

PCCP

Accepted Manuscript

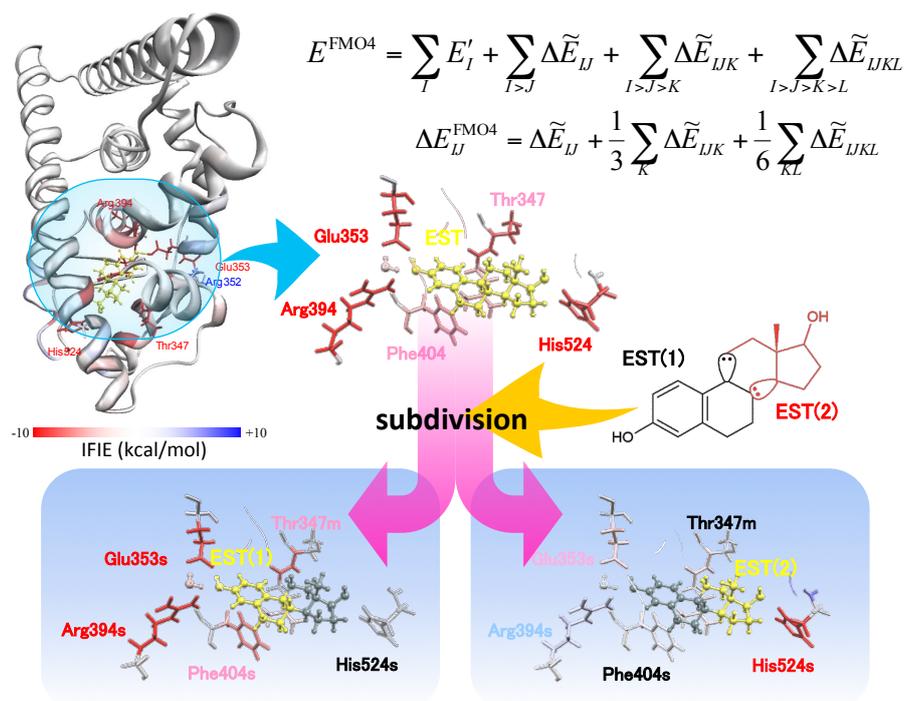


This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



One can perform the interaction energy analysis of protein-ligand system in atomic details on the basis of the fragment molecular orbital method.

Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems

Shigenori Tanaka,^{*a} Yuji Mochizuki,^{*b,d} Yuto Komeiji,^c Yoshio Okiyama,^d Kaori Fukuzawa^{e,d}

^aGraduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan; tanaka2@kobe-u.ac.jp

^bDepartment of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan; fullmoon@rikkyo.ac.jp

^cBiomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, Tsukuba, Ibaraki 305-8568, Japan

^dInstitute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

^eMizuho Information and Research Institute Inc., 2-3 Kanda Nishiki-cho, Chiyoda-ku, Tokyo 101-8442, Japan

Abstract

Recent developments in the fragment molecular orbital (FMO) method for theoretical formulation, implementation, and application to nano and biomolecular systems are reviewed. The FMO method has enabled *ab initio* quantum-mechanical calculations for large molecular systems such as protein-ligand complex at reasonable computational cost in a parallelized way. There have been a wealth of application outcomes by the FMO method in the fields of biochemistry, medicinal chemistry and nanotechnology, in which the electron correlation effects play vital roles. With the aid of the advances in high-performance computing, the FMO method promises larger, faster, and more accurate simulations of biomolecular and related systems, including the descriptions of dynamical behaviors in solvent environments. The current status and future prospects of the FMO scheme are addressed in these contexts.

1. Introduction

The fragment molecular orbital (FMO) method was proposed in 1999 by Kitaura *et al.* [1] to realize fully quantum-mechanical (QM) calculations for large-scale molecular systems like proteins at the practical cost of computations through good parallelism. FMO-related methodological developments and associated applications have been achieved extensively so far, as reviewed in the literature [2-9]. In particular, various other methods based on fragmentation schemes were also reviewed by Gordon *et al.* [9] in a comparative fashion, and thus we will focus the present Perspective Article only on the FMO method except for some specific cases.

In the original two-body FMO treatment [1], the target molecular system is divided into appropriate fragments, where the respective amino acid residues in proteins or water molecules in hydration clusters are typically defined as a fragment. As described in detail later, a series of parallelized molecular orbital (MO) calculations are performed on the fragment monomers and dimers under the presence of environmental electrostatic potential (ESP) to ensure chemical reliability [5-7]. The Hartree-Fock (HF) calculations of all monomers are repeated until the self-consistent charge (SCC) condition is satisfied, thus describing the electronic polarization of system. The electron delocalizations are then taken into account by the dimer calculations.

The development and distribution of the FMO methodology have been promoted by its implementation in various quantum chemistry programs. FMO was originally implemented with a local version of GAUSSIAN [1], but soon ABINIT-MP [10] and later GAMESS [2] became the main development platforms of FMO. In the ABINIT-MP program, originally developed for the FMO-HF calculations by Nakano *et al.* [10-12], the standard message passing interface (MPI) was adopted for parallelization. It is notable that the crucial techniques of bond-detachment atom (BDA) to cut single bonds [10] as well as efficient ESP approximation [11] were first devised in ABINIT-MP and that the first attempt to obtain the entire MO set from the FMO-based density (FMO-MO) [12] was made as well. The FMO methodology was implemented into the GAMESS program [13] first for HF calculations [14]. The parallel control for both the fragments and the Fock matrix constructions in FMO-HF was done with the generalized distributed-data interface (GDDI) [15]. The electron correlation methods such as the second-order Møller-Plesset perturbation theory (MP2) have been augmented to obtain the quantitative reliability of FMO calculations in both GAMESS and ABINIT-MP, as discussed later. The FMO method has also been implemented in several other programs including NWChem (development version) [16-18], PAICS [19],

and OpenFMO [20,21].

One of the most useful quantities obtained by FMO calculations is a set of the fragment-fragment interaction energies, by which chemical insights into residue-residue and residue-ligand interactions in protein-ligand systems can be illuminated. Such detailed views are informative in biochemistry and pharmacology [5,7], and this may be a key factor for why the FMO scheme has been utilized widely [9]. Conventionally, two different names have been used for the interaction energies between fragments; pair interaction energy (PIE) for GAMESS [22] and inter-fragment interaction energy (IFIE) for ABINIT-MP [23,24]. A couple of modifications for the IFIEs were also attempted later [25-28].

On the basis of current status of the methodologies, this Perspective Article addresses the recent developments in basic theories, implementations, accelerations and applications in the FMO method. Dependable treatments of electron correlation effects beyond the mean-field HF approximation play vital roles for the quantitative descriptions of molecular recognition and excited states in biomolecular systems. The extension to the description of time-dependent behaviors by *ab initio* (or first principles) molecular dynamics (MD) should be also essential for understanding the biomolecular functions such as protein folding, aggregation [29], ligand binding [30-33] and enzymatic reaction [34,35]. In order to obtain useful results with given computational resources, some acceleration and parallelization techniques should be explored. A number of intriguing methods for eliciting useful information from the FMO calculations have also been developed. All these combinations have been extensively employed in actual applications in the fields of biochemistry, medicinal chemistry and nanotechnology. This review article will demonstrate them through a variety of examples, many of which have been taken from the outcomes by our own groups.

In Sec. 2, the basic framework of the FMO theory will be reviewed, where the descriptions are performed in somewhat different manners from those seen in other related review articles and books [2-9]. Section 3 is then devoted to the illustration of application examples, in which the readers will find how to apply the FMO method to get the solution to actual problems; future directions inspired through actual researches will also be addressed. Section 4 will conclude this article with a summary.

2. Basics of fragment molecular orbital theory

2.1. General formulation of FMO method

The FMO method was originally formulated on the basis of the fragment pair approximation, which is currently referred to as the FMO2 method [7], and later extended to higher-order (FMO3 and FMO4) ones [2,27,36]. Theoretical background of the FMO method is briefly summarized below.

In the FMO method for a molecular system divided into N_f fragments (monomers), the Schrödinger equation for monomer I is solved to obtain eigenfunction Ψ_I associated with the eigenvalue E_I of the Hamiltonian H_I as follows [10]:

$$H_I \Psi_I = E_I \Psi_I \quad (1)$$

$$H_I = \sum_{i \in I} \left[\left(-\frac{1}{2} \nabla_i^2 \right) + \sum_A \left(-\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \right) + \sum_{J \neq I} \int \frac{\rho_J(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' \right] + \sum_{i \in I} \sum_{i > j \in I} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2)$$

where $\rho_J(\mathbf{r}')$ shows the electron density at position \mathbf{r}' on monomer J and \mathbf{R}_A refers to the nuclear coordinate of atom A with positive charge Z_A . This Hamiltonian includes the electrostatic potential generated by $N_f - 1$ monomers surrounding the fragment monomer I . In solving eqn (1), all the MOs on a monomer are optimized by self-consistent-field (SCF) theory [37] and should be localized within the monomer [10], which will be addressed later. At the same time, all the electron densities on all monomers are also self-consistently solved and converged through the SCC iterations [38]. In the same way as for the monomer, the Schrödinger equation for fragment pair (dimer) IJ is also solved so that the MOs are localized within the dimer:

$$H_{IJ} \Psi_{IJ} = E_{IJ} \Psi_{IJ}. \quad (3)$$

The only difference of the Hamiltonian of dimer from that of monomer is that the former includes the electrostatic potential generated by surrounding $N_f - 2$ monomers. In the two-body approximation including the effect of fragment pairs, FMO2, the total electron energy of the whole molecule is expressed using the total electron energies of monomers and dimers (E_I and E_{IJ}) as

$$E \cong \sum_{I > J} E_{IJ} - (N_f - 2) \sum_I E_I. \quad (4)$$

Additionally, the total electron density $\rho(\mathbf{r})$ is also obtained in a similar way [38] as

$$\rho(\mathbf{r}) \cong \sum_{I > J} \rho_{IJ}(\mathbf{r}) - (N_f - 2) \sum_I \rho_I(\mathbf{r}). \quad (5)$$

We will first describe the detail of the computational scheme in the case of HF calculation with the FMO method (FMO-HF) below. All the monomers and dimers are here assumed to have the closed-shell structures for their electrons, while the FMO method itself is applicable to the electronic systems with open-shell structure or to excited states as well. The Hartree-Fock-Roothaan equations for monomers ($x=D$) and dimers ($x=LD$) are represented using the modified Fock operator as follows [11]:

$$\mathbf{F}^x \mathbf{C}^x = \mathbf{S}^x \mathbf{C}^x \boldsymbol{\varepsilon}^x \quad (6)$$

$$\mathbf{F}^x = \mathbf{H}^x + \mathbf{G}^x \quad (7)$$

$$H_{pq}^x = H_{pq}^{\text{core } x} + V_{pq}^x + \sum_k B_k \langle p | \theta_k \rangle \langle \theta_k | q \rangle \quad (8)$$

$$V_{pq}^x = \sum_{K \neq x} (u_{pq}^K + v_{pq}^K) \quad (9)$$

$$u_{pq}^K = \sum_{A \in K} \langle p | (-Z_A / |\mathbf{r} - \mathbf{R}_A|) | q \rangle \quad (10)$$

$$v_{pq}^K = \sum_{rs \in K} P_{rs}^K(pq, rs) \quad (11)$$

$$P_{pq}^x = \sum_i^{\text{occ}} 2C_{pi}^{x*} C_{qi}^x \quad (12)$$

$$G_{pq}^x = \sum_{rs} P_{rs}^x \left[(pq, rs) - \frac{1}{2}(ps, rq) \right] \quad (13)$$

where \mathbf{V}^x is the electrostatic potential from surrounding fragments, which consists of the potential generated by nuclei u^K and that by electrons v^K ; \mathbf{P}^x and $pqrs$ for four-center two-electron integral refer to the density matrix and the indices of atomic orbitals (AOs) as the basis functions to expand MOs. The third term on the right-hand side of eqn (8) represents a projection operator to remove the orbital θ_k from the variational space with parameter B_k set to 10^6 . Regarding the fragmentation for a molecule, the sp^3 carbon site is usually employed for the division. The atom at the division site is called the bond detached atom (BDA). As for θ_k , the localized MOs for a methane molecule obtained with the natural localized MO method are used, in which the C-H distance is fixed at 1.09Å. We then consider a fragment I to which the BDA belongs and a neighboring fragment J to which the atom bonding to the BDA belongs. By rotating the MOs so that one of the orbitals θ_l is directed toward the atom in fragment J bonding to the BDA, the

contribution of θ_I to fragment I and that of other MOs to fragment J are removed by the shift operator, respectively. This procedure for the restriction of basis set enables the localization of MOs within a fragment [5,39] (Fig. 1).

In order to accelerate the FMO calculations, two additional approximations, one for the environmental electrostatic potentials and the other for the electrostatic interactions between distant fragment pairs, are very helpful [11]. Let us consider the expression for the electronic energy in the FMO2 method as

$$\begin{aligned}
 E^{\text{FMO2}} &= \sum_{I>J} E_{IJ} - (N_f - 2) \sum_I E_I \\
 &= \sum_{I>J} (E_{IJ} - E_I - E_J) + \sum_I E_I \\
 &= \sum_{I>J} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J} (V_{IJ} - V_I - V_J) + \sum_I E'_I + \sum_I V_I \\
 &= \sum_I E'_I + \sum_{I>J} [\Delta E'_{IJ} + \text{Tr}(\Delta \mathbf{P}^{IJ} \mathbf{V}^{IJ})]
 \end{aligned} \tag{14}$$

with

$$E'_x = E_x - V_x \tag{15}$$

$$V_x = \text{Tr}(\mathbf{P}^x \mathbf{V}^x) \tag{16}$$

$$\Delta E'_{IJ} = E'_{IJ} - E'_I - E'_J \tag{17}$$

$$\Delta \mathbf{P}^{IJ}_{pq} = \mathbf{P}^{IJ}_{pq} - \mathbf{P}^I_{pq} \oplus \mathbf{P}^J_{pq} \tag{18}$$

$$\begin{aligned}
 \mathbf{P}^I_{pq} \oplus \mathbf{P}^J_{pq} &= \mathbf{P}^I_{pq} + \mathbf{P}^J_{pq} \quad \text{for } p, q \in I, J, \\
 &= \mathbf{P}^I_{pq} \quad \text{for } p, q \in I \text{ and } p, q \notin J, \\
 &= \mathbf{P}^J_{pq} \quad \text{for } p, q \in J \text{ and } p, q \notin I, \\
 &= 0 \quad \text{for all other cases.}
 \end{aligned} \tag{19}$$

Since eqn (14) contains the electrostatic potentials only for the fragment dimers, it enables us to use different approximations between the monomers and dimers for the electrostatic potentials. We then define the IFIE or PIE [22-24], $\Delta \tilde{E}'_{IJ}$, as

$$\Delta \tilde{E}'_{IJ} = \Delta E'_{IJ} + \text{Tr}(\Delta \mathbf{P}^{IJ} \mathbf{V}^{IJ}) \tag{20}$$

which can be regarded as an effective interaction between fragments I and J , and useful for, *e.g.*, the interaction analysis between a ligand molecule and amino acid residues.

The electrostatic potentials from the fragments located at relatively short distances can be approximated (esp-aoc approximation [11]) as

$$v_{pq}^K \cong \sum_{r \in K} (\mathbf{P}^K \mathbf{S}^K)_{rr} (pq, rr) \text{ for } R_{\min}(x, K) \geq L_{\text{aoc}} \quad (21)$$

by using the Mulliken approximation. On the other hand, the electrostatic potentials from the fragments located at long distances can be approximated (esp-ptc approximation [11]) as

$$v_{pq}^K \cong \sum_{A \in K} \langle p | (Q_A / |\mathbf{r} - \mathbf{R}_A|) | q \rangle \text{ for } R_{\min}(x, K) \geq L_{\text{ptc}} \quad (22)$$

by using the Mulliken atomic charge

$$Q_A = \sum_{r \in A} (\mathbf{P}^K \mathbf{S}^K)_{rr} . \quad (23)$$

We can also employ an approximation for the electrostatic interaction between distant fragment pairs (dimer-es approximation [11]) as

$$E'_{IJ} \cong E'_I + E'_J + \text{Tr}(\mathbf{P}^I \mathbf{u}^J) + \text{Tr}(\mathbf{P}^J \mathbf{u}^I) + \sum_{pq \in I} \sum_{rs \in J} P_{pq}^I P_{rs}^J (pq, rs) \quad (24)$$

which enables us to skip the SCF calculations for many dimers and significantly reduces the computational cost from $\mathcal{O}(N^{3.4})$ to $\mathcal{O}(N^2)$. It would further be possible to reduce the computational cost to $\mathcal{O}(N)$ through the introduction of some acceleration techniques such as the continuous multipole method (CMM) and the fast multipole method (FMM) [40].

In the cases when one considers the electron correlation effects by means of, *e.g.*, size-consistent MP2 calculations, the FMO2 energy can be expressed as (see Sec. 2.2)

$$E^{\text{MP2}} \cong \sum_{I>J} E_{IJ}^{\text{MP2}} - (N_f - 2) \sum_I E_I^{\text{MP2}} \quad (25)$$

It is also straightforward to extend the FMO2 method so that the contributions from the fragment trimers [2,36] and tetramers [27] are taken into account as

$$\begin{aligned}
E^{\text{FMO3}} &= \sum_{I>J>K} E_{IJK} - (N_f - 3) \sum_{I>J} E_{IJ} + \frac{(N_f - 2)(N_f - 3)}{2} \sum_I E_I \\
&= \sum_{I>J>K} [(E_{IJK} - E_I - E_J - E_K) - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK}] + \sum_{I>J} \Delta E_{IJ} + \sum_I E_I
\end{aligned} \tag{26}$$

$$\begin{aligned}
E^{\text{FMO4}} &= \sum_{I>J>K>L} E_{IJKL} - (N_f - 4) \sum_{I>J>K} E_{IJK} + \frac{(N_f - 3)(N_f - 4)}{2} \sum_{I>J} E_{IJ} - \frac{(N_f - 2)(N_f - 3)(N_f - 4)}{6} \sum_I E_I \\
&= \sum_{I>J>K>L} \{ (E_{IJKL} - E_I - E_J - E_K - E_L) \\
&\quad - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{IL} - \Delta E_{JK} - \Delta E_{JL} - \Delta E_{KL} \\
&\quad - [(E_{IJK} - E_I - E_J - E_K) - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK}] \\
&\quad - [(E_{IJL} - E_I - E_J - E_L) - \Delta E_{IJ} - \Delta E_{IL} - \Delta E_{JL}] \\
&\quad - [(E_{IKL} - E_I - E_K - E_L) - \Delta E_{IK} - \Delta E_{IL} - \Delta E_{KL}] \\
&\quad - [(E_{JKL} - E_J - E_K - E_L) - \Delta E_{JK} - \Delta E_{JL} - \Delta E_{KL}] \} \\
&\quad + \sum_{I>J>K} [(E_{IJK} - E_I - E_J - E_K) - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK}] \\
&\quad + \sum_{I>J} \Delta E_{IJ} + \sum_I E_I
\end{aligned} \tag{27}$$

for the total electronic energies, which are referred to as the FMO3 and FMO4 methods, respectively (see below). Further, the FMO calculations can be performed under the periodic boundary condition [41], in which the cell energy can be given by

$$\begin{aligned}
E^{\text{cell}} &= \sum_I E_I + \sum_{I>J} \Delta \tilde{E}_{IJ} + \sum_{I>J>K} \Delta \tilde{E}_{IJK} \\
&\quad + \sum_I \left[\sum_{n_z=-l}^{-1} \sum_{n_x=-l}^l \sum_{n_y=-l}^l \sum_{J'} \Delta \tilde{E}_{IJ'} + \sum_{n_x=-l+1}^l \sum_{n_y=-l}^{n_x-1} \sum_{J'} (\Delta \tilde{E}_{IJ'})_{n_z=0} + \sum_{n_x=-l}^{-1} \sum_{J'} (\Delta \tilde{E}_{IJ'})_{n_z=0}^{n_y=n_x} \right]
\end{aligned} \tag{28}$$

where n_x , n_y and $n_z = 0, \pm 1, \pm 2, \dots$ specify the real and image cells, and l means the number of layers of image cells.

Here, we will address in more detail the FMO4 method, a recent important development [27]. Note that, prior to FMO4, the four-body expansion of system energy was reported in the literature of incremental scheme (IS) [42], systematic molecular fragmentation (SMF) [43] and kernel energy method (KEM) [44], and that these studies stimulated the development of FMO4. The FMO method, based on the many-body expansion theory, has been often applied to large-scale computations of biomolecular systems with two-body approximation (FMO2) and for water clusters with three-body

one (FMO3) [2,36], thus describing the properties of the whole systems well. Usually, upon fragmentation of a protein-ligand complex system, an amino acid residue is regarded as a fragment, with its ligand also constituting a fragment. The IFIE analysis between these fragments in the complex has provided information on local energetic stability [24]. Nevertheless, some of these semantic fragments may contain multiple chemically meaningful functional groups (*e.g.*, -OH or -Ph) or biochemical functional sites. For example, an amino acid residue consists of a main-chain part peptide-bonded with neighbouring residues and a side-chain part showing non-covalent, specific interactions with various residues and ligands. In addition, a ligand substrate, *e.g.*, a potential drug, is virtually an aggregation of functional groups. Hence, more detailed and effective interaction analyses with well-defined site-specific views could be obtained by subdividing the fragments into smaller functional parts, but such additional division should seriously diminish energetic accuracy of FMO2 and even FMO3.

Recent introduction of FMO4 [27] has resolved this dilemma by enabling us to divide a molecular system more finely with keeping good accuracy. The formulation for FMO4 is written as

$$E^{\text{FMO4}} = \sum_I E'_I + \sum_{I>J} \Delta\tilde{E}_{IJ} + \sum_{I>J>K} \Delta\tilde{E}_{IJK} + \sum_{I>J>K>L} \Delta\tilde{E}_{IJKL} \quad (29)$$

where $\Delta\tilde{E}_{IJ}$, $\Delta\tilde{E}_{IJK}$ and $\Delta\tilde{E}_{IJKL}$ show the second, third and fourth terms of IFIE, respectively, as in eqn (27), and the pairwise IFIEs corrected with these many-body terms can be defined as follows:

$$\Delta E_{IJ}^{\text{FMO2}} = \Delta\tilde{E}_{IJ} \quad (30)$$

$$\Delta E_{IJ}^{\text{FMO3}} = \Delta E_{IJ}^{\text{FMO2}} + \frac{1}{3} \sum_K \Delta\tilde{E}_{IJK} \quad (31)$$

$$\Delta E_{IJ}^{\text{FMO4}} = \Delta E_{IJ}^{\text{FMO3}} + \frac{1}{6} \sum_{KL} \Delta\tilde{E}_{IJKL} \quad (32)$$

These many-body corrected IFIE analysis provides inter-fragment interactions with higher resolution which are favorable to the structure-based drug design (SBDD) or the fragment-based drug design (FBDD) [28]. In addition to biomolecular systems, inorganic systems such as silica cluster can be now treated as aggregations of fragments by using the FMO4 method, and the hybrid systems of biomolecules and inorganic materials are also targeted for the advancement of nano-biotechnology [45] (see Sec.

3.4).

2.2. Electron correlations

2.2.1. MP2, higher-order methods, and density functional theory (DFT)

The iterative SCF procedure [37] is necessary for the electronic structure calculations of molecular systems to optimize the occupied MO set of the ground state under a mean-field approximation at the single configuration condition at the HF level. Semiquantitative results for a given target system might be obtained by the HF-SCF calculations, but the electron correlation effect as a deviation from mean-field description should be taken into account to improve accuracy. Especially, the dispersion (or van der Waals) interactions, which play key roles in proteins as well as in molecular clusters, could be straightforwardly incorporated through the methods of wave-function expansions based on the HF reference configuration. The MP2 theory is the simplest wave-function approach, where the perturbing operator is set just as a fluctuation potential [37,46]. The size-consistency, an essential requirement of many-body perturbation methods, is maintained in the MP series [37,46-48]. The coupled-cluster (CC) approach [37,46,49] can be regarded as an infinite-order perturbation theory, and thus the corresponding results are expected to be better than those by MP2 at the increased computational cost. Since the FMO scheme provides an additive formula of fragment terms [1,5,6], the size-consistency must be ensured. Hence, MP2 is the first option of correlated FMO calculations (recall eqn (25) for instance).

The MP2 correlation energy is given by the following expression with the spin-orbital notation [37]

$$E^{\text{MP2}} = \frac{1}{4} \sum_{ijab} \frac{\langle ij || ab \rangle \langle ab || ij \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \quad (33)$$

where ij and ab indicate the HF occupied (internal) and virtual (external) orbitals, respectively. As can be seen, MP2 is described by the double excitations (hereafter simplified as doubles). The denominator of eqn (33) is the difference in occupied-virtual orbital energies, whereas the numerator is the square of anti-symmetrized two-electron integrals. As the result, the MP2 energy should be negative, yielding 80-90 % of the total non-relativistic correlation energy [46,48]. For actual implementations, eqn (33) is converted to a spatial orbital form after the spin-integration as

$$E^{\text{MP2}} = \sum_{ijab} \frac{(ia, jb)[2(ia, jb) - (ib, ja)]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}$$

$$, \quad (34)$$

and the integral list is generated by the transformation from AOs to MOs

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

This transformation can be carried out at the N^5 (N is the number of orbitals as a formal size parameter of a given system) cost via the quarter steps as below

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

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

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

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

These quarter transformations completely dominates the cost of MP2 calculations, and thus various algorithms and related codes including the parallelism as well as integral-direct processing have been developed as reviewed in the literature [48].

The FMO-MP2 calculation was reported in 2004 with GAMESS [50]. The importance of three-body fragment corrections (FMO3) [4] was addressed in Ref. [51], and the FMO3-MP2 ability [52] was implemented in GAMESS by incorporating a new efficient MP2 module [53]. At almost the same time as Ref. [50], the FMO-MP2 calculation was made available in ABINIT-MP [54-56]. By using a parallelized (by MPI) integral-direct (or IO-less) MP2 module, a large complex system of the HIV-1 protease (with 198 amino acid residues) and lopinavir (anti-viral agent) was computed at the FMO-MP2/6-31G level in only 14.3 hours by a total of 64 cores of that age [55], and this timing data could be of a breakthrough that realistic proteins can be routinely calculated with FMO-MP2 on in-house class parallel computers. Owing to this efficient parallelization, a number of application calculations were performed with ABINIT-MP (see, *e.g.*, Ref. [5]).

In 2008, the MP2 module in ABINIT-MP was modified in a style of matrix-matrix products with the highly optimized DGEMM routine of level-3 BLAS libraries [57]. The FMO-MP2/6-31G job of an influenza hemagglutinin (HA) monomer bound with a single Fab fragment antibody was completed within 1 hour with 4096 processors of the first Earth Simulator (ES), a vector-parallel supercomputer, where the total number of amino acid residues in the HA complex was 921. This achievement can be considered as

the first touchstone of massively parallel execution of correlated FMO calculations. Detailed IFIE analyses for the HA-Fab system were carried out to identify the potential mutations in HA [58] (see Sec. 3.2).

An analysis of local MO-wise inter-fragment interactions named Fragment Interaction based on Local MP2 (FILM) [59,60] was implemented for the ABINIT-MP on the basis of Pulay's technique [61]. FILM neither accelerates nor hinders the computation, but it helps a visual understanding of specific attractive interaction such as CH- π . A certain dimer MP2 energy can be decomposed into the intra-monomer and the inter-monomer components according to the localization situations of the respective local occupied MOs, and then the inter-monomer energies are treated as the orbital-wise stabilization contributions of FILM.

Ishikawa *et al.* [19] developed the PAICS program for FMO-MP2 calculations, where the counterpoise correction for the basis set superposition error (BSSE-CP) [62] was incorporated for the first time in the FMO context. Additionally, Ref. [19] presented a statistical discussion for protein-ligand interaction energies through MD-sampled structures [33]. Such a statistical sampling is essential to incorporate thermal fluctuations of protein structures at the incremented computational costs. Although the FMO scheme itself is effective to handle large molecular systems as mentioned above, the so-called integral approximation through proper factorization [63] should be useful to obtain further acceleration. The resolution-of-identity (RI) technique was introduced for FMO-MP2 in PAICS [64], where the dimer part of MP2 was accelerated by a factor of ten. The Cholesky decomposition (CD) was also utilized [65] for ABINIT-MP (see Sec. 2.5). Currently, the CD acceleration is available for FMO4-MP2 (as well as HF), and the calculation steps of fragment tetramers and trimers are more accelerated than that of dimers as expected. In GAMESS, the RI(MP2) scheme was implemented for FMO-MP2 [66] and FMO3-MP2 [67].

The electron pair-pair interactions can be considered beyond the MP2 treatments [37,46-48]. The MP2 correlation correction overestimates the attractive interaction energies in certain cases, and higher-order treatments might provide some remediation. The third-order or MP3 energy has three contributions identified by the numbers of holes (occupied orbitals) and particles (virtual orbitals) in a diagrammatic sense, but still consists of only doubles. The MP3 formula are given in a spin-orbital form as [37]

$$E^{\text{MP3}(4\text{p}-2\text{h})} = \frac{1}{8} \sum_{ijkl} \frac{\langle ij || ab \rangle \langle ab || cd \rangle \langle cd || ij \rangle}{(\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b)(\epsilon_i + \epsilon_j - \epsilon_c - \epsilon_d)} \quad (40)$$

$$E^{\text{MP3}(2\text{p}-4\text{h})} = \frac{1}{8} \sum_{ijklab} \frac{\langle ij || ab \rangle \langle ab || kl \rangle \langle kl || ij \rangle}{(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)(\varepsilon_k + \varepsilon_l - \varepsilon_a - \varepsilon_b)} \quad (41)$$

$$E^{\text{MP3}(3\text{p}-3\text{h})} = \sum_{ijkabc} \frac{\langle ij || ab \rangle \langle kb || cj \rangle \langle ac || ik \rangle}{(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)(\varepsilon_i + \varepsilon_k - \varepsilon_a - \varepsilon_c)} \quad (42)$$

These three equations are of N^6 order, in contrast to N^4 order of MP2, and thus the cost of MP3 calculations should be governed not by the integral transformation but by the tensor contractions. As in the case of MP2, the actual implementations should be done with a spatial orbital form, and the costliest “4p-2h” contribution may be rewritten as

$$E^{\text{MP3}(4\text{p}-2\text{h})} = \sum_{ij} \left(\sum_{ab} \frac{(ia, jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \left(\sum_{cd} (ac, bd) \frac{2(ic, jd) - (id, jc)}{\varepsilon_i + \varepsilon_j - \varepsilon_c - \varepsilon_d} \right) \right) \quad (43)$$

In ABINIT-MP, the MP3 module in a fully integral-direct fashion with shared-memory parallelism was developed for the fragment processing of monomers and dimers under the OpenMP technology [68]. The key points of efficient implementation were the use of DGEMM routines and also an AO-MO mixed strategy for “4p-2h” frequently called as the External Exchange Operator (EEO) technique [69]. According to the EEO, the first-order (MP1) amplitude in eqn (43) is back-transformed as

$$t_{ij}^{cd} = \frac{2(ic, jd) - (id, jc)}{\varepsilon_i + \varepsilon_j - \varepsilon_c - \varepsilon_d} \quad (44)$$

$$t_{ij}^{qs} = \sum_{cd} C_{qc} C_{sd} t_{ij}^{cd} \quad (45)$$

and the Fock-like contraction with the AO integrals

$$X_{ij}^{pr} = \sum_{qs} (pq, rs) t_{ij}^{qs} \quad (46)$$

is performed under the proper screening to effectively reduce the operation counts. Once the successive transformation

$$X_{ij}^{ab} = \sum_{pr} C_{pa} C_{rb} X_{ij}^{pr} \quad (47)$$

is performed, the remaining processing for eqn (43) is rather trivial. The new MP3 engine made the FMO-MP3 calculations possible even for a huge HA trimer with two

Fab fragments consisting of as many as 2351 amino acid residues in total [68]. With 1024 vector processors on the second Earth Simulator (ES2), the FMO-MP3/6-31G job for this HA trimer complex was completed in 5.8 hours, whereas the corresponding MP2 job time was 4.3 hours, indicating a small incremental factor of 1.3 and showing a practical applicability of FMO-MP3. An empirical scaling method with the MP3 correlation energy proposed by Pitoňák *et al.* (MP2.5) [70] could be an attractive option to remedy the overestimation of interaction energies by MP2, and this procedure was adopted in analyzing the HA trimer system in detail [71] (see Sec. 3.2). The BSSE-CP procedure at the FMO-MP3 level was also implemented [72].

The CC method [37,46,49] is based on the exponential expansion starting from the HF reference wave function Ψ_0 ,

$$|\Psi_{\text{CC}}\rangle = e^T |\Psi_0\rangle, \quad (48)$$

$$T = T_1 + T_2 + \dots, \quad (49)$$

and thus higher-order excitations are taken into account as the products of excitation operators to ensure size-consistency as exemplified below. The most widely used CCSD method is defined when the excitation operators are limited to singles (S) and doubles (D) in eqn (49). The projection equations for cluster amplitudes are given as

$$\left\langle \Psi_0 | H - E_0 | \left(1 + T_1 + T_2 + \frac{1}{2!} T_1^2 \right) \Psi_0 \right\rangle = E^{\text{CCSD}}, \quad (50)$$

$$\left\langle \Psi_i^a | H - E_0 | \left(1 + T_1 + T_2 + \frac{1}{2!} T_1^2 + T_1 T_2 + \frac{1}{3!} T_1^3 \right) \Psi_0 \right\rangle = 0, \quad (51)$$

$$\left\langle \Psi_{ij}^{ab} | H - E_0 | \left(1 + T_1 + T_2 + \frac{1}{2!} T_1^2 + \frac{1}{2!} T_2^2 + T_1 T_2 + \frac{1}{3!} T_1^3 + \frac{1}{2!} T_1^2 T_2 + \frac{1}{4!} T_1^4 \right) \Psi_0 \right\rangle = 0, \quad (52)$$

where Ψ_i^a and Ψ_{ij}^{ab} are the singles and doubles configurations, respectively, and E_0 is the reference HF energy [46,49]. The tensor contractions among cluster amplitudes, integrals and intermediate arrays for CCSD are of \mathcal{N}^6 but are much more complicated than those of MP3, and they are iterated until the convergence; typically 8 to 12 times more expensive computations are necessary to achieve six digit precision. To enhance the reliability, the perturbative triples correction is sometimes added at the cost of \mathcal{N}^7

after the CCSD iteration, and this recipe is termed as CCSD(T) [46,49].

Fedorov *et al.* [73] reported the FMO-CCSD(T) calculations with GAMESS, where water clusters and glycine polymers were used as test systems. A general CC module for protein calculations was developed and incorporated into ABINIT-MP under the OpenMP parallelism [74]. The fundamental working equations of this implementation owed to Scuseria's formulation [75] as performed in Ref. [76], while several modifications were made to carry out other correlated methods like the full fourth-order MP (MP4) (including the singles, doubles, quadruples and triples) [46-48] as well as the quadratic configuration interaction with singles and doubles (QCISD) [77]. The AO-MO mixed processing such as EEO [69,76] was again employed for efficient and memory-reducing computations as for the MP3 case [68]. In Ref. [74], a series of test calculations were performed on the ES2 supercomputer. The FMO-CCSD(T)/6-31G job for the HIV-1 protease-lopinavir complex, whose cost was 64 times larger than that of MP2, was finished in 9.8 hours with 512 processors, and the visualized IFIE results are illustrated in Fig. 2.

The intermediate array for the doubles part of (T) correction is constructed by the contraction of N^2 order

$$W_{ijk}^{abc} = P_{ijk}^{abc} \left(\sum_d (ia, bd) t_{kj}^{cd} - \sum_l (ia, jl) t_{lk}^{bc} \right), \quad (53)$$

where this equation is six-fold so that the sum of the following permutations should be taken on the right-hand side of eqn (53) [46,49]:

$$P_{ijk}^{abc} \binom{abc}{ijk} = \binom{abc}{ijk} + \binom{bac}{jik} + \binom{cba}{kji} + \binom{acb}{ikj} + \binom{cab}{kij} + \binom{bca}{jki}. \quad (54)$$

Namely, DGEMM is called 12 times in total in the most inner part of (T) routine in the CC module of ABINIT-MP [74], and thus they can be efficiently processed, in particular on the vector computers. With 1024 processors of ES2, a remarkably high performance rate of 38.6 % (relative to the theoretical peak speed) was observed for the FMO-MP4/6-31G job of a Trp₁₂₇-His model protein, being consistent with the fact that the (T) computation dominates the total cost in the MP4 case [74]. Recently, the Brueckner doubles (BD) [78], whose memory requirement is much smaller than that of CCSD, was made available in the ABINIT-MP program as well.

Finally, the FMO calculations of density functional theory (DFT) are addressed as effectively correlated approaches. Fedorov *et al.* [79] implemented the FMO-DFT ability in GAMESS. With the NWChem program [80], Sekino *et al.* [16-18] reported several

FMO-DFT calculations. There is a reference [81] concerning the DFT implementation in ABINIT-MP as well. The standard B3LYP hybrid functional [82] is available in these three programs. Some actual applications of FMO-DFT have been made by using GAMESS. In particular, Grimme's empirical dispersion correction [83] is available even for the geometry optimizations in GAMESS [84]. Incorporation of other advanced DFT schemes [85-88] into the FMO method is an important future issue. The DFT-based excited state calculations with GAMESS are addressed later in Sec. 2.3.

2.2.2. Gradient calculation

The first derivative of total energy with respect to the nuclear coordinates (energy gradient or minus force) is fundamental for geometry optimization and MD simulation. In the original two-body FMO scheme, the total energy is given as

$$E^{\text{Tot.}} = \sum_{I>J} E_{IJ} - (N_f - 2) \sum_I E_I + \sum_{A>B} V_{AB}^{\text{Nuc.}} \quad (55)$$

where N_f is the number of fragments and $IJ(AB)$ specify the fragments (nuclei). The corresponding force on coordinate q_A may thus be written as

$$-\frac{\partial E^{\text{Tot.}}}{\partial q_A} = -\sum_{I>J} \frac{\partial E_{IJ}}{\partial q_A} + (N_f - 2) \sum_I \frac{\partial E_I}{\partial q_A} - \sum_{B \neq A} \frac{\partial V_{AB}^{\text{Nuc.}}}{\partial q_A} . \quad (56)$$

Since the early stage of FMO developments [1,2], the importance of FMO-based geometry optimization for proteins was recognized, and the FMO-HF gradient was reported in 2001 [39]. This development led to the first FMO-MD work [89] in 2003. (See Sec. 2.4 and Ref. [90] for further details of FMO-MD implementations and applications.) In 2007, the first FMO geometry optimization of a protein was performed with GAMESS at the HF/6-31G* level [91], where the gradient of dimer-ES approximation [11] was derived. Nagata *et al.* [92] then examined the accuracy of ESP derivatives for FMO-HF. Komeiji *et al.* [93] reported the implementation of FMO3-HF gradient in ABINIT-MP and its application to MD, in which the importance of three-body correction (FMO3) [36] was illuminated for water clusters with and without protonation.

While the accuracy of FMO-MP2 gradients was shown in Ref. [50], any formulations and corresponding equations were not presented unfortunately. Mochizuki *et al.* [94] developed a parallelized integral-direct MP2 gradient module for efficient MD simulations at the FMO3-MP2 level, where the ESP-related equations as well as coupled perturbed HF (CPHF) equation for MP2-relaxed density were given. The

importance of MP2 correlation in MD simulations was illustrated in reproducing the peak position of O-O distribution function for water in Ref. [94].

Ishikawa *et al.* [95] extended the FMO-MP2(RI) implementation in PAICS [64] to enable efficient gradient calculations for the partial geometry optimization of proteins [96]. With ABINIT-MP, Tsukamoto *et al.* [97] showed that the FMO-MP2 gradient [94] is necessary to optimize the central structure entailing π/π or CH/ π interactions with Trp in the TrpCage mini-protein; the structure was deformed relative to the experimental structure at the FMO-HF level. A similar situation was observed for the pharmacophore structure of estrogen receptor (ER). Implementation of the CD-based MP2 [65] gradient in ABINIT-MP have been underway. In contrast to PAICS and ABINIT-MP, the current geometry optimizer with the frozen domain (FD) restriction in GAMESS has adopted the HF or DFT gradient calculations with dispersion correction [83] as addressed previously [84].

Nagata *et al.* achieved systematic improvements of numerical accuracy of FMO2 gradients and implemented them in GAMESS [98-104]. Ref. [99] reported the derivative term involving the BDA projection operator, which was omitted in the first paper on the FMO-HF gradient [39]. Another more critical omission in Ref. [39] was the response term arising from the fact that the ESP or fragment charge sets are self-consistently determined for the fragment monomers (or the SCC condition). In other words, the contribution from charge reorganizations against the nuclear displacement can be sizable for the fragment dimers. However, a full system should be formally treated to compensate the response contributions from dimers through a set of CPHF equations, which however seemed at first too costly to implement. Fortunately, a smart way to solve the CPHF equations was derived in Ref. [102], where the self-consistent Z-vector (SCZV) technique was devised for the FMO-HF gradient, by which the dimer CPHF equations were decoupled into monomer ones and effectively solved with the minimal increment of computational cost after the usual dimer calculations. As a result, the accuracy of gradient was greatly improved. This technique was further extended for the FMO-MP2 gradient [103], where the corresponding set of equations became lengthy relative to the HF case. The hybrid gradient of effective fragment potential (EFP) and FMO was developed as well [98,100,101,104]. Additionally, the influence of omitting the response term on the statistical quantities by FMO-HF/MD simulations was examined [101] and it was found that its artifact would not be quantitatively negligible. When FMO3 and FMO4 gradient schemes are employed for MD without the response term, however, the situations become better than that of the naïve FMO2. This observation might be explained by many-body charge delocalization or reorganization. Anyhow, the

FMO-HF gradient with SCZV [102] has now been implemented in ABINIT-MP as well. It is also noted that the FMO-HF second derivative (corresponding to force constants) based on Nagata's equations was recently implemented by Nakata *et al.* [105] in GAMESS.

2.2.3. Environmental effects and open-shell treatments

In some cases, heavy elements in which the relativistic effect [106] is essential are contained in biomolecular systems. For example, *cis*-platin is known as a representative anti-cancer drug, which should form a cross-link with DNA through Pt-N bonds. Although a number of relativistic MO schemes have been developed, the model core potential (MCP) [107,108] was employed as a cost-effective and reliable method for the FMO calculations with ABINIT-MP [109], where the hydrated Hg(II) ion and *cis*-platin-DNA complex were used as test cases. The MCP gradient was also available [110]. There were a couple of realistic MCP applications: The first is the MP2-correlated FMO-MD study for the difference in hydration structures of *cis*-platin and *trans*-platin [111] (see Sec. 3.4.2 below). The second is the base-base interaction energy analyses in a hydrated *cis*-platin-DNA complex at the FMO-MP2.5 level (unpublished work). The MCP gradient was incorporated also in GAMESS [104].

The hydration can affect the electronic structure of biomolecules to some extent. At the FMO-HF level, a systematic study was made on this issue for the ubiquitin protein explicitly hydrated by several layers of explicit water molecules [112]. It was observed that charge distributions even of internal regions in ubiquitin are influenced via the electronic polarization and charge transfer (CT) through interaction with water molecules. Further studies concerning the hydration effects on proteins with explicit water molecules were performed in the FMO framework [113,114]. The explicit solvent is realistic but expensive, and hence implicit solvent models have also been sought. The most widely used implicit solvent is the polarizable continuum model (PCM) [115], and this model has been made available for various FMO calculations with GAMESS [116-119]. The largest example by FMO-MP2(PCM) was an influenza HA trimer complex reported by Sawada *et al.* [120]. In ABINIT-MP, the Poisson-Boltzmann (PB) approach was adopted for an implicit model of hydration effect [121]. When the CT from surrounding water molecules is substantial for biomolecular systems such as DNA, however, the inherent limitation of implicit models should be kept in mind.

The GAMESS program has an advantage to perform open-shell FMO calculations. Based on the multi-layer FMO (MFMO) framework [122], the multi-configuration SCF

(MCSCF) ability including restricted open-shell HF (ROHF) was supported [123]. The ROHF-based MP2, CC [124], gradient [125] and PIE [126] schemes are available in GAMESS. The unrestricted HF (UHF) type calculations are available both in GAMESS [127] and in ABINIT-MP.

Finally in this subsection, some other FMO-related approaches are addressed as follows. A quantum Monte Carlo (QMC) calculation coupled with the FMO scheme was attempted [128] for the accurate description of electron correlation. The nucleus-electron orbital (NEO) set was evaluated [129] under the FMO framework. Recently, an interesting multi-state extension of FMO was made in the MD context [130]. Many-body FMO expansion was re-formulated [131] in the Green's function (GF) framework. Future developments of FMO/GF can be expected, especially for the improvement of ESP treatments. Besides the FILM tool for dispersion interaction [59,60], the orbital-wise CTs in systems such as hydrogen-bonding network can be grasped by the configuration analysis for fragment interaction (CAFI) in ABINIT-MP [132]. The concurrent electron relaxation functional (CERF) [133] was utilized for the CAFI calculations, where a partial renormalization [134] was employed to ensure the size-consistency. It has been known that the renormalization techniques are useful to take higher-order effects into account without significantly changing the computational cost [37,46]. With CAFI, the pair list of electron donor and acceptor orbitals can be obtained with the respective energy lowerings due to the CTs associated with those interactions such as hydrogen bonding.

2.2.4. Other correlated fragmentation methods

Although this Perspective Article is focused on the FMO-related methods and demonstrative applications [2-7], some other correlation and excited-state methods with conceptual similarity with FMO are worth addressing in the following.

The elongation method has been developed by Aoki *et al.* [135] for general polymer systems, where MP2 correlation [136] and time-dependent-Hartree-Fock (TD-HF) approximation for excitation energy [137] and static second-harmonic-generation (SHG) [138] have been studied. All these elongation calculations were available in the GAMESS program. Kobayashi and Nakai have revisited the divide-and-conquer (D&C) method [139] in a modern MO fashion [140]. Their D&C methods at the levels of MP2 [141-143], CC [144] and so on were implemented in GAMESS. The correlated calculations with cluster-in-molecule (CIM) approach by Li *et al.* [145-148] were carried out on GAMESS as well. For convenience of readers, it is noticed that a special issue

entitled “Fragment and localized orbital methods in electronic structure theory” (Phys. Chem. Chem. Phys. Vol. 14, 2012) is available, which reviews FMO [7], elongation [135], D&C [140] and CIM [145] methods.

In addition, the divide-expand-consolidate (DEC) method was developed in the DALTON program [149,150], and also the molecule-in-molecule (MIM) method was implemented in MO modules of the GAUSSIAN suite [151,152]. Finally, it should be noted that IS proposed by Stoll in 1992 [153] has applicability to three-dimensional systems such as solid silicon, where a three-body expansion of units was first attempted for the correlation energy. Recently, efficient re-implementations of IS have been made for the CC calculations [42,154-156] as well as excited-state calculations [157].

2.3. Excited states: Excitation energy and property evaluation

There are a number of photoactive proteins (FPs) such as green fluorescent protein (GFP), whose low-lying excited states of π - π^* transitions are of primary interest. The simplest approach to describe such states is the configuration interaction singles (CIS) [158] scheme, and the corresponding Hamiltonian matrix (of spin singlet) to be diagonalized is given with the spatial orbital integrals as

$$A_{ia,jb} = 2(ia, jb) - (ij, ab) + \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) \quad (57)$$

The diagonalization can be iteratively done at N^4 cost through the direct AO-integral processing. The parallelized CIS ability in the ABINIT-MP program was implemented [159] based on the MFMO framework [122] in which the photoactive pigment or chromophore (including the surrounding amino acid residues) should be the target region of excitation and the remaining part was kept frozen at the HF stage. Namely, the environmental effect of protein was incorporated. Although the excitation energies by CIS could suffer from overestimations as large as 1-2 eV, the doubles correction denoted as CIS(D) [160] can reduce such energetic errors effectively. In CIS(D), both the relaxation energy (negative value) for singly excited states and the differential correlation energy (positive) from the ground state are evaluated at the MP2 level, where the gross correction is mostly negative. The actual processing of CIS(D) is more complicated and costly than that of MP2, but it is still of N^5 cost. Mochizuki *et al.* [161] reported the development of a fully integral-direct CIS(D) module with MPI parallelism, and then substantial improvements were demonstrated in comparison with the CIS results of Ref. [159]. This MFMO-CIS(D) method was applied to the DsRed, a famous

Red FP [162], and also to the explicitly hydrated formaldehyde whose geometrical configurations were generated by the FMO-MD simulation [163]. Several modified CIS(D) variants through the renormalization techniques were proposed and implemented in ABINIT-MP [164-166], which were successfully used for mFruits [167], Yellow and Blue FPs [168] and firefly luciferase [169]. MFMO-CIS(D) calculations for the Fenna-Matthews-Olson protein were also carried out in order to estimate the excitation energies as well as the electronic couplings among porphyrins embedded in the protein (unpublished work). In ABINIT-MP, the CIS energy gradient [158] is available for excited-state FMO-MD as well [90].

Chiba *et al.* [170-174] utilized the time-dependent (TD)-DFT methods in the FMO scheme on GAMESS, where the long-range correction (LC) was employed in order to avoid the energy collapse in CT-type excited states. It is notable that the PCM implicit hydration [173] and the geometry optimization [174] were supported for TD-DFT excited states. CIS has also been available in GAMESS, and Ikegami *et al.* [175] performed a series of CIS calculations for a large protein subunit in the photosynthetic reaction center, where several porphyrin units are embedded. Though the calculated energies were not in good agreement with the experimental values due to the lack of correlation corrections, the directionality of electron flow was analyzed in detail [175].

Molecular properties can be analytically evaluated according to the response theory, in which the equation of explicit sum-over-states (SOS) is converted to an appropriate set of linear equations [176]. The linear response (LR) equation for frequency (ω)-dependent polarizability (the second derivative with respect to electric field) implemented in ABINIT-MP [177] has a matrix form as

$$\begin{pmatrix} \mathbf{A} - \omega \mathbf{1} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} + \omega \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{N}_\mu^+ \\ \mathbf{N}_\mu^- \end{pmatrix} = - \begin{pmatrix} \mathbf{V}_\mu \\ \mathbf{V}_\mu \end{pmatrix}, \quad (58)$$

where \mathbf{A} is the CIS matrix of eqn (57) and \mathbf{B} describes a sort of correlation effect through the de-excitation from doubles

$$B_{ia,jb} = 2(ia, jb) - (ib, ja) . \quad (59)$$

\mathbf{N} and \mathbf{V} in eqn (58) are the response and polarization vectors, respectively, with the Cartesian components. The polarizability tensor is given by the dot products of these vectors. The actual computation can be performed with the AO integral list at N^4 cost under parallelization as in the case of CIS [158]. The two-body sum of fragment monomers and dimers [1]

$$\alpha_{\mu\nu} = \sum_{I>J} \alpha_{\mu\nu,IJ} - (N_f - 2) \sum_I \alpha_{\mu\nu,I} \quad (60)$$

is usable for the polarizability as well. An extension to hyperpolarizability involving the SHG is straightforward due to the response theory [176].

The LR equation of (58) has a close connection to the TD-HF or random phase approximation (RPA) as well as CPHF. In the evaluation of nuclear magnetic resonance (NMR) parameters [178], a magnetic CPHF equation appears, where special care should be paid for the gauge invariance. The FMO-NMR method at the HF level was implemented by Gao *et al.* [179-181], where a set of FMO-derived charges was obtained by GAMESS and then the gauge-invariant CPHF equation was solved by the GAUSSIAN program. Sekino *et al.* [182] also reported the NMR calculation with DFT on NWChem. Further developments and implementations are highly expected for the property evaluations with the FMO scheme.

2.4. Molecular dynamics: FMO-MD

FMO-based MD simulation, "FMO-MD", is an *ab initio* or first principles MD method [89] that can simulate molecular phenomena involving electronic structure changes such as polarization, electron transfer, and chemical reaction at finite temperature. FMO-MD has been successfully applied to analyses of ion-solvent interaction and chemical reactions of organic molecules. It is hoped to extend the field of FMO-MD to solvated proteins and nucleic acids in the near future. In this subsection, methodological aspects of FMO-MD are reviewed. We have previously published several review articles of the method [5,90,183], but we include here more recent advances.

2.4.1. Principle and software

FMO-MD is a kind of Born-Oppenheimer dynamics, in which the motion of the electrons and that of the nuclei are separated (Fig. 3). In FMO-MD, the electronic state is solved by FMO using the instantaneous 3D coordinates of the nuclei (\mathbf{r}) to obtain the energy (E) and force (\mathbf{F} , minus the energy gradient) acting on each nucleus, which are then used to update \mathbf{r} classical-mechanically by MD.

We implemented FMO-MD by combining the PEACH MD program [184] and the ABINIT-MP (F)MO program [10]. The resultant FMO-MD simulator, PEACH/ABINIT-MP system, has been revised several times [185-187]. In the first

version built in 2002 [185], the ABINIT-MP program was converted into a subroutine library for FMO force calculation and linked with PEACH. This version 2002 was disadvantageous for the software developers because the interface library must be reconstructed at every version-up of ABINIT-MP, and hence in the next FMO-MD implementation in 2006 we took another way of merger. In version 2006 [186] both programs were treated independently; at every time step of MD, PEACH executes ABINIT-MP via the system call command, and the data (coordinate, fragments, force, energy, *etc.*) are transmitted between the programs via data files. Version 2006 facilitated independent development of both programs; nonetheless, the complicated software structure hindered transportation and execution of FMO-MD on massively parallel computers. In the latest version of 2013 [187], therefore, all the FMO-MD-related functions have been imported from PEACH to ABINIT-MP, so that ABINIT-MP can stand alone as an FMO-MD simulator. Most of the FMO-MD studies presented in this article have been performed by a version of the PEACH/ABINIT-MP system.

Like any Born-Oppenheimer dynamics method, implementation of FMO-MD is simple and straightforward. Therefore, besides the PEACH/ABINIT-MP system, several FMO-MD software systems have been reported in the literature, using ABINIT-MP [41,188-192], GAMESS [92,101,102,110,193], or PAICS [194].

Methods in FMO-MD are roughly classified into those of FMO, those of MD, and those of the interface. Accuracy and performance of FMO-MD depend mostly on the FMO gradient calculation. Many of standard algorithms of classical MD can be used for FMO-MD. Particular algorithms necessary to interface FMO and MD also exist. Below we present the three methodological aspects of FMO-MD .

2.4.2. FMO methods important to FMO-MD

The capability and performance of FMO-MD mostly depend on the FMO portion. In particular, the accuracy of the FMO energy gradient determines the precision of the FMO-MD calculation. The methods in FMO are reviewed in other sections of this article and here only those important to FMO-MD are briefly described.

FMO n - Accuracy of FMO, both energy and force, can be improved by adding 3-body, 4-body, ..., and n -body terms (FMO n) at the expense of the computation cost of $\mathcal{O}(1)$. The improvement by FMO3 is especially apparent in FMO-MD, as illustrated by a simulation of H⁺-transfer in water, though FMO3 was five times more costly than FMO2 [93].

MP2 - MP2 is an easy way to incorporate dispersion forces and was applied to FMO-MD as soon as the energy gradient formula for MP2/FMO was first derived [94]. Simultaneously, a new option called "FMO(3)" was implemented, in which FMO3 is applied to HF but FMO2 is applied to MP2 to reduce computation time, based on the relatively short-ranged nature of the electron correlation compared to the electrostatic interactions. The MP2/FMO-MD was soon applied to a droplet of water molecules [94]. The water was simulated with the 6-31G* basis set with and without MP2. The first peak position of water oxygen in the radial distribution function by MP2/FMO-MD was closer to the experimental value than that of HF/FMO-MD was, indicating the importance of the dispersion term.

MCP - MCP is an inexpensive way to incorporate relativistic effect needed for (F)MO calculation of heavy metals [107,108,195-197] (Sec. 2.2.3). The MCP method has been combined with FMO and implemented in ABINIT-MP [109] and used in the comparative MCP/FMO(3)-MD simulations of hydrated *cis*-platin and *trans*-platin (see Sec. 3.4.2).

UHF - UHF, a way to handle open-shell molecular systems, has been implemented in ABINIT-MP. FMO-MD simulation of hydrated Cu(II) has been underway at the FMO3-UHF level, and the Jahn-Teller distortion of hexa-hydration has been reasonably reproduced.

Periodic boundary condition (PBC) – PBC was originally introduced to FMO-MD in the TINKER/ABINIT-MP system [41], in which layers of image solvent boxes were placed around the real box. Later, Brorsen *et al.* [193] implemented PBC via minimum image convention.

Fully analytic energy gradient - The FMO energy gradient originally developed by Kitaura *et al.* [39] was incomplete because of neglect of response term to avoid solution of CPHF equation. This problem was resolved and fully analytic energy gradient is now available for HF and MP2 [102,103]. The analytic gradient has been implemented in GAMESS and has been proved to improve the accuracy of FMO-MD simulation [193]. (See Sec.2.2.2 for details.)

2.4.3. MD methods important to FMO-MD

Many of standard methods of MD can be combined with FMO-MD, including thermostats, bond constraint, enhanced sampling, and so on. Below we present only two algorithms closely related to FMO-MD.

Blue moon ensemble - Most chemical reactions are rare events that seldom occur

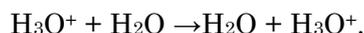
during standard FMO-MD or other *ab initio* MD simulations. A way to simulate a rare event is the blue moon ensemble method [198], which calculates the free energy profile along a reaction coordinate (RC) while constraining RC to a specified value. The method was interfaced with FMO-MD [199] and was successfully applied to drawing a free energy profile of the Menschutkin reaction [186] and the amination and hydration of H₂CO [200,201].

Path Integral Molecular Dynamics (PIMD) - The nuclei were handled by the classical mechanics in most of the FMO-MD simulations performed to date (Fig. 3), but PIMD [202] has been introduced into FMO-MD to incorporate the nuclear quantum effect [191]. FMO-PIMD consumes tens of times more computational resource than the classical FMO-MD does, but is necessary for a better description of, for example, proton transfer reaction.

2.4.4. Dynamic fragmentation algorithms

Dynamic Fragmentation (DF) is the important interface algorithm between FMO and MD. DF redefines fragment data upon structure change of molecules during FMO-MD [93,186,187].

The necessity for DF can be understood by the following example. If one simulates H₃O⁺ hydrated by H₂O solvent, the following H⁺ transfer sometimes occurs:



Even this simplest reaction is difficult to simulate by FMO-MD unless H₃O⁺ and H₂O have been united into a fragment from the beginning. It is usually impossible, however, to know *a priori* where the reaction will take place, and hence one has to define H₃O⁺ and H₂O as separate fragments. Upon a spontaneous H⁺-transfer, the FMO calculation ends in an abrupt failure because the fragment definition does not match the chemical structure of the molecules. To avoid this failure, one needs an algorithm for redefinition of fragments dependent on the molecular configuration, namely the DF algorithm.

DF first appeared as an *ad hoc* algorithm whose program needed to be newly developed for each new simulated molecules [163,186,203], but it was soon modified to handle general molecular systems [93]. This generalized DF algorithm was intended for simulation of only small molecules each of which constitutes a fragment [187]. This DF algorithm has been successfully used for FMO-MD simulations of H⁺-transfer in water [93], amination and hydration of formaldehyde [200,201], dynamics of *cis/trans*-platins in water [111], and hydration of metal ions [204]. DF has been also implemented in the GAMESS software [193].

Though fairly effective, the generalized DF still had a limitation of handling only small molecules. Many macromolecules including proteins and polynucleotides should be divided into several fragments, however. The limitation of the method has been overcome by a further modification of the algorithm named “Dynamic fragmentation with static fragments (DF/SF algorithm, [187]).” The gist of DF/SF is to classify the atoms of the simulated system into dynamic (variable) and static (invariable) ones prior to an FMO-MD simulation, and at each time step of the simulation a dynamic atom is assigned to a suitable fragment by the DF algorithm while a static one always to a fragment initially specified by the user.

Below we explain the DF/SF algorithm using a molecular system made up of $\text{C}_2\text{H}_5\text{CO}_2\text{H}$, CH_3OH , and several H_2O molecules as an example (Fig. 4). $\text{C}_2\text{H}_5\text{CO}_2\text{H}$ is divided into two static fragments; one is C_2H_5 with a formal charge of +1 and the other is CO_2H with a formal charge of -1. CH_3OH is set dynamic. The DF/SF algorithm has several modes (Table 1, [93,187]), but here we outline the DF1/SF2 option, the most robust one.

To run an FMO-MD simulation with DF/SF, the user should input some data and parameters as follows. First, static fragments should be specified among the default fragmentation data (dotted circles in Fig. 4 A). Heavy atoms belonging to the static fragments are classified as static, *i.e.*, always assigned to the same fragments during MD, but hydrogen atoms are regarded as dynamic. Note that hydrogen atoms are regarded as static in SF1 mode in Table 1, but this mode is unstable and not recommended [187]. Second, a table of atomic compositions and formal charges of possible fragment species should be provided. Table 2 is an example intended for simulation of CH_3OH in water. No species related to static fragments, $\text{C}_2\text{H}_5\text{CO}_2\text{H}$ in this case, are necessary. Third, the threshold parameter ρ^{im} , explained below, should be input. Using these data, the DF/SF algorithm is executed at every time step of the FMO-MD simulation.

In the DF/SF algorithm, the dynamic atoms are assigned to appropriate fragments by the following principle: a group of atoms connected either by covalent bonds or coordinate bonds are united into a fragment. Hence, a fragment consists of a molecule or a coordination compound. A pair of atoms is regarded as connected if their distance is considerably shorter than the addition of their van der Waals radii (Fig. 4 B). Specifically, a reduced distance between atoms i and j , ρ_{ij} , is calculated by the following equation,

$$\rho_{ij} = r_{ij} / (R_i + R_j) \quad (61)$$

where r_{ij} is their distance and R_i and R_j are their van der Waals radii. Atoms i and j are regarded as bonded if $\rho_{ij} < \rho^{\text{lim}}$, where ρ^{lim} (denoted as ρ_1^{lim} in [93]) is the threshold parameter pre-set to 0.8-0.9.

After all the heavy atoms, both dynamic and static, have been classified into appropriate fragments, each hydrogen atom is attached to its closest heavy atom, again using the reduced distance given by eqn (61) (Fig. 4 C). In the DF1/SF2 mode, all the hydrogen atoms are subject to reassignment, including those initially belonging to static fragments. Note that in the example (Fig. 4), H^+ has been transferred from $-\text{CO}_2\text{H}$ to a nearby H_2O so that the final fragments are $-\text{CO}_2^-$ and H_3O^+ (Fig. 4 D).

Finally, a formal charge is assigned to each fragment thus constructed either by looking up the fragment species list for the dynamic fragments (Table 2) or by using the initial formal charge with possible readjustment dependent on increase/decrease of hydrogen atoms for the static fragments (Fig. 4 D). This is the end of the DF1/SF2 mode of the algorithm. The fragmentation data thus constructed are used in the FMO calculation to obtain new E , \mathbf{F} , q , and other data related to the electronic structure of the instantaneous configuration (Fig. 3).

The DF and DF/SF algorithms so far presented have a problem of producing energy gap upon fragment rearrangement. That is, the energy sometimes shows discontinuity if a set of fragments different from the previous one is produced at a snapshot of FMO-MD. At least three ways have been so far published to alleviate this energy gap problem. The first way is to use FMO3 instead of FMO2, which way worked effectively in simulation of the H^+ -transfer reaction in water, though with five times more computational cost [93]. The second is the use of the fully analytic energy gradient already discussed in Sec. 2.4.2, which presumably may reduce artifactual fragment rearrangement and thereby decrease the production of the energy gap itself. The third is the use of a smoothing function to avoid the discontinuity in the potential energy surface [130]. Basically, these three ways may be combined.

As so far presented DF is a rather complicated algorithm. Nonetheless, DF or a similar interface algorithm is absolutely necessary to perform MD with any fragment-based quantum method. Further, DF/SF has finally enabled us to simulate proteins and/or polynucleotides in explicit solvation by FMO-MD.

2.5. Some remarks on acceleration: Cholesky decomposition with adaptive metric (CDAM)

The huge amounts of MO calculations for multiple combinations of fragments must

be processed to compute three-body and four-body correction energies in the FMO3 and FMO4 methods, and the large number of orbitals on each combination also increase the computational cost significantly. In order to reduce the cost for these MO calculations, one needs to introduce a factorization approximation for processing the two-electron repulsion integrals (ERIs), a dominant bottleneck in any MO calculations both with and without the electron correlation.

Several factorizations of ERIs such as density fitting, RI and CD approaches are recently used as principal techniques for accelerating MO calculations [63] and applied also to the FMO method [64,65,67,94]. All these approaches approximately factorize four-center ERIs to three-center ones to reduce the order of the computational cost from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^3)$.

Beebe and Linderberg [205] proposed the essential equation for the CD formulation as follows:

$$(pq,rs) \approx \sum_K^M L_{K,pq} L_{K,rs} \quad (62)$$

where the four-center AO-integral matrix is approximately represented by the products of the CD vectors $L_{K,pq}$ with the Cholesky bases K and the effective rank M . The decomposition is achieved to satisfy the relationship

$$\left| (pq,rs) - \sum_K^M L_{K,pq} L_{K,rs} \right| \leq \delta \quad (63)$$

with removing the linear dependency among the Cholesky bases, which are obtained from AO pairs on-the-fly, and the accuracy is well controlled by the decomposition threshold δ .

On the other hand, the density fitting and RI methods expand a four-centered ERI by three-centered ERIs with using externally-introduced auxiliary bases I, J ,

$$(pq,rs) \approx \sum_{IJ} (pq|I) G_{IJ}^{-1} (J|rs). \quad (64)$$

When the matrix $G_{IJ} = (I|J)$ is Cholesky-decomposed as $\mathbf{G}^{-1} = \mathbf{Z}\mathbf{Z}^T$, the relation

$$(pq,rs) \approx \sum_{IJ} \sum_K^M (pq|I) Z_{IK} Z_{KJ} (J|rs) = \sum_K^M L_{K,pq} L_{K,rs} \quad (65)$$

is obtained, and the density fitting (or RI) and CD methods are found to be intimately connected with each other [63].

If the \mathbf{Z} vector is stored on memory, Cholesky vector \mathbf{L} is also constructed on memory with pseudo-three-center integral calculations directly. As a preprocess for achieving direct CD, diagonal elements of \mathbf{G} matrix in the Coulomb metric are first screened with newly introduced threshold τ as

$$(I|I) = (\overline{pq} | \overline{pq}) \geq \tau \quad (66)$$

and size-reduced \mathbf{G}' matrix is obtained, which consists of selected candidate Cholesky bases $I' = \overline{pq}$. Inverse factorization for \mathbf{G}' provides \mathbf{Z} and substantial amounts of \mathbf{L} . \mathbf{L} is stored on memory by the distribution for only thousands of orbitals for fragment monomer, dimer, and so on in the FMO framework. It is also important to be free from disk I/O. Eqns (62)-(66) above constitute the Cholesky decomposition with adaptive metric (CDAM) [65].

Once the Cholesky vectors are obtained, the Coulomb and exchange integrals are approximately written by

$$J_{pq} = \sum_{rs} (pq, rs) P_{rs} \approx \sum_{rs, I} L_{I, pq} L_{I, rs} P_{rs} \quad (67)$$

$$K_{pq} = \frac{1}{2} \sum_{rs} (pq, rs) P_{rs} \approx \sum_{i, I} X_{I, pi} X_{I, qi} \quad (68)$$

respectively, using $L_{I, pq}$ and $X_{I, pi} = \sum_r L_{I, pr} C_{ri}$; AO-MO transformation needed in the MP2 calculation is also represented by

$$(ia, jb) = \sum_{pqrs} C_{pi} C_{qa} C_{rj} C_{sb} (pq, rs) \approx \sum_I B_{I, ia} B_{I, jb} \quad (69)$$

using the relation,

$$B_{I, ia} = \sum_q X_{I, qi} C_{qa} \approx \sum_{pq} L_{I, pq} C_{pi} C_{qa}. \quad (70)$$

They virtually have the forms of simple matrix product. Thus, once the AO-based \mathbf{L} array is available, most of the operations can be processed with the DGEMM routine with good acceleration, and the parallelization due to the CD basis I is straightforward. As a whole, the CDAM is a key technique for FMO4 (and FMO3) calculations with ABINIT-MP.

2.6. Massively parallel processing

Using the GDDI environment of GAMESS [15], Ikegami *et al.* [206] achieved a

massively parallel (using 600 scalar CPUs of a cluster computer) FMO-HF benchmark on the photosynthetic reaction center protein and reported it at the SuperComputing 2005 conference (SC05). This report demonstrated the inherent potential of FMO calculation for efficient parallel executions [2,5]. In addition, a heuristic static load-balancing algorithm [207] was applied to the FMO method. Recently, the FMO-MP2 gradient calculations of water clusters were carried out on the BlueGene/P computer, where up to 131072 processors were used [208]. In contrast, Katouda *et al.* [209] performed the efficient FMO-MP2(RI)/6-31G* calculations for large influenza protein in only 20 minutes on the K-computer by using 12288 nodes and 86016 processor cores, and further 196608 processors were successfully used. Along a different direction from the massive parallelism, Talamudupula *et al.* [210] attempted a heterogeneous computing for the FMO-HF and FMO-MP2 calculations with GAMESS. The OpenFMO program has been developed [20,21] for the massively parallel FMO-HF calculations, where its design policy was oriented to the efficient usage of a few hundred thousands of processors.

In earlier studies, the vector-parallel supercomputers (ES and ES2) were used as the platform for the correlated FMO calculations with ABINIT-MP [57,68,74]; for example, up to 4096 processors of ES were employed in Ref. [57]. The usage of the K-computer started in 2012, and preliminary results have been obtained for the FMO4-MP2(CD)/6-31G calculations (where the HF-SCF steps were carried out with CD) [65]. Fig. 5 shows a couple of illustrative timings with 600 and 6000 nodes (or 4800 and 48000 processor cores) for the HIV-1 protease-lopinavir complex with the fragmentation of main and side chains of amino acid residues (a total of 363 fragments). As can be seen from both bars in Fig. 5, the costs of MP2 are small and those of HF-SCF virtually dominate the total job time. In particular, the monomer SCC step is costly, for at this stage many processors were kept idle since each monomer was computed in a single node (8 cores). The times for the tetramer and trimer HF-SCF calculations are almost negligible in the case of 6000 nodes due to the rich amounts of processors. Apparently, works for more efficient calculations on the K-computer are required especially for the monomer SCC.

A comprehensive review on the high-performance MO calculations with parallelism was written by de Jong *et al.* [211], but fragmentation-based methods [9] were not addressed there. The intra-fragment calculations (*e.g.*, for monomers to tetramers in the FMO4 case [27]) should be more accelerated by combining various low-cost MO methods and algorithms devised for non-fragmentation approaches. Such future efforts can be useful to perform more chemically relevant FMO calculations on peta-scale computers,

which can cover, *e.g.*, the statistical averaging over structural fluctuations of hydrated proteins [19] and the heterogeneous situations of solid surface-protein boundary problems.

2.7. Data analysis

In addition to its high precision and performance, an important feature of FMO method is the capability of evaluating the IFIE or PIE between each fragment pair. Systematically calculated IFIE values can quantitatively describe the interactions between amino acid residues and small compounds on the energetic bases. In the cases of proteins and nucleic acids, it is thus possible to perform fragment-based analyses of those interactions such as residue-residue, residue-substrate, Watson-Crick base pair, stacking pair, and residue-base interactions; the intermolecular binding energies can also be evaluated in terms of the combinations of IFIEs. In addition, the FMO method enables the theoretical analyses on charge distribution and transfer since it can calculate the atomic charges varying according to the environmental and structural changes, in contrast to the classical-mechanical methods in which the atomic charges are fixed. Concerning the interaction analysis at the orbital level, the CAFI scheme [132] can describe the CT interactions between MOs for the hydrogen bond, in which the configuration analysis technique was employed for the FMO method. Further, the FILM method [59] can analyze the π/π and CH/ π interactions at the orbital level, thus leading to more detailed information on biomolecular interactions through combinations of these methodologies. In this subsection, these and related approaches are addressed.

2.7.1. IFIE map

The IFIEs are obtained for all the combinations of fragment pairs and can be depicted comprehensively on a two-dimensional IFIE map [24]. Fig. 6 shows an example for cyclic-AMP receptor (CRP) protein-DNA complex [212], in which the attractive (red) and repulsive (blue) interactions between the numbered (1-200 for CRP protein and 201-244 for DNA) fragments are illustrated with varied deepness of hue for the magnitude on the upper-left and lower-right sides, respectively. One may expect that the molecular interactions such as the hydrogen bonding are colored in red as in the distance matrix, and it is observed in fact that the α helices and the β strands in protein are represented as the patterns parallel and perpendicular to the diagonal line, respectively [24]. In the case of the base-base interactions in DNA duplex, the stacking

and the Watson-Crick hydrogen bonding interactions are represented as the patterns parallel and perpendicular to the diagonal line, thus reflecting the secondary structures in proteins and the double-helix structures in DNAs, respectively.

2.7.2. Visualized cluster analysis of protein-ligand interaction (VISCANA)

In silico analysis of protein-ligand docking is essential for structure-based, rational drug design. For the virtual ligand screening (VLS), a visualized cluster analysis of protein-ligand interaction (VISCANA) method [23] was developed to carry out a cluster analysis on the basis of IFIEs obtained with the FMO calculations. VISCANA makes it possible to classify structurally similar ligand molecules through the interaction pattern of a ligand and amino acid residues of the receptor protein as well as the interaction energy of the ligands and the protein. This method is thus different from the conventional clustering method based on ligand structures. VISCANA can also estimate the correct docking conformation by analyzing the patterns of the receptor-ligand interactions of some conformations through the docking calculation. VISCANA thus provides a powerful tool for assessing the VLS of xenobiotic chemicals prior to biochemical experiments.

The first application of VISCANA [23] was performed to human estrogen receptor α based on the FMO-HF method (Fig. 7), and it provided reasonable binding energies and binding patterns of ligand-protein interactions. Incorporation of electron correlation effects by MP2 or higher-order methods can provide us with more reliable descriptions of the van der Waals interactions and hydrogen bonds that are important for receptor-ligand binding. It is also desirable to perform the geometry optimization of entire complexes or pharmacophores, particularly for induced-fit complexes. The visualization software for molecule and molecular-interaction analyses based on FMO calculations with ABINIT-MP is available on BioStation Viewer [213], which includes the VISCANA function with a supporting CSV format of the receptor-ligand interaction energy table.

2.7.3. Statistically corrected IFIE (SCIFIE)

The IFIE provides a useful measure for the effective interactions between the fragments such as those between a ligand molecule and surrounding amino acid residues in the FMO calculations. The calculated IFIEs, however, often show unphysically large values primarily due to electrostatic interactions, because the FMO

method only considers the electronic polarization for a specific molecular configuration and does not take into account the screening effects in terms of inter-fragment correlations. In most molecular systems, nonetheless, positively and negatively charged fragments are located alternately to retain the charge neutrality. To partially correct for these collective effects, the statistically corrected IFIE (SCIFIE) has then been introduced [26] on the basis of many-body techniques used in classical mechanics.

Let us consider a situation that the (symmetric) IFIEs u_{ij} ($1 \leq i, j \leq N$) for a system with N fragments have been given through an FMO calculation. Each fragment i may be regarded as a classical particle embedded in the network of mutual interactions by many fragments. The effective interaction or the potential of mean force w_{ij} is then related to the pair correlation function h_{ij} between the fragments via [214]

$$h_{ij} = e^{-\beta w_{ij}} - 1. \quad (71)$$

The parameter β in eqn (71) is usually related to the absolute temperature T and the Boltzmann constant k_B via $\beta = (k_B T)^{-1}$, but, in this case, may be regarded as an optimization parameter to control the degree of screening. In the actual implementation, the optimal value of β is determined so as to maximize the degree of total screening.

The pair correlation function h_{ij} is related to the direct correlation function c_{ij} in terms of the Ornstein-Zernike relation [214]:

$$h_{ij} = c_{ij} + \sum_{k \neq i, j} c_{ik} h_{kj} \quad (72)$$

The direct correlation function is conceptually introduced as a difference between the total and indirect parts of inter-particle correlation, and may be expressed in the Percus-Yevick (PY) approximation [214] in classical many-body theories as

$$c_{ij} = e^{-\beta w_{ij}} - e^{-\beta(w_{ij} - u_{ij})} \quad (73)$$

providing a closure equation to determine w_{ij} for a given set of u_{ij} . There is another well-known approximation often employed in classical many-body theories, that is, the hypernetted-chain (HNC) approximation [214]. In this approximation, the direct correlation function is expressed as

$$c_{ij} = e^{-\beta w_{ij}} - 1 + \beta(w_{ij} - u_{ij}) \quad (74)$$

instead of eqn (73) for the PY approximation. In Coulombic systems, it has been remarked [214] that the HNC approximation would provide a more dependable description of interparticle correlations than the PY approximation.

As an example, a calculated result [26] for the SCIFIEs between Glu23 and other fragments in crambin protein with 46 residues is illustrated in Fig. 8, in which the values obtained in the PY scheme are compared with the bare IFIEs. The screening effect due to other residues is thus observed, reflecting the many-body effect between fragments.

2.7.4. MO-based analyses

As addressed in Sec. 2.2.3, the CAFI method [132] provides a convenient tool to analyze the hydrogen bond network in biomolecular systems. In this method the MO sets of fragments are merged and then orthonormalized by the use of a weighted Löwdin orthonormalization technique. The energy calculation is performed with the CERF [133]. The relaxation energy is thus obtained in an orbital-wise fashion and the CAFI can analyze the inter-fragment CT and the intra-fragment polarization interactions at the MO level. Fig. 9 illustrates an example for the hydrogen bond between an estrogen receptor and its ligand molecule, estradiol (EST). It is observed [30] that there takes place a CT from the oxygen lone-pair orbital of carboxyl group of Glu353 to the σ^*_{OH} orbital of hydroxyl group of ligand. The CAFI can thus describe the direction of CT as well as the magnitude of interaction (energy).

In contrast to CAFI, the FILM method [59] can primarily evaluate the dispersion-type interactions such as the CH/ π and π/π interactions at the orbital level. This method enables us to decompose the interaction energy associated with dispersion interactions into the contributions of localized occupied orbitals. Fig. 10 shows the T-type π/π interaction between the phenyl ring of ligand and the neighboring Phe in an estrogen receptor-ligand complex [215]. In this case, Phe plays a role as a CH donor instead of a π donor, thus indicating the CH/ π interaction with the phenyl ring. The FILM scheme was also employed [33] to describe the dispersion interactions between RXR protein and 9cRA ligand (see Sec. 3.1.2).

One important issue in the FMO analysis is how to construct the MOs of the whole system, since FMO calculations give only the MOs of the fragments and their combinations. This issue is essential when one would like to discuss the frontier orbitals associated with chemical reactions in biomolecular systems such as enzymatic reactions,

where the highest occupied MO (HOMO) and lowest unoccupied MO (LUMO) play vital roles. The well delocalized MOs over wide region of biomolecules should also play essential roles in transferring electrons or holes. Inadomi *et al.* [12] attempted the reproduction of entire MO sets from the AO density matrices calculated at the FMO-HF level as the FMO-MO approach [12,216]. The bottleneck of this scheme is manifestly the Fock matrix construction of the entire system, which should be carried out with parallelism [217]. The full diagonalization would be demanding for a larger Fock matrix, and thus the eigenvalues only near the occupied-virtual MO boundary with chemical relevance might be extracted [218,219]. To obtain the entire MO sets from FMO calculations more effectively, Tsuneyuki *et al.* [220] devised an alternative way called the linear combination of MOs (FMO-LCMO). In the FMO-LCMO procedure, the MO-based Fock matrix can be constructed by filling the monomer Fock matrices and the projected dimer matrices, where the spurious components of virtual MOs involving BDA [10] should be removed even for the minimal basis set. Very recently, FMO-LCMO was extended to DFT calculations with the three-body correction and also made usable for basis sets better than the minimal quality, yielding the greater practicality [221]. Nishioka *et al.* [222,223] applied the FMO-LCMO method to estimate the rate of CTs in several molecular complexes, and such a usage is a good example to handle long-range phenomena. It is noted that Fedorov *et al.* [224] confirmed accuracy of the total Fock matrix, especially on the influence of exchange contribution to ESP.

3. Applications

3.1. Protein-ligand interactions

The FMO method provides a very useful tool to theoretically analyze the molecular recognition in biomolecular systems. A benefit comes from its ability to specify the effective inter-fragment interactions between amino acid residue and ligand molecule or between their partial components, which are referred to as IFIE or PIE (see Sec. 1 and Sec. 2.1). In addition to the electrostatic, polarization, charge transfer, exchange repulsion interactions, the dispersion interaction can also be described when one considers the electron correlation effect appropriately. Through the following examples in this section, we will demonstrate how the FMO-IFIE and associated analyses have succeeded in elucidating the molecular recognition and related phenomena in biomolecular systems. In particular, the FMO-IFIE analysis is very helpful for SBDD since it enables the calculation of ligand-binding energy and the prediction of binding

affinity in protein-ligand systems. Further, the IFIE analysis can provide the information on the molecular interactions at the level of each amino acid residue, thus facilitating the rational drug design. It is also remarked that the accurate information on molecular structures is essential for these energetic analysis, and that primarily enthalpic part is discussed in the FMO framework. These difficulties still remain to be overcome (see also Sec. 3.5.3).

3.1.1. Estrogen receptor (ER)

The estrogen receptor (ER) is a member of the nuclear receptor (NR) superfamily and acts as a ligand-inducible transcriptional regulator that regulates expression of many genes involved in various physiological actions of its ligands at the transcriptional level. Its natural ligand, 17 β -estradiol, is associated with reproduction, homeostasis, differentiation and morphogenesis during embryonal development and postnatal life. ER is related to several diseases including breast cancer, endometrial cancer and osteoporosis, and therefore is a highly important drug target. The understanding of the molecular mechanism concerning the ligand binding to ER is thus essential for various issues such as transcriptional regulation, drug design, and tissue-specific selective estrogen receptor modulator (SERM). We here address the interaction analyses of the ER-ligand binding on the basis of the FMO method.

In conventional theoretical methods, the binding energy (ΔE) can be obtained by supermolecule calculation:

$$\Delta E = E_{\text{complex}} - (E_{\text{protein}} + E_{\text{ligand}}) \quad (75)$$

where E_{protein} , E_{ligand} and E_{complex} refer to the energies of protein, ligand and complex, respectively. In addition to this binding energy, the FMO method can provide the IFIE sum which represents an alternative evaluation for binding energy. When one regards a ligand molecule as fragment I , the ligand binding energy can be expressed by the sum of IFIEs between fragment I and all other fragments J as

$$IFIE_{sum}^I = \sum_J \Delta E_{IJ} \quad (76)$$

It has been found that there is a good correlation between ΔE by eqn (75) and experimental binding affinity [225] and that there is a similar tendency between ΔE and

the IFIE sum by eqn (76) [30]. This result indicates that the prediction of ligand binding affinity by the FMO method is feasible. It has also been found [30,225] through the charge distribution and CAFI analyses that the CT interaction play a pivotal role in the ligand binding with ER.

In terms of the IFIEs, one can quantitatively analyze the interaction energies between a ligand and each amino acid residue. For example, as seen in Fig. 11, those important residues which show strong interactions are found such as Glu353, Arg394 and His524 that form the hydrogen bond network with a ligand (17 β -estradiol; EST) and Phe404 that interacts with the hydrophobic part at the center of ligand. It was observed here that inclusion of the electron correlation is essential for the theoretical description. As observed in Fig. 12, there is a significant difference in IFIEs between the HF and MP2 results concerning the interactions between the EST and surrounding amino acid residues [30]. In particular, most of the interactions between the ligand and surrounding hydrophobic residues are attractive in the MP2 calculation, while they are repulsive in the HF calculation, thus demonstrating the importance of dispersion interactions associated with electron correlations.

In drug design, it is desirable to evaluate the ligand-binding interactions by dividing them into contributions from multiple functional groups. As for the amino acid residues, the division into the main and side chains enables molecular design at higher resolutions. Though the divisions of ligand molecule and amino acid residue should be avoided in the FMO2 method to maintain computational accuracy, it is possible to divide them in the FMO3/FMO4 method without the significant loss of accuracy [27,28]. In the case of ER system, it has been found [28] that one can perform the FMO4 analysis at the units of functional groups with chemical accuracy, in which a ligand molecule and amino acid residues are divided into a couple of pieces (see Fig. 13).

3.1.2. Retinoid X receptor (RXR)

The retinoid X receptor (RXR) is also a member of the NR superfamily and acts as a ligand-inducible transcriptional regulator as well as the ER. Its natural ligand 9-cis retinoic acid (9cRA) is the metabolite of vitamin A that controls morphogenesis, differentiation, and homeostasis during embryonal development and postnatal life. 9cRA is also an effective inhibitor of tumor cell growth, and this antitumor activity is useful in therapy and prevention of cancers such as human immunodeficiency virus (HIV) associated Kaposi's sarcoma [226,227]. RXR not only forms a homodimeric DNA complex, but also can form heterodimeric DNA complexes with various NRs, such as

retinoic acid receptor (RAR), thyroid hormone receptor (TR), vitamin D receptor (VDR) and peroxisome proliferator-activated receptor (PPAR) [228]. Because RXR has diverse important biological roles associated with human life and diseases, it has been one of the primary targets of drug discovery. To date, many experimental studies have been devoted to elucidate the transcriptional activation mechanism of RXR so as to efficiently exploit the functions of RXR. It is now widely accepted that the transcriptional activity of RXR, as well as many NRs, is induced by the binding of a ligand to RXR ligand-binding domain (LBD), which contains the ligand-dependent transcriptional activation function 2 (AF2), and controlled by the exchange of the binding of transcriptional coregulators including coactivator and corepressor to RXR [229].

On the basis of the FMO method, the molecular interactions of liganded RXR with steroid receptor co-activating factor-1 (SRC1) coactivator were evaluated at MP2/6-31G level [31] to examine the contribution of helix 12 (H12), which contains the core of the transcriptional AF2 activating domain, to the coactivator binding of RXR (see Fig. 14). The interaction between H12 and SRC1 was proved to be the main cause for the stabilization of the coactivator binding. In particular, highly conserved charged (Glu453) and hydrophobic (Phe450) residues in H12 were found to have stronger electrostatic and dispersion interactions with SRC1 than the other charged and hydrophobic residues in H12, respectively. In addition, the CT from RXR to SRC1 was found to occur mainly by the changes in charges of H12 residues. Large positive and negative charge changes were observed especially for Glu453 and for Lys631 and Ile632 in SRC1, respectively, indicating that Glu453 is an electron donor for Lys631 and Ile632 in this CT. These findings quantitatively demonstrated that H12 and its highly conserved residues significantly contribute to the coactivator binding not only by the electrostatic (ES) and dispersion interactions but also by the CT described with the quantum-mechanical framework.

The *ab initio* FMO calculations with MP2, spin-component scaled MP2 [230] and scaled opposite-spin MP2 [231] methods were also performed [32] for RXR complexes with its ligand 9cRA and SRC1 to examine the influence of mutations in transcriptional activation function 2 activating domain core (AF2C) of RXR on molecular interactions between 9cRA liganded RXR and SRC1 coactivator. The RXR-SRC1 interactions in three types of RXR-9cRA-SRC1 complexes, namely, the wild type (WT), a mutant whose Glu453 of AF2C was substituted by Lys (E453K), and another mutant whose Glu456 of AF2C was substituted by Lys (E456K), were compared. Comparison of WT, E453K and E456K suggested possible causes for a marked decrease in the transcriptional activity of RXR by the mutation of Glu453, a highly conserved charged residue of AF2C. It was

quantitatively demonstrated that the strength of the RXR-SRC1 interaction correlates with the degree of the transcriptional activation (WT > E456K > E453K). In E453K, the RXR-SRC1 interaction was substantially reduced by the AF2C-SRC1 repulsive interaction, and the CT from RXR to SRC1 was also inhibited by the decreased electron donation from AF2C to SRC1. These findings suggested that the inhibitions of the local RXR-SRC1 interaction via AF2C and of the local CT from RXR to SRC1 via AF2C may be the possible causes for the marked decrease in the transcriptional activity of RXR.

Further, FMO calculations were performed [33] for the α -subtype of the human RXR (hRXR α) complex with its natural ligand 9cRA to quantitatively specify the key residues with important roles for the ligand inducible information transmission of RXR. In the RXR-9cRA complex, the transactivation H12 adopts a canonical agonist conformation, which just corresponds to the AF2C. Through the analyses of molecular interactions with the FMO-MP2 method, Trp305 and Leu436 of the AF2C binding pocket were proved to stabilize the H12 canonical agonist conformation, and, at the same time, to participate in the recognition of the 9cRA molecule. Besides, through the analyses of orbital interactions by the FILM method [59], Trp305 and Leu436 were found to recognize the 9cRA molecule especially at its C19 methyl group, which has been most notably targeted to modify for agonist and antagonist design. Moreover, on the basis of the relationships of molecular interactions, it was suggested that the interactions of Trp305 and Leu436 with AF2C residues should be significantly influenced by the interactions of Trp305 and Leu436 with 9cRA. These findings quantitatively demonstrated that Trp305 and Leu436 are the possible key residues for the information transmission in liganded RXR, accounting for their importance suggested by experiments. Taken together, these results substantiated that the FMO approach is useful for the understanding of the detailed molecular mechanism underlying the transcriptional regulation of RXR and related nuclear receptors at the quantum mechanical level.

3.1.3. Ribulose biphosphate carboxylase/oxygenase (Rubisco)

Ribulose biphosphate carboxylase/oxygenase (Rubisco) catalyzes a part of reactions in plant photosynthesis. Two molecules of 3-phosphoglycerate (3PGA) are produced from ribulose 1,5-biphosphate (RuBP), a carbon dioxide and a water molecule. This reaction is the rate limiting step of the carbon fixation process. The improvement of this enzyme would therefore lead to producing crops with high quality, and consequently, would somewhat contribute to the prevention of the global warming. It is known that

the specificity factor, a good measure of the CO₂ fixation efficiency, takes almost the same values among the higher plant Rubiscos [232-235]. However, Rubisco of the red algae, *Galdieria partita*, has a twice higher specificity factor [236]. Thus, it is important to investigate the difference between *G. partita* Rubisco and others at the molecular level.

Clues to this difference have been suggested in the X-ray structures of Rubisco complexed with its substrate, RuBP. RuBP has two phosphates at the both termini of its carbon chain. When the substrate is bound, positions of each phosphate group are situated at two anion binding sites, P1 and P2, respectively. X-ray crystal structures with various ligands such as anions or chemical compounds including phosphate groups were reported, and interestingly, the anion binding sites of Rubisco change when the different ligands are bound. The anion binding site P1 has two different possible positions, P1 α and P1 β , and the P2 anion binding site also has two possible positions, P2 α and P2 β (Fig. 15). When the transition-state analogue to the substrate (RuBP), 2-carboxyarabinitol-1,5-bisphosphate (2CABPs), is bound, the phosphate groups are situated at P1 α and P2 α sites, while located at P1 β and P2 β sites when the reaction products 3PGA, inorganic anions HPO₄²⁻ or SO₄²⁻, are bound. Further, the occupation of anion binding site is correlated with the position of the loop6, which is well conserved among different species and composed of Val331 to Gln337. When the anions are at P1 α and P2 α sites, loop6 is at the closed position. When anions are at P1 β and P2 β sites, loop6 is at the open position. Observed positions of loop6 for the closed and open forms are also shown in Fig. 15. When loop6 is located at the closed position, it covers the substrate or ligand, where Lys334 is faced to P1 anion binding site. It is suggested that loop6 is situated at the closed position when carboxylation reaction is occurring because the Rubisco is at the closed position when the transition-state analogue 2CABP is bound. It is also suggested that loop6 is at the open position when the products 3PGA are bound. It seems that the loop6 catches the substrate during the chemical reaction and releases the products by opening it. Interestingly, only *G. partita* Rubisco has tendency to remain at the closed position even at the situation when other Rubiscos are at the open position. Even when only SO₄²⁻ is bound to *G. partita* Rubisco at P1 α site, loop6 is at the closed position. As mentioned above, during the chemical reaction, it seems that loop6 is closed as if it caught the substrate. Additionally, position of Lys334 in the closed state loop6 is faced with substrate and close to the position where the chemical reaction occurs. Thus, this difference in conformational change of loop6 may be related to the high specificity factor of *G. partita*.

Rubisco from one of the thermophilic red algae *Galdieria partita* with a high

specificity factor thus shows a characteristic difference from higher plant Rubiscos in structural change. The FMO analysis was performed [34] to investigate this difference by evaluating the IFIE values in comparison to experimental structural studies. Some important residues were found, which determined the loop6 stability or which made difference in the structure between the higher plant and *G. partita* Rubiscos. It was found that amino acid change of Lys18 to Ile18 was important for the difference in location at which anion binding site was occupied, P1 α or P1 β , when inorganic anions were bound to the enzyme. Occupation of P2 anion binding site made the stabilizing interaction between Lys128 and the loop6 stronger. Amino acid change of His386 to Gln386 contributed to the difference in the loop6 stability, while amino acid change of Met472 to Thr472 did not contribute to it. It was confirmed that the patterns of interactions among Thr65, Thr67 and Thr462 were consistent with previous experimental discussions. However, a case was found that Thr65 was not stabilized with anion at P1 α binding site in a closed-state structure of *G. partita* Rubisco. Some important amino acid residues for the stability of the loop6 closure were thus found and confirmed through the FMO calculations. In the future study, it would be desirable to confirm that these residues actually induce the open to close transition of the loop6 with dynamical simulations.

3.1.4. L-2-haloacid dehalogenase (L-DEX)

Halogenated compounds are widely used as herbicides, insecticides, plastics, solvents, and the starting materials for the production of these substances. However, they also cause serious environmental pollution and health problems owing to their toxicity, persistence, and transformation into hazardous metabolites. Dehalogenases detoxify various halogen compounds and seem to be promising from the viewpoint of environmental technology. Various microorganisms are known to produce dehalogenases that detoxify halogenated compounds by cleaving carbon-halogen bonds [237-239]. These enzymes are highly attractive candidates that can be used for the biodegradation of environmental pollutants.

L-2-haloacid dehalogenase (L-DEX) catalyzes the hydrolytic dehalogenation of L-2-haloalkanoic acids to produce the corresponding D-2-hydroxyalkanoic acids. This enzyme is expected to be applicable to bioremediation of environments contaminated with halogenated organic compounds. *In silico* molecular modeling was performed [35] for the analysis of the reaction mechanism of L-DEX from *Pseudomonas* sp. YL (L-DEX YL). The complexes of wild-type L-DEX YL and its Lys151Ala and Asp180Ala mutants

with its typical substrate, L-2-chloropropionate, were constructed by docking simulation. Subsequently, classical MD and *ab initio* FMO calculations of the complexes were performed, where the FMO method was applied at the MP2/6-31G level to estimate the IFIEs. Lys151 and Asp180, which were experimentally shown to be important to enzyme activity, interacted particularly strongly with L-2-chloropropionate, catalytic water, nucleophile (Asp10), and with each other. These calculations suggested that Lys151 stabilizes substrate orientation and balances the electric charge around the active site, while Asp180 stabilizes the rotation of the nucleophile Asp10, fixes catalytic water around Asp10, and prevents Lys151 from approaching Asp10. Further, Asp180 may activate catalytic water on its own or with Lys151, Ser175 and Asn177. These roles were consistent with those found in earlier studies. Thus, it was found that combination of classical MD and *ab initio* FMO calculations is a powerful tool for the elucidation of the mechanism of enzymatic reaction at the molecular level and can be applied to other catalytically important residues. These results may play an important role in elucidating the reaction mechanism and rational design of L-DEX YL with improved enzymatic activity or substrate specificity.

3.1.5. Influenza virus hemagglutinin (HA) complexed with saccharide receptor

Influenza is one of the most important infectious diseases of humans [240-243]. The hemagglutinin (HA) protein of the influenza virus binds to the host cell receptor in the early stage of viral infection. A change in binding specificity from avian $\alpha 2\text{-}3$ to human $\alpha 2\text{-}6$ receptor is essential for optimal human-to-human transmission and pandemics. Therefore, it is important to reveal the key factors governing the binding affinity of HA-receptor complex at the molecular level for the understanding and prediction of influenza pandemics. The FMO calculations for the interaction energy analysis of HA-receptor complexes were carried out [244] to quantitatively study the binding specificity of HAs to avian and human receptors. To discuss the binding property of influenza HA comprehensively, a number of HAs from human H1, swine H1, avian H3 and avian H5 viruses were analyzed. Detailed investigations about the interaction patterns of complexes of various HAs and receptor analogues then revealed that intra-molecular interactions between conserved residues in HA played an important role for HA-receptor binding. These results may provide a hint to understand the role of conserved acidic residues at the receptor binding site which were destabilized by the electrostatic repulsion with sialic acid. The calculated binding energies and interaction patterns between receptor and HAs were consistent with the binding specificities of

each HA and thus explained the receptor binding mechanism. The calculated results in the FMO-IFIE analysis also provided a number of viewpoints regarding the models for the HA-receptor binding specificity associated with mutated residues. Examples included the important roles of Glu190 and Gln226 for the binding specificity of H5 HA. Since the adaptation of H5 HA to human receptor is a serious issue and the mechanism of the specificity change is unknown, this result would be helpful for the prediction of the change in receptor specificity associated with forthcoming possible pandemics.

Similar approaches are also applicable to various viral subspecies. The FMO analysis on the binding between HA and α 2-6 linked sialyloligosaccharides was performed [245] for a pandemic influenza virus 2009/H1N1pdm, which was of swine-origin. The FMO-MP2 calculations were carried out to study the differences in the binding affinity with human receptors between the 2009 pandemic virus and those previous ones such as A/swine/Iowa/1930 and A/Puerto Rico/8/1934. As the results, the strongest receptor binding affinity was observed for the 2009/H1N1pdm. The IFIE analysis revealed that the amino acid mutation of 2009/H1N1pdm, Ser145Lys, was a major cause of such strong binding affinity. Therefore, the HA of pandemic 2009/H1N1pdm has been found to recognize the α 2-6 receptor much stronger than the 1930-swine and 1934-human HAs. This result may also be useful for the prediction of mutation and associated change in receptor specificity toward possible pandemics.

3.2. Protein-protein interactions: Influenza virus HA antigen-antibody complex

Recent concerns about the avian and swine influenza viruses highlight its threat and the need to understand its evolutionary dynamics. The influenza virus has a remarkable ability to escape host defense mechanisms by altering its binding characters through changes of amino acid residues in the pertinent proteins. This property is referred to as antigenic drift [240-243] and has been thought to result from the accumulation of a series of amino acid changes in antigenically important regions of proteins. In addition, the viral resistance against some drugs is associated with analogous mutation properties as well. It is thus essential to elucidate the molecular mechanisms by which viruses alter their ligand binding characters in order to find an efficient way to suppress the pandemics and epidemics of influenza.

The FMO-MP2/6-31G calculations for an antigen-antibody system consisting of the HA monomer and the Fab fragment (totally 14086 atoms, 921 residues and 78390 AOs) were performed with cluster computers and 4096 vector processors of ES (see Sec. 2.2.1). Later, the calculation with the extended 6-31G* basis set (121314 AOs) was also carried

out on cluster computers, and some specific residues associated with probable mutations were successfully identified through the IFIE analysis, thus providing a method to predict the forthcoming mutations in HA [58,246]. Fig. 16 illustrates the structure of HA-Fab antigen-antibody system of H3N2 A/Aichi/2/68 influenza virus (PDB entry: 1EO8), in which the IFIEs of the residues in HA monomer with the whole Fab fragment of antibody (colored in yellow) are shown with red and blue colors denoting the interaction energies of stabilization (negative) and destabilization (positive), respectively.

The FMO-MP2 and FMO-MP3 electronic-state calculations were also carried out on ES2 for a protein complex consisting of HA trimer and two Fab fragments to obtain the FMO-MP2.5 energy of the complex (Fig. 17, see Sec. 2.2.1 for MP2.5). The FMO-MP3/6-31G calculation, which was a correlated all-electron calculation for the world's largest target system (36160 atoms) at that time, was completed in only 5.8 hours with 128 nodes on ES2 [68]. Thus, a realistic applicability of FMO-MP3 calculations to large-scale proteins has been demonstrated when supercomputers are available.

We postulated that, for a mutation in HA to take place, it should satisfy two conditions. First, after the mutation, that HA should preserve the function for the infection into host cell. This property can be evaluated in terms of the hemadsorption experiment by amino acid substitution [242,243], in which the hemadsorption activity of mutant associated with the binding to sialic acid moieties on the host cell is measured. Second, the amino acid site at which the residue is substantially attracted by the antibody would be preferentially mutated to escape the antibody pressure. Thus we expected that residues satisfying these two conditions would have a high probability of the mutation.

The FMO-IFIE results were thus interpreted based upon the above assumption [58]. Fig. 16 illustrates the IFIE sums between the residues in the antigenic region E of HA monomer and the whole Fab fragment of antibody for (A) charged, (B) hydrophobic, and (C) polar residues. The green, red, and blue bars in Fig. 16 represent that they are located at the positive (allowed), negative (prohibited), and unspecified (not measured) residue sites in the hemadsorption experiments, respectively. The numbers below the bars denote the years of the occurrence of the mutation. In (B), among the five residues showing the attractive interaction, two of them actually mutated in the past, which are observed to be located at the allowed or positive positions in the hemadsorption experiments. In (A) as well, only the residue sites which show the strong attraction and whose mutation is not prohibited in the hemadsorption experiment mutated in the past.

An exceptional case for Thr83 in (C) can also be accounted for by considering a special manner of fragmentation employed in the FMO method [246]. Thus, through the combination of hemadsorption experiment and FMO-IFIE analysis, we have provided a novel protocol for predicting the residues in HA likely to mutate in the future.

Molecular interaction analyses based on the FMO-MP2.5 calculation were carried out for the HA trimer-Fab fragment complex system as well (Fig. 17). The FMO-IFIE analysis again provided a reasonable result for the prediction of probable mutations in HA associated with the escape from antibody pressure [71].

3.3. Protein-DNA/RNA interactions

3.3.1. Cyclic-AMP receptor protein (CRP)

CRP is one of the most extensively studied sequence-specific DNA-binding proteins. Acting as a transcription factor, CRP regulates the gene expression related to sugar metabolism. When cyclic-AMP (cAMP), an intercellular signaling molecule, binds to CRP, following the receptor dimerization, the conformation changes and binding to specific DNA sequences near promoters are induced, thus the transcription is activated. CRP is an activator of transcription in the lactose operon in *Escherichia coli*. In order to reveal this transcriptional regulation mechanism, one should understand the specific interactions between DNA and the transcription factor. Since experiments show that CRP specifically recognizes a conserved region 5'-(T₄G₅T₆G₇A₈)-3' in the consensus DNA sequence, the FMO calculations concerning this binding specificity were carried out [212] (see also Fig. 6), where a complex model of DNA-CRP-cAMP was employed for the HF and MP2 calculations. The main chain (sugar-phosphate) and base parts of DNA were separated in the fragmentation, and the IFIE analysis was performed for the units of base and main chain.

It was observed in the IFIEs between the CRP-cAMP complex and each sequence of DNA that the interactions with not the main chain but the base part represent the sequence specificity obtained in experiments. As for the interaction between CRP and the base part of DNA, that with the conserved sequence 5'-(T₄G₅T₆G₇A₈)-3' was the strongest, and the sequence specificity was particularly significant in the G:C base pairs at the 5-th and 7-th sites. It is noted here that the interaction energy analysis by the Amber94 force field did not reproduce the sequence specificity found in the FMO analysis, thus showing the necessity of quantum chemical calculations.

The FMO calculations were also performed to analyze the interactions inside the

DNA such as those for the hydrogen bonding between the Watson-Crick base pairs and for the stacked bases [212]. While the results based on the HF and MP2 methods showed a similar tendency concerning the hydrogen bonding, the HF and the MP2 methods gave the repulsive and attractive interactions, respectively, concerning the stacking interactions, again indicating the importance of dispersion forces associated with the electron correlations in DNA systems.

3.3.2. Neuro-oncological ventral antigen (NOVA)

RNA-binding proteins (RBPs) express their proper function and play essential roles in the RNA metabolism when they bind to the target RNA molecules [247]. The mechanism of stable complex formation with RNA is an important issue to understand how an RBP acts on only its targets. Since 1990s, more than 600 RBP-RNA complex structures have been resolved, and the observations of atomic details of RBP-RNA interactions and of global structural change upon complex formation have provided qualitative insights into the mechanism of complex stabilization [248,249]. The FMO method was applied to the problem of the protein-RNA complex stabilization, in which neuro-oncological ventral antigen (NOVA) was selected as a model system. NOVA contains three K-homology (KH) domains [250,251] which are known as canonical RNA-binding domains. One of NOVA KH domains (NOVA-2 KH3) bound to an RNA was resolved by X-ray crystallographic analysis [252] (Fig. 18), which consists of *ca.* 2000 atoms (87 amino acid and 20 RNA residues). Upon the complex formation with RNA, NOVA-2 KH3 (denoted by NOVA below, for simplicity) specifically recognizes the "UCAY" (Y: pyrimidine base) quartet (Fig. 18). An observation of the X-ray structure indicates that specific and non-specific amino acid residue-RNA base interactions, pinioning an RNA by the GxxG motif and the variable loop named as "molecular vise", are also important for the stabilization of complex formation (Fig. 18). The FMO-IFIE analysis at MP2/6-31G level [253] demonstrated that the change of electronic state occurs over the whole protein structure upon the NOVA-RNA complex formation, and is intimately related to the secondary structure of NOVA. The FMO results also suggested that the pattern of change of IFIEs is dependent on the positional relation between NOVA and RNA (Fig. 19). It is expected that the change of electronic state under RBP-RNA interactions reflects the specific or non-specific RBP-RNA recognition condition and thus FMO can be a useful tool to predict specific RBP-RNA complex.

3.4. Inorganic nano systems

3.4.1. Extension to inorganic and solid-state systems

As addressed in Refs. [2-9], the FMO calculations have been used mostly for biochemical molecules or hydrated clusters. However, the FMO method itself should be potentially applicable to non-biochemical or inorganic systems, for example, solids with moderate band gaps, if appropriate extensions are made. Stoll [153] developed the IS as an MO-based many-body approach for solid, and a series of related developments (as addressed above) and applications (*e.g.*, [254,255]) have been reported. Collins *et al.* [43,256] proposed the SMF, oriented to solid modeling. Fedorov *et al.* [257] devised a constraint version of FMO scheme named the adaptive frozen orbital (AFO) method, where the BDA-related hybrid orbitals were kept unrelaxed to avoid the potential overpolarization at the HF stage. With AFO, they modeled the adsorption of toluene and phenol molecules onto zeolite surface. The AFO method of Ref. [257] was extended to the gradient evaluation, and the geometries of silicon rod model were optimized; no actual formulation of gradient was provided unfortunately [258].

The total energies of adamantane-shaped carbon and silicon were reasonably reproduced by FMO4 only with ABINIT-MP [27], in comparison with the reference MO values. It is notable that no restriction on orbital relaxation was imposed in the FMO4 calculations as in the cases of two-body and three-body approximations, in contrast to the AFO scheme [257,258]. This feature of FMO4 can be advantageous in modeling large-scale hetero-boundary systems where both electronic polarization and CT are essential. The complexes composed of solid surfaces and proteins [259] under the hydration condition can be such demanding targets. For the adsorption of an artificially designed peptide (RKLPGA: Arg-Lys-Leu-Pro-Asp-Ala) [260] onto the silica surface, the FMO4-MP2(CD)/6-31G calculations were performed [45], where the silica cluster contained as many as 257 silicon atoms and the silicon BDA technique [261] was employed for three-dimensional Si-O bonding networks (Fig. 20). Ref. [45] revealed detailed interaction pictures of the electrostatic stabilizations as well as electron delocalizations. The nano-biotechnology field will become more important in industrial applications and thus be of attractive target for the FMO calculations, where the size of proteins on the surfaces will be of a few hundred amino acid residues.

In Ref. [261], the FMO-HF and -MP2 energies as well as frequency-dependent polarizability (see the LR equation in Sec. 2.3) were systematically evaluated by using ABINIT-MP for the linear silane, carbosilane and siloxane. As a consequence, the reasonable results were obtained as long as the fragmentation unit was set properly,

demonstrating the applicability to silicon-based engineering polymers. The following FMO calculations were performed with GAMESS, where the excited states in molecular solid were modeled [262] and the geometry of hetero-atomic nano-ribbons was optimized [263]. The ionic liquids have attracted great industrial and military interests, and thus there have been several reported attempts [8,264,265]. Very recently, the cellulose aggregates [266] and the molecular sieves [267] were successfully modeled. The FMO applications to problems of inorganic and material fields will thus get popularity before long.

3.4.2. Comparison on hydration dynamics of *cis*- and *trans*-platins

FMO-MD with electron correlation has given important insight into the hydration dynamics of *cis*- and *trans*-platins [111]. *Cis*-platin (*cis*-[Pt^{II}Cl₂(NH₃)₂]) has a high anticancer activity, but this activity is not shared by its geometrical isomer, *trans*-platin (Fig. 21). This difference in their anticancer activity used to be explained by a hypothesis that *trans*-platin should have no ability to form a complex with DNA leading to anticancer activity, until recent findings that some *trans*-type Pt-complexes show antitumor activities. Thus, the reason for their difference in biochemical activity remains obscure. Some steps earlier than the final DNA binding might play an important role in the efficacy of the medicine. One of the earlier steps, hydration, was thus focused by Mori *et al.* [111], who analysed the difference in hydration of *cis*- and *trans*-platins by FMO-MD to understand their difference in anticancer activity.

Hydrated *cis*- and *trans*-platins were subjected to FMO-MD simulations. From a methodological viewpoint, this study features use of MCP and MP2/FMO-MD. Each platin complex was hydrated with a spherical droplet of water centred at the Pt atom. FMO calculations were performed at the FMO(3)-MP2 level (see Sec. 2.4.2). The generalized version of DF (see Sec. 2.4.4) was applied to allow for the generation of proton-transferred species during the production MD runs, but mostly DF defined the platin and water molecules as independent fragments. After 1-ps equilibration, 2-ps production MD run were performed at 300 K.

The natural charge via NBO on each ligand in *cis*- and *trans*-platin and the Pt-Cl bond lengths are plotted as functions of simulation time (Fig. 22). The Pt/Cl sites had larger fluctuation than the NH₃ sites had in both platins. This difference should be due to the fact that NH₃ has no amplitude in the highest occupied molecular orbital. The two graphs in Fig. 22 showed a correlation between fluctuation of the Pt/Cl sites and that of the Pt-Cl bond. The frequency of the charge fluctuation was 334 cm⁻¹, which can be

assigned to the Pt-Cl stretching mode coupled with intermolecular vibrations between the platinum and the water molecules. The correlation in charge fluctuation on Pt and Cl sites revealed existence of CT interaction between them. The CT interaction coupled with the fluctuation of the solvent water should have induced a Pt-Cl bonds fluctuation because the frontier MO that took part in the CT was a Pt-Cl antibonding orbital. *Trans*-platin has inversion symmetry, and hence its dipole moment is much smaller than that of *cis*-platin. This follows that more water molecules coordinated to *cis*-platin than to *trans*-platin. Thus, the CT interaction coupled with the solvent motion was stronger in *cis*-platin than in *trans*-platin. As a result, the Pt-Cl bonds became easier to elongate to cleave in the hydrated *cis*-platin than in the hydrated *trans*-platin.

This study thus exemplified the ability of FMO-MD to give a microscopic picture of molecular function in solvent.

3.5. Future directions

3.5.1. Highly correlated methods on high performance computing

In this subsection, modern theoretical and computational MO approaches, which have not yet been used in FMO, are briefly sketched along three future directions. The incorporation of these elements into the FMO methodologies will enrich the validity and usefulness of the FMO method for a variety of applications.

First, the so-called explicitly correlated methods like R12 or F12 [268-270] would be attractive to implement, if absolute accuracy is pursued. Although the computational costs of these methods should be demanding even with various integral factorization and quadrature techniques (typical order N^8), ongoing developments of massively parallel computers [211] may lead to the practical usability, as discussed by Ten-no [270]. The correlation energy convergence against the basis set size in the R12 and F12 approaches is much faster than that in the conventional higher-order methods [46,49], and this feature may be convenient for the FMO calculations where the ESP treatment itself [5,6,8] has a vulnerability for large basis sets due to interferences of charge clouds [271]. Anyhow, much more cores of computers will be needed for the intra-fragment processing of actual calculations.

Second, the approximation of correlating amplitude tensors by lower-rank ones is addressed. To our knowledge, such approach was practically introduced by Olsen *et al.* [272,273] as low-rank CI in 1987. In 2000s, the tensor decomposition or factorization techniques, *e.g.*, higher-order singular value decomposition (HOSVD), have been

revisited and extended for general correlation methods (MP2 or CC) by several groups [274-280]. The motivation of these approaches was put on the reductions of both the computational costs of a series of tensor contractions and the memory spaces of higher-excitation amplitudes and associated intermediates, as summarized in the literature [274,275]. The potential applicability of tensor decompositions for FMO correlated calculations was pointed out in Ref. [278]. The methods of density matrix renormalization group (DMRG) or complete graph tensor network (CGTN) [280-286] have proved the powerful ability in handling very large complete active space of MCSCF expansions, especially in the multi-nuclear transition metal complexes with serious near-degeneracy, where the problems with a few dozen of active MOs were made tractable through efficient renormalizations. Very recently, Kurashige *et al.* [287] reported an illustrative DMRG study on the Mn_4CaO_5 cluster taken from the redox-center of photosystem II. One more approach to be addressed in this context is the distinguishable CC (DCC) method [288] as a variant of parameter-scaled CC [289]. Ref. [288] showed that multiple bond dissociation processes can be properly described with only doubles amplitudes under the Brueckner orbital optimization [78]. DCC may be effective for the FMO calculation in such a material design as singlet radicals.

Third, the computational approaches with general purpose graphic processor unit (GPGPU) are briefed below. The usages of GPGPU were reported for the integral generation [290-292], MP2(RI) [293], CC [294,295] and CIS [296]. The actual GPGPU calculations should be performed at a tradeoff between speed-up with factor of more than ten and diminished accuracy due to single precisions, while one can use the double precisions on many GPUs now. It is notable that the TeraChem program for efficient GPGPU computation has been developed and the geometry optimizations of several proteins have been demonstrated at the HF level [297]. Reduction of memory requirement is another serious demand in GPGPU computation because a GPU usually has only a small-sized memory. This memory-reduction will be necessary also in the next-generation supercomputers, whose memory resource per processor must be made smaller to minimize power consumption and failure probability. This demand may be fulfilled by the tensor hyper-contraction (THC) approach, a recently proposed memory-saving algorithm [298-300]. THC or similar tensorial factorization methods may reduce the memory requirement in FMO and enable practical FMO calculations both on GPU and on the next-generation supercomputers. Continuous efforts should be made for future improvement.

3.5.2. Nuclear quantum effects

Biomolecular systems contain plenty of hydrogen atoms or protons, which often show significant nuclear quantum effects. It is possible to describe the nuclear quantum effects in terms of the path integral formulation [301]. *Ab initio* path integral MD simulations coupled with the FMO method were carried out for water monomer, trimer, and glycine pentamer systems [191]. *Ab initio* path integral Monte Carlo simulations for water trimer system [302] were also performed in which the electron correlation effects were taken into account up to the MP3 level. Geometrical structures and molecular interactions in hydrogen-bonded systems were thus analyzed in light of the interplays among the nuclear quantum, thermal, and electron correlation effects. The use of the FMO method would then be essential to enable the *ab initio* simulations for larger systems including water molecules, for which the incorporation of PBC [41] would be desirable.

3.5.3. Pharmaceutical applications

The FMO method is highly expected to play vital roles in pharmaceutical applications [303-310]. The problems that remain to be overcome or investigated are as follows. First, since the FMO calculations have primarily been focused on accurate evaluation of the enthalpic contributions to free energy in protein-ligand systems, more efforts for evaluating the entropic contributions and associated solvent effects should be made extensively in the near future. Regarding the entropy evaluations, it may be more practical to take them into account through classical-mechanical means instead of expensive quantum-mechanical approaches, considering the high cost of current *ab initio* calculations such as FMO-MD. Further, a number of statistical or information-science techniques, including conventional QSAR (quantitative structure-activity relationship) approaches [311-313], would be helpful to perform actual applications to drug design on the basis of *ab initio* calculations. In addition, in order to enhance the success rate in pharmaceutical applications, the accurate determination of the three-dimensional structures of receptor-ligand complexes including the protonation states would be the key issue. The FMO methodology is then expected to play helpful or essential roles for practical applications with the aid of those advanced techniques such as the partial geometry optimization [97], the FMO4 [27,28], and the pair interaction energy decomposition analysis (PIEDA) [304,314].

3.5.4. Coarse graining

Since biomolecular systems are inherently huge and their dynamical properties play pivotal roles for their functions, *ab initio* approaches based on the electronic structure calculations should be combined with or provide microscopic information to more macroscopic descriptions. The quantum mechanical / molecular mechanical (QM/MM)-like multi-scale approaches [315-317] are typical examples, and the FMO method can be employed in this context as well. The MFMO approach, as addressed above, would provide a useful tool for this direction. In this connection, a method to obtain atomic charges for classical force-field simulations on the basis of the FMO method has been proposed [318,319]. In this FMO-RESP method, the RESP (restrained electrostatic potential) charges in biomolecular systems can be determined so that the electric charges assigned to each atom accurately reproduce the electrostatic potentials given by the FMO calculations, where some restraints to prevent the fitted charges from deviating from the reference charges employed in the original force field are imposed. This methodology would thus enable classical-mechanical force-field simulations with varied electric charges according to the structural and environmental changes in biomolecular systems [320].

As for the dynamical behaviors in biomolecular systems, the charge and energy transfer issues are of great interest in theoretical analyses. Beyond the conventional approaches due to Marcus and Förster [321], a number of more modern, *ab initio* methodologies have been developed, in which the effects of dynamical electronic coupling, surrounding quantum phonons, inelastic interactions, temporal interference and quantum coherence can be described quantitatively for the rate constants and the population dynamics [322-324]. The FMO method should then provide basic, microscopic information underlying these analyses for multi-scale or macroscopic dynamics.

4. Summary

Since its proposal in 1999, through vast of investigations on theory, implementation and applications, the *ab initio* FMO method has extensively been developed as a convenient tool for many researchers in the fields of biochemistry, medicinal chemistry and nanotechnology. Currently a number of FMO associated programs are freely available, in which users can select a variety of quantum-chemical methods according to individual problems and GUIs for visual analyses. One of the most significant advantages of the FMO method is its capabilities to be easily combined with other

quantum- and classical-mechanical theories developed independently. In addition, its execution on parallelized high-performance computing platforms is promising. The FMO method has thus been confirmed to be an indispensable tool for many applications in molecular sciences, and is still in progress toward more accurate and faster computations in a wide variety of research fields.

Acknowledgements

We would like to thank Shinji Amari, Hiroshi Chuman, Kuniyoshi Ebina, Takatoshi Fujita, Takeshi Ishikawa, Mika Ito, Akifumi Kato, Chisachi Kato, Kazuo Kitaura, Ikuo Kurisaki, Hirotohi Mori, Tadashi Murase, Katsuhisa Nakajima, Takashi Nakamura, Tatsuya Nakano, Eri Nobusawa, Stuart Rothstein, Yasuteru Shigeta, Jewgeni Starikov, Kiyoshi Tanaka, Chiduru Watanabe, Hirofumi Watanabe, Katsumi Yamashita, and other many persons for various useful discussions, comments, and collaborations during the years through which we have developed the FMO related work. We also thank Mieko Nakamura for the help for editing the present manuscript. Financial supports for the present and related work have been provided by JST-CREST, FSIS and RISS at University of Tokyo (IIS), SFR at Rikkyo University, Health and Labour Sciences Research Grants of Japan, and MEXT-Kakenhi (Nos. 18066009 and 23540451).

Notes and references

- 1 K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **313**, 701.
- 2 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **120**, 6832.
- 3 D.G. Fedorov and K. Kitaura, in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, ed. E.B. Starikov, J.P. Lewis and S. Tanaka, Elsevier, Amsterdam, 2006, pp. 3.
- 4 D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 6904.
- 5 D. G. Fedorov and K. Kitaura, ed., *The FRAGMENT MOLECULAR ORBIAL METHOD: Practical Applications to Large Molecular Systems*, CRC Press, Boca Raton, 2009.
- 6 T. Nagata, D. G. Fedorov, and K. Kitaura, in *Linear-Scaling Techniques in Computational Chemistry and Physics*, ed. R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski, Springer Netherlands, 2011, pp. 17.
- 7 D. G. Fedorov, T. Nagata, and K. Kitaura, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7562.
- 8 M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V Slipchenko, and J. A. Boatz, *J. Phys. Chem. B*, 2009, **113**, 9646.
- 9 M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V Slipchenko, *Chem. Rev.*, 2012, **112**, 632.
- 10 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, *Chem. Phys. Lett.*, 2000, **318**, 614.
- 11 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, *Chem. Phys. Lett.*, 2002, **351**, 475.
- 12 Y. Inadomi, T. Nakano, K. Kitaura, and U. Nagashima, *Chem. Phys. Lett.*, 2002, **364**, 139.
- 13 M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.
- 14 D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, *J. Comput. Chem.*, 2004, **25**, 872.
- 15 G. D. Fletcher, M. W. Schmidt, B. M. Bode, and M. S. Gordon, *Comput. Phys. Commun.*, 2000, **128**, 190.
- 16 H. Sekino, Y. Sengoku, S. Sugiki, and N. Kurita, *Chem. Phys. Lett.*, 2003, **378**, 589.

- 17 S. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, *Chem. Phys. Lett.*, 2003, **382**, 611.
- 18 S. Sugiki, M. Matsuoka, R. Usuki, Y. Sengoku, N. Kurita, H. Sekino, and S. Tanaka, *J. Theor. Comput. Chem.*, 2005, **1**, 183.
- 19 T. Ishikawa, T. Ishikura, and K. Kuwata, *J. Comput. Chem.*, 2009, **30**, 2594.
- 20 T. Takami, J. Maki, J. Ooba, Y. Inadomi, H. Honda, R. Susukita, K. Inoue, T. Kobayashi, R. Nogita, M. Aoyagi, T. E. Simos, and G. Maroulis, in *AIP Conference Proceedings*, AIP, 2007, vol. 963, pp. 122.
- 21 T. Takami, J. Maki, J. Ooba, Y. Inadomi, H. Honda, T. Kobayashi, R. Nogita, and M. Aoyagi, *arXiv: cs/0701075*, 2007.
- 22 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2009, **131**, 171106.
- 23 S. Amari, M. Aizawa, J. Zhang, K. Fukuzawa, Y. Mochizuki, Y. Iwasawa, K. Nakata, H. Chuman, and T. Nakano, *J. Chem. Inf. Model.*, 2006, **46**, 221.
- 24 I. Kurisaki, K. Fukuzawa, Y. Komeiji, Y. Mochizuki, T. Nakano, J. Imada, A. Chmielewski, S. M. Rothstein, H. Watanabe, and S. Tanaka, *Biophys. Chem.*, 2007, **130**, 1.
- 25 D.G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2012, **116**, 704.
- 26 S. Tanaka, C. Watanabe, and Y. Okiyama, *Chem. Phys. Lett.*, 2013, **556**, 272.
- 27 T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, and S. Tanaka, *Chem. Phys. Lett.*, 2012, **523**, 128.
- 28 C. Watanabe, K. Fukuzawa, Y. Okiyama, T. Tsukamoto, A. Kato, S. Tanaka, Y. Mochizuki, and T. Nakano, *J. Mol. Graph. Model.*, 2013, **41**, 31.
- 29 B. M. B. Van Schouwen, M. Nakano, H. Watanabe, S. Tanaka, H. L. Gordon, and S. M. Rothstein, *J. Mol. Struct. THEOCHEM*, 2010, **944**, 12.
- 30 K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura, and T. Nakano, *J. Phys. Chem. B*, 2006, **110**, 24276.
- 31 M. Ito, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, *J. Phys. Chem. B*, 2007, **111**, 3525.
- 32 M. Ito, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, *J. Phys. Chem. A*, 2008, **112**, 1986.
- 33 M. Ito, K. Fukuzawa, T. Ishikawa, Y. Mochizuki, T. Nakano, and S. Tanaka, *J. Phys. Chem. B*, 2008, **112**, 12081.
- 34 H. Watanabe, T. Enomoto, and S. Tanaka, *Biochem. Biophys. Res. Commun.*, 2007, **361**, 367.
- 35 T. Nakamura, A. Yamaguchi, H. Kondo, H. Watanabe, T. Kurihara, N. Esaki, S. Hirono, and S. Tanaka, *J. Comput. Chem.*, 2009, **30**, 2625.

- 36 T. Fujita, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2009, **478**, 295.
- 37 A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Macmillan, New York, 1982.
- 38 K. Kitaura, T. Sawai, T. Asada, T. Nakano, and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **312**, 319.
- 39 K. Kitaura, S. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, *Chem. Phys. Lett.*, 2001, **336**, 163.
- 40 C. H. Choi and D. G. Fedorov, *Chem. Phys. Lett.*, 2012, **543**, 159.
- 41 T. Fujita, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2011, **506**, 112.
- 42 J. Friedrich, M. Hanrath, and M. Dolg, *J. Chem. Phys.*, 2007, **126**, 154110.
- 43 H. M. Netzloff and M. A. Collins, *J. Chem. Phys.*, 2007, **127**, 134113.
- 44 L. Huang, L. Massa, and J. Karle, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1849.
- 45 Y. Okiyama, T. Tsukamoto, C. Watanabe, K. Fukuzawa, S. Tanaka, and Y. Mochizuki, *Chem. Phys. Lett.*, 2013, **566**, 25.
- 46 I. Shavitt and R. Bartlett, *Many-Body Methods in Chemistry and Physics*, Cambridge University Press, Cambridge, 2009.
- 47 W. Kutzelnigg, *Int. J. Quantum Chem.*, 2009, **109**, 3858.
- 48 D. Cremer, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 509.
- 49 R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.*, 2007, **79**, 291.
- 50 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **121**, 2483.
- 51 D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2006, **433**, 182.
- 52 D. G. Fedorov, K. Ishimura, T. Ishida, K. Kitaura, P. Pulay, and S. Nagase, *J. Comput. Chem.*, 2007, **28**, 1476.
- 53 K. Ishimura, P. Pulay, and S. Nagase, *J. Comput. Chem.*, 2006, **27**, 407.
- 54 Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima, and K. Kitaura, *Theor. Chem. Acc.*, 2004, **112**, 442.
- 55 Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari, and K. Kitaura, *Chem. Phys. Lett.*, 2004, **396**, 473.
- 56 Although another name ABINIT-MPX was often used to refer to the “extended (X) version”, “X” is omitted in the present article for simplicity.
- 57 Y. Mochizuki, K. Yamashita, T. Murase, T. Nakano, K. Fukuzawa, K. Takematsu, H. Watanabe, and S. Tanaka, *Chem. Phys. Lett.*, 2008, **457**, 396.

- 58 K. Takematsu, K. Fukuzawa, K. Omagari, S. Nakajima, K. Nakajima, Y. Mochizuki, T. Nakano, H. Watanabe, and S. Tanaka, *J. Phys. Chem. B*, 2009, **113**, 4991.
- 59 T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, H. Tokiwa, S. Tanaka, and K. Tanaka, *Theor. Chem. Acc.*, 2007, **118**, 937.
- 60 T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, S. Tanaka, and K. Tanaka, *Chem. Phys. Lett.*, 2008, **463**, 189.
- 61 S. Sæbø and P. Pulay, *Annu. Rev. Phys. Chem.*, 1993, **44**, 213.
- 62 S. F. Boys and F. Bernardi, *Mol. Phys.*, 2002, **100**, 65.
- 63 T. B. Pedersen, F. Aquilante, and R. Lindh, *Theor. Chem. Acc.*, 2009, **124**, 1.
- 64 T. Ishikawa and K. Kuwata, *Chem. Phys. Lett.*, 2009, **474**, 195.
- 65 Y. Okiyama, T. Nakano, K. Yamashita, Y. Mochizuki, N. Taguchi, and S. Tanaka, *Chem. Phys. Lett.*, 2010, **490**, 84.
- 66 A. P. Rahalkar, M. Katouda, S. R. Gadre, and S. Nagase, *J. Comput. Chem.*, 2010, **31**, 2405.
- 67 M. Katouda, *Theor. Chem. Acc.*, 2011, **130**, 449.
- 68 Y. Mochizuki, K. Yamashita, K. Fukuzawa, K. Takematsu, H. Watanabe, N. Taguchi, Y. Okiyama, M. Tsuboi, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2010, **493**, 346.
- 69 S. Sæbø and J. Almlöf, *Chem. Phys. Lett.*, 1989, **154**, 83.
- 70 M. Pitoňák, P. Neogrády, J. Cerný, S. Grimme, and P. Hobza, *Chem. Phys. Chem.*, 2009, **10**, 282.
- 71 A. Yoshioka, K. Fukuzawa, Y. Mochizuki, K. Yamashita, T. Nakano, Y. Okiyama, E. Nobusawa, K. Nakajima, and S. Tanaka, *J. Mol. Graph. Model.*, 2011, **30**, 110.
- 72 Y. Okiyama, K. Fukuzawa, H. Yamada, Y. Mochizuki, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2011, **509**, 67.
- 73 D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2012, **116**, 704.
- 74 Y. Mochizuki, K. Yamashita, T. Nakano, Y. Okiyama, K. Fukuzawa, N. Taguchi, and S. Tanaka, *Theor. Chem. Acc.*, 2011, **130**, 515.
- 75 G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, *J. Chem. Phys.*, 1988, **89**, 7382.
- 76 R. Kobayashi and A. P. Rendell, *Chem. Phys. Lett.*, 1997, **265**, 1.
- 77 J. A. Pople, M. Head-Gordon, and K. Raghavachari, *J. Chem. Phys.*, 1987, **87**, 5968.
- 78 N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, *Chem. Phys. Lett.*, 1989, **164**, 185.

- 79 D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2004, **389**, 129.
- 80 M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477.
- 81 Y. Shimodo, K. Morihashi, and T. Nakano, *J. Mol. Struct. THEOCHEM*, 2006, **770**, 163.
- 82 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
- 83 S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 84 D. G. Fedorov, Y. Alexeev, and K. Kitaura, *J. Phys. Chem. Lett.*, 2011, **2**, 282.
- 85 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787.
- 86 Y. Zhao and D.G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
- 87 J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.
- 88 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.
- 89 Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, *Chem. Phys. Lett.*, 2003, **372**, 342.
- 90 Y. Komeiji, Y. Mochizuki, T. Nakano, and H. Mori, in *Molecular Dynamics - Theoretical Developments and Applications in Nanotechnology and Energy*, ed. Lichang Wang, InTech, 2012.
- 91 D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 2722.
- 92 T. Nagata, D. G. Fedorov, and K. Kitaura, *Chem. Phys. Lett.*, 2010, **492**, 302.
- 93 Y. Komeiji, Y. Mochizuki, and T. Nakano, *Chem. Phys. Lett.*, 2010, **484**, 380.
- 94 Y. Mochizuki, T. Nakano, Y. Komeiji, K. Yamashita, Y. Okiyama, H. Yoshikawa, and H. Yamataka, *Chem. Phys. Lett.*, 2011, **504**, 95.
- 95 T. Ishikawa and K. Kuwata, *J. Phys. Chem. Lett.*, 2012, **3**, 375.
- 96 T. Ishikawa, N. Yamamoto, and K. Kuwata, *Chem. Phys. Lett.*, 2010, **500**, 149.
- 97 T. Tsukamoto, Y. Mochizuki, N. Watanabe, K. Fukuzawa, and T. Nakano, *Chem. Phys. Lett.*, 2012, **535**, 157.
- 98 T. Nagata, D. G. Fedorov, K. Kitaura, and M. S. Gordon, *J. Chem. Phys.*, 2009, **131**, 024101.
- 99 T. Nagata, D. G. Fedorov, and K. Kitaura, *Chem. Phys. Lett.*, 2009, **475**, 124.
- 100 T. Nagata, D. G. Fedorov, T. Sawada, K. Kitaura, and M. S. Gordon, *J. Chem. Phys.*, 2011, **134**, 034110.
- 101 T. Nagata, D. G. Fedorov, and K. Kitaura, *Theor. Chem. Acc.*, 2012, **131**, 1136.

- 102 T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, *J. Chem. Phys.*, 2011, **134**, 124115.
- 103 T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, *J. Chem. Phys.*, 2011, **135**, 044110.
- 104 T. Nagata, D. G. Fedorov, and K. Kitaura, *Chem. Phys. Lett.*, 2012, **544**, 87.
- 105 H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, *J. Chem. Phys.*, 2013, **138**, 164103.
- 106 P. Pyykko, *Chem. Rev.*, 1988, **88**, 563.
- 107 E. Miyoshi, H. Mori, R. Hirayama, Y. Osanai, T. Noro, H. Honda, and M. Klobukowski, *J. Chem. Phys.*, 2005, **122**, 074104.
- 108 H. Mori, K. Ueno-Noto, Y. Osanai, T. Noro, T. Fujiwara, M. Klobukowski, and E. Miyoshi, *Chem. Phys. Lett.*, 2009, **476**, 317.
- 109 T. Ishikawa, Y. Mochizuki, T. Nakano, S. Amari, H. Mori, H. Honda, T. Fujita, H. Tokiwa, S. Tanaka, and Y. Komeiji, *Chem. Phys. Lett.*, 2006, **427**, 159.
- 110 T. Fujiwara, H. Mori, Y. Mochizuki, Y. Osanai, and E. Miyoshi, *Chem. Phys. Lett.*, 2011, **510**, 261.
- 111 H. Mori, N. Hirayama, Y. Komeiji, and Y. Mochizuki, *Comput. Theor. Chem.*, 2012, **986**, 30.
- 112 Y. Komeiji, T. Ishida, D. G. Fedorov, and K. Kitaura, *J. Comput. Chem.*, 2007, **28**, 1750.
- 113 T. Nagata, D.G. Fedorov, T. Sawada, and K. Kitaura, *J. Phys. Chem. A*, 2012, **116**, 9088.
- 114 T. Ishikawa, R.R. Burri, Y.O. Kamatari, S. Sakuraba, N. Matubayasi, A. Kitao, and K. Kuwata, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3646.
- 115 J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999.
- 116 D. G. Fedorov, K. Kitaura, H. Li, J. H. Jensen, and M. S. Gordon, *J. Comput. Chem.*, 2006, **27**, 976.
- 117 H. Li, D. G. Fedorov, T. Nagata, K. Kitaura, J. H. Jensen, and M. S. Gordon, *J. Comput. Chem.*, 2010, **31**, 778.
- 118 T. Nagata, D. G. Fedorov, H. Li, and K. Kitaura, *J. Chem. Phys.*, 2012, **136**, 204112.
- 119 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2005, **123**, 134103.
- 120 T. Sawada, D. G. Fedorov, and K. Kitaura, *J. Am. Chem. Soc.*, 2010, **132**, 16862.
- 121 H. Watanabe, Y. Okiyama, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2010, **500**, 116.
- 122 D. G. Fedorov, T. Ishida, and K. Kitaura, *J. Phys. Chem. A*, 2005, **109**, 2638.

- 123 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2005, **122**, 054108.
- 124 S. R. Pruitt, D. G. Fedorov, K. Kitaura, and M. S. Gordon, *J. Chem. Theory Comput.*, 2010, **6**, 1.
- 125 S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, *J. Phys. Chem. A*, 2012, **116**, 4965.
- 126 M. C. Green, D. G. Fedorov, K. Kitaura, J. S. Francisco, and L. V Slipchenko, *J. Chem. Phys.*, 2013, **138**, 074111.
- 127 H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, *J. Chem. Phys.*, 2012, **137**, 044110.
- 128 R. Maezono, H. Watanabe, S. Tanaka, M. D. Towler, and R. J. Needs, *J. Phys. Soc. Japan*, 2007, **76**, 064301.
- 129 B. Auer, M. V Pak, and S. Hammes-Schiffer, *J. Phys. Chem. C*, 2010, **114**, 5582.
- 130 A. W. Lange and G. A. Voth, *J. Chem. Theory Comput.*, 2013, **9**, 4018.
- 131 K. Yasuda and D. Yamaki, *J. Chem. Phys.*, 2006, **125**, 154101.
- 132 Y. Mochizuki, K. Fukuzawa, A. Kato, S. Tanaka, K. Kitaura, and T. Nakano, *Chem. Phys. Lett.*, 2005, **410**, 247.
- 133 Y. Mochizuki, *Chem. Phys. Lett.*, 2005, **410**, 165.
- 134 R. Ahlrichs, P. Scharf, and C. Ehrhardt, *J. Chem. Phys.*, 1985, **82**, 890.
- 135 Y. Aoki and F. L. Gu, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7640.
- 136 M. Makowski, J. Korchowiec, F. L. Gu, and Y. Aoki, *J. Comput. Chem.*, 2010, **31**, 1733.
- 137 M. Miura and Y. Aoki, *Mol. Phys.*, 2010, **108**, 205.
- 138 G.-T. Yu, W. Chen, F. L. Gu, Y. Orimoto, and Y. Aoki, *Mol. Phys.*, 2009, **107**, 81.
- 139 W. Yang, *Phys. Rev. Lett.*, 1991, **66**, 1438.
- 140 M. Kobayashi and H. Nakai, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7629.
- 141 M. Kobayashi, T. Akama, and H. Nakai, *J. Chem. Phys.*, 2006, **125**, 204106.
- 142 M. Kobayashi, Y. Imamura, and H. Nakai, *J. Chem. Phys.*, 2007, **127**, 074103.
- 143 T. Yoshikawa, M. Kobayashi, and H. Nakai, *Theor. Chem. Acc.*, 2011, **130**, 411.
- 144 M. Kobayashi and H. Nakai, *J. Chem. Phys.*, 2008, **129**, 044103.
- 145 W. Li, Y. Guo, and S. Li, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7854.
- 146 S. Li, J. Ma, and Y. Jiang, *J. Comput. Chem.*, 2002, **23**, 237.
- 147 W. Li, P. Piecuch, J. R. Gour, and S. Li, *J. Chem. Phys.*, 2009, **131**, 114109.
- 148 W. Li and P. Piecuch, *J. Phys. Chem. A*, 2010, **114**, 8644.
- 149 M. Ziólkowski, B. Jansík, T. Kjaergaard, and P. Jørgensen, *J. Chem. Phys.*, 2010, **133**, 014107.
- 150 I.-M. Høyvik, K. Kristensen, B. Jansik, and P. Jørgensen, *J. Chem. Phys.*, 2012, **136**, 014105.

- 151 N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2012, **8**, 2669.
- 152 N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2011, **7**, 1336.
- 153 H. Stoll, *Chem. Phys. Lett.*, 1992, **191**, 548.
- 154 J. Friedrich and M. Dolg, *J. Chem. Phys.*, 2008, **129**, 244105.
- 155 H. Stoll, *Mol. Phys.*, 2010, **108**, 243.
- 156 J. Zhang and M. Dolg, *J. Chem. Theory Comput.*, 2013, **9**, 2992.
- 157 R. A. Mata and H. Stoll, *J. Chem. Phys.*, 2011, **134**, 034122.
- 158 J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.*, 1992, **96**, 135.
- 159 Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura, and T. Nakano, *Chem. Phys. Lett.*, 2005, **406**, 283.
- 160 M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, *Chem. Phys. Lett.*, 1994, **219**, 21.
- 161 Y. Mochizuki, K. Tanaka, K. Yamashita, T. Ishikawa, T. Nakano, S. Amari, K. Segawa, T. Murase, H. Tokiwa, and M. Sakurai, *Theor. Chem. Acc.*, 2006, **117**, 541.
- 162 Y. Mochizuki, T. Nakano, S. Amari, T. Ishikawa, K. Tanaka, M. Sakurai, and S. Tanaka, *Chem. Phys. Lett.*, 2007, **433**, 360.
- 163 Y. Mochizuki, Y. Komeiji, T. Ishikawa, T. Nakano, and H. Yamataka, *Chem. Phys. Lett.*, 2007, **437**, 66.
- 164 Y. Mochizuki and K. Tanaka, *Chem. Phys. Lett.*, 2007, **443**, 389.
- 165 Y. Mochizuki, *Chem. Phys. Lett.*, 2008, **453**, 109.
- 166 Y. Mochizuki, *Chem. Phys. Lett.*, 2009, **472**, 143.
- 167 N. Taguchi, Y. Mochizuki, T. Nakano, S. Amari, K. Fukuzawa, T. Ishikawa, M. Sakurai, and S. Tanaka, *J. Phys. Chem. B*, 2009, **113**, 1153.
- 168 N. Taguchi, Y. Mochizuki, and T. Nakano, *Chem. Phys. Lett.*, 2011, **504**, 76.
- 169 A. Tagami, N. Ishibashi, D. Kato, N. Taguchi, Y. Mochizuki, H. Watanabe, M. Ito, and S. Tanaka, *Chem. Phys. Lett.*, 2009, **472**, 118.
- 170 M. Chiba, D. G. Fedorov, and K. Kitaura, *Chem. Phys. Lett.*, 2007, **444**, 346.
- 171 M. Chiba, D. G. Fedorov, and K. Kitaura, *J. Chem. Phys.*, 2007, **127**, 104108.
- 172 M. Chiba and T. Koido, *J. Chem. Phys.*, 2010, **133**, 044113.
- 173 M. Chiba, D. G. Fedorov, and K. Kitaura, *J. Comput. Chem.*, 2008, **29**, 2667.
- 174 M. Chiba, D. G. Fedorov, T. Nagata, and K. Kitaura, *Chem. Phys. Lett.*, 2009, **474**, 227.
- 175 T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa, and S. Sekiguchi, *J. Comput. Chem.*, 2010, **31**, 447.

- 176 J. Olsen and P. Jørgensen, *J. Chem. Phys.*, 1985, **82**, 3235.
- 177 Y. Mochizuki, T. Ishikawa, K. Tanaka, H. Tokiwa, T. Nakano, and S. Tanaka, *Chem. Phys. Lett.*, 2006, **418**, 418.
- 178 T. Helgaker, M. Jaszuński, and K. Ruud, *Chem. Rev.*, 1999, **99**, 293.
- 179 Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, *Chem. Phys. Lett.*, 2007, **445**, 331.
- 180 Q. Gao, S. Yokojima, D. G. Fedorov, K. Kitaura, M. Sakurai, and S. Nakamura, *J. Chem. Theory Comput.*, 2010, **6**, 1428.
- 181 Q. Gao, S. Yokojima, D.G. Fedorov, K. Kitaura, M. Sakurai, and S. Nakamura, *Chem. Phys. Lett.*, 2014, **593**, 165.
- 182 H. Sekino, Y. Sengoku, and N. Matsumura, *Comput. Lett.*, 2007, **3**, 423.
- 183 Y. Komeiji, Y. Mochizuki, T. Nakano, and D. G. Fedorov, *J. Mol. Struct. THEOCHEM*, 2009, **898**, 2.
- 184 Y. Komeiji, M. Uebayasi, R. Takata, A. Shimizu, K. Itsukashi, and M. Taiji, *J. Comput. Chem.*, 1997, **18**, 1546.
- 185 Y. Komeiji, Y. Inadomi, and T. Nakano, *Comput. Biol. Chem.*, 2004, **28**, 155.
- 186 Y. Komeiji, T. Ishikawa, Y. Mochizuki, H. Yamataka, and T. Nakano, *J. Comput. Chem.*, 2009, **30**, 40.
- 187 Y. Komeiji, T. Fujiwara, Y. Okiyama, and Y. Mochizuki, *Chem-Bio Informatics J.*, 2013, **13**, 45.
- 188 T. Ishimoto, H. Tokiwa, H. Teramae, and U. Nagashima, *Chem. Phys. Lett.*, 2004, **387**, 460.
- 189 T. Ishimoto, H. Tokiwa, H. Teramae, and U. Nagashima, *J. Chem. Phys.*, 2005, **122**, 094905.
- 190 K. Tamura, T. Watanabe, T. Ishimoto, and U. Nagashima, *Bull. Chem. Soc. Jpn.*, 2008, **81**, 110.
- 191 T. Fujita, H. Watanabe, and S. Tanaka, *J. Phys. Soc. Japan*, 2009, **78**, 104723.
- 192 T. Fujiwara, Y. Mochizuki, Y. Komeiji, Y. Okiyama, H. Mori, T. Nakano, and E. Miyoshi, *Chem. Phys. Lett.*, 2010, **490**, 41.
- 193 K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, *J. Chem. Theory Comput.*, 2012, **8**, 5008.
- 194 T. Okamoto, T. Ishikawa, Y. Koyano, N. Yamamoto, K. Kuwata, and M. Nagaoka, *Bull. Chem. Soc. Jpn.*, 2013, **222**, 210.
- 195 Y. Sakai, E. Miyoshi, M. Klobukowski, and S. Huzinaga, *J. Comput. Chem.*, 1987, **8**, 226.

- 196 Y. Osanai, M. S. Mon, T. Noro, H. Mori, H. Nakashima, M. Klobukowski, and E. Miyoshi, *Chem. Phys. Lett.*, 2008, **452**, 210.
- 197 Y. Osanai, E. Soejima, T. Noro, H. Mori, M. S. Mon, M. Klobukowski, and E. Miyoshi, *Chem. Phys. Lett.*, 2008, **463**, 230.
- 198 M. Sprik and G. Ciccotti, *J. Chem. Phys.*, 1998, **109**, 7737.
- 199 Y. Komeiji, *Chem-Bio Informatics J.*, 2007, **7**, 12.
- 200 M. Sato, H. Yamataka, Y. Komeiji, Y. Mochizuki, and T. Nakano, *Chem. Eur. J.*, 2010, **16**, 6430.
- 201 M. Sato, H. Yamataka, Y. Komeiji, and Y. Mochizuki, *Chem. Eur. J.*, 2012, **18**, 9714.
- 202 D. Marx and M. Parrinello, *J. Chem. Phys.*, 1996, **104**, 4077.
- 203 M. Sato, H. Yamataka, Y. Komeiji, Y. Mochizuki, T. Ishikawa, and T. Nakano, *J. Am. Chem. Soc.*, 2008, **130**, 2396.
- 204 T. Fujiwara, H. Mori, Y. Mochizuki, H. Tatewaki, and E. Miyoshi, *J. Mol. Struct. THEOCHEM*, 2010, **949**, 28.
- 205 N. Beebe and J. Linderberg, *Int. J. Quantum Chem.*, 1977, **12**, 683.
- 206 T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa, and S. Sekiguchi, *ACM/IEEE SC 2005 Conf.*, 2005, 10.
- 207 Y. Alexeev, A. Mahajan, S. Leyffer, G. Fletcher, and D.G. Fedorov, *Proc. Supercomputing 2012*, IEEE Computer Society, Salt Lake City, 2012.
- 208 G. D. Fletcher, D. G. Fedorov, S. R. Pruitt, T. L. Windus, and M. S. Gordon, *J. Chem. Theory Comput.*, 2012, **8**, 75.
- 209 M. Katouda, T. Nakajima, and S. Nagase, in *Proceedings of JSST 2012*, 2012, pp. 338.
- 210 S. K. Talamudupula, M. Sosonkina, A. Gaenko, and M. W. Schmidt, *Procedia Comput. Sci.*, 2012, **9**, 489.
- 211 W. A. de Jong, E. Bylaska, N. Govind, C. L. Janssen, K. Kowalski, T. Müller, I. M. B. Nielsen, H. J. J. van Dam, V. Veryazov, and R. Lindh, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6896.
- 212 K. Fukuzawa, Y. Komeiji, Y. Mochizuki, A. Kato, T. Nakano, and S. Tanaka, *J. Comput. Chem.*, 2006, **27**, 948.
- 213 ABINIT-MP and its pre/post processing program, BioStation Viewer, is available at: <http://www.ciss.iis.u-tokyo.ac.jp/english/dl/>
- 214 J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic Press, London, 2006.

- 215 K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, in *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, eds. K. Kitaura and D. G. Fedorov, CRC Press, Boca Raton, 2009, p. 133.
- 216 T. Watanabe, Y. Inadomi, H. Umeda, K. Fukuzawa, S. Tanaka, T. Nakano, and U. Nagashima, *J. Comput. Theor. Nanosci.*, 2009, **6**, 1328.
- 217 H. Umeda, Y. Inadomi, T. Watanabe, T. Yagi, T. Ishimoto, T. Ikegami, H. Tadano, T. Sakurai, and U. Nagashima, *J. Comput. Chem.*, 2010, **31**, 2381.
- 218 T. Sakurai and H. Sugiura, *J. Comput. Appl. Math.*, 2003, **159**, 119.
- 219 T. Sakurai, H. Tadano, T. Ikegami, and U. Nagashima, *Taiwan. J. Math.*, 2010, **14**, 855.
- 220 S. Tsuneyuki, T. Kobori, K. Akagi, K. Sodeyama, K. Terakura, and H. Fukuyama, *Chem. Phys. Lett.*, 2009, **476**, 104.
- 221 T. Kobori, K. Sodeyama, T. Otsuka, Y. Tateyama, and S. Tsuneyuki, *J. Chem. Phys.*, 2013, **139**, 094113.
- 222 H. Nishioka and K. Ando, *J. Chem. Phys.*, 2011, **134**, 204109.
- 223 H. Kitoh-Nishioka and K. Ando, *J. Phys. Chem. B*, 2012, **116**, 12933.
- 224 D. G. Fedorov and K. Kitaura, *J. Comput. Chem.*, 2007, **28**, 222.
- 225 K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma, and T. Nakano, *J. Comput. Chem.*, 2005, **26**, 1.
- 226 S. Ross, P. McCaffery, U. C. Drager, and L. M. De Luca, *Physiol. Rev.*, 2000, **80**, 1021.
- 227 S. Miles, B. Dezube, J. Lee, S. Krown, M. Fletcher, M. Saville, L. Kaplan, J. Groopman, D. Scadden, T. Cooley, J. Von Roenn, and A. Friedman-Kien, *Aids*, 2002, **16**, 421.
- 228 D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R. M. Evans, *Cell*, 1995, **83**, 835.
- 229 L. Nagy and J. W. R. Schwabe, *Trends Biochem. Sci.*, 2004, **29**, 317.
- 230 S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095.
- 231 Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 9793.
- 232 M. Parry, P. Andralojc, R. Mitchell, P. Madgwick, and A. Keys, *J. Exp. Bot.*, 2003, **54**, 1321.
- 233 T. John Andrews and S. M. Whitney, *Arch. Biochem. Biophys.*, 2003, **414**, 159.
- 234 R. Spreitzer, *Photosynth. Res.*, 1999, **60**, 29.
- 235 D. B. Jordan and W. L. Ogren, *Nature*, 1981, **291**, 513.

- 236 K. Uemura, Anwaruzzaman, S. Miyachi, and A. Yokota, *Biochem. Biophys. Res. Commun.*, 1997, **233**, 568.
- 237 R. M. de Jong and B. W. Dijkstra, *Curr. Opin. Struct. Biol.*, 2003, **13**, 722.
- 238 D. B. Janssen, I. J. T. Dinkla, G. J. Poelarends, and P. Terpstra, *Environ. Microbiol.*, 2005, **7**, 1868.
- 239 T. Kurihara and N. Esaki, *Chem. Rec.*, 2008, **8**, 67.
- 240 R.G. Webster, W.J. Bean, O.T. Gorman, T.M. Chambers and Y. Kawaoka, *Microbiol. Rev.*, 1992, **56**, 152.
- 241 D. Fleury, R.S. Daniels, J.J. Skehel, M. Knossow and T. Bizebard, *Proteins*, 2000, **40**, 572.
- 242 K. Nakajima, E. Nobusawa, K. Tonegawa, S. Nakajima, *J. Virol.* 2003, **77**, 10088.
- 243 K. Nakajima, E. Nobusawa, A. Nagy, S. Nakajima, *J. Virol.* 2005, **79**, 6472.
- 244 T. Iwata, K. Fukuzawa, K. Nakajima, S. Aida-Hyugaji, Y. Mochizuki, H. Watanabe, and S. Tanaka, *Comput. Biol. Chem.*, 2008, **32**, 198.
- 245 K. Fukuzawa, K. Omagari, K. Nakajima, E. Nobusawa, and S. Tanaka, *Protein Pept. Lett.*, 2011, **18**, 530.
- 246 A. Yoshioka, K. Takematsu, I. Kurisaki, K. Fukuzawa, Y. Mochizuki, T. Nakano, E. Nobusawa, K. Nakajima, and S. Tanaka, *Theor. Chem. Acc.*, 2011, **130**, 1197.
- 247 H. Hieronymus and P. A. Silver, *Genes Dev.*, 2004, **18**, 2845.
- 248 J. Pérez-Cañadillas and G. Varani, *Curr. Opin. Struct. Biol.*, 2001, **11**, 53.
- 249 S. D. Auweter, F. C. Oberstrass, and F. H.-T. Allain, *Nucleic Acids Res.*, 2006, **34**, 4943.
- 250 S. Adinolfi, C. Bagni, M. Morelli, F. Fraternali, G. Musco, and A. Pastore, *Biopolymers*, 1999, **51**, 153.
- 251 R. Valverde, L. Edwards, and L. Regan, *FEBS J.*, 2008, **275**, 2712.
- 252 H. Lewis, K. Musunuru, K. Jensen, C. Edo, H. Chen, R. B. Darnell, and S. K. Burley, *Cell*, 2000, **100**, 323.
- 253 I. Kurisaki, K. Fukuzawa, T. Nakano, Y. Mochizuki, H. Watanabe, and S. Tanaka, *J. Mol. Struct. THEOCHEM*, 2010, **962**, 45.
- 254 B. Paulus, P. Fulde, and H. Stoll, *Phys. Rev. B*, 1995, **51**, 10572.
- 255 H. Stoll, B. Paulus, and P. Fulde, *Chem. Phys. Lett.*, 2009, **469**, 90.
- 256 M. A. Collins and V. A. Deev, *J. Chem. Phys.*, 2006, **125**, 104104.
- 257 D. G. Fedorov, J. H. Jensen, R. C. Deka, and K. Kitaura, *J. Phys. Chem. A*, 2008, **112**, 11808.

- 258 D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, *Chem. Phys. Lett.*, 2009, **477**, 169.
- 259 A. Vallee, V. Humblot, and C.-M. Pradier, *Acc. Chem. Res.*, 2010, **43**, 1297.
- 260 K.-I. Sano and K. Shiba, *J. Am. Chem. Soc.*, 2003, **125**, 14234.
- 261 T. Ishikawa, Y. Mochizuki, K. Imamura, T. Nakano, H. Mori, H. Tokiwa, K. Tanaka, E. Miyoshi, and S. Tanaka, *Chem. Phys. Lett.*, 2006, **430**, 361.
- 262 H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii, and K. Kitaura, *J. Phys. Chem. A*, 2008, **112**, 10887.
- 263 P. Avramov and D. Fedorov, *J. Phys. Chem. Lett.*, 2012, **3**, 2003.
- 264 E. I. Izgorodina, J. Rigby, and D. R. MacFarlane, *Chem. Commun.*, 2012, **48**, 1493.
- 265 P. Carlson, S. Bose, D. Armstrong, T. Hawkins, M. S. Gordon, and J. Petrich, *J. Phys. Chem. B*, 2011, **116**, 503.
- 266 A. Devarajan, S. Markutsya, M. H. Lamm, X. Cheng, J. C. Smith, J. Y. Baluyut, Y. Kholod, M. S. Gordon, and T. L. Windus, *J. Phys. Chem. B*, 2013, **117**, 10430.
- 267 L. Roskop, D. G. Fedorov, and M. S. Gordon, *Mol. Phys.*, 2013, **111**, 1622.
- 268 L. Kong, F. A. Bischoff, and E. F. Valeev, *Chem. Rev.*, 2012, **112**, 75.
- 269 C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, *Chem. Rev.*, 2012, **112**, 4.
- 270 S. Ten-no, *Theor. Chem. Acc.*, 2012, **131**, 1070.
- 271 D. G. Fedorov, L. V Slipchenko, and K. Kitaura, *J. Phys. Chem. A*, 2010, **114**, 8742.
- 272 J. Olsen, P.A. Malmqvist, B. O. Roos, R. Lindh, and P.-O. Widmark, *Chem. Phys. Lett.*, 1987, **133**, 91.
- 273 R. Lindh, J. Olsen, and B. O. Roos, *Chem. Phys. Lett.*, 1988, **148**, 276.
- 274 T. Kinoshita, O. Hino, and R. J. Bartlett, *J. Chem. Phys.*, 2003, **119**, 7756.
- 275 F. Bell, D. S. Lambrecht, and M. Head-Gordon, *Mol. Phys.*, 2010, **108**, 2759.
- 276 J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, *J. Chem. Phys.*, 2011, **134**, 044123.
- 277 D. Kats and F. R. Manby, *J. Chem. Phys.*, 2013, **138**, 144101.
- 278 W. Uemura and O. Sugino, *Phys. Rev. Lett.*, 2012, **109**, 253001.
- 279 C. A. Schwerdtfeger and D. A. Mazziotti, *J. Chem. Phys.*, 2012, **137**, 244103.
- 280 G. Moritz and M. Reiher, *J. Chem. Phys.*, 2007, **126**, 244109.
- 281 K. H. Marti and M. Reiher, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6750.
- 282 K. H. Marti, B. Bauer, M. Reiher, M. Troyer, and F. Verstraete, *New J. Phys.*, 2010, **12**, 103008.
- 283 T. Yanai and G. K.-L. Chan, *J. Chem. Phys.*, 2007, **127**, 104107.

- 284 Y. Kurashige and T. Yanai, *J. Chem. Phys.*, 2009, **130**, 234114.
- 285 S. Tanaka, *J. Phys. Soc. Japan*, 2013, **82**, 075001.
- 286 U. Benedikt, H. Auer, M. Espig, W. Hackbusch, and A. A. Auer, *Mol. Phys.*, 2013, **111**, 2398.
- 287 Y. Kurashige, G. K.-L. Chan, and T. Yanai, *Nat. Chem.*, 2013, **5**, 660.
- 288 D. Kats and F. R. Manby, *J. Chem. Phys.*, 2013, **139**, 021102.
- 289 L. M. J. Huntington and M. Nooijen, *J. Chem. Phys.*, 2010, **133**, 184109.
- 290 K. Yasuda, *J. Comput. Chem.*, 2008, **29**, 334.
- 291 A. Asadchev and M. S. Gordon, *J. Chem. Theory Comput.*, 2012, **8**, 4166.
- 292 Y. Miao and K. M. Merz, *J. Chem. Theory Comput.*, 2013, **9**, 965.
- 293 R. Olivares-Amaya, M. A. Watson, R. G. Edgar, L. Vogt, Y. Shao, and A. Aspuru-Guzik, *J. Chem. Theory Comput.*, 2010, **6**, 135.
- 294 A. E. DePrince and J. R. Hammond, *J. Chem. Theory Comput.*, 2011, **7**, 1287.
- 295 W. Ma, S. Krishnamoorthy, O. Villa, and K. Kowalski, *J. Chem. Theory Comput.*, 2011, **7**, 1316.
- 296 C. M. Isborn, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, *J. Chem. Theory Comput.*, 2011, **7**, 1814.
- 297 H. J. Kulik, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, *J. Phys. Chem. B*, 2012, **116**, 12501.
- 298 E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, *J. Chem. Phys.*, 2012, **137**, 044103.
- 299 R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, *J. Chem. Phys.*, 2012, **137**, 224106.
- 300 R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, *J. Chem. Phys.*, 2013, **138**, 194107.
- 301 R. Feynman and A. Hibbs, *Quantum mechanics and path integrals*, McGraw-Hill Companies, Inc., New York, 1965.
- 302 T. Fujita, S. Tanaka, T. Fujiwara, M.-A. Kusa, Y. Mochizuki, and M. Shiga, *Comput. Theor. Chem.*, 2012, **997**, 7.
- 303 H. Watanabe, S. Tanaka, N. Okimoto, A. Hasegawa, M. Taiji, Y. Tanida, T. Mitsui, M. Katsuyama, and H. Fujitani, *Chem-Bio Informatics J.*, 2010, **10**, 32.
- 304 O. Ichihara, J. Barker, R.J. Law and M. Whittaker, *Mol. Inf.*, 2011, **30**, 298.
- 305 K. Fujimura and Y. Sasabuchi, *Chem. Med. Chem.*, 2010, **5**, 1254.
- 306 M.P. Mazanetz, O. Ichihara, R.J. Law, and M. Whittaker, *J. Cheminf.*, 2011, **3**, 2.
- 307 H. Mori and K. Ueno-Noto, *J. Phys. Chem. B*, 2011, **115**, 4774.

- 308 K. Ohno, K. Mori, M. Orita, and M. Takeuchi, *Curr. Med. Chem.*, 2011, **18**, 220.
- 309 Y. Alexeev, M.P. Mazanetz, O. Ichihara, and D.G. Fedorov, *Curr. Top. Med. Chem.*, 2012, **12**, 2013.
- 310 K. Ueno-Noto, S. Ise, and K. Takano, *J. Theor. Comput. Chem.*, 2013, **12**, 1350060.
- 311 T. Yoshida, Y. Munei, S. Hitaoka, and H. Chuman, *J. Chem. Inf. Model.*, 2010, **50**, 850.
- 312 S. Hitaoka, M. Harada, T. Yoshida, and H. Chuman, *J. Chem. Inf. Model.*, 2010, **50**, 1796.
- 313 S. Hitaoka, H. Matoba, M. Harada, T. Yoshida, D. Tsuji, T. Hirokawa, K. Itoh, and H. Chuman, *J. Chem. Inf. Model.*, 2011, **51**, 2706.
- 314 D.G. Fedorov and K. Kitaura, *J. Comput. Chem.*, 2007, **28**, 222.
- 315 A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612.
- 316 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
- 317 J. Åqvist and A. Warshel, *Chem. Rev.*, 1993, **93**, 2523.
- 318 Y. Okiyama, H. Watanabe, K. Fukuzawa, T. Nakano, Y. Mochizuki, T. Ishikawa, S. Tanaka, and K. Ebina, *Chem. Phys. Lett.*, 2007, **449**, 329.
- 319 Y. Okiyama, H. Watanabe, K. Fukuzawa, T. Nakano, Y. Mochizuki, T. Ishikawa, K. Ebina, and S. Tanaka, *Chem. Phys. Lett.*, 2009, **467**, 417.
- 320 L. Chang, T. Ishikawa, K. Kuwata, and S. Takada, *J. Comput. Chem.*, 2013, **34**, 1251.
- 321 V. May and O. Kuhn, *Charge and Energy Transfer Dynamics in Molecular Systems*, Wiley-VCH, Berlin, 2000.
- 322 S. Tanaka and E. B. Starikov, *Phys. Rev. E*, 2010, **81**, 027101.
- 323 S. Tanaka, *Chem. Phys. Lett.*, 2011, **508**, 139.
- 324 Y. Suzuki and S. Tanaka, *Phys. Rev. E*, 2012, **86**, 021914.

Figure Captions

Fig. 1. A sp^3 carbon atom C_α located at the interface between the fragments I and $I+1$ is selected as a bond detached atom (BDA) at which the MOs are distributed between the two fragments in the case of polypeptide.

Fig. 2. Visualized IFIE results of HIV-1 protease-lopinavir at the FMO-CCSD(T)/6-31G level. Lopinavir is shown by yellow balls. Red and blue colors for amino acid residues mean stabilization and destabilization, respectively. (Reproduced [74] with permission from Springer.)

Fig. 3. Schematic presentation of FMO-MD explained by Fortran-like pseudo codes (lower) and by a model molecular system (upper) where black circles represent atomic nuclei and the grey shade the electron cloud. At each step of MD time integration (velocity-Verlet), subroutine "CALFOR" is called to execute dynamic fragmentation (DYNFRAG) and FMO calculation (CALFMO) for the instantaneous molecular coordinates " \mathbf{r} ", then the calculated force " \mathbf{F} ", energy " E ", atomic charge " q ", *etc.* are returned, and \mathbf{F} is used for further time integration.

Fig. 4. DF algorithm exemplified by a configuration of $C_2H_5CO_2H$ and CH_3OH in water solvent.

Fig. 5. Timing data (in minutes) of FMO4-MP2(CD)/6-31G jobs for HIV-1 protease-lopinavir complex by using 600 and 6000 nodes on the K-computer. Breakdowns for HF-SCF and MP2 steps are shown.

Fig. 6. IFIE map for CRP-DNA complex. Red and blue spots refer to the attractive and repulsive interactions between the fragments, respectively.

Fig. 7. VISCANA for complexes of human estrogen receptor α and ligand molecules. Attractive and repulsive interactions between the ligands (ordinate) and each amino acid residue (abscissa) are depicted by red and blue colors, respectively. A clustering concerning the dissimilarity of interaction patterns among the ligand molecules has been carried out.

Fig. 8. (Upper) Calculated result for the SCIFIEs between Glu23 and other fragments in

crambin. The values obtained in the Percus-Yevick (PY) scheme (red bars) are compared with the bare IFIEs (blue bars). The interactions with Thr1, Arg10, Arg17, Asp43 and Asn46 are denoted by the numbers 1, 10, 17, 39 and 42 on the abscissa, respectively. (Lower) Structure of crambin (PDB entry: 1EJG) employed in this analysis is also shown. Locations of six charged residues are designated in color; Thr1 (blue), Arg10 (cyan), Arg17 (yellow), Glu23 (magenta), Asp43 (green) and Asn46 (red). (Reproduced [26] with permission from Elsevier.)

Fig. 9. Visualization of the CT interactions from Glu353 to the ligand and from the ligand to His524 in a complex of estrogen receptor (ER) and estradiol (EST).

Fig. 10. CH/ π interaction between Phe (upper) and ligand molecule (lower) described by FILM in an ER-EST complex.

Fig. 11. Visualization of the IFIEs between the EST and each amino acid residue fragment of the ER calculated at the MP2/6-31G* level. Around the domain indicated in yellow (EST), the red and blue fragments illustrate stabilized and destabilized interactions, respectively.

Fig. 12. Hydrophobic residues that showed IFIE values with the EST stronger than 1 kcal/mol: left, at the HF/6-31G* level; right, at the MP2/6-31G* level. The ball-and-stick representation refers to EST and a water molecule; the stick representation refers to hydrophobic residues with significant IFIE values. The line representation is the C $_{\alpha}$ backbone.

Fig. 13. FMO4-IFIEs between EST and each amino acid of the ER at the MP2/6-31G level: (a) FMO4-IFIEs of EST by the conventional main-chain fragmentation, (b) FMO4-IFIEs of EST(1) by the main/side chain fragmentation, and (c) FMO4-IFIEs of EST(2) by the main/side chain fragmentation.

Fig. 14. Ribbon display of the X-ray crystal structure (PDB entry: 1FM9) of the hRXR α LBD (green) complexed with 9cRA (purple) and SRC1 peptide (blue). The position of AF2C (red) in H12 is also displayed. (Reproduced [32] with permission from American Chemical Society.)

Fig. 15. P1 and P2 anion binding sites of the higher plant Rubiscos. X-ray crystal

structure of spinach Rubisco with transition-state analogue 2CABP (PDB ID: 8RUC) and that of tobacco Rubisco with inorganic phosphate anions (PDB ID: 1EJ7) are shown. In the structure of 8RUC, phosphate groups of 2CABP are situated at P1 α and P2 α anion binding sites, and the loop6 is at the closed position (red curve). In the structure of 1EJ7, phosphate anions are situated at the P1 β and P2 β anion binding sites, and the loop6 is at the open position (blue curve). Position of Lys18 which strongly interacts with P1 anion and is mutated to Ile in *Galdieria partita* Rubisco is also shown.

Fig. 16. (Left) IFIEs of the residues in HA with the whole Fab fragment of antibody (colored in yellow) calculated with the FMO-MP2/6-31G* method. Red and blue colors refer to the interaction energies of stabilization (negative) and destabilization (positive), respectively. (Right) IFIE sums between the residues in the antigenic region E and the whole Fab fragment of antibody. (A) Charged residues. (B) Hydrophobic residues. (C) Polar residues. The green, red, and blue bars represent that they are located at the positive (allowed), negative (prohibited), and unspecified (not measured) residue sites in the hemadsorption experiments, respectively. The numbers below the bars denote the years of the emergence of the mutation. (Reproduced [58] with permission from American Chemical Society.)

Fig. 17. (A) Visualization of IFIEs between HA trimer (I, II, III) and Fab dimer (I, II) calculated at the FMO-MP2.5/6-31G level. The color represents the sign and strength of the interactions between each residue in the HA trimer and the Fab dimer. For the Fab domain indicated in yellow, the red and blue fragments refer to stabilized and destabilized interactions, respectively, and the deepness of the hue indicates the strength of the interaction. (B) Visualization of antigenic regions A (pink) and B (light blue) by sphere representation. The illustration was created with BioStation Viewer. (Reproduced [71] with permission from Elsevier.)

Fig. 18. Primary and secondary structures of NOVA-2 third KH domain, tertiary structure of NOVA-2 third KH domain–RNA complex and the bound RNA sequence. (A) Primary and secondary structures of NOVA-2 containing third KH domain (residues #10–72) and C-terminal region (residues #73–90). The numbers in the parenthesis are the original residue numbers of full sequence. The locations of GXXG motif (residues #22–25) and variable loop (residues #43–52) are indicated by green bold letters. Orange cylinders and green arrows represent α -helix and β -sheet, respectively. (B) The RNA sequence in the crystal structure of NOVA–RNA complex (PDB ID: 1EC6). The digits on

the symbols of RNA base represent the position in the RNA. The specifically recognized sequence is indicated by underscore. (C) Tertiary structure of NOVA-2 third KH domain and the bound RNA (PDB ID: 1EC6). Gly-X-X-Gly motif and variable loop, RNA binding interface of the KH domain, and specifically recognized bases are indicated by green surface, red surface and blue stick, respectively. (Reproduced [253] with permission from Elsevier.)

Fig. 19. Changes of intra-molecular IFIEs upon the complex formation of NOVA and RNA. The minimum distance matrix (A) and difference (complex– free) IFIE map (B). Each square frame along the diagonal line represents each α -helix forming region, and each rectangle frame represents β -sheet forming region. Secondary structure is indicated above and on the right of map by orange circle and green triangle dots for α -helix and β -strand, respectively. The closeness of distance and the strength of IFIE are represented by the deepness of hue. For the IFIE map, an upper triangle is used to show the plots of negative IFIE values, whereas a lower triangle the plots of positive IFIE values. (Reproduced [253] with permission from Elsevier.)

Fig. 20. Visualized IFIE results of silica cluster - RKLPGA peptide complex at FMO4-MP2(CD)/6-31G level. In the left figure, the contributions from silica are summed for peptide (*vice versa* for the right figure). Red and blue colors mean stabilization and destabilization, respectively. Hydrating water molecules are partly visible as well. (Reproduced [45] with permission from Elsevier.)

Fig. 21. Structures of *cis*- and *trans*-platins and schematic representations of their DNA complexes. (Reproduced [111] with permission from Elsevier.)

Fig. 22. (Left) Time evolution of natural charge on the Pt, NH₃, and Cl sites in the *cis*- and *trans*-platin. Solid and dotted lines indicate *cis*- and *trans*- isomers, respectively. (Right) Time evolution of Pt-Cl bond lengths. (Reproduced [111] with permission from Elsevier.)

TablesTable 1. Modes of the DF/SF algorithm specified by combination of DF and SF modes.

DF modes:

DF0: No dynamic fragmentation.

The default fragmentation set s is used throughout the simulation.

DF1: Basic dynamic fragmentation.

Each molecule or coordinate complex constitutes a fragment.

DF2: Advanced dynamic fragmentation.

The fragments made by DF1 are further merged if they are hydrogen bonded.

SF modes:

SF0: No static fragments.

All the atoms are regarded as dynamic.

SF1: Complete fixation of static fragments.

All the atoms including hydrogen are assigned to the same static fragments.

SF2: Marginal fixation of static fragments.

Only heavy atoms within the static fragments are fixed to the same fragment.

Hydrogen atoms are treated dynamic (variable).Table 2. List of possible fragment species for FMO-MD of CH₃OH in water.

<u>Name</u>	<u>Atomic composition</u>	<u>Formal charge</u>
CH ₃ O [·]	C O H H H	-1
CH ₃ OH	C O H H H H	0
CH ₃ OH ₂ ⁺	C O H H H H H	+1
OH [·]	O H	-1
H ₂ O	O H H	0
H ₃ O ⁺	O H H H	+1
H ₃ O ₂ [·]	O O H H H	-1
H ₄ O ₂	O O H H H H	0
H ₅ O ₂ ⁺	O O H H H H H	+1

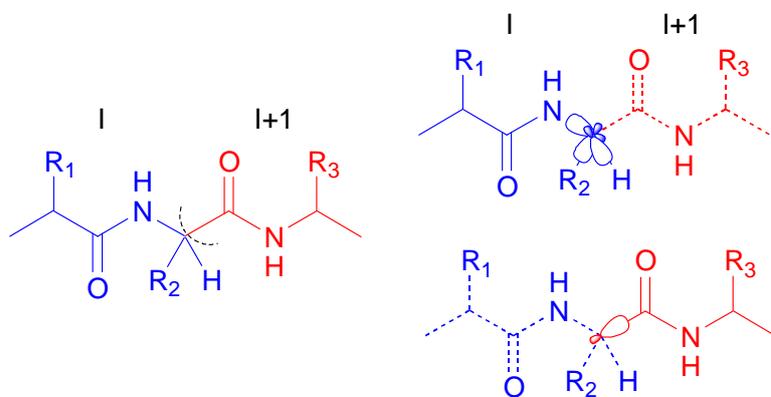
Figures

Fig. 1

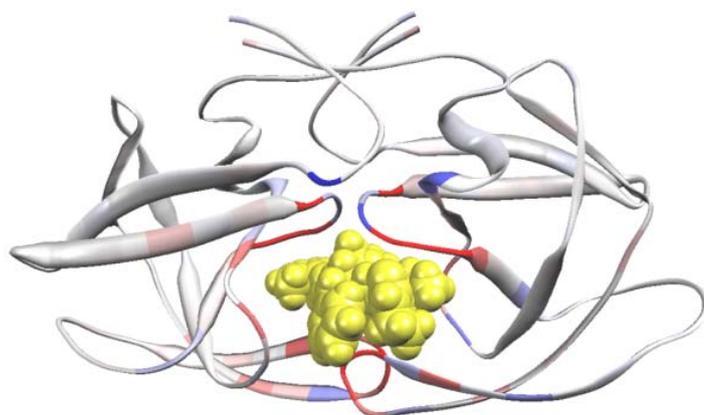


Fig. 2

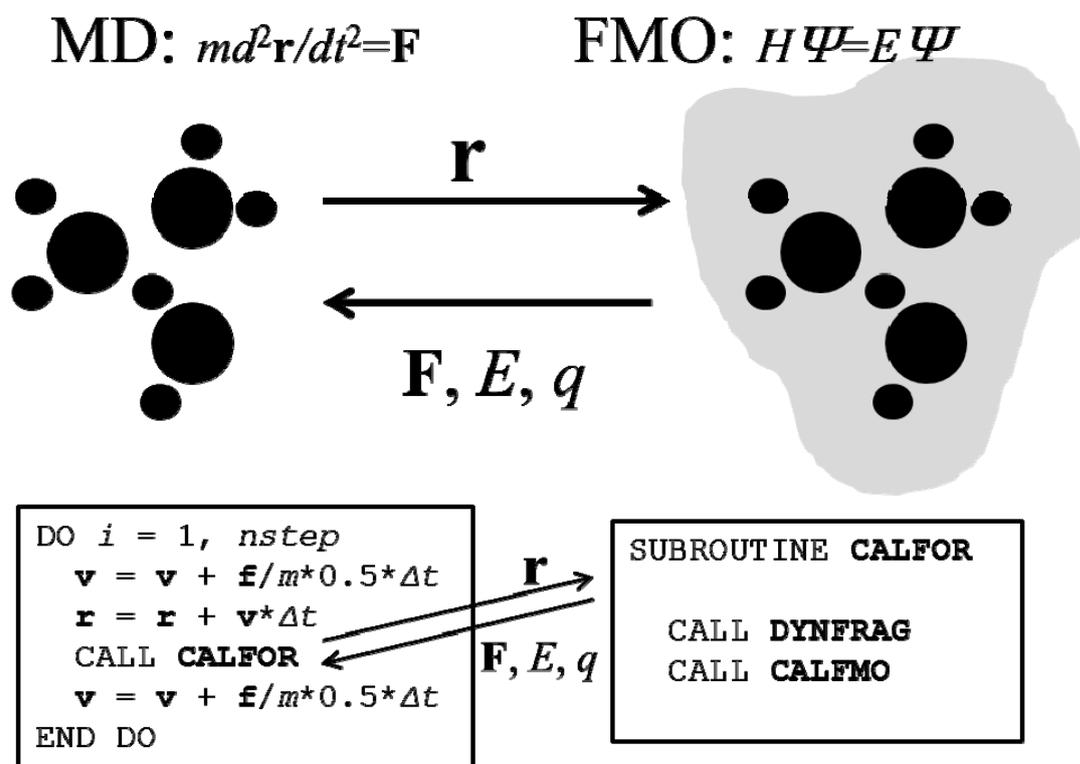


Fig. 3

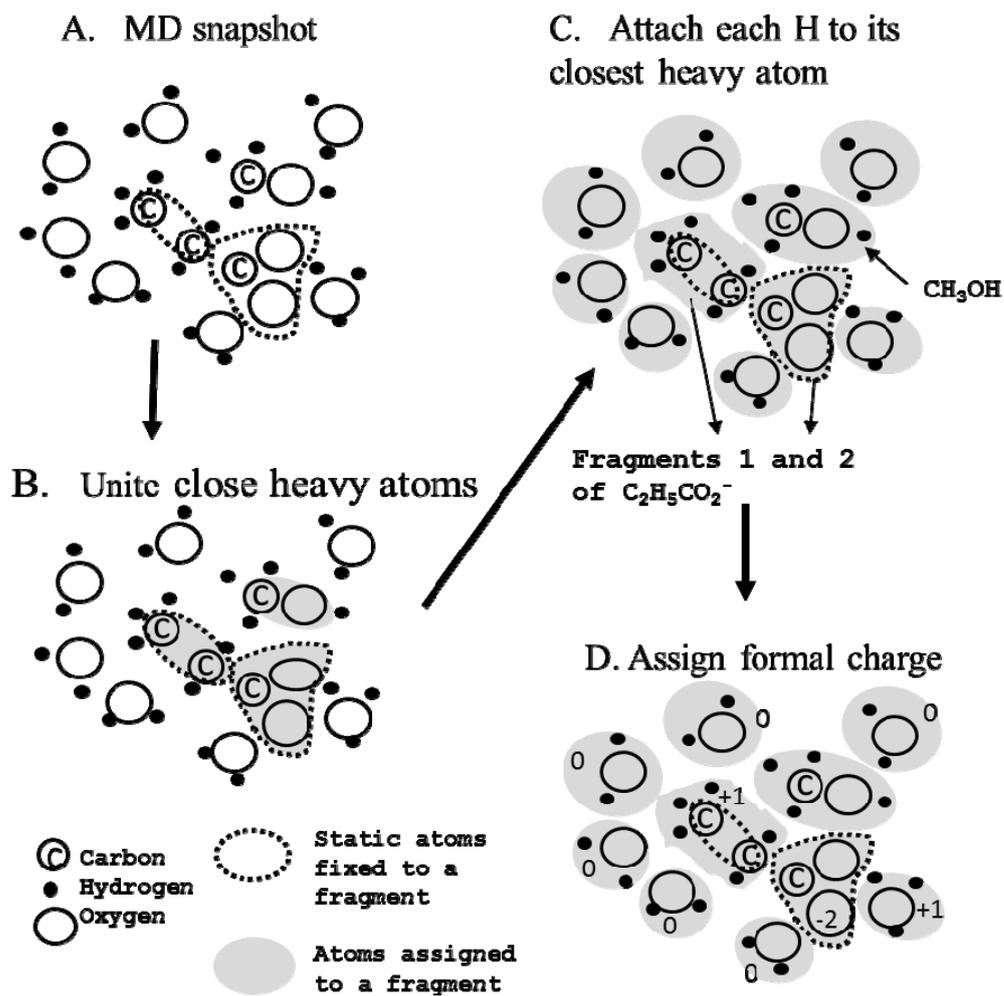


Fig. 4

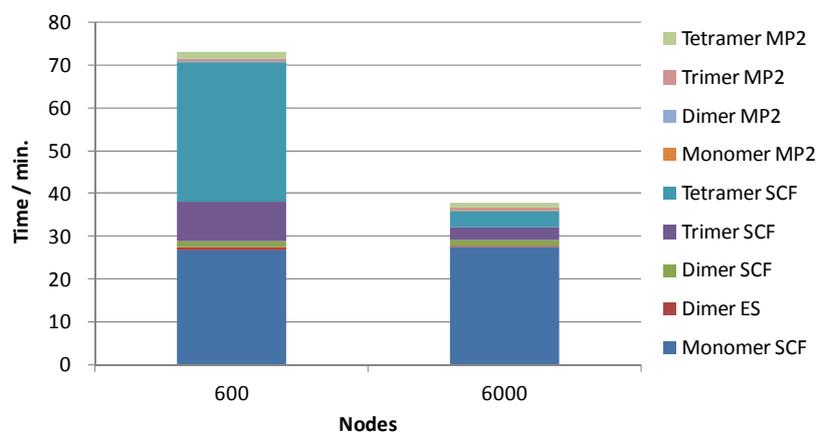


Fig. 5

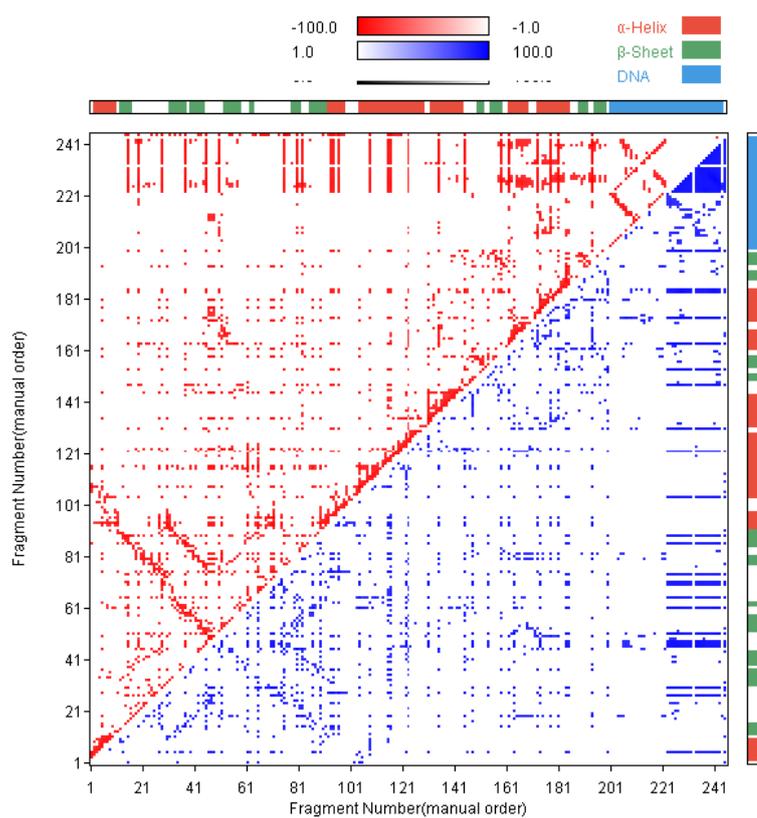


Fig. 6

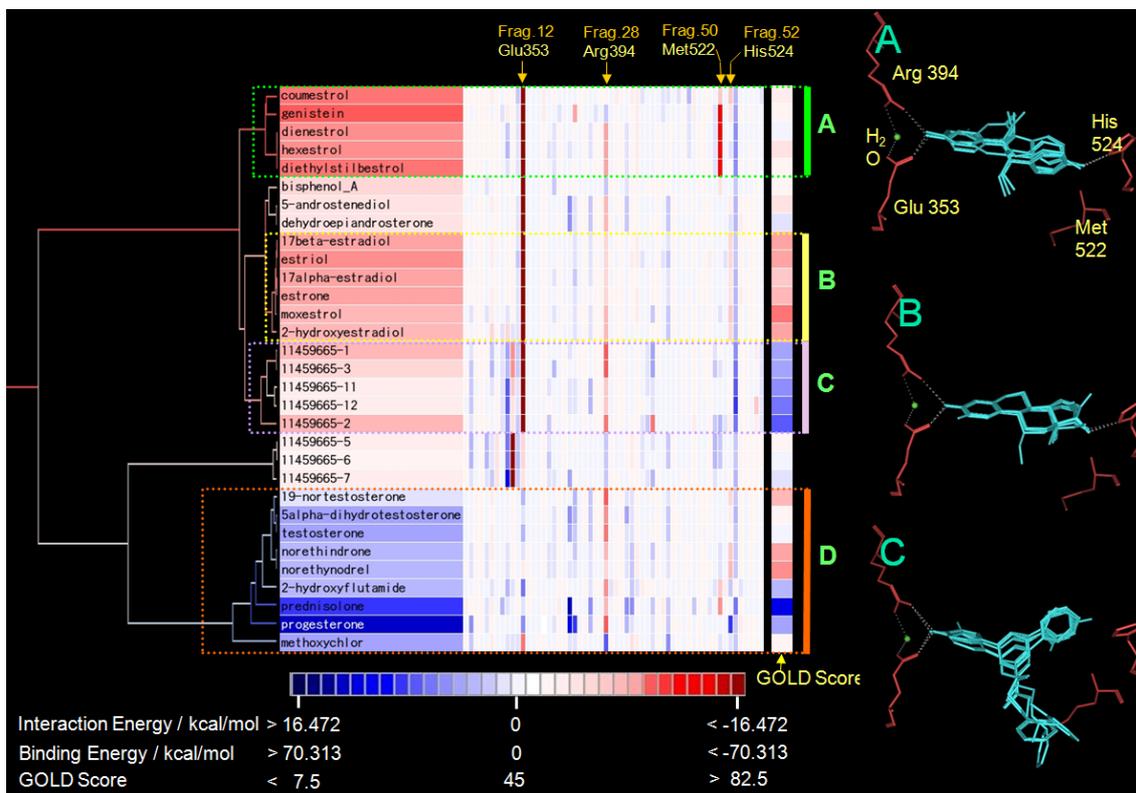


Fig. 7

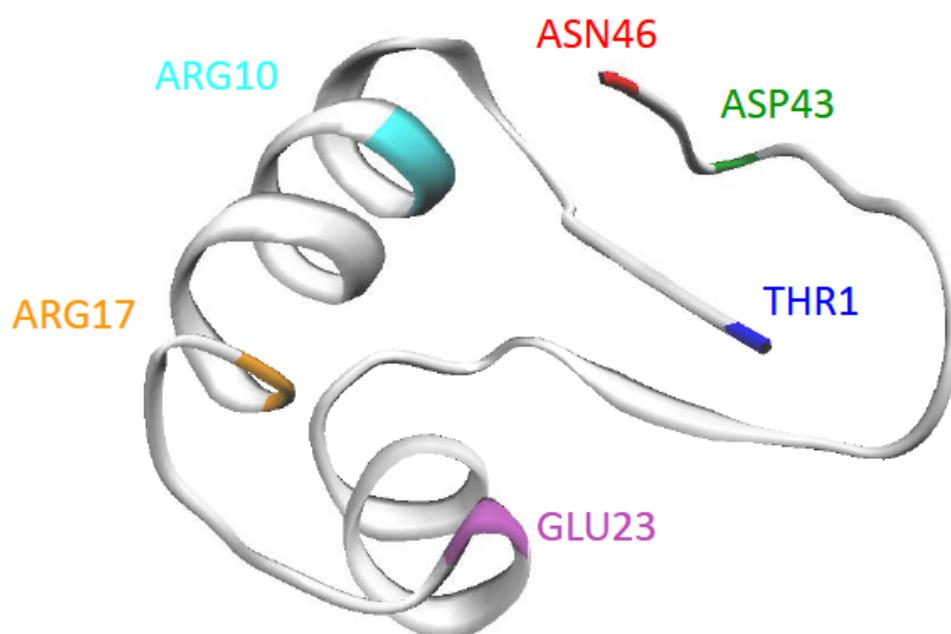
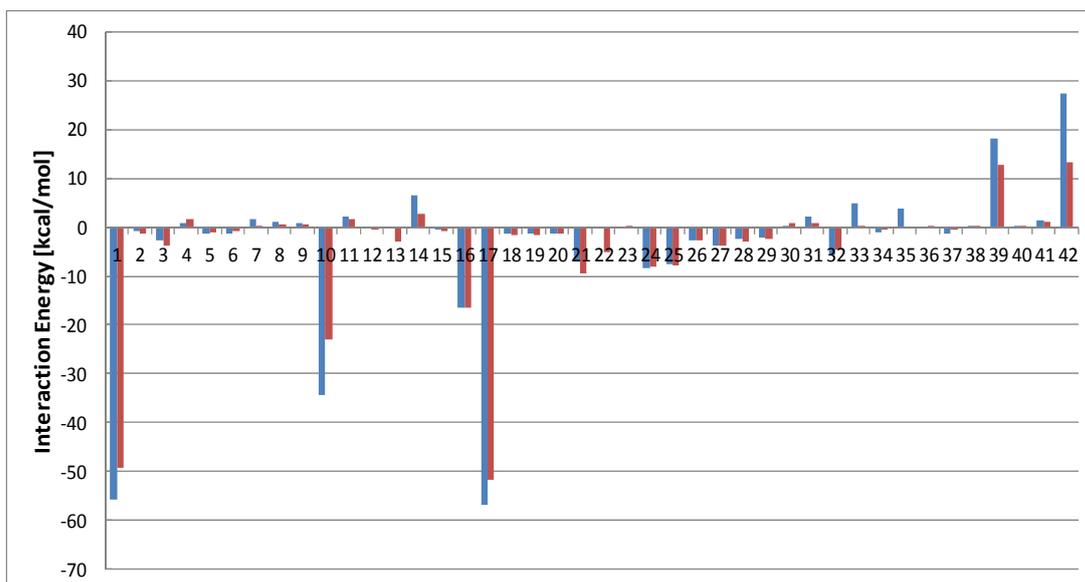


Fig. 8

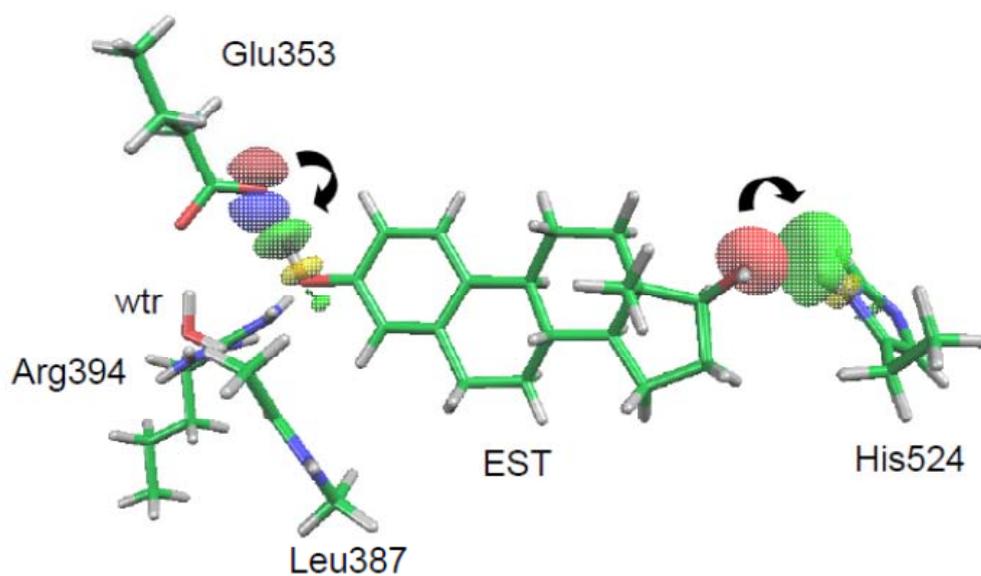


Fig. 9

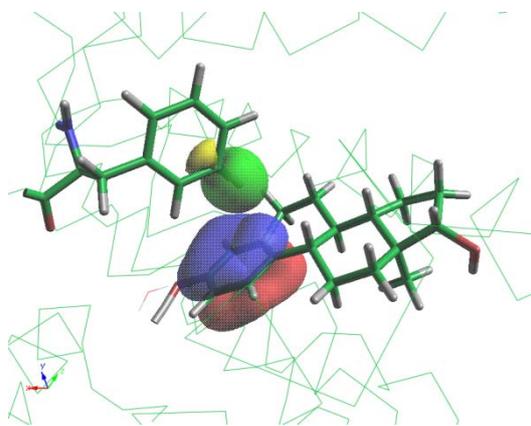


Fig. 10

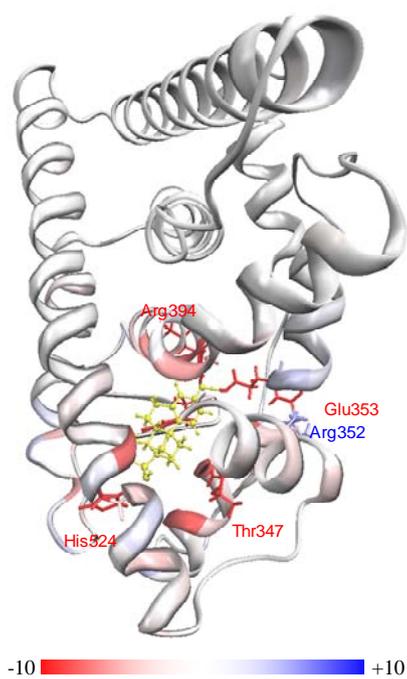


Fig. 11

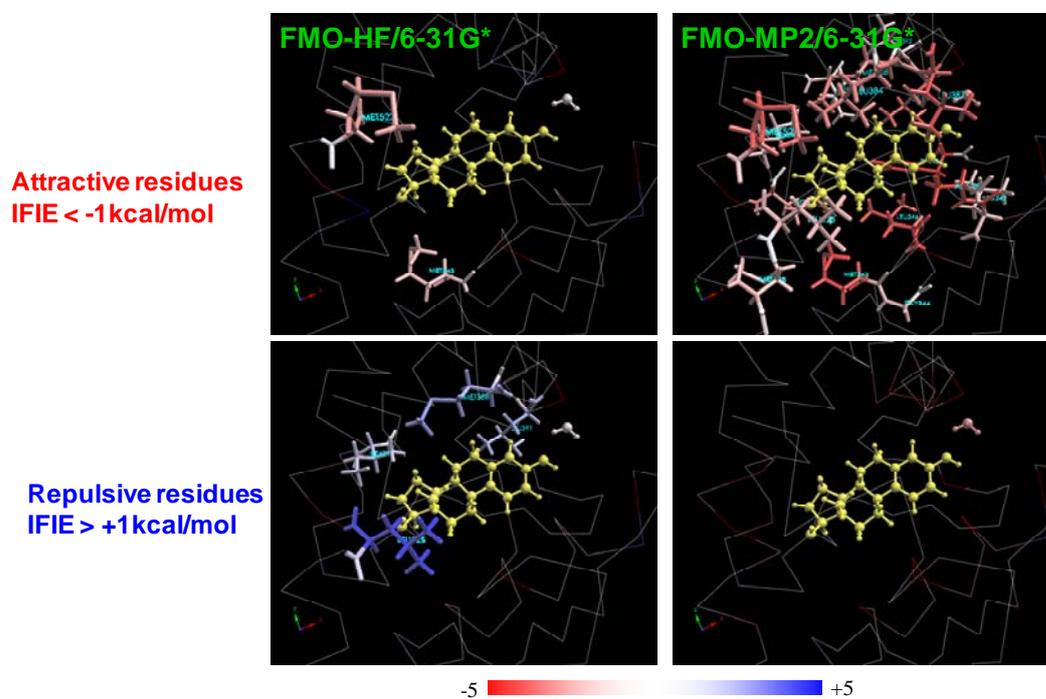


Fig. 12

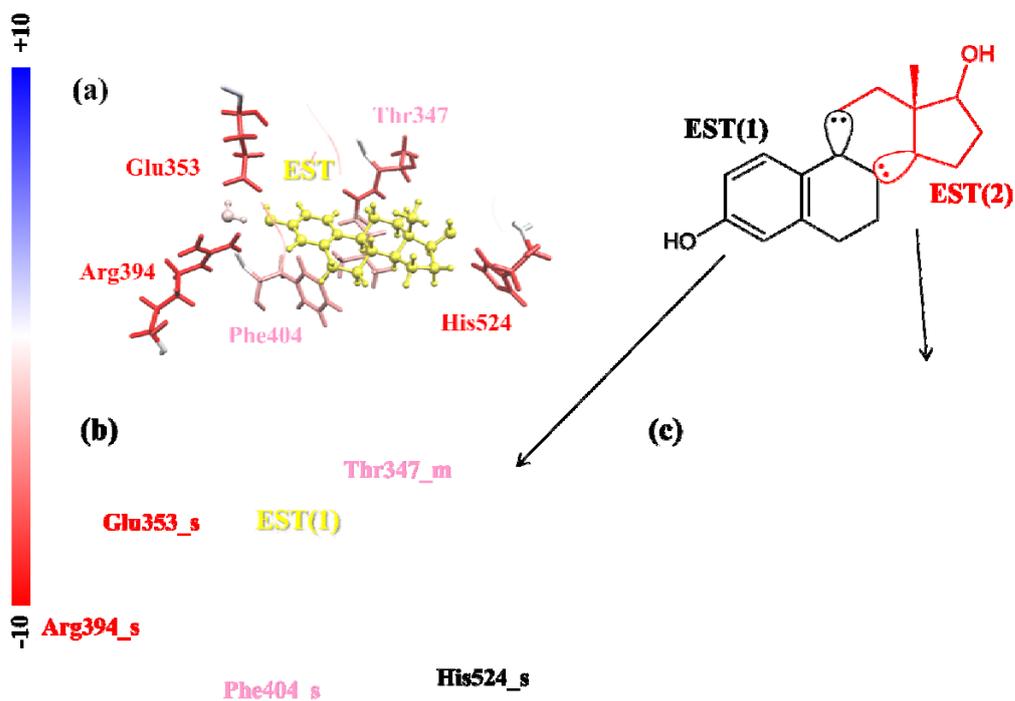


Fig. 13

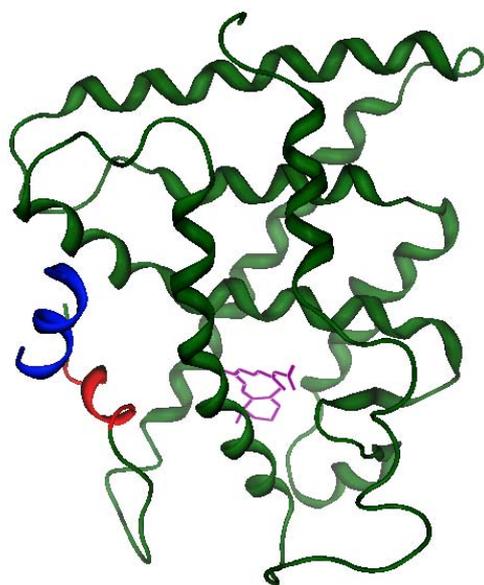


Fig. 14

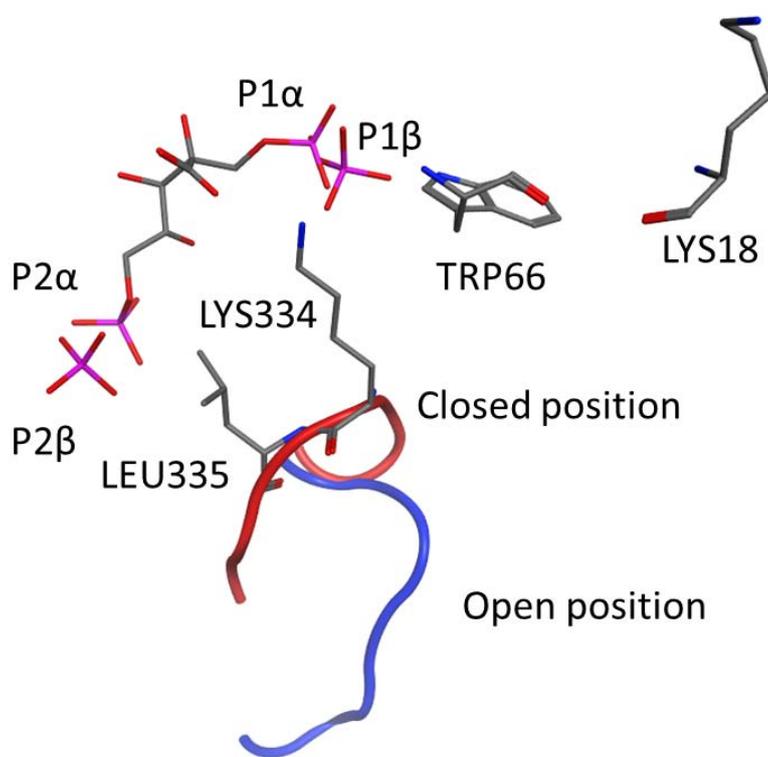


Fig. 15

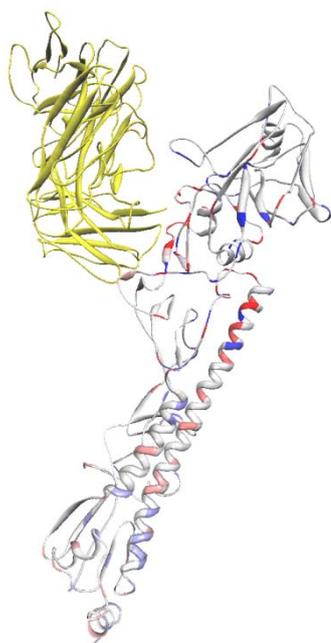
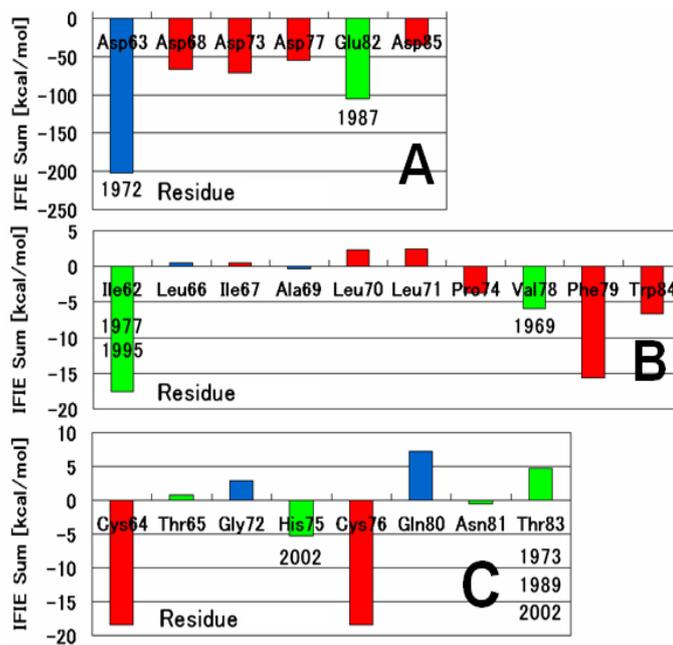


Fig. 16



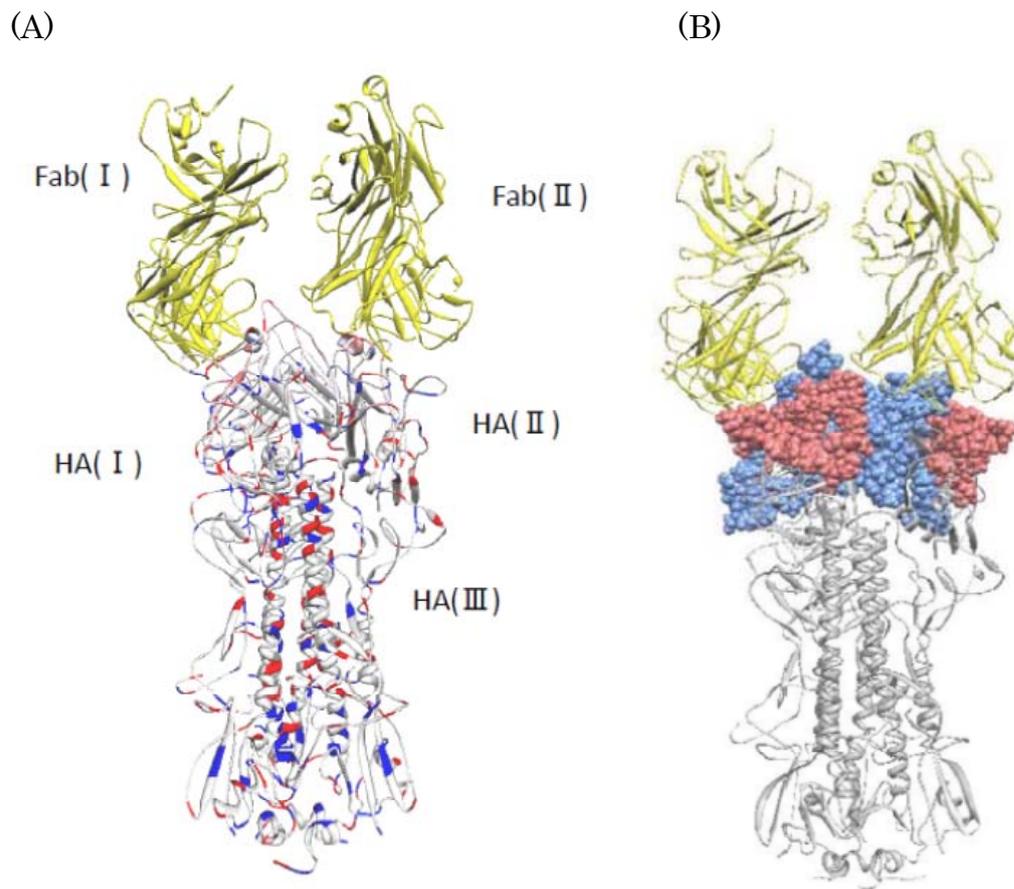


Fig. 17

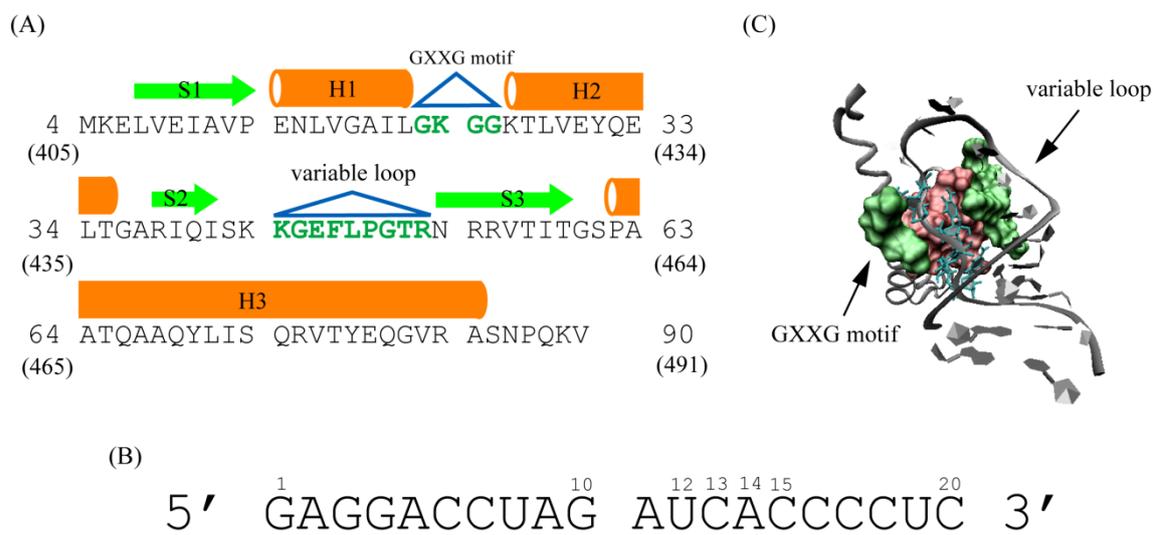


Fig. 18

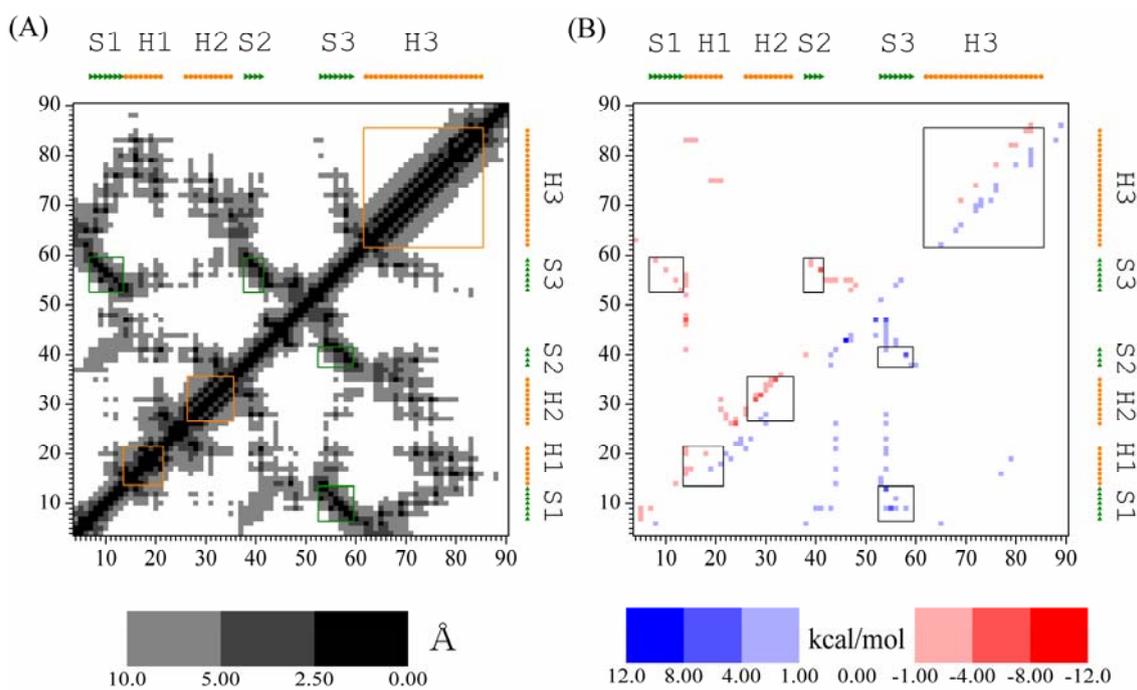


Fig. 19

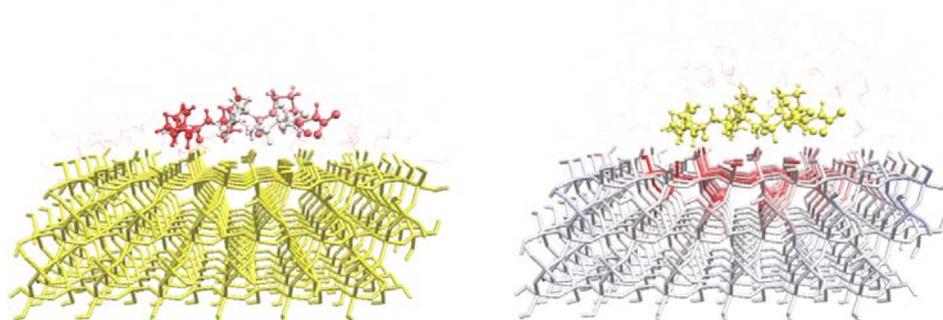


Fig. 20

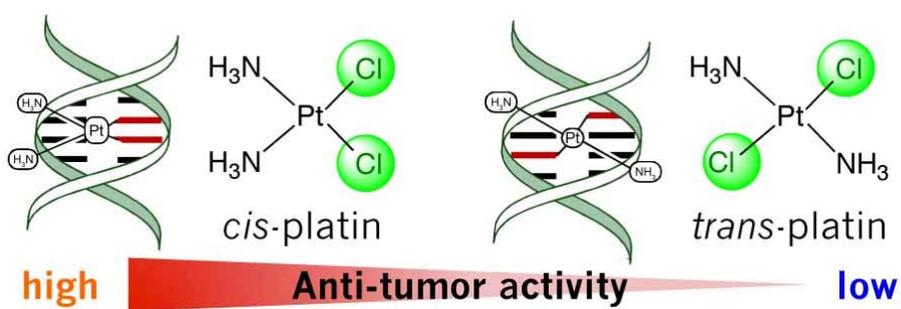


Fig. 21

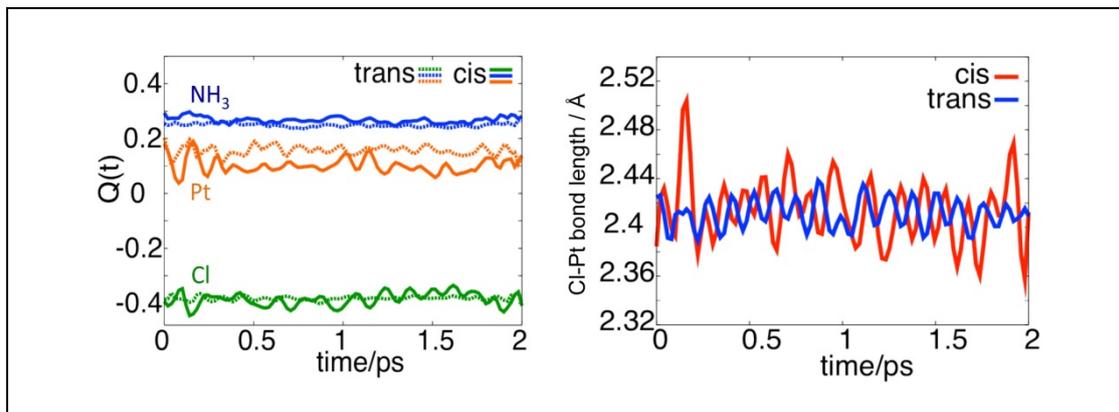


Fig. 22