

## The Use of Many-Body Expansions and Geometry Optimizations in Fragment-Based Methods

Dmitri G. Fedorov,<sup>†</sup> Naoya Asada,<sup>‡</sup> Isao Nakanishi,<sup>⊥</sup> and Kazuo Kitaura<sup>\*,§</sup>

<sup>†</sup>NRI, National Institute of Advanced Industrial Science and Technology (AIST), Central 2, Umezono 1-1-1, Tsukuba, 305-8568, Japan

<sup>‡</sup>Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

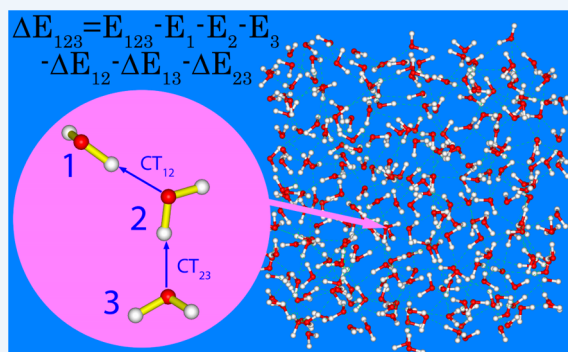
<sup>⊥</sup>Department of Pharmaceutical Sciences, Kinki University, 3-4-1, Kowakae, Higashi-Osaka, Osaka 577-8502, Japan

<sup>§</sup>Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

**CONSPECTUS:** Chemists routinely work with complex molecular systems: solutions, biochemical molecules, and amorphous and composite materials provide some typical examples. The questions one often asks are what are the driving forces for a chemical phenomenon? How reasonable are our views of chemical systems in terms of subunits, such as functional groups and individual molecules? How can one quantify the difference in physicochemical properties of functional units found in a different chemical environment? Are various effects on functional units in molecular systems additive? Can they be represented by pairwise potentials? Are there effects that cannot be represented in a simple picture of pairwise interactions? How can we obtain quantitative values for these effects?

Many of these questions can be formulated in the language of many-body effects. They quantify the properties of subunits (fragments), referred to as one-body properties, pairwise interactions (two-body properties), couplings of two-body interactions described by three-body properties, and so on. By introducing the notion of fragments in the framework of quantum chemistry, one obtains two immense benefits: (a) chemists can finally relate to quantum chemistry, which now speaks their language, by discussing chemically interesting subunits and their interactions and (b) calculations become much faster due to a reduced computational scaling. For instance, the somewhat academic sounding question of the importance of three-body effects in water clusters is actually another way of asking how two hydrogen bonds affect each other, when they involve three water molecules. One aspect of this is the many-body charge transfer (CT), because the charge transfers in the two hydrogen bonds are coupled to each other (not independent).

In this work, we provide a generalized view on the use of many-body expansions in fragment-based methods, focusing on the general aspects of the property expansion and a contraction of a many-body expansion in a formally two-body series, as exemplified in the development of the fragment molecular orbital (FMO) method. Fragment-based methods have been very successful in delivering the properties of fragments, as well as the fragment interactions, providing insights into complex chemical processes in large molecular systems. We briefly review geometry optimizations performed with fragment-based methods and present an efficient geometry optimization method based on the combination of FMO with molecular mechanics (MM), applied to the complex of a subunit of protein kinase 2 (CK2) with a ligand. FMO results are discussed in comparison with experimental and MM-optimized structures.



### 1. INTRODUCTION

Many molecular systems have a large number of atoms, for example, proteins, DNA, nanomaterials, etc. Whether truncated models provide an adequate representation of them is always a big question, necessitating an analysis of convergence with respect to increasing the model size.<sup>1</sup> Important chemical properties in many cases strongly depend on the truncation level. For instance, by extracting the active center of a protein–ligand complex, one alters the long-range electrostatic environment, which can exhibit a strong influence on the binding. Another problem is solvent effects. In a truncated model, treatment of solvent can be very unrealistic, and mechanistic and dynamic properties of large systems can be strongly

influenced by truncation. Likewise, using small models of nanomaterials can result in very different electronic properties because of the quantum confinement effect,<sup>2</sup> among other causes. An alternative of using periodic boundary conditions for materials that are not periodic can bring about profound effects related to improper symmetry and concentration of components (because of using small elementary cells).

**Special Issue:** Beyond QM/MM: Fragment Quantum Mechanical Methods

**Received:** June 13, 2014

**Published:** August 21, 2014

Therefore, it is necessary to develop accurate and efficient methods capable of treating large molecular systems. Traditional force fields in the molecular mechanics (MM) framework have been highly tuned in terms of both computational efficiency<sup>3</sup> and accuracy, especially for systems composed of standard building blocks, such as proteins made of amino acid residues. The development of new force fields is an active area of research, and the distinct trend for this development is to heavily rely on quantum-chemical (QM) calculations.<sup>4</sup>

In order to improve the accuracy of force fields and deal with the breaking and creation of chemical bonds, an interface of QM and MM has been developed, which can be done with a mechanical embedding in the integrated molecular orbital (MO) molecular mechanics (IMOMM)<sup>5,6</sup> or an electrostatic embedding, QM/MM.<sup>7</sup> A generalization of IMOMM, our own *N*-layer integrated molecular orbital molecular mechanics (ONIOM) methodology,<sup>8,9</sup> is also available.<sup>10</sup>

Another route toward a more accurate treatment of large systems is given by fragment-based methods,<sup>4</sup> which have a long history and are becoming increasingly popular.<sup>11–25</sup> There is also a continuous effort to develop linear scaling<sup>26,27</sup> semiempirical approaches<sup>28,29</sup> and other approaches.<sup>30</sup>

In the fragment molecular orbital (FMO) method,<sup>31–35</sup> a molecular system is divided into fragments, which are calculated using an *ab initio* QM method in the presence of the embedding electrostatic potential (ESP). Using ESP derived from converged fragment densities, fragment pair (dimer) and, optionally, fragment triple (trimer) calculations are performed.

Any molecular study in QM relies on molecular geometry. Apart from MD simulations, where the dynamic effects on the structure are investigated, a typical QM study deals with obtaining stationary points (energy minima and transition states) and evaluating properties for them. Optimizing structures of large systems is a major challenge, because they are often flexible and require many hundreds or thousands of single point gradient calculations. There is some progress in applying traditional<sup>36</sup> as well as novel<sup>37–40</sup> and fragment-based algorithms<sup>14,23,41–43</sup> for QM-based optimizations of structures. An alternative is to perform QM calculations using the structures obtained with other methods, for instance, MM-optimized structures.

Using the analytic gradient<sup>44,45</sup> has enabled optimization of polypeptides and very small proteins<sup>46</sup> as well as radical systems,<sup>47</sup> in FMO. To improve the efficiency, it has been proposed to save computational efforts by focusing on a part of the system in the partial<sup>48</sup> and frozen domain<sup>49</sup> methods, later also adopted for effective FMO (EFMO).<sup>50–53</sup> An implementation of QM/MM where QM is described by FMO has been developed.<sup>54</sup> Recently, transition state search with FMO has become possible, and some chemical reactions have been studied with FMO.<sup>55,56</sup>

By these methods, the structures of chignolin (PDB 1UAO),<sup>46,57</sup> Trp-cage protein construct (PDB 1L2Y),<sup>46,58</sup> crambin (PDB 1CRN),<sup>59</sup> silicon nanowire,<sup>60</sup> boron nitride nanoribbons,<sup>61</sup> and polymer radicals<sup>62</sup> have been optimized using FMO. The present state is that full geometry optimizations of relatively rigid inorganic systems, requiring a manageable number of optimization steps, are possible for systems of considerable size, as exemplified by an optimization of BN nanoribbons containing 7878 atoms. However, very flexible biological systems such as proteins or water clusters can only be realistically treated when the number of optimized

atoms is on the order of several hundred or less. A practical way is resort to partial optimizations, that is, freezing the coordinates of the larger part of atoms, in which case one can handle large systems, as demonstrated in the partial optimization of the protein–ligand complex consisting of 19471 atoms.<sup>49</sup> Other alternatives to sampling the configurational space are given by Monte Carlo<sup>63</sup> or genetic<sup>64</sup> algorithms.

In the Methodology section, we summarize and generalize the applications of the many-body property expansion in fragment-based methods. Extending an earlier work,<sup>46</sup> we enable a covalent boundary between MO and MM regions in the FMO-based IMOMM method (FMO/MM) and apply it to the optimization of a protein–ligand complex.

## 2. METHODOLOGY

In 1970, a many-body expansion of the interaction potential was proposed,<sup>65</sup> and in 1977, it was suggested<sup>66</sup> to use a similar expression in the framework of a subsystem-based method. In 1992, the same idea was used to define the correlation energy based on contributions of groups of localized orbitals<sup>67</sup> and another usage of it was made to define the FMO energy in 1999.<sup>31</sup> Although these expressions look rather alike, there is a considerable difference in the physical picture. In the original FMO,<sup>31</sup> the following expression is used to define the energy of a system divided into *N* fragments,

$$E^{\text{FMO2}} = \sum_I E_I + \sum_{I>J} (E_{IJ} - E_I - E_J) \quad (1)$$

where  $E_I$  and  $E_{IJ}$  are the energies of fragments *I* and their pairs *IJ*, respectively. The essential feature of FMO2 is the incorporation of many-body effects in the expression, which is formally limited to two-body contributions. This is accomplished by the self-consistent treatment of the fragment polarization, so that the electrostatic effects are treated at the full *N*-body level, as shown numerically in the pair interaction energy decomposition analysis (PIEDA).<sup>68</sup> One can also see that every subsystem energy in eq 1 includes *N*-body effects because the energies are computed in the electrostatic field of all fragments, thus their electrostatic coupling at the level of *N*-body terms is accounted for. The self-consistent polarization loop used in FMO and other methods<sup>13</sup> is necessary to incorporate many-body polarization. In FMO, only quantum effects such as charge transfer and exchange-repulsion are truncated at the two-body level. A diagrammatic representation of FMO clearly showed the details of counting many-body contributions.<sup>69</sup>

The SCF energies, which appear in eq 1, include the interaction energy with the embedding potential. It was suggested<sup>70</sup> in 2002 to separate this interaction for two purposes: (a) a better interaction picture, in which explicit many-body effects are separated and (b) introduction of approximations. The modified expression, which is entirely equivalent to eq 1 provided that no approximations are used, is

$$E^{\text{FMO2}} = \sum_I E'_I + \sum_{I>J} [(E'_{IJ} - E'_I - E'_J) + \text{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ})] \quad (2)$$

Here, the internal energies  $E'_X$  of monomers  $X = I$  and dimers  $X = IJ$  are

$$E'_X = E_X - \text{Tr}(\mathbf{D}^X \mathbf{V}^X) \quad (3)$$

where  $\mathbf{D}^X$  and  $\mathbf{V}^X$  are the electron density and embedding ESP of  $X$ , respectively.  $\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^I \oplus \mathbf{D}^J)$  is the density difference matrix describing the interfragment charge (density) transfer in dimer  $IJ$ .

There are two main approximations<sup>70</sup> applicable to eq 2. The first is the separated dimer approximation (ES-DIM), in which the energy of the separated dimer  $IJ$  is approximated for fragments sufficiently separated from each other as the sum of the two monomer energies plus the electrostatic (ES) interaction between the monomers,

$$E'_{IJ} = E'_I + E'_J + \Delta E_{IJ}^{\text{ES}} \quad (4)$$

The other is the approximation of ESP, composed of the potentials  $\mathbf{V}^{X(I)}$  of individual fragments  $I$  acting on  $X$

$$\mathbf{V}^X = \sum_{I \neq X}^N \mathbf{V}^{X(I)} \quad (5)$$

In the ES-PTC approximation,  $\mathbf{V}^{X(I)}$  for fragments  $I$  far separated from  $X$  is computed using atomic charges rather than electron density. The original expression in eq 1, as can be inferred also from the diagrammatic treatment,<sup>69</sup> cannot be used for introducing either of these approximations, because of the double counting of the electrostatics (the ES-DIM approximation) and misbalance<sup>71</sup> of the approximations in monomer and dimer ESPs (the ES-PTC approximation). One can rewrite eq 1 using combinatorics in terms of monomer and dimer energies as

$$E^{\text{FMO2}} = (2 - N) \sum_I^N E_I + \sum_{I>J}^N E_{IJ} \quad (6)$$

but this expression is only useful without approximations; otherwise it is misleading and its use should be avoided. Reiterating, eqs 1, 2, and 6 representing FMO2 are only equivalent when no approximations are employed; and otherwise only eq 2 is appropriate.

In addition to reducing the effective scaling because the number of SCF dimers scales linearly with  $N$  when the ES-DIM approximation is used, eq 2 separates explicit many-body effects. Namely, the  $M$ -body effects ( $M > 2$ ) in FMO2 appear because of the embedding potentials, and the SCF energies  $E_X$  in eq 1 include them. However, in eq 2, internal energies are used, which reflect the implicit (impl) many-body effects, whereas explicit (expl) many-body effects are now separated.

$$\begin{aligned} E^{\text{FMO2}} &= E^{\text{FMO2,impl}} + E^{\text{FMO2,expl}} \\ E^{\text{FMO2,impl}} &= \sum_I^N E'_I + \sum_{I>J}^N (E'_{IJ} - E'_I - E'_J) \\ E^{\text{FMO2,expl}} &= \sum_{I>J}^N \text{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \end{aligned} \quad (7)$$

$E^{\text{FMO2,impl}}$  describes the internal energy contribution and reflects many-body polarization effects on the two-body interactions. The explicit many-body effects are given by the  $E^{\text{FMO2,expl}}$  term, which is the energy cost to accomplish interfragment charge (density) transfer,  $\Delta \mathbf{D}^{IJ}$ , under the influence of the embedding potential,  $\mathbf{V}^{IJ}$ , for fragments polarized by ESP. In other words, it quantifies whether

surrounding charge distributions of fragments other than  $I$  and  $J$  (in ESP  $\mathbf{V}^{IJ}$ ) promote or demote the charge transfer between  $I$  and  $J$  in dimer  $IJ$ . The charge transfer energy has two contributions, one described by  $E^{\text{FMO2,expl}}$  and the other is a fraction of  $E^{\text{FMO2,impl}}$ .

The energy decomposition analysis (EDA)<sup>72</sup> was developed in 1976. Its incorporation into FMO in the form of PIEDA provides further insights into the physical picture of FMO. A simple energy expansion in eq 1, when applied to the energies of isolated fragments and their pairs (without the embedding potential), can be thought of as a purely two-body truncation of the many-body interaction.<sup>65</sup> Its accuracy may be good for nonpolar systems but is not very satisfactory in the case of polar fragments, for instance, water.<sup>73</sup> By introduction of the electrostatic embedding, it is possible to bring many-body effects into the expression, which is formally two-body (eq 1). PIEDA clearly shows that the electrostatics and polarization in FMO are the same as in the full EDA, applied to the system of  $N$  fragments.<sup>74</sup> It is the exchange-repulsion and charge-transfer effects that are truncated at the two-body level in FMO2.

Thus, FMO may be viewed as a truncation of EDA, with a high-order treatment of the electrostatic effects. Although this preferential treatment of the electrostatics may be seen as a distinct improvement, because the electrostatic effects are long-ranged and very important in polar systems, the lack of the high-order exchange-repulsion and charge transfer becomes a serious problem either when large basis sets are used<sup>75</sup> or when charge transfer is large, for instance, in the case of metal ions.<sup>76</sup> Adding the exchange potential<sup>77,78</sup> to the ESP (which usually only has Coulomb terms) does not seem to improve the accuracy in general.

The only consistent way to improve the accuracy is to use three-<sup>69</sup> or four-body<sup>79</sup> expansions. A number of other fragment-based methods have also chosen to incorporate them in one way or another.<sup>14,16,19,22,24,80–86</sup> Many-body effects of QM nature are very important in the systems with extensive hydrogen bonding, in particular, in water.<sup>87–91</sup> FMO2 lacks many-body quantum effects as revealed by PIEDA,<sup>68</sup> which is why FMO2 shows some errors in water clusters.<sup>69</sup> Note that the actual errors reflect to some extent the effect of error cancellation due to various factors: for instance, in FMO2, when point charges are used to describe ESP, the errors are very considerably reduced<sup>75</sup> because the point charge treatment introduces errors that favorably cancel to a large extent the errors due to neglected many-body quantum effects. The importance of many-body effects for protein folding has been stressed,<sup>92</sup> as well as the importance of charge polarization.<sup>93</sup>

The many-body expansion of properties is not limited to the energy. In a more general form, a property  $\mathcal{A}$  can be written in an  $M$ -body expansion as

$$\mathcal{A}^M = \mathcal{A}^1 + \sum_{m=2}^M \Delta^m \mathcal{A} \quad (8)$$

where the properties  $\mathcal{A}_I$  of  $N$  subsystems (fragments) are summed in the one-body total property,

$$\mathcal{A}^1 = \sum_{I=1}^N \mathcal{A}_I \quad (9)$$

and the  $m$ -body corrections are

$$\Delta^m \mathcal{A} = \sum_{I_1 > \dots > I_m}^N \Delta \mathcal{A}_{I_1 \dots I_m} \quad (10)$$

For instance,

$$\Delta \mathcal{A}_{IJ} = \mathcal{A}_{IJ} - \mathcal{A}_I - \mathcal{A}_J \quad (11)$$

$$\begin{aligned} \Delta \mathcal{A}_{IJK} &= \mathcal{A}_{IJK} - \mathcal{A}_I - \mathcal{A}_J - \mathcal{A}_K - \Delta \mathcal{A}_{IJ} - \Delta \mathcal{A}_{IK} \\ &\quad - \Delta \mathcal{A}_{JK} \end{aligned} \quad (12)$$

and so on. These can be written in a general recursive form as

$$\Delta \mathcal{A}_{I_1 \dots I_m} = \mathcal{A}_{I_1 \dots I_m} - \sum_{i=1}^m \mathcal{A}_{I_i} - \sum_{i=2}^{m-1} \Delta \mathcal{A}_{\{I_i\}} \quad (13)$$

where  $\Delta \mathcal{A}_{\{I_i\}}$  is the sum of the  $i$ -body terms  $\Delta \mathcal{A}_{I_1 \dots I_i}$  involving all unique combinations of sets  $I_1 \dots I_i$ .

Because the contributions in eq 8 usually decrease in magnitude with the many-body level (i.e., typically,  $|\Delta^m \mathcal{A}| < |\Delta^{m-1} \mathcal{A}|$ ) there is no need to compute all  $m$ -body contributions in eq 8 at the same level of theory, although the combination of theories should be compatible (i.e., mixing different levels of treating the electron correlation). Similar ideas have been used in the context of computing the electron correlation with localized orbitals.<sup>94</sup> For fragment-based methods, there are many multilayer formulations of FMO<sup>95</sup> and other methods,<sup>96</sup> some of which explicitly use different levels of theory for many-body corrections.<sup>97</sup>

It can be noted that the expansion in eq 8 resembles at least formally, a perturbation expression, where one defines a zero-order level as the sum  $\mathcal{A}_I$  of all monomer properties, to which higher-order corrections are added. It was argued<sup>77</sup> that FMO can be viewed as a perturbation theory, based on the Green's function formalism. On the other hand, symmetry-adapted perturbation theory<sup>98</sup> has been applied to analyze interactions in a variety of systems.

Because FMOM includes  $N$ -body effects (where  $N$  is the number of fragments), for its energy in eq 2 one cannot designate such a single  $\mathcal{A}$  to recast eq 2 in the form of eq 8. It is possible instead to use

$$E^{\text{FMOM}} = \mathcal{A}^M + \mathcal{B}^M \quad (14)$$

where  $\mathcal{A}^M$  and  $\mathcal{B}^M$  correspond to  $E^{\text{FMO2, impl}}$  and  $E^{\text{FMO2, expl}}$ , respectively. For completeness, the FMO3 equation is given below, and it can be seen that it also features a combination of two properties.

$$\begin{aligned} E^{\text{FMO3}} &= E^{\text{FMO2}} + \sum_{I>J>K} (E'_{IJK} - E'_I - E'_J - E'_K - \Delta E'_{IJ} \\ &\quad - \Delta E'_{IK} - \Delta E'_{JK}) + \sum_{I>J>K} [\text{Tr}(\Delta \mathbf{D}^{JK} \mathbf{V}^{JK}) \\ &\quad - \text{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) - \text{Tr}(\Delta \mathbf{D}^{IK} \mathbf{V}^{IK}) \\ &\quad - \text{Tr}(\Delta \mathbf{D}^{JK} \mathbf{V}^{JK})] \end{aligned} \quad (15)$$

where

$$\begin{aligned} \Delta E'_{IJ} &= E'_{IJ} - \Delta E'_I - \Delta E'_J \\ \Delta \mathbf{D}^{JK} &\equiv \mathbf{D}^{JK} - (\mathbf{D}^I \oplus \mathbf{D}^J \oplus \mathbf{D}^K) \end{aligned} \quad (16)$$

Naturally, two questions arise: (a) what physical conditions should be satisfied in order to make the expansion in eq 8 valid

and (b) does the expansion converge and how quickly? Although a general answer to neither of these questions is attempted to be given here, as a comment pertaining to the first question, it appears that the property  $\mathcal{A}$  should be size-extensive in the usually accepted sense that the property of a system AB can be written as a sum of A and B, if there is no interaction between them.

$$\mathcal{A}(AB) = \mathcal{A}(A) + \mathcal{A}(B) \quad (17)$$

Clearly, not all properties satisfy this criterion. If  $\mathcal{A}$  is the energy of a method, which is not size-extensive, for instance, the energy of configuration interaction with singles and doubles (CISD), then one cannot use this expansion. Therefore, one cannot expect that FMOM-CISD ( $M > 1$ ) should be useful in general to obtain the total CISD energy. Fortunately, many QM methods are size-extensive. However, even for a size-extensive QM method the expansion of  $\mathcal{A}$  cannot be used, if  $\mathcal{A}$  is not the energy but some other property that does not satisfy eq 17. For instance, if  $\mathcal{A}$  is the intensity of a transition between states 1 and 2 under the influence of an operator  $\hat{A}$  such as dipole moment,

$$\mathcal{A} \approx |\langle 1 | \hat{A} | 2 \rangle|^2 \quad (18)$$

then it does not appear to be possible to expand this property in eq 8. Another example of a property that is difficult to decompose into many-body contributions is the polarization with a nonlinear coupling.

There is considerable experience built in FMO<sup>69,76</sup> and other methods,<sup>24</sup> which shows by actual examples although not in any way by a general proof that the properties of molecular systems do converge and do so very quickly. In fact, even one-body methods are already useful,<sup>13</sup> and three and higher order methods very often have negligible errors, whereas two-body methods typically but not always have a satisfactory accuracy.

Many fragment-based methods impose a fixed integer electron count on fragments and their pairs, although there are some schemes that allow fractional occupations.<sup>12,99</sup> Because it is a restriction, it was suggested<sup>68</sup> to use the values of the interfragment charge transfer as a measure of the accuracy, although partly because of possible fortuitous error cancellations the criterion is by no means decisively reliable. In terms of convergence, it should be noted that the expansion in eq 8 possesses the following important property by construction

$$\mathcal{A}^N = \mathcal{A}_{I_1 \dots I_N} \quad (19)$$

In other words, for the full series extended to  $N$ -body terms, the property  $\mathcal{A}$  is determined by the single value computed for the whole system,  $\mathcal{A}_{I_1 \dots I_N}$  (all other lower-body terms cancel out). In other words, when a system is divided into  $N$ -fragments, then the  $N$ -body expansion of the property is always exact, and an  $M$ -body truncation for  $M < N$  is in general approximate. A number of different properties have been expanded in FMO in the series in eq 8. A representative summary is given in Table 1.

Finally, we turn to the physical importance and practical use of the many-body corrections, such as those defined in eqs 11 and 12. First of all, a division of a system into fragments is subjective and is typically based on convenience of further discussion rather than on well-defined mathematical principles. Frequently, the fragment definition is done in such a way as to try to minimize the fragmentation error  $e^M$ .

**Table 1. Properties Expanded in the Many-Body Series in FMO**

property	expansion order	refs
energy	2, 3, 4	31, 69, 79
energy gradient	2, 3	44, 45, 119
Hessian	2	55
electron density	2	120
multipole moments	2, 3	69, 121
atomic charges	2, 3	122
dynamic polarizabilities	2	123
solvent potential	2, 3	124
molecular electrostatic potential	2	125
Fock matrix	2, 3	78, 126
chemical shifts	2	127

$$e^M = \mathcal{A}^M - \mathcal{A}^N \quad (20)$$

The definition of the error may include taking its absolute value, although the sign is often kept to see in which way the error is biased. Most but not all<sup>100</sup> fragment-based methods are not fully variational, and thus it cannot be expected that the error in the energy is always positive. In fact, in FMO2 it is typically negative, because of the neglected many-body exchange-repulsion. Typically, the fragmentation strategy reflects a compromise between computational cost (larger fragments take more time), convenience of the result analysis, and accuracy.

The fragment properties are influenced not only by the fragment size but also by the details of the method such as the treatment of fragment boundaries.<sup>57</sup> In the framework of FMO, a few studies deal with the properties of fragments  $\mathcal{A}_{IJ}$ , such as the polarization energy<sup>101</sup> or the excitation energy.<sup>102</sup> Most applications of FMO focus on pair corrections (pair interaction energies, PIE) in eq 11, which in FMO are usually defined as

$$\Delta E_{IJ}^{\text{PIE}} = (E'_{IJ} - E'_I - E'_J) + \text{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \quad (21)$$

Although formally two-body, they also include explicit many-body effects in the last term in eq 21.

Three- and four-body corrections can be added<sup>103</sup> to eq 12 in an averaged way to modify the pair interactions in what we call mean-averaged generalized PIE (MAGPIE). Such division is rather formal and divides the many-body corrections in an equal way, which is not physical in general. However, we recognize that in the case when FMO2 does not provide desirable accuracy, it is not appropriate to limit the discussion of interactions to two-body PIEs, and if a discussion involving three-body terms is not practical, MAGPIE may be a useful alternative. Generalizing MAGPIE for a general property  $\mathcal{A}$ , produces

$$\Delta \mathcal{A}'_{IJ} = \Delta \mathcal{A}_{IJ} + \frac{1}{3} \sum_{K \neq I, J}^N \Delta \mathcal{A}_{IJK} + \frac{1}{6} \sum_{K > L}^N \sum_{K, L \neq I, J} \Delta \mathcal{A}_{IJKL} + \dots \quad (22)$$

The coefficient of 1/3 (1/6) appears because a trimer (tetramer) includes 3 (6) unique dimers, and the many-body corrections,  $\Delta \mathcal{A}_{IJK}$  ( $\Delta \mathcal{A}_{IJKL}$ ), are equally divided among all dimers. The main justification for using effective two-body corrections,  $\Delta \mathcal{A}'_{IJ}$ , in place of  $\Delta \mathcal{A}_{IJ}$  is to compensate for the insufficient accuracy of the latter, which can occur because of a large basis set or the use of small fragments, for instance, side

chains in amino acid residues.<sup>103</sup> With the aid of eq 22, eq 8 can be written in a formally two-body way as

$$\mathcal{A}^M = \mathcal{A}^1 + \sum_{I > J}^N \Delta \mathcal{A}'_{IJ} \quad (23)$$

where

$$\Delta \mathcal{A}'_{IJ} = \Delta \mathcal{A}_{IJ} + \sum_{m=3}^M \frac{1}{C_m^2} \sum_{\substack{I_1 > \dots > I_{m-2} \\ I, J \neq I_1 \dots I_{m-2}}}^N \Delta \mathcal{A}_{IJ, I_1 \dots I_{m-2}} \quad (24)$$

and  $C_m^2$  is the binomial coefficient.

Just like there are different atomic charge definitions based on dividing the electron density into atoms, with Mulliken charges using an equal division of the contributions belonging to two atoms, it is conceivable although not clearly advantageous to generalize the definition of the effective pair corrections in eq 22 in the following way,

$$\Delta \mathcal{A}'_{IJ} = \Delta \mathcal{A}_{IJ} + \sum_{m=3}^M a_{IJ}^m \sum_{\substack{I_1 > \dots > I_{m-2} \\ I, J \neq I_1 \dots I_{m-2}}}^N \Delta \mathcal{A}_{IJ, I_1 \dots I_{m-2}} \quad (25)$$

For equal partitioning in eq 22, the coefficients  $a_{IJ}^m$  are independent of fragment indices  $I, J$ .

$$a_{IJ}^m = 1/C_m^2 \quad (26)$$

It can be argued though that this is unphysical. Consider a trimer  $IJK$  such that fragments  $I$  and  $J$  are close to each other and strongly interact whereas the third fragment  $K$  interacts with  $I$  and  $J$  weakly. According to the equal partitioning recipe, all three effective pair interactions  $IJ$ ,  $IK$ , and  $JK$  include the same contribution of  $\Delta \mathcal{A}_{IJK}/3$ , whereas one can argue that it is mainly due to  $IJ$ . A somewhat similar argument was made in the definition of the interfragment charge transfer in PIEDA, related to the division of charges on the fragment boundary.<sup>68</sup> An even more general definition is

$$\Delta \mathcal{A}'_{IJ} = \Delta \mathcal{A}_{IJ} + \sum_{m=3}^M \sum_{\substack{I_1 > \dots > I_{m-2} \\ I, J \neq I_1 \dots I_{m-2}}}^N (a_{IJ, I_1 \dots I_{m-2}} \Delta \mathcal{A}_{IJ, I_1 \dots I_{m-2}}) \quad (27)$$

where one could partition pair corrections for an  $m$ -mer, based on the pair contributions in it as

$$a_{IJ, I_1 \dots I_{m-2}} = \frac{|\Delta \mathcal{A}_{IJ}|}{\sum_{\substack{K > L \\ K, L \in I_1 \dots I_m}}^N |\Delta \mathcal{A}_{KL}|} \quad (28)$$

The connection between the definitions in eqs 26 and 28 is easy to see by assuming that all  $|\Delta \mathcal{A}_{IJ}|$  are nonzero and equal to each other and noticing that the number of independent pairs in the sum appearing in the denominator of eq 28 is  $C_m^2$ . If all  $|\Delta \mathcal{A}_{IJ}| = 0$  (the limit of noninteracting fragments), then one has to use eq 26.

To improve the reliability of pair interactions, it was suggested to add basis set superposition error corrections<sup>104,105</sup> or to do calculations in the complete basis set limit.<sup>106</sup> An analysis of pair interactions in FMO was used in quantitative structure–activity relationship (QSAR) studies.<sup>107</sup> There are also a number of ways to perform the analysis, either in terms

of total PIEs or with PIEs decomposed into EDA-like contributions of electrostatic (ES), exchange-repulsion (EX), charge transfer plus mixed terms (CT+mix), and dispersion (DI) to which recently solvent screening (SOLV)<sup>108</sup> and thermal fluctuations<sup>109</sup> were added. For instance, in PIEDA with polarizable continuum model (PCM)<sup>108</sup> the pair interaction in solution is defined as

$$\Delta E_{IJ}^{\text{PIE/PCM}} = \Delta E_{IJ}^{\text{ES}} + \Delta E_{IJ}^{\text{EX}} + \Delta E_{IJ}^{\text{CT+mix}} + \Delta E_{IJ}^{\text{DI}} + \Delta E_{IJ}^{\text{SOLV}} \quad (29)$$

To gain more insights, it is useful to decompose PIEs into orbital contributions, associated with individual functional groups, which is done in the configuration analysis for fragment interaction (CAFI)<sup>110</sup> and the fragment interaction analysis based on local MP2 (FILM).<sup>111</sup> Among properties other than the energy, one can name the interfragment charge transfer<sup>68</sup> based on the density matrix difference  $\Delta D^I$ , fragment charges and dielectric constants in solution,<sup>108</sup> and atomic charges.

### 3. OPTIMIZATIONS OF PROTEIN–LIGAND COMPLEXES WITH FMO/MM

Prior to this work, only MO calculations in IMOMM as implemented in GAMESS<sup>112</sup>/Tinker<sup>113</sup> interface were parallelized (note that GAMESS uses a somewhat old sequential version of Tinker adapted for it). This is not a big problem for typical IMOMM calculations, because MM is fast and MO is slow. However, MO calculations done with FMO are fast<sup>114</sup> and compatible in computational cost with MM on medium-sized or large computer clusters. Therefore, we parallelized MM in Tinker: the charge–charge interaction energy and its first derivative (echarge1 module) and the van der Waals energy and its first derivatives (elj1 module) were parallelized, which were found to constitute the main computational bottlenecks for our simulations.

In this work, we extend FMO/MM<sup>46,115</sup> by introducing link atoms (usually hydrogens) at the boundary between FMO and MM regions.<sup>6</sup> There are no FMO-specific changes to the treatment of link atoms, and the general recipe is used.<sup>6</sup> We cap dangling bonds with hydrogen atoms, and use them in fragment calculations. The total energy  $E$  and its gradient are divided into FMO and MM contributions,

$$E = E_{\text{FMO}}(\mathbf{R}_1) + E_{\text{MM}}(\mathbf{R}_1, \mathbf{R}_2) \quad (30)$$

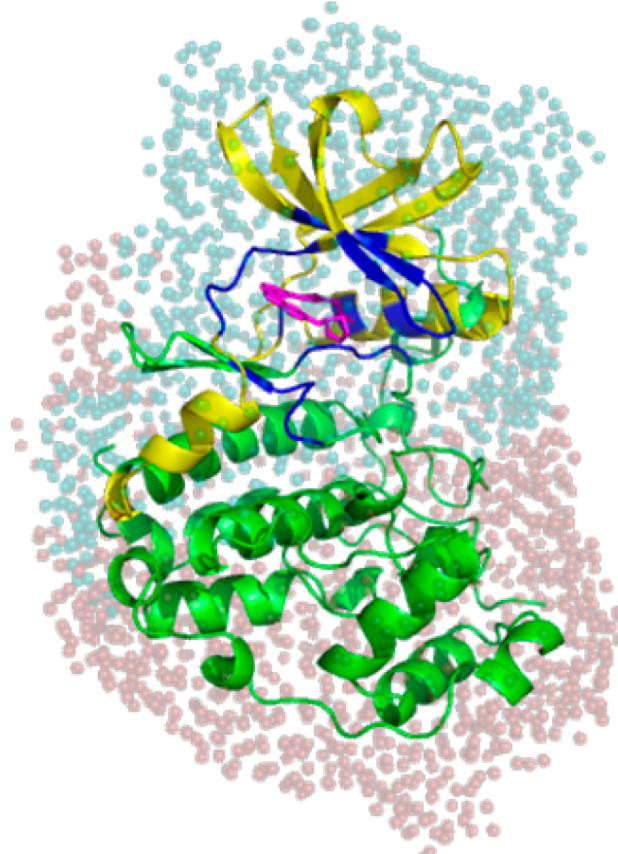
$$\frac{\partial E}{\partial \mathbf{R}_1} = \frac{\partial E_{\text{FMO}}}{\partial \mathbf{R}_1} + \frac{\partial E_{\text{MM}}}{\partial \mathbf{R}_1}$$

$$\frac{\partial E}{\partial \mathbf{R}_2} = \frac{\partial E_{\text{MM}}}{\partial \mathbf{R}_2} \quad (31)$$

where  $\mathbf{R}_1$  and  $\mathbf{R}_2$  are a set of atomic coordinates in the FMO region and atomic coordinates in the MM region, respectively. Note that the interaction energy between FMO and MM atoms is included in  $E_{\text{MM}}$  and the link atoms in FMO region do not feel MM forces.

The following procedure is applied for a geometry optimization: Step 1, provide initial coordinates. Step 2, optimize MM atoms with FMO atoms frozen and compute  $\partial E_{\text{MM}}/\partial \mathbf{R}_1$  and  $E_{\text{MM}}$ . Step 3, compute  $E_{\text{FMO}}$  and  $\partial E_{\text{FMO}}/\partial \mathbf{R}_1$ . Step 4, compute the energy  $E$  and its energy gradient  $\partial E/\partial \mathbf{R}_1$ . Step 5, if  $\partial E/\partial \mathbf{R}_1$  is not small enough, update the geometry of FMO atoms and return to step 2.

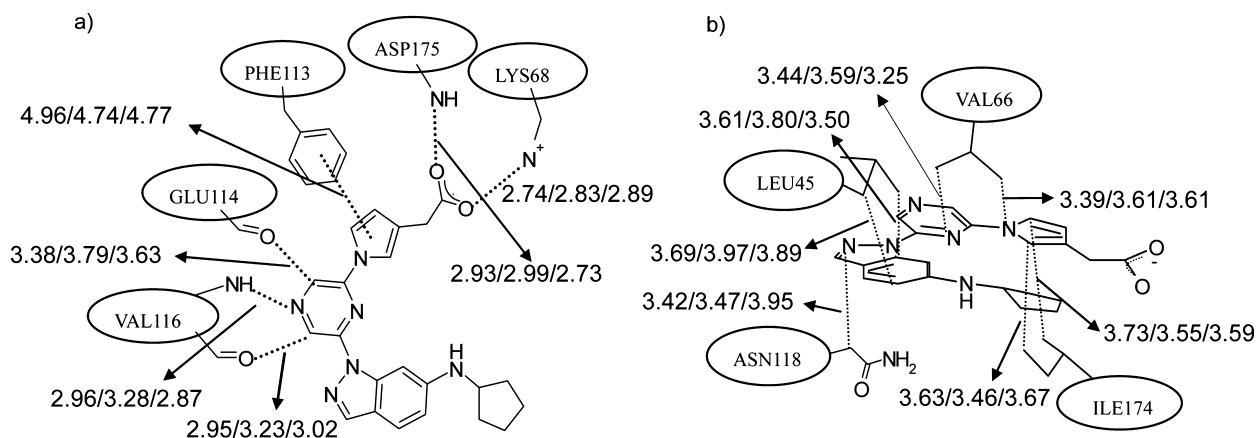
Test calculations were performed on the complex of  $\alpha$  subunit of protein kinase 2 (CK2) with its ligand (Figure 1). In



**Figure 1.** Solvated complex of CK2 $\alpha$  and its ligand ((1-{6-[6-(cyclopentylamino)-1H-indazol-1-yl]pyrazin-2-yl}-1H-pyrrol-3-yl)-acetic acid). FMO atoms are colored in purple (ligand) and blue (the binding pocket of the protein). MM atoms are shown as yellow (optimized) and green (frozen) for the protein, and cyan (optimized) and pink (frozen) for water.

this complex various kinds of molecular interactions are involved: a salt bridge, NH $\cdots$ N, CH $\cdots$ O, and CH $\cdots$  $\pi$  hydrogen bonds, and  $\pi\cdots\pi$  stacking. The interactions vary from strong to weak, and a well balanced description of them is needed to get a good geometry. The experimental structure (PDB 3AT3) was used to construct the model system; hydrogen atoms were added assuming standard protonation states, and the complex was immersed in a solvating water shell of 6 Å.

Geometry optimizations were performed with the convergence threshold OPTTOL set to  $5.0 \times 10^{-4}$  au for FMO and 0.01 kcal/(mol·Å) for MM atoms. The protein in FMO was divided as one residue per fragment, and the ligand was treated as one fragment. The hybrid orbital projection operator was used for treating fragment boundaries.<sup>116</sup> We used FMO2 at the RHF-D/6-31G level (D is the empiric dispersion).<sup>117</sup> AMBER f99 (protein) and gaff (ligand) force fields and TIP3P (water) were employed for MM. The FMO region was defined to include the ligand and the amino acid residues and water molecules separated by the unitless FMO distance<sup>70</sup> of 2.0 from the ligand. We optimized all FMO atoms (667) and 1980 MM atoms. A single point calculation took about 36 min on the Heian cluster (112 CPU cores of Xeon 3.0 GHz), and 470 FMO steps were required for convergence.



**Figure 2.** Selected interatomic distances between amino acid residues and ligand atoms in the optimized structure. The left and right panels show the top and side views of the binding site, respectively. For each distance between heavy atoms (Å), three numbers are given, corresponding to the FMO/MM minimum, MM minimum, and experimental values, respectively.

The optimized structure, *in toto*, is in good agreement with experiment (rmsd is 0.49 Å). Some selected interatomic distances between amino acid residues and ligand atoms are shown in Figure 2. The ligand forms one salt bridge with the NZ atom of Lys68 and two hydrogen bonds with the main chain NH atoms of Val116 and Asp175, stabilizing the binding. The position and relative orientation of these groups are almost the same as in experiment: a small variation of distances ( $\sim 0.2$  Å) is observed. One of the characteristic features of this ligand binding is the CH $\cdots$ O hydrogen bond between the CH group of the pyrazine ring of the ligand and the main chain O atom of Glu114 and Val116. The typical CH $\cdots$ O distance is around 3.6 Å, and the relatively short bonding distances in the experimental structure became even shorter after the geometry optimization. It is plausible because there are multiple polar interactions between the ligand and the hinge region allowing a close contact.

The distances and arrangement of CH $\cdots$  $\pi$  interactions between the hydrophobic moieties of indazole and cyclohexane ring agree well with experiment. The side chain of Leu45 is found to back away from the ligand compared with experiment, but the position of this residue is reasonable. A  $\pi$  $\cdots$  $\pi$  interaction is found between the pyrazine ring of ligand and the benzene ring of Phe113. The two aromatic rings formed a T-shaped conformation. The interaction distance between the two ring centers (4.96 Å) is in agreement with experiment.

As shown in Figure 2, the FMO and MM structures show in general somewhat similar deviations from experiment. However, the NH $\cdots$ N and CH $\cdots$ O hydrogen bonds between the pyrazine moiety of the ligand and Val116 are not well described by the force field. In this particular complex, multiple weak interactions are reasonably described by both FMO/MM and MM methods.

#### 4. CONCLUSIONS

The many-body expansion of properties in fragment-based methods is very useful. Not only does it allow a systematic way to improve the accuracy by increasing the level of the many-body expansion, but it also provides very useful information about the interactions in large molecular systems, which can be further enhanced with the additional analytic techniques.

Many fragment-based methods, which can be called “chemical”, are based on the calculation of individual fragments and the interactions between them, which appeals to chemists,

who like to operate with chemically well-defined subunits; for such methods, the many-body analyses are in particular well suited. An example is FMO, with many useful analyses and applications heavily using the many-body expansion. The other group can be called “physical”; in the first step, fragments are computed, and in the second, their density is used in calculating the energy of the whole system, often with a common Fermi level for all fragments. An example is a variety of divide-and-conquer methods.<sup>12,15,21</sup> The latter category is better suited for describing delocalized systems, and the former is useful in analyzing properties of systems with some localization.

From the practical point of view, for most systems one can apply various methods from both categories. Even delocalized systems such as graphene<sup>118</sup> and boron nitride nanoribbons<sup>61</sup> have been successfully treated with chemical methods, although the focus was on structure rather than the full electronic properties. The choice of a method for actual applications is made based on such considerations as the ease of use, efficiency, and available features in computation software, as well as the availability of graphical user interfaces for making input files and visualizing the results.

In this work, we have summarized and generalized the definition of many-body effects, including a general form for the effective expressions, which provide a contraction of many-body effects in a lower-body formalism. In other words, by using an effective two-body expression, one can discuss complicated many-body effects such as hydrogen bond coupling in a simple two-body representation in the form of a correction to pair interactions (for instance, as coupling corrections to individual hydrogen bond energies). Up to now, these many-body effects have been mainly discussed in terms of the energy, but clearly they can also be considered for other properties as suggested above.

We have developed the FMO/MM method for efficient geometry optimizations of large molecular systems and applied it to the CK2 $\alpha$ –ligand complexes. The optimized structures are in reasonable agreement with experiment. Hydrogen bonding and CH $\cdots$  $\pi$  interactions are important for protein–ligand binding. In addition, other weak interactions such as  $\pi$  $\cdots$  $\pi$  stacking and CH $\cdots$ O bonding also play an important role, determining the conformation and orientation of the ligand. It is important to consider all interactions in a balanced manner to perform geometry optimizations, and FMO/MM can be expected to be a useful method.

Significant progress in geometry optimizations of large molecular systems has been achieved, driven both by the methodological development and the steady growth in computational hardware. There are common problems though, such as the problem of finding the global minimum, and the issue of having to sample the whole energy surface rather than a single minimum. These problems are especially severe for very flexible biological systems such as proteins, typically studied at room temperature in solution. Thus, future development of fragment-based methods should address these concerns for more reliable simulations.

## AUTHOR INFORMATION

### Corresponding Author

\*Author e-mail address: kkitaura@people.kobe-u.ac.jp.

### Funding

We thank the Next Generation Super Computing Project, Nanoscience Program (MEXT, Japan), and Computational Materials Science Initiative (CMSI, Japan) for financial support.

### Notes

The authors declare no competing financial interest.

### Biographies

**Dmitri G. Fedorov** received his M.S. in quantum chemistry at Saint Petersburg State University in Russia in 1993. He got his Ph.D. in physical chemistry at Iowa State University (USA) in 1999, working under the guidance of Prof. Mark Gordon. He spent two years at the University of Tokyo as a JSPS postdoctoral research fellow, working with Kimihiko Hirao and then moved to the National Institute of Advanced Industrial Science and Technology (AIST) in Japan in 2002, where he is currently employed as senior research scientist. His main research interests are focused on developing quantum-mechanical methods for large molecular systems.

**Naoya Asada** received his Ph.D. in 2012 from Kyoto University in Japan under the guidance of Prof. Kazuo Kitaura. Since 2012, he has been employed by SHIONOGI & CO., Ltd.

**Isao Nakanishi** received his Ph.D. from Osaka University, Japan in 1999. After working for Fujisawa Pharmaceutical CO., Ltd from 1985 to 2003, he joined the laboratory of Prof. K. Kitaura at Kyoto University as an associate professor. He had an experience of working with Prof. I. D. Kuntz at Molecular Design Institute of UCSF in 2000. He has been a professor in the department of pharmaceutical sciences at Kinki University since 2008.

**Kazuo Kitaura** received his Ph.D. in quantum chemistry in 1976 from Osaka City University. He worked as assistant professor at IMS and Osaka City University and later became associate professor at IMS and then professor at Osaka Prefecture University. After serving as principal investigator at the National Institute of Advanced Industrial Science and Technology, he became professor at Kyoto University. Currently, he is project professor at Kobe University. His primary research interests are the development of quantum-chemical methods for large molecules and their applications to structures, properties, and reactions of biomolecules.

## REFERENCES

(1) Hu, L.; Söderhjelm, P.; Ryde, U. On the convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.  
(2) Avramov, P. V.; Fedorov, D. G.; Sorokin, P. B.; Chernozatonskii, L. A.; Ovchinnikov, S. G. Quantum dots embedded into silicon nanowires effectively partition electron confinement. *J. Appl. Phys.* **2008**, *104*, No. 054305.

(3) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497*, 643–646.

(4) Gordon, M. S.; Pruitt, S. R.; Fedorov, D. G.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632–672.

(5) Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

(6) Shoemaker, J. R.; Burggraf, L. W.; Gordon, M. S. SIMOMM: An integrated molecular orbital/molecular mechanics optimization scheme for surfaces. *J. Phys. Chem. A* **1999**, *103*, 3245–3251.

(7) *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series 712, American Chemical Society: Washington, DC, 1998.

(8) Svensson, M.; Humble, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and Pt(P(*t*-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(9) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. Development of a 3-body:many-body integrated fragmentation method for weakly bound clusters and application to water clusters (H<sub>2</sub>O)<sub>n=3–10,16,17</sub>. *J. Chem. Phys.* **2011**, *135*, No. 044123.

(10) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.* **2003**, *24*, 760–769.

(11) Otto, P.; Ladik, J. Investigation of the interaction between molecules at medium distances. I. SCF LCAO MO supermolecule, perturbational and mutually consistent calculations for two interacting HF and CH<sub>2</sub>O molecules. *Chem. Phys.* **1975**, *8*, 192–200.

(12) Yang, W. Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **1991**, *66*, 1438–1441.

(13) Gao, J. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997**, *101*, 657–663.

(14) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. Geometry optimizations and vibrational spectra of large molecules from a generalized energy-based fragmentation approach. *J. Phys. Chem. A* **2008**, *112*, 10864–10872.

(15) Kobayashi, M.; Nakai, H. How does it become possible to treat delocalized and/or open-shell systems in fragmentation-based linear-scaling electronic structure calculations? The case of the divide-and-conquer method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7629–7639.

(16) Söderhjelm, P.; Ryde, U. How accurate can a force field become? A polarizable multipole model combined with fragment-wise quantum-mechanical calculations. *J. Phys. Chem. A* **2009**, *113*, 617–627.

(17) Kiewisch, K.; Jacob, C. R.; Visscher, J. Quantum-chemical electron densities of proteins and of selected protein sites from subsystem density functional theory. *J. Chem. Theory Comput.* **2013**, *9*, 2425–2440.

(18) Frank, A.; Möller, H. M.; Exner, T. E. Toward the quantum chemical calculation of NMR chemical shifts of proteins. 2. Level of theory, basis set, and solvents model dependence. *J. Chem. Theory Comput.* **2012**, *8*, 1480–1492.

(19) Collins, M. A. Systematic fragmentation of large molecules by annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744–7751.

(20) Yu, H.; Leverentz, H. R.; Bai, P.; Siepmann, J. I.; Truhlar, D. G. Water 26-mers drawn from bulk simulations: Benchmark binding energies for unprecedentedly large water clusters and assessment of the electrostatically embedded three-body and pairwise additive approximations. *J. Phys. Chem. Lett.* **2014**, *5*, 660–670.

(21) He, X.; Merz, K. M., Jr. Divide and conquer Hartree–Fock calculations on proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.

(22) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective fragment molecular orbital method: A merger of the effective fragment potential



and fragment molecular orbital methods. *J. Phys. Chem. A* **2010**, *114*, 8705–8712.

(23) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.* **2006**, *125*, No. 104109.

(24) Huang, L.; Massa, L. Quantum kernel applications in medicinal chemistry. *Future Med. Chem.* **2012**, *4*, 1479–1494.

(25) Gao, J.; Wang, Y. Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method. *J. Chem. Phys.* **2012**, *136*, No. 071101.

(26) Pulay, P.; Saebo, S.; Meyer, W. An efficient reformulation of the closed-shell self-consistent electron pair theory. *J. Chem. Phys.* **1984**, *81*, 1901–1905.

(27) Scuseria, G. E. Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.

(28) Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical calculations of binding enthalpy for protein–ligand complexes. *Int. J. Quantum Chem.* **2004**, *97*, 747–763.

(29) Stewart, J. J. P. Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765–805.

(30) Koskinen, P.; Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **2009**, *47*, 237–253.

(31) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(32) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.

(33) *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds.; CRC Press: Boca Raton, FL, 2009.

(34) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562–7577.

(35) Tanaka, S.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Fukuzawa, K. Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10310–10344.

(36) Van Alsenoy, C.; Yu, C.-H.; Peeters, A.; Martin, J. M. L.; Schäfer, L. Ab initio geometry determinations of proteins. I. Crambin. *J. Phys. Chem. A* **1998**, *102*, 2246–2251.

(37) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. Density-functional geometry optimization of the 150000-atom photo-system-I trimer. *J. Chem. Phys.* **2006**, *124*, No. 024301.

(38) Liu, K.; Inerbaev, T.; Korchowicz, J.; Gu, F. L.; Aoki, Y. Geometry optimization for large systems by the elongation method. *Theor. Chem. Acc.* **2012**, *131*, 1277–1279.

(39) Guo, W.; Wu, A.; Xua, X. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203–208.

(40) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. Analytic gradients for the multicentred integrated QM:QM method for weakly bound clusters: Efficient and accurate 2-body:many-body geometry optimizations. *Mol. Phys.* **2007**, *105*, 2777–2782.

(41) Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Fully quantum mechanical energy optimization for protein-ligand structure. *J. Comput. Chem.* **2004**, *25*, 1431–1437.

(42) Bates, D. M.; Smith, J. R.; Tschumper, G. S. Efficient and accurate methods for the geometry optimization of water clusters: Application of analytic gradients for the two-body:many-body QM:QM fragmentation method to  $(\text{H}_2\text{O})_n$ ,  $n = 3-10$ . *J. Chem. Theory Comput.* **2011**, *7*, 2753–2760.

(43) Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. WebMTA: A web-interface for ab initio geometry optimization of large molecules using molecular tailoring approach. *J. Comput. Chem.* **2009**, *30*, 1167–1173.

(44) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. Fragment molecular orbital method: analytical energy gradients. *Chem. Phys. Lett.* **2001**, *336*, 163–170.

(45) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.* **2011**, *134*, No. 124115.

(46) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. The fragment molecular orbital method for geometry optimizations of polypeptides and proteins. *J. Phys. Chem. A* **2007**, *111*, 2722–2732.

(47) Pruitt, S. R.; Fedorov, D. G.; Gordon, M. S. Geometry optimizations of open-shell systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2012**, *116*, 4965–4974.

(48) Ishikawa, T.; Yamamoto, N.; Kuwata, K. Partial energy gradient based on the fragment molecular orbital method: Application to geometry optimization. *Chem. Phys. Lett.* **2010**, *500*, 149–154.

(49) Fedorov, D. G.; Alexeev, Y.; Kitaura, K. Geometry optimization of the active site of a large system with the fragment molecular orbital method. *J. Phys. Chem. Lett.* **2011**, *2*, 282–288.

(50) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective fragment molecular orbital method: A merger of the effective fragment potential and fragment molecular orbital methods. *J. Phys. Chem. A* **2010**, *114*, 8705–8712.

(51) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Mapping enzymatic catalysis using the effective fragment molecular orbital method: Towards all ab initio biochemistry. *PLoS One* **2013**, *8*, No. e60602.

(52) Christensen, A. S.; Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Hybrid RHF/MP2 geometry optimizations with the effective fragment molecular orbital method. *PLoS One* **2014**, *9*, No. e88800.

(53) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully integrated effective fragment molecular orbital method. *J. Chem. Theory Comput.* **2013**, *9*, 2235–2249.

(54) Okamoto, T.; Ishikawa, T.; Koyano, Y.; Yamamoto, N.; Kuwata, K.; Nagaoka, M. A minimal implementation of the AMBER-PAICS interface for ab initio FMO-QM/MM-MD simulation. *Bull. Chem. Soc. Jpn.* **2013**, *86*, 210–222.

(55) Nakata, H.; Nagata, T.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Analytic second derivatives of the energy in the fragment molecular orbital method. *J. Chem. Phys.* **2013**, *138*, No. 164103.

(56) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Efficient vibrational analysis for unrestricted Hartree-Fock based on the fragment molecular orbital method. *Chem. Phys. Lett.* **2014**, *603*, 67–74.

(57) Fedorov, D. G.; Kitaura, K. Use of an auxiliary basis set to describe the polarization in the fragment molecular orbital method. *Chem. Phys. Lett.* **2014**, *597*, 99–105.

(58) Tsukamoto, T.; Mochizuki, Y.; Watanabe, N.; Fukuzawa, K.; Nakano, T. Partial geometry optimization with FMO-MP2 gradient: Application to TrpCage. *Chem. Phys. Lett.* **2012**, *535*, 157–162.

(59) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Simulations of Raman spectra using the fragment molecular orbital method. *J. Chem. Theory Comput.* **2014**, DOI: 10.1021/ct5003829.

(60) Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. Analytic gradient for the adaptive frozen orbital bond detachment in the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *477*, 169–175.

(61) Avramov, P. V.; Fedorov, D. G.; Sorokin, P. B.; Sakai, S.; Entani, S.; Ohtomo, M.; Matsumoto, Y.; Naramoto, H. Intrinsic edge asymmetry in narrow zigzag hexagonal heteroatomic nanoribbons causes their subtle uniform curvature. *J. Phys. Chem. Lett.* **2012**, *3*, 2003–2008.

(62) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Derivatives of the approximated electrostatic potentials in unrestricted Hartree-Fock based on the fragment molecular orbital method and an application to polymer radicals. *Theor. Chem. Acc.* **2014**, *133*, 1477.

- (63) Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
- (64) Deaven, D. M.; Ho, K. M. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **1995**, *75*, 288–291.
- (65) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water molecule interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.
- (66) Stoll, H.; Preuß, H. On the direct calculation of localized HF orbitals in molecule clusters, layers and solids. *Theor. Chim. Acta* **1977**, *46*, 11–21.
- (67) Stoll, H. Correlation energy of diamond. *Phys. Rev. B* **1992**, *46*, 6700–6704.
- (68) Fedorov, D. G.; Kitaura, K. Pair interaction energy decomposition analysis. *J. Comput. Chem.* **2007**, *28*, 222–237.
- (69) Fedorov, D. G.; Kitaura, K. The importance of three-body terms in the fragment molecular orbital method. *J. Chem. Phys.* **2004**, *120*, 6832–6840.
- (70) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: use of approximate electrostatic potential. *Chem. Phys. Lett.* **2002**, *351*, 475–480.
- (71) Fedorov, D. G.; Kitaura, K. The three-body fragment molecular orbital method for accurate calculations of large systems. *Chem. Phys. Lett.* **2006**, *433*, 182–187.
- (72) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree - Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.
- (73) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. Pair interaction molecular orbital method: An approximate computational method for molecular interactions. *Chem. Phys. Lett.* **1999**, *312*, 319–324.
- (74) Chen, W.; Gordon, M. S. Energy decomposition analyses for many-body interaction and applications to water complexes. *J. Chem. Phys.* **1996**, *100*, 14316–14328.
- (75) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic study of the embedding potential description in the fragment molecular orbital method. *J. Phys. Chem. A* **2010**, *114*, 8742–8753.
- (76) Fujita, T.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Accuracy of fragmentation in ab initio calculations of hydrated sodium cation. *Chem. Phys. Lett.* **2009**, *478*, 295–300.
- (77) Yasuda, K.; Yamaki, D. The extension of the fragment molecular orbital method with the many-particle Green's function. *J. Chem. Phys.* **2006**, *125*, No. 154101.
- (78) Fedorov, D. G.; Kitaura, K. The role of the exchange in the embedding electrostatic potential for the fragment molecular orbital method. *J. Chem. Phys.* **2009**, *131*, No. 171106.
- (79) Nakano, T.; Mochizuki, Y.; Yamashita, K.; Watanabe, C.; Fukuzawa, K.; Segawa, K.; Okiyama, Y.; Tsukamoto, T.; Tanaka, S. Development of the four-body corrected fragment molecular orbital (FMO4) method. *Chem. Phys. Lett.* **2012**, *523*, 128–133.
- (80) Paulus, B. The method of increments - a wavefunction-based ab initio correlation method for solids. *Phys. Rep.* **2006**, *428*, 1–52.
- (81) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. Fast electron correlation methods for molecular clusters in the ground and excited states. *Mol. Phys.* **2005**, *103*, 2255–2265.
- (82) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.
- (83) Suárez, E.; Díaz, N.; Suárez, D. Thermochemical fragment energy method for biomolecules: application to a collagen model peptide. *J. Chem. Theory Comput.* **2009**, *5*, 1667–1679.
- (84) Huang, L.; Massa, L.; Karle, J. The kernel energy method of quantum mechanical approximation carried to fourth-order terms. *Proc. Nat. Acad. Sciences U.S.A.* **2008**, *105*, 1849–1854.
- (85) Richard, R. M.; Herbert, J. M. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.* **2012**, *137*, No. 064113.
- (86) Mayhall, N. J.; Raghavachari, K. Many-overlapping-body (MOB) expansion: A generalized many body expansion for non-disjoint monomers in molecular fragmentation calculations of covalent molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.
- (87) Dang, L. X.; Chang, T.-M. Molecular dynamics study of water clusters, liquid, and liquid–vapor interface of water with many-body potentials. *J. Chem. Phys.* **1997**, *106*, 8149–8159.
- (88) Xantheas, S. S. Ab initio studies of cyclic water clusters (H<sub>2</sub>O)<sub>n</sub>, n = 1–6. II. Analysis of many-body interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.
- (89) Gregory, J. K.; Clary, D. C. Structure of water clusters. The contribution of many-body forces, monomer relaxation, and vibrational zero-point energy. *J. Phys. Chem.* **1996**, *100*, 18014–18022.
- (90) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. *J. Phys. Chem. A* **1997**, *101*, 9163–9168.
- (91) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. Hydrogen bonding in water clusters: Pair and many-body interactions from symmetry-adapted perturbation theory. *J. Phys. Chem. A* **1999**, *103*, 6811–6819.
- (92) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M., Jr. Are many-body effects important in protein folding? *J. Phys. Chem. B* **2000**, *104*, 9554–9563.
- (93) Ji, C. G.; Zhang, J. Z. H. Electronic polarization is important in stabilizing the native structures of proteins. *J. Phys. Chem. B* **2009**, *113*, 16059–16064.
- (94) Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster. *J. Chem. Phys.* **2001**, *114*, 661–681.
- (95) Fedorov, D. G.; Ishida, T.; Kitaura, K. Multilayer formulation of the fragment molecular orbital method (FMO). *J. Phys. Chem. A* **2005**, *109*, 2638–2646.
- (96) Wang, Y.; Sosa, C. P.; Cembran, A.; Truhlar, D. G.; Gao, J. Multilevel X-Pol: A fragment-based method with mixed quantum mechanical representations of different fragments. *J. Phys. Chem. B* **2012**, *116*, 6781–6788.
- (97) Mochizuki, Y.; Nakano, T.; Komeiji, Y.; Yamashita, K.; Okiyama, Y.; Yoshikawa, H.; Yamataka, H. Fragment molecular orbital-based molecular dynamics (FMO-MD) method with MP2 gradient. *Chem. Phys. Lett.* **2011**, *504*, 95–99.
- (98) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (99) Nafziger, J.; Wu, Q.; Wasserman, A. Molecular binding energies from partition density functional theory. *J. Chem. Phys.* **2011**, *135*, No. 234101.
- (100) Gao, J.; Cembran, A.; Mo, Y. Generalized X-Pol theory and charge delocalization states. *J. Chem. Theory Comput.* **2010**, *6*, 2402–2410.
- (101) Churchill, C. D. M. The fragment molecular orbital method and understanding monomer polarization. *Chem. Phys. Lett.* **2012**, *554*, 185–189.
- (102) Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. Configuration interaction singles method with multilayer fragment molecular orbital scheme. *Chem. Phys. Lett.* **2005**, *406*, 283–288.
- (103) Watanabe, C.; Fukuzawa, K.; Okiyama, Y.; Tsukamoto, T.; Kato, A.; Tanaka, S.; Mochizuki, Y.; Nakano, T. Three- and four-body corrected fragment molecular orbital calculations with a novel subdividing fragmentation method applicable to structure-based drug design. *J. Mol. Graphics Modell.* **2013**, *41*, 31–42.
- (104) Kamiya, M.; Hirata, S.; Valiev, M. Fast electron correlation methods for molecular clusters without basis set superposition errors. *J. Chem. Phys.* **2008**, *128*, No. 074103.
- (105) Ishikawa, T.; Ishikura, T.; Kuwata, K. Theoretical study of the prion protein based on the fragment molecular orbital method. *J. Comput. Chem.* **2009**, *30*, 2594–2601.
- (106) Asada, N.; Fedorov, D. G.; Kitaura, K.; Nakanishi, I.; Merz, K. M., Jr. An efficient method to evaluate intermolecular interaction energies in large systems using overlapping multicenter ONIOM and

the fragment molecular orbital method. *J. Phys. Chem. Lett.* **2012**, *3*, 2604–2610.

(107) Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *J. Cheminf.* **2011**, *3*, No. 2.

(108) Fedorov, D. G.; Kitaura, K. Energy decomposition analysis in solution based on the fragment molecular orbital method. *J. Phys. Chem. A* **2012**, *116*, 704–719.

(109) Tanaka, S.; Watanabe, C.; Okiyama, Y. Statistical correction to effective interactions in the fragment molecular orbital method. *Chem. Phys. Lett.* **2013**, *556*, 272–277.

(110) Mochizuki, Y.; Fukuzawa, K.; Kato, A.; Tanaka, S.; Kitaura, K.; Nakano, T. A configuration analysis for fragment interaction. *Chem. Phys. Lett.* **2005**, *410*, 247–253.

(111) Ishikawa, T.; Mochizuki, Y.; Amari, S.; Nakano, T.; Tokiwa, H.; Tanaka, S.; Tanaka, K. Fragment interaction analysis based on local MP2. *Theor. Chem. Acc.* **2007**, *118*, 937–945.

(112) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(113) Ponder, J. W.; Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8*, 1016–1024.

(114) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A new hierarchical parallelization scheme: Generalized distributed data interface (GDDI), and an application to the fragment molecular orbital method (FMO). *J. Comput. Chem.* **2004**, *25*, 872–880.

(115) Asada, N. Application of quantum-chemical methods to drug design. Ph.D. thesis, Kyoto University, 2011 (in Japanese).

(116) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: Application to polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614–618.

(117) Grimme, S. Semiempirical GGA - type density functional constructed with a long - range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(118) Yeole, S. D.; Gadre, S. R. On the applicability of fragmentation methods to conjugated p systems within density functional framework. *J. Chem. Phys.* **2010**, *132*, No. 094102.

(119) Komeiji, Y.; Mochizuki, Y.; Nakano, T. Three-body expansion and generalized dynamic fragmentation improve the fragment molecular orbital-based molecular dynamics (FMO-MD). *Chem. Phys. Lett.* **2010**, *484*, 380–386.

(120) Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. Definition of molecular orbitals in fragment molecular orbital method. *Chem. Phys. Lett.* **2002**, *364*, 139–143.

(121) Sekino, H.; Sengoku, Y.; Sugiki, S.-I.; Kurita, N. Molecular orbital analysis based on fragment molecular orbital scheme. *Chem. Phys. Lett.* **2003**, *378*, 589–597.

(122) Fischer, B.; Fukuzawa, K.; Wenzel, W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1264–1273.

(123) Mochizuki, Y.; Ishikawa, T.; Tanaka, K.; Tokiwa, H.; Nakano, T.; Tanaka, S. Dynamic polarizability calculation with fragment molecular orbital scheme. *Chem. Phys. Lett.* **2006**, *418*, 418–422.

(124) Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (FMO). *J. Comput. Chem.* **2006**, *27*, 976–985.

(125) Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. DNA and estrogen receptor interaction revealed by fragment molecular orbital calculations. *J. Phys. Chem. B* **2007**, *111*, 9621–9627.

(126) Kobori, T.; Sodeyama, K.; Otsuka, T.; Tateyama, Y.; Tsuneyuki, S. Trimer effects in fragment molecular orbital-linear

combination of molecular orbitals calculation of one-electron orbitals for biomolecules. *J. Chem. Phys.* **2013**, *139*, No. 094113.

(127) Sekino, H.; Matsumura, N.; Sengoku, Y. Evaluation of NMR chemical shift by fragment molecular orbital method. *Comput. Lett.* **2007**, *3*, 423–430.