ provides a prototypical case. Figure 8 then illustrates the distribution of the LCMO energy levels near the HOMO$^{(d)}$/LUMO$^{(d)}$ in the DM

Table III. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO−n:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for pseudo-glycine pentamer.

|	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	334	334	10	20	30	27	32	35
LUMO+2	0.159	0.156	0.059	0.025	0.018	−0.084	−0.085	−0.100
LUMO+1	0.144	0.140	0.029	−0.022	−0.041	−0.108	−0.123	−0.152
LUMO	0.135	0.128	−0.048	−0.057	−0.100	−0.169	−0.169	−0.176
HOMO	−0.344	−0.332	−0.333	−0.333	−0.333	−0.333	−0.333	−0.333
HOMO−1	−0.360	−0.351	−0.384	−0.360	−0.358	−0.360	−0.359	−0.360
HOMO−2	−0.385	−0.384	−0.428	−0.392	−0.391	−0.395	−0.394	−0.394
HOMO/LUMO gap	0.478	0.460	0.285	0.275	0.233	0.164	0.164	0.157
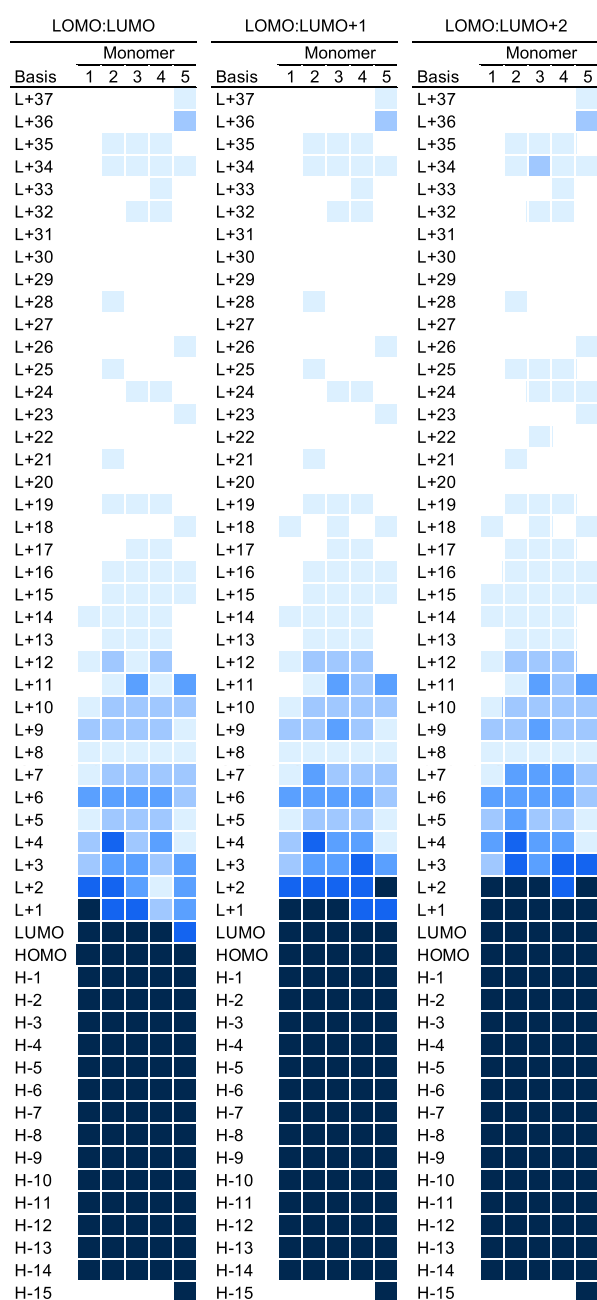
Fig. 7. (Color online) Monomer MO distribution as selected bases in pseudo-glycine pentamer, determined through DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0$, 1, 2). The color coding follows the legend in Fig. 3. The basis labels of monomer MO energy levels below the HOMO and above the LUMO are abbreviated as H−k and L+k ($k = 1, 2, ...$), respectively.

scheme. Although the occupied orbitals are almost identical in all cases, the distribution on the unoccupied side varies considerably according to the conditions of the target and threshold values, but with the converged LUMO[(d)] level below $c^{min} = 0.1$. Figure 9 illustrates the dependence of the LCMO energy levels near LUMO[(d)] on the threshold of the contribution index in the DM-based LOMO:LUMO+n target schemes, which shows the relevance of $c^{min} \leq 0.1$ in comparison with the energies of the canonical MOs. By considering all occupied MOs of the monomer bases, the FMO-LCMO orbitals near HOMO[(d)]/LUMO[(d)] are drastically improved in the DM scheme. Figure 10 shows the dependence of the delocalized MO shapes near LUMO[(d)] on the threshold of the contribution index in the DM based LOMO:LUMO target schemes, and Fig. 11 shows those in the DM-based LOMO:LUMO+n target schemes with $c^{min} = 0.1$. The delocalized MO shapes with $c^{min} = 0.1$ reproduce those of the canonical shapes well, which also indicates the relevance of the threshold parameter. The delocalized MO energy levels also appear to be recovered through the EL scheme by including all occupied orbitals into the bases (see Table IV). However, the actual shapes of the delocalized orbitals, particularly LUMO[(d)] and its uppers, are different from those in the DM scheme, indicating that the canonical orbitals are not well reproduced, as shown in Fig. 12. This shows, as suggested by the distribution of the monomer bases in the DM scheme in Fig. 7, that the monomer bases necessary for the description of the orbitals near LUMO[(d)] are insufficiently considered. Thus, the EL scheme cannot provide a relevant solution to how many and in what manner high-energy orbitals of monomer bases should be considered, which can be replicated in the DM scheme.

To advance the discussion, we now perform a model calculation in which the target orbitals are confined to those near the LUMO, which is referred to as the None:LUMO+n target scheme. The distribution of the monomer MOs selected as bases is shown in Fig. 13, and the energy levels of corresponding monomer MOs are found in Fig. S8 in the Supplemental Material section.[31] The monomer MO bases selected in the DM scheme are then only unoccupied orbitals. However, as shown in Fig. 14, some of the LCMO orbitals obtained with these monomer MO bases appear in the occupied orbital region, thus reflecting the orthogonality problem. Through a comparison with Fig. 8, we can see that the orbitals well above LUMO[(d)] are brought into the occupied orbital region owing to a lack of orthogonality in the occupied orbitals.
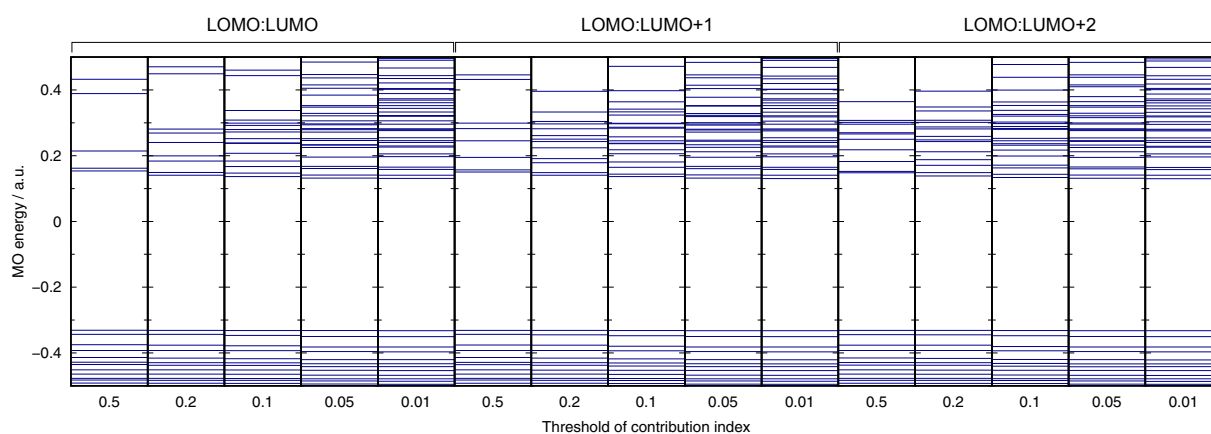
Fig. 8. (Color online) Delocalized MO energy levels determined through DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for pseudo-glycine pentamer.
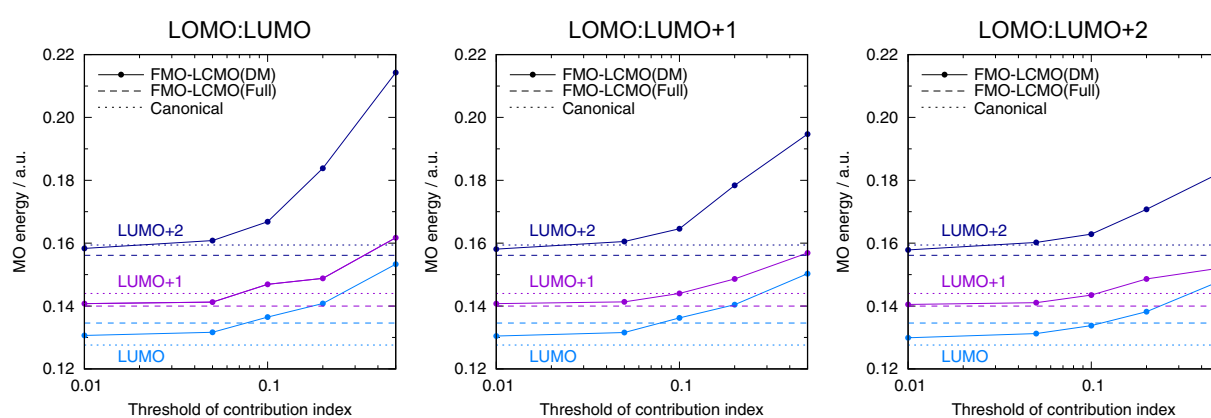

Fig. 9. (Color online) Dependence of delocalized MO energy levels near LUMO on the threshold of contribution index in DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for pseudo-glycine pentamer.
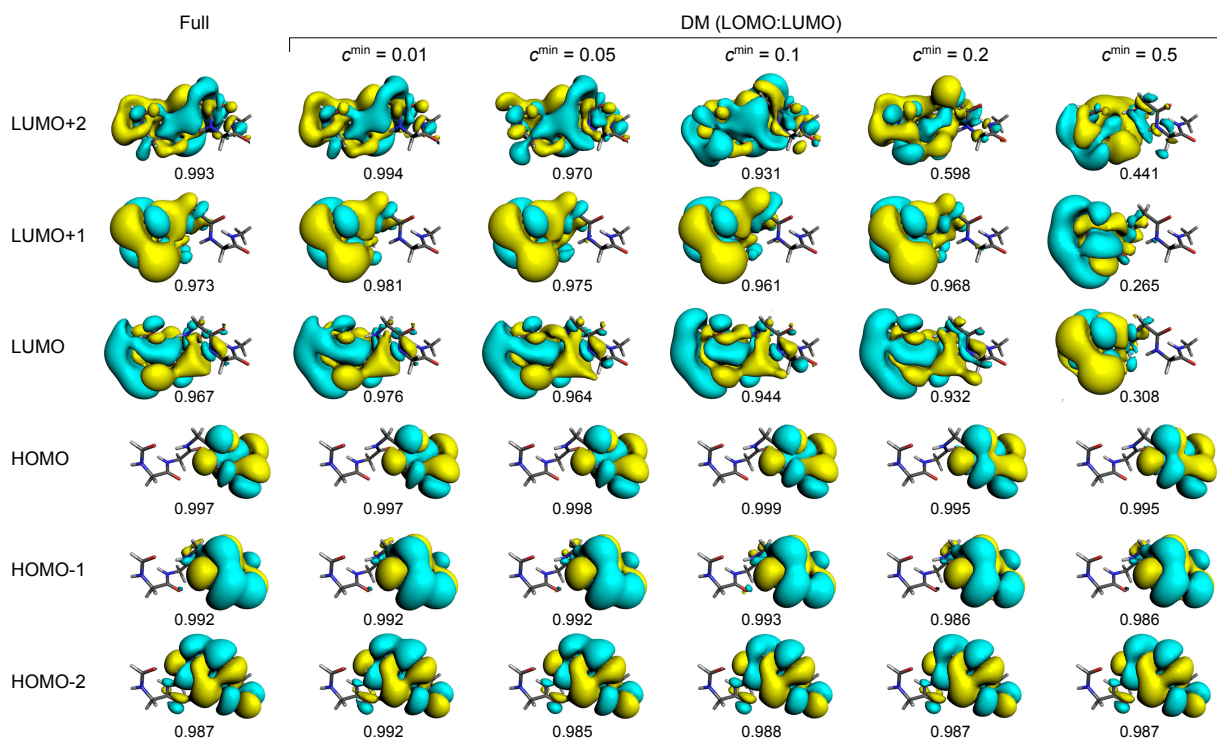

Fig. 10. (Color online) Delocalized MOs near frontier MOs generated by DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO while changing threshold parameter c^{min} for pseudo-glycine pentamer. The overlap integrals with canonical MOs are shown below each molecular orbital.
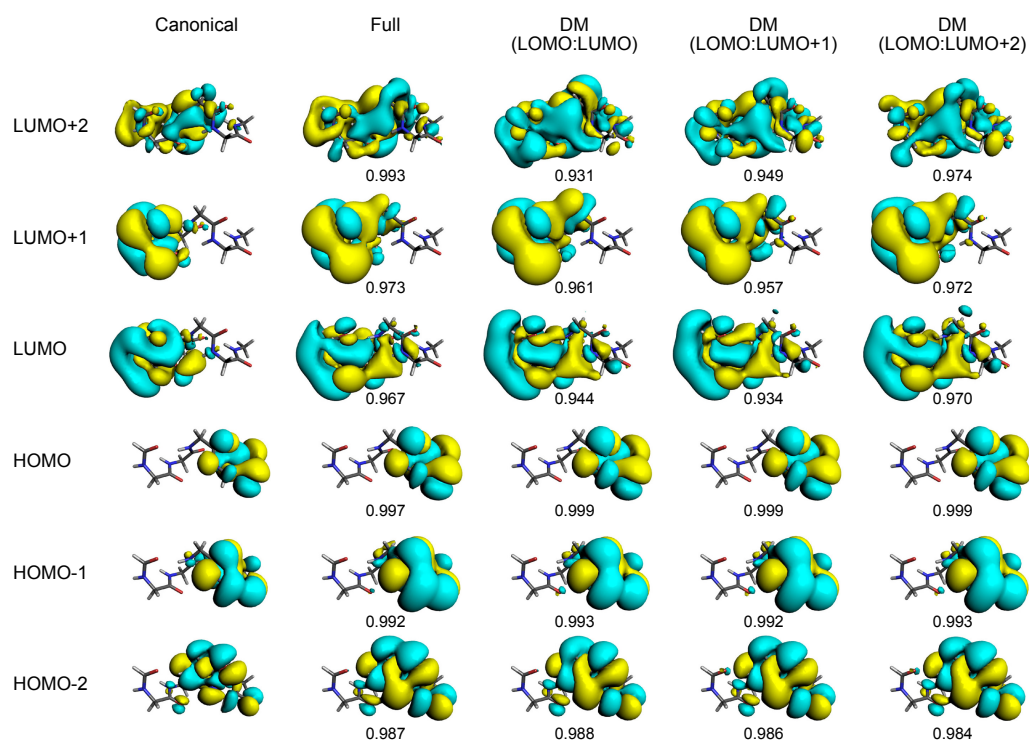
Fig. 11. (Color online) Delocalized MOs near frontier MOs generated by DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) with $c^{\min} = 0.1$ for pseudo-glycine pentamer. The overlap integrals with canonical MOs are shown below each molecular orbital.
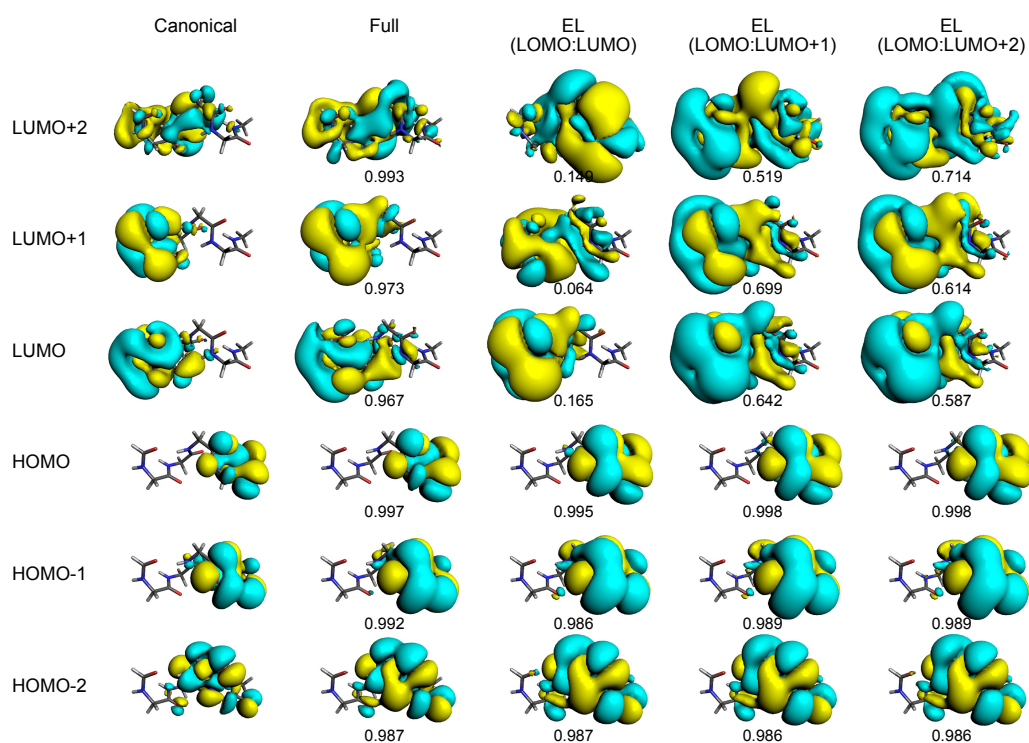

Fig. 12. (Color online) Delocalized MOs near frontier MOs generated by EL-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for pseudo-glycine pentamer. The overlap integrals with canonical MOs are shown below each molecular orbital.

Table IV. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for pseudo-glycine pentamer.

|	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	334	334	81	86	91	100	106	110
LUMO+2	0.159	0.156	0.388	0.195	0.182	0.167	0.165	0.163
LUMO+1	0.144	0.140	0.213	0.157	0.152	0.147	0.144	0.143
LUMO	0.135	0.128	0.154	0.150	0.148	0.136	0.136	0.134
HOMO	−0.344	−0.332	−0.331	−0.332	−0.332	−0.332	−0.332	−0.332
HOMO−1	−0.360	−0.351	−0.344	−0.345	−0.345	−0.347	−0.348	−0.348
HOMO−2	−0.385	−0.384	−0.375	−0.376	−0.376	−0.378	−0.379	−0.380
HOMO/LUMO gap	0.478	0.460	0.485	0.482	0.479	0.468	0.468	0.466
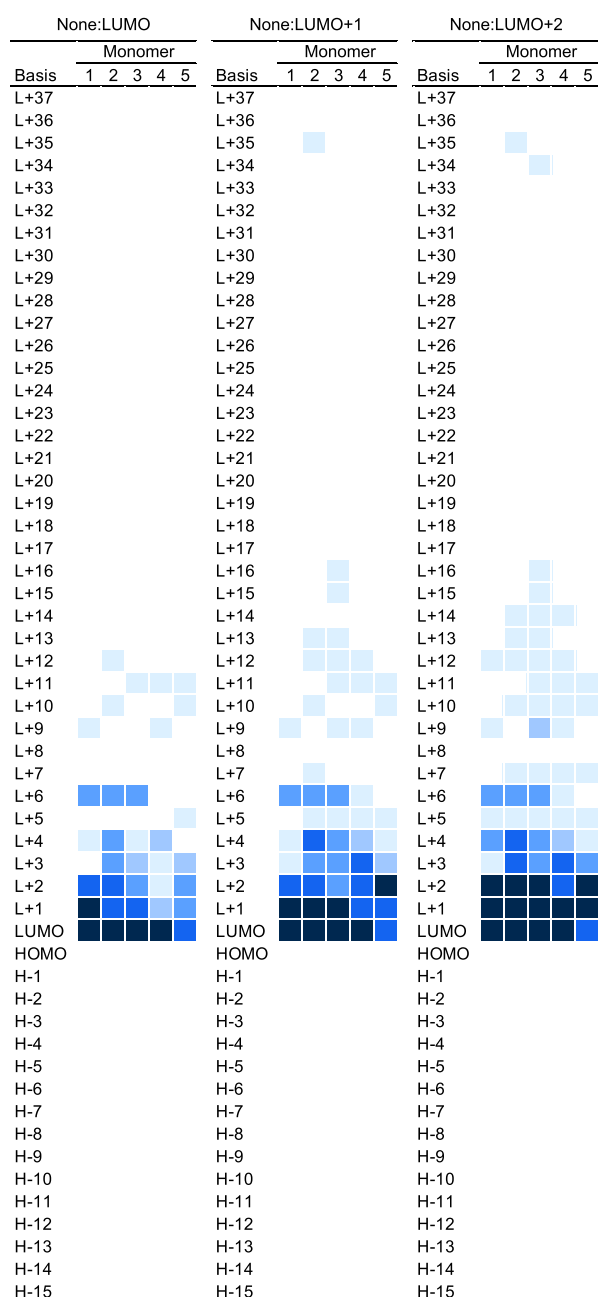
Fig. 13. (Color online) Monomer MO distribution as selected bases in pseudo-glycine pentamer, determined through DM-based FMO-LCMO scheme targeting the MO energy window of None:LUMO+n ($n = 0, 1, 2$). The color coding follows the legend in Fig. 3. The basis labels of MO energy levels below the HOMO and above the LUMO are abbreviated as H−k and L+k ($k = 1, 2, \ldots$), respectively.

3.3 Water clusters

Finally, we consider water clusters of dimer, trimer, and tetramer,[32] as shown in Fig. 15. Note that the water molecules as fragments are situated in a heterogeneous environment in the dimer and trimer cases.

Figures 16–18 illustrate the calculated results for LCMO energy levels near LUMO$^{(d)}$ as functions of the threshold of the contribution index (c^{min}) in the DM scheme for LOMO:LUMO+n targets. It is apparent that different behaviors were observed between the water dimer and trimer/tetramer cases, and the c^{min} dependence in the latter is not smooth but somewhat discrete. This suggests that there are MO bases that are not regarded as important in the fragment dimer level but play a significant role in the LCMO orbitals in the cases of trimers and tetramers. This fact may also indicate the necessity of a description beyond the FMO2 approximation for water trimer and tetramer systems.[18]

We carried out the HOMO−n:LUMO+n and LOMO:LUMO+n target calculations in both the EL- and DM-based schemes, the results of which are compiled in Tables V–X. Thus, it can be observed that the reproduction of the canonical results is unsatisfactory near LUMO$^{(d)}$ when we employ the HOMO−n:LUMO+n target. However, we can achieve a satisfactory reproduction if we employ the LOMO:LUMO+n target to accommodate LCMO orbitals that should be appropriately described. Because there still remain certain ambiguities in the present study owing to the neglect of higher-order effects beyond the fragment dimer approximation, we may need further investigations concerning this issue.

4. Conclusion

Whereas the present basic study has focused on relatively small molecular systems, it is becoming increasingly difficult to apply diagonalization with monomer-based orbitals when one constructs an FMO-LCMO for large-scale molecular systems, in which the full inclusion of the monomer MOs will be intractable. One of our strategies is to confine the basis space to that near the HOMO/LUMO, although we need to fully account for the occupied basis orbitals when the orthogonality with respect to the delocalized occupied orbitals is important. Under these circumstances, it is essential to efficiently select the basis space necessary for the relevant description of an LCMO. As observed above for the distribution of monomer MO bases selected in the DM scheme, the basis selection in the EL scheme lacks a rational justification for reproducing the orbitals near HOMO$^{(d)}$/LUMO$^{(d)}$; That is, the orbital energy
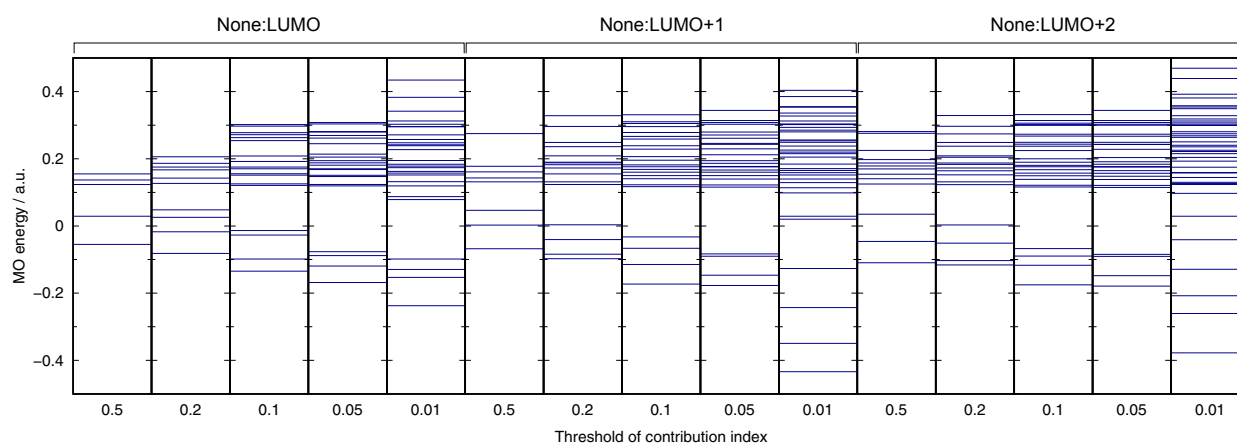
Fig. 14. (Color online) Delocalized MO energy levels determined by DM based FMO-LCMO scheme targeting the MO energy window of None:LUMO+n ($n = 0, 1, 2$) for pseudo-glycine pentamer.
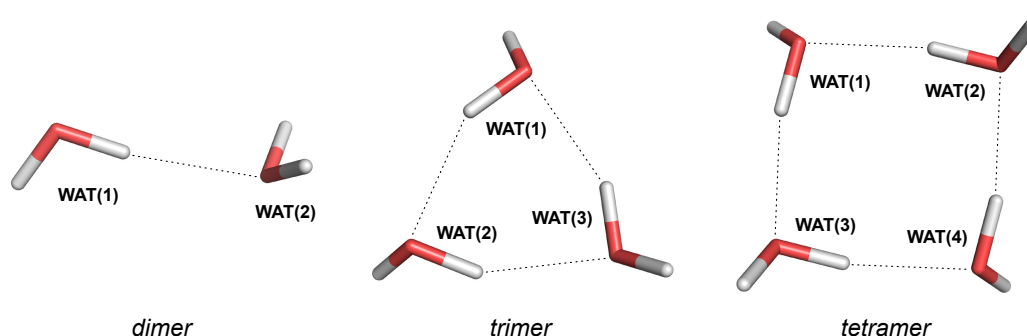

Fig. 15. (Color online) Structures of water dimer, trimer, and tetramer. They are in the C_s symmetry with a hydrogen bond, in the C_1 symmetry with three hydrogen bonds, and in the S_4 symmetry with four hydrogen bonds, respectively. Water molecules are written with the fragment number in parentheses.
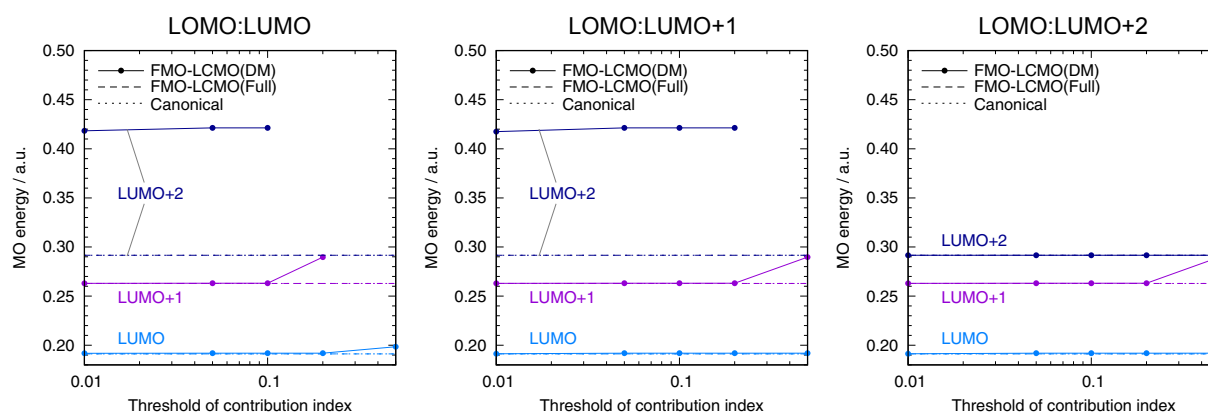

Fig. 16. (Color online) Dependence of delocalized MO energy levels near LUMO on the threshold of contribution index in DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for water dimer.

range necessary for a relevant description varies according to the specific molecular systems. By contrast, the DM scheme can provide a more robust methodology for selecting the monomer MO bases because it relies on a universal threshold of contribution index, irrespective of molecular systems. Thus, we found an advantage of the DM scheme such that it enables a rational selection of monomer MO bases to accurately construct the canonical orbitals for the entire system. Through the present analysis, we found a threshold value of approximately $c^{\min} = 0.1$ to be a reasonable choice in light of the compromise between the accuracy and cost of diagonalization.

Thus, the present density-matrix-based FMO-LCMO method is superior to the energy-based method in terms of the efficiency of data compression of monomer MO bases adopted to construct the delocalized MOs for the entire system. Our criterion for selecting the localized MOs of fragment monomers can provide a systematic way to control the cost performance of an approximation.

Fig. 17. (Color online) Dependence of delocalized MO energy levels near LUMO on the threshold of contribution index in DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for water trimer.
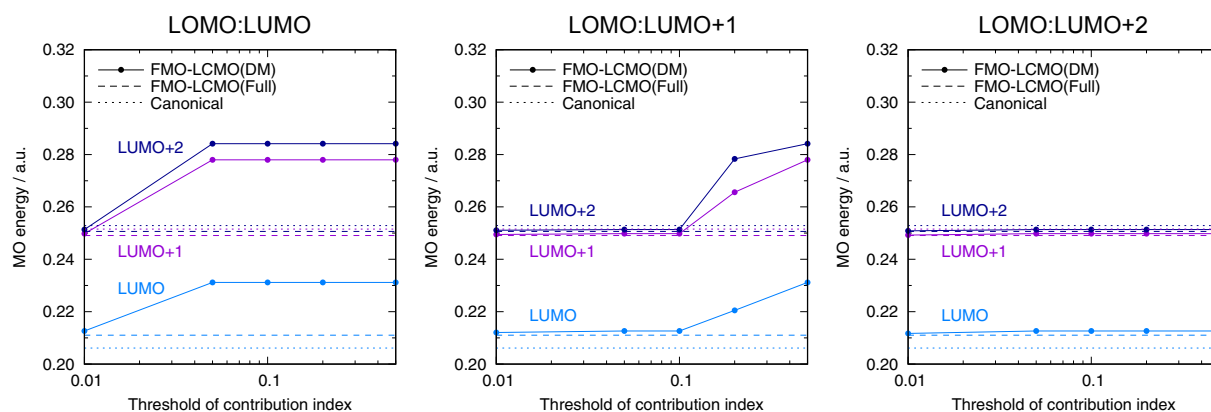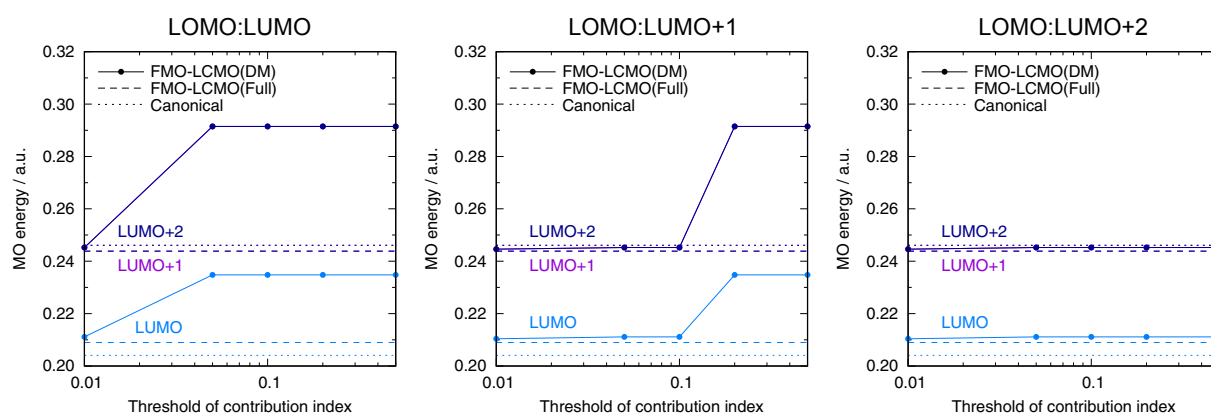

Fig. 18. (Color online) Dependence of delocalized MO energy levels near LUMO on the threshold of contribution index in DM-based FMO-LCMO scheme targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) for water tetramer.

Table V. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO−n:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water dimer.

| | Canonical | Full | EL | | | DM ($c^{\min} = 0.1$) | | |
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	38	38	4	8	12	3	6	7
LUMO+2	0.292	0.292		0.292	0.292		0.303	0.292
LUMO+1	0.263	0.263	0.231	0.248	0.249	0.215	0.230	0.230
LUMO	0.191	0.191	0.182	0.185	0.186	0.171	0.178	0.178
HOMO	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470
HOMO−1	−0.520	−0.520	−0.531	−0.523	−0.520		−0.524	−0.524
HOMO−2	−0.548	−0.548		−0.549	−0.548		−0.551	−0.551
HOMO/LUMO gap	0.662	0.662	0.652	0.656	0.657	0.641	0.648	0.649

Table VI. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water dimer.

| | Canonical | Full | EL | | | DM ($c^{\min} = 0.1$) | | |
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	38	38	12	14	16	13	13	14
LUMO+2	0.292	0.292		0.292	0.292	0.421	0.421	0.292
LUMO+1	0.263	0.263	0.290	0.263	0.263	0.263	0.263	0.263
LUMO	0.191	0.191	0.192	0.192	0.192	0.192	0.192	0.192
HOMO	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470	−0.470
HOMO−1	−0.520	−0.520	−0.520	−0.520	−0.520	−0.520	−0.520	−0.520
HOMO−2	−0.548	−0.548	−0.548	−0.548	−0.548	−0.548	−0.548	−0.548
HOMO/LUMO gap	0.662	0.662	0.662	0.662	0.662	0.662	0.662	0.662

Y. Okiyama et al.

Table VII. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO−n:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water trimer.

	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	57	57	6	12	18	6	9	12
LUMO+2	0.253	0.251	0.202	0.219	0.228	0.202	0.182	0.219
LUMO+1	0.252	0.249	0.199	0.216	0.225	0.199	0.179	0.216
LUMO	0.206	0.211	0.164	0.177	0.184	0.164	0.162	0.177
HOMO	−0.493	−0.493	−0.496	−0.495	−0.493	−0.496	−0.497	−0.495
HOMO−1	−0.496	−0.495	−0.500	−0.499	−0.496	−0.500	−0.501	−0.499
HOMO−2	−0.508	−0.507	−0.516	−0.514	−0.508	−0.516	−0.516	−0.514
HOMO/LUMO gap	0.699	0.704	0.660	0.671	0.676	0.660	0.658	0.671

Table VIII. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water trimer.

	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	57	57	18	21	24	18	21	21
LUMO+2	0.253	0.251	0.284	0.251	0.251	0.284	0.251	0.251
LUMO+1	0.252	0.249	0.278	0.250	0.250	0.278	0.250	0.250
LUMO	0.206	0.211	0.231	0.213	0.212	0.231	0.213	0.213
HOMO	−0.493	−0.493	−0.492	−0.492	−0.492	−0.492	−0.492	−0.492
HOMO−1	−0.496	−0.495	−0.495	−0.495	−0.495	−0.495	−0.495	−0.495
HOMO−2	−0.508	−0.507	−0.507	−0.507	−0.507	−0.507	−0.507	−0.507
HOMO/LUMO gap	0.699	0.704	0.723	0.705	0.704	0.723	0.705	0.705

Table IX. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of HOMO−n:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water tetramer.

	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	76	76	8	16	24	8	12	16
LUMO+2	0.246	0.244	0.171	0.220	0.221	0.171	0.150	0.220
LUMO+1	0.246	0.244	0.161	0.170	0.181	0.161	0.128	0.170
LUMO	0.204	0.209	0.135	0.162	0.166	0.135	0.110	0.162
HOMO	−0.493	−0.493	−0.499	−0.498	−0.493	−0.499	−0.499	−0.498
HOMO−1	−0.496	−0.496	−0.502	−0.499	−0.496	−0.502	−0.502	−0.499
HOMO−2	−0.496	−0.496	−0.502	−0.499	−0.496	−0.502	−0.502	−0.499
HOMO/LUMO gap	0.698	0.702	0.633	0.660	0.659	0.633	0.609	0.660

Table X. Comparison of delocalized MO energy levels in a.u. by EL/DM-based FMO-LCMO schemes targeting the MO energy window of LOMO:LUMO+n ($n = 0, 1, 2$) to those by full FMO-LCMO scheme using all monomer MO bases, and canonical MO scheme, for water tetramer.

	Canonical	Full	EL			DM ($c^{min} = 0.1$)		
			$n = 0$	$n = 1$	$n = 2$	$n = 0$	$n = 1$	$n = 2$
Number of basis	76	76	24	28	32	24	28	28
LUMO+2	0.246	0.244	0.291	0.245	0.245	0.291	0.245	0.245
LUMO+1	0.246	0.244	0.291	0.245	0.245	0.291	0.245	0.245
LUMO	0.204	0.209	0.235	0.211	0.211	0.235	0.211	0.211
HOMO	−0.493	−0.493	−0.492	−0.493	−0.493	−0.492	−0.493	−0.493
HOMO−1	−0.496	−0.496	−0.495	−0.495	−0.495	−0.495	−0.495	−0.495
HOMO−2	−0.496	−0.496	−0.495	−0.495	−0.495	−0.495	−0.495	−0.495
HOMO/LUMO gap	0.698	0.702	0.727	0.704	0.704	0.727	0.704	0.704

*okiyama@nihs.go.jp

†tanaka2@kobe-u.ac.jp

1) *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, ed. M. S. Gordon (Wiley, Chichester, U.K., 2017).
2) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
3) T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).
4) D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).
5) S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, and K. Fukuzawa, Phys. Chem. Chem. Phys. **16**, 10310 (2014).
6) Y. Inadomi, T. Nakano, K. Kitaura, and U. Nagashima, Chem. Phys. Lett. **364**, 139 (2002).
7) S. Tsuneyuki, T. Kobori, K. Akagi, K. Sodeyama, K. Terakura, and H. Fukuyama, Chem. Phys. Lett. **476**, 104 (2009).
8) T. Kobori, K. Sodeyama, T. Otsuka, Y. Tateyama, and S. Tsuneyuki, J. Chem. Phys. **139**, 094113 (2013).
9) H. Nishioka and K. Ando, J. Chem. Phys. **134**, 204109 (2011).
10) H. Kitoh-Nishioka and K. Ando, Chem. Phys. Lett. **621**, 96 (2015).
11) H. Kitoh-Nishioka, K. Welke, Y. Nishimoto, D. G. Fedorov, and S. Irle, J. Phys. Chem. C **121**, 17712 (2017).
12) R. Sato, H. Kitoh-Nishioka, T. Yanai, and Y. Shigeta, Chem. Lett. **46**, 873 (2017).
13) H. Kitoh-Nishioka, Y. Shigeta, and K. Ando, J. Chem. Phys. **153**, 104104 (2020).
14) T. Fujita and Y. Noguchi, Phys. Rev. B **98**, 205140 (2018).
15) Y. Suzuki, K. Ebina, and S. Tanaka, Chem. Phys. **474**, 18 (2016).
16) Y. Suzuki, H. Watanabe, Y. Okiyama, K. Ebina, and S. Tanaka, Chem. Phys. **539**, 110903 (2020).
17) D. G. Fedorov and K. Kitaura, Chem. Phys. Lett. **433**, 182 (2006).
18) T. Fujita, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **478**, 295 (2009).
19) T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, and S. Tanaka, Chem. Phys. Lett. **523**, 128 (2012).
20) D. G. Fedorov and K. Kitaura, J. Chem. Phys. **147**, 104106 (2017).
21) The idempotency is not necessarily required for this TFDM.
22) S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
23) S. R. White, Phys. Rev. B **48**, 10345 (1993).
24) T. Xiang, Phys. Rev. B **53**, R10445 (1996).
25) V. Klema and A. Laub, IEEE Trans. Autom. Control **25**, 164 (1980).
26) L. De Lathauwer, B. De Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1253 (2000).
27) F. Kleibergen and R. Paap, J. Econ. **133**, 97 (2006).
28) M. Brand, Linear Algebra Appl. **415**, 20 (2006).
29) T. Nakano, Y. Mochizuki, K. Fukuzawa, S. Amari, and S. Tanaka, Developments and Applications of ABINIT-MP Software Based on the Fragment Molecular Orbital Method, in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, ed. E. B. Starikov, J. P. Lewis, and S. Tanaka (Elsevier, Amsterdam, 2006) Chap. 2, p. 39.
30) K. Fukuzawa, C. Watanabe, I. Kurisaki, N. Taguchi, Y. Mochizuki, T. Nakano, S. Tanaka, and Y. Komeiji, Comput. Theor. Chem. **1034**, 7 (2014).
31) (Supplemental Material) The monomer MO energy levels of CG stacked base pair of DNA; the molecular structures and FMO-LCMO results of AA, AT, GA stacked base pairs of DNA; the monomer MO energy levels of pseudo-glycine pentamer are provided online.
32) S. Maheshwary, N. Patel, N. Sathyamurthy, A. D. Kulkarni, and S. R. Gadre, J. Phys. Chem. A **105**, 10525 (2001).