

# Simulations of Chemical Reactions with the Frozen Domain Formulation of the Fragment Molecular Orbital Method

Hiroya Nakata,<sup>†,‡,§</sup> Dmitri G. Fedorov,<sup>\*,§</sup> Takeshi Nagata,<sup>||,§,⊥</sup> Kazuo Kitaura,<sup>||</sup> and Shinichiro Nakamura<sup>‡</sup>

<sup>†</sup>Department of Biomolecular Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan

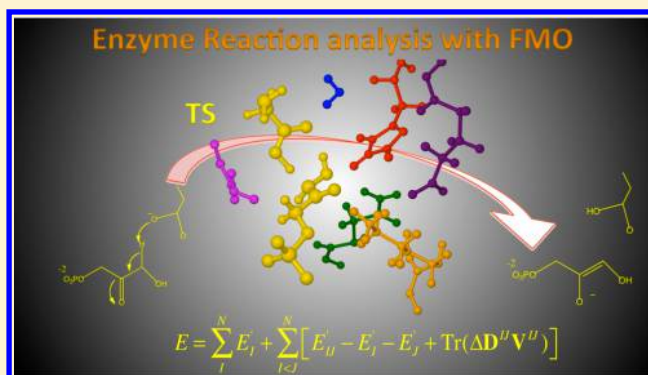
<sup>‡</sup>Research Cluster for Innovation, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

<sup>#</sup>Japan Society for the Promotion of Science, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan

<sup>§</sup>Nanosystem Research Institute, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

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

**ABSTRACT:** The fully analytic first and second derivatives of the energy in the frozen domain formulation of the fragment molecular orbital (FMO) were developed and applied to locate transition states and determine vibrational contributions to free energies. The development is focused on the frozen domain with dimers (FDD) model. The intrinsic reaction coordinate method was interfaced with FMO. Simulations of IR and Raman spectra were enabled using FMO/FDD by developing the calculation of intensities. The accuracy is evaluated for  $S_N2$  reactions in explicit solvent, and for the free binding energies of a protein–ligand complex of the Trp cage protein (PDB: 1L2Y). FMO/FDD is applied to study the keto–enol tautomeric reaction of phosphoglycolohydroxamic acid and the triosephosphate isomerase (PDB: 7TIM), and the role of amino acid residue fragments in the reaction is discussed.



## 1. INTRODUCTION

The study of chemical reaction is the central subject in chemistry. In contrast to chemical reactions in gas phase, which often can be modeled by small clusters of molecules, chemical reactions in liquid and solid states typically involve large molecular systems, and a straightforward application of quantum mechanics (QM) methods is difficult.

A practical compromise between cost and accuracy is given by hybrids of QM and molecular mechanics (MM);<sup>1–4</sup> one can also represent the bulk part of a protein by a continuum model.<sup>5,6</sup> Applications of purely QM approaches to large molecular systems can be divided into three groups: (a) linear scaling methods,<sup>7–10</sup> (b) various partitioning schemes, involving subsystems or fragments,<sup>11–29</sup> and (c) model approaches, often of semiempirical nature.<sup>30,31</sup>

Despite very significant progress in both method development and computation hardware efficiency, an application of fully *ab initio* methods to the study of chemical reactions of large molecular systems is still difficult, as it requires large computational resources as well as necessary methodological and software development to evaluate the matrix of second derivatives (the Hessian), do transition state search, map a reaction path, or perform molecular dynamics (MD)

simulations. A number of partial schemes have emerged, which describe the whole system quantum-mechanically, but reduce the number of degrees of freedom in terms of atomic coordinates, by building a partial Hessian for a subset of atoms.<sup>32–35</sup>

The fragment molecular orbital (FMO) method<sup>36–40</sup> is a fragmentation approach, in which a molecular system is divided into a set of fragments. The energy, its gradient, or Hessian of each fragment are calculated in the presence of the embedding electrostatic potential (ESP). The ESP is determined from the density of fragments, and it is self-consistently updated until the electronic states of fragments mutually converge. After that fragment pair or triple calculations are performed in the embedding ESP, which is determined from the electron densities of monomers. The analytic energy gradient<sup>41–43</sup> and the second derivative of the energy<sup>44,45</sup> have been developed. The FMO method has been applied to proteins,<sup>46–48</sup> DNA,<sup>49</sup> and inorganic systems.<sup>50–53</sup>

The FMO method can be used for geometry optimizations,<sup>54,55</sup> transition state searches,<sup>44</sup> and MD simulations.<sup>56–58</sup>

**Received:** March 23, 2015

**Published:** June 4, 2015

Full geometry optimizations of small proteins<sup>54,59,60</sup> and relatively short MD of chemical reactions<sup>61,62</sup> have been reported using FMO. However, geometry optimizations of flexible proteins consisting of many thousands of atoms take many optimization steps to converge and are expensive.

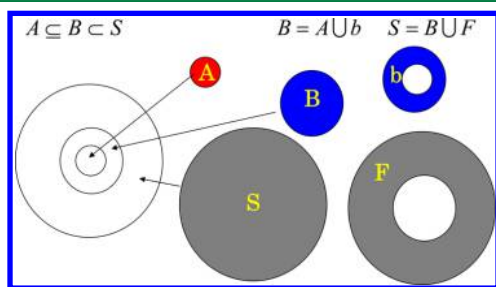
To improve efficiency of a QM geometry optimization,<sup>63</sup> one can freeze some part of the system, and only optimize the rest. In the FMO framework, partial optimizations<sup>64,65</sup> have been performed for protein–ligand complexes.<sup>66</sup> FMO and effective FMO (EFMO)<sup>67,68</sup> constrained optimizations of enzymatic reactions have been done to determine reaction paths.<sup>69,70</sup> FMO has also been used to refine enzymatic reaction energetics for structures obtained with other methods.<sup>71–73</sup>

In this work, first previously<sup>65</sup> neglected terms in the analytic gradient are derived and implemented for the frozen domain with dimers (FDD) formulation of FMO (FMO/FDD), achieving a fully analytic gradient. Second, analytic second derivatives are developed for FMO/FDD. Third, FMO is interfaced with the intrinsic reaction coordinate (IRC) method.<sup>74–76</sup> Using FMO-based IRC, reaction paths are mapped for several chemical reactions. IRC is a general approach applicable when the reaction coordinate is unknown, compared to the alternative of constrained geometry optimizations for a reaction coordinate known a priori.<sup>69,70</sup> The accuracy tests in comparison with full QM are performed for S<sub>N</sub>2 reactions. Finally, FMO/FDD is applied to map the path and simulate infrared (IR) and Raman spectra for a chemical reaction involving triosephosphate isomerase (TIM).<sup>77–79</sup>

## 2. THEORY

**2.1. Analytic Energy Gradient for FMO/FDD.** A detailed description of the FMO methodology can be found elsewhere.<sup>37,38</sup> Briefly, in FMO a system is divided into a set of fragments, which are calculated in the embedding ESP, representing the Coulomb potential of the whole system. Individual fragment monomer calculations are followed by fragment pairs (dimers), and, optionally, trimers. Here, two-body expansion FMO2 is used, combined with the hybrid orbital projection operators for treating interfragment boundaries.<sup>80</sup> Throughout this paper, restricted Hartree–Fock (RHF) is used. It has been argued that for proteins it is a good approach for geometry optimizations.<sup>81</sup>

A detailed description of FMO/FDD is given elsewhere,<sup>65</sup> and here it is briefly outlined (see Figure 1). FMO/FDD uses a multilayer formulation of FMO,<sup>82</sup> with two layers,  $L_1$  and  $L_2$ , into which all fragments are assigned. For the initial structure, a



**Figure 1.** Definition of regions in FMO/FDD: total system (S), frozen (F), polarizable buffer (B) and active (A) domains, shown combined and separately for clarity. Domain *b* is defined as B excluding A. Layers 1 and 2 coincide with domains F and B, respectively.

monomer calculation (FMO1) for the whole system S is performed at the level of  $L_1$  in order to obtain the fragment densities. Then, for the initial and consequent geometry optimizations, FMO2 calculations of the polarizable domain B are performed at the level of layer  $L_2$  in the presence of the embedding ESP for the whole system S, that includes the contributions from fragments in F and B; the former are frozen and the latter updated for each geometry. Thus, an FMO/FDD calculation, variationally speaking, is a one-layer computation (layer  $L_2$ ) of B using constant embedding ( $L_1$ ) from F (see Figure 1). The active domain A, whose coordinates are permitted to change during geometry optimization, forms the central part of B.

The total energy of FMO/FDD is defined as<sup>65</sup>

$$E^{\text{FMO/FDD}} = \sum_{I \in \mathbf{B}} E_I^{L_2} + \sum_{\substack{I > J \\ I \in \mathbf{A}, J \in \mathbf{B}}} \Delta E_{IJ}^{L_2} + \sum_{I \in \mathbf{A}, J \in \mathbf{F}} \Delta E_{IJ}^{L_2, L_1} \quad (1)$$

where  $E_X^{L_2}$  ( $X = I, J$ , or  $IJ$ ) is the internal fragment energy of  $X$ , i.e., the RHF energy  $E_X^{L_2}$  of  $X$  from which the embedding energy is subtracted *a posteriori*.

$$\begin{aligned} E_X^{L_2} &= E_X^{L_1} - \text{Tr}(\mathbf{D}^X \mathbf{V}^X) \\ &= \sum_{\mu, \nu} D_{\mu\nu}^X h_{\mu\nu}^X \\ &\quad + \frac{1}{2} \sum_{\mu, \nu, \lambda, \sigma} D_{\mu\nu}^X D_{\lambda\sigma}^X \left[ (\mu\nu|\sigma\lambda) - \frac{1}{2} (\mu\sigma|\nu\lambda) \right] + E_{\text{NR}}^X \end{aligned} \quad (2)$$

Here,  $\mathbf{D}^X$  and  $\mathbf{V}^X$  are the density and ESP of  $X$ , respectively. The layer is not explicitly indicated in  $\mathbf{D}^X$  and  $\mathbf{V}^X$ : If  $X = I$  is in layer  $L_1$ , then it is  $L_1$ , and  $L_2$  otherwise. To avoid confusion, the layer is shown in fragment energies.  $\mu, \nu, \lambda$ , and  $\sigma$  run over atomic orbitals,  $h_{\mu\nu}^X$  and  $(\mu\nu|\sigma\lambda)$  are the core and two-electron integrals, respectively.  $E_{\text{NR}}^X$  is the nuclear repulsion energy.  $\Delta E_{IJ}^{L_2}$  is the RHF dimer contribution for two nearby fragments,  $I$  and  $J$ :

$$\Delta E_{IJ}^{L_2} = E_{IJ}^{L_2} - E_I^{L_2} - E_J^{L_2} + \sum_{\mu, \nu} \Delta D_{\mu\nu}^{IJ} V_{\mu\nu}^{IJ} \quad (3)$$

where  $\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^I \oplus \mathbf{D}^J)$  is the density transfer matrix. The electrostatic (ES) interaction between far separated fragments  $I$  and  $J$  is given by

$$\begin{aligned} \Delta E_{IJ}^{L_2, L_1} &= \sum_{\mu, \nu \in I} \sum_{A \in J} D_{\mu\nu}^I \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle \\ &\quad + \sum_{\mu, \nu \in J} \sum_{A \in I} D_{\mu\nu}^J \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle \\ &\quad + \sum_{\mu, \nu \in I} \sum_{\lambda, \sigma \in J} D_{\mu\nu}^I D_{\lambda\sigma}^J (\mu\nu|\lambda\sigma) \end{aligned} \quad (4)$$

where  $Z_A$  is the nuclear charge, and  $\mathbf{r}$  and  $\mathbf{R}_A$  are the coordinates of electron and atom  $A$ , respectively. The embedding ESP  $V_{\mu\nu}^{IJ}$  for dimer  $IJ$  is the sum of all fragment contributions for  $K \neq I, J$ :

$$V_{\mu\nu}^{IJ} = \sum_{K \neq I, J} V_{\mu\nu}^{IJ(K)} \quad (5)$$

$$V_{\mu\nu}^{X(I)} = \sum_{A \in J} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle + \sum_{\lambda, \sigma \in J} D_{\lambda\sigma}^I(\mu\nu|\lambda\sigma) \quad (6)$$

The total energy in eq 1 can be divided into internal energy (int) and ESP-related contributions,

$$E^{\text{FMO/FDD}} = E^{\text{int}} + E^{\text{ESP}} \quad (7)$$

where

$$E^{\text{int}} = \sum_{I \in \mathbf{B}} E_I'^{L_2} + \sum_{I \in \mathbf{A}, J \in \mathbf{B}} (E_{IJ}'^{L_2} - E_I'^{L_2} - E_J'^{L_2}) \quad (8)$$

$$E^{\text{ESP}} = \sum_{I > J} \text{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) + \sum_{I \in \mathbf{A}, J \in \mathbf{F}} \Delta E_{IJ}'^{L_2, L_1} \quad (9)$$

The first sum over  $I$  in  $E^{\text{int}}$  is referred to as the FMO1 internal energy below. The derivative of  $E^{\text{FMO/FDD}}$  with respect to a nuclear coordinate  $a$  is

$$\frac{\partial E^{\text{FMO/FDD}}}{\partial a} = E^{\text{int}, a} + E^{\text{ESP}, a} + R^a \quad (10)$$

where atomic coordinates  $a$  in the above equation are for atoms in fragments in  $\mathbf{B}$ .

The derivatives of the internal energy  $E^{\text{int}, a}$  are given by the gradients of the internal energies  $E_X'$  as described in detail elsewhere.<sup>42</sup> Here, the other two terms specific to FMO/FDD are described in detail,

$$E^{\text{ESP}, a} = \sum_{I > J} \sum_{\mu\nu \in IJ} \Delta D_{\mu\nu}^{IJ} V_{\mu\nu}^{a, IJ} + \sum_{I \in \mathbf{A}, J \in \mathbf{F}} \frac{\partial \Delta E_{IJ}'^{L_2, L_1}}{\partial a} \quad (11)$$

$$\begin{aligned} \frac{\partial \Delta E_{IJ}'^{L_2, L_1}}{\partial a} = & \sum_{\mu, \nu \in I} \sum_{A \in J} D_{\mu\nu}^I \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle^a \\ & + \sum_{\mu, \nu \in J} \sum_{A \in I} D_{\mu\nu}^J \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle^a \\ & + \sum_{\mu, \nu \in I} \sum_{\lambda, \sigma \in J} D_{\mu\nu}^I D_{\lambda\sigma}^J (\mu\nu|\lambda\sigma)^a \end{aligned} \quad (12)$$

where  $V_{\mu\nu}^{a, IJ}$  is the integral derivative contribution to the derivative of the ESP in eq 6:

$$V_{\mu\nu}^{a, IJ} = \sum_{K \neq I, J} \left[ \sum_{A \in K} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle^a + \sum_{\lambda, \sigma} D_{\lambda\sigma}^K (\mu\nu|\lambda\sigma)^a \right] \quad (13)$$

$R^a$  is the total response contribution to the gradient including response terms  $U_{ri}^{a, K}$ :

$$\begin{aligned} R^a = & \sum_{I > J} \left[ - \sum_{m \in I} \sum_{i \in I}^{\text{vir}} \sum_{i \in I}^{\text{occ}} 4U_{mi}^{a, I} (V_{mi}^{IJ} - V_{mi}^I) \right. \\ & - \sum_{m \in J} \sum_{i \in J}^{\text{vir}} \sum_{i \in J}^{\text{occ}} 4U_{mi}^{a, J} (V_{mi}^{IJ} - V_{mi}^J) \\ & + \sum_{\mu, \nu \in IJ} \sum_{K \neq I, J} \sum_{m \in K} \sum_{i \in K}^{\text{vir}} \sum_{i \in K}^{\text{occ}} 4U_{mi}^{a, K} \Delta D_{\mu\nu}^{IJ} (\mu\nu|mi) \\ & - \sum_I \sum_{m \in I} \sum_{i \in I}^{\text{vir}} \sum_{i \in I}^{\text{occ}} 4U_{mi}^{a, I} V_{mi}^I \\ & \left. + \sum_{I \in \mathbf{A}, J \in \mathbf{F}} \left[ \sum_{m \in I} \sum_{i \in I}^{\text{vir}} \sum_{i \in I}^{\text{occ}} 4U_{mi}^{a, I} V_{mi}^{I(I)} + \sum_{m \in J} \sum_{i \in J}^{\text{vir}} \sum_{i \in J}^{\text{occ}} 4U_{mi}^{a, J} V_{mi}^{J(I)} \right] \right] \quad (14) \end{aligned}$$

If the point-charge (PC) approximation is used to calculate the ESP contributions of far separated fragments (ESP-PC),<sup>83</sup> all terms in eq 14 should be calculated. Moreover, ESP themselves, their derivatives as well as the equations to obtain the responses  $U_{mi}^{a, I}$  have to be modified for the ESP-PC approximation, as described in detail elsewhere.<sup>43</sup> In this paper, we only derived and implemented the ESP-PC approximation for the analytic gradient, whereas for the Hessian it has not been derived due to the complexity of the response terms. The approximation will be useful for geometry optimizations using FMO/FDD.

Without the ESP-PC approximation, there are cancellations of ESP contributions  $U_{mi}^{a, I} V_{mi}^I$  and the response contributions in eq 14 are

$$\begin{aligned} R^a = & \sum_{I \notin \mathbf{A}} \sum_{m \in I} \sum_{i \in I}^{\text{vir}} \sum_{i \in I}^{\text{occ}} 4U_{mi}^{a, I} V_{mi}^I \\ & + \sum_{I > J} \sum_{\mu, \nu \in IJ} \sum_{K \neq I, J} \sum_{m \in K} \sum_{i \in K}^{\text{vir}} \sum_{i \in K}^{\text{occ}} 4U_{mi}^{a, K} \Delta D_{\mu\nu}^{IJ} (\mu\nu|mi) \end{aligned} \quad (15)$$

When the ESP-PC approximation is used, numeric tests show the total response terms are very important and should not be neglected. Without the approximation they are smaller but still needed for the fully analytic gradient. As discussed in elsewhere,<sup>43</sup> even with the inclusion of the response terms, in order to get accurate FMO gradient with the ESP-PC approximation, sufficiently large values of the ESP-PC threshold should be used.

**2.2. Analytic Hessian for FMO/FDD.** The second derivative of the energy with respect to nuclear coordinates  $a$  and  $b$  is

$$\begin{aligned} \frac{\partial^2 E^{\text{FMO/FDD}}}{\partial a \partial b} = & \sum_{I \in \mathbf{B}} \frac{\partial^2 E_I'^{L_2}}{\partial a \partial b} + \sum_{I > J} \frac{\partial^2 \Delta E_{IJ}'^{L_2}}{\partial a \partial b} \\ & + \sum_{I \in \mathbf{A}, J \in \mathbf{F}} \frac{\partial^2 \Delta E_{IJ}'^{L_2, L_1}}{\partial a \partial b} \end{aligned} \quad (16)$$

Most terms in the FMO/FDD second derivative are the same as in the general FMO Hessian,<sup>44,45</sup> and they are omitted here

for brevity, except for one term described in some detail below, as it is specific to FMO/FDD. It should be noted that similar to previous work<sup>44</sup> some terms in the derived fully analytic Hessian were not implemented, whose magnitude is expected to be small, in particular, second-order responses.

From the derivative with respect to  $b$  of  $U_{mi}^{a,I} V_{mi}^I$  in eq 15 (in the Hessian calculations the ESP-PC approximation is not used), one obtains the following term,<sup>44,45</sup>

$$\begin{aligned} \bar{U}^{ab} = & \sum_{I \in B} \bar{U}^{ab,I,I} \\ & + \sum_{I > J} (\bar{U}^{ab,I,IJ} + \bar{U}^{ab,J,IJ} - \bar{U}^{ab,I,I} - \bar{U}^{ab,J,J}) \\ & - \sum_{I \in A, J \in F} (\bar{U}^{ab,I,I(J)} + \bar{U}^{ab,J,J(I)}) \end{aligned} \quad (17)$$

where for  $Y = I, J$ , and  $IJ$  the contributions  $\bar{U}^{ab,X,Y(Z)}$  of fragment  $Z$  to their sum  $\bar{U}^{ab,X,Y}$  are

$$\begin{aligned} \bar{U}^{ab,X,Y} = & \sum_{Z \neq Y} \bar{U}^{ab,X,Y(Z)} \\ \bar{U}^{ab,X,Y(Z)} = & -4 \sum_{i \in X} \sum_{j \in X} \sum_{m \in X}^{\text{occ}} U_{mi}^{b,X} (V_{mj}^{Y(Z)} S_{ij}^{a,X} + V_{ij}^{Y(Z)} S_{jm}^{a,X}) \\ & + \sum_{i \in X} \sum_{j \in X} \sum_{k \in X}^{\text{occ}} S_{ki}^{b,X} V_{ij}^{Y(Z)} S_{jk}^{a,X} \\ & - \sum_{i \in X} \sum_{j \in X}^{\text{occ}} S_{ki}^{ab,X} V_{ij}^{Y(Z)} + \sum_{m \in X} \sum_{i \in X}^{\text{vir}} 4 \frac{\partial V_{mi}^X U_{mi}^{a,X}}{\partial b} \end{aligned} \quad (18)$$

Without the ESP-PC approximation, many response contributions to the second derivative of energy cancel out, and one obtains the following equation:

$$\bar{U}^{ab} = \sum_{I \in B, I \notin A} \bar{U}^{ab,I,I} - \sum_{I \in A, J \notin A} \bar{U}^{ab,J,J(I)} \quad (20)$$

In this study, the Hessian matrix is constructed following the partial Hessian vibrational analysis,<sup>33,84</sup> where only the Hessian for atoms in the active domain **A**,  $\mathbf{H}_{AA}$ , is explicitly constructed and diagonalized. This is equivalent to padding the Hessian in the polarizable domain **B** as follows:

$$\begin{aligned} \mathbf{H} = & \begin{pmatrix} \mathbf{H}_{AA} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{bb} \end{pmatrix}, \\ \text{where } \mathbf{H}_{bb} = & \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \epsilon \end{pmatrix}, \quad \epsilon = 10^{-8} \text{ au} \end{aligned} \quad (21)$$

Because only Hessian elements for atoms  $a$  and  $b$  in **A** are explicitly required, the evaluation of  $\bar{U}^{ab}$  in eq 20 is not needed.

In IRC, the initial Hessian for atoms in **B** is constructed according to eq 21, and consequently updated using the analytic gradient for atoms in **B**.

**2.3. IR and Raman Spectra with FMO/FDD.** The IR intensities and Raman activities for FMO/FDD were derived in a trivial extension of the general FMO formulation.<sup>60</sup> Clearly, there are limitations as to what spectral features can and cannot be captured with FMO/FDD. Namely, FMO/FDD can only describe vibrations local to domain **A** in the embedding

potential of the whole system. The interactions between domains **A** and **b** as well as **A** and **F** also contribute to the Hessian, but the vibrations are strongly localized in **A**.

Therefore, FMO/FDD cannot be used to simulate modes delocalized over the whole protein, which should be clear anyway, given the formulation of FMO/FDD. However, the method is very efficient in treating local vibrational modes.

**2.4. Extension of the Gradient to FMO/FD.** In addition to the FMO/FDD analytic gradient, FMO/FD gradient is also developed in this work. FMO/FD differs from FMO/FDD in the inclusion of additional dimers. Namely, the total energy of FMO/FD is<sup>65</sup>

$$E^{\text{FMO/FD}} = \sum_{I \in B} E_I^{L_2} + \sum_{I > J} \Delta E_{IJ}^{L_2} + \sum_{I \in A, J \in F} \Delta E_{IJ}^{L_2, L_1} \quad (22)$$

FD differs from FDD in the second sum (compare to eq 1), where fragments  $I$  are confined to **B** in FD and **A** in FDD. As this is a trivial change in the formulation, it is not elaborated further in detail. Because ultimately one is only interested in the gradients and second derivatives for the atoms in **A**, the neglect of the dimers in FDD lying outside of **A** and polarized by the fragments in it should make a relatively small numeric difference, especially for the Hessian, so the FMO/FD Hessian was not derived. It should also be noted that in both FD and FDD one only optimizes atoms in **A**, so it is consistent with evaluating only  $H_{AA}$ .

### 3. COMPUTATIONAL DETAILS

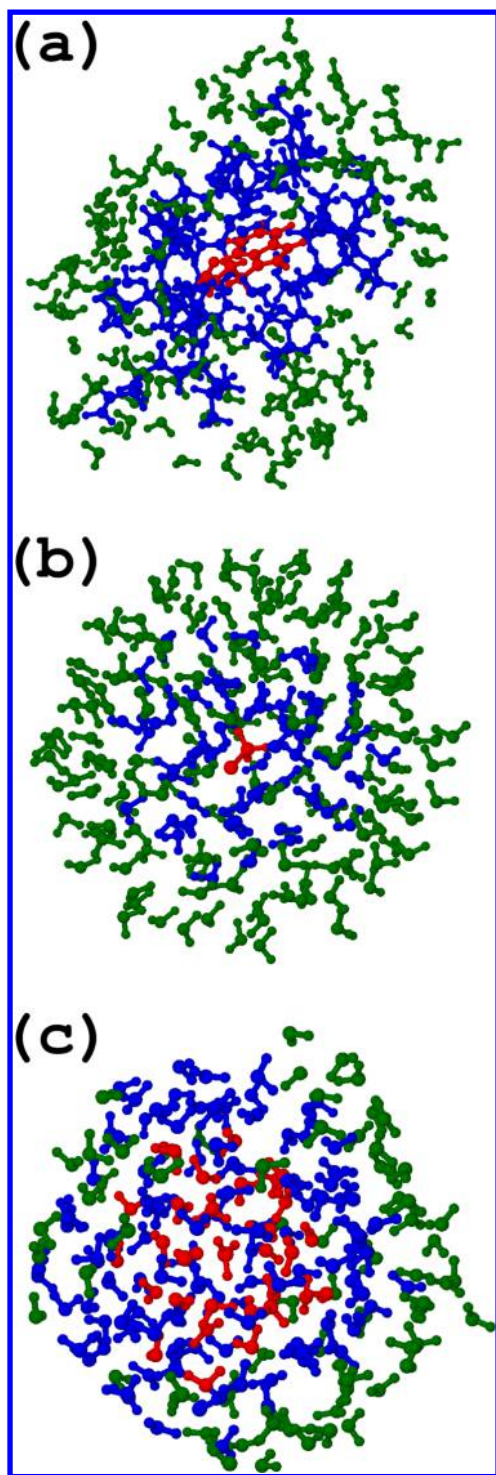
The first and second analytic derivatives of the FMO/FDD energy were implemented into the FMO code<sup>85</sup> in GAMESS,<sup>86,87</sup> and parallelized using generalized distributed data interface (GDDI).<sup>88</sup> FMO was interfaced with the IRC engine in GAMESS (for FMO both with and without FDD). The Gibbs free energies in this work include rotational, vibrational and translational contributions according to the standard harmonic oscillator and rigid rotor models.<sup>89</sup>

In the beginning of a FMO/FDD calculation, one has to compute the whole system at the FMO1 level for the initial geometry; after that the electronic state of fragments in domain **F** is frozen for other geometries during a geometry optimization or IRC. Therefore, it is necessary to allow for a calculation restart, whereby the molecular orbitals of the fragments in **F** are read. Such restarts were implemented and used as necessary. For instance, during IRC, the starting point to generate the electronic state of fragments in **F** was the transition state geometry.

FragIt program was used for making FMO input files.<sup>90</sup> In all FMO/FDD calculations, STO-3G and 6-31G(d) were used for layers  $L_1$  and  $L_2$ , respectively, and the method is denoted by FMO-RHF/FDD/STO-3G:6-31G(d). Numerical energy gradient was calculated by the double differencing with the coordinate shift of 0.005 Å for the system where fragments are connected by covalent bonds, and 0.0005 Å for molecular clusters. The Trp cage (PDB: 1L2Y) protein was solvated in 161 water molecules with the total of 787 atoms (Figure 2a), and CH<sub>3</sub>Cl was solvated with 219 water molecules (Figure 2b) with the total of 662 atoms.

The accuracy of the free energy barrier and the reaction path obtained with IRC were compared to full *ab initio* results for the S<sub>N</sub>2 reaction CH<sub>3</sub>Cl + OH<sup>−</sup> → Cl<sup>−</sup> + CH<sub>3</sub>OH in explicit solvent (41 water molecules). The dependence of the free





**Figure 2.** Systems used for FMO/FDD accuracy calculations (domains F, b, and A are shown in green, blue, and red, respectively). (a) Trp cage protein (PDB: 1L2Y) solvated in 161 water molecules. (b)  $\text{CH}_3\text{Cl}$  solvated in 219 water molecules. (c)  $\text{XCH}_3 + \text{OH}^-$  solvated in 212 water molecules.

energy differences on the size of the domains A and B was studied for (a) the  $\text{S}_\text{N}2$  reaction in 212 water molecules (Figure 2c, total 643 atoms) and (b) for protein–ligand binding  $\text{P} + \text{L} \rightarrow \text{PL}$ , for the complex of the Trp cage protein (P) with the ligand (L), *p*-hydroxybenzoic acid.

Finally, as an application of FMO/FDD, using the analytic gradient, Hessian, and IRC developed in this work, a reaction

involving TIM was studied (Figure 3a). TIM has two reactive sites, and two phosphoglycolhydroxamic acid (PGH) molecules are included in the complex. The PGH adjacent to Glu410 is denoted as PGHB in the Protein Data Bank file; this PGHB undergoes the chemical reaction, while PGHA remains unchanged. The initial structure (PDB: 7TIM)<sup>79</sup> was solvated in a layer of water (9227 atoms in the solvated enzyme and reactant), and the geometry was optimized with AMBER99 force field<sup>91</sup> using AMBER12.<sup>92</sup> In spectra simulations, the broadening parameter of  $10 \text{ cm}^{-1}$  is used for the line width.<sup>60</sup>

## 4. RESULTS AND DISCUSSION

**4.1. Accuracy of the FMO/FDD Analytic Gradient.** The gradient errors for FMO/FDD are evaluated by computing the difference between analytic and numeric gradients for the coordinates  $a$  of atoms in A. For the solvated Trp cage, the central Trp-6 residue fragment is chosen as the domain A, and the surrounding amino acid residues and water molecules within  $5.2 \text{ \AA}$  are assigned to B. For the solvated  $\text{CH}_3\text{Cl}$ , the solute is chosen as domain A, and the surrounding water molecules within  $5.2 \text{ \AA}$  are assigned to B. The protein is divided into 19 fragments (one residue per fragment, except that the N-terminus was merged with its neighbor), while water molecules and  $\text{CH}_3\text{Cl}$  are assigned as one molecule per fragment.

The unitless distance of  $r_{AB}$  between atoms A and B used in the thresholds<sup>83</sup> in FMO is defined as  $r_{AB} = |\mathbf{R}_A - \mathbf{R}_B| / (W_A + W_B)$ . The unitless value of 2.0 for distance-related thresholds corresponds to  $5.2 \text{ \AA}$ , if two closest atoms are O and H, because the van der Waals radii  $W_A$  are equal to 1.4 and  $1.2 \text{ \AA}$  for O and H, respectively.

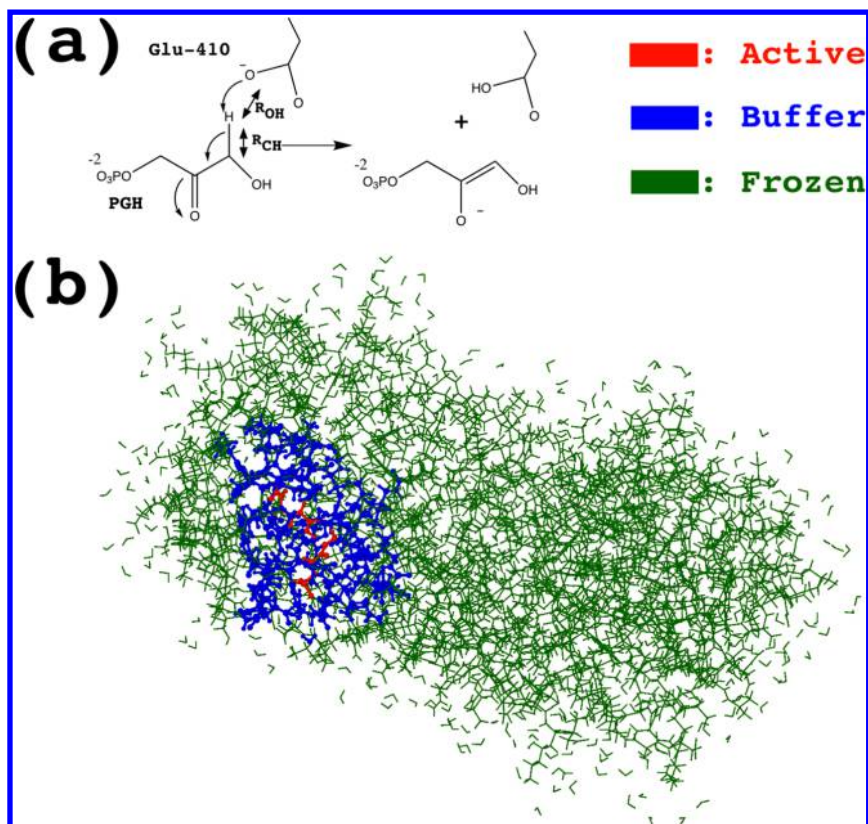
The systems are shown in Figure 2a,b, and the results are summarized in Figure 4 and Table 1. For both the Trp cage and  $\text{CH}_3\text{Cl}$  systems, a drastic improvement in the gradient error is obtained by including the response terms developed in this work. The error trends are similar to those found<sup>42</sup> for FMO without FDD, and the threshold of the ESP-PC approximation should be set to 3.5 or larger for a better accuracy, although the value of 2.0 appears to be sufficient for molecular clusters. In order to calculate the Hessians, the ESP-PC approximation was not used in the study of chemical reactions.

**Table 1.** Maximum Differences and Root-Mean-Square Deviation between the Analytic and Numeric Energy Gradients for FMO-RHF/FDD/STO-3G:6-31G(d)<sup>a</sup>

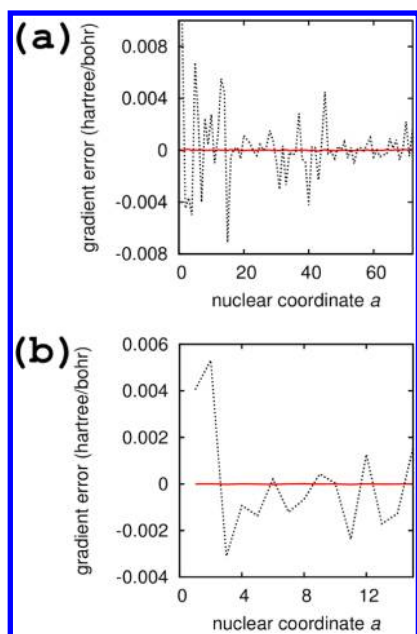
system	with response		no response	
	max	rmsd	max	rmsd
solvated 1L2Y <sup>b</sup>	0.000081	0.000028	0.009749	0.002440
solvated $\text{CH}_3\text{Cl}$ <sup>c</sup>	0.000021	0.000009	0.005304	0.002208

<sup>a</sup>The analytic gradient is computed with and without response terms in eq 14. <sup>b</sup>RESPPC=3.5, RESDIM=2.0. <sup>c</sup>RESPPC=2.0, RESDIM=2.0.

**4.2. Accuracy of the Reaction Paths Mapped with FMO/FDD.** The accuracies of the FMO/FDD Hessian and reaction path (IRC) are evaluated for the  $\text{S}_\text{N}2$  reaction  $\text{CH}_3\text{X} + \text{OH}^- \rightarrow \text{X}^- + \text{CH}_3\text{OH}$  in explicit solvent,  $\text{X} = \text{F}, \text{Cl}, \text{and Br}$ . The active domain A was chosen to be the reaction fragment, containing the solute as well as 6 surrounding water molecules merged into one fragment, and the other water molecules formed separate fragments. In domain B all fragments within  $5.2 \text{ \AA}$  were assigned, which for this system containing 41 solvent (water) molecules resulted in B coinciding with S.



**Figure 3.** Enzymatic reaction studied with FMO/FDD: (a) schematic representation and (b) the whole system. The reactants or products form the active domain A.



**Figure 4.** FMO/FDD gradient errors (the difference between analytic and numeric gradients) for atoms in A at the level of RHF/STO-3G:6-31G(d) shown as a red solid line for fully analytic gradients, (developed in this work) and as a black dotted line for approximate gradients neglecting the response terms in eq 14 (following the approximate gradient developed earlier<sup>65</sup>), computed for (a) solvated IL2Y shown in Figure 2a (using the thresholds of RESDIM=2.0 and RESPPC=3.5) and (b) solvated CH<sub>3</sub>Cl shown in Figure 2b (using the thresholds of RESDIM=2.0 and RESPPC=2.0).

The results are shown in Table 2. FMO/FDD correctly reproduces the trend for the three halogen atoms, and the error in the Gibbs free energy does not exceed 1.6 kcal/mol. The largest error in the imaginary frequency is 7.9 cm<sup>-1</sup>. Overall, FMO/FDD has slightly larger errors than FMO without FDD, compared to full RHF; when FDD is not used, the largest error in the Gibbs free energies and imaginary frequencies are 0.6 kcal/mol and 4.9 cm<sup>-1</sup>, respectively, which shows how much improvement can be obtained by enlarging the active domain. The reaction paths mapped with IRC are shown in Figure 5, and a reasonable agreement is observed between all three levels, FMO-RHF/FDD, FMO-RHF and full RHF.

**Table 2.** Gibbs Free Energy Barrier  $\Delta G$  (kcal/mol) at 298.15 K, and Imaginary Frequency  $\omega$  (cm<sup>-1</sup>) for the S<sub>N</sub>2 Reaction CH<sub>3</sub>X + OH<sup>-</sup> → X<sup>-</sup> + CH<sub>3</sub>OH in Explicit Solvent, with X = F, Cl, and Br ((6-31G(d))

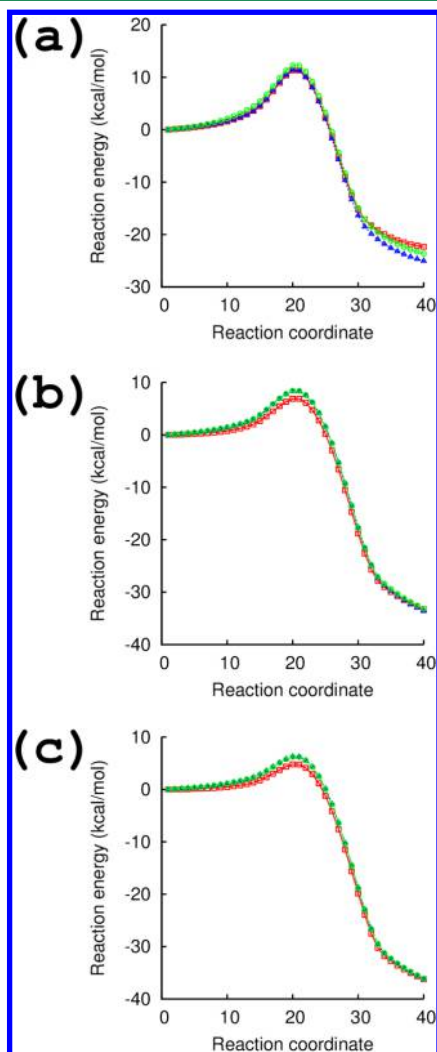
method	F		Cl		Br	
	$\Delta G$	$\omega$	$\Delta G$	$\omega$	$\Delta G$	$\omega$
FMO-RHF/FDD	15.3	590.1	12.4	464.0	8.6	426.8
FMO-RHF <sup>a</sup>	17.1	585.8	13.0	470.3	10.2	428.5
full RHF	16.5	590.7	13.0	472.1	10.2	427.6

<sup>a</sup>From ref 44.

**4.3. Dependence of the Free Energy on the Domain Size.** The free energy of the reaction barrier of the S<sub>N</sub>2 reaction CH<sub>3</sub>Cl + OH<sup>-</sup> → Cl<sup>-</sup> + CH<sub>3</sub>OH solvated in 212 water molecules was calculated (643 atoms in total). Also the free energy of binding of the ligand (*p*-hydroxybenzoic acid) to the Trp cage protein was evaluated in gas phase. The fragmentation

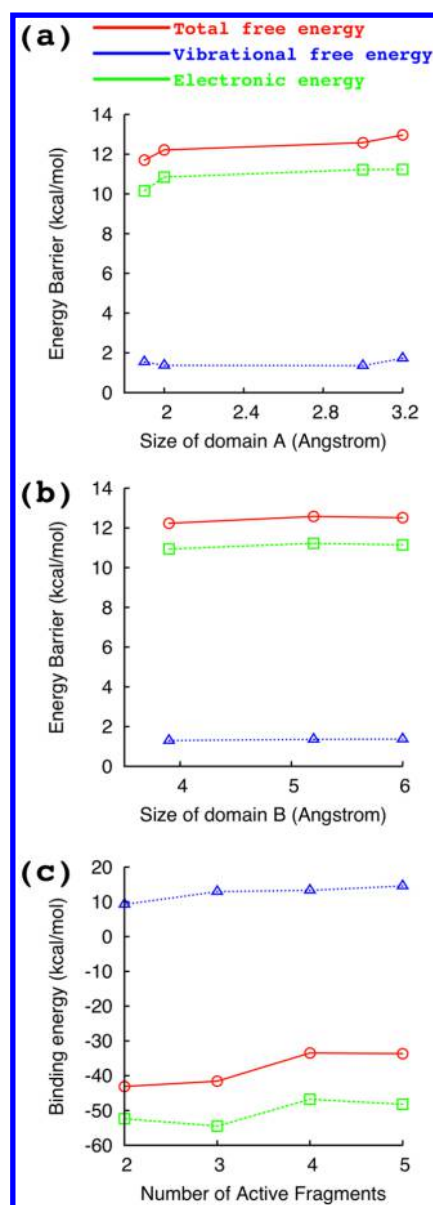
of these systems was done as above. For the  $S_N2$  reaction, four definitions of **A** were used for 0, 1, 2, and 3 solvent shells (1, 8, 16, and 19 fragments in **A**, respectively), whereas for the Trp cage complex, 2, 3, 4, and 5 manually selected fragments were assigned to **A** (namely, Tyr-3, Gln-5, Pro-18, and Arg-16 were gradually added to the ligand). For the  $S_N2$  reaction the size of **B** was also varied, while for the protein–ligand complex the size of **B** was constant (5.2 Å).

The results are shown in Figure 6. For both systems, the vibrational contribution to the Gibbs free energy can be seen to be almost constant with a very small dependence on the radius of either **A** or **B**.



**Figure 5.** Reaction paths for the  $S_N2$  reaction ( $\text{CH}_3\text{X} + \text{OH}^- \rightarrow \text{X}^- + \text{CH}_3\text{OH}$ ) in explicit solvent, shown as green circles for full RHF, blue triangles for FMO-RHF, and red squares for FMO-RHF/FDD: (a)  $\text{X} = \text{F}$ , (b)  $\text{X} = \text{Cl}$ , and (c)  $\text{X} = \text{Br}$ . The energy plotted is the electronic contribution only.

For the  $S_N2$  reaction, the central fragment already includes six water molecules, so even for the smallest size of **A** the results are already essentially converged. The difference in the total reaction barrier is 1.3 kcal/mol between the smallest and largest sizes of **A**, of which 0.2 kcal/mol is the vibrational free energy contribution. The difference in the free energy barrier between the smallest and largest size of **B** is 0.2 kcal/mol for either the electronic or vibrational free energy.



**Figure 6.** Dependence of the Gibbs free energy at 298.15 K on the domain size in FMO/FDD. (a) Reaction barrier for the  $S_N2$  reaction of  $\text{CH}_3\text{Cl}$  in explicit solvent as a function of the size of **A**. (b) Reaction barrier for the  $S_N2$  reaction of  $\text{CH}_3\text{Cl}$  as a function of the size of **B**. (c) Protein–ligand binding energy for the Trp cage protein in the gas phase and *p*-hydroxybenzoic acid ligand, as a function of the size of **A**.

For the protein–ligand complex, it is clearly necessary to include some fragments from the binding pocket in the protein into **A** to obtain better results. Between the smallest and largest choices of **A**, the differences of the free energy of the protein–ligand binding are 4.1 and 5.3 kcal/mol for the electronic and vibrational contributions, respectively. The three residue fragments Tyr-3, Gln-5, and Pro-18 are located near the ligand, and their addition to **A** changes the total free energy of binding from  $-41.6$  (two fragments in **A**) to  $-33.5$  kcal/mol (four fragments in **A**). A further inclusion of Arg-16 residue into **A** changes the total binding energy only by 0.2 kcal/mol.

**4.4. Application of FMO/FDD to an Enzymatic Reaction.** The keto–enol tautomeric reaction of PGH and TIM, previously investigated experimentally<sup>78</sup> and theoretically,<sup>77</sup> was studied in this work using FMO/FDD. The system



is shown in Figure 3, the energies are summarized in Table 3, and the results are plotted in Figure 7. The fragmentation was,

**Table 3. Changes of the Electronic Energy  $\Delta E_{\text{elec}}$ , Zero-Point Energy  $\Delta E_{\text{ZPE}}$ , Vibrational Gibbs Free Energy  $\Delta G_{\text{vib}}$ , and Total Gibbs Free Energy  $\Delta G_{\text{tot}}$  at 298.15 K (kcal/mol),<sup>a</sup> Relative to the Reactant for the Keto–Enol Tautomerism Reaction Shown in Figure 3, Computed at the FMO-RHF/D/FDD/STO-3G:6-31(d) Level**

	$\Delta E_{\text{elec}}$	$\Delta E_{\text{ZPE}}$	$\Delta G_{\text{vib}}$	$\Delta G_{\text{tot}}$
reactant	0.00	0.0	0.00	0.00
TS	24.60	3.66	−3.18	21.43
product	12.29	3.20	−0.45	11.85

<sup>a</sup> $\Delta G_{\text{tot}}$  is the total value, including other contributions not shown in the table, computed for a single minimum in the harmonic oscillator and rigid rotor model.  $\Delta G_{\text{vib}}$  adds temperature corrections to ZPE from the partition function.

as a rule, one residue (or water molecule) per fragment except that the reaction center comprising the reactant PGH and Glu-410 were merged into one fragment, denoted as PGHGLU-410, and FMO properties are discussed for this merged fragment having the total charge of −3. The active domain A was chosen to be fragment PGHGLU-410 as shown in Figure 3, and the amino acid residues and water molecules within 5.2 Å distance from A were included in the polarizable buffer B. After locating the transition state from a trial geometry using

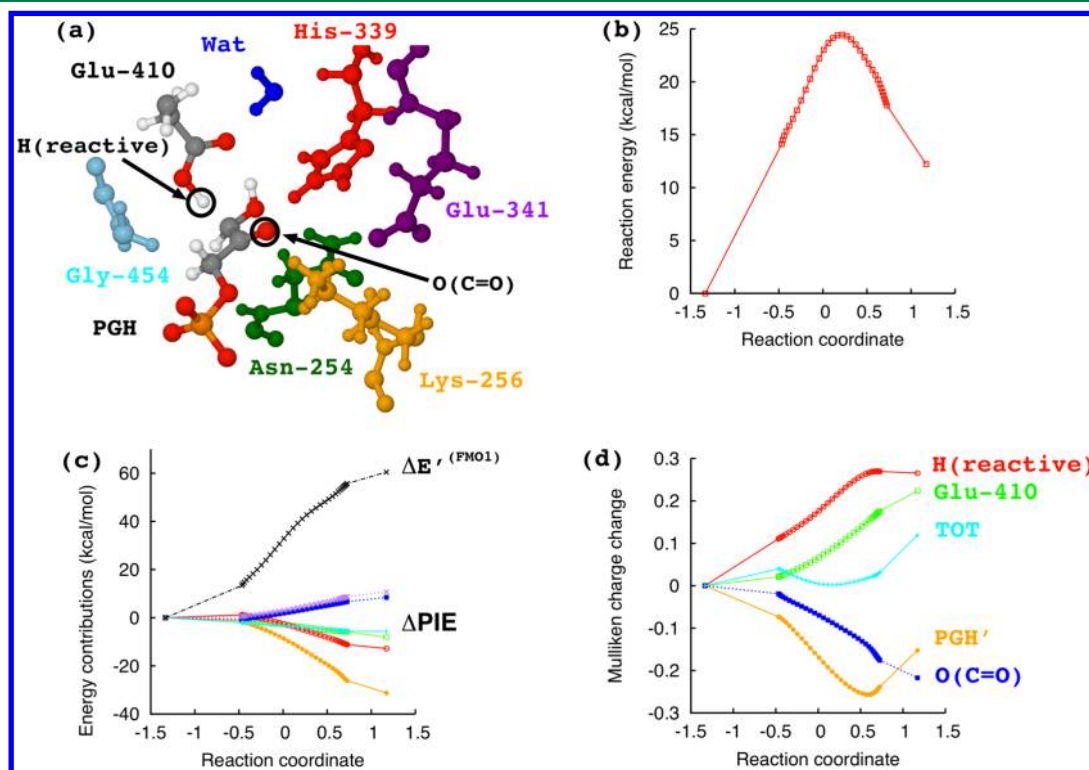
the analytic Hessian developed in this work, IRC simulations were performed to map the reaction path.

The free energy barrier obtained in this study is 21.4 kcal/mol at the RHF/6-31G(d) with dispersion,<sup>93</sup> comparable to 21.9 kcal/mol reported by others<sup>77</sup> at the level of RHF/3-21G, whereas the experimental estimate is 14 kcal/mol.<sup>78</sup> The reaction coordinate is mainly the difference in the two bond distances,  $R_{\text{react}} = R_{\text{C-H}} - R_{\text{O-H}}$ , see Figure 3a. For the transition state this difference obtained from the IRC path (Figure 7b) is 0.27 Å, which can be compared to 0.36 Å in the other study.<sup>77</sup>

A byproduct of FMO calculations (see eq 1) is pair interaction energies (PIEs),  $\Delta E_{ij}^{L_2}$  and  $\Delta E_{ij}^{L_2L_1}$  (the latter has no charge transfer contribution due to the ES-DIM approximation for these dimers; thus,  $\Delta E_{ij}^{L_2L_1} = \Delta E_{ij}^{L_2L_1}$ ), collectively called  $\Delta E_{ij}$ . PIEs can be used to identify the residue fragments, which are important factors in the catalytic activity.<sup>71–73</sup> It should be noted that here, as in other FMO publications, fragments are defined by detaching C $\alpha$ –C bonds at C $\alpha$ , so that fragment residues in FMO are shifted compared to conventional residues by a carboxyl group. The names of residue fragments in FMO have a dash to indicate the difference from conventional residues (GLU-410 is the fragment that corresponds to the amino acid residue Glu410).

The change in the energy during a chemical reaction in FMO is

$$\Delta E^{\text{FMO/FDD}}(R) = E^{\text{FMO/FDD}}(R) - E^{\text{FMO/FDD}}(R_0) \quad (23)$$



**Figure 7.** Keto–enol tautomeric reaction of phosphoglycolohydroxamic acid (PGH) and the triosephosphate isomerase (7TIM). (a) The binding pocket at the transition state geometry. (b) Reaction path obtained with IRC using FMO-RHF/FDD; the reaction coordinate is mainly given by  $R_{\text{C-H}} - R_{\text{O-H}}$  (see Figure 3a). The energy of the reactant is set to 0. Only the electronic contribution is plotted. (c) Change in the components of the total energy change in the Keto–enol tautomeric reaction relative to the reactant values: the sum of all monomer energies  $\sum_i E'_i$  (the FMO1 internal energy, shown as black dashed line) and pair interaction energies  $\Delta E_{ij}$  between the reactant and important residue fragments of the enzyme, shown as lines colored as atoms in (a). (d) Change in the Mulliken charges of the subsystems defined within the reaction fragment: PGH, Glu-410 and two atoms O(C=O) and H(reactive). The charges at the reactant geometry are set to 0 (the absolute values of charges are given in Table 5).



where the total FMO/FDD energy is computed at the geometry  $R$  relative to a reference geometry  $R_0$ . Then, omitting FMO/FDD for brevity, one obtains from eq 23, using eq 1,

$$\Delta E = \sum_{I \in B} \Delta E_I'^{L_2} + \sum_{I>J} \Delta \Delta E_{IJ}^{L_2} + \sum_{I \in A, J \in F} \Delta \Delta E_{IJ}^{L_2, L_1} \quad (24)$$

where  $\Delta E_I'^{L_2}$  and  $\Delta \Delta E_{IJ}^{L_2}$  ( $\Delta \Delta E_{IJ}^{L_2, L_1}$ ) are the changes in the monomer energies and PIEs, respectively. In FMO/FDD the energies of  $L_1$  do not appear, and for brevity the layer is omitted ( $\Delta E_I'$  denotes  $\Delta E_I'^{L_2}$ ) in some discussions below.  $\Delta E_I' = E_I'(R) - E_I'(R_0)$ , and similarly the change in PIEs is defined by taking the difference between two geometries. Thus, one can analyze the reasons why the reaction barrier and heat take the calculated values by discussing the values of their components,  $\Delta E_I'$  and  $\Delta \Delta E_{IJ}$ .

The change in the internal energy  $E_I'$  along the reaction path has two contributions: (a) the deformation energy, which shows how fragment  $I$  is either stabilized or destabilized when its geometry changes, here when the reference state is for the reactant complex), and (b) the destabilization polarization effect on fragment  $I$  by the rest of the system;<sup>94</sup> the stabilization component of the polarization forms a part of the electrostatic component of PIEs. This change in the polarization arises from a change in the rest of the system. In other words, the two sides to the changes of  $E_I'$  are internal (deformation) and external (polarization). It is actually possible to further analyze these contributions with the aid of the isolated state analysis in the pair interaction energy decomposition analysis (PIEDA),<sup>94–96</sup> but it is not attempted it here.

Below, the reference geometry  $R_0$  is taken to be that of the reactant. The most important geometries  $R$  are those of the transition state and products, in which case  $\Delta E$  plotted in Figure 7b corresponds to the reaction barrier and heat, respectively.

For the isomerization reaction, the changes in the energies  $\Delta E_I'$  and  $\Delta \Delta E_{IJ}$  are summarized in Table 4 and Figure 7c, which show the driving factors for the reaction.

The changes in the PIEs  $\Delta \Delta E_{IJ}$  along the reaction to some extent reduce the large destabilization in the internal energy of the reactant fragment. Typically, the stabilization polarization contribution, which forms a part of PIEs, is two times larger

**Table 4. Changes<sup>a</sup> in the Internal Energies  $E_I'$  and Pair Interaction Energies  $\Delta E_{IJ}$ , Contributing to the Total Energy Change during the Keto–Enol Tautomerism Reaction Involving TIM, Shown in Figure 3, and Computed at the FMO-RHF/FDD/D/STO-3G:6-31(d) Level**

property	component <sup>b</sup>	TS	product
$\Delta E_I'$	$I = \text{PGHGLU-410}$	38.53	52.04
$\Delta E_I'$	$I = \text{Gly-475}$	1.82	4.86
$\Delta \sum_i E_i'^c$	other monomers	0.72	3.59
$\Delta \Delta E_{IJ}$	$I = \text{PGHGLU-410}, J = \text{Lys-256}$	−13.04	−31.27
$\Delta \Delta E_{IJ}$	$I = \text{PGHGLU-410}, J = \text{His-339}$	−4.81	−12.82
$\Delta \sum_{I>J} \Delta E_{IJ}^d$	other PIEs	1.38	−4.11
$\Delta E$	total	24.60	12.29

<sup>a</sup>Relative to the reactant. For example,  $\Delta E_I' = E_I'(\text{TS or product}) - E_I'(\text{reactant})$ . <sup>b</sup>PGHGLU-410 denotes the reaction fragment composed of phosphoglycolohydroxamic acid (PGH) and Glu-410. <sup>c</sup>The sum over  $I$  excludes PGHGLU-410 and Gly-475. <sup>d</sup>The sum over pairs of  $IJ$  excludes PGHGLU-410,Lys-256 and PGHGLU-410,His-339.

than the destabilization component.<sup>94</sup> PIEs change less than  $E_I'$ , and assuming the polarization is generally less than PIEs, it can be inferred that the main contribution to the reaction energy is the deformation rather than the polarization. PIEs partially reduce this dramatic destabilization.

In this system, the largest PIEs are found for the ligand interacting with Gly-454, Asn-254, Lys-256, Glu-341, and His-339 in the binding pocket, and one water molecule (see Figure 7a). The difference between these PIEs along the reaction is plotted in Figure 7c relative to the values for the reactants, which were −19.20, −35.73, −220.09, 134.09, −37.87, and −17.44 kcal/mol for Gly-454, Asn-254, Lys-256, Glu-341, His-339, and Wat, respectively. Lys-256 is the main stabilizing factor, which has both a large absolute value of PIE and a large change of it along the reaction path. His-339 is second in importance (Glu-410, which directly participates in the reaction, forms a part of the reactant fragment, and it is not easily possible to discuss its PIE).

Except for Glu-341 and Wat, other PIEs become more exothermic as the reaction proceeds. The cationic Lys-256 shows the largest stabilization: in the product vs reactant the change in the PIE is about −31 kcal/mol. The reactant and product charges change from −3 to −2, while that of the enzyme changes from 0 to −1. A change in the charge distribution in the pocket along the reaction path makes a big difference on the PIE. On the other hand, the PIEs for the anionic Glu-341 and neutral Wat become more endothermic. These results indicate which residues may be mutated in an attempt to change the catalytic activity.

The reactant PGH molecule for the purpose of analysis was divided into three subunits: atom H(reactive), atom O(C=O), and the rest of atoms PGH'. To avoid ambiguity, the subunit of Glu-410 in the discussion of charges means the ionized state of Glu-410 in the reactant (an anion). The subunit charges for the reaction fragment (including PGH and Glu-410) are given in Table 5 and Figure 7d. The definition of the charges for a

**Table 5. Mulliken Charges of the Subunits of the Reaction Fragment at the Reactant, TS, and Product Geometry for the Keto–Enol Tautomerism Reaction Involving TIM, Obtained from FMO-RHF/FDD/D/STO-3G:6-31(d) Calculations**

	reactant	TS	product
Glu-410	−0.571	−0.476	−0.348
H(reactive)	0.260	0.474	0.525
O(C=O)	−0.641	−0.735	−0.859
PGH'	−1.714	−1.927	−1.866
total <sup>a</sup>	−2.667	−2.664	−2.547

<sup>a</sup>The sum of the charges of all subunits.

subunit C (where C is Glu-410, PGH', H(reactive), and O(C=O)) is<sup>97</sup>

$$Q_A = \sum_{I:A \in I} Q_A^I + \sum_{I>J:A \in IJ} (Q_A^{IJ} - Q_A^I \delta_{A \in I} - Q_A^J \delta_{A \in J}) \quad (25)$$

$$Q^C = \sum_{A:A \in C} Q_A \quad (26)$$

$Q_A^X$  are Mulliken atomic charges of atom  $A$  in  $X = I, J$ , and  $IJ$ , and  $\delta_{A \in I} = 1$  if atom  $A$  is found in  $I$  and 0 otherwise.

The formal charge on the reactive fragment is −3; however, FMO takes into account interfragment charge transfer via

dimer calculations, and the actual charge is about  $-2.6$ , as shown in Table 5. Adding the charges of Glu-410 and H(reactive) for the product gives  $+0.177$  (the formal charge is 0), while for the reactant the charge of Glu-410 is  $-0.571$  (the formal charge is  $-1$ ). The subunit charges in the transition state are between those in the reactant and product, with the exception of that of PGH', which features a maximum (see also Figure 7d) somewhere between the TS and product. The sum of the charges of PGH' and  $O(C=O)$  gives the charge of PGH except for the reactive H, and this charge changes from  $-2.4$  (reactant) to  $-2.7$  (TS and products).

The reaction can be viewed as a transfer of  $H^+$ , and H(reactive) has its charge increased from  $0.260$  to  $0.474$  and  $0.525$ , in the reactant, TS, and product states. The total charge on the reactive fragment changes relatively little, in particular insignificantly until the reaction proceeds beyond the TS, which indicates that the charge changes are internal to the reaction fragment, and there is relatively small charge transfer from it to the rest of the enzyme and solvent.

The IR and Raman spectra simulated with FMO/FDD are shown in Figure 8. It can be seen how the formation and breaking of covalent bonds changes the spectra. For instance, the  $COO^-$  group in the products has intense IR and Raman peaks in Figure 8c.

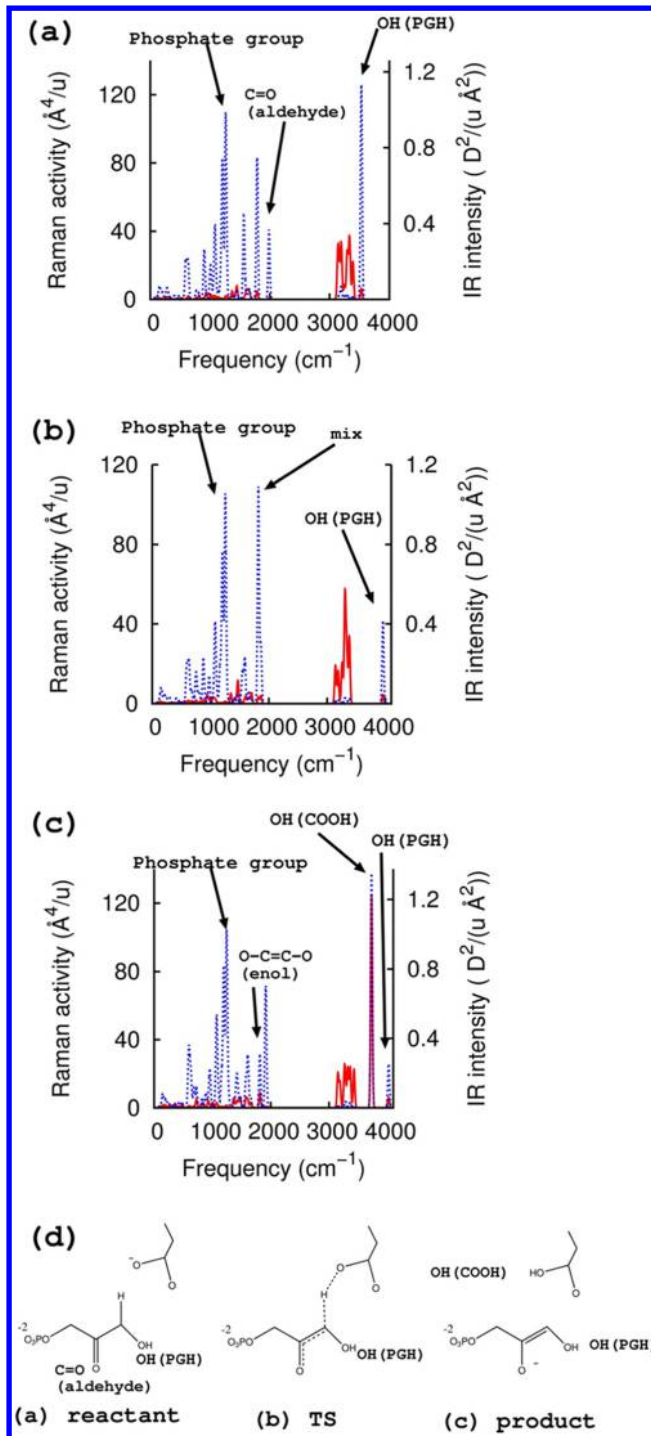
All enzymatic calculations were performed on 16 PC nodes equipped with 2.93 GHz Xeons (8 CPU cores and 12 GB RAM per node). First, to obtain the initial TS structure, a preliminary constrained geometry optimization was performed, and it took 12 h; then one Hessian calculation took 24 h and the transition state search with RUNTYP=SADPOINT took 12 h. IRC calculations were performed in two directions to obtain the path toward the reactant and product. In total, IRC took 42 h for 16 points in each direction on the path (one point required 40 single gradient calculations). Then, to obtain converged geometries at the two minima, corresponding to the reactant and product, two geometry optimizations (15 single gradient calculations) were performed, which took 12 h, respectively. Finally, the free energy vibrational analysis have been performed for transition state, and two minima, and each Hessian calculation took 24 h. Thus, to map the reaction path for this system containing 9227 atoms with FMO/FDD, it took about 102 h, and to obtain free energies for the reactants, products, and TS, 72 h were required for the three Hessian calculations.

## 5. CONCLUSIONS

The analytic energy gradient for FMO/FD and FMO/FDD, as well as analytic Hessian for FMO/FDD, have been derived and implemented. As a result, geometry optimizations, transition state search, IRC following, and IR and Raman spectra simulations for large molecular systems can be performed. The developed Hessian allows one to evaluate free energies at a given stationary point, which can be used for refining the energetics of chemical reactions.

The accuracy of the FMO/FDD gradient and Hessian has been demonstrated on a number of representative systems, including  $S_N2$  reactions and a protein–ligand complex for the Trp cage protein. It has been shown that FMO/FDD accurately reproduces reaction paths obtained with IRC. The dependence of the energetics on the domain size has also been studied.

The efficiency of FMO/FDD has been shown in the application to an enzyme reaction involving triosephosphate isomerase. The reaction path of the reaction has been mapped,



**Figure 8.** Spectra for the keto–enol tautomeric reaction of phosphoglycolohydroxamic acid and the triosephosphate isomerase, calculated with FMO/FDD at the RHF/D/6-31G(d) level: (a) reactants, (b) TS, and (c) products. Chemical structures are shown in (d). The “mix” mode corresponds to the coupling between the vibrational modes of  $C=O$  (aldehyde) and  $O=C=C-O$  (enol). Raman and IR spectra are plotted as red solid and blue dashed lines, respectively.

the free energies have been evaluated using the Hessians, and IR and Raman spectra have been simulated, for the system containing 9227 atoms using QM.

The factors leading to the observed energetics have been elucidated using the fragment energies as well as pair

interaction energies. It has been shown that the latter only partially compensate for the large destabilization of the reaction fragment consisting of the reactant and Glu-410, which is destabilized mainly due to the deformation. The main stabilizing residue fragments in the enzyme have been found to be Lys-256 and His-339, whose stabilization effect becomes stronger as one proceeds from the reactants to the products. The charge transfer between PGH and Glu-410 has been discussed in detail.

For biological systems such as enzymes, a single reaction path does not always provide a reliable basis for studying chemical reactions, because due to the flatness and high complexity of the energy surface multiple reaction paths may be competitive, and ultimately one has to perform MD simulations, which are possible using the FMO/FDD gradient developed in this work. In many cases a single transition state and reaction path may be sufficiently representative to be useful, especially in rigid inorganic systems, such as chemical reactions or adsorption occurring on a surface or in porous materials.<sup>98,99</sup> The free energies evaluated with FMO/FDD can be used to improve the energetics of chemical reactions, for example, in the evaluation of  $pK_a$  values in proteins.<sup>84</sup>

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: d.g.fedorov@aist.go.jp.

### Funding

This work has been supported by the Next Generation Super Computing Project, Nanoscience Program (MEXT, Japan) and Computational Materials Science Initiative (CMSI, Japan). Calculations were performed on TSUBAME2.5 at the Global Scientific Information and Computing Center of Tokyo Institute of Technology, RIKEN Integrated Cluster of Clusters (RICC).

### Notes

The authors declare no competing financial interest.

<sup>†</sup>Deceased June 11, 2012.

## REFERENCES

- (1) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.
- (2) Hai, L.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185–199.
- (3) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (4) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- (5) Siegbahn, P. E. M.; Borowski, T. *Acc. Chem. Res.* **2006**, *39*, 729–738.
- (6) Si, D.; Li, H. *J. Chem. Phys.* **2009**, *131*, 044123.
- (7) Goedecker, S.; Scuseria, G. E. *Comp. Sci. Eng.* **2003**, *5*, 14–21.
- (8) Zalesny, R.; Papadopoulos, M. G.; Mezey, P. G.; Leszczynski, J., Eds. *Linear-Scaling Techniques in Computational Chemistry and Physics*; Springer: New York, 2011.
- (9) Nakata, A.; Bowler, D. R.; Miyazaki, T. *J. Chem. Theory Comput.* **2014**, *10*, 4813–4822.
- (10) Wilkinson, K. A.; Hine, N. D. M.; Skylaris, C.-K. *J. Chem. Theory Comput.* **2014**, *10*, 4782–4794.
- (11) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632–672.
- (12) Otto, P.; Ladik, J. *Chem. Phys.* **1975**, *8*, 192–200.
- (13) Gao, J. L. *J. Phys. Chem. B* **1997**, *101*, 657–663.
- (14) He, X.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.
- (15) Friedrich, J.; Yu, H.; Leverentz, H. R.; Bai, P.; Siepmann, J. I.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2014**, *5*, 660–670.
- (16) Tong, Y.; Mei, Y.; Zhang, J. Z. H.; Duan, L. L.; Zhang, Q. G. *J. Theor. Comput. Chem.* **2009**, *8*, 1265.
- (17) Söderhjelm, P.; Kongsted, J.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 1726–1737.
- (18) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. *Acc. Chem. Res.* **2014**, *47*, 2837–2845.
- (19) Frank, A.; Möller, H. M.; Exner, T. E. *J. Chem. Theory Comput.* **2012**, *8*, 1480–1492.
- (20) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553–578.
- (21) Kiewisch, K.; Jacob, C. R.; Visscher, J. *J. Chem. Theory Comput.* **2013**, *9*, 2425–2440.
- (22) Sakai, S.; Morita, S. *J. Phys. Chem. A* **2005**, *109*, 8424–8429.
- (23) Hua, W.; Fang, T.; Li, W.; Yu, J.-G.; Li, S. *J. Phys. Chem. A* **2008**, *112*, 10864–10872.
- (24) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- (25) Collins, M. A. *J. Chem. Phys.* **2015**, *141*, 094108.
- (26) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.
- (27) Sahu, N.; Gadre, S. R. *J. Chem. Phys.* **2015**, *142*, 014107.
- (28) Howard, J. C.; Tschumper, G. S. *J. Chem. Phys.* **2013**, *139*, 184113.
- (29) Yoshikawa, T.; Nakai, H. *J. Comput. Chem.* **2014**, *36*, 164–170.
- (30) Stewart, J. J. P. *J. Mol. Model.* **2009**, *15*, 765–805.
- (31) Cui, Q.; Elstner, M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368–14377.
- (32) Zheng, W. J.; Brooks, B. R. *Biophys. J.* **2005**, *89*, 167.
- (33) Li, H.; Jensen, J. H. *Theor. Chem. Acc.* **2002**, *107*, 211–219.
- (34) Ghysels, A.; van Neck, D.; van Speybroeck, V.; Verstraeten, T.; Waroquier, M. *J. Chem. Phys.* **2007**, *126*, 224102.
- (35) Ghysels, A.; Speybroeck, V. V.; Pauwels, E.; Catak, S.; Brooks, B. R.; Neck, D. V.; Waroquier, M. *J. Comput. Chem.* **2010**, *31*, 994–1007.
- (36) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (37) Fedorov, D. G.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009.
- (38) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (39) Fedorov, D. G.; Nagata, T.; Kitaura, K. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562–7577.
- (40) Tanaka, S.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Fukuzawa, K. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10310–10344.
- (41) Kitaura, K.; Sugiki, S. I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163–170.
- (42) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Phys.* **2011**, *134*, 124115.
- (43) Nagata, T.; Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2012**, *544*, 87–93.
- (44) Nakata, H.; Nagata, T.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. *J. Chem. Phys.* **2013**, *138*, 164103.
- (45) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. *Chem. Phys. Lett.* **2014**, *603*, 67.
- (46) Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. J. *Cheminf.* **2011**, *3*, 2.
- (47) Sawada, T.; Fedorov, D. G.; Kitaura, K. *J. Am. Chem. Soc.* **2010**, *132*, 16862–16872.
- (48) He, X.; Fusti-Molnar, L.; Cui, G.; Merz, K. M. *J. Phys. Chem. B* **2009**, *113*, 5290–5300.
- (49) Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. *J. Phys. Chem. B* **2007**, *111*, 9621–9627.
- (50) Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. *Chem. Phys. Lett.* **2009**, *477*, 169–175.
- (51) Carlson, P. J.; Bose, S.; Armstrong, D. W.; Hawkins, T.; Gordon, M. S.; Petrich, J. W. *J. Phys. Chem. B* **2012**, *116*, 503–512.
- (52) Avramov, P. V.; Fedorov, D. G.; Sorokin, P. B.; Sakai, S.; Entani, S.; Ohtomo, M.; Matsumoto, Y.; Phys, H. N. *J. Chem. Lett.* **2012**, *3*, 2003–2008.
- (53) Nishimoto, Y.; Fedorov, D. G.; Irle, S. *J. Chem. Theory Comput.* **2014**, *10*, 4801–4812.



- (54) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722–2732.
- (55) Fedorov, D. G.; Asada, N.; Nakanishi, I.; Kitaura, K. *Acc. Chem. Res.* **2014**, *47*, 2846–2856.
- (56) Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Fedorov, D. G. *J. Mol. Struct. (THEOCHEM)* **2009**, *898*, 2.
- (57) Okamoto, T.; Ishikawa, T.; Koyano, Y.; Yamamoto, N.; Kuwata, K.; Nagaoka, M. *Bull. Chem. Soc. Jpn.* **2013**, *86*, 210–222.
- (58) Nakata, H.; Schmidt, M. W.; Fedorov, D. G.; Kitaura, K.; Nakamura, S.; Gordon, M. S. *J. Phys. Chem. A* **2014**, *118*, 9762–9771.
- (59) Nagata, T.; Fedorov, D. G.; Li, H.; Kitaura, K. *J. Chem. Phys.* **2012**, *136*, 204112.
- (60) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. *J. Chem. Theory Comput.* **2014**, *10*, 3689–3698.
- (61) Sato, M.; Yamataka, H.; Komeiji, Y.; Mochizuki, Y.; Ishikawa, T.; Nakano, T. *J. Am. Soc. Chem.* **2008**, *130*, 2396–2397.
- (62) Sato, M.; Yamataka, H.; Komeiji, Y.; Mochizuki, Y. *Chem.—Eur. J.* **2012**, *18*, 9714–9721.
- (63) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, *124*, 024301.
- (64) Ishikawa, T.; Yamamoto, N.; Kuwata, K. *Chem. Phys. Lett.* **2010**, *500*, 149–154.
- (65) Fedorov, D. G.; Alexeev, Y.; Kitaura, K. *J. Phys. Chem. Lett.* **2011**, *2*, 282–288.
- (66) Tsukamoto, T.; Mochizuki, Y.; Watanabe, N.; Fukuzawa, K.; Nakano, T. *Chem. Phys. Lett.* **2012**, *535*, 157–162.
- (67) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. *J. Phys. Chem. A* **2010**, *114*, 8705–8712.
- (68) Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. *Acc. Chem. Res.* **2014**, *47*, 2786–2794.
- (69) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. *PLoS One* **2013**, *8*, e60602.
- (70) Christensen, A. S.; Steinmann, C.; Fedorov, D. G.; Jensen, J. H. *PLoS One* **2014**, *9*, e88800.
- (71) Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457.
- (72) Ito, M.; Brinck, T. *J. Phys. Chem. B* **2014**, *118*, 13050–13058.
- (73) Jensen, J. H.; Willemoes, M.; Winther, J. R.; Vico, L. D. *PLoS One* **2014**, *9*, e95833.
- (74) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523–5527.
- (75) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1991**, *95*, 5853–5860.
- (76) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161.
- (77) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.
- (78) Alberly, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5627–5631.
- (79) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. *Biochemistry* **1991**, *30*, 5821–5826.
- (80) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614–618.
- (81) Goerigk, L.; Collyer, C. A.; Reimers, J. R. *J. Phys. Chem. B* **2014**, *118*, 14612–14626.
- (82) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638–2646.
- (83) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475–480.
- (84) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.
- (85) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832–6840.
- (86) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (87) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry, the first forty years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 1167–1189.
- (88) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872–880.
- (89) Davidson, N. *Statistical Mechanics*; McGraw-Hill: New York, 1962.
- (90) Steinmann, C.; Ibsen, M. W.; Hansen, A. S.; Jensen, J. H. *PLoS One* **2012**, *7*, e44480.
- (91) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (92) Case, D. A.; Darden, T.; Cheatham, T. III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. *AMBER 12*; University of California, San Francisco, 2012.
- (93) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- (94) Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 222–237.
- (95) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2012**, *116*, 704–719.
- (96) Alexeev, Y.; Mazanetz, M. P.; Ichihara, O.; Fedorov, D. G. *Curr. Top. Med. Chem.* **2012**, *12*, 2013–2033.
- (97) Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. *Chem. Phys. Lett.* **2002**, *364*, 139–143.
- (98) Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 11808–11816.
- (99) Roskopf, L.; Fedorov, D. G.; Gordon, M. S. *Mol. Phys.* **2013**, *111*, 1622–1629.