Analytic Second Derivative of the Energy for Density Functional Theory Based on the Three-Body Fragment Molecular Orbital Method

Hiroya Nakata
Tokyo Institute of Technology

Dmitri G. Fedorov
National Institute of Advanced Industrial Science and Technology

Federico Zahariev
Iowa State University, fzahari@iastate.edu

Michael Schmidt
Iowa State University, mike@si.fi.ameslab.gov

Kazuo Kitaura
Kobe University

Follow this and additional works at: http://lib.dr.iastate.edu/chem_pubs
Part of the Chemistry Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/chem_pubs/581. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Analytic second derivative of the energy for density functional theory based on the three-body fragment molecular orbital method

Hiroya Nakata, Dmitri G. Fedorov, Federico Zahariev, Michael W. Schmidt, Kazuo Kitaura, Mark S. Gordon, and Shinichiro Nakamura

Citation: The Journal of Chemical Physics 142, 124101 (2015); doi: 10.1063/1.4915068
View online: http://dx.doi.org/10.1063/1.4915068
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/12?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in

Analytic second derivatives of the energy in the fragment molecular orbital method

Unrestricted Hartree-Fock based on the fragment molecular orbital method: Energy and its analytic gradient

Time-dependent density functional theory based upon the fragment molecular orbital method

Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods

Ab initio molecular orbital and density functional characterization of the potential energy surface of the N₂O+Br reaction
Analytic second derivative of the energy for density functional theory based on the three-body fragment molecular orbital method

Hiroya Nakata (中田浩弥),1,2,3,a) Dmitri G. Fedorov,4,b) Federico Zahariev,5 Michael W. Schmidt (守美都舞空),5 Kazuo Kitaura (北浦和夫),6 Mark S. Gordon (強首領真空),3 and Shinichiro Nakamura (中村振一郎)2

1Center for Biological Resources and Informatics, Tokyo Institute of Technology, B-62 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan
2RIKEN, Research Cluster for Innovation, Nakamura Lab, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
3Japan Society for the Promotion of Science, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan
4NRI, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
5Department of Chemistry and Ames Laboratory, US-DOE, Iowa State University, Ames, Iowa 50011, USA
6Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

(Rceived 26 January 2015; accepted 4 March 2015; published online 23 March 2015)

Analytic second derivatives of the energy with respect to nuclear coordinates have been developed for spin restricted density functional theory (DFT) based on the fragment molecular orbital method (FMO). The derivations were carried out for the three-body expansion (FMO3), and the two-body expressions can be obtained by neglecting the three-body corrections. Also, the restricted Hartree-Fock (RHF) Hessian for FMO3 can be obtained by neglecting the density-functional related terms. In both the FMO-RHF and FMO-DFT Hessians, certain terms with small magnitudes are neglected for computational efficiency. The accuracy of the FMO-DFT Hessian in terms of the Gibbs free energy is evaluated for a set of polypeptides and water clusters and found to be within 1 kcal/mol of the corresponding full (non-fragmented) ab initio calculation. The FMO-DFT method is also applied to transition states in SN2 reactions and for the computation of the IR and Raman spectra of a small Trp-cage protein (PDB: 1L2Y). Some computational timing analysis is also presented. © 2015 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4915068]

I. INTRODUCTION

The first and second derivatives of the energy with respect to nuclear coordinates are indispensable for quantum-chemical (QM) modelling in chemistry; for instance, for geometry optimizations, transition state searches, and the prediction of vibrational frequencies. The first derivative (gradient) is used in molecular dynamics (MD) simulations, because the forces on the atoms are given by the gradient. The second derivative (Hessian) is used to calculate harmonic frequencies, and many properties related to them, such as the zero point energy (ZPE), the vibrational contribution to the free energy, infrared (IR) intensities, and Raman activities.

The Hessian of the QM energy has been introduced by pioneers such as Gerratt and Mills,1 and Pulay.2 It is usually evaluated by solving coupled perturbed Hartree-Fock (CPHF) equations,3 that are also formulated for other approaches such as the multi-configurational self-consistent field,4 coupled cluster theory with both closed and open shells,5,7 as well as for second order Møller-Plesset perturbation theory (MP2),8 configuration interaction,9 and density functional theory (DFT).10–12

The cost of computing the Hessian is very high even for single reference methods, thus, the calculations are limited to small and medium sized systems. Among single reference approaches, one of the most widely used is DFT,13 which typically involves the numeric integration of exchange-correlation functionals on a grid.14

For computing the Hessian in large molecular systems, there are several options: combined QM and molecular mechanics (MM),15 n-layered integrated molecular orbital and molecular mechanics (ONIOM),16 and density-functional tight-binding (DFTB).17 To accelerate the Hessian evaluation for a subset of atoms, usually, in a numeric fashion, a number of partial schemes is available.18–21 The linear-scaling QM algorithms22–25 have primarily been developed for energy and gradient evaluations; the efficiency of Hessian computations has also been improved.26–28 Some fragment based approaches29–40 feature analytic second derivatives.41–46

The fragment molecular orbital (FMO) method17–51 is a fragment-based approach. In the FMO method, the system is divided into a set of fragments (also called monomers), the electronic state of a fragment is embedded into an electrostatic potential (ESP) that is determined by the nuclei and electron densities of the other fragments. The monomer calculations are thus coupled and iterated to self-consistency. Next, dimer and trimer calculations are performed in the ESP of monomers (fixed at this point). The fragmentation

1Electronic mail: nakata.h.ab@m.titech.ac.jp. Fax: +81-48-467-8503.
2Electronic mail: d.g.fedorov@aist.go.jp
reduces the computational timing significantly. Hence, the FMO method has been applied to proteins, DNA, and inorganic systems.

The discussion in this work focuses on the three-body FMO method (FMO3) combined with DFT. The corresponding two-body FMO expressions can be easily obtained by ignoring the three-body correction terms. A comparison of the accuracy of two and three-body expansions of DFT for the energy and gradient is found elsewhere, and the cost scaling of FMO with system size has already been discussed in general.

The FMO analytic energy gradient has been developed for closed and open-shell Hartree-Fock (HF) and DFT. The analytic second derivative, which can be used to evaluate Raman activities, has been developed only for HF at the two-body level (FMO2). The FMO-RHF Hessian is not fully analytic, because some terms (e.g., contributions from second order responses) that are expected to be small are neglected. All of these terms are also neglected in FMO-DFT Hessians developed here, and no further approximations are introduced.

In Secs. II–V, the FMO3 Hessian is first formulated for the general case. Then, the detailed derivations, specifically for DFT, are presented. All of the DFT-related exchange-correlation terms introduced in this work are fully analytic. The accuracy of the FMO Hessian is evaluated by comparing some properties to those obtained with full (non-fragmented) DFT. Simulations of the IR and Raman spectra of the Trp cage protein (1L2Y) and a study of S2 reactions in solution are also presented. The computational efficiency of FMO/DFT Hessians is illustrated for water clusters.

II. THEORY

A. Analytic second derivative for FMO3

The FMO3 total energy is given by

$$E_{\text{FMO3}} = \sum_I E'_I + \sum_{I<J} \Delta E_{IJ} + \sum_{I>J>K} \left[ \Delta E_{IJK} - \Delta E_{IL} - \Delta E_{JK} - \Delta E_{KI} \right],$$

where the two- and three-body energy contributions are, respectively,

$$\Delta E_{IJ} = E'_I - E'_J - \chi_{IJ},$$

$$\Delta E_{IJK} = E'_{IJK} - E'_K - E'_J - E'_I - \chi_{IJK},$$

where $$E'_I$$ is the internal energy of monomers ($$X = I$$), dimers ($$X = IJ$$), or trimers ($$X = IJK$$), and $$\chi_{IJ}$$ is the corresponding electrostatic potential. The difference between the density matrices of a dimer (trimer) and monomers, defined as

$$\Delta D_{IJ} = D_{IJ} - \left( D'_I \otimes D'_J \right),$$

$$\Delta D_{IJK} = D_{IJK} - \left( D'_I \otimes D'_J \otimes D'_K \right).$$

The second derivative of the energy with respect to nuclear coordinates $$a$$ and $$b$$ is

$$\frac{\partial^2 E_{\text{FMO3}}}{\partial a \partial b} = \sum_I \frac{\partial^2 E'_I}{\partial a \partial b} + \sum_{I<J} \frac{\partial^2 \Delta E_{IJ}}{\partial a \partial b} + \sum_{I>J>K} \frac{\partial^2}{\partial a \partial b} \left[ \Delta E_{IJK} - \Delta E_{IL} - \Delta E_{JK} - \Delta E_{KI} \right].$$

To obtain the analytic second derivative of the FMO3 total energy, one needs to evaluate the X-mer quantities $$\frac{\partial^2 E'_X}{\partial a \partial b}$$. The derivatives of the many-body terms $$\Delta E_{IJ}$$ and $$\Delta E_{IJK}$$ include two types of contributions: (a) differences of $$E'_X$$ values, and (b) density transfer terms $$\text{Tr}(\Delta D_{X}^{IJK} \chi_{IJK})$$. For $$X = IJ$$, the corresponding second derivative expressions can be found in the FMO2-HF Hessian derivation.

The following focuses on the new terms that arise for FMO3 and for DFT. DFT specific terms arise only in the derivatives of the internal energies $$E'_X$$, because the ESP does not contain exchange-correlation terms.

B. Analytic second derivative of the internal energy for DFT

The second derivative of the internal fragment energy is

$$\frac{\partial^2 E'_X}{\partial a \partial b} = \sum_{i \in X} [h_{i}^{a'b} + p_{i}^{a'b} + F_{ij}^{a'b} - \delta_{i}^{a'b} \epsilon_i^X].$$

$$+ \sum_{i,j \in X} [s_{ij}^{a'b} \epsilon_i^X].$$

$$+ \sum_{m \in X} \left[ 4F_{im}^{a'b} - 2s_{ik}^{a'b} \epsilon_i^X - 2s_{jk}^{a'b} \epsilon_j^X \right].$$

$$+ \sum_{k,l \in X} \left[ 2s_{kl}^{a'b} \epsilon_i^X \right].$$

$$+ \sum_{i,j \in X} \left[ 2s_{ij}^{a'b} \epsilon_i^X \right].$$

$$+ \sum_{k,l \in X} \left[ 2s_{kl}^{a'b} \epsilon_i^X \right].$$

$$= \frac{\partial^2 E_{\text{NR}}}{\partial a \partial b} + \frac{\partial^2 E_{\text{ES}}}{\partial a \partial b} - \sum_{i,j \in X} \left[ 2s_{ij}^{a'b} \epsilon_i^X \right].$$

where $$i, j, k, l$$, and $$m$$ denote molecular orbitals (MO) ($$m$$ is used for virtual orbitals, while $$i, j, k$$, and $$l$$ are used for occupied orbitals), $$h_{i}^{a'b}$$ and $$p_{i}^{a'b}$$ are the second derivatives of the one-electron core integrals and the hybrid projection operator integrals, respectively, with respect to coordinates $$a$$ and $$b$$. $$s_{ij}^{a'b}$$ and $$s_{ij}^{a'b}$$ are the first and second derivatives of the overlap integrals, respectively. $$F_{ij}^{a'b}$$ and $$F_{ij}^{a'b}$$ are matrices formed from the first and second integral derivatives of the internal Fock matrix, respectively. $$F_{ij}^{a'b}$$ is obtained from the conventional Fock matrix $$F_{ij}^{a'b}$$ by subtracting the ESP.
the energy of MO $i$ in fragment $X$. $U_{\text{HF},X}^{ab,X}$ is the second order orbital response, whose definition is provided in Subsection II C.

To extend the HF formulation to DFT, one has to modify the internal energy expression by adding the exchange-correlation energy derivatives,

$$\frac{\partial^2 E^X}{\partial a \partial b} = \frac{\partial^2}{\partial a \partial b} \left[ \sum_{A \in X} \sum_{t \in A} w_{At} f_{At}(\rho_\alpha, \rho_\beta, \gamma_{\alpha \alpha}, \gamma_{\alpha \beta}, \gamma_{\beta \beta}) \right],$$

(9)

where $w_{At}$ and $f_{At}$ are the quadrature weights and the exchange-correlation functional at the grid point $t$ belonging to atom $A$. $\alpha$ and $\beta$ are spin labels. The exchange-correlation contributions are evaluated by numerical integration over all atoms $A$ and grid points $t$. Consequently, the internal Fock matrix $F_{ij}^X$ and the orbital Hessian matrix $A_{ijkl}'$ for DFT become

$$F_{ij}^X = h_{ij}^X + \sum_{k \in X} (2(ij|kk) - c_{HF}(ik|jk)) + P_{ij} + (V_{\text{oC}})_{ij},$$

(10)

$$A_{ijkl}' = 4(ij|kl) - c_{HF}((ik|jl) - (il|jk)) + 4(f_{nO})_{ijkl},$$

(11)

where $c_{HF}$ represents the fraction of HF exchange in the functional: $c_{HF} = 0$ for pure DFT (no HF exchange), and $c_{HF}$ is between 0 and 1 for hybrid functionals. $(V_{\text{oC}})_{ij}$ and $(f_{nO})_{ijkl}$ are the matrix elements of the first and second functional derivatives of the exchange-correlation functional with respect to the electron density. In this study, the Becke multicenter numerical integration scheme was applied, but with some minor modifications according to Murray et al.

The second derivative of the exchange-correlation energy contribution can be written as

$$\frac{\partial^2 E^X}{\partial a \partial b} = \sum_{A \in X} \sum_{t \in A} \left[ \frac{\partial^2 w_{At}}{\partial a \partial b} f_{At}(\rho_\alpha, \rho_\beta, \gamma_{\alpha \alpha}, \gamma_{\alpha \beta}, \gamma_{\beta \beta}) 
+ w_{At} \frac{\partial^2 f_{At}(\rho_\alpha, \rho_\beta, \gamma_{\alpha \alpha}, \gamma_{\alpha \beta}, \gamma_{\beta \beta})}{\partial a \partial b} 
+ \frac{\partial w_{At}}{\partial a} \frac{\partial f_{At}(\rho_\alpha, \rho_\beta, \gamma_{\alpha \alpha}, \gamma_{\alpha \beta}, \gamma_{\beta \beta})}{\partial b} 
+ \frac{\partial w_{At}}{\partial b} \frac{\partial f_{At}(\rho_\alpha, \rho_\beta, \gamma_{\alpha \alpha}, \gamma_{\alpha \beta}, \gamma_{\beta \beta})}{\partial a} \right].$$

(12)

All of these terms, including the second derivatives of the quadrature weights, have been implemented in the quantum chemistry program GAMESS, and the other terms are the same as our previous FMO-HF Hessian. It is important to note that the contribution of the weight derivative terms should be included even when very fine grids are used, contrary to what is suggested in other studies. Because of minor differences to other implementations, additional details of the derivations are provided for interested readers.

C. Analytic second derivative of the density transfer terms

The second derivative of the three-body ESP contribution is

$$\frac{\partial^2 \text{Tr} \left( \Delta D_{IJK} V_{IJK} \right)}{\partial a \partial b} = \frac{\partial^2}{\partial a \partial b} \left[ \sum_{\mu,\nu \in IJK} D_{\mu \nu}^{IJK} V_{\mu \nu} - \sum_{\mu,\nu \in IJK} D_{\mu \nu}^{IJK} V_{\mu \nu} \right] - \sum_{\mu,\nu \in IJK} D_{\mu \nu}^{IJK} V_{\mu \nu} - \sum_{\mu,\nu \in IJK} D_{\mu \nu}^{IJK} V_{\mu \nu} ,$$

(13)

where $\mu$ and $\nu$ are the atomic orbital indices. Eq. (13) is a combination of four terms, each of which is given by (for $X = IJK$, $I$, $J$, or $K$)

$$\sum_{\mu,\nu \in X} \frac{\partial^2}{\partial a \partial b} \left[ \sum_{\mu,\nu \in X} \left( \sum_{\mu,\nu \in Y} V_{\mu \nu}^{ab, IJK} \right) - \sum_{\mu,\nu \in X} \left( \sum_{\mu,\nu \in Y} V_{\mu \nu}^{ab, IJK} \right) \right]$$

(14)

where $R_{ab,X}^{ab, X,Y}$ is a two-electron integral response term, $U_{ab,X}^{ab, X,Y}$ is the orbital response term, whose definition is given below. Inserting Eqs. (7)–(14) into Eq. (6), one can calculate the FMO3 Hessian.

The second order orbital response $U_{ab,X}^{ab, X,Y}$ (like $X$, denotes some monomer, dimer, or trimer) appears in the above expressions only in the $Y = X$ or $Y = IJK$ forms, $U_{ab,X}^{ab, X,X}$ and $U_{ab,X}^{ab, X,IJK}$. The dimer terms for $Y = IJ$ are also present in the two-body expressions, which are not explicitly listed above and can be found elsewhere.

The second order responses arise from the derivative of $\sum_{\mu,\nu \in X} U_{\mu \nu}^{ab, X} Y$ with respect to nuclear coordinate $b$. An important finding in the previous studies is that the sum of all orbital response terms cancels out in FMO2 gradients and Hessians if the ESP point charge (PC) approximation, ESP-PC, is applied to all ESPs, or if the ESP-PC approximation is not used at all; otherwise, the orbital response terms need to be evaluated. Now, it is shown that the orbital response terms $U_{ab,X}^{ab, X,Y}$ also cancel out for FMO3, for the same two conditions.

The combination of all three-body orbital response terms appearing in the last sum of Eq. (6) is as follows:

$$\Delta U_{ab,\text{FMO3}} = \sum_{I,J,K} \left[ -U_{ab,1,1JK} - U_{ab,1,1JK} - U_{ab,1,1JK} 
+ \sum_{I,J,K} \left( U_{ab,1,1JK} + U_{ab,1,1JK} + U_{ab,1,1JK} 
+ U_{ab,1,1JK} + U_{ab,1,1JK} \right) \right].$$

(15)

Note that $U_{ab,1,1JK}$ cancels out, because it appears with opposite signs in the internal energy and the density transfer terms, see Eqs. (7) and (14). This cancellation occurs because the ESP is included in the DFT (HF) trimer calculations, and the ESP term was only formally separated into a different term.
in the above expressions. By separating out the contribution $V_{\mu\nu}^{IJ}$ of a single external fragment $J$, 

$$V_{\mu\nu}^{IJ} = \sum_{A \in \mu} \left[ \int \frac{-Z_A}{|r - R_A|} v \right] + \sum_{i,j \in J} D_{ij}^e \partial \tilde{\varphi}_{\mu\nu}(\lambda\sigma),$$  
(16)

one can identify a response contribution $\mathbf{U}^{ab,1,1(J)}$ for the individual fragments $I$, 

$$\mathbf{U}^{ab,1,1(J)} = \sum_{J \in X} \mathbf{U}^{ab,1,1(J)} = \sum_{J \in X} \mathbf{U}^{ab,1,1(J)},$$  
(17)

where 

$$\mathbf{U}^{ab,1,1(J)} = -4 \sum_{i \in X} \sum_{j \in X} \sum_{m \in X} U_{mi}^b X_i (V_{ij}^{IJ}) S_m^a X_i + V_{ij}^{IJ} S_m^a X_i + 4 \sum_{i \in X} \sum_{j \in X} \sum_{m \in X} S_m^a X_i V_{ij}^{IJ} S_m^a X_i - 2 S_m^a X_i V_{ij}^{IJ} + \sum_{m \in X} \sum_{i \in X} \sum_{j \in X} 4 (V_{ij}^{IJ}) S_m^a X_i (\partial V_{ij}^{IJ} / \partial b).$$  
(18)

The orbital response terms in Eq. (15), related to fragment $I$, can be expanded as 

$$\mathbf{U}^{ab,1,1J} = \mathbf{U}^{ab,1,1K} = \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1L} = \sum_{L \in I,J} \mathbf{U}^{ab,1,1(L)} + \sum_{L \in I,K} \mathbf{U}^{ab,1,1(L)} - \sum_{L \in I,J,K} \mathbf{U}^{ab,1,1(L)} = 0.$$  
(19)

Thus, the orbital response terms related to $J$ and $K$ likewise cancel. Therefore, the sum of all three-body terms in Eq. (15) is zero, and it is not necessary to calculate these terms explicitly (unless the ESP-PC approximation is used for some, but not all, ESPs).

If two fragments $I$ and $J$ are far from each other, the ES-DIM approximation may be used. In this case, $ΔE_{IJ}$ obtained from a DFT dimer $IJ$ is approximated and the contribution to the orbital responses in the last term of Eq. (6), $-\mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J}$, is replaced by $-\mathbf{U}^{ab,1,1J}$ as derived elsewhere for the gradient. A similar relation holds for the Hessian, and $-\mathbf{U}^{ab,1,1J} + \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J}$ are the same, because 

$$-\mathbf{U}^{ab,1,1J} + \mathbf{U}^{ab,1,1J} - \mathbf{U}^{ab,1,1J} = \mathbf{U}^{ab,1,1J},$$  
(20)

Therefore, the use of the ES-DIM approximation does not affect the fact that $\Delta \mathbf{U}^{ab,FMO}$ is equal to zero.

If the ESP-PC approximation is not used, the second order responses are not needed. The first order responses must be calculated for all monomers, dimers, and trimers (they appear in Eqs. (7) and (14)). To obtain the FMO3 Hessian, one must add terms related to the internal energy derivatives in Eq. (7), as well as the density transfer in Eq. (14). The second order response terms in each of these equations are separated in Eq. (15), where they cancel out. To obtain the first order response terms $U_{\mu\nu}$, one has to solve the CPHF equations. The CPHF equations in GAMESS have been extended to DFT by using an appropriate DFT form of the orbital Hessian $\mathbf{A}_{ijkl}^{FMO}$ and Fock matrix $F_{ij}^{X}$, following Eqs. (10) and (11), respectively.

IR intensities and Raman activities in FMO-DFT are calculated analogously to FMO-RHF using DFT instead of RHF in a trivial extension.

Finally, note that there is a FMO3 specific approximation, which neglects the contribution of those trimers in which fragments are far apart. This approximation may be used with all of the above derivations, as it simply restricts the sum of $\sum_{I \neq J \neq K}$ from running over all distinct fragment triples (trimers) to a subset of compact trimers.

### III. COMPUTATIONAL DETAILS

First, a conventional analytic DFT Hessian program was implemented in GAMESS and parallelized similarly to the existing RHF Hessian. Then, the FMO-DFT analytic Hessian was implemented in GAMESS and was parallelized using the generalized distributed data interface (GDDI) and spherical harmonics was used.

First, the accuracy of the FMO3-DFT Hessian is evaluated in comparison to the full DFT Hessian. For $(\text{H}_2\text{O})_\text{10}$, three different types of exchange-correlation functionals (B3LYP, BLYP, and LC-BLYP) were used with the default grid (96 radial points in the Euler-MacLaurin quadrature and 302 angular points in the Lebedev grid). For a test of the accuracy in systems in which fragments were connected by covalent bonds, a capped alanine decamer (ALA)$_\text{10}$ and a styrene oligomer were also calculated (Fig. 1).

Second, the FMO-DFT Hessian (with the LC-BLYP functional) was calculated to obtain the vibrational spectrum of a small Trp-cage miniprotein construct (PDB: 1L2Y), consisting of 304 atoms, whose geometry was optimized for aqueous solution using the polarizable continuum model (PCM) at the FMO3-DFT/PCM(1) level until the root mean square gradient dropped below $4 \times 10^{-5}$ hartree/bohrs.

Third, the transition states for the S$_\text{2}$2 reaction XCH$_2$ + OH$^-$ → CH$_3$ + X$^-$, X = Br, Cl, or F, were studied in explicit solvent, namely, 41 waters (the total number of atoms is 130), using B3LYP. In the calculations of the PCM solvated Trp-cage protein and explicitly solvated S$_\text{2}$2 reactions, an empirical dispersion (D) correction version 3 was used (denoted DFT-D), with the standard grid (SGI) in DFT.

Finally, the computational efficiency was demonstrated on a cubic box of water molecules, containing 16, 21, 54, 122, 136, and 165 water molecules using BLYP.

With the exception of the styrene oligomer, all FMO calculations were performed with FMO3; in styrene, the magnitude of the three-body corrections is small, so FMO2 was used. Hessian calculations were performed on stationary points. Unless otherwise indicated, geometry optimizations and transition state searches used a convergence criterion of 0.0001 hartree/bohr, the ES-DIM approximation was used.
FIG. 1. Fragmentation of (a) capped alanine decamer, (b) styrene oligomer, (c) Trp-cage protein (1L2Y).

with a unitless threshold of 2.0, and the default long distance trimer cutoff values were applied for all calculations. The ESP-PC approximation was not used during all Hessian calculations. In the simulations of vibrational spectra, Gaussian fitting was used with a peak width of 20 cm$^{-1}$. For comparison to experiment, all frequencies for the Trp-cage protein were scaled by 0.9915.

Molecular systems were divided into 1 water molecule or 1 amino acid residue per fragment. For styrene, the system was separated into two monomer units per fragment. For SN$_2$ reactions, the reaction center was treated as 1 fragment, and each water molecule was treated as a fragment. FragIt was used to generate input files.

In order to calculate the ZPE, Helmholtz free energy ($H$), and Gibbs free energy ($G$), statistical thermodynamics based on the harmonic oscillator/rigid rotor approximation is applied. The vibrational (vib) contributions are given by

$$\text{ZPE} = \frac{1}{2} \sum_i \nu_i,$$

$$H_{\text{vib}} = \sum_i \frac{\nu_i}{\exp(\frac{\nu_i}{kT}) - 1} + \text{ZPE},$$

$$S_{\text{vib}} = \sum_i \left[ \frac{\nu_i}{T(\exp(\frac{\nu_i}{kT}) - 1)} - k \ln \left( 1 - \exp(-\frac{\nu_i}{kT}) \right) \right],$$

$$G_{\text{vib}} = H_{\text{vib}} - TS_{\text{vib}},$$

where $\nu_i$ is the vibrational frequency of normal mode $i$, $k$ is the Boltzmann constant, and $T$ is the temperature. Thus, the higher the frequency, the larger is the contribution to the ZPE. On the other hand, as discussed elsewhere, for the vibrational partition function and quantities derived from it (the entropy and free energies), low frequency vibrations make a large contribution. The vibrational contributions shown above are added to the translational and rotational terms to obtain the full free energies $H$ and $G$. Therefore, comparisons between the full ab initio and FMO predictions for ZPE, $H$, and $G$ are good benchmarks for the accuracy of high and low vibrational frequencies.

IV. RESULTS AND DISCUSSION

A. FMO accuracy

A summary of prominent IR peaks for (H$_2$O)$_{10}$ is given in Table I, and IR spectra are shown in Fig. 2(a) for the B3LYP density functional. The three peaks in Table I were chosen to represent high, medium, and low frequency modes. Both frequencies and intensities obtained with FMO3-DFT agree reasonably well with the full DFT for all three exchange-correlation functionals used in this study. For example, in the case of B3LYP, the FMO frequency errors compared to the full DFT Hessian are 2, 4, and 19 cm$^{-1}$ for $\nu_1$, $\nu_2$, and $\nu_3$, respectively, and the corresponding intensity errors...
IR spectra of (a) water cluster. Three prominent IR peaks, whose normal modes extend over multiple water molecules according to the full DFT calculation. By including trimers of water molecules in FMO3, it is possible to accurately evaluate all of the frequencies, and the error in the intensity is 20% or less (depending on the functional). In general, Fig. 2(a) shows that FMO3 reliably predicts the IR spectra of water clusters.

Results for the polyalanine and styrene oligomer are shown in Figs. 2(b) and 2(c), and Table II. Once again, the FMO-DFT results are in reasonable agreement with full DFT. The errors of the three prominent peaks in polyalanine are 8, 2, and 13 cm\(^{-1}\), and the errors in IR intensity are 0.07, 0.19, and 0.13 D\(^2/(u \, \AA^2)\), respectively. In the styrene oligomer, the errors in vibrational frequency and IR intensity do not exceed 1 cm\(^{-1}\) and 0.01 D\(^2/(u \, \AA^2)\), respectively. Water and the polypeptides have hydrogen bonding with significant three-body effects, whose accurate treatment necessitates FMO3, whereas in the styrene oligomer, the three-body effects are relatively small and FMO2 was used. The root mean square deviation (RMSD) between FMO-DFT and full DFT for all vibrational frequencies for the alanine decamer is 2 cm\(^{-1}\), and 0.5 cm\(^{-1}\) for the styrene oligomer. The maximum error among all frequencies is 10 cm\(^{-1}\) for the alanine decamer, and 5 cm\(^{-1}\) for the styrene oligomer.

A comparison of the vibrational energies is given in Table III. The differences between FMO-DFT and full DFT quantities for ZPE, Gibbs free energy (\(G\)), and Helmholtz free energies (\(H\)) are small: the errors in polyalanine are 0.19, 0.47, and 0.09 kcal/mol, respectively; for the styrene oligomer, they are 0.01, 0.98, and 0.55 kcal/mol, respectively. For the water cluster, the largest error is 0.56 kcal/mol. Using FMO3 may presumably reduce the error of 0.98 kcal/mol in the FMO2 Gibbs free energy for the styrene oligomer. The fragmentation errors in the FMO3 thermodynamic quantities described here are smaller than those previously reported\(^{46}\) for FMO2-HF. The Gibbs free energies reported here are accurate to 1 kcal/mol, although they depend on the low frequency vibrations via the entropy, and the low frequency vibrations are often delocalized and notoriously difficult to estimate.

### Table II. Three prominent IR peaks \(\nu_i\), computed with LC-BLYP-D/6-31G(d). The units for frequencies and intensities are cm\(^{-1}\) and D\(^2/(u \, \AA^2)\), respectively.

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>IR intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\nu_1)</td>
<td>(\nu_2)</td>
</tr>
<tr>
<td>Polyalanine</td>
<td>645</td>
<td>1795</td>
</tr>
<tr>
<td>Styrene oligomer</td>
<td>721</td>
<td>1523</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>IR intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\nu_1)</td>
<td>(\nu_2)</td>
</tr>
<tr>
<td>Polyalanine</td>
<td>653</td>
<td>1797</td>
</tr>
<tr>
<td>Styrene oligomer</td>
<td>722</td>
<td>1523</td>
</tr>
</tbody>
</table>

\(^{46}\)FMO3.
\(^{46}\)FMO2.

### B. IR and Raman spectra of the Trp-cage protein

The IR and Raman spectra of the Trp-cage protein were simulated with the FMO3-DFT-D Hessian, computed at the FMO3-DFT-D/PCM(1) optimized geometry. The RMSD of the optimized structure from the NMR experiment (the first
structure in the PDB75 is 0.8343 Å. The largest contributions to the RMSD are in flexible side chains and the two termini, whereas the backbones are similar. The calculated and experimental structures are shown in Fig. 3(a).

A comparison of the calculated and experimental vibrational frequencies is given in Table IV, where a range of calculated frequencies for each type of vibration is specified for normal modes of the same nature (because of the coupling of multiple groups of the same kind). Most of the calculated vibrational frequencies agree well with the experiment.76 The difference is 7 cm$^{-1}$ for the Tyr symmetric mode, and the Amide II and III peaks very closely reproduce the experimental values. On the other hand, the Amide I peak has a deviation of about ±50 cm$^{-1}$ from the experiment. The reason for the discrepancy may be the lack of solvent in the Hessian calculations. It was previously reported that Amide I peaks are shifted by ~30 cm$^{-1}$ by adding explicit solvent in a small polypeptide,74 because of the coupling of the Amide I normal modes to water bending modes. Likewise, COO$^{-}$ modes are strongly affected by solvent (~59 cm$^{-1}$ as reported elsewhere76), whereas Amide II were found76 to be weakly affected.

The calculated IR and Raman spectra of the Trp-cage protein are shown in Figure 3(b). The vibrational modes around 1611 cm$^{-1}$ are mainly localized on the ammonium cation in Arg, and the modes around 1769 cm$^{-1}$ are Amide I vibrational modes that reflect inter-residue hydrogen bond coupling in the protein. The most Raman active vibrational peak is the CH$\_2$ stretching mode along the backbone main chain; these CH$\_2$ modes are not IR active modes, since the dipole derivative contributions are expected to be small. The symmetric and antisymmetric COO$^{-}$ stretching modes are observed at 1445 and 1718 cm$^{-1}$, respectively, whose frequencies may be underestimated somewhat by the omission of solvent, if the trend found in another polypeptide74 holds in general.

C. Transition state analysis for S$\_\text{n2}$ reactions

The activation free energies for the S$\_\text{n2}$ reactions between XCH$_3$ and OH$^{-}$ (X = Br, Cl, or F) immersed in 41 water molecules are summarized in Table V. The errors in the free energy barrier compared to full DFT are 2-3 kcal/mol. The trends within the series of halogen atoms are correctly described by the FMO3-DFT method. In order to improve the accuracy in practical applications, one could try to merge some water molecules of the first solvation shell with the reaction center into one fragment, however, this was not investigated in the present work.

To understand their origin, some components of the total free energy barriers $\Delta G$ are listed in Table V. High frequencies that make large contributions to the ZPE and low frequencies that make large contributions to the vibrational free energy have some effect. Also, the error due to vibrational terms is comparable to the error in the electronic energy. This system
TABLE V. Imaginary vibrational frequencies $\omega$ (cm$^{-1}$) and contributions (kcal/mol) to the S$_2$C reaction barriers for XCH$_3$ and OH$^-$ (X = Br, CI, and F) solvated in 41 water molecules: electronic energy $\Delta E$, zero point energy $\Delta ZPE$, vibrational contribution $\Delta G_{\text{vib}}$ to the Gibbs free energy $\Delta G$ (computed with B3LYP/6-31G(d) at T = 298.15 K).

<table>
<thead>
<tr>
<th>Method</th>
<th>F</th>
<th>Cl</th>
<th>Br</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMO3-DFT $\Delta E$</td>
<td>24.57</td>
<td>14.18</td>
<td>7.33</td>
</tr>
<tr>
<td>Full-DFT $\Delta E$</td>
<td>23.88</td>
<td>15.05</td>
<td>5.66</td>
</tr>
<tr>
<td>FMO3-DFT $\Delta ZPE$</td>
<td>0.59</td>
<td>0.28</td>
<td>3.14</td>
</tr>
<tr>
<td>Full DFT $\Delta ZPE$</td>
<td>0.01</td>
<td>1.33</td>
<td>3.10</td>
</tr>
<tr>
<td>FMO3-DFT $\Delta G_{\text{vib}}$</td>
<td>3.63</td>
<td>3.38</td>
<td>4.56</td>
</tr>
<tr>
<td>Full DFT $\Delta G_{\text{vib}}$</td>
<td>1.39</td>
<td>4.55</td>
<td>3.84</td>
</tr>
<tr>
<td>FMO3-DFT $\Delta G$</td>
<td>28.20</td>
<td>17.56</td>
<td>11.89</td>
</tr>
<tr>
<td>Full DFT $\Delta G$</td>
<td>25.27</td>
<td>19.60</td>
<td>9.50</td>
</tr>
<tr>
<td>FMO3-DFT $\omega$</td>
<td>486</td>
<td>419</td>
<td>324</td>
</tr>
<tr>
<td>Full DFT $\omega$</td>
<td>470</td>
<td>419</td>
<td>374</td>
</tr>
</tbody>
</table>

is an anion in explicit solvent; due to solute-solvent charge delocalization, some FMO-DFT error is observed.

The imaginary vibrational frequencies obtained with the FMO-DFT Hessian agree reasonably well with those obtained with the full DFT Hessian. The larger error found for Br is attributed to the fact that the energy barrier is small, and the potential energy surface is flat. Nevertheless, the monotonic decrease of frequency from F to Cl to Br is properly reproduced by the FMO method. The decrease in the barrier height and imaginary frequency in the series F, Cl, and Br is attributed to the larger atomic radius and the core shielding effect on the valence electrons, as well as the increase in the atomic mass.

D. Computational efficiency

The timings for the computation of the FMO3 Hessians were measured on a PC cluster containing 16 2.93 GHz Xeon nodes (8 CPU cores and 12 GB memory per node). For this purpose, several water molecular clusters were used, with system sizes ranging from 16 to 165 water molecules. The results are shown in Fig. 4. The FMO3-BLYP Hessian took about 21 h for the largest system (495 atoms) in this study. For comparison, the timing of the largest cluster for B3LYP is 20.5 h, while for BLYP it is 21.4 h, as self-consistent monomer calculations for B3LYP converge faster than for BLYP.

The largest water cluster in this study (495 atoms) took 21 and 18 h with FMO3-DFT and FMO3-HF, respectively, and the overall trend in Fig. 4 shows that DFT and HF Hessians for FMO are comparable in terms of computational time requirements. This is because the main cost is in the evaluation of two-electron integrals and orbital responses, while the extra cost of DFT-related grid integration is relatively minor.

Finally, it is of interest to compare the time requirements for FMO3-DFT to full DFT Hessian calculations. The latter demands a significant amount of memory, so a different single node was used, with 24 cores and 128 GB of memory. Total run times for FMO3-DFT or full DFT Hessians are 50 or 187 min for H$_2$O$_{16}$, respectively, and 248 or 5142 min for H$_2$O$_{31}$. Thus, FMO3 decreases the run time by a factor of 21 for 31 waters, which is roughly the largest full Hessian run that may be done with the memory that is available on the nodes used for this study.

V. CONCLUSIONS

Analytic second derivatives of the energy with respect to nuclear coordinates have been derived for closed shell spin-restricted FMO3-DFT and implemented into GAMESS. As described in a previous paper,$^{46,74}$ certain terms that are expected to be small are neglected in the FMO-RHF and FMO-DFT Hessians, so the FMO Hessian code implemented in GAMESS is not fully analytic but the magnitude of the neglected terms is typically very small. FMO2-DFT and FMO3-HF Hessian calculations are enabled as well. Full DFT Hessians have also been coded, including an implementation for long-range corrected functionals,$^{90}$ such as LC-BLYP, which improve the treatment of the electrostatics. The proper account of electrostatics is thought to be important in polar systems, such as water or proteins. The new development of FMO Hessians permits simulations of IR and Raman spectra and vibrational contributions to the energetics, as well as DFT transition state searches.

It has been demonstrated on a representative set of test systems (IR spectra and chemical reactions in solution) that the accuracy of FMO3-DFT properties derived from the Hessian is reasonably high. The FMO approach significantly reduces the computational timing (by a factor of 21 for 93 atoms as reported in this study), and the memory requirements, both of which preclude full DFT Hessian evaluations in large molecular systems.

As a demonstrative application, a FMO3-DFT geometry optimization and Hessian evaluation for the small Trp-cage protein was performed. The vibrational frequencies obtained show reasonable agreement with experiment.

The present development enables DFT Hessian calculations for large systems. As opposed to the gradient, where individual orbital responses are not needed owing to the formulation of the Z-vector$^{3,66}$ equations, the Hessian requires
these responses, and this makes such Hessian calculations rather expensive. Nevertheless, as demonstrated in this work, the FMO3-DFT Hessian is accurate and tractable for systems containing several hundreds of atoms.

ACKNOWLEDGMENTS

This work was in part supported by the Next Generation Super Computing Project, Nanoscience Program (MEXT, Japan) and Computational Materials Science Initiative (CMSI, Japan), and JSPS KAKENHI Grant No. 262235. H.N. acknowledges the use of the following computer resources: Japan Super Computing Project, Nanoscience Program (MEXT, Japan), and JSPS KAKENHI Grant No. 262235. H.N.


10See supplementary material at http://dx.doi.org/10.1063/1.4915068 for the details of the grid related derivative terms, two-electron integral response terms, IR spectra of water using other functionals.