T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	振動分光および生体金属分子系の計算を目指したＦＭＯ法の開発
Title(English)	Development of Fragment Molecular Orbital Method toward the Application to Vibration Spectra and Inorganic Biochemistry
著者(和文)	中田浩弥
Author(English)	Hiroya Nakata
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第9811号, 授与年月日:2015年3月26日, 学位の種別:課程博士, 審査員:櫻井　実,藤井　正明,小長谷　明彦,田口　英樹,大谷　弘之,中村　振一郎
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9811号, Conferred date:2015/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Science Tokyo Research Repository)

Development of Fragment Molecular Orbital Method toward the Application to Vibration Spectra and Inorganic Biochemistry

Hiroya Nakata

Development of Fragment Molecular Orbital Method toward the Application to Vibration Spectra and Inorganic Biochemistry

Hiroya Nakata[1,2]

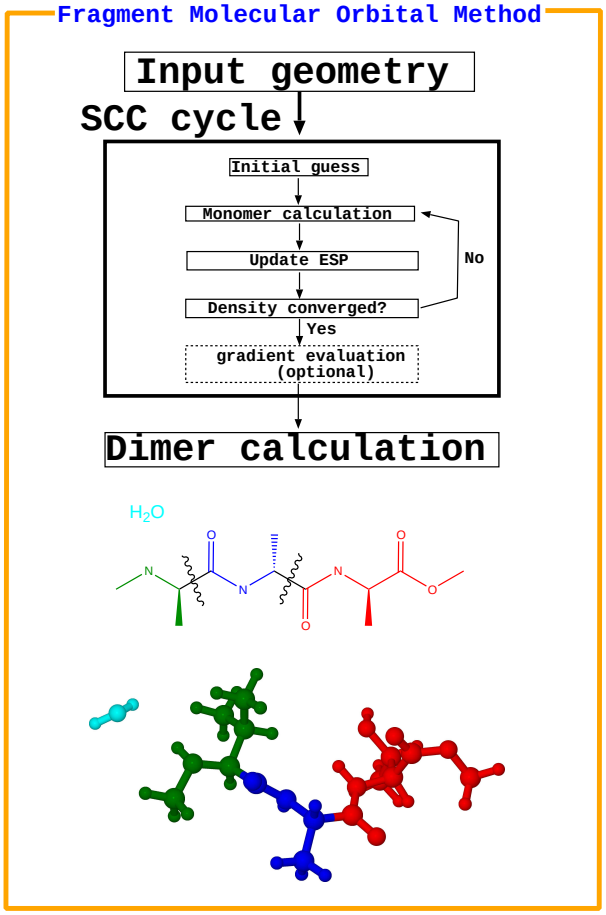
1, *Center for Biological Resources and Informatics Tokyo Institute of Technology B-62 4259 Nagatsuta-cho, Midori-ku Yokohama 226-8501, Japan*

2, *RIKEN, Research Cluster for Innovation, Nakamura Lab, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*

CONTENTS

Chapter I: Introduction

Why the fragment molecular orbital method is necessary?



I. INTRODUCTION: WHY THE FRAGMENT MOLECULAR ORBITAL METHOD IS NECESSARY?

A. Quantum chemistry in biomolecule

The application of quantum chemistry is widely varied over various field on chemistry such as physical, organic, analytical, inorganic, and biological chemistry. The wide range of applicability perhaps comes from the ability to estimate significantly abundant physical property within a single program package. The molecular structure is obtained by the geometry optimization, many spectroscopic properties such as UV-Vis, NMR, IR, Raman scattering, and ESR become available. The reaction free energy barrier can be estimated within several kcal/mol accuracy by the reaction coordinate analysis. Moreover, the combination with other various methods such as multi scale physics and macroscopic simulations can be applied to wide range of problems in material and life science.

The wide spread use of quantum chemistry is due to the simple and easy way toward applying the quantum chemistry programs. There are more than a dozen of free or commercialized quantum chemical software package such as GAMESS[1,2], NwChem[3], QChem[4], and GAUSSIAN[5]. Organic chemist may use for evaluation of the structure of organic molecule, the consideration of reaction mechanism, and prediction of reaction product. The analytical chemists use for the interpretation of the property of spectra such as vibrational frequency and intensity. Inorganic chemist use quantum chemistry combined with ligand field theory, and evaluation of the synthesized molecules. Most of these calculations can be routinely performed with the available QM software. However, there is not yet the established approach that can be applied to biological systems especially containing the transition metals to predict various physical properties within the single QM program.

Applying the QM method to biomolecular systems, the hybrid approach of QM and molecular mechanic (MM) is perhaps most widely used method,[6–8] and typically called QM/MM approach. QM/MM have been applied to quite a large number of systems in order to gain an insight for protein structure and the reaction mechanism in the biological systems.[9,10] The QM/MM method is widely used, because of its simple structure, and most of the quantum

chemical software packages have been interfaced with the major MM programs, especially in biological field, for examples, the GROMACS[11], CHARMM[12], AMBER[13], Tinker[14], and AMOEBA[15] force field are applied many interesting physical problems.

Despite the above success of QM/MM approaches, several difficulties are still remained for the application of proteins especially to the metal containing enzyme, or the property evaluation of extensively complex biological phenomena. Firstly, the MM method requires the parameters for the force field, and sometimes appropriate parameters are not available for transition metals. Secondly, the boundary region between QM and MM is not always reliable because of the imbalance of the theory. Polarization of MM field is often ignored, and only the point charge contribution is considered. Thus the self consistent field between QM and MM is not satisfied. Due to these failure of QM/MM approach, the accuracy of QM/MM method depends significantly on the size of QM region, and the application of the QM/MM method to biological system is not yet straightforward compared to the other field of chemistry. There is a possibility that these disadvantage of QM/MM approach may be completely solved in future by the researcher in the field of quantum chemistry. Several recent papers are providing the perspective for the further discussion of QM/MM method toward the poralizable QM/MM force field[16], and size boundary treatments[17–19].

Alternatively, the fully quantum chemical treatment of the entire protein system is also attractive. The difficulty to apply the quantum chemistry to large biomolecular system is the steep increase in the computational cost with the system size, while there are several advantages in the fully quantum chemical treatment: the parameter for MM is not necessary, and boundary problem as noted in QM/MM method dose not matter. This motivates to develop linear-scaling algorithms[20–24] and fragment-based methods,[25–38] toward various biological systems.

Fragment molecular orbital (FMO) method is one of the fragmentation approach[39,40]. The system is divided into small subset of fragments, and energy, gradient, or Hessian is calculated independently in a presence of bath of electrostatic potential (ESP). The ESP is iteratively updated until the respective fragment electron densities are mutually converged. The FMO can be performed with many wave functions,[41–47] and have been applied to biological protein,[48–50] DNA,[51] organic,[52] and inorganic systems.[53–57] The fully analytic energy

gradient is available for Hartree-Fock (HF)[58], Møller-Plesset perturbation theory (MP2)[59]. For HF and Density functional theory (DFT), open-shell calculation is also possible with restricted open-shell method[60,61].

However, the applications of these fragmentation approach including FMO have been mostly limited within only the single point energy, the properties of electronic structure, and in some part geometry optimizations. While there have been quite a few examples for the application to the normal mode analysis for the vibrational spectra, reaction analysis, and the protein enzyme reactions.[62,63] Especially, if one focuses on the system containing the transition metal that typically possesses the unpaired electrons, there have been no application examples for the vibrational analysis, and analysis of reaction mechanism. These limitations of application territory in fully QM approach come from the difficulty to develop the theory and methodology for the first and second order derivative of the energy with respect to the molecular coordinates.

The aim in this thesis is the new theory and methodological development to calculate fully analytic first and second order derivative of the energy (gradient and Hessian) to obtain the absorption spectrum, vibration spectrum, and reaction analysis in biomoleculer systems, and extend the ability to calculate the metal enzyme containing many unpaired electrons. In this introductory chapter I, first, the fundamental derivative technique for the conventional *ab initio* calculation is introduced, then the difficulty to apply the method directly to the biological systems is explained. Secondly, the basis of FMO method is introduced, and explained how the implementation of analytic gradient and Hessian is difficult. Finally, the aim of this thesis is explained again in detail, and outline of this thesis is briefly described.

B. The derivative technique for Hartree-Fock theory

The Hartree-Fock (HF) approximation is the central to quantum chemistry. Although the accuracy of HF energy is not always good for a practical application purpose, HF is the good starting points as an initial guess before inclusion of electronic correlation. Furthermore, the HF theory also plays an important role as the starting point for the methodology developments (See more detail for HF theory[64]). Because, from both practical and methodology

development point of view, the HF equation is solved for most of the wave functions such as Møller-Plesset perturbation theory (MP2), Coupled Cluster (single and double (CC(SD)), Configuration interaction (CI), and Density functional theory (DFT). Then the specific terms for respective wave functions are additionally calculated for each method. Therefore, HF is the starting point of theoretical quantum chemistry, and this is valid not only for the energy calculations but also the gradient, Hessian, and excited state energies. Even though the gradient and Hessian in HF method are necessary knowledge for any methodological development studies of energy derivatives, the derivative techniques for HF theory are not well known techniques compared to the HF energy itself. Therefore, in this section, the first and second order derivative of energy are briefly introduced (For further reading of the energy derivatives, see the ref.[65].).

1. HF energy gradient

The HF energy E^{HF} is described as

$$E^{\mathrm{HF}} = \sum_{i}^{\mathrm{occ}} 2h_{ii} + \sum_{i}^{\mathrm{occ}} \sum_{j}^{\mathrm{occ}} \{2(ii|jj) - (ij|ij)\}, \tag{1}$$

where i and j denote the molecular orbitals: the linear combination of atomic orbitals, (throughout this thesis, the Greek indices (i, j, k, and l) are used for the molecular orbitals, while Roman indices (μ, ν, λ, and σ) are used for atomic orbitals). The occ denotes for the occupied molecular orbital basis, h_{ii} is the one-electron integral (the contribution of kinetic, and atomic nuclear attraction), $(ii|jj)$ is the electronic coulomb interaction, and $(ij|ij)$ is the exchange repulsion.

Therefore, the derivative of the HF energy E^{HF} with respect to the nuclear coordinate a is

$$\frac{\partial E^{\mathrm{HF}}}{\partial a} = \sum_{i}^{\mathrm{occ}} 2h_{ii}^{a} + \sum_{i}^{\mathrm{occ}} \sum_{j}^{\mathrm{occ}} \{2(ii|jj)^{a} - (ij|ij)^{a}\} - 2\sum_{i,j}^{\mathrm{occ}} S_{ij}^{a} F_{ij} + \sum_{m}^{\mathrm{vir}} \sum_{i}^{\mathrm{occ}} 4U_{mi}^{a} F_{mi}, \tag{2}$$

where h^a_{ii}, $(ii|jj)^a$, $(ij|ij)^a$ and S^a_{ij} is the derivative of integral terms as described

$$h^a_{ii} = \sum_{\mu,\nu} c_{\mu i} c_{\nu i} h^a_{\mu\nu} \tag{3}$$

$$(ij|kl)^a = \sum_{\mu,\nu} c_{\mu i} c_{\nu j} c_{\lambda k} c_{\sigma l} (\mu\nu|\lambda\sigma) \tag{4}$$

$$S^a_{ij} = \sum_{\mu,\nu} c_{\mu i} c_{\nu i} S^a_{\mu\nu} \tag{5}$$

where $S_{\mu\nu}$ is the overlap integral in atomic orbital basis, and F_{mi} and U^a_{mi} are the Fock matrix elements and unknown response terms that is directly related with the derivative of the molecular orbital coefficient $\frac{\partial c_{\mu i}}{\partial a}$ as

$$F_{ij} = h_{ij} + \sum_{k}^{\text{occ}} \{2(ij|kk) - (ik|jk)\}, \tag{6}$$

$$\frac{\partial c_{\mu i}}{\partial a} = \sum_{m}^{\text{all}} U^a_{mi} c_{\mu m}, \tag{7}$$

where all denotes the all molecular orbitals, while vir denotes the virtual molecular orbitals.

In the variation condition, the energy is the minimum with respect to the molecular orbital coefficients $c_{\mu i}$. In other words, the derivative of the energy with respect to the molecular orbital coefficients $c_{\mu i}$ is always zero, thus the Fock matrix satisfies the following relationship as

$$F_{ij} = \delta_{ij} \epsilon_i, \tag{8}$$

where ϵ_i is the orbital energy. Another important technique is the orthogonality of the molecular orbitals:

$$S_{ij} = \delta_{ij}, \text{ where } S_{ij} = \sum_{\mu,\nu} c_{\mu i} S_{\mu\nu} c_{\nu j}. \tag{9}$$

By taking the derivative of the both side of equation with respect to the atomic coordinate a, one obtains

$$U^a_{ij} + U^a_{ji} = -\sum_{\mu,\nu} c_{\mu i} c_{\nu j} \frac{S_{\mu\nu}}{\partial a}. \tag{10}$$

Using the relationship described in Eq. (8), and (10), one can obtain the final formulation for the analytic energy gradient of HF energy as following,

$$\frac{\partial E^{\mathrm{HF}}}{\partial a} = \sum_{i}^{\mathrm{occ}} 2h_{ii}^{a} + \sum_{i}^{\mathrm{occ}}\sum_{j}^{\mathrm{occ}} \{2(ii|jj)^{a} - (ij|ij)^{a}\} - 2\sum_{i,j}^{\mathrm{occ}} S_{ii}^{a}\epsilon_{i}. \tag{11}$$

Comparison between the Eq. (2) and (11) implies that the contribution of unknown response terms U_{mi}^{a} disappeared because the HF satisfy the variation condition described in Eq. (8). The physical meaning of unknown response terms in analytical energy gradient is the contribution of first order perturbation arising due to the rotation between occupied and virtual molecular orbitals. Not only the calculation of analytic energy gradient, but also most of the quantum chemical calculation methods use the variation condition, in order to simplify the computational programing.

2. HF energy second order derivative

The derivation of the energy second order derivative with respect to the nuclear coordinate a and b can be obtained formally from Eq. (2),

$$\frac{\partial^{2} E^{\mathrm{HF}}}{\partial a \partial b} = \frac{\partial}{\partial b}\left[\sum_{i}^{\mathrm{occ}} 2h_{ii}^{a} + \sum_{i}^{\mathrm{occ}}\sum_{j}^{\mathrm{occ}} \{2(ii|jj)^{a} - (ij|ij)^{a}\} - 2\sum_{i,j}^{\mathrm{occ}} S_{ij}^{a}F_{ij} + \sum_{m}^{\mathrm{vir}}\sum_{i}^{\mathrm{occ}} 4U_{mi}^{a}F_{mi}\right], \tag{12}$$

and it can be expanded to

$$\begin{aligned}\frac{\partial^{2} E^{\mathrm{HF}}}{\partial a \partial b} =& 2\sum_{i}^{\mathrm{occ}} h_{ii}^{ab} + \sum_{i,j}^{\mathrm{occ}} \left\{2(ii|jj)^{ab} - (ij|ij)^{ab}\right\} + \sum_{i}^{\mathrm{all}}\sum_{j}^{\mathrm{occ}} U_{ij}^{ab}F_{ij} + 4\sum_{i}^{\mathrm{all}}\sum_{j}^{\mathrm{occ}} \left(U_{ij}^{a}F_{ij}^{b} + U_{ij}^{b}F_{ij}^{a}\right) \\ &+4\sum_{i,j}^{\mathrm{all}}\sum_{k}^{\mathrm{occ}} F_{ij}U_{ik}^{a}U_{jk}^{b} + \sum_{i,k}^{\mathrm{all}}\sum_{j,l}^{\mathrm{occ}} U_{ij}^{a}U_{kl}^{b}A_{ij,kl},\end{aligned} \tag{13}$$

where U_{ij}^{ab} is the second order derivative of the unknown response terms, and $A_{ij,kl}$ and F_{ij}^{a} are the orbital Hessian matrix (the second order derivative of the energy with respect to the molecular orbital energy), and Fock matrix derivative as following,

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk), \tag{14}$$

$$F_{ij}^{a} = h_{ij}^{a} + \sum_{k}^{\mathrm{occ}} \{2(ij|kk)^{a} - (ik|jk)^{a}\}. \tag{15}$$

The second order derivative of orbital response terms U_{ij}^{ab} in Eq. (13) is the contribution of mixing virtual and occupied molecular orbital arising due to the second order perturbation of the energy with respect to the molecular coordinate.

Similar to the energy gradient derivation, the variation condition in Eq. (8) can be used to simplify the above equations, and one obtained the final equation used in the practical computer programs:

$$\begin{aligned}\frac{\partial^2 E^{\mathrm{HF}}}{\partial a \partial b} =& 2\sum_{i}^{\mathrm{occ}} h_{ii}^{ab} + \sum_{i,j}^{\mathrm{occ}} \left\{ 2(ii|jj)^{ab} - (ij|ij)^{ab} \right\} - 2\sum_{i}^{\mathrm{occ}} S_{ii}^{ab}\epsilon_i + 4\sum_{i,j}^{\mathrm{occ}} S_{ij}^{a} S_{ij}^{b}\epsilon_i \\ & - \sum_{i}^{\mathrm{occ}}\sum_{j}^{\mathrm{occ}} S_{ij}^{b} \left[4F_{ij}^{a} - 2\sum_{k,l}^{\mathrm{occ}} A_{ij,kl} S_{kl}^{a} \right] - \sum_{i}^{\mathrm{occ}}\sum_{j}^{\mathrm{occ}} S_{ij}^{a} \left[4F_{ij}^{b} - 2\sum_{k,l}^{\mathrm{occ}} A_{ij,kl} S_{kl}^{b} \right] \\ & + \sum_{m}^{\mathrm{vir}}\sum_{i}^{\mathrm{occ}} U_{mi}^{b} \left[4F_{mi}^{a} - 4S_{mi}^{a}\epsilon_i - 2\sum_{j,l}^{\mathrm{occ}} A_{jl,im} S_{jl}^{a} \right]. \end{aligned} \tag{16}$$

Consequently, the complex terms related to the second order derivative of response terms U_{ij}^{ab} can be neglected in the practical implementations. The remaining terms in the Eq. (16) are the first order response contributions from U_{ij}^{a}, and it is necessary to solve the Coupled perturbed Hartree-Fock (CPHF) equation to obtain the U_{ij}^{a}.

3. Coupled perturbed Hartree-Fock equation

The remaining terms are now the derivative contribution from the orbital response U_{ij}^{a} running from virtual i to occupied j molecular orbitals, therefore solving the CPHF equation is necessary. The CPHF equation is also obtained from the variational condition in Eq. (8). By taking the derivative of the Fock matrix F_{ij} with respect to the nuclear coordinate a, one obtains the following equation,

$$\frac{\partial F_{ij}}{\partial a} = F_{ij}^{a} - (\epsilon_j - \epsilon_i)U_{ij}^{a} - S_{ij}^{a}\epsilon_j - \frac{1}{2}\sum_{k,l}^{\mathrm{occ}} S_{kl}^{a} A_{ij,kl} + \sum_{k}^{\mathrm{vir}}\sum_{l}^{\mathrm{occ}} U_{kl}^{a} A_{ij,kl}, \tag{17}$$

and using the relationship $F_{ij} = \delta_{ij}$, one obtains the set of linear equations (CPHF equations)

$$\sum_{k}^{\mathrm{vir}}\sum_{l}^{\mathrm{occ}} A'_{ij,kl} U_{kl}^{a} = B_{0,ij}^{a}, \tag{18}$$

where the $\mathbf{A}'$ and $\mathbf{B}_0^a$ are

$$A'_{ij,kl} = \delta_{ik,jl}(\epsilon_j - \epsilon_i) - A_{ij,kl}, \tag{19}$$

$$B^a_{0,ij} = F^a_{ij} - S^a_{ij}\epsilon_j - \frac{1}{2}\sum_{k,l}^{\mathrm{occ}} S^a_{kl} A_{ij,kl}. \tag{20}$$

therefore, by solving the set of Eq. (18), one obtains the response term U^a_{ij}.

4. Summary for HF energy derivative technique

Summarising, the HF energy derivative technique is practically used under the variational condition. The overall flow of calculation is depicted in FIG. (1).

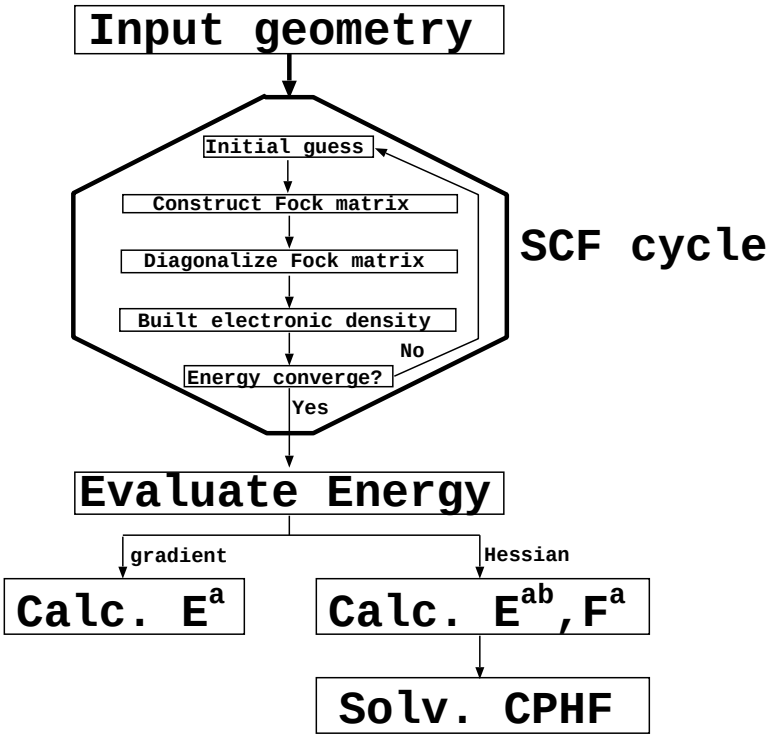


FIG. 1. Schematic illustration for how the HF calculation works

First, the electronic structure is determined self consistently by the diagonalization of the Fock matrix, and the total energy is evaluated according to the Eq. (1). If the gradient or Hessian is requested, the calculation continues to evaluate the respective properties according

to Eq. (11) or (16). The overall scheme in FIG. (1) is general for any level of theory or calculation properties for the excited state energy, gradient or Hessian, thus same subroutine is used for most of the quantum chemical calculations.

The second order derivative and vibrational analysis are pioneered by Purlay,[66], and the efficiency has been improved by introducing the analytic techniques for restricted Hartree-Fock (RHF), restricted open-shell (ROHF), and unrestricted (UHF) wave functions.[65] Following the initial implementation of analytic second order derivative based on HF level, wide range of electronic correlation method has been developed such as Multi-configurational self-consistent field,[67] coupled-cluster single-double for both closed and open shell systems,[68–70] Møller-Plesset perturbation theory,[71] configuration interaction,[72] and Density Functional Theory (DFT).[73–75] Therefore analytic second order derivative can routinely be performed for most level of quantum chemical theory.

For the energy and gradient, the most expensive step for computational timing is the SCF iteration due to the diagonalization of Fock matrix, and the scaling is cubically increasing with the system size. On the other hand in the Hessian calculation, the time consuming step is the solving CPHF equation, because of the set of linear equation is running over virtual to occupied molecular orbital for respective atomic coordinates. Thus, the computational timing for Hessian increases with the fourth power of the system size. Consequently, the *ab initio* calculations become practically difficult for the system containing more than hundreds of atoms. Therefore, reaction analysis, excited state calculation, or vibrational analysis for large biological system is virtually impossible.

C. Introduction of fragment molecular orbital method

In the previous section, the theory how the *ab initio* calculation works in the QM program is discussed, and there is the limitation toward applying the large molecular systems. The problem is the steep increase in the computational effort with the system size, and practical application for large system become difficult. This drives the development of fragmentation approach, such as FMO method.

In the FMO, the total system is separated into the small subset of fragment, and the

respective fragment is iteratively calculated to obtain the self consistent ESP.[39,76] The independent fragment calculation is called monomer calculation, and the self consistent cycle is called SCC. After obtaining the mutually converged ESP, the pair of fragment (dimer) calculations are performed to gain the total property in the system. This level of FMO method is called FMO2. If the three body contribution (trimer) is included, the level of FMO method is called FMO3. The overall scheme how the FMO calculation works is depicted in FIG. (2)-(a).

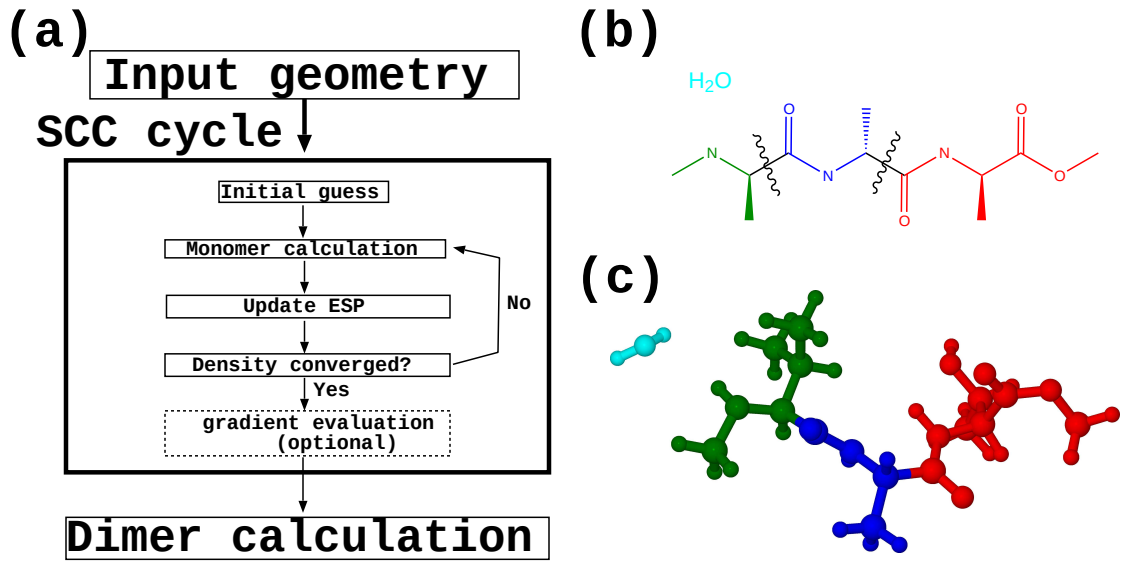


FIG. 2. (a) Schematic illustration for how the FMO calculation works (The solid line is for all the FMO calculation, while dashed line is only for the gradient run). (b) Schematic illustration of fragmentation scheme for the fragmentation across the covalent bonds. (c) The picture for the fragmentation of amino acid residues.

If the system contains the fragmentation across the covalent bonds (See FIG. (2)-(b)), the system is typically divided by the carbon atom on the fragment boundary (not the bond between the fragments). At the carbon atom on the fragment boundary, one proton is treated as the cap of the fragments, while the other five are treated as the ESP. The five electrons are projected out by the universal constant 10^5, while they are allowed some extent of freedom by the projection operator.

The success of FMO method is the introduction of self consistent ESP, and the sophisti-

cated treatment of the fragmentation boundary by the hybrid projection operator (HOP), and FMO2 reproduces the conventional full *ab initio* calculation within a few kcal/mol. The FMO scheme removes the steep increase of the computational timing with the system size, because only the small subset of *ab initio* calculation is necessary. Since the respective calculation can be independently performed with each other, therefore the parallelization of the program accelerates the computational timing ideally[77].

However, the fragmentation scheme affects the variation condition in the system, and additional extra terms are arise when one considers the nuclear gradient or second order derivative of the energy. In this section, hereafter, the FMO2 energy equation and the analytic energy gradient will be introduced, and the derivations will be discussed how the full *ab initio* energy gradient and Hessian should be modified for the FMO calculations.

1. FMO total energy and its analytic energy gradient

The total energy of FMO2 in the system is described as the following equation,

$$E = \sum_I E'_I + \sum_{I<J} \left[E'_{IJ} - E'_I - E'_J + \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \right], \tag{21}$$

where E'_X (X denote the IJ, I or J) is the internal energy without the contribution of ESP:

$$E'_X = \sum_{i \in X}^{\mathrm{occ}} 2h_{ii}^X + \sum_{i \in X}^{\mathrm{occ}} \sum_{j \in X}^{\mathrm{occ}} \{2(ii|jj) - (ij|ij)\}, \tag{22}$$

and the density difference ESP terms $\mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ})$ are

$$\mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) = \sum_{\mu,\nu \in IJ} D_{\mu\nu}^{IJ} V_{\mu\nu}^{IJ} - \sum_{\mu,\nu \in I} D_{\mu\nu}^{I} V_{\mu\nu}^{IJ} - \sum_{\mu,\nu \in J} D_{\mu\nu}^{J} V_{\mu\nu}^{IJ}, \tag{23}$$

where $D_{\mu\nu}^X$ is the density matrix, and $V_{\mu\nu}^X$ is the ESP terms:

$$V_{\mu\nu}^X = \sum_{K \neq IJ} \sum_{A \in K} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle + \sum_{K \neq IJ} \sum_{\lambda,\sigma \in K} D_{\lambda\sigma}^K (\mu\nu|\lambda\sigma), \tag{24}$$

where Z_A is nuclear charge of atom A, $\mathbf{r}$ and $\mathbf{R}_A$ are the position of electron and nuclear respectively.

Taking the derivative of the energy in Eq. (21) with respect to the nuclear coordinate a, one obtain the analytic energy gradient in the system,

$$\frac{\partial E}{\partial a} = \sum_I \frac{\partial E'_I}{\partial a} + \sum_{I<J} \left[\frac{\partial E'_{IJ}}{\partial a} - \frac{\partial E'_I}{\partial a} - \frac{\partial E'_J}{\partial a} + \frac{\partial \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ})}{\partial a} \right], \tag{25}$$

and the derivative of internal energy is

$$\frac{\partial E_X^{\mathrm{HF}}}{\partial a} = \sum_{i\in X}^{\mathrm{occ}} 2h_{ii}^{a,X} + \sum_{i\in X}^{\mathrm{occ}} \sum_{j\in X}^{\mathrm{occ}} \{2(ii|jj)^a - (ij|ij)^a\} - 2\sum_{i,j\in X}^{\mathrm{occ}} S_{ij}^{a,X} F_{ij}^{\prime,X} - \sum_{m\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} 4U_{mi}^{a,X} V_{mi}^X, \tag{26}$$

where $F_{ij}^{\prime,I}$ is the Fock matrix element in the Fragment X without the ESP contributions. All the other indices and matrices are already introduce in previous sections, and the superscript X denotes the operator and basis sets on the fragments X. If one compares between the HF gradient in Eq. (11) and internal fragment derivative in Eq. (26), the only difference is the orbital relaxation terms arising due to the ESP, and Eq. (26) contains the unknown response terns $U_{mi}^{a,X}$.

Likewise the derivatives of the ESP potential produce the contribution of ESP to the total analytic energy gradient as,

$$\begin{aligned} 2\sum_{i\in X}^{\mathrm{occ}} \frac{\partial V_{ii}^{IJ}}{\partial a} =& 2\sum_{i\in X}^{\mathrm{occ}} V_{ii}^{a,IJ} + 4\sum_{m\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} U_{mi}^{a,X} V_{mi}^{IJ} - 2\sum_{m\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} U_{mi}^{a,X} V_{mi}^{IJ} \\ & -4\sum_{k\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{i,j\in X}^{\mathrm{occ}} S_{ij}^{a,K} (kk|ij) + 8\sum_{k\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{m\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} U_{mi}^{a,K} (kk|mi), \end{aligned} \tag{27}$$

where the $V_{mi}^{a,IJ}$ is the derivative of ESP potential integral with respect to the nuclear coordinate a:

$$V_{mi}^{a,IJ} = \sum_{K\neq IJ} \sum_{A\in K} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle^a + \sum_{K\neq IJ} \sum_{\lambda,\sigma\in K} D_{\lambda\sigma}^K (\mu\nu|\lambda\sigma)^a. \tag{28}$$

Therefore the derivative of the ESP potential also contains the orbital relaxation terms $U_{mi}^{a,K}$, and solving the CPHF equation is necessary for the first order derivative of the energy with respect to the coordinate a.

2. CPHF equation in FMO

The previous subsection demonstrates that the analytic energy gradient requires the contribution of orbital response $U_{ri}^{a,I}$, thus solving the CPHF equation is necessary for the respective subset of fragments. The CPHF equation can be derived in analogous way with full *ab initio* calculation by using the following relationship:

$$\frac{\partial F_{ij}^{I}}{\partial a} = 0. \tag{29}$$

By taking the derivative of the Fock matrix in the fragments I, one obtains

$$\begin{aligned}\frac{\partial F_{ij}^{I}}{\partial a} =& F_{ij}^{a,I} - (\epsilon_j^I - \epsilon_i^I)U_{ij}^{a,I} - S_{ij}^{a,I}\epsilon_j^I - \frac{1}{2}\sum_{k,l}^{\text{occ}} S_{kl}^{a,I} A_{ij,kl} + \sum_{k}^{\text{vir}}\sum_{l}^{\text{occ}} U_{kl}^{a} A_{ij,kl} \\ &+ \frac{1}{2}\sum_{K\neq I}\sum_{k,l\in K}^{\text{occ}} S_{kl}^{a,K} A_{ij,kl}^{I,K} - \sum_{K\neq I}\sum_{k\in K}^{\text{vir}}\sum_{l\in K}^{\text{occ}} U_{kl}^{a,K} A_{ij,kl}^{I,K},\end{aligned} \tag{30}$$

where the orbital Hessian matrix in the FMO method is

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|kj), \tag{31}$$

$$A_{ij,kl}^{I,K} = -4(ij|kl). \tag{32}$$

Inserting the Eq. (30) into the Eq. (29), one obtains the CPHF equation in FMO

$$\sum_{k\in I}^{\text{vir}}\sum_{l\in I}^{\text{occ}} A_{ij,kl}^{I,I} U_{kl}^{a,I} + \sum_{K\neq I}\sum_{k\in K}^{\text{vir}}\sum_{l\in K}^{\text{occ}} A_{ij,kl}^{I,K} U_{kl}^{a,K} = B_{0,ij}^{a,I}, \tag{33}$$

where $A_{ij,kl}^{I,I}$ and $B_{0,ij}^{a,I}$ are

$$A_{ij,kl}^{I,I} = \delta_{ik}\delta_{jl}(\epsilon_j^I - \epsilon_i^I) - A_{ij,kl} \tag{34}$$

$$B_{0,ij}^{a,I} = F_{ij}^{a,I} - S_{ij}^{a,I}\epsilon_j^I - \frac{1}{2}\sum_{k,l\in I}^{\text{occ}} S_{kl}^{a,I} A_{ij,kl} + \frac{1}{2}\sum_{K\neq I}\sum_{k,l\in K}^{\text{occ}} S_{kl}^{a,K} A_{ij,kl}^{I,K}. \tag{35}$$

By taking the comparison between the full *ab initio* in Eq. (18) and FMO CPHF equation in Eq. (33), one can realize that the set of response terms $U_{kl}^{a,I}$ couples with each others, and the linear equation must be solved together over all the fragments in the system. Consequently, the dimension of linear equation increases steeply with the system size, and some sophisticated techniques are necessary for solving the CPHF equations.

Practically, the set of CPHF equation may be decomposed into the respective fragment basis equation, as

$$\sum_{k\in I}^{\mathrm{vir}}\sum_{l\in I}^{\mathrm{occ}} A^{I,I}_{ij,kl}U^{a,I}_{kl} = B^{a,I}_{0,ij} - \sum_{K\neq I}^{\mathrm{vir}}\sum_{k\in K}^{\mathrm{vir}}\sum_{l\in K}^{\mathrm{occ}} A^{I,K}_{ij,kl}U^{a,K}_{kl}, \tag{36}$$

and iteratively solved the equation for only the fragment I until mutually converged $U^{a,I}_{ri}$ is obtained (See more detail for Chapter II).[58]

3. Summary of FMO gradient

In this section, the FMO energy and gradient have been introduced. Summarising, there are two important differences between the FMO energy derivatives and full *ab initio* calculations. First, the FMO energy dose not satisfy the variation condition, and first order orbital response terms $U^{a,I}_{mi}$ appeared in gradient (thus, second order response term ($U^{ab,I}_{mi}$) is appeared in Hessian). Secondary, therefore, solving the CPHF equation is necessary, and the problem is that the CPHF equations couple with each other from virtual to occupied molecular orbital over all the fragments in the system. Thus, the extraordinary large dimension of linear equation should be solved. Practically, the CPHF equation is solved by decomposing the set of CPHF equations into the respective fragments.

The large dimension of linear equation can be solved iteratively by separating the set of equations into the independent fragments as shown in Eq. (36). The decomposition of a set of linear equation is called the self consistent Z-vector method, and it has been developed for only the closed shell systems (FMO-RHF, MP2) by Nagata et. al.[58,59].

D. Aim and outline of this thesis

Summarising in this Chapter I, the derivative technique for both *ab initio* and FMO energy with respect to the nuclear coordinate are briefly introduced. Despite the simple structure of *ab initio* analytic energy gradient and Hessian, the computational scaling steeply increases with the system size, and it drives us to develop the linear scaling algorithm and fragmentation approaches. Excellent performance of FMO approaches were observed

in many previous studies,[48,48,49,49,50,50–52] because of the high efficiency of parallelization. However, the fragmentation of the system affects the variation condition that is usually satisfied in the full *ab initio* calculations. Thus one requires higher order orbital response terms for both gradient and Hessian in FMO than the gradient and Hessian in the full HF.

The fully analytic energy gradient technique was developed recently, while the method was limited within energy gradient for closed-shell systems. Therefore, the analysis of metal enzyme typically containing the unpaired electron was not possible, and any vibrational analysis was not available.

The aim in this thesis is the development of the FMO method to calculate any biological systems in order to predict the UV-Vis spectrum, IR spectrum, Raman scattering, and the reaction free energy. By developing these new method, many biological application territories of quantum chemistry are becoming open, and that enables us understanding the molecular level in the complex biological phenomena, which was virtually impossible within the current conventional computational techniques.

For this purpose, six new FMO methods were developed in this thesis;
(1) The Unrestricted Hartree-Fock (UHF) calculation based on the FMO (FMO-UHF) for the open-shell systems (for radical molecules) such as metal enzyme.
(2) The FMO-UHF is accelerated by the point charge approximation.
(3) DFT is interfaced for both ground and excited state of open-shell calculation.
(4) The analytic energy second order derivative of energy (Hessian) is interfaced with FMO method.
(5) The Hessian is accelerated to calculate more than hundreds atoms in the system, and interfaced with open-shell system.
(6) The FMO calculation is combined with Raman scattering.

The respective Chapter corresponds to the new development of method noted in the above. Therefore,
Chapter I is this introduction and the basic technique used in the entire parts in this thesis.
Chapter II is the description of analytic energy gradient technique especially focusing on the open-shell method, and it explains the detailed technique that is used for FMO gradient.
Chapter III is the acceleration technique to improve the FMO performance to achieve linear

scaling by using point charge approximation for ESP when the distances between fragments are far separated with each other.

Chapter IV is the development for DFT and Time Dependent Density Functional Theory (TD-DFT) for calculation of excited state energy of open-shell systems, and the method is applied to metal protein containing Cu.

Chapter V is the development of analytic second order derivative for closed-shell system, and the performance is evaluated by simple amino acid polymers, and some organic reaction analyses.

Chapter VI is the extension of the second order derivative to open-shell system, and the computational timing is accelerated to calculate more than 700 hundreds atoms in the system.

Chapter VII is the extension of FMO to calculate Raman, and applications to the moderately large protein system (Crambin protein), and the performance is evaluated.

Chapter VIII is the conclusion of this thesis, the achievement and future objective are discussed.

REFERENCES

[1]N. W. Schmidt, K. K. Baldridge, J. A. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[2]M. S. Gordon and M. W. Schmidt, Advances in electronic structure theory: GAMESS a decade later, in *Theory and Applications of Computational Chemistry, the first forty years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria, pages 1167–1189, Elsevier, Amsterdam, 2005.

[3]M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, Comput. Phys. Commun. **181**, 1477 (2010).

[4]Y. Shao, L. Fusti-Molnar, Y. Jung, J. rg Kussmann, C. Ochsenfeld, S. T. Brown, A. T. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. ONeill, R. A. D. Jr., R. C. Lochan,

T. Wang, G. J. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Daschel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khaliullin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. W. III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. S. III, J. Kong, A. I. Krylov, P. M. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys., 8, 3172 (2006). **8**, 3172 (2006).

[5]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. A. andJ. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, , and D. J. Fox, Gaussian, Inc., Wallingford CT (2010), Gaussian 09, Revision C.01.

[6]A. Warshel and M. Karplus, J. Am. Chem. Soc. **94**, 5612 (1972).

[7]J. A. McCammon, B. R. Gelin, and M. Karplus, Nature **267**, 585 (1977).

[8]F. Maseras and K. Morokuma, Journal of Computational Chemistry **16**, 1170 (1995).

[9]L. Hai and D. Truhlar, Theoretical Chemistry Accounts: Theory, Computation, and Modeling **117**, 185 (2007).

[10]H. M. Senn and W. Thiel, Angewandte Chemie International Edition **48**, 1198 (2009).

[11]D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen,

Journal of Computational Chemistry **26**, 1701 (2005).

[12]A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, The Journal of Physical Chemistry B **102**, 3586 (1998).

[13]D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, Journal of Computational Chemistry **26**, 1668 (2005).

[14]P. Ren, C. Wu, and J. W. Ponder, Journal of Chemical Theory and Computation **7**, 3143 (2011).

[15]J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, The Journal of Physical Chemistry B **114**, 2549 (2010), PMID: 20136072.

[16]N. M. Thellamurege, D. Si, F. Cui, H. Zhu, R. Lai, and H. Li, Journal of Computational Chemistry **34**, 2816 (2013).

[17]S. Sumner, P. Sderhjelm, and U. Ryde, Journal of Chemical Theory and Computation **9**, 4205 (2013).

[18]L. Hu, P. Sderhjelm, and U. Ryde, Journal of Chemical Theory and Computation **9**, 640 (2013).

[19]A. Monari, J.-L. Rivail, and X. Assfeld, Accounts of Chemical Research **46**, 596 (2013), PMID: 23249409.

[20]S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[21]G. E. Scuseria, J. Phys. Chem. A **103**, 4782 (1999).

[22]X. Li, J. M. Milliam, G. E. Scuseria, M. J. Frisch, and H. B. Schlegel, J. Chem. Phys. **119**, 7651 (2003).

[23]P. G. Mezey and J. Leszczynski, *Linear-Scaling Techniques in Computational Chemistry and Physics.*, Springer, New York, 2011.

[24]J. R. Reimers, *Computational Methods for Large Systems: Electronic Structure Approaches*

for Biotechnology and Nanotechnology, Wiley, New York, 2011.

[25]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[26]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[27]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[28]M. Kobayashi, T. Yoshikawa, and H. Nakai, Chem. Phys. Lett. **500**, 172 (2010).

[29]X. He and K. M. Merz, J. Chem. Theory Comput. **6**, 405 (2010).

[30]M. Kobayashi and H. Nakai, J. Chem. Phys. **138**, 044102 (2013).

[31]H. Yu, H. R. Leverentz, P. Bai, J. I. Siepmann, and D. G. Truhlar, J. Phys. Chem. Lett. **5**, 660 (2014).

[32]Y. Tong, Y. Mei, J. Z. H. Zhang, L. L. Duan, and Q. G. Zhang, J. Theor. Comp. Chem. **8**, 1265 (2009).

[33]P. Söderhjelm, J. Kongsted, and U. Ryde, J. Chem. Theory Comput. **6**, 1726 (2010).

[34]J. Gao and Y. Wang, J. Chem. Phys. **136**, 071101 (2012).

[35]M. A. Collins, Phys. Chem. Chem. Phys. **14**, 7744 (2012).

[36]A. Frank, H. M. Möller, and T. E. Exner, J. Chem. Theory Comput. **8**, 1480 (2012).

[37]M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Ann. Rev. Phys. Chem. **64**, 553 (2013), PMID: 23561011.

[38]K. Kiewisch, C. R. Jacob, and J. Visscher, J. Chem. Theory Comput. **9**, 2425 (2013).

[39]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[40]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A. **111**, 6904 (2007).

[41]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[42]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **122**, 0541081 (2005).

[43]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).

[44]Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **406**, 283 (2005).

[45]S.-I. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, Chem. Phys. Lett. **382**, 611 (2003).

[46]M. Chiba, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **444**, 346 (2007).

[47]H. Nakata, D. G. Fedorov, S. Yokojima, K. Kitaura, M. Sakurai, and S. Nakamura, J.

Chem. Phys **140** (2014).

[48]M. P. Mazanetz, O. Ichihara, R. J. Law, and M. Whittaker, J. Cheminf. **3**, 2 (2011).

[49]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[50]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Phys. Chem. B **114**, 15700 (2010).

[51]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B **111**, 9621 (2007).

[52]H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii, and K. Kitaura, J. Phys. Chem. A **112**, 10887 (2008).

[53]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).

[54]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B **116**, 503 (2012).

[55]P. V. Avramov, D. G. Fedorov, P. B. Sorokin, S. Sakai, S. Entani, M. Ohtomo, and H. N. Y. Matsumoto, J. Phys. Chem. Lett. **3**, 2003 (2012).

[56]Y. Okiyama, T. Tsukamoto, C. Watanabe, K. Fukuzawa, S. Tanaka, and Y. Mochizuki, Chem. Phys. Lett. **566**, 25 (2013).

[57]L. Roskop, D. G. Fedorov, and M. S. Gordon, Mol. Phys. **111**, 1622 (2013).

[58]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[59]T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, J. Chem. Phys. **135**, (2011).

[60]S. R. Pruitt, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Theory Comp. **6**, 1 (2010).

[61]S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, J. Phys. Chem. A **116**, 4965 (2012).

[62]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101 (2008).

[63]S. Sakai and S. Morita, J. Phys. Chem. A **109**, 8424 (2005).

[64]A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, DOVER PUBLICATION, INC., New York, 1982.

[65]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, and J. Goddard, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press, New York, 1994.

[66]P. Pulay, Mol. Phys. **17**, 197 (1969).

[67]T. U. Helgaker, J. Almlf, H. J. A. Jensen, and P. Jorgensen, J. Chem. Phys. **84**, 6266 (1986).

[68]H. Koch, H. J. A. Jensen, P. Jo/rgensen, T. Helgaker, G. E. Scuseria, and H. F. Schaefer, J. Chem. Phys. **92**, 4924 (1990).

[69]J. Gauss and J. F. Stanton, Chem. Phys. Lett. **276**, 70 (1997).

[70]P. G. Szalay, J. Gauss, and J. F. Stanton, Theor. Chem. Acc. **100**, 5 (1998).

[71]J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. **16**, 225 (1979).

[72]R. Krishnan, H. B. Schlegel, and J. A. Pople, J. Chem. Phys. **72**, 4654 (1980).

[73]A. Komornicki and G. Fitzgerald, J. Chem. Phys. **98**, 1398 (1993).

[74]P. Deglmann, F. Furche, and R. Ahlrichs, Chem. Phys. Lett. **362**, 511 (2002).

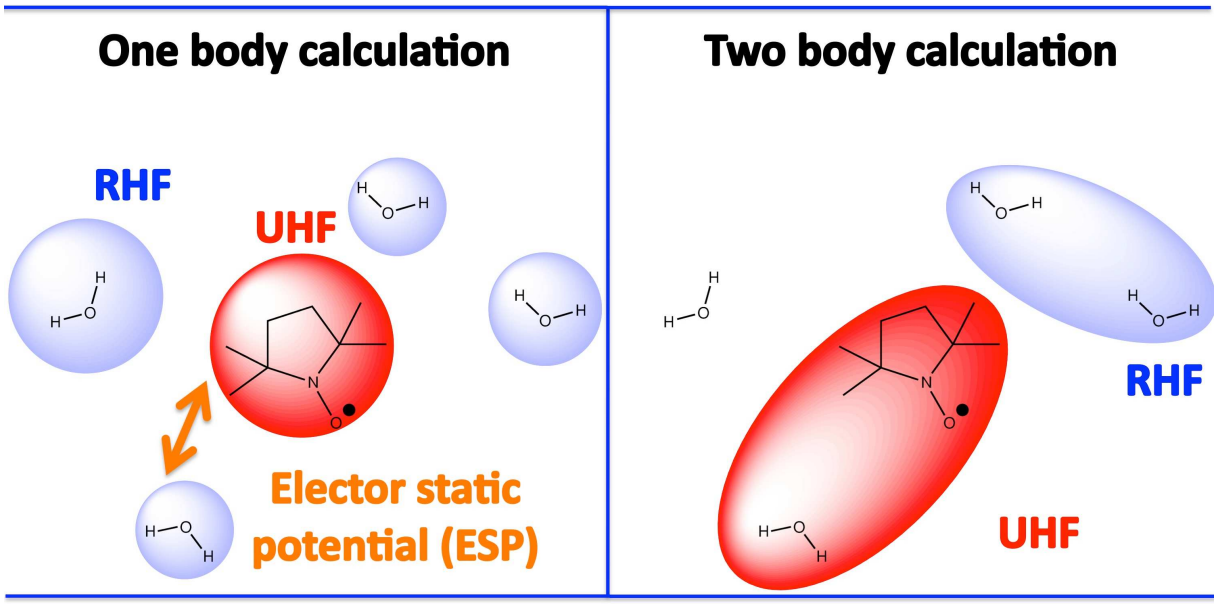
[75]B. G. Johnson and M. J. Frisch, Chem. Phys. Lett. **216**, 133 (1993).

[76]D. G. Fedorov and K. Kitaura, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC press, Boca Raton, FL, 2009.

[77]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004).

Chapter II

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



H. Nakata, D. G. Fedorov, T. Nagata *et. al.*

J. Chem. Phys. **2012,** *137,* 044110.

DOI : 10.1063/1.4737860
View online: http://dx.doi.org/10.1063/1.4737860

II. UNRESTRICTED HARTREE-FOCK BASED ON THE FRAGMENT MOLECULAR ORBITAL METHOD: ENERGY AND ITS ANALYTIC GRADIENT

A. Introduction

Biomolecules, polymers, and nano systems with a large number of atoms are challenging targets of computational chemistry. For example, enzymes catalyze the synthesis of common cellular products or take part in the generation of energy. By understanding these reactions at the molecular level, complicated behavior and fundamental questions about cellular activity can be clarified.

The difficulty in treating large molecular systems comes from the steep scaling of their computational cost, increasing approximately cubically with the system size even for the most basic Hartree-Fock (HF) method. Very significant effort has been invested by theoreticians in improving the efficiency and reducing the scaling. Two basic directions can be followed: linear scaling *ab initio* methods[1–4] or methods based on fragmenting the system[5]. The latter approach has been taken in a number of different ways[6–16].

The fragment molecular orbital (FMO) method[17] is based on dividing a molecular system into fragments and performing *ab initio* calculations of fragments and their dimers in the electrostatic field of the whole system[18–20]. A number of theoretical methods and models have been interfaced with FMO (see Chapter 2 of Ref.[3]). The method has been applied to nanowires[21], protein-ligand complexes[22–24], ionic liquids[25], molecular crystals[26], DNA[27] and other systems.

In comparison to the major efforts invested in developing ground state closed shell methods, considerably less has been accomplished for fragment-based approaches to calculating excited states[28–31] or open-shell systems[14,15,32–34]. Transition metals involved in biological systems often possess an open-shell character, which in principle should be treated with multireference methods[35], but a simpler approach often taken is unrestricted HF (UHF), with which one hopes to describe fractionally occupied d-orbitals in the transition metal complexes[36,37]. UHF can also be useful in describing radical chemistry or electron transfer

processes.

The original formulation of the energy gradient for FMO[38] has not been exactly analytic neglecting several contributions, added later: Mulliken charge derivatives[39], hybrid orbital projection operator gradient[40] and the response terms associated with the coupling of the embedding electrostatic potential and the electronic state of dimers. The fully analytic gradient including the response terms has been developed only recently for restricted HF (RHF)[41] and Møller-Plesset perturbation theory[42,43]; it has been demonstrated in molecular dynamics simulations (MD)[44] that the gradient is fully analytic. It is very important to have the analytic gradient in order to perform geometry optimizations[45] or MD simulations[46–48].

In this work, we develop the energy and its analytic gradient for FMO-based UHF, following the initial FMO-UHF energy implementation by Hino et. al.[49]. We evaluate the accuracy of FMO-UHF in comparison to *ab initio* UHF by studying solvated organic radical molecules, polypeptides and a transition metal complex. As an important motivation of this work, the photosynthetic reaction center PSII contains several manganese atoms and thus the ability to use spin unrestricted wave function in large molecular systems is very desirable.

B. Mathematical formulation

1. Energy expression for FMO-UHF

Individual fragment (monomer) calculations are performed in the presence of the electrostatic potentials (ESP) due to the remaining fragments in the first step of FMO, until their mutual self-consistency. In the presence of the converged electrostatic potentials, dimer calculations are performed once for each dimer.

We formulate the FMO-UHF method by assigning some of the fragments as UHF and others as RHF. If one or more fragments involved in a dimer is UHF, then the dimer is also computed with UHF and otherwise with RHF.

We note that although we here generalize the method for any number of UHF fragments, our actual program implementation is limited to only one UHF fragment. Also, in this work,

we take the approach of the local spin polarization. In principle, it is conceivable to have a delocalized spin-polarization, that is, even if only one fragment is designed as UHF, it would exert a spin-polarized ESP on RHF fragments making the Fock matrices for the α and β spins different, so that all fragments have to be treated with UHF. Instead, we enforce that ESP is spin-restricted, in other words, the embedding potential is based on the sum of α and β densities, so that RHF fragments are immersed in the spin-restricted ESP and thus can be described with RHF. The UHF fragment is also immersed in the spin-restricted ESP of all RHF fragments, and its spin-unrestricted nature is local because of its UHF wave function, not ESP.

There is a question of how to choose the spin multiplicity for those dimers which are composed of two UHF monomers. The simplest case is when at least one UHF monomer is a singlet. In this case the multiplicity of the dimer is uniquely determined by the other UHF monomer. If both UHF monomers are not singlets, the dimer can have several multiplicities according to the angular momentum addition rules. We leave it to a future study to determine the strategy for handling the ambiguity of the dimer multiplicities in the general case. The safest choice in this case is the highest spin multiplicity, whereas other spin states should be treated with caution as they often have an inherent multireference character.

The total energy in the two-body FMO2-UHF method is given by

$$
\begin{aligned}
E = & \sum_{I \in \mathrm{RHF}}^{N_{\mathrm{RHF}}} E'_I + \sum_{K \in \mathrm{UHF}}^{N_{\mathrm{UHF}}} E'_K + \sum_{\substack{I>J \\ I,J \in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} (E'_{IJ} - E'_I - E'_J) \\
& + \sum_{\substack{K>L \\ K,L \in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} (E'_{KL} - E'_K - E'_L) \\
& + \sum_{K \in \mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{I \in \mathrm{RHF}}^{N_{\mathrm{RHF}}} (E'_{IK} - E'_I - E'_K) \\
& + \sum_{\substack{I>J \\ I,J \in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \\
& + \sum_{\substack{K>L \\ K,L \in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \left[\mathrm{Tr}(\Delta \mathbf{D}^{\alpha,KL} \mathbf{V}^{KL}) + \mathrm{Tr}(\Delta \mathbf{D}^{\beta,KL} \mathbf{V}^{KL}) \right] \\
& + \sum_{K \in \mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{I \in \mathrm{RHF}}^{N_{\mathrm{RHF}}} \big[\mathrm{Tr}(\Delta \mathbf{D}^{\alpha,IK} \mathbf{V}^{IK}) \\
& + \mathrm{Tr}(\Delta \mathbf{D}^{\beta,IK} \mathbf{V}^{IK}) \big], \qquad (1)
\end{aligned}
$$

where N_{RHF} and N_{UHF} are the numbers of fragments for RHF and UHF, respectively. Thus, the total number of fragments N is $N = N_{\mathrm{RHF}} + N_{\mathrm{UHF}}$. In our study, we use I, J to denote RHF fragments, while K, L run over UHF fragments. E'_I, E'_J, E'_{IJ}, E'_K, E'_L, E'_{IK}, and E'_{KL} are internal energies (i.e., RHF or UHF energies with the ESP contribution subtracted).

In this work we also develop FMO3-UHF, based on the three-body expansion of FMO[50,51]. The corresponding energy expression can be easily written based on FMO3-RHF[51] and FMO2-UHF (Eq. (1)). Here, we omit its explicit expression for simplicity.

The dimer density matrix difference $\Delta \mathbf{D}^{XY}$ for RHF is

$$
\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^I \oplus \mathbf{D}^J) \qquad (2)
$$

and for UHF

$$\Delta\mathbf{D}^{\alpha,KL} = \mathbf{D}^{\alpha,KL} - (\mathbf{D}^{\alpha,K} \oplus \mathbf{D}^{\alpha,L}), \tag{3}$$

$$\Delta\mathbf{D}^{\beta,KL} = \mathbf{D}^{\beta,KL} - (\mathbf{D}^{\beta,K} \oplus \mathbf{D}^{\beta,L}), \tag{4}$$

$$\Delta\mathbf{D}^{\alpha,K,I} = \mathbf{D}^{\alpha,KI} - \left(\mathbf{D}^{\alpha,K} \oplus \frac{1}{2}\mathbf{D}^{I}\right), \tag{5}$$

$$\Delta\mathbf{D}^{\beta,K,I} = \mathbf{D}^{\beta,KI} - \left(\mathbf{D}^{\beta,K} \oplus \frac{1}{2}\mathbf{D}^{I}\right), \tag{6}$$

where the density matrix $\mathbf{D}^{XY}$ is divided into two subspaces (α and β spins) for the UHF calculations and those are denoted by $\mathbf{D}^{\alpha,XY}$ and $\mathbf{D}^{\beta,XY}$, respectively.

The matrix $\mathbf{V}^{XY}$ is the ESP on the dimer XY.

$$V_{\mu\nu}^{XY} = \sum_{\substack{I\neq X,Y \\ I\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} (u_{\mu\nu}^{I} + v_{\mu\nu}^{I}) + \sum_{\substack{K\neq X,Y \\ K\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} (u_{\mu\nu}^{K} + v_{\mu\nu}^{K}), \tag{7}$$

where the one-electron u and two-electron v integrals are given by

$$u_{\mu\nu}^{W} = \sum_{A\in W} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle, \tag{8}$$

$$v_{\mu\nu}^{I} = \sum_{\lambda,\sigma\in I} D_{\lambda\sigma}^{I} (\mu\nu|\lambda\sigma), \tag{9}$$

$$v_{\mu\nu}^{K} = \sum_{\lambda,\sigma\in K} D_{\lambda\sigma}^{\alpha+\beta,K} (\mu\nu|\lambda\sigma), \tag{10}$$

where $W = I$ for RHF and $W = K$ for UHF. The indices μ, ν, λ, and σ run over atomic orbitals, and $(\mu\nu|\lambda\sigma)$ is a two-electron integral. Throughout this study, we use Greek indices for atomic orbitals basis functions. The electron and nuclear positions are $\mathbf{r}$ and $\mathbf{R}_A$, respectively. A numbers atoms, and the atomic charge is denoted by Z_A. $D_{\lambda\sigma}^{\alpha+\beta,K}$ is the sum of the α and β density matrix, $D_{\lambda\sigma}^{\alpha,K} + D_{\lambda\sigma}^{\beta,K}$.

The internal energies of RHF fragments in Eq. (1) are

$$\begin{aligned} E_I' = & \sum_{\mu,\nu\in I} D_{\mu\nu}^{I} h_{\mu\nu}^{I} \\ & + \frac{1}{2} \sum_{\mu,\nu,\lambda,\sigma\in I} \left[D_{\mu\nu}^{I} D_{\lambda\sigma}^{I} - \frac{1}{2} D_{\mu\lambda}^{I} D_{\nu\sigma}^{I} \right] (\mu\nu|\lambda\sigma) \\ & + E_I^{\mathrm{NR}}, \end{aligned} \tag{11}$$

and for UHF

$$
\begin{aligned}
E'_K = & \sum_{\mu,\nu\in K} D^{\alpha+\beta,K}_{\mu\nu} h^K_{\mu\nu} \\
& + \sum_{\mu,\nu,\lambda,\sigma\in K} \Bigg[\frac{1}{2}(D^{\alpha,K}_{\mu\nu} D^{\alpha,K}_{\lambda\sigma} + D^{\beta,K}_{\mu\nu} D^{\beta,K}_{\lambda\sigma}) + D^{\alpha,K}_{\mu\nu} D^{\beta,K}_{\lambda\sigma} \\
& - \frac{1}{2}\left(D^{\alpha,K}_{\mu\lambda} D^{\alpha,K}_{\nu\sigma} + D^{\beta,K}_{\mu\lambda} D^{\beta,K}_{\nu\sigma} \right) \Bigg] (\mu\nu|\lambda\sigma) \\
& + E^{\mathrm{NR}}_K . \qquad (12)
\end{aligned}
$$

Here $h^K_{\mu\nu}$ is sum of the one electron core Hamiltonian including the hybrid orbital projection[41] $P^K_{\mu\nu}$. E^{NR}_I is the nuclear repulsion energy of I.

2. Analytic energy gradient in FMO-UHF

Here we derive the analytic first derivatives with respect to a nuclear displacement a.

$$
\begin{aligned}
\frac{\partial E}{\partial a} = & \sum_{I\in\mathrm{RHF}}^{N_{\mathrm{RHF}}} \frac{\partial E'_I}{\partial a} + \sum_{K\in\mathrm{UHF}}^{N_{\mathrm{UHF}}} \frac{\partial E'_K}{\partial a} \\
& + \sum_{\substack{I>J \\ I,J\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \frac{\partial (E'_{IJ} - E'_I - E'_J)}{\partial a} \\
& + \sum_{\substack{K>L \\ K,L\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \frac{\partial (E'_{KL} - E'_K - E'_L)}{\partial a} \\
& + \sum_{\substack{I \\ I\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\substack{K \\ K\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \frac{\partial (E'_{IK} - E'_I - E'_K)}{\partial a} \\
& + \sum_{\substack{I>J \\ I,J\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \frac{\partial \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a} \\
& + \sum_{\substack{K>L \\ K,L\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \frac{\partial \left(\mathrm{Tr}(\Delta\mathbf{D}^{\alpha,KL}\mathbf{V}^{KL}) + \mathrm{Tr}(\Delta\mathbf{D}^{\beta,KL}\mathbf{V}^{KL}) \right)}{\partial a} \\
& + \sum_{\substack{I \\ I\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\substack{K \\ K\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \frac{\partial \left(\mathrm{Tr}(\Delta\mathbf{D}^{\alpha,IK}\mathbf{V}^{IK}) + \mathrm{Tr}(\Delta\mathbf{D}^{\beta,IK}\mathbf{V}^{IK}) \right)}{\partial a} . \qquad (13)
\end{aligned}
$$

The differentiations of E'_X with respect to a nuclear coordinate a lead for RHF to

$$\begin{aligned}\frac{\partial E'_I}{\partial a} = &-2\sum_{i}^{\text{occ}}\sum_{j}^{\text{occ}} S_{ij}^{a,I} F_{ij}^{\prime I} + \frac{\partial E_I^{\text{NR}}}{\partial a} \\ &- 4\sum_{r}^{\text{vir}}\sum_{i}^{\text{occ}} U_{ri}^{a,I} V_{ri}^{I} \\ &+ \sum_{\mu,\nu} D_{\mu\nu}^{I} \frac{\partial h_{\mu\nu}^{I}}{\partial a} \\ &+ \frac{1}{2}\sum_{\mu,\nu,\lambda,\sigma\in I}\left[D_{\mu\nu}^{I} D_{\lambda\sigma}^{I} - \frac{1}{2} D_{\mu\lambda}^{I} D_{\nu\sigma}^{I}\right]\frac{\partial(\mu\nu|\lambda\sigma)}{\partial a}\end{aligned} \tag{14}$$

and for UHF,

$$\begin{aligned}\frac{\partial E'_K}{\partial a} = &-\sum_{\sigma}\sum_{i^\sigma}^{\text{occ}}\sum_{j^\sigma}^{\text{occ}} S_{i^\sigma j^\sigma}^{a,K} F_{i^\sigma,j^\sigma}^{\prime K} \\ &- \sum_{\sigma}\sum_{r^\sigma}^{\text{vir}}\sum_{i^\sigma}^{\text{occ}} 2U_{r^\sigma i^\sigma}^{a,K} V_{r^\sigma i^\sigma}^{K} \\ &+ \sum_{\mu,\nu} D_{\mu\nu}^{\alpha+\beta,K}\frac{\partial h_{\mu\nu}^{K}}{\partial a} \\ &+ \sum_{\mu,\nu,\lambda,\sigma\in K}\left[\frac{1}{2}(D_{\mu\nu}^{\alpha,K} D_{\lambda\sigma}^{\alpha,K} + D_{\mu\nu}^{\beta,K} D_{\lambda\sigma}^{\beta,K}) + D_{\mu\nu}^{\alpha,K} D_{\lambda\sigma}^{\beta,K}\right. \\ &\left. - \frac{1}{2}\left(D_{\mu\lambda}^{\alpha,K} D_{\nu\sigma}^{\alpha,K} + D_{\mu\lambda}^{\beta,K} D_{\nu\sigma}^{\beta,K}\right)\right]\frac{\partial(\mu\nu|\lambda\sigma)}{\partial a} \\ &+ \frac{\partial E_K^{\text{NR}}}{\partial a},\end{aligned} \tag{15}$$

where i, j, k, l, and m are molecular orbitals (linear combination of atomic orbitals). Note that we use i^α, and j^β to distinguish the α and β subspaces in UHF. The spin label σ can be α or β. $F_{ij}^{\prime X}$ are internal Fock matrix elements[41] of fragment X.

$$F_{ij}^{\prime X} = h_{ij}^{X} + \sum_{k\in X}[2(ij|kk) - (ik|jk)] \tag{16}$$

$S_{ij}^{a,X}$ is the overlap derivative matrix element

$$S_{ij}^{a,X} = \sum_{\mu,\nu\in X} C_{\mu i}^{X*}\frac{\partial S_{\mu\nu}^{X}}{\partial a} C_{\nu j}^{X} \tag{17}$$

where S^X is the AO overlap integral. $U_{ri}^{a,I}$ is the orbital response[52] for nuclear displacement a. The response terms $U_{ri}^{a,I}$ associated with ESP (V_{ri}^I) are gathered and calculated in the combined coupled-perturbed Hartree-Fock (CPHF) equations, after all the dimer calculations are finished. This is discussed in the next subsection. It is straightforward to obtain the derivatives of $\mathrm{Tr}(\Delta\mathbf{D}^{XY}\mathbf{V}^{XY})$ with respect to a nuclear coordinate a (see Appendix A and supporting information.[53]).

Inserting each term into the differentiation of E in Eq. (13), one obtains the analytic first derivatives with respect to a nuclear displacement a. The gradient can be written as

$$\frac{\partial E}{\partial a} = \frac{\partial E'}{\partial a} + R^a \tag{18}$$

where E' is the collection of the derivative in Eq. (13) without response terms ($U_{ri}^{a,I}$). R^a is the collection of all the response terms in Eq. (13):

$$\begin{aligned}
R^a =& 4 \sum_{\substack{I>J \\ I,J\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\substack{X\neq I,J \\ X\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in I,J} \sum_{r\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} \Delta D_{\mu\nu}^{IJ} U_{ri}^{a,X} (\mu\nu|ri) \\
&+2 \sum_{\substack{I>J \\ I,J\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\substack{X\neq I,J \\ X\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in I,J} \sum_{\sigma} \sum_{r^\sigma\in X}^{\mathrm{vir}} \sum_{i^\sigma\in X}^{\mathrm{occ}} \Delta D_{\mu\nu}^{IJ} U_{r^\sigma i^\sigma}^{a,X} (\mu\nu|r^\sigma i^\sigma) \\
&+4 \sum_{\substack{K>L \\ K,L\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\substack{X\neq K,L \\ X\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in K,L} \sum_{r\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} \Delta D_{\mu\nu}^{\alpha+\beta,KL} U_{ri}^{a,X} (\mu\nu|ri) \\
&+2 \sum_{\substack{K>L \\ K,L\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\substack{X\neq K,L \\ X\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in K,L} \sum_{\sigma} \sum_{r^\sigma\in X}^{\mathrm{vir}} \sum_{i^\sigma\in X}^{\mathrm{occ}} \\
&\Delta D_{\mu\nu}^{\alpha+\beta,KL} U_{r^\sigma i^\sigma}^{a,X} (\mu\nu|r^\sigma i^\sigma) \\
&+4 \sum_{I\in \mathrm{RHF}}^{N_{\mathrm{RHF}}} \sum_{K\in \mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{\substack{X\neq K,I \\ X\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in K,I} \sum_{r\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} \\
&\Delta D_{\mu\nu}^{\alpha+\beta,KI} U_{ri}^{a,X} (\mu\nu|ri) \\
&+2 \sum_{I\in \mathrm{RHF}}^{N_{\mathrm{RHF}}} \sum_{K\in \mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{\substack{X\neq K,I \\ X\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in K,I} \sum_{\sigma} \sum_{r^\sigma\in X}^{\mathrm{vir}} \sum_{i^\sigma\in X}^{\mathrm{occ}} \\
&\Delta D_{\mu\nu}^{\alpha+\beta,KI} U_{r^\sigma i^\sigma}^{a,X} (\mu\nu|r^\sigma i^\sigma).
\end{aligned} \tag{19}$$

In Eq. (19), six different types of summation are generated by combining RHF and UHF, leading to the three types of orbital responses $U_{ri}^{a,I}$, $U_{r^{\alpha}i^{\alpha}}^{a,K}$, and $U_{r^{\beta}i^{\beta}}^{a,K}$.

We treat the response terms R^a as the scalar product ($\mathbf{L^T U}$) of two super vectors

$$\begin{aligned} R^a = & \sum_{I\in\mathrm{RHF}} \sum_{r\in I}^{\mathrm{vir}} \sum_{i\in I}^{\mathrm{occ}} L_{ri}^{I} U_{ri}^{a,I} \\ & + \sum_{K\in\mathrm{UHF}} \sum_{\sigma} \sum_{r^{\sigma}\in K}^{\mathrm{vir}} \sum_{i^{\sigma}\in K}^{\mathrm{occ}} L_{r^{\sigma}i^{\sigma}}^{K} U_{r^{\sigma}i^{\sigma}}^{a,K} \end{aligned} \tag{20}$$

where L_{ri}^{I} and L_{ri}^{K} are

$$\begin{aligned} L_{ri}^{I} = & 4 \sum_{\substack{X>Y \\ X,Y\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in X,Y} \Delta D_{\mu\nu}^{\alpha+\beta,XY} (\mu\nu|ri) \\ & + 4 \sum_{\substack{X\in\mathrm{RHF} \\ X\neq I}}^{N_{\mathrm{RHF}}} \sum_{Y\in\mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in X,Y} \Delta D_{\mu\nu}^{\alpha+\beta,XY} (\mu\nu|ri) \\ & + 4 \sum_{\substack{X>Y \\ X,Y\in\mathrm{RHF} \\ X,Y\neq I}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in X,Y} \Delta D_{\mu\nu}^{XY} (\mu\nu|ri) \end{aligned} \tag{21}$$

$$\begin{aligned} L_{r^{\sigma}i^{\sigma}}^{K} = & 2 \sum_{\substack{X>Y \\ X,Y\in\mathrm{UHF} \\ X,Y\neq K}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in X,Y} \left(\Delta D_{\mu\nu}^{\alpha+\beta,XY}\right) (\mu\nu|r^{\sigma}i^{\sigma}) \\ & + 2 \sum_{\substack{X\in\mathrm{UHF} \\ X\neq K}}^{N_{\mathrm{UHF}}} \sum_{Y\in\mathrm{RHF}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in X,Y} \left(\Delta D_{\mu\nu}^{\alpha+\beta,XY}\right) (\mu\nu|r^{\sigma}i^{\sigma}) \\ & + 2 \sum_{\substack{X>Y \\ X,Y\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in X,Y} \Delta D_{\mu\nu}^{XY} (\mu\nu|r^{\sigma}i^{\sigma}). \end{aligned} \tag{22}$$

3. CPHF equations and Z-vector method for FMO-UHF

The unknowns $U_{ri}^{a,I}$, $U_{r^{\alpha}i^{\alpha}}^{a,K}$, and $U_{r^{\beta}i^{\beta}}^{a,K}$ are the virtual-occupied orbital responses, which can be obtained by solving the monomer CPHF equations in FMO. The UHF CPHF equations[54] are generated from the derivation of the Fock matrix elements with respect to a nuclear displacement a. The Fock matrix for the RHF fragment I is

$$F_{ij}^{I} = h_{ij}^{I} + V_{ij}^{I} + \sum_{k\in I} [2(ij|kk) - (ik|jk)] \tag{23}$$

and for UHF

$$\begin{aligned} F_{i^\alpha j^\alpha}^{K} =& h_{i^\alpha j^\alpha}^{K} + V_{i^\alpha j^\alpha}^{K} + \sum_{k^\alpha \in K} [(i^\alpha j^\alpha | k^\alpha k^\alpha) \\ & -(i^\alpha k^\alpha | j^\alpha k^\alpha)] + \sum_{k^\beta \in K} (i^\alpha j^\alpha | k^\beta k^\beta). \end{aligned} \tag{24}$$

Therefore, the first derivatives with respect to a are for RHF

$$\begin{aligned} \frac{\partial F_{ij}^{I}}{\partial a} =& F_{ij}^{a,I} - (\epsilon_j^I - \epsilon_i^I) U_{ij}^{a,I} - S_{ij}^{a,I} \epsilon_j^I \\ & - \sum_{k,l\in I}^{\text{occ}} S_{kl}^{a,I} [2(ij|kl) - (ik|jl)] + \sum_{r\in I}^{\text{vir}} \sum_{k\in I}^{\text{occ}} U_{rk}^{a,I} A'_{ij,rk} \\ & - \sum_{\substack{X\neq I \\ X\in \text{UHF}}}^{N_{\text{UHF}}} \sum_{\sigma} \sum_{k^\sigma \in X}^{\text{occ}} \sum_{l^\sigma \in X}^{\text{occ}} 2 S_{k^\sigma l^\sigma}^{a,X} (ij|k^\sigma l^\sigma) \\ & + \sum_{\substack{X\neq I \\ X\in \text{UHF}}}^{N_{\text{UHF}}} \sum_{\sigma} \sum_{r^\sigma \in X}^{\text{vir}} \sum_{k^\sigma \in X}^{\text{occ}} 4 U_{r^\sigma k^\sigma}^{a,X} (ij|r^\sigma k^\sigma) \\ & - \sum_{\substack{X\neq I \\ X\in \text{RHF}}}^{N_{\text{RHF}}} \sum_{k\in X}^{\text{occ}} \sum_{l\in X}^{\text{occ}} 2 S_{kl}^{a,X} (ij|kl) \\ & + \sum_{\substack{X\neq I \\ X\in \text{RHF}}}^{N_{\text{RHF}}} \sum_{r\in X}^{\text{vir}} \sum_{k\in X}^{\text{occ}} 4 U_{rk}^{a,X} (ij|rk), \end{aligned} \tag{25}$$

and for UHF

$$\begin{aligned}
\frac{\partial F^K_{i^\alpha j^\alpha}}{\partial a} =& F^{a,K}_{i^\alpha j^\alpha} - (\epsilon^{\alpha,K}_j - \epsilon^{\alpha,K}_i) U^{a,K}_{i^\alpha j^\alpha} - S^{a,K}_{i^\alpha j^\alpha} \epsilon^{\alpha,K}_j \\
& - \frac{1}{2} \sum_{k^\alpha, l^\alpha \in K}^{\text{occ}} S^{a,K}_{k^\alpha l^\alpha} A'_{i^\alpha j^\alpha, k^\alpha l^\alpha} \\
& + \sum_{r^\alpha \in K}^{\text{vir}} \sum_{k^\alpha \in K}^{\text{occ}} U^{a,K}_{r^\alpha k^\alpha} A'_{i^\alpha j^\alpha, r^\alpha k^\alpha} \\
& - \frac{1}{2} \sum_{k^\beta, l^\beta \in K}^{\text{occ}} S^{a,K}_{k^\beta l^\beta} A'_{i^\alpha j^\alpha, k^\beta l^\beta} \\
& + \sum_{r^\beta \in K}^{\text{vir}} \sum_{k^\beta \in K}^{\text{occ}} 2 U^{a,K}_{r^\beta k^\beta} A'_{i^\alpha j^\alpha, r^\beta k^\beta} \\
& - \sum_{\substack{X \neq K \\ X \in \text{UHF}}}^{N_{\text{UHF}}} \sum_\sigma \sum_{k^\sigma \in X}^{\text{occ}} \sum_{l^\sigma \in X}^{\text{occ}} 2 S^{a,X}_{k^\sigma l^\sigma} (i^\alpha j^\alpha | k^\sigma l^\sigma) \\
& + \sum_{\substack{X \neq K \\ X \in \text{UHF}}}^{N_{\text{UHF}}} \sum_\sigma \sum_{r^\sigma \in X}^{\text{vir}} \sum_{k^\sigma \in X}^{\text{occ}} 4 U^{a,X}_{r^\sigma k^\sigma} (i^\alpha j^\alpha | r^\sigma k^\sigma) \\
& - \sum_{\substack{X \neq K \\ X \in \text{RHF}}}^{N_{\text{RHF}}} \sum_{k \in X}^{\text{occ}} \sum_{l \in X}^{\text{occ}} 2 S^X_{kl} (i^\alpha j^\alpha | kl) \\
& + \sum_{\substack{X \neq K \\ X \in \text{RHF}}}^{N_{\text{RHF}}} \sum_{r \in X}^{\text{vir}} \sum_{k \in X}^{\text{occ}} 4 U^{a,X}_{rk} (i^\alpha j^\alpha | rk). \qquad (26)
\end{aligned}$$

Note that $\frac{\partial F^K_{i^\beta j^\beta}}{\partial a}$ is obtained by substituting α for β and β for α in Eq. (26).

The molecular orbital Hessian contribution is given by

$$A'_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk), \qquad (27)$$

$$\begin{aligned}
A'_{i^\alpha j^\alpha, k^\alpha l^\alpha} =& 2(i^\alpha j^\alpha | k^\alpha l^\alpha) - (i^\alpha k^\alpha | j^\alpha l^\alpha) \\
& - (i^\alpha l^\alpha | j^\alpha k^\alpha), \qquad (28)
\end{aligned}$$

$$A'_{i^\alpha j^\alpha, k^\beta l^\beta} = 2(i^\alpha j^\alpha | k^\beta l^\beta). \qquad (29)$$

Because the Fock matrix is diagonal, the derivatives of the Fock matrix are also diagonal: $\frac{\partial F_{ij}}{\partial a} = \delta_{ij} \frac{\partial \epsilon_i}{\partial a}$. Therefore, the expressions in Eq. (25) and Eq. (26) are equal to 0 for the offdiagonal elements.

We obtain the following equations for RHF:

$$\begin{aligned}
&\sum_{r\in I}^{\text{vir}}\sum_{k\in I}^{\text{occ}} A^{I,I}_{ij,rk}U^{a,I}_{rk} + \sum_{\substack{K\neq I\\ K\in\text{RHF}}}^{N_{\text{RHF}}}\sum_{r\in K}^{\text{vir}}\sum_{k\in K}^{\text{occ}} A^{I,K}_{ij,rk}U^{a,K}_{rk}\\
&+\sum_{\substack{X\neq I\\ X\in\text{UHF}}}^{N_{\text{UHF}}}\sum_{\sigma}\sum_{r^\sigma\in X}^{\text{vir}}\sum_{k^\sigma\in X}^{\text{occ}} A^{I,X}_{ij,r^\sigma k^\sigma}U^{a,X}_{r^\sigma k^\sigma}\\
&=B^{a,I}_{0,ij}, \qquad (30)
\end{aligned}$$

and for UHF there are two sets of equations, one for the α spin

$$\begin{aligned}
&\sum_{r^\alpha\in K}^{\text{vir}}\sum_{k^\alpha\in K}^{\text{occ}} A^{K,K}_{i^\alpha j^\alpha,r^\alpha k^\alpha}U^{a,K}_{r^\alpha k^\alpha} + \sum_{r^\beta\in K}^{\text{vir}}\sum_{k^\beta\in K}^{\text{occ}} A^{K,K}_{i^\alpha j^\alpha,r^\beta k^\beta}U^{a,K}_{r^\beta k^\beta}\\
&+\sum_{\substack{X\neq K\\ X\in\text{RHF}}}^{N_{\text{RHF}}}\sum_{r\in X}^{\text{vir}}\sum_{k\in X}^{\text{occ}} A^{K,X}_{i^\alpha j^\alpha,rk}U^{a,X}_{rk}\\
&+\sum_{\substack{X\neq K\\ X\in\text{UHF}}}^{N_{\text{UHF}}}\sum_{\sigma}\sum_{r^\sigma\in X}^{\text{vir}}\sum_{k^\sigma\in X}^{\text{occ}} A^{K,X}_{i^\alpha j^\alpha,r^\sigma k^\sigma}U^{a,X}_{r^\sigma k^\sigma}\\
&=B^{a,K}_{0,i^\alpha j^\alpha}, \qquad (31)
\end{aligned}$$

and the other for the β spin:

$$\begin{aligned}
&\sum_{r^\alpha\in K}^{\text{vir}}\sum_{k^\alpha\in K}^{\text{occ}} A^{K,K}_{i^\beta j^\beta,r^\alpha k^\alpha}U^{a,K}_{r^\alpha k^\alpha} + \sum_{r^\beta\in K}^{\text{vir}}\sum_{k^\beta\in K}^{\text{occ}} A^{K,K}_{i^\beta j^\beta,r^\beta k^\beta}U^{a,K}_{r^\beta k^\beta}\\
&+\sum_{\substack{X\neq K\\ K\in\text{RHF}}}^{N_{\text{RHF}}}\sum_{r\in K}^{\text{vir}}\sum_{k\in K}^{\text{occ}} A^{K,K}_{i^\beta j^\beta,rk}U^{a,K}_{rk}\\
&+\sum_{\substack{X\neq K\\ X\in\text{UHF}}}^{N_{\text{UHF}}}\sum_{\sigma}\sum_{r^\sigma\in X}^{\text{vir}}\sum_{k^\sigma\in X}^{\text{occ}} A^{K,X}_{i^\beta j^\beta,r^\sigma k^\sigma}U^{a,X}_{r^\sigma k^\sigma}\\
&=B^{a,K}_{0,i^\beta j^\beta} \qquad (32)
\end{aligned}$$

with the following definitions,

$$A^{I,I}_{ij,rk} = \delta_{ir}\delta_{jk}(\epsilon^I_j - \epsilon^I_i) - [4(ij|rk) - (ir|jk) - (ik|jr)] \tag{33}$$

$$\begin{aligned} A^{K,K}_{i^\alpha j^\alpha, r^\alpha k^\alpha} &= \delta_{i^\alpha r^\alpha}\delta_{j^\alpha k^\alpha}(\epsilon^K_{j^\alpha} - \epsilon^K_{i^\alpha}) \\ &\quad - [2(i^\alpha j^\alpha|r^\alpha k^\alpha) - (i^\alpha r^\alpha|j^\alpha k^\alpha) - (i^\alpha k^\alpha|j^\alpha r^\alpha)] \end{aligned} \tag{34}$$

$$A^{K,K}_{i^\alpha j^\alpha, r^\beta k^\beta} = 2(i^\alpha j^\alpha|r^\beta k^\beta) \tag{35}$$

$$A^{K,K}_{i^\beta j^\beta, r^\alpha k^\alpha} = 2(i^\beta j^\beta|r^\alpha k^\alpha) \tag{36}$$

$$\begin{aligned} A^{K,K}_{i^\beta j^\beta, r^\beta k^\beta} &= \delta_{i^\beta r^\beta}\delta_{j^\beta k^\beta}(\epsilon^K_{j^\beta} - \epsilon^K_{i^\beta}) \\ &\quad - \left[2(i^\beta j^\beta|r^\beta k^\beta) - (i^\beta r^\beta|j^\beta k^\beta) - (i^\beta k^\beta|j^\beta r^\beta)\right] \end{aligned} \tag{37}$$

$$A^{K,X}_{i^\alpha j^\alpha, rk} = -4(i^\alpha j^\alpha|rk) \tag{38}$$

$$A^{K,X}_{i^\beta j^\beta, rk} = -4(i^\beta j^\beta|rk) \tag{39}$$

$$A^{K,X}_{ij,rk} = -4(ij|rk) \tag{40}$$

$$\begin{aligned} B^{a,K}_{0,i^\alpha j^\alpha} &= F^{a,K}_{i^\alpha j^\alpha} - S^{a,K}_{i^\alpha j^\alpha}\epsilon^K_{j^\alpha} \\ &\quad - \sum_{k^\alpha l^\alpha \in K}^{\text{occ}} S^{a,K}_{k^\alpha l^\alpha}\left[(i^\alpha j^\alpha|k^\alpha l^\alpha) - (i^\alpha k^\alpha|j^\alpha l^\alpha)\right] \\ &\quad - \sum_{k^\beta l^\beta \in K}^{\text{occ}} S^{a,K}_{k^\beta l^\beta}(i^\alpha j^\alpha|k^\beta l^\beta) \\ &\quad - \sum_{\substack{X \neq K \\ X \in \text{UHF}}}^{N_{\text{UHF}}} \sum_\sigma \sum_{k^\sigma l^\sigma \in X}^{\text{occ}} 2S^{a,X}_{k^\sigma l^\sigma}(i^\alpha j^\alpha|k^\sigma l^\sigma) \\ &\quad - \sum_{\substack{X \neq K \\ X \in \text{RHF}}}^{N_{\text{RHF}}} \sum_{kl \in X}^{\text{occ}} 2S^{a,X}_{kl}(i^\alpha j^\alpha|kl) \end{aligned} \tag{41}$$

$$
\begin{aligned}
B^{a,K}_{0,i^\beta j^\beta} = {} & F^{a,K}_{i^\beta j^\beta} - S^{a,K}_{i^\beta j^\beta}\epsilon^K_{j^\beta} \\
& - \sum_{k^\beta l^\beta \in K}^{\mathrm{occ}} S^{a,K}_{k^\beta l^\beta}\left[(i^\beta j^\beta|k^\beta l^\beta) - (i^\beta k^\beta|j^\beta l^\beta)\right] \\
& - \sum_{k^\alpha l^\alpha \in K}^{\mathrm{occ}} S^{a,K}_{k^\alpha l^\alpha}(i^\beta j^\beta|k^\alpha l^\alpha) \\
& - \sum_{\substack{X\neq K \\ X\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_\sigma \sum_{k^\sigma l^\sigma \in X}^{\mathrm{occ}} 2S^{a,X}_{k^\sigma l^\sigma}(i^\beta j^\beta|k^\sigma l^\sigma) \\
& - \sum_{\substack{X\neq K \\ X\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{kl\in X}^{\mathrm{occ}} 2S^{a,X}_{kl}(i^\beta j^\beta|kl).
\end{aligned} \tag{42}
$$

$$
\begin{aligned}
B^{a,I}_{0,ij} = {} & F^{a,I}_{ij} - S^{a,I}_{ij}\epsilon^I_j \\
& - \sum_{kl\in I}^{\mathrm{occ}} S^{a,I}_{kl}\left[2(ij|kl) - (ik|jl)\right] - 2S^{a,I}_{kl}(ij|kl) \\
& - \sum_{\substack{X\neq K \\ X\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_\sigma \sum_{k^\sigma l^\sigma \in X}^{\mathrm{occ}} 2S^{a,X}_{k^\sigma l^\sigma}(ij|k^\sigma l^\sigma). \\
& - \sum_{\substack{X\neq K \\ X\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{kl\in X}^{\mathrm{occ}} 2S^{a,X}_{kl}(ij|kl).
\end{aligned} \tag{43}
$$

The CPHF equations and the response terms for the fully analytic energy gradient in FMO-UHF are

$$
\mathbf{A}\mathbf{U}^a = \mathbf{B}^a_0, \tag{44}
$$

and the gradient contribution in Eq. (18) is

$$
R^a = \mathbf{L}^T\mathbf{U}^a, \tag{45}
$$

where $\mathbf{L}^T$ is defined in Eq. (20).

By using the **Z**-vector method of Handy and Schaefer[55], the following set of simultaneous equations is solved for **Z**:

$$
\mathbf{A}^T\mathbf{Z} = \mathbf{L}, \tag{46}
$$

and the contribution to the FMO energy gradient is,

$$R^a = \mathbf{Z}^T \mathbf{B}_0^a. \tag{47}$$

In FMO one can separate[41] the diagonal (I,I), (K,K) and off-diagonal (K,I) blocks of $\mathbf{A}$ in Eq. (46): for RHF fragments as

$$\begin{aligned}
&\sum_{r\in I}^{\text{vir}} \sum_{i\in I}^{\text{occ}} A^{I,I}_{ri,kl} Z^I_{ri} \\
=&L^I_{kl} - \sum_{\substack{X\neq I \\ X\in \text{RHF}}}^{N_{\text{RHF}}} \sum_{r\in X}^{\text{vir}} \sum_{i\in X}^{\text{occ}} A^{X,I}_{ri,kl} Z^X_{ri} \\
&- \sum_{\substack{X\neq I \\ X\in \text{UHF}}}^{N_{\text{UHF}}} \sum_{\sigma} \sum_{r^\sigma\in X}^{\text{vir}} \sum_{i^\sigma\in X}^{\text{occ}} A^{X,I}_{r^\sigma i^\sigma,kl} Z^X_{r^\sigma i^\sigma},
\end{aligned} \tag{48}$$

and for UHF as

$$\begin{aligned}
&\sum_{\sigma} \sum_{r^\sigma\in K}^{\text{vir}} \sum_{i^\sigma\in K}^{\text{occ}} A^{K,K}_{r^\sigma i^\sigma,k^\alpha l^\alpha} Z^K_{r^\sigma i^\sigma} \\
=&L^K_{k^\alpha l^\alpha} - \sum_{\substack{X\neq K \\ X\in \text{RHF}}}^{N_{\text{RHF}}} \sum_{r\in X}^{\text{vir}} \sum_{i\in X}^{\text{occ}} A^{X,K}_{ri,k^\alpha l^\alpha} Z^X_{ri} \\
&- \sum_{\substack{X\neq K \\ X\in \text{UHF}}}^{N_{\text{UHF}}} \sum_{\sigma} \sum_{r^\sigma\in X}^{\text{vir}} \sum_{i^\sigma\in X}^{\text{occ}} A^{X,K}_{r^\sigma i^\sigma,k^\alpha l^\alpha} Z^X_{r^\sigma i^\sigma},
\end{aligned} \tag{49}$$

and

$$\begin{aligned}
&\sum_{\sigma} \sum_{r^\sigma\in K}^{\text{vir}} \sum_{i^\sigma\in K}^{\text{occ}} A^{K,K}_{r^\sigma i^\sigma,k^\beta l^\beta} Z^K_{r^\sigma i^\sigma} \\
=&L^K_{r^\beta i^\beta} - \sum_{\substack{X\neq K \\ X\in \text{RHF}}}^{N_{\text{RHF}}} \sum_{r\in X}^{\text{vir}} \sum_{i\in X}^{\text{occ}} A^{X,K}_{ri,k^\beta l^\beta} Z^X_{ri} \\
&- \sum_{\substack{X\neq K \\ X\in \text{UHF}}}^{N_{\text{UHF}}} \sum_{\sigma} \sum_{r^\sigma\in X}^{\text{vir}} \sum_{i^\sigma\in X}^{\text{occ}} A^{X,K}_{r^\sigma i^\sigma,k^\beta l^\beta} Z^X_{r^\sigma i^\sigma}.
\end{aligned} \tag{50}$$

Because the number of $\mathbf{Z}$-vector elements is usually so large that the linear equation system cannot be solved directly, iterative techniques are employed in practice. The external

unknowns Z_{ri}^{K} in the last two terms of Eq. (48), (49), and (50) are frozen, and the diagonal elements Z_{ri}^{I} are solved for all fragments independently. Consequently, Z_{ri}^{I} and Z_{ri}^{K} frozen above are updated, and this process is iterated until full self-consistency in the self-consistent **Z**-vector method (SCZV)[41] for both RHF and UHF fragments together. SCZV typically converges in 10-20 iterations. In practice, the response operators are calculated in the atomic orbital basis and transformed back to the molecular orbital basis[56].

C. Computational Details

The developed equations for energy and fully analytic energy gradients of FMO-UHF were implemented into GAMESS.[57] All developed code was parallelized with the generalized distributed data interface (GDDI)[58]. Cartesian basis set were used in our calculations (ISPHER=-1).

To evaluate the errors in energies and analytic energy gradients in comparison with *ab initio* UHF calculations, we used the 2,2,6,6-tetramethylpiperidine-N-oxide (TEMPO)[59] (FIG. 1 and TABLE S1) and the active part of the protein *Pseudomonas aeruginosa* azurin (Az)[60] (FIG. 2 and TABLE S2) for the test of the FMO-UHF energies and analytic energy gradients (PDB ID of Az is 1R1C.). The neutral doublet state is experimentally observed in TEMPO and Az. The unpaired electron is localized on the N-oxide site in TEMPO, and the unpaired electron is localized on the tryptophan (residue 108) in Az.

The geometry of TEMPO surrounded by 5, 11, and 17 dichloromethane molecules (in which the radical is usually solvated in experiments) was used in order to test the dependence of energy errors on the system size. We first optimized the structure of TEMPO surrounded by 17 dichloromethane molecules using the default geometry optimizer in GAMESS based on the numeric updates of the Hessian. The geometry of TEMPO surrounded by 5 and 11 dichloromethane were then obtained by removing 6 or 12 dichloromethane molecules from the equilibrium geometry of TEMPO surrounded by 17 dichloromethane molecules. The residues 104-109 (around the tryptophan resides) of the X-ray structure[60] were used for Az. The hydrogen atoms were added to the protein by AMBER[62].

Secondly, we examined the errors of FMO-UHF analytic energy gradients in comparison

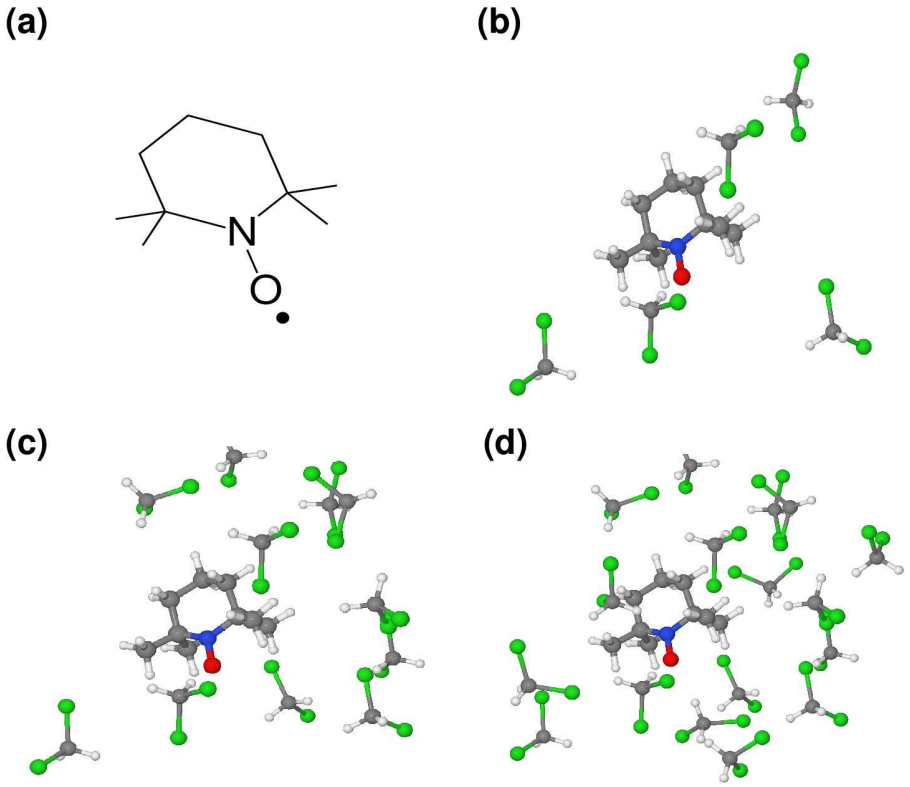


FIG. 1. (a) Structure of the stable organic radical, 2,2,6,6-tetramethylpiperidine-N-oxide (TEMPO). TEMPO is calculated in solution, immersed in (b) 5, (c) 11 and (d) 17 dichloromethane molecules.

with numerical energy gradients. The analytic energy gradients were calculated both with the response terms (including Eq. (19) i.e. fully analytic energy gradients) and excluding them to test the FMO-UHF energy gradients. The numerical gradients were computed with double differencing and a coordinate step of 0.0005 Å. For the polypeptide, we only considered numeric derivatives for the bond attached and detached atoms (BAA and BDA, respectively), which are on the fragment border and are typically found to have the largest errors[41].

Finally, we applied the FMO-UHF analytic energy gradients for the optimization of the structure of Manganese(III) acetylacetonate ($Mn(acac)_3$) (FIG. 3) solvated in 6 dichloromethane molecules. The spin state is a quintet. The equilibrium geometry is compared with the

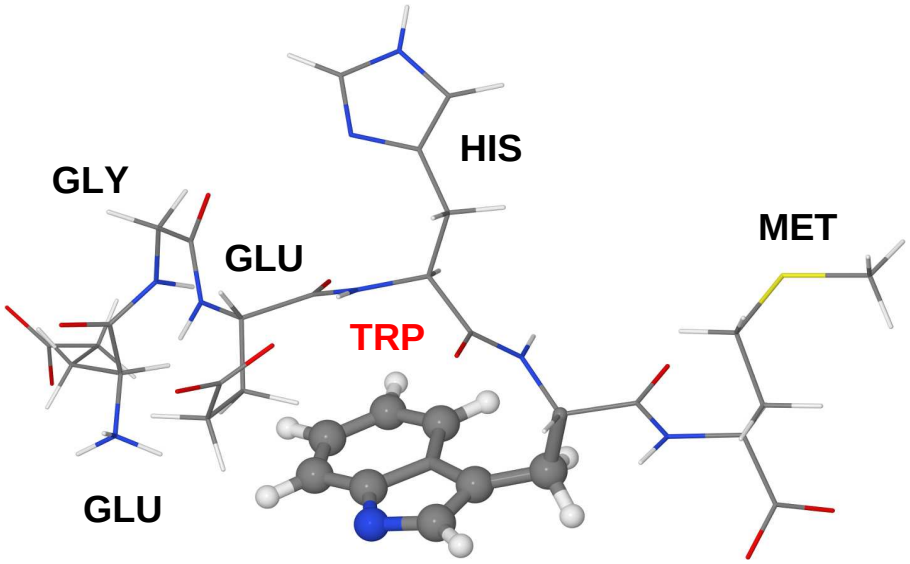


FIG. 2. Geometry of the Az polypeptide, the active site of the 1R1C protein, which is a neutral radical in the doublet spin state.

experimental X-ray structure[63].

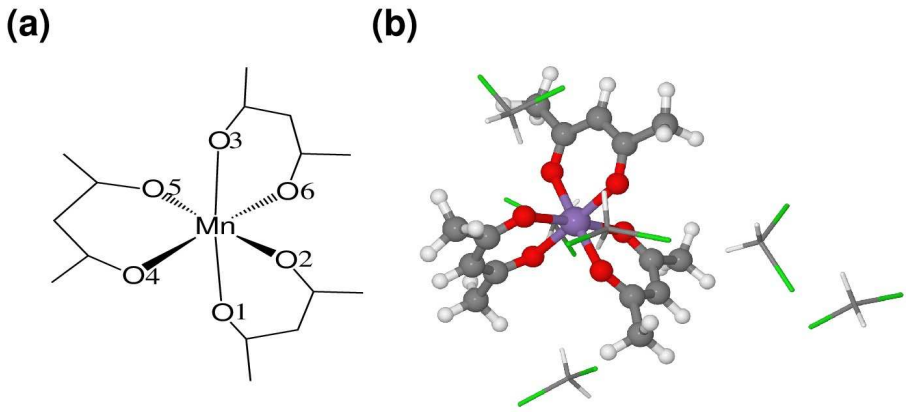


FIG. 3. Mn(III) acetylacetonate ($Mn(acac)_3$) complex: (a) chemical structure, (b) $Mn(acac)_3$ immersed in six dichloromethane molecules.

In TEMPO and $Mn(acac)_3$, each molecule is assigned as one fragment, e.g. TEMPO surrounded by 5 dichloromethane molecules has 6 fragments. In Az, the polypeptide is divided as amino acid residue per fragment (6 fragments total).

We do not use ESP approximations in this work, and other thresholds have their default values.

D. Results and Discussion

1. Accuracy of the FMO-UHF energies

We examine the accuracy of the energy by comparing FMO-UHF to *ab initio* UHF calculations. The errors in FMO2-UHF and FMO3-UHF calculations in comparison with the corresponding *ab initio* calculations are shown in TABLE I. For TEMPO, the errors

TABLE I. Energy differences (kcal/mol) between *ab initio* and FMO-based UHF.

method	6-31G(d)	6-311G(d)
	TEMPO solvated with 17 dichloromethane molecules	
FMO2	0.166	-0.608
FMO3	0.002	0.029
	TEMPO solvated with 11 dichloromethane molecules	
FMO2	-0.033	0.145
FMO3	-0.000	0.001
	TEMPO solvated with 5 dichloromethane molecules	
FMO2	0.005	-0.025
FMO3	0.000	-0.000
	Az polypeptide	
FMO2	-1.667	10.354
FMO3	0.049	0.109

for FMO2-UHF calculations do not exceed 0.166 kcal/mol for 6-31G(d) and 0.608 kcal/mol for 6-311G(d). The addition of the three-body terms in FMO3-UHF significantly reduces the error to 0.002 kcal/mol for 6-31G(d) and 0.029 kcal/mol for 6-311G(d). As in FMO-RHF, the three-body effects[50] in FMO-UHF are important, i.e., the errors in FMO3-UHF are more than one order of magnitude smaller than those in FMO2-UHF. Similarly, the dipole moments obtained by FMO-UHF agree well with those by the *ab initio* calculations (TABLE II). The maximum absolute error is only 0.0042 Debye for 6-31G(d) and -0.0111 Debye for 6-311G(d). The errors are greatly reduced when FMO3-UHF is used.

TABLE II. Dipole moment differences (Debye) between *ab initio* and FMO-based UHF.

method	6-31G(d)	6-311G(d)
	TEMPO solvated with 17 dichloromethane molecules	
FMO2	0.0033	-0.0111
FMO3	-0.0028	0.0019
	TEMPO solvated with 11 dichloromethane molecules	
FMO2	0.0042	-0.0021
FMO3	0.0003	-0.0002
	TEMPO solvated with 5 dichloromethane molecules	
FMO2	-0.0026	-0.0004
FMO3	0.0004	0.0005
	Az polypeptide	
FMO2	-0.0549	1.3005
FMO3	-0.0111	-0.1811

The errors of FMO-UHF in Az are -1.667 kcal/mol in FMO2, and 0.049 kcal/mol in FMO3 for the 6-31G(d) basis set. We obtained a relatively large error in energy (10.354

kcal/mol) for the two-body FMO-UHF with the 6-311G(d) basis set. The larger basis set has a more extended radical delocalization; the FMO2 error can be reduced either with FMO3 (to 0.109 kcal/mol) or by enlarging the UHF fragment and using FMO2: when tryptophan and histidine are merged, the FMO2-UHF error becomes 3.385 kcal/mol. A similar behaviour was also observed for FMO-ROHF[32].

2. Accuracy of the FMO2-UHF energy gradient

The FMO2-UHF analytic energy gradients are compared with analytic energy gradients of *ab initio* UHF calculations. Also, we examine the response contribution in Eq. (19).

TABLE III. Maximum (MAX) and RMS gradients (hartree/bohr) for *ab initio* UHF and fully analytic FMO-UHF, using 6-31G(d).

FMO-UHF		UHF	
MAX	RMS	MAX	RMS
	TEMPO solvated with 17 dichloromethane molecules		
0.000094	0.000022	0.000095	0.000021
	TEMPO solvated with 11 dichloromethane molecules		
0.003023	0.000650	0.003021	0.000649
	TEMPO solvated with 5 dichloromethane molecules		
0.003463	0.000886	0.003463	0.000886
	Az polypeptide		
0.081978	0.020847	0.081997	0.020919

The summary of the results are shown in TABLE III, in terms of the maximum absolute energy gradient values, and the root mean square (RMS) values. The fully analytic energy

gradients agree well with those of the *ab initio* UHF calculations (the errors are shown in TABLE IV). The energy errors are within 1.0×10^{-4} hartree/bohr in organic molecules, and 5.0×10^{-3} in Az polypeptide.

TABLE IV. Maximum (MAX) and RMS differences between UHF and FMO-UHF fully analytic energy gradients (hartree/bohr), using 6-31G(d).

MAX	RMS
TEMPO solvated with 17 dichloromethane molecules	
0.000051	0.000007
TEMPO solvated with 11 dichloromethane molecules	
0.000014	0.000002
TEMPO solvated with 5 dichloromethane molecules	
0.000008	0.000001
Az polypeptide	
0.002320	0.001177

We compare the analytic energy gradients with the numerical energy gradients to examine the importance of the response terms in Eq. (19). The absolute energy gradient errors and RMS error are shown in TABLE V. The maximum gradient error is reduced from 4.3×10^{-5} to 3.0×10^{-6} by the consideration of the response terms in Eq. (19). Similarly, the consideration of response term also reduced both the maximum gradient errors and RMS errors in Az polypeptide. There is some indication that the residual difference between the analytic and numeric gradients is due to some inherent numeric behaviour of UHF, so that the numeric gradient is inaccurate.

To elucidate the accuracy of the fully analytic energy gradients furthermore, we plotted the errors between the analytic and numeric energy gradients in FIG. 4. The reduction of the

TABLE V. Maximum (MAX) differences and RMSD between the numeric and analytic energy gradients for FMO-UHF/6-31G(d), in hartree/bohr.

SCZV[a]		No SCZV[b]	
MAX	RMSD	MAX	RMSD
	TEMPO solvated with 11 molecules		
0.000003	0.000001	0.000043	0.000007
	Az polypeptide		
0.000050	0.000024	0.000501	0.000206

[a] with the response terms in Eq. (19), [b] without them.

errors by the consideration of the response terms is clear. The errors of all gradient elements are reduced by taking into account the response terms (FIG. 4), which are important to get accurate gradients.

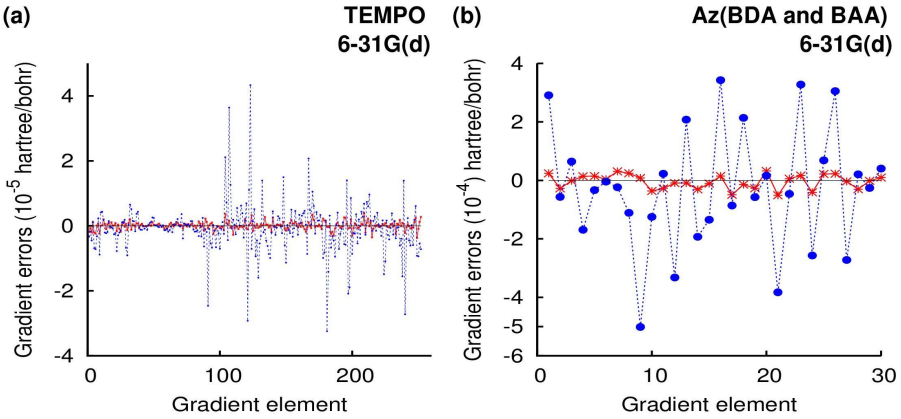


FIG. 4. Difference between analytic and numerical gradients in hartree/bohr. Red solid line shows the error between fully analytic FMO-UHF gradient and numeric values. Blue dashed line denotes the errors for the approximate gradient (no response terms Eq. (19)). (a) TEMPO immersed in 11 dichloromethane molecules (the errors are shown for all atoms) (b) Az polypeptide, see FIG. 2 (the errors are shown only for atoms on the fragment border, BDAs and BAAs).

3. Geometry optimization by FMO2-UHF

Geometry optimizations were carried out using the developed FMO2-UHF fully analytic energy gradients. We compare the optimized structure of $Mn(acac)_3$ (FIG. 3) obtained with FMO-UHF and *ab initio* UHF calculations.

The root mean-square deviations (RMSD), RMS bond lengths, and RMS bond angles are summarized in TABLE VI. The energy difference between FMO2-UHF and *ab initio* UHF

TABLE VI. Properties at the geometries of $Mn(acac)_3$ optimized with each method separately using 6-31G(d). The total energies at the minima are in Hartree. In RMSD (Å) to X-ray experiment only heavy atoms are used. The maximum (MAX) and RMS deviations of bond lengths (Å) and angles (deg.) are between *ab initio* UHF and FMO-UHF.

	ab initio UHF	FMO-UHF, no SCZV[a]	FMO-UHF, SCZV[b]
Total Energy	-7927.2412337	-7927.2419148	-7927.2419125
RMSD to exp[63]	0.2924	0.2926	0.2906
	Difference between *ab initio* UHF and FMO-UHF		
Bond RMS	...	0.0004	0.0007
Bond MAX	...	0.0015	0.0024
Angle RMS	...	0.0851	0.1068
Angle MAX	...	0.2756	0.3911
RMSD	...	0.1224	0.1540

[a] without the response terms in Eq. (19), [b] with the terms.

calculations at their respective minima is only 0.427 kcal/mol for both the fully analytic and approximate (no response terms) gradients. The RMS deviations from X-ray are 0.2924 Å for *ab initio*, 0.2926 Å for FMO-UHF without response terms, and 0.2906 Å for FMO-UHF with the response terms (see TABLE VI).

We also compare the difference between FMO-UHF and *ab initio* UHF calculations. The

RMSD from *ab initio* structures is 0.1540 Å and 0.1224 Å with and without the response terms, respectively. Thus, FMO-UHF quite accurately reproduces the geometry obtained by *ab initio* UHF calculations.

To elucidate the effect of the response terms, we compare the minima obtained by optimizing the geometry with and without the response terms. The total energy differed by 0.001 kcal/mol, while the RMSD between the optimized geometries was 0.0580 Å.

4. Calculation efficiency of FMO2-UHF

The efficiency of FMO2-UHF calculations was tested by comparing the CPU timings for *ab initio* UHF and FMO2-UHF calculation, separately for the energy and the energy with the fully analytic gradient.

All the calculations were performed on a single node equipped with dual 2.93 GHz Xeons (6 cores per CPU) and 64 GB RAM per node. The results are shown in FIG. 5. The CPU

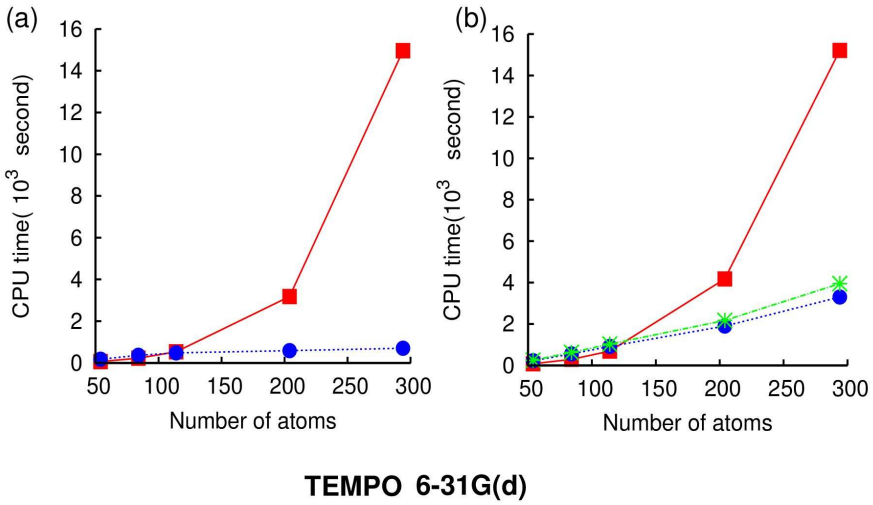


FIG. 5. Comparison of timings for UHF and FMO-UHF calculations in second. Dashed and solid lines denote FMO-UHF and UHF, respectively. The total of 12 cores of two 2.93 GHz Xeon CPUs is used. (a) energy, (b) gradient. For the gradient, stars and circles show the timings for the FMO-UHF gradient with and without the response terms in Eq. (19).

times of FMO2-UHF are less than those of *ab initio* UHF when the fragment size is larger

than 36 both for the energy and gradient calculations. For larger clusters, FMO2-UHF outperforms *ab initio* UHF. For 54 fragments, FMO2-UHF fully analytic energy gradients took 3948 seconds, while UHF analytic energy gradients took 15213 seconds (about a four-fold decrease in the timing). Response terms take about 16 % of the total time (Table VII). Although this fraction is not negligible, a geometry optimization with response terms usually takes fewer iterations (by 10-25 % depending on the system), because the gradient is accurate.

TABLE VII. Computational time (seconds) for reponse terms (SCZV) and the total FMO-UHF time for TEMPO solvated in n dichloromethane molecules, 6-31G(d).

n	SCZV	Total
5	22.6	248.3
11	70.6	626.1
17	110.6	1028.9
35	275.3	2174.0
53	647.9	3948.5

E. Conclusions

We have developed the spin-unrestricted Hartree-Fock formulation of the fragment molecular orbital method. Both energy and the fully analytic energy gradients have been successfully derived, implemented and parallelized in this study. The accuracy of both the energy and gradient has been evaluated in comparison with *ab initio* UHF without fragmentation. In addition, the accuracy of the gradient has been demonstrated in comparison to numeric gradients. We have also performed a geometry optimization using UHF and FMO-UHF demonstrating the reasonable agreement between the two structures, as well as their comparison to experiment. We note that neglecting response terms can be a reasonable approach in many cases, excluding perhaps molecular dynamics where it is desirable to have very accurate analytic gradients.

Pair interaction energies can be easily defined for UHF (Eq. (1)) analogously to RHF[20], providing useful information about the interaction in large systems and giving some insight about chemical processes, such as solute-solvent interactions or protein-ligand binding.

The FMO2-UHF method is an efficient approach for large chemical systems with unpaired electrons, and we expect it to be useful in optimizing geometries and studying energetics in large molecular systems.

Appendix A

The differentiation of $\mathrm{Tr}(\Delta\mathbf{D}^{XY}\mathbf{V}^{XY})$ with respect to a nuclear coordinate a leads for RHF to

$$
\begin{aligned}
&\frac{\partial \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a} \\
&= \sum_{\mu\nu\in I,J} \Delta D^{IJ}_{\mu\nu} \sum_{X\neq I,J} \sum_{\lambda,\sigma\in X} \left[\frac{\partial u^X_{\mu\nu}}{\partial a}(\mu\nu|\lambda\sigma) + D^X_{\lambda\sigma}\frac{\partial(\mu\nu|\lambda\sigma)}{\partial a} \right] \\
&+4 \sum_{r\in I,J}^{\mathrm{vir}} \sum_{i\in I,J}^{\mathrm{occ}} U^{a,IJ}_{ri} V^{IJ}_{ri} \\
&-4 \sum_{r\in I}^{\mathrm{vir}} \sum_{i\in I}^{\mathrm{occ}} U^{a,I}_{ri} V^{IJ}_{ri} \\
&-4 \sum_{r\in J}^{\mathrm{vir}} \sum_{i\in J}^{\mathrm{occ}} U^{a,J}_{ri} V^{IJ}_{ri} \\
&-2 \sum_{\mu,\nu\in I,J}^{\mathrm{occ}} \frac{\partial S^{IJ}_{\mu\nu}}{\partial a} W^{IJ,\mathrm{RHF}}_{\mu\nu} \\
&+2 \sum_{\mu,\nu\in I}^{\mathrm{occ}} \frac{\partial S^{I}_{\mu\nu}}{\partial a} W^{I,\mathrm{RHF}}_{\mu\nu} \\
&+2 \sum_{\mu,\nu\in J}^{\mathrm{occ}} \frac{\partial S^{J}_{\mu\nu}}{\partial a} W^{J,\mathrm{RHF}}_{\mu\nu} \\
&-2 \sum_{\substack{X\neq I,J \\ X\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu\in X} \Delta X^{X(IJ,\mathrm{RHF})}_{\mu\nu} \frac{\partial S^X_{\mu\nu}}{\partial a} \\
&-2 \sum_{\substack{X\neq I,J \\ X\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu\in X} \Delta X^{X(IJ,\mathrm{UHF})}_{\mu\nu} \frac{\partial S^X_{\mu\nu}}{\partial a} \\
&+2 \sum_{\mu,\nu\in I,J} \Delta D^{IJ}_{\mu\nu} \sum_{\substack{X\neq I,J \\ X\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\sigma} \sum_{r^\sigma}^{\mathrm{vir}} \sum_{i^\sigma}^{\mathrm{occ}} U^{a,X}_{r^\sigma i^\sigma}(\mu\nu|r^\sigma i^\sigma) \\
&+4 \sum_{\mu,\nu\in I,J} \Delta D^{IJ}_{\mu\nu} \sum_{\substack{X\neq I,J \\ X\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{r}^{\mathrm{vir}} \sum_{i}^{\mathrm{occ}} U^{a,X}_{ri}(\mu\nu|ri),
\end{aligned}
\tag{51}
$$

and for UHF one has

$$
\begin{aligned}
\frac{\partial \mathrm{Tr}(\Delta \mathbf{D}^{KL}\mathbf{V}^{KL})}{\partial a} =& 2\sum_{\sigma}\sum_{r^\sigma\in K,L}^{\mathrm{vir}}\sum_{i^\sigma\in K,L}^{\mathrm{occ}} U^{a,KL}_{r^\sigma i^\sigma}V^{KL}_{r^\sigma i^\sigma} \\
&-2\sum_{\sigma}\sum_{r^\sigma\in K}^{\mathrm{vir}}\sum_{i^\sigma\in K}^{\mathrm{occ}} U^{a,K}_{r^\sigma i^\sigma}V^{KL}_{r^\sigma i^\sigma} \\
&-2\sum_{\sigma}\sum_{r^\sigma\in L}^{\mathrm{vir}}\sum_{i^\sigma\in L}^{\mathrm{occ}} U^{a,L}_{r^\sigma i^\sigma}V^{KL}_{r^\sigma i^\sigma} \\
&-2\sum_{\mu,\nu\in KL}^{\mathrm{occ}}\frac{\partial S^{KL}_{\mu\nu}}{\partial a}W^{KL,\mathrm{UHF}}_{\mu\nu} \\
&+2\sum_{\mu,\nu\in K}^{\mathrm{occ}}\frac{\partial S^{K}_{\mu\nu}}{\partial a}W^{K,\mathrm{UHF}}_{\mu\nu} \\
&+2\sum_{\mu,\nu\in L}^{\mathrm{occ}}\frac{\partial S^{L}_{\mu\nu}}{\partial a}W^{L,\mathrm{UHF}}_{\mu\nu} \\
&+\sum_{\mu,\nu,\in K,L}\Delta D^{\alpha+\beta,KL}_{\mu\nu} \\
&\quad\left[\sum_{X\neq K,L}\sum_{\lambda,\sigma\in X}\frac{\partial u^{X}_{\mu\nu}}{\partial a}(\mu\nu|\lambda\sigma)+D^{K}_{\lambda\sigma}\frac{\partial(\mu\nu|\lambda\sigma)}{\partial a}\right] \\
&-2\sum_{\substack{X\neq K,L\\ X\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}}\sum_{\mu,\nu\in X}\Delta X^{X(KL,\mathrm{RHF})}_{\mu\nu}\frac{\partial S^{X}_{\mu\nu}}{\partial a} \\
&-2\sum_{\substack{X\neq K,L\\ X\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}}\sum_{\mu,\nu\in X}\Delta X^{X(KL,\mathrm{UHF})}_{\mu\nu}\frac{\partial S^{X}_{\mu\nu}}{\partial a} \\
&+2\sum_{\mu,\nu\in K,L}\Delta D^{\alpha+\beta,KL}_{\mu\nu} \\
&\quad\sum_{\substack{X\neq K,L\\ X\in\mathrm{UHF}}}^{N_{\mathrm{UHF}}}\sum_{\sigma}\sum_{r^\sigma}^{\mathrm{vir}}\sum_{i^\sigma}^{\mathrm{occ}}U^{a,X}_{r^\sigma i^\sigma}(\mu\nu|r^\sigma i^\sigma) \\
&+4\sum_{\mu,\nu\in K,L}\Delta D^{\alpha+\beta,KL}_{\mu\nu} \\
&\quad\sum_{\substack{X\neq K,L\\ X\in\mathrm{RHF}}}^{N_{\mathrm{RHF}}}\sum_{r}^{\mathrm{vir}}\sum_{i}^{\mathrm{occ}}U^{a,X}_{ri}(\mu\nu|ri), \qquad (52)
\end{aligned}
$$

and for the mixed case of RHF and UHF,

$$
\begin{aligned}
\frac{\partial \mathrm{Tr}(\Delta \mathbf{D}^{KI} \mathbf{V}^{KI})}{\partial a} =& 2 \sum_{\sigma} \sum_{r^\sigma \in K,I}^{\mathrm{vir}} \sum_{i^\sigma \in K,I}^{\mathrm{occ}} U^{a,KI}_{r^\sigma i^\sigma} V^{KI}_{r^\sigma i^\sigma} \\
& -2 \sum_{\sigma} \sum_{r^\sigma \in K}^{\mathrm{vir}} \sum_{i^\sigma \in K}^{\mathrm{occ}} U^{a,K}_{r^\sigma i^\sigma} V^{KI}_{r^\sigma i^\sigma} \\
& -2 \sum_{r \in I}^{\mathrm{vir}} \sum_{i \in I}^{\mathrm{occ}} U^{a,I}_{ri} V^{KI}_{ri} \\
& -2 \sum_{\mu,\nu \in K,I}^{\mathrm{occ}} \frac{\partial S^{KI}_{\mu\nu}}{\partial a} W^{KI,\mathrm{UHF}}_{\mu\nu} \\
& +2 \sum_{\mu,\nu \in K}^{\mathrm{occ}} \frac{\partial S^{K}_{\mu\nu}}{\partial a} W^{K,\mathrm{UHF}}_{\mu\nu} \\
& +2 \sum_{\mu,\nu \in I}^{\mathrm{occ}} \frac{\partial S^{I}_{\mu\nu}}{\partial a} W^{I,\mathrm{RHF}}_{\mu\nu} \\
& + \sum_{\mu,\nu \in K,I} \Delta D^{\alpha+\beta,KI}_{\mu\nu} \\
& \quad \left[\sum_{X \neq K,I} \sum_{\lambda,\sigma \in X} \frac{\partial u^{X}_{\mu\nu}}{\partial a} (\mu\nu|\lambda\sigma) + D^{X}_{\lambda\sigma} \frac{\partial (\mu\nu|\lambda\sigma)}{\partial a} \right] \\
& \\
& -2 \sum_{\substack{X \neq K,I \\ X \in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{\mu,\nu \in X} \Delta X^{X(KI,\mathrm{RHF})}_{\mu\nu} \frac{\partial S^{X}_{\mu\nu}}{\partial a} \\
& -2 \sum_{\substack{X \neq K,I \\ X \in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\mu,\nu \in X} \Delta X^{X(KI,\mathrm{UHF})}_{\mu\nu} \frac{\partial S^{X}_{\mu\nu}}{\partial a} \\
& +2 \sum_{\mu,\nu \in K,I} \Delta D^{\alpha+\beta,KI}_{\mu\nu} \\
& \quad \sum_{\substack{X \neq K,I \\ X \in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} \sum_{\sigma} \sum_{r^\sigma}^{\mathrm{vir}} \sum_{i^\sigma}^{\mathrm{occ}} U^{a,X}_{r^\sigma i^\sigma} (\mu\nu|r^\sigma i^\sigma), \\
& +4 \sum_{\mu,\nu \in K,I} \Delta D^{\alpha+\beta,KI}_{\mu\nu} \\
& \quad \sum_{\substack{X \neq K,I \\ X \in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} \sum_{r}^{\mathrm{vir}} \sum_{i}^{\mathrm{occ}} U^{a,X}_{ri} (\mu\nu|ri)
\end{aligned}
\tag{53}
$$

with the following definitions:

$$W_{\mu\nu}^{X,\mathrm{RHF}} = \frac{1}{4}\sum_{\lambda,\sigma\in X}^{\mathrm{occ}} D_{\mu\lambda}^{X} D_{\nu\sigma}^{X} V_{\lambda\sigma}^{IJ}, \tag{54}$$

$$W_{\mu\nu}^{X,\mathrm{UHF}} = \frac{1}{2}\sum_{\lambda,\sigma\in K,L}^{\mathrm{occ}} D_{\mu\lambda}^{\alpha,X} D_{\nu\sigma}^{\alpha,X} V_{\lambda\sigma}^{KL} + D_{\mu\lambda}^{\beta,X} D_{\nu\sigma}^{\beta,X} V_{\lambda\sigma}^{KL}, \tag{55}$$

$$\Delta X_{\mu\nu}^{X(IJ,\mathrm{RHF})} = \frac{1}{4}\sum_{\zeta,\eta\in I,J}\sum_{\lambda,\sigma\in X} \Delta D_{\zeta\eta}^{IJ} D_{\mu\lambda}^{X} D_{\nu\sigma}^{X} (\zeta\eta|\lambda\sigma), \tag{56}$$

$$\Delta X_{\mu\nu}^{X(KL,\mathrm{UHF})} = \frac{1}{2}\sum_{\zeta,\eta\in K,L}\sum_{\lambda,\sigma\in X} \Delta D_{\zeta\eta}^{KL}\Big(D_{\mu\lambda}^{\alpha,X} D_{\nu\sigma}^{\alpha,X} + D_{\mu\lambda}^{\beta,X} D_{\nu\sigma}^{\beta,X}\Big)(\zeta\eta|\lambda\sigma). \tag{57}$$

REFERENCES

[1]Goedecker, S. Rev. Mod. Phys. **71**, 1085 (1999).

[2]Scuseria, G. E. J. Phys. Chem. A.**103**, 4782 (1999).

[3]Linear-Scaling Techniques in Computational Chemistry and Physics. R. Zalesny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski (Eds.), Springer, New York, 2011.

[4]J. R. Reimers, editor, Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology, Wiley, New York, 2011.

[5]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[6]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[7]J. L. Gao, J. Phys. Chem. B. **101**, 657 (1997).

[8]D. W. Zhang, Y. Xiang, J. Z. H. Zhang, J. Phys. Chem. B. **107**, 12039 (2003).

[9]H. R. Leverentz and D. G. Truhlar, J. Chem. Theory Comput. **5**, 1573 (2009).

[10]M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, and J. A. Boatz, J. Phys. Chem. B. **113**, 9646 (2009).

[11]L. Huang, L. Massa, I. Karle, and J. Karle, Proc. Nat. Acad. Sc. U.S.A. **106**, 3664 (2009).

[12]P. Söderhjelm, J. Kongsted, and U. Ryde, J. Chem. Theory Comput. **6**, 1726 (2010).

[13]S. D. Yeole and S. Gadre, J. Chem. Phys. **132**, 094102 (2010).

[14]J. Korchowiec, F. L. Gu, and Y. Aoki, Int. J. Quantum Chem. **105**, 875 (2005).

[15]M. Kobayashi, T. Yoshikawa, and H. Nakai, Chem. Phys. Lett. **500**, 172 (2010).

[16]X. He and K. M. Merz, J. Chem. Theory Comput. **6**, 405 (2010).

[17]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi Chem. Phys. Lett. **313**, 701 (1999).

[18]D. G. Fedorov, K. Kitaura (Eds.), The Fragment Molecular Orbital Method: practical applications to large molecular systems (CRC press, Boca Raton, FL), 2009.

[19]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).

[20]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[21]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).

[22]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[23]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Phys. Chem. B. **114**, 15700 (2010).

[24]M. P. Mazanetz, O. Ichihara, R. J. Law, and M. Whittaker, J. Cheminf. **3**, 2 (2011).

[25]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B. **116**, 503 (2012).

[26]H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii, and K. Kitaura, J. Phys. Chem. A. **112**, 10887 (2008).

[27]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B. **111**, 9621 (2007).

[28]S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, Mol. Phys. **103**, 2255 (2005).

[29]Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **406**, 283 (2005).

[30]R. A. Mata, H. Stoll, and B. J. C. Cabral, J. Chem. Theory Comput. **5**, 1829 (2009).

[31]A. D. Fusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, J. Phys. Chem. Lett. **2**, 2184 (2011).

[32]S. R. Pruitt, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Theory Comp. **6**, 1 (2010).

[33]S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, J. Phys. Chem. A. **116**, 4965 (2012).

[34]T. Yoshikawa, M. Kobayashi, and H. Nakai, Theor. Chem. Acc. **130**, 411-417 (2011).

[35]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **122**, 054108 (2005).

[36]A. Hansen, G. L. Dimitrios, and F. Neese, J. Chem. Phys. **135**, 214102 (2011).

[37]L. Noodleman, J. Chem. Phys. **74**, 5737 (1981).

[38]K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, Chem. Phys. Lett. **336**, 163 (2001).

[39]T. Nagata, D. G. Fedorov and K. Kitaura, Chem. Phys. Lett. **475**, 124 (2009).

[40]T. Nagata, D. G. Fedorov and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).

[41]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[42]T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, J. Chem. Phys. **135**, 044110 (2011).

[43]T. Nagata, D. G. Fedorov, H. Li, and K. Kitaura, J. Chem. Phys. **136**, 204112 (2012).

[44]T. Nagata, D. G. Fedorov, and K. Kitaura. Theor. Chem. Acc. **131**, 1136 (2012).

[45]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A. **111**, 2722 (2007).

[46]Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **372**, 342 (2003).

[47]T. Fujita, H. Watanabe, and S. Tanaka, J. Phys. Soc. Jpn. **78**, 104723 (2009).

[48]T. Fujita, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **506**, 112 (2011).

[49]O. Hino, S. Tanimori, T. Ogawa, M. Kobayashi, and K. Hasegawa, poster presentation 2P075, Molecular Structure Conference, Sendai, Japan, 2007 (in Japanese). http://www.wdc-jp.biz/msf/ps2007/jp/download_file.php?flag=1&subject_no=2P075

[50]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[51]D. G. Fedorov and K. Kitaura, Chem. Phys. Lett. **433**, 182 (2006).

[52]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, J. Goddard, A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic Structure Theory, Oxford University Press, New York, 1994.

[53]See supplementary material at

ftp : //ftp.aip.org/epaps/journ_chem_phys/E − JCPSA6 − 137 − 002229/ for a complete derivation of the analytical gradients of internal fragment energies, ESP, and Fock matrix elements as well as the Cartesian coordinates of all systems.

[54]C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, Theor. Chem. Acc. **110**, 233 (2003).

[55]N. C. Handy, and H. F. Schaefer III, J. Chem. Phys. **81**, 5031 (1984).

[56]C. Ochsenfeld, and M. S. Gordon. Chem. Phys. Lett. **270**, 399 (1997).

[57]N. W. Schmidt, K. K. Baldridge, J. A. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen , S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[58]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **2**5, 872 (2004).

[59]J. Clayden, N. Greeves, S. Warren, and P. Wothers, Organic Chemistry (Oxford university press, New York 2001)

[60]J. E. Miller, C. Gradinaru, B. R. Crane, A. J. D. Bilio, W. A. Wehbi, S. Un, J. R. Winkler, and H. B. Gray J. Am. Chem. Soc. **125**, 14220 (2003).

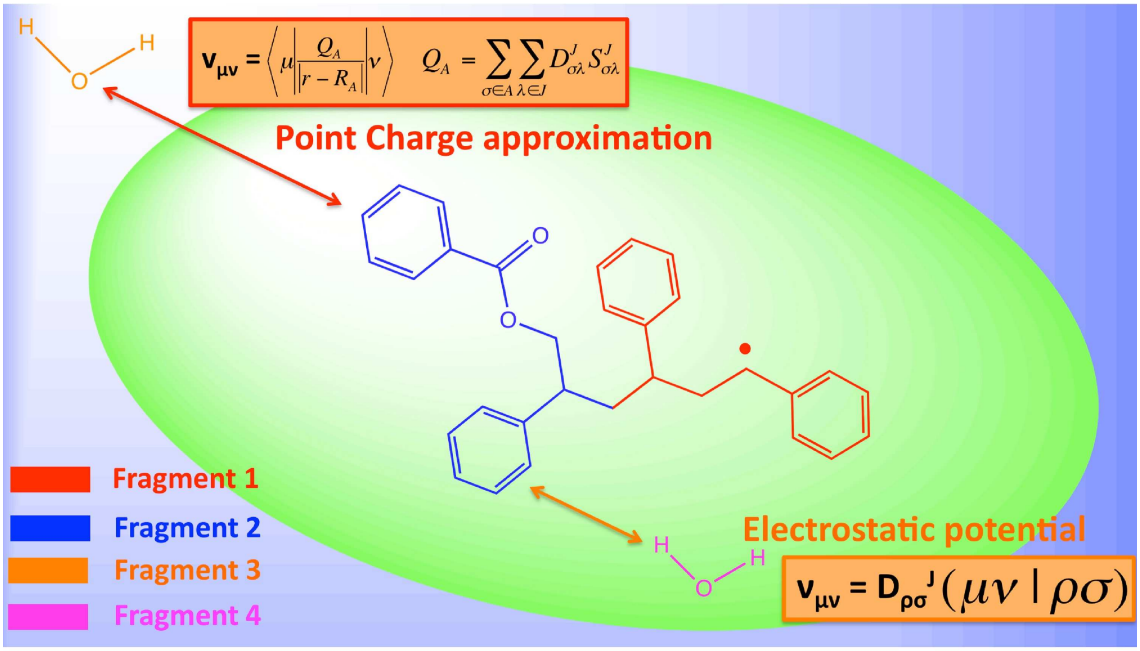
[61]J. Wang, P. Cieplak, and P.A. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[62]D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2010), AMBER 11, University of California, San Francisco.

[63]B. Morosin and J. R. Brathovde, Acta Cryst. **17**, 705 (1964).

Chapter III

Derivatives of the approximated electrostatic potentials in unrestricted Hartree-Fock based on the fragment molecular orbital method and an application to polymer radicals



H. Nakata, D. G. Fedorov, S. Yokojima *et. al.*

Theor. Chem. Acc. (2014) 133: 1477.

DOI : 10.1007/s00214-014-1477-6

View online: http://link.springer.com/article/10.1007/s00214-014-1477-6

III. DERIVATIVES OF THE APPROXIMATED ELECTROSTATIC POTENTIALS IN UNRESTRICTED HARTREE-FOCK BASED ON THE FRAGMENT MOLECULAR ORBITAL METHOD AND AN APPLICATION TO POLYMER RADICALS

A. Introduction

An accurate prediction of physical properties is the key to understanding physical phenomena. Quantum chemistry can describe a wide range of processes such as chemical reactions, however, it is difficult to apply traditional quantum-chemical approaches to large systems because the computational time and resources increase steeply with system size. The need to do accurate calculations of large systems is the strong motivation to develop linear-scaling algorithms[1–5] and fragment-based methods.[6–25]

In the latter category one does quantum-mechanical (QM) calculations of fragments, into which a large system is divided, and the total properties are computed from the values of subsystems. There is a considerable variation in the details of the choice of fragments, treatment of fragment boundaries, electrostatic embedding and the total property evaluation.[6] Some examples of fragment based methods are: effective fragment potential (EFP),[16] mutually consistent field (MCF),[7] explicit polarization (X-Pol) potential,[9,10] molecular fragmentation with conjugated caps (MFCC),[13] the electrostatically embedded many-body method (EEMB),[25] systematic molecular fragmentation (SMF),[20] the kernel energy method (KEM),[21] the polarizable multipole interaction with supermolecular pairs (PMISP),[22] the molecular tailoring approach (MTA),[23] the elongation method,[11,12] and divide-and-conquer approaches (DC).[8,17–19]

Another example is the fragment molecular orbital (FMO) method,[26] in which one calculates each fragment in the presence of the embedding electrostatic potential (ESP) due to the remaining fragments. Once the fragment electronic states are self-consistent with respect to ESP, fragment pair calculations are performed.[27–29] The effective fragment molecular method (EFMO),[30] is similar to FMO and differs mainly in using a polarizability model instead of ESP. FMO has been interfaced with many kinds of wave functions for closed[31–33] and open

shells,[34–38] and applied to a variety of chemical systems: silicon nanowires,[39] protein-ligand complexes,[40,41] DNA,[42] ionic liquids,[43] organic pigments,[44] boron nitride nanorings,[45] silica nanoparticles[46] and surfaces.[47] Properties such as chemical shifts[48–50] can also be calculated by FMO. An important by-product of fragment-based calculations is the pair interaction energies, which can be decomposed into physically meaningful components.[51–55]

Several approximations[56] have been proposed in FMO to improve efficiency. Instead of the charge density distribution, one can use point charges (PC) to represent ESP of far separated fragments in the ESP-PC approximation. Quantum effects are obviously important for the interaction between fragments in a neighborhood. However, it decreases as the distance increases. Therefore, one can replace QM calculations for dimers (DIM), which are far separated, by the electrostatic (ES) interaction on between the two charge distributions. The analytic gradient for FMO has been recently developed,[38,57,58] but the gradient for the ESP-PC[59,60] and ES-DIM[61] approximations has only been developed for closed shells. FMO can be used to perform geometry optimizations,[61,62] MD simulations,[63–67] or vibrational analysis.[68]

Polymers are very important in industrial materials. Thus, there is a lot of scientific problems for theoretical modelling of these systems. In order to account for realistic polymer sizes and times scale of dynamic processes, multiscale approaches are attractive,[69–71] such as coarse-grained methods on the mesoscale,[72] fine-grain atomistic molecular dynamics (MD) simulations on the nanoscale,[72] and QM calculations in the cluster modeling of the macroscopic systems.[73–79] Periodic boundary conditions[75–77] may be used to model periodic polymers with various wave functions.[78] The elongation method[11,12] has been applied to polymers and the hybrid QM/QM approach has been used to evaluate kinetic constants for free radical polymerization of polystyrene.[79]

In this study, we develop the energy and its analytic gradient for FMO based on unrestricted Hartree-Fock (UHF) with the point charge approximation for ESP (ESP-PC) and the separated dimer approximation (ES-DIM). As pointed out by Xie et al.[80] and also in other related studies,[81,82] we note that the use of point charges for a self-consistent embedding potential in a fully variational way requires an appropriate modification of the Fock matrices. The consideration of variational treatment makes the expressions for molecular

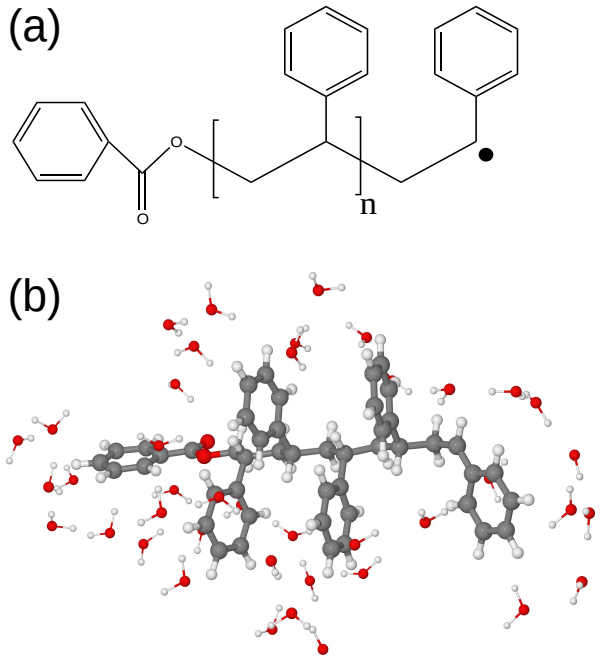


FIG. 1. Structures of decamer polystyrene radical (a) in vacuum and (b) solvated in a 3 Å water shell.

properties much easier derive and implement. However, as found out by Nagata et al.,[60] when charge distributions are used for close fragments and only far separated fragments are represented by point charges, then the numerical error due to the loss of a fully variational treatment of the embedding potentials is essentially zero (provided that the threshold is set to an appropriately large value).

First, we demonstrate by comparison to numerical gradients that it is also valid for open shells, while Nagata et al.[60] have shown it for closed shells. Secondly, we apply FMO-UHF to polymer radicals. First we study the accuracy of fragmentations and then optimize the reactant and product geometries, and calculate the reaction heat for a living free radical reaction. To improve the efficiency of geometry optimizations, we interfaced FMO with the delocalized internal coordinates.[83]

B. Mathematical formulation

1. *Energy and gradient in FMO-UHF*

In FMO, one divides a molecular system into fragments and performs QM calculations of fragments with the inclusion of the embedding ESP determined by the electronic state of all fragments. In general, several fragments may be described by UHF, and the rest by restricted Hartree-Fock (RHF), although in the current practical implementation only one fragment is of UHF type. After fragment (monomer) calculations converge with respect to ESP, fragment pair (dimer) calculations are performed once in the ESP determined at the monomer step. For a dimer, UHF is used only when at least one of the two fragments is described by UHF, otherwise RHF is employed. More details of FMO-UHF with the exact ESP based on charge distributions can be found elsewhere.[38] In this work we use the hybrid orbital projection (HOP) treatment of fragment boundaries.[84]

The energy equation in the two-body FMO (FMO2) based UHF is

$$
\begin{aligned}
E = & \sum_{I \in \mathrm{RHF}}^{N_{\mathrm{RHF}}} E'_I + \sum_{K \in \mathrm{UHF}}^{N_{\mathrm{UHF}}} E'_K + \sum_{\substack{I>J \\ I,J \in \mathrm{RHF} \\ R_{IJ} < R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} \left[E'_{IJ} - E'_I - E'_J + \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \right] \\
& + \sum_{\substack{K>L \\ K,L \in \mathrm{UHF} \\ R_{KL} < R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{UHF}}} \left[E'_{KL} - E'_K - E'_L + \mathrm{Tr}(\Delta \mathbf{D}^{KL} \mathbf{V}^{KL}) \right] \\
& + \sum_{K \in \mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{\substack{I \in \mathrm{RHF} \\ R_{KI} < R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} \left[E'_{IK} - E'_I - E'_K + \mathrm{Tr}(\Delta \mathbf{D}^{IK} \mathbf{V}^{IK}) \right] + \Delta E_{\mathrm{ES}}. \qquad (1)
\end{aligned}
$$

N_{RHF} and N_{UHF} are the numbers of RHF or UHF fragments, respectively. We use I, J to denote RHF fragments, and K, L to denote UHF fragments. R_{IJ} is the interfragment distance, i.e. the distance between two closest atoms. E'_X is the internal energy of monomers ($X = I$) or dimers ($X = IJ$). $\mathbf{V}^{IJ}$ is the ESP used for dimer IJ. ΔE_{ES} stands for the sum of all separated dimer energies, defined later. $\Delta \mathbf{D}^{IJ}$ is the density transfer matrix, $\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^I \oplus \mathbf{D}^J)$, where $\mathbf{D}^X$ is the density matrix of X. The electron densities $D_{\mu\nu}$ without a spin label correspond to the sum over the densities for spin α and β, i.e.,

$D^{\alpha}_{\mu\nu} + D^{\beta}_{\mu\nu}$.

The internal fragment energy E'_I for RHF is

$$E'_I = \sum_{\mu,\nu\in I} D^I_{\mu\nu} h^I_{\mu\nu} + \frac{1}{2} \sum_{\mu,\nu,\lambda,\rho\in I} \left[D^I_{\mu\nu} D^I_{\lambda\rho} - \frac{1}{2} D^I_{\mu\lambda} D^I_{\nu\rho} \right] (\mu\nu|\lambda\rho) + E^{\mathrm{NR}}_I, \tag{2}$$

while E'_K for UHF is

$$E'_K = \sum_{\mu,\nu\in K} D^K_{\mu\nu} h^K_{\mu\nu} + \frac{1}{2} \sum_{\mu,\nu,\lambda,\rho\in K} \Bigg[D^K_{\mu\nu} D^K_{\lambda\rho} - \left(D^{\alpha,K}_{\mu\lambda} D^{\alpha,K}_{\nu\rho} + D^{\beta,K}_{\mu\lambda} D^{\beta,K}_{\nu\rho} \right) \Bigg] (\mu\nu|\lambda\rho) + E^{\mathrm{NR}}_K, \tag{3}$$

where $(\mu\nu|\lambda\rho)$ is a two-electron Coulomb integral. We use Greek indices $(\mu, \nu, \lambda, \rho)$ to denote the atomic basis function throughout this study, except that σ and τ label spin. E^{NR}_I is the nuclear repulsion energy of fragment I. $h^K_{\mu\nu}$ is the sum of one-electron integrals including HOP.[84]

The embedding electrostatic potential $V^X_{\mu\nu}$ has one-electron u and two-electron v contributions,

$$V^X_{\mu\nu} = \sum_{\substack{I\neq X \\ I\in \mathrm{RHF}}}^{N_{\mathrm{RHF}}} (u^{X(I)}_{\mu\nu} + v^{X(I)}_{\mu\nu}) + \sum_{\substack{K\neq X \\ K\in \mathrm{UHF}}}^{N_{\mathrm{UHF}}} (u^{X(K)}_{\mu\nu} + v^{X(K)}_{\mu\nu}), \tag{4}$$

where

$$u^{X(I)}_{\mu\nu} = \sum_{A\in I} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle, \tag{5}$$

$$v^{X(I)}_{\mu\nu} = \sum_{\lambda,\rho\in I} D^I_{\lambda\rho} (\mu\nu|\lambda\rho). \tag{6}$$

The electron and nuclear positions are $\mathbf{r}$ and $\mathbf{R}_A$, respectively; A labels atoms, and Z_A is the atomic nuclear charge.

The Mulliken point charge (ESP-PC) approximation is applied to the embedding ESP contributions for fragments separated by $R_{\mathrm{ESP-PC}}$ or longer, otherwise, the exact ESP is used. The computational time is reduced drastically, because two-electron integrals are

replaced by one-electron integrals:[60]

$$v_{\mu\nu}^{X(I)} = \sum_{\lambda,\rho\in I} D_{\lambda\rho}^{I}(\mu\nu|\lambda\rho) \approx \sum_{A\in I} \left\langle \mu \left| \frac{Q_A}{|\mathbf{r}-\mathbf{R}_A|} \right| \nu \right\rangle, \tag{7}$$

where Q_A is the Mulliken atomic population of atom A in fragment I

$$Q_A = \sum_{\rho\in A} \sum_{\lambda\in I} D_{\rho\lambda}^{I} S_{\lambda\rho}^{I}, \tag{8}$$

$S_{\lambda\rho}^{I}$ is the overlap matrix element of fragment I.

In the electrostatic dimer (ES-DIM) approximation, if the distance between two monomers is longer than a threshold $R_{\mathrm{ES-DIM}}$, then the dimer energy is approximated as the electrostatic interaction between two monomers. The energy $\Delta E_{IJ}^{\mathrm{ES}}$ for a far separated dimer IJ is

$$\begin{aligned} \Delta E_{IJ}^{\mathrm{ES}} = & \sum_{\mu,\nu\in I} D_{\mu\nu}^{I} u_{\mu\nu}^{I(J)} + \sum_{\mu,\nu\in J} D_{\mu\nu}^{J} u_{\mu\nu}^{J(I)} \\ & + \sum_{\mu,\nu\in I} \sum_{\lambda,\rho\in J} D_{\mu\nu}^{I} D_{\lambda\rho}^{J} (\mu\nu|\lambda\rho) + \Delta E_{IJ}^{\mathrm{NR}}, \end{aligned} \tag{9}$$

where $\Delta E_{IJ}^{\mathrm{NR}} = E_{IJ}^{\mathrm{NR}} - E_{I}^{\mathrm{NR}} - E_{J}^{\mathrm{NR}}$. ΔE_{ES} is the sum of all contributions (Eq. (9)) for dimers in which the two monomers are separated by more than $R_{\mathrm{ES-DIM}}$. We do not use point charge approximations in Eq. (9). Eqs. (5), (6), (7), (8), and 9 are also applicable to UHF fragments because these equations are formulated for the total density summed over spin. For the $R_{\mathrm{ES-DIM}}$ and $R_{\mathrm{ESP-PC}}$, we use 2.5 throughout this study. While we sometimes use 1.5 to perform the geometry optimization of large molecule within the reasonable time.

Next, we take an analytic derivative of the FMO energy in Eq. (1) with respect to an atomic coordinate a. The expression has three groups of contributions: (a) internal energy E' derivatives, (b) derivatives of the charge transfer coupling with ESP (the trace terms), and (c) derivatives of separated dimers (the last term).

The second term (b) has the following form:

$$\frac{\partial \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a} = \sum_{\mu,\nu\in IJ} \left(\frac{\partial \Delta D_{\mu\nu}^{IJ}}{\partial a} V_{\mu\nu}^{IJ} + \frac{\partial V_{\mu\nu}^{IJ}}{\partial a} \Delta D_{\mu\nu}^{IJ} \right). \tag{10}$$

The internal energy E' derivatives and the density derivative term in Eq. (10) are defined elsewhere[38] and for completeness we give them in SI. In the next subsection we focus on

the remaining terms, the ESP derivatives in Eq. (10) and the derivatives of the separated dimers, which are new development of this work.

2. Derivative of ESP and ES dimers

The derivative of ESP and ES dimers for RHF fragments is given by Nagata et al.,[60] and we reformulate it for UHF. More details of these derivations are given in the supporting information. The derivative of the approximated ESP of fragment X due to UHF fragment K is

$$\begin{aligned}\frac{\partial v_{\mu\nu}^{X(K)}}{\partial a} &\approx \sum_{A\in K}\frac{\partial}{\partial a}\left\langle\mu\left|\frac{Q_A}{|\mathbf{r}-\mathbf{R}_A|}\right|\nu\right\rangle \\ &= \sum_{A\in K}\left[Q_A\left\langle\frac{\partial\mu}{\partial a}\left|\frac{1}{|\mathbf{r}-\mathbf{R}_A|}\right|\nu\right\rangle + Q_A\left\langle\mu\left|\frac{1}{|\mathbf{r}-\mathbf{R}_A|}\right|\frac{\partial\nu}{\partial a}\right\rangle\right. \\ &\left. +Q_A\left\langle\mu\left|\frac{\partial}{\partial a}\left(\frac{1}{|\mathbf{r}-\mathbf{R}_A|}\right)\right|\nu\right\rangle + \frac{\partial Q_A}{\partial a}\left\langle\mu\left|\frac{1}{|\mathbf{r}-\mathbf{R}_A|}\right|\nu\right\rangle\right]. \end{aligned} \tag{11}$$

The first three terms are the derivative of one-electron integrals. The last term is the Mulliken charge derivative contribution and it can be expanded as:

$$\begin{aligned}\frac{\partial Q_A}{\partial a} = \sum_{\eta\in A}\sum_{\lambda\in K}\Bigg[&-\sum_{\rho\in K}\sum_{\gamma\in K}D_{\gamma\lambda}^{\sigma,K}S_{\lambda\eta}^{K}D_{\eta\rho}^{\sigma,K}S_{\rho\gamma}^{a,K} \\ &+\sum_{i^\sigma\in K}^{\text{occ}}\sum_{m^\sigma\in K}^{\text{vir}}U_{m^\sigma i^\sigma}^{a,K}(C_{m^\sigma\lambda}^{K}C_{i^\sigma\eta}^{K}+C_{m^\sigma\eta}^{K}C_{i^\sigma\lambda}^{K})S_{\lambda\eta}^{K} \\ &+S_{\lambda\eta}^{a,K}D_{\eta\lambda}^{K}\Bigg], \end{aligned} \tag{12}$$

where i and m denote occupied and virtual molecular orbitals, respectively, expanded in terms of atomic orbitals with the coefficients $C_{i\mu}$. σ labels spin (α or β). $S_{\lambda\eta}^{a,K}$ are the derivatives of the overlap integrals. $U_{m^\sigma i^\sigma}^{a,K}$ are the orbital responses, obtained from coupled-perturbed Hartree-Fock (CPHF) equations.[85]

The derivative of the ES dimer energy for a dimer composed of an RHF (I) and UHF

(K) fragment is

$$
\begin{aligned}
E_{KI}^{\mathrm{ES},a} = & 2\sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{m^\sigma\in K}^{\mathrm{vir}} U_{m^\sigma i^\sigma}^{a,K} V_{m^\sigma i^\sigma}^{K(I)} \\
& - \sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{m^\sigma\in K}^{\mathrm{occ}} S_{m^\sigma i^\sigma}^{a,K} V_{m^\sigma i^\sigma}^{K(I)} + 4\sum_{i\in I}^{\mathrm{occ}}\sum_{m\in I}^{\mathrm{vir}} U_{mi}^{a,I} V_{mi}^{I(K)} \\
& - 2\sum_{i\in I}^{\mathrm{occ}}\sum_{m\in I}^{\mathrm{occ}} S_{mi}^{a,I} V_{mi}^{I(K)} + \sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}} V_{i^\sigma i^\sigma}^{a,K(I)} \\
& + 2\sum_{i\in I}^{\mathrm{occ}} V_{ii}^{a,I(K)} - 2\sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{j\in I}^{\mathrm{occ}} (i^\sigma i^\sigma|jj)^a,
\end{aligned} \tag{13}
$$

and for two UHF fragments K and L

$$
\begin{aligned}
E_{KL}^{\mathrm{ES},a} = & 2\sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{m^\sigma\in K}^{\mathrm{vir}} U_{m^\sigma i^\sigma}^{a,K} V_{m^\sigma i^\sigma}^{K(L)} - \sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{m^\sigma\in K}^{\mathrm{occ}} S_{m^\sigma i^\sigma}^{a,K} V_{m^\sigma i^\sigma}^{K(L)} \\
& + 2\sum_{\sigma}\sum_{i^\sigma\in L}^{\mathrm{occ}}\sum_{m^\sigma\in L}^{\mathrm{vir}} U_{m^\sigma i^\sigma}^{a,L} V_{m^\sigma i^\sigma}^{L(K)} - \sum_{\sigma}\sum_{i^\sigma\in L}^{\mathrm{occ}}\sum_{m^\sigma\in L}^{\mathrm{occ}} S_{m^\sigma i^\sigma}^{a,L} V_{m^\sigma i^\sigma}^{L(K)} \\
& + \sum_{\sigma}\sum_{i^\sigma\in K}^{\mathrm{occ}} V_{i^\sigma i^\sigma}^{a,K(L)} + \sum_{\sigma}\sum_{i^\sigma\in L}^{\mathrm{occ}} V_{i^\sigma i^\sigma}^{a,L(K)} \\
& - \sum_{\sigma,\tau}\sum_{i^\sigma\in K}^{\mathrm{occ}}\sum_{j^\tau\in L}^{\mathrm{occ}} (i^\sigma i^\sigma|j^\tau j^\tau)^a,
\end{aligned} \tag{14}
$$

where τ labels spin, and the partial contributions to the ESP of fragment K due to UHF fragment L and RHF fragment I are

$$
V_{i^\sigma i^\sigma}^{K(L)} = \sum_{A\in L}\left\langle i^\sigma \left| \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|} \right| i^\sigma \right\rangle + \sum_{\tau}\sum_{j^\tau\in L}^{\mathrm{occ}} (i^\sigma i^\sigma|j^\tau j^\tau), \tag{15}
$$

$$
V_{i^\sigma i^\sigma}^{K(I)} = \sum_{A\in I}\left\langle i^\sigma \left| \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|} \right| i^\sigma \right\rangle + 2\sum_{j\in I}^{\mathrm{occ}} (i^\sigma i^\sigma|jj), \tag{16}
$$

respectively. In Eqs. (13) and (14), the overlap derivative in the MO basis is

$$
S_{ij}^{a,X} = \sum_{\mu,\nu\in X} C_{\mu i}^{X*} \frac{\partial S_{\mu\nu}^{X}}{\partial a} C_{\nu j}^{X}. \tag{17}
$$

Inserting the terms in Eq. (10), (11), (12), 13, and 14 into the total energy derivative E^a, one can obtain the analytic energy gradient. In the compact form, it can be written as

$$
\frac{\partial E}{\partial a} = \frac{\partial E'}{\partial a} + \overline{U}^a + R^a, \tag{18}
$$

where the first term is the derivative of the total internal energy (the sum of all E' terms in Eq. (1)), the second is the ESP-related response term arising from the differences in the approximations of ESPs in monomers and dimers. The last term comes from the ESP coupled to the density difference matrix.

$\overline{U}^a$ in UHF is calculated from the occupied-virtual orbital responses $U^{a,I}$, similar to RHF.[60]

$$\begin{aligned}
\overline{U}^a = &-4\sum_{I}^{N_{\mathrm{RHF}}}\sum_{r}^{\mathrm{vir}}\sum_{i}^{\mathrm{occ}} U_{ri}^{a,I}V_{ri}^{I} - 2\sum_{\sigma}\sum_{K}^{N_{\mathrm{UHF}}}\sum_{r^\sigma}^{\mathrm{vir}}\sum_{i^\sigma}^{\mathrm{occ}} U_{r^\sigma i^\sigma}^{a,K}V_{r^\sigma i^\sigma}^{K} \\
&+ 4\sum_{\substack{I>J \\ R_{IJ}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} \Big[\sum_{r\in I}^{\mathrm{vir}}\sum_{i\in I}^{\mathrm{occ}} U_{ri}^{a,I}(V_{ri}^{I} - V_{ri}^{IJ}) \\
&+ \sum_{r\in J}^{\mathrm{vir}}\sum_{i\in J}^{\mathrm{occ}} U_{ri}^{a,J}(V_{ri}^{J} - V_{ri}^{IJ})\Big] + 2\sum_{\substack{K>L \\ R_{KL}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{UHF}}}\sum_{\sigma}\Big[\sum_{r^\sigma\in K}^{\mathrm{vir}}\sum_{i^\sigma\in K}^{\mathrm{occ}} U_{r^\sigma i^\sigma}^{a,K} \\
&\times (V_{r^\sigma i^\sigma}^{K} - V_{r^\sigma i^\sigma}^{KL}) + \sum_{r^\sigma\in L}^{\mathrm{vir}}\sum_{i^\sigma\in L}^{\mathrm{occ}} U_{r^\sigma i^\sigma}^{a,L}(V_{r^\sigma i^\sigma}^{L} - V_{r^\sigma i^\sigma}^{KL})\Big] \\
&+ 2\sum_{K}^{N_{\mathrm{UHF}}}\sum_{\substack{I \\ R_{KI}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} \Big[\sum_{\sigma}\sum_{r^\sigma\in K}^{\mathrm{vir}}\sum_{i^\sigma\in K}^{\mathrm{occ}} U_{r^\sigma i^\sigma}^{a,K}(V_{r^\sigma i^\sigma}^{K} - V_{r^\sigma i^\sigma}^{KI}) \\
&+ 2\sum_{r\in I}^{\mathrm{vir}}\sum_{i\in I}^{\mathrm{occ}} U_{ri}^{a,I}(V_{ri}^{I} - V_{ri}^{KI})\Big]. \qquad (19)
\end{aligned}$$

$\overline{U}^a$ can be represented by the scalar product ($\mathbf{L}^T\mathbf{U}$) of two super vectors with components corresponding to the virtual r and occupied i molecular orbitals,

$$\overline{U}^a = \sum_{I\in\mathrm{RHF}}^{N_{\mathrm{RHF}}}\sum_{r}^{\mathrm{vir}}\sum_{i}^{\mathrm{occ}} L_{ri}^{I,U}U_{ri}^{a,I} + \sum_{K\in\mathrm{UHF}}^{N_{\mathrm{UHF}}}\sum_{\sigma}\sum_{r^\sigma}^{\mathrm{vir}}\sum_{i^\sigma}^{\mathrm{occ}} L_{r^\sigma i^\sigma}^{K,U}U_{r^\sigma i^\sigma}^{a,K}, \qquad (20)$$

where the Lagrangian for RHF fragment I is

$$\begin{aligned}
L_{ri}^{I,U} = &-4V_{ri}^{I} + 4\sum_{\substack{J\neq I \\ R_{IJ}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} (V_{ri}^{I} - V_{ri}^{IJ}) \\
&+ 4\sum_{\substack{K\neq I \\ R_{KI}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{UHF}}} (V_{ri}^{I} - V_{ri}^{KI}) \qquad (21)
\end{aligned}$$

and for UHF fragment K

$$L^{K,U}_{r^\sigma i^\sigma} = -2V^K_{r^\sigma i^\sigma} + 2 \sum_{\substack{L\neq K \\ R_{KL}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{UHF}}} (V^K_{r^\sigma i^\sigma} - V^{KL}_{r^\sigma i^\sigma}) + 2 \sum_{\substack{I\neq K \\ R_{KI}<R_{\mathrm{ES-DIM}}}}^{N_{\mathrm{RHF}}} (V^K_{r^\sigma i^\sigma} - V^{KI}_{r^\sigma i^\sigma}). \quad (22)$$

It is convenient to transform the Lagrangians and R^a from the molecular to atomic orbital basis, following Nagata et al.[60] In addition to the internal energy related Lagrangians in Eqs. (21) and (22), we define three types of Lagrangians arising from the derivatives of the exact ESP (Eq. (10)) point charge ESP (Eq. (12)) and ES dimers (Eqs. (13) and (14)), denoted by ESP, ESP-PC and ES-DIM, respectively. The explicit formulation is shown in the supporting information.

The total (tot) Lagrangians for RHF fragment I and UHF fragment K are

$$L^{I,\mathrm{tot}}_{ri} = L^{I,U}_{ri} + L^{I,\mathrm{ESP}}_{ri} + L^{I,\mathrm{ESP-PC}}_{ri} + L^{I,\mathrm{ES-DIM}}_{ri} \quad (23)$$

$$L^{K,\mathrm{tot}}_{r^\sigma i^\sigma} = L^{K,U}_{r^\sigma i^\sigma} + L^{K,\mathrm{ESP}}_{r^\sigma i^\sigma} + L^{K,\mathrm{ESP-PC}}_{r^\sigma i^\sigma} + L^{K,\mathrm{ES-DIM}}_{r^\sigma i^\sigma}. \quad (24)$$

The contribution of the response terms to the total energy gradient is

$$\overline{U}^a + R^a = \sum_{I\in\mathrm{RHF}}^{N_{\mathrm{RHF}}} \sum_{r}^{\mathrm{vir}} \sum_{i}^{\mathrm{occ}} L^{I,\mathrm{tot}}_{ri} U^{a,I}_{ri} + \sum_{K\in\mathrm{UHF}}^{N_{\mathrm{UHF}}} \sum_{\sigma} \sum_{r^\sigma}^{\mathrm{vir}} \sum_{i^\sigma}^{\mathrm{occ}} L^{K,\mathrm{tot}}_{r^\sigma i^\sigma} U^{a,K}_{r^\sigma i^\sigma}. \quad (25)$$

The scalar product $(\mathbf{L}^{\mathrm{tot}})^T\mathbf{U}$ is obtained by solving the CPHF equations as discussed next.

3. CPHF equation for approximated ESP

The set of CPHF equations in FMO-UHF involves all fragments (both RHF and UHF),

$$\mathbf{A}\mathbf{U}^a = \mathbf{B}^a_0. \quad (26)$$

The RHF definitions can be found in Ref.[58] and we focus on UHF. $\mathbf{A}$ in Eq. (26) is the orbital Hessian matrix for the whole system. The diagonal block for UHF fragment K and

spin σ is

$$\begin{aligned} A^{K,K}_{i^\sigma j^\sigma, r^\sigma k^\sigma} =& \delta_{i^\sigma r^\sigma}\delta_{j^\sigma k^\sigma}(\epsilon^K_{j^\sigma} - \epsilon^K_{i^\sigma}) - [2(i^\sigma j^\sigma|r^\sigma k^\sigma) \\ & - (i^\sigma r^\sigma|j^\sigma k^\sigma) - (i^\sigma k^\sigma|j^\sigma r^\sigma)], \end{aligned} \tag{27}$$

where ϵ_i is orbital energy, and for $\tau \neq \sigma$,

$$A^{K,K}_{i^\sigma j^\sigma, r^\tau k^\tau} = -2(i^\sigma j^\sigma|r^\tau k^\tau). \tag{28}$$

The off-diagonal blocks for the exact ESP of RHF fragment I and UHF fragments K and L are

$$A^{K,I}_{i^\sigma j^\sigma, rk} = -4(i^\sigma j^\sigma|rk), \tag{29}$$

where $i, j \in K$ and $r, k \in I$.

$$A^{K,L}_{i^\sigma j^\sigma, r^\tau k^\tau} = -2(i^\sigma j^\sigma|r^\tau k^\tau), \tag{30}$$

where $i, j \in K$ and $r, k \in L$.

For the point charge ESP of RHF fragment I and UHF fragment L

$$A^{K,I}_{i^\sigma j^\sigma, rk} \approx -2\sum_{A\in I}\sum_{\lambda\in A}\sum_{\eta\in K}\left\langle i^\sigma \left| \frac{C^I_{\lambda k}C^I_{\eta r} + C^I_{\eta k}C^I_{\lambda r}}{|\mathbf{r} - \mathbf{R}_A|} \right| j^\sigma \right\rangle, \tag{31}$$

$$A^{K,L}_{i^\sigma j^\sigma, r^\tau k^\tau} \approx -\sum_{A\in L}\sum_{\lambda\in A}\sum_{\eta\in K}\left\langle i^\sigma \left| \frac{C^L_{\lambda k^\tau}C^L_{\eta r^\tau} + C^L_{\eta k^\tau}C^L_{\lambda r^\tau}}{|\mathbf{r} - \mathbf{R}_A|} \right| j^\sigma \right\rangle. \tag{32}$$

$\mathbf{B}^{a,K}_0$ is the derivative integral term,

$$\begin{aligned} B^{a,K}_{0,i^\sigma j^\sigma} =& F^{a,K}_{i^\sigma j^\sigma} - S^{a,K}_{i^\sigma j^\sigma}\epsilon^K_{j^\sigma} - \sum_{k^\tau, l^\tau \in K}^{\text{occ}} S^{a,K}_{k^\tau l^\tau}(i^\sigma j^\sigma|k^\tau l^\tau) \\ & - \sum_{k^\sigma, l^\sigma \in K}^{\text{occ}} S^{a,K}_{k^\sigma l^\sigma}\left[(i^\sigma j^\sigma|k^\sigma l^\sigma) - (i^\sigma k^\sigma|j^\sigma l^\sigma)\right] \\ & + \frac{1}{2}\sum_{\substack{L\neq K \\ L\in \text{UHF}}}^{N_{\text{UHF}}}\sum_{\tau}\sum_{k^\tau, l^\tau \in L}^{\text{occ}} S^{a,L}_{k^\tau l^\tau}A^{K,L}_{i^\sigma j^\sigma, k^\tau l^\tau} \\ & + \frac{1}{2}\sum_{\substack{I\neq K \\ I\in \text{RHF}}}^{N_{\text{RHF}}}\sum_{k,l\in I}^{\text{occ}} S^{a,I}_{kl}A^{K,I}_{i^\sigma j^\sigma, kl}, \end{aligned} \tag{33}$$

where F_{ij}^{a} is derivative of Fock matrix with respect to an atomic coordinate a.

The Z-vector equations[86,87] are solved using the self-consistent Z-vector method,[58]

$$\mathbf{AZ} = \mathbf{L}^{\mathrm{tot}}. \tag{34}$$

Finally, the response contribution to the gradient is obtained as

$$\overline{U}^{a} + R^{a} = \mathbf{Z}^{T}\mathbf{B}_{0}^{a}. \tag{35}$$

C. Computational Details

The accuracy of the analytic FMO-UHF gradient developed in this work is compared to numerical energy gradients obtained with double differencing with the coordinate shifts of 0.0005 Å. To elucidate the importance of the response terms in Eq. (18) clearly, we performed NVE MD simulations for 100 steps with the time steps of 0.1, 0.2, 0.5 and 1.0 fs, starting from a pre-equilibrated structure. The velocity Verlet algorithm was used for time integration throughout this work.

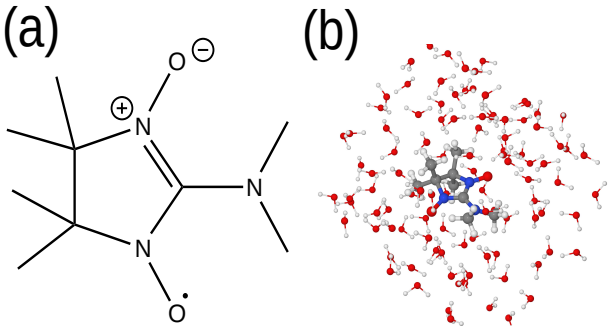


FIG. 2. Structures of dimethyl-amino-nitronyl-nitroxide (DMANN) radical (a) in vacuum and (b) solvated in a 8 Å water shell.

Several test systems were chosen. A dimethyl-amino-nitronyl-nitroxide (DMANN) radical[88] was solvated in an 8 Å box of water molecules (FIG. 2, the total number of atoms is 338). A polystyrene radical was solvated in a 3 Å shell of water molecules (FIG. 1, the total number of atoms is 217). An OH radical was solvated in several water boxes with the size from 10 to 30 Å (FIG. 3, the total number of atoms is 134, 401, 917, 1742, and

3068). Small clusters of solvated OH were also studied by Pruitt et al.[34,35] using restricted open-shell FMO.

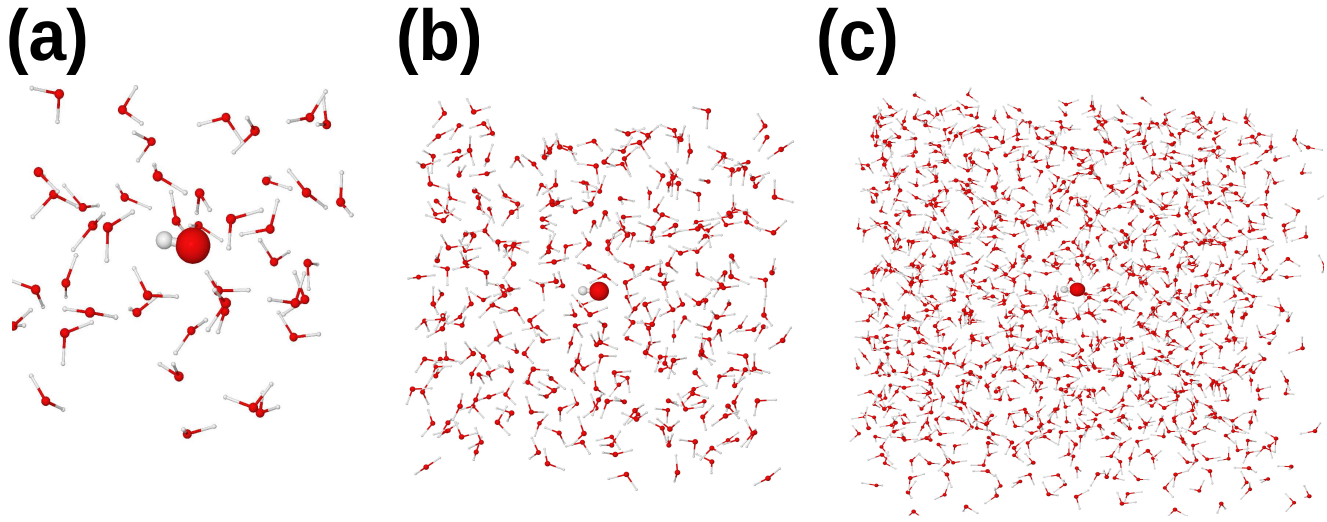


FIG. 3. OH radical solvated in a box of water with the linear size of (a) 10 Å, (b) 20 Å and (c) 30 Å.

To determine the initial structures, all molecules were optimized using QM and molecular mechanics (QM/MM) with NWChem 6.1.[89] The solute and solvent were treated with QM (B3LYP/6-31G(d)) and Amber force field,[90] respectively. For pre-equilibration in MD, we reoptimized the geometry using FMO-UHF and then carried out a short 100 fs NVT MD simulation at 300 K. The final geometry and velocities were used in the following NVE MD simulations.

FMO-UHF derivatives were implemented in a development version of GAMESS[91,92] and parallelized with the generalized distributed data interface (GDDI).[93] We used the 6-31G(d) basis set with spherical harmonics (ISPHER=1). In FMO, the ES dimer and point charge ESP approximations were applied with the thresholds of $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$, respectively. We note that both thresholds are unitless,[56] because they are applied to the interatomic distances normalized by the sum of the atomic van-der-Waals radii.

In order to merge 2 water molecules into 1 fragment, we performed a preliminary FMO calculation with the fragmentation of 1 water molecule per fragment. Then we selected the pair with the largest pair interaction energy and merged it, then another pair among the

remaining fragments and so forth, until all water molecules were paired.

D. Results and discussion

1. Gradient accuracy

First, we compare the analytic energy gradient obtained with FMO-UHF (with the ESP-PC approximation and SCZV response terms) and *ab initio* UHF (TABLE I). The maximum deviation from *ab initio* are 0.000899 and 0.000039 hartree/bohr in DMANN and polystyrene, respectively. The difference of the rms gradients is 0.000202 and 0.000049 hartree/bohr in DMANN and polystyrene, respectively.

TABLE I. Maximum (MAX) and root mean square (rms) analytic energy gradients (hartree/bohr) in FMO2-UHF and *ab initio* UHF for 6-31G(d), using $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$.

	FMO-UHF, SCZV		*ab initio* UHF	
system	MAX	rms	MAX	rms
DMANN[a]	0.067930	0.025277	0.068829	0.025479
polystyrene[b]	0.068677	0.021708	0.068638	0.021757

[a] Dimethyl-amino-nitronyl-nitroxide (DMANN) solvated in 102 water molecules (divided into 51 fragments) FIG. 2. [b] Polystyrene (divided into 3 fragments) solvated in 45 water molecules (divided into 23 fragments), FIG. 1.

The consideration of the response terms coming from SCZV (Eq. (18)) is important to get an accurate gradient. This is clearly seen by looking at the difference between the analytic and numerical energy gradients, summarized in TABLE II. All gradient elements are dramatically improved, as shown in FIG. 4. The consideration of the unknown response terms decreased the maximum gradient error by two orders of magnitude, from 10^{-3} to 10^{-5} hartree/bohr in both polystyrene and DMANN (TABLE II). Note that 10^{-6} is comparable to numerical gradient error, because we used default convergence criteria for FMO-UHF (10^{-8} *a.u.* for SCF convergence and 10^{-7} *a.u.* for self consistent charge to determine the electrostatic embedding potential).

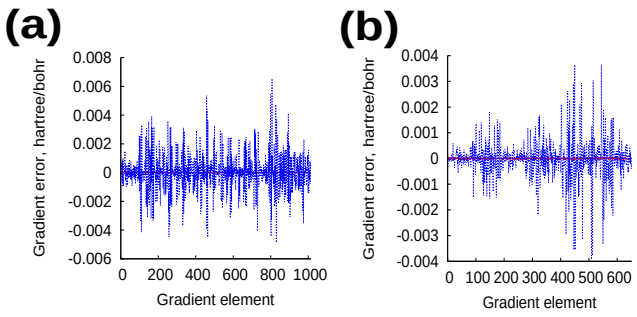


FIG. 4. Deviations between the analytic and numeric FMO-UHF/6-31G(d) gradients for the thresholds of $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$ are shown for the complete analytic gradient (red solid line) including the response terms (Eq. (18)) and an incomplete analytic gradient (blue dashed line) neglecting them. The gradient elements are plotted sequentially for all atoms in (a) DMANN solvated in 102 water molecules and (b) polystyrene decamer solvated in 45 water molecules.

2. Molecular dynamics

NVE MD simulations of DMANN solvated in 102 water molecules were performed using the FMO-UHF energy gradient with and without the response terms (Eq. (18)). The energy fluctuations for the time steps of 0.1 fs and 1.0 fs are shown in FIG. 5. They are larger when the response terms are ignored for both time steps. For the time step of 0.1 fs, the root-

TABLE II. Maximum (MAX) and root-mean-square deviations (rmsd) between the analytic and numerical FMO2-UHF gradients (hartree/bohr) for 6-31G(d), using $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$.

	FMO-UHF, SCZV[a]		FMO-UHF, no SCZV[b]	
system	MAX	rmsd	MAX	rmsd
DMANN[c]	-0.000039	0.000006	0.006536	0.001277
polystyrene[c]	-0.000051	0.000016	-0.003930	0.000800

[a] The rmsd with the response terms in Eq. (18), and [b] without them. [c] The same systems as in TABLE I.

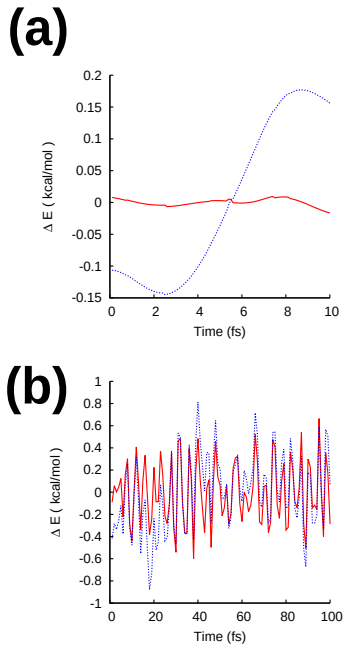


FIG. 5. Energy fluctuations (ΔE) of DMANN solvated in 102 water molecules using $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$ thresholds in NVE MD with FMO-UHF/6-31G(d), for the time steps of (a) 0.1 and (b) 1.0 fs. The data for the complete analytic gradient including the response terms in Eq. (18) are shown as a red solid line whereas the incomplete gradient data are drawn with a black dashed line.

mean-square error (RMSE) of the energy with and without the response terms are 0.0056 and 0.13 kcal/mol, respectively. The consideration of the response terms reduces the error by the factor of 22.

For the time step of 1.0 fs, even including the response terms results in a substantial error (RMSE), so that this time step is too large. A similar finding was reported by Gordon and coworkers for FMO-RHF based MD.[67] We did not use the RATTLE algorithm,[64,94] which is not implemented in GAMESS. RATTLE can increase the time step while maintaining the accuracy.

Next, the gradient accuracy is verified by looking at the energy fluctuation in MD as a function of the time step (FIG. 6). The slopes are 2.0 and 0.45 for MD performed with and

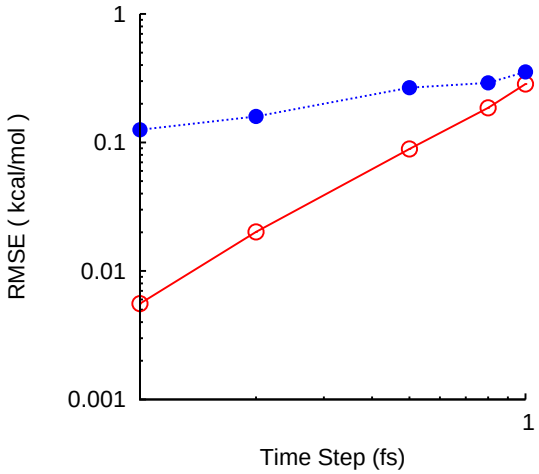


FIG. 6. Double logarithmic plots of the root-mean-square error (RMSE) of the total FMO-UHF/6-31G(d) energy for 100 steps in NVE MD, performed for DMANN solvated in 102 water molecules using $R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$ thresholds. The data for the complete analytic gradient including the response terms in Eq. (18) are shown as red open circles with a solid line whereas the incomplete gradient data are drawn with blue filled circles with a dashed line.

without the response terms, while the correct theoretical value should be 2 for the velocity Verlet analysis.[95] It is also clear from FIG. 6 that the approximate gradient (without the response terms) is not accurate even for small time steps. We conclude by suggesting to use the values of time steps of 0.1-0.3 fs and the accurate gradient with response terms, in

agreement with the FMO-RHF study.[67]

3. Application of FMO-UHF to organic polymer radicals

The earlier polymer-related FMO study was conducted only for a small radical system of the initiation step in the reversible addition-fragmentation transfer (RAFT) reaction by Pruitt et al.[34,35] using restricted open shell FMO, although it is too small to be a representative polymer model. We chose a living free radical reaction as a pilot test system for polymer radicals. The polymerization reaction rate depends on the choice of monomers, such as maleic anhydride and styrene.[96] To model a copolymer reactant consisting of the 1:2 ratio of maleic anhydride and styrene, we constructed the system with the repeating unit of two styrenes and one maleic anhydride with appropriate terminating group (the top part of FIG. 7). In the model of the polymerization reaction, we added one styrene or maleic anhydride to the above reactant, resulting in a larger polymer (the bottom of FIG. 7). The polymer reaction occurs between a 14-mer reactant and a monomer reactant, resulting in the 15-mer product.

First, we check the fragmentation accuracy for single point energy calculations. The test systems are a decamer of polystyrene, and a decamer copolymer of maleic anhydride and styrene, divided into fragments by assigning two monomer units per fragment as shown in FIG. 8. The error of FMO-UHF relative to *ab initio* UHF was 0.223 and 0.205 kcal/mol for the pure polystyrene and its copolymer with maleic anhydride, respectively. It is not possible to fragment these systems as 1 monomer unit per fragment, because in this case the bond detachment points are separated just by one other atom.

The same fragmentation of 2 monomer units per fragment was done for the reaction (FIG. 9). The polymer molecules are very flexible, and geometry optimization in Cartesian coordinates failed to converge after a few hundreds of steps. To overcome this, we switched to delocalized internal coordinates (DLC),[83] which proved to be efficient for these systems. We optimized the structures of the reactants and products, and calculated the reaction heat as the energy difference at the minima obtained with FMO and *ab initio*. In geometry optimizations we added the empiric dispersion[97] (D) correction to UHF (UHF/D), because

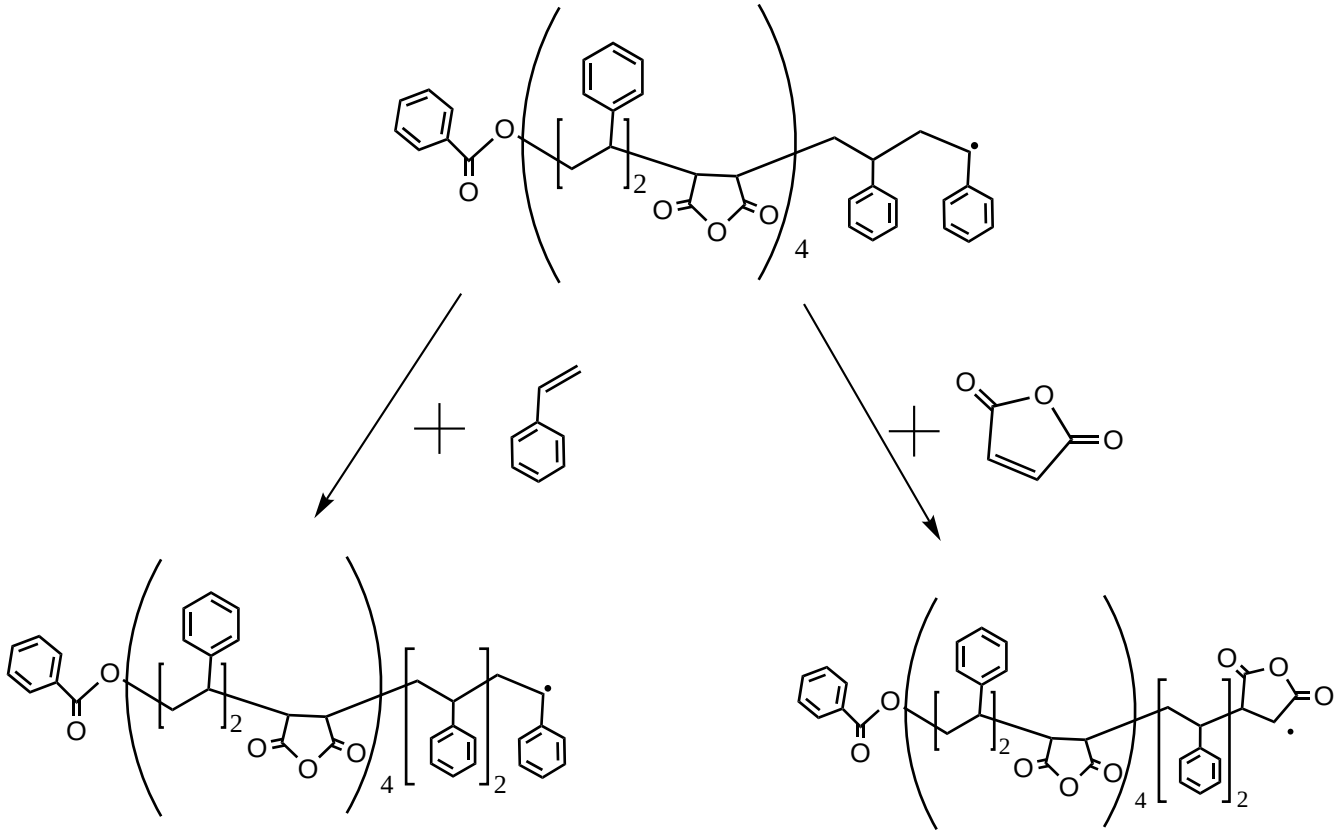


FIG. 7. Living free radical reaction. The copolymer reactant shown on the top reacts with the monomer reactant, styrene (left) or maleic anhydride (right) and the reaction products are shown at the bottom.

these molecules have aromatic rings and the dispersion plays an important role in their interactions.

The results are shown in TABLE III. It can be seen that the root-mean-square deviations (rmsd) between FMO-UHF and *ab initio* UHF optimized structures do not exceed 0.16 Å for all systems. The FMO errors in reproducing the reaction heat do not exceed 0.17 kcal/mol. The reaction is exothermicity(-29.91 and -15.84 for styrene and maleic anhydride). Therefore it is possible to react iteratively until the large polymer is constructed. Such potential to iteratively happen the living radical polymerization agrees with the experiments as it was observed,[96] although the thermodynamic quantities are not found in the experiments.

Finally, we discuss the pair interaction energies (PIE), shown in FIG. 9. The PIEs ΔE_{IJ}

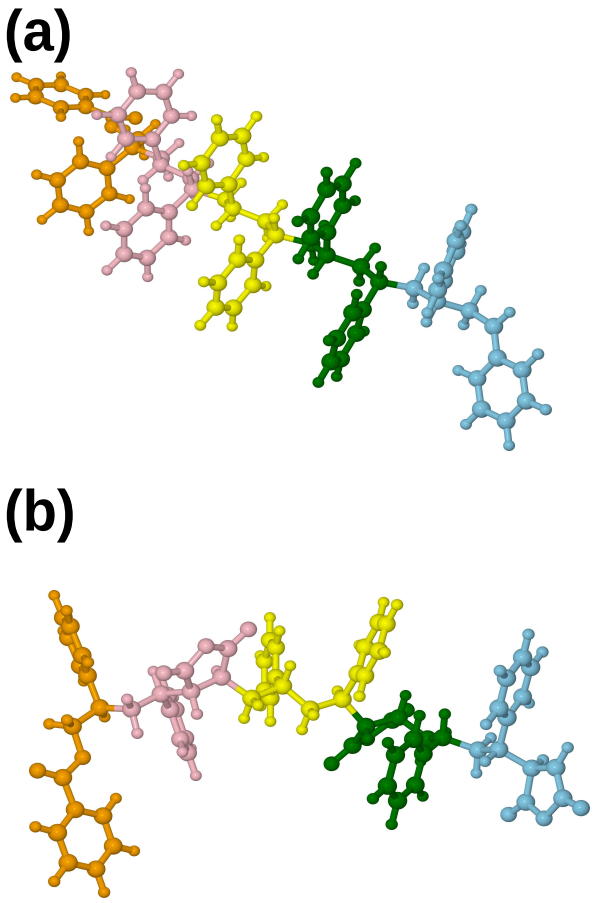


FIG. 8. Polymer fragmentation in FMO (fragments are shown in different colors): (a) polystyrene decamer, (b) decamer copolymer of styrene and maleic anhydride.

appear as dimer corrections in Eq. (1), i.e.,

$$\Delta E_{IJ} = E'_{IJ} - E'_I - E'_J + \mathrm{Tr}(\Delta \mathbf{D}^{IJ}\mathbf{V}^{IJ}) \tag{36}$$

The values of PIEs for covalently connected fragments are large (about -15 *a.u.*). This is because the bonds are fractioned at an atom (Called bond detached atom, BDA). Two electrons from the bond and one proton from the BDA are assigned to one fragment, and the other electrons and protons are in the other fragment. The bond detachment effect appeared in the $E'_{IJ} - E'_I - E'_J$ term, for which the BDA proton in other fragment attraction is neglected in the monomer, since external field is subtracted from E'_I[51,98]. In the following discussion, we exclude this term from the PIEs of connected fragment pairs.

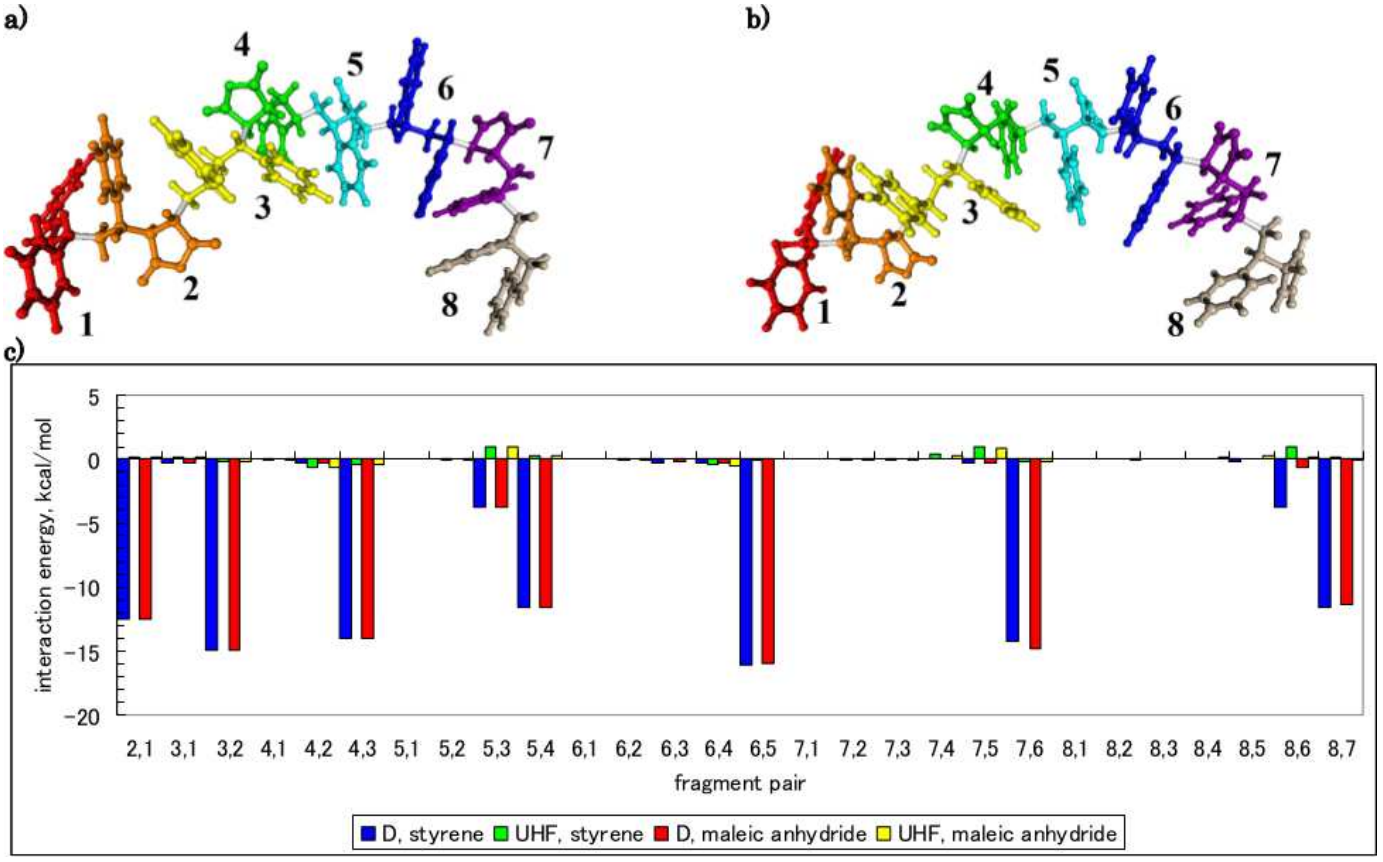


FIG. 9. Fragmentation of the reaction products for (a) styrene and (b) maleic anhydride. (c) Pair interaction energies, divided into the UHF and dispersion (D) contributions. The radical fragment (the reaction center) is number 8.

In case of UHF/D, one can decompose PIEs into the dispersion contribution (D), and the rest (denoted by UHF),

$$\Delta E_{IJ}^{\mathrm{UHF/D}} = \Delta E_{IJ}^{\mathrm{UHF}} + \Delta E_{IJ}^{\mathrm{D}}, \tag{37}$$

where the dispersion contribution (D) is computed as the difference between the dispersion energies of dimer IJ and monomers I and J. It can be easily computed for connected pairs.

PIEs can be of very considerable aid in judging the quality of fragmentation. Because FMO treats the electrostatic and dispersion (for the empiric model) interactions to the full order,[51] and only the charge transfer and exchange-repulsion interactions are truncated (e.g., to the second oder in FMO2), large non-electrostatic UHF contributions to PIEs can indicate

TABLE III. FMO errors (kcal/mol) in the total energies (ΔE) and in reaction heats ($\Delta\Delta E$) as well as rmsd (Å) between the optimized structures, FMO-UHF/D vs *ab initio* UHF/D (6-31G(d)). The polymerization reaction heat at the *ab initio* UHF level is -29.91 and -15.84 kcal/mol for styrene and maleic anhydride, respectively.

property	reactant[a]	product1[b]	product2[c]
ΔE	-0.11	-0.18	0.06
$\Delta\Delta E$		-0.07	0.17
rmsd	0.16	0.11	0.06

[a] The reactant is shown in FIG. 7. [b] Reaction with styrene [c] Reaction with maleic anhydride.

that the FMO error may be considerable. The small values of the UHF contribution to PIEs (the largest is 1.0 kcal/mol) shown in FIG. 9 are in agreement with the high accuracy of FMO total energies discussed above.

As can be seen in FIG. 9, there should be a large interaction of the $i, i+2$ nature, where i numbers monomer units. The $i, i+1$ interactions include a steric repulsion between rings, so that the neighboring rings in pairs $i, i+1$ are staggered, especially the pairs of benzene rings. Because we had to divide the system as two monomer units per fragment, the $i, i+2$ interactions are of the $I, I+1$ type, where I numbers fragments. It is clear from FIG. 9, that there is a very strong attractive dispersion interaction of the $I, I+1$ type. One can observe that it is affected by the ring stacking geometry and the ring kind (benzene or maleic anhydride ring). The strongest attraction is for the 6,5 pair (about -16 kcal/mol). The second in importance is the $I, I+2$ interaction, for the 5,3 and 8,6 pairs it is -2.8 kcal/mol.

The main difference between the two products comes from fragment 8 (which is the reaction center). The 8,6 pair interaction is substantial in styrene (-2.8 kcal/mol) but small in maleic anhydride (-0.5 kcal/mol). As discussed elsewhere,[6] the PIEs shown in FIG. 9 are not the only contributions to the reaction enthalpy. The other two factors are the polarization of fragments (related to the internal energies E'_I) and the internal energy contributions for connected fragment pairs.

It is easy to compare the polarization of fragments 1-7, which are defined identically for the two products. It should be remembered that in FMO the polarization is divided into two parts, the destabilization (contained in E'_I) and stabilization contributions. The latter is a part of PIEs ΔE_{IJ} and can be explicitly extracted.[51] Most internal energies E'_I, I=1-7 are similar between the two products, and a large difference is observed for fragment 6, which is by 1.7 kcal/mol less destabilized in the maleic anhydride product compared to styrene, and fragment 7, which is more destabilized for maleic anhydride by 3.6 kcal/mol. Maleic anhydride is more polar, and it perturbs (polarizes) the nearest fragment more (but less the second nearest). The total destabilization polarization energy for fragments 1-7 in the product is larger by 2.1 kcal/mol for maleic anhydride than for styrene. Fragment 8 is different in size so we cannot discuss its relative polarization (E'_I are not comparable).

These two factors (-2.6 kcal/mol from all PIEs and -2.1 kcal/mol from polarization) explain about 1/3 of the total reaction energy difference of -14.1 kcal/mol between styrene and maleic anhydride. The rest is not so easy to analyze, as it is contained in the connected fragment pairs and the polarization of the reaction center fragment. The main part of the reaction energy difference between the styrene and maleic anhydride does not come from dispersion. The sum of all $\Delta E_{IJ}^{\mathrm{D}}$ is more exothermic by -3.2 kcal/mol for the styrene product.

In addition to the energetic factors, one can look at the dipole moments. The dipole moment of the reaction center (fragment 8) in the styrene and maleic anhydride reaction products is 1.702 and 6.535 Debye, respectively. The total dipole moments of the products at the FMO2 level are 6.358 and 8.123 Debye, respectively. This shows that the maleic anhydride product is more polar, and that the difference is very pronounced in the reaction center. It will have implications in the production of these polymers, for example, in the solubility and in general for solvent effects, which should be larger for the maleic anhydride polymerization.

4. Computational time

We evaluated the total timings for the hydroxyl radical solvated in the water boxes of the linear size of 10, 15, 20, 25 and 30 Å. The timings are measured on 8 nodes equipped

with dual 2.93 GHz Xeons (6 cores per CPU) and 64 GB RAM. It is clear that the ESP-PC approximation is very efficient in reducing the computational costs: for the largest system the exact ESP calculation took 41.3 hours whereas the ESP-PC approximation took only 5.7 hours (in both calculations, we used $R_{\mathrm{ES-DIM}} = 2.5$).

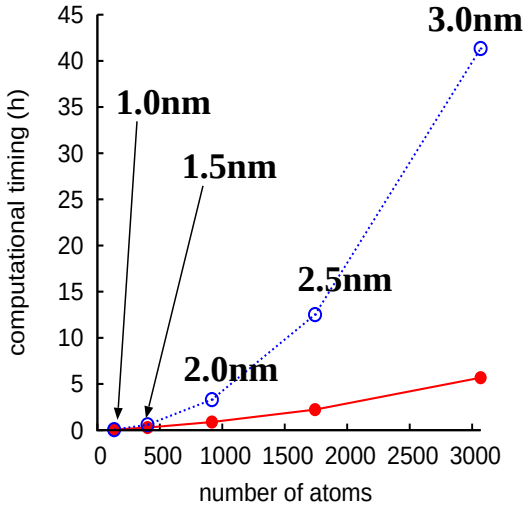


FIG. 10. Computational time of the complete analytic energy gradient in FMO-UHF/6-31G(d) for the OH radical solvated in water boxes, obtained on eight nodes equipped with dual 2.93 GHz Xeons (6 cores per CPU) and 64 GB RAM. The data for the approximated ESP ($R_{\mathrm{ES-DIM}} = 2.5$ and $R_{\mathrm{ESP-PC}} = 2.5$) are shown as red open circles with a solid line, whereas blue filled circles with a dashed line is used to depict the data for the exact ESP. The linear size of water boxes is shown in nm.

For these systems, FMO both with and without the ESP-PC approximation outperformed *ab initio*. The *ab initio* calculation took 1.2 hours for the 15 Å box (401 atoms) and FMO with and without the ESP-PC approximation took 0.28 and 0.59 hours, respectively. The *ab initio* computation took more than 48 hours for the 20 Å box (917 atoms), whereas FMO with and without the ESP-PC approximation took 0.87 and 3.30 hours, respectively. Compared to the high cost of *ab initio* calculations (with approximately cubic scaling in general), it can be seen in FIG. 10 that the scaling of FMO with approximated ESP is nearly linear.

E. Conclusions

We have developed the energy and analytic derivatives of the approximated electrostatic potential in FMO-UHF, including the rigorous contributions from response terms. We have demonstrated the good accuracy of FMO-UHF in comparison to both *ab initio* analytic gradients and FMO numerical gradients. For the sufficiently large values of the point charge threshold, the developed analytic gradient is highly accurate. The importance of response terms is clearly demonstrated by comparison to numerical energy gradients and in the energy fluctuations in MD. The measured timings show that the scaling of FMO-UHF gradient is approximately linear.

Secondly, we have extended the application field of FMO to polymers. By studying the living free radical polymerization reaction, we have shown that FMO has a good accuracy in both the structures and reaction energies, in comparison to full *ab initio* calculations. Some insights into the nature of intramolecular interactions in polymers have been gained with the aid of fragment polarization and pair interaction energies in FMO. We expect that FMO-UHF will become a useful tool to study large molecular systems with open-shell character.

Supporting Information

A detailed mathematical derivation of the gradients and the Cartesian coordinates of molecular systems are available via the Internet at
http : //link.springer.com/content/esm/art : 10.1007
/s00214 − 014 − 1477 − 6/file/MediaObjects/214_2014_1477_FMOESM1_ESM.pdf

REFERENCES

[1]S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[2]G. E. Scuseria, J. Phys. Chem. A **103**, 4782 (1999).

[3]X. Li, J. M. Milliam, G. E. Scuseria, M. J. Frisch, and H. B. Schlegel, J. Chem. Phys. **119**, 7651 (2003).

[4]P. G. Mezey and J. Leszczynski, *Linear-Scaling Techniques in Computational Chemistry and Physics.*, Springer, New York, 2011.

[5]J. R. Reimers, *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, Wiley, New York, 2011.

[6]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[7]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[8]W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[9]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[10]Y. Wang, C. P. Sosa, A. Cembran, D. G. Truhlar, and J. Gao, J. Phys. Chem. B **116**, 6781 (2012).

[11]J. Korchowiec, F. L. Gu, and Y. Aoki, Int. J. Quantum Chem. **105**, 875 (2005).

[12]Y. Aoki and F. L. Gu, Phys. Chem. Chem. Phys. **14**, 7640 (2012).

[13]X. H. Chen and J. Z. H. Zhang, J. Theor. Comput. Chem. **3**, 277 (2004).

[14]S. Hua, W. Li, and S. Li, Chem. Phys. Chem. **14**, 108 (2013).

[15]M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, and J. A. Boatz, J. Phys. Chem. B **113**, 9646 (2009).

[16]J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, J. Chem. Theory Comput. **8**, 2835 (2012).

[17]M. Kobayashi, T. Yoshikawa, and H. Nakai, Chem. Phys. Lett. **500**, 172 (2010).

[18]X. He and K. M. Merz, J. Chem. Theory Comput. **6**, 405 (2010).

[19]M. Kobayashi and H. Nakai, Phys. Chem. Chem. Phys. **14**, 7629 (2012).

[20]M. A. Collins, Phys. Chem. Chem. Phys. **14**, 7744 (2012).

[21]L. Huang and L. Massa, Future Medicinal Chemistry **4**, 1479 (2012).

[22]P. Söderhjelm, J. Kongsted, and U. Ryde, J. Chem. Theory Comput. **6**, 1726 (2010).

[23]N. Sahu, S. D. Yeole, and S. R. Gadre, J. Chem. Phys. **138**, 104101 (2013).

[24]A. Frank, H. M. Möller, and T. E. Exner, J. Chem. Theory Comput. **8**, 1480 (2012).

[25]E. K. Kurbanov, H. R. Leverentz, D. G. Truhlar, and E. A. Amin, J. Chem. Theory Comput. **8**, 1 (2012).

[26]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[27]D. G. Fedorov and K. Kitaura, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC press, Boca Raton, FL, 2009.

[28]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A. **111**, 6904 (2007).

[29]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[30]C. Steinmann, D. G. Fedorov, and J. H. Jensen, PLOS One **8**, e60602 (2013).

[31]S.-I. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, Chem. Phys. Lett. **382**, 611 (2003).

[32]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[33]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).

[34]S. R. Pruitt, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Theory Comp. **6**, 1 (2010).

[35]S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, J. Phys. Chem. A **116**, 4965 (2012).

[36]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **122**, 0541081 (2005).

[37]Y. Komeiji, Y. Mochizuki, T. Nakano, and H. Mori, *Recent Advances in Fragment Molecular Orbital-Based Molecular Dynamics(FMO-MD) Simulations.*, InTech, 2012.

[38]H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, J. Chem. Phys. **137**, 044110 (2012).

[39]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).

[40]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[41]Y. Alexeev, M. P. Mazanetz, O. Ichihara, and D. G. Fedorov, Curr. Top. Med. Chem. **12**, 2013 (2012).

[42]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B **111**, 9621 (2007).

[43]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B **116**, 503 (2012).

[44]H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii, and K. Kitaura, J. Phys. Chem. A **112**, 10887 (2008).

[45]P. V. Avramov, D. G. Fedorov, P. B. Sorokin, S. Sakai, S. Entani, M. Ohtomo, and H. N. Y. Matsumoto, J. Phys. Chem. Lett. **3**, 2003 (2012).

[46]L. Roskop, D. G. Fedorov, and M. S. Gordon, Mol. Phys. , in press (2013).

[47]Y. Okiyama, T. Tsukamoto, C. Watanabe, K. Fukuzawa, S. Tanaka, and Y. Mochizuki, Chem. Phys. Lett. **566**, 25 (2013).

[48]H. Sekino, N. Matsumura, and Y. Sengoku, Comput. Lett. **3**, 423 (2007).

[49]Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, Chem. Phys. Lett. **445**, 331 (2007).

[50]Q. Gao, S. Yokojima, D. G. Fedorov, K. Kitaura, M. Sakurai, and S. Nakamura, J. Chem. Theory. Comput. **6**, 1428 (2010).

[51]D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).

[52]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **116**, 704 (2012).

[53]Y. Mochizuki, K. Fukuzawa, A. Kato, S. Tanaka, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **410**, 247 (2005).

[54]T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, H. Tokiwa, S. Tanaka, and K. Tanaka, Theor. Chem. Acc. **118**, 937 (2007).

[55]M. C. Green, D. G. Fedorov, K. Kitaura, J. S. Francisco, and L. V. Slipchenko, J. Chem. Phys. **138**, 074111 (2013).

[56]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **351**, 475 (2002).

[57]K. Kitaura, S. I. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, Chem. Phys. Lett. **336**, 163 (2001).

[58]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[59]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **475**, 124 (2009).

[60]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **544**, 87 (2012).

[61]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[62]D. G. Fedorov, Y. Alexeev, and K. Kitaura, J. Phys. Chem. Lett. **2**, 282 (2011).

[63]Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **372**, 342 (2003).

[64]Y. Komeiji, T. Ishikawa, Y. Mochizuki, H. Yamataka, and T. Nakano, J. Comput. Chem. **30**, 40 (2009).

[65]T. Fujita, H. Watanabe, and S. Tanaka, J. Phys. Soc. Jpn. **78**, 104723 (2009).

[66]T. Fujita, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **506**, 112 (2011).

[67]K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 5008 (2012).

[68]H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, J. Chem. Phys. **138**, 164103 (2013).

[69]G. Kacar, C. Atilgan, and A. S. Özen, J. Phys. Chem. C **114**, 370 (2010).

[70]M. Nagaoka, Y. Ohta, and H. Hitomi, Coord. Chem. Rev. **251**, 2522 (2007).

[71]J. A. Elliott and S. J. Paddison, Phys. Chem. Chem. Phys. **9**, 2602 (2007).

[72]M. Karttunen, I. Vattulainen, and A. Lukkarinen, *Novel Methods in Soft Matter Simulations*, Springer-Verlag, Berlin, Heidelberg, 2004.

[73]G. Morales and R. Martinez, J. Phys. Chem. A **113**, 8683 (2009).

[74]S. S. Zade and M. Bendikov, Chem. Eur. J. **13**, 3688 (2007).

[75]S. Suhai, J. Chem. Phys. **73**, 3843 (1980).

[76]S. Hirata, Phys. Rev. B **57**, 11994 (1998).

[77]Y. Aoki, A. Imamura, and T. Sasaki, Bull. Chem. Soc. Jpn. **61**, 1063 (1988).

[78]S. Hirata, Phys. Chem. Chem. Phys. **11**, 8397 (2009).

[79]D. Moscatelli, C. Cavallotti, and M. Morbidelli, Macromolecules **39**, 9641 (2006).

[80]W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108 (2008).

[81]H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, J. Chem. Phys. **128**, 034107 (2008).

[82]N. J. Mayhall, K. Raghavachari, and H. P. Hratchian, J. Chem. Phys. **132**, 114107 (2010).

[83]J. Baker, A. Kessi, and B. Delley, J. Chem. Phys. **105**, 192 (1996).

[84]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).

[85]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, and J. Goddard, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic*

Structure Theory, Oxford University Press, New York, 1994.

[86]N. C. Handy and H. F. Schaefer III, J. Chem. Phys. **81**, 5031 (1984).

[87]C. Ochsenfeld and M. S. Gordon, Chem. Phys. Lett. **270**, 399 (1997).

[88]J. Clayden, N. Greeves, S. Warren, and P. Wothers, *Organic Chemistry*, Oxford university press, New York, 2001.

[89]M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, Comput. Phys. Commun. **181**, 1477 (2010).

[90]J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[91]N. W. Schmidt, K. K. Baldridge, J. A. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[92]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[93]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004).

[94]H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).

[95]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, 1987.

[96]D. Benoit, C. J. Hawker, E. E. Huang, Z. Lin, and T. P. Russell, Macromolecules **33**, 1505 (2000).

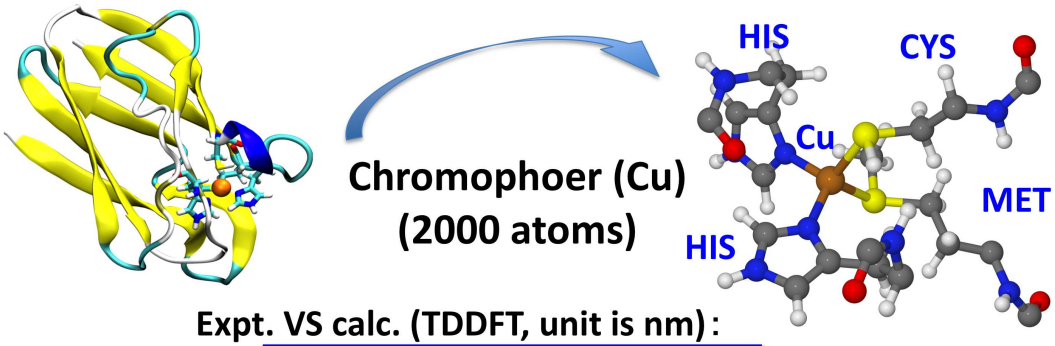
[97]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[98]D. G. Fedorov and K. Kitaura, *In Modern Methods for Theoretical Physical Chemistry and Biopolymers; Starikow, E; Tanaka, S; Lewis, J.*, Elsevier, Amsterodam, 2006.

Chapter IV

Unrestricted density functional theory based on the fragment molecular orbital method for the ground and excited state calculations of large systems

Absorption spectrum for plastocyanine



Expt. VS calc. (TDDFT, unit is nm):

expt.	781	598
calc.	796	549

H. Nakata, D. G. Fedorov *et. al.*

J. Chem. Phys. **2014,** *140,* 144101.

DOI :10.1063/1.4870261
View online: http://dx.doi.org/10.1063/1.4870261

IV. UNRESTRICTED DENSITY FUNCTIONAL THEORY BASED ON THE FRAGMENT MOLECULAR ORBITAL METHOD FOR THE GROUND AND EXCITED STATE CALCULATIONS OF LARGE SYSTEMS

A. Introduction

Photochemical reactions are important and interesting features of many chemical processes in biochemistry[1,2]. They are initiated by the absorption of a photon from visible or ultraviolet radiation followed by the transition to an electronically excited state. In biological systems such as photosynthesis[4], electronic excitations trigger oxidation-reduction processes, often catalyzed by transition metals[3].

The photosynthetic center involves several transition metals (Cu, Fe or Mn) with unpaired electrons and in order to describe electronic excitations several methods can be applied, such as configuration interaction,[5] symmetry-adapted cluster configuration interaction (SAC-CI),[6,7] equation-of-motion coupled-cluster with single and double excitations (EOM-CCSD),[8,9] multireference configuration interaction,[10] and time-dependent density functional theory (TDDFT)[11–16]. TDDFT can be used for the geometry optimization of excited states and molecular dynamics (MD) simulations[17] to predict fluorescence spectra and mechanisms of photochemical processes.

Although TDDFT is not always very accurate and can have problems with charge transfer states[18] often attributed to the treatment of long-range electrostatics, it is relatively inexpensive and capable of providing a satisfactory description of small and medium-sized systems with several transition metals when more advanced methods are difficult to apply[2,19,20]. Unrestricted DFT and TDDFT are also prone to possible spin-contamination (i.e., electronic states do not possess proper integer or half-integer spins).

In solution, solvent can shift the excitation energy of a solute, known as the solvatochromic shift[21]. It can be evaluated using either continuum approaches, such as the polarizable continuum model (PCM)[22–24], typically applicable to the solvatochromic shift caused by the electrostatic solvent effect, or with explicit solvation, in which solvent molecules can also describe other kinds of shifts, for instance due to solute-solvent charge transfer. Combined

quantum mechanics and molecular mechanics (QM/MM),[25–28], especially with accurate force fields such as the effective fragment potential[29–31], as well as fragmentation approaches[32–45] can be used to treat solvent explicitly. Recently, a number of efficient inexpensive excited state methods have been developed based on fragmentation[46,47], linear-scaling TDDFT[48,49] and other ideas[50–54].

In the fragment molecular orbital (FMO) method[55–58], each fragment is polarized with the embedding electrostatic potential (ESP), evaluated in the iterative fashion using the charge distribution of all other fragments; followed by fragment pair calculations, also done in the presence of ESP. Geometry optimizations,[59,60] molecular dynamics,[61,62] and vibrational analysis[63] can be performed and nuclear magnetic resonance chemical shifts[64–66] can be evaluated. FMO has been applied to study protein-ligand complexes,[67], DNA[68], cellulose[69] and inorganic systems[70–73]. FMO is based on the many-body expansion of the total energy and it provides interaction energies in various analyses[74–78]. For closed shells, FMO formulations have been developed for DFT,[79] second order Møller-Plesset perturbation theory (MP2),[80] coupled cluster,[81] configuration interaction,[82,83] PCM,[84] and effective fragment potential (EFP).[85] For open shells, only Hartree-Fock (HF) and multi-configuration self-consistent field formulations are available[86–90]. FMO-TDDFT has been developed[91–95] only for closed shell DFT and applied to a number of systems[96–98].

In this work, we develop the energy and energy gradient for the ground state using unrestricted DFT (FMO-UDFT) as well as the energy for the excited states (FMO-UTDDFT). The accuracy of these methods is evaluated in comparison with both conventional *ab initio* calculations and experiment. For test calculations, we chose benzoquinone anion radical, solvated in water, which is a typical organic molecule in biological systems[99] and an important example for a solvated quinone[100]. For a system with fragments connected by covalent bonds, we calculated the active part of a protein, *Pseudomonas aeruginosa* azurin[101]. In addition, we also calculated absorption spectra of tetramethylpiperidine N-oxide (TEMPO) in different solvents and compared them to experimental results[102,103]. For a larger test, we chose the plastocyanin protein[104], for which both experimental[104,105] and other theoretical studies[106,107] have been published. The computational efficiency is evaluated on solvated TEMPO.

B. Theory

In unrestricted FMO-DFT, similar to unrestricted HF (UHF)[90], we select several open-shell fragments and describe them with UDFT, whereas other fragments are treated with restricted DFT (RDFT). The ground state total energy E is given by

$$\begin{aligned} E = & \sum_{I=1}^{N_{\mathrm{RDFT}}} E_I + \sum_{I>J}^{N_{\mathrm{RDFT}}} (E_{IJ} - E_I - E_J) \\ & + \sum_{K=1}^{N_{\mathrm{UDFT}}} E_K + \sum_{K>L}^{N_{\mathrm{UDFT}}} (E_{KL} - E_K - E_L) \\ & + \sum_{K=1}^{N_{\mathrm{UDFT}}} \sum_{I=1}^{N_{\mathrm{RDFT}}} (E_{KI} - E_K - E_I) , \end{aligned} \tag{1}$$

where I and J denote RDFT fragments, whereas K and L number UDFT fragments. N_{RDFT} and N_{UDFT} are the numbers of RDFT and UDFT fragments, respectively. E_I and E_{IJ} are the monomer and dimer ground state energies, respectively. Closed-shell fragments have zero spin density, because it is obtained from a single Slater determinant with identical α and β orbitals.

In the actual implementation we only allow one open-shell (UDFT) fragment. The gradient in FMO-UDFT is calculated by taking analytic derivative of Eq. (1), and explicit derivations are essentially the same as in FMO-UHF[90] so we omit them here, except that in the present formulation we neglect the response terms, whose magnitude for restricted Hartree-Fock (RHF)[108] and UHF[90] has been evaluated and which can reduce the efficiency of geometry optimizations.

Assuming one open-shell chromophore fragment K, the excited state energy E^* in FMO-UTDDFT is given by

$$\begin{aligned} E^* = & \sum_{I=1}^{N_{\mathrm{RDFT}}} E_I + \sum_{I>J}^{N_{\mathrm{RDFT}}} (E_{IJ} - E_I - E_J) \\ & + E_K^* + \sum_{I=1}^{N_{\mathrm{RDFT}}} (E_{KI}^* - E_K^* - E_I) , \end{aligned} \tag{2}$$

where E_K^* and E_{KI}^* are the monomer and dimer fragment excited state energies, respectively. By subtracting the ground state energy in Eq. (1) from the excited state energy in Eq. (2),

the energies of the RDFT fragments cancel out and one gets the excitation energy ω_i for an excited state i

$$\omega_i = \omega_i^K + \sum_{I=1}^{N_{\mathrm{RDFT}}} \left(\omega_i^{KI} - \omega_i^K\right), \tag{3}$$

where ω_i^K and ω_i^{KI} are the UTDDFT excitation energies of monomer K and dimer KI, respectively.

$$\omega_i^X = E_X^{i,*} - E_X \tag{4}$$

where X is K or KI; $E_X^{i,*}$ is the same as E_X^* above with the explicitly specified excited state (i).

Equation (3) corresponds to the two-body FMO expansion (FMO2), and if the sum describing dimer corrections is neglected, then the method is FMO1. It is the addition of the pair corrections in Eq. (3) which eliminates edge effects in covalent systems because in dimers the detached bonds are present intact; also, pair corrections describe quantum effects upon the excitations such as charge transfer and exchange-repulsion. In fact, in FMO-based TDDFT, one does not need to calculate the energy of RDFT dimers at all to obtain the excitation energy ω_i: only dimers KI are needed.

The limitation of assigning only one fragment as open-shell is that the accuracy of this approach can be expected to be reasonably good for mainly local excitations, when pair corrections efficiently bring in non-local effects at the two-body level; this approach may have problems with completely delocalized excitations. It is customary in excited state studies to look at the main orbitals describing the excitation. In FMO one should also pay attention as to whether these orbitals might be localized on fragment borders with a possible artificial effects due to fragmentation. In addition, large pair corrections (i.e., $\omega_i^{KI} - \omega_i^K$ in Eq. (3)) indicate that the excitation is not very localized and a redefinition of the chromophore fragment may be necessary.

To improve computational efficiency, we introduce an approximation, which we denote by its keyword, RCORSD. In this approximation, only those dimers KI are computed with UTDDFT, for which the interfragment distance is smaller than the threshold value. This makes the number of UTDDFT dimers to be independent of the total number of fragments and we achieve $O(N^0)$ scaling in terms of UTDDFT monomer and dimers and nearly linear

scaling $O(N^1)$ for UDFT monomers (N is the number of fragments). Because UTDDFT step is often the most time consuming and determines the scaling, FMO-UTDDFT is potentially of $O(N^0)$ scaling, although it certainly has a higher scaling for the ground state part (UDFT). In practice, for small open-shell fragments the ground state calculation is usually dominant, but for large open-shell fragments (which is often the case) the excited state calculation is the most time consuming.

The excited state energies of fragments are calculated solving the non-Hermitian eigenvalue equation,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{5}$$

where

$$\begin{aligned} A_{i^\alpha a^\alpha, j^\alpha b^\alpha} &= \delta_{i^\alpha j^\alpha} \delta_{a^\alpha b^\alpha} (\epsilon_{a^\alpha} - \epsilon_{i^\alpha}) + (i^\alpha a^\alpha | j^\alpha b^\alpha) \\ &\quad - c_{\mathrm{HF}} (i^\alpha j^\alpha | a^\alpha b^\alpha) + 2 f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\alpha b^\alpha}, \qquad (6) \\ A_{i^\alpha a^\alpha, j^\beta b^\beta} &= (i^\alpha a^\alpha | j^\beta b^\beta) + 2 f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\beta b^\beta}, \qquad (7) \\ B_{i^\alpha a^\alpha, j^\alpha b^\alpha} &= (i^\alpha a^\alpha | j^\alpha b^\alpha) - c_{\mathrm{HF}} (i^\alpha b^\alpha | a^\alpha j^\alpha) + f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\alpha b^\alpha}, \qquad (8) \\ B_{i^\alpha a^\alpha, j^\beta b^\beta}, &= (i^\alpha a^\alpha | b^\beta j^\beta) + f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\beta b^\beta}, \qquad (9) \end{aligned}$$

and we use i,j for occupied and a,b for virtual orbitals. α and β are the spin labels. The coefficient c_{HF} is the fraction of the HF exchange in hybrid functionals. $(ia|jb)$ is the two electron integral, and ϵ_i is the energy of the molecular orbital i. Ω is the eigenvalue of the matrix on the left hand side of Eq. (5) and it corresponds to the excitation energy from the ground state to an excited state. $f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\beta b^\beta}$ is the exchange-correlation (XC) term given by

$$\begin{aligned} f^{\mathrm{XC}}_{i^\alpha a^\alpha, j^\beta b^\beta} &= \int\int d\mathbf{r} d\mathbf{r}' \psi_{i^\alpha}(\mathbf{r}) \psi_{j^\beta}(\mathbf{r}') \\ &\quad \times \frac{\delta^2 E^{\mathrm{XC}}[n]}{\delta n_\alpha \delta n_\beta} \psi_{a^\alpha}(\mathbf{r}) \psi_{b^\beta}(\mathbf{r}'). \end{aligned} \tag{10}$$

The matrix elements for $A_{i^\beta a^\beta, j^\beta b^\beta}$, $A_{i^\beta a^\beta, j^\alpha b^\alpha}$, $B_{i^\beta a^\beta, j^\beta b^\beta}$, and $B_{i^\beta a^\beta, j^\alpha b^\alpha}$ can be obtained by replacing α with β in Eqs. (6), (7), (8), and (9).

The only change in these equations due to FMO is the fact that they are applied to fragments and their dimers rather than the whole system and, secondly, that the molecular

orbitals and their energies are obtained from fragment calculations in the presence of the embedding potential. We carefully monitored spin values for the states in unrestricted DFT and TDDFT to ensure that spin-contamination was small.

C. Computational details

We implemented the energy and gradient of FMO-UDFT and the energy of FMO-UTDDFT in the FMO code[109,110] in our development version of GAMESS.[111,112] All calculations are parallelized using two-level generalized distributed data interface (GDDI)[113]. In all systems, doublet spin multiplicities were calculated in the ground and excited states.

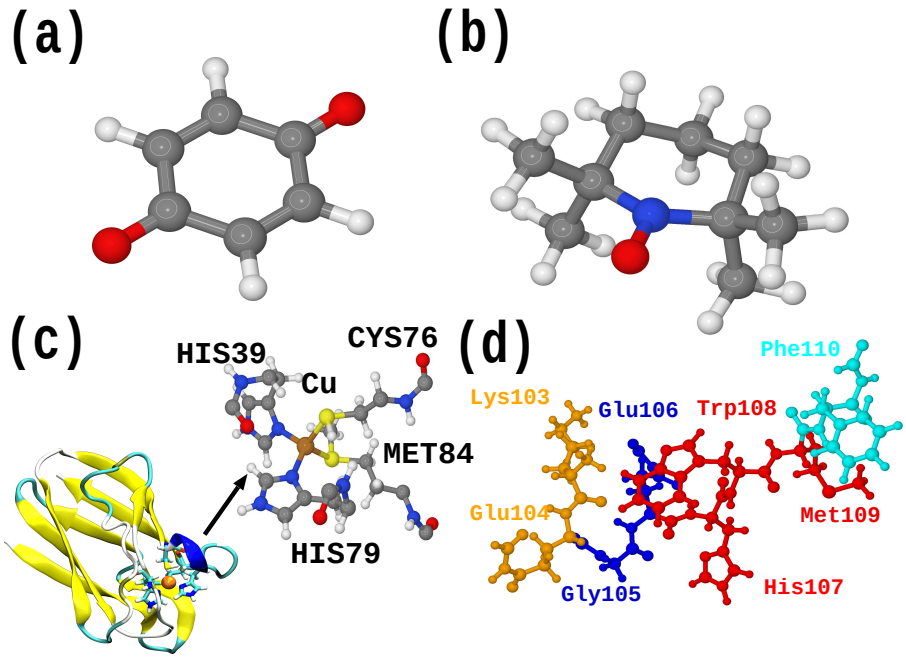


FIG. 1. Geometries of (a) 1,4-benzoquinone anion radical, (b) TEtraMethylPiperidine N-Oxide (TEMPO) and (c) Plastocyanin (PDB: 1BXV). (d) Polypeptide constructed from azurin (PDB: 1R1C).

The geometry of the 1,4-benzoquinone anion radical (FIG. 1-(a)) solvated in 43 water molecules (141 atoms total) was optimized for the ground state using *ab initio* UDFT/6-31G(d) (B3LYP functional[114,115]), and single point excited state calculations were performed

with FMO and *ab initio* TDDFT. In FMO, this system was fragmented by assigning one molecule per fragment (44 fragments total).

The geometry of the azurin protein was obtained from the X-ray structure (PDB: 1R1C)[101]. The residues from 103 to 110 around the tryptophan (NH-Lys-Glu-Gly-Glu-His-Trp-Met-Phe-CO) were extracted to form a polypeptide for the accuracy tests. Hydrogen atoms were added using Leap in Amber[116]. The system was fragmented as one residue per fragment, and by performing a preliminary FMO calculation it was found that the lowest excited state in the dimer Trp-Met and Trp-His has a lower energy than that in the Trp fragment, so that it became clear that one has to merge tryptophan (Trp), methionine (Met) and histidine (His) into one chromophore fragment (the number of fragments was 4). We note that in FMO the fragmentation of polypeptides is performed at Cα atoms (more details are given elsewhere[56]), so that residue fragments are slightly shifted relative to conventional residues defined by dividing peptide bonds.

Tetra-methylpiperidine N-oxide (TEMPO, FIG. (1)-(b)) was solvated in hexane, carbon tetrachloride, ethanol, methanol, and water, using 22, 70, 46, 128 and 164 solvent molecules, respectively (the number of atoms varied from 379 to 521). Initial geometries were generated with a 10 Å solvent shell around TEMPO. For these systems, we took several (5 or 6) minima for each solvent, obtained with Monte-Carlo EFP[117] simulations. Then all atoms in the cluster were optimized using FMO-UHF/6-31G(d) with the empiric dispersion (D)[118] correction (FMO-UHF/D).

Finally, we refined the structures using the frozen domain (FD) formulation at the UDFT/D level with dimer approximation (FDD)[60], using B3LYP and LC-BLYP[119] functionals. In FMO/FDD we defined TEMPO as domain A, and optimized all atoms in it. The polarizable domain B in FMO/FDD coincided with the whole system, thus in this case the whole system was treated at the UDFT/6-31G(d) level. The computational timings were evaluated using the TEMPO solvated in 3, 39, 55, 105, 143, and 162 water molecules (38-514 atoms total).

We took the X-ray structure of plastocyanin (PDB code: 1BXV[120]) and protonated it using LEaP[116]. The structure was optimized with the AMBER99 force field[121] and then we carried out a short 1 ns NPT MD simulation at 300 K and 1 atm. The isotropic position

scaling (ntp=1) and the Langevin thermostat (ntt=3) were used to perform the NPT ensemble simulation. The nonbonded cutoff (cut=12) was set to be 12 Å. The final geometry was used in the following QM/MM optimizations at the UB3LYP/6-31G(d) level. The QM/MM optimized structure was used for the final optimization with FMO/FDD. The fragmentation of plastocyanin in FMO was based on 1 amino acid residue per fragment except that the first layer of residues (HIS39, CYS76, HIS79 and MET84) forming dative bonds with Cu was merged into one large fragment containing Cu.

In the above simulations, solvent water molecules were explicitly considered: the plastocyanin (1358 atoms) was solvated in a 50 Å box of water using periodic boundary conditions in MD, and from that system we extracted a 5 Å solvation shell (196 water molecules) around the solute and used it in QM/MM and FMO (1946 atoms in total). All MD and QM/MM simulation were performed with Amber 12 package[121,122] interfaced with Gaussian 09[123]. In FMO/FDD, domain A was chosen to be the fragment including Cu and domain B includes all fragments within the interfragment distance of 1.5 from the fragment containing copper (the distances in FMO are unitless[126]). There were 63 and 458 atoms in domains A and B, respectively. All atoms in domain A were optimized. FMO/FDD is based on multilayer FMO[124], in which fragments are divided into layers in the following way: fragments in domain B (which by construction includes domain A) are assigned to layer 2, whereas the rest of the system (frozen domain F) belongs to layer 1. We used UHF/STO-3G for layer 1 and UDFT/6-31G(d) (LC-BLYP functional) in layer 2.

We employed the double-ζ MCP basis set for Cu[125], and the 6-31G(d) basis set for all other atoms, using spherical harmonics (ISPHER=1). In FMO, the electrostatic dimer and point charge ESP approximations were applied with the thresholds of $R_{\mathrm{ES-DIM}} = 2.0$ and $R_{\mathrm{ESP-PC}} = 2.0$, respectively, except that we used $R_{\mathrm{ES-DIM}} = 1.5$ and $R_{\mathrm{ESP-PC}} = 1.5$ in FMO/FDD geometry optimizations[60] for computational efficiency and $R_{\mathrm{ESP-PC}} = 2.5$ in the excited state calculations of the plastocyanin to improve the accuracy. We note that both thresholds are unitless,[126] because they are applied to the interatomic distances divided by the sum of the van-der-Waals radii of the two atoms. The van-der-Waals radii for each atom types were taken from Bondi's compilation.[127] Geometry optimizations were performed with the 10^{-4} hartree/bohr convergence criterion (OPTTOL). Explicit solvent was divided

as one molecule per fragment except where otherwise indicated. For making FMO input files, we used Fragit program[128].

To add the inhomogeneous broadening in the absorption spectra, we used the gaussian profile, i.e., the absorption intensity I is taken to be

$$I(\omega) \propto \frac{I_i}{A} \exp\left[-\left(\frac{\omega - \omega_i}{A}\right)^2\right], \tag{11}$$

where the absorption intensity I_i and frequency ω_i for an excited state i are obtained from *ab initio* calculations. A is the parameter determining the band width (0.1 eV is used throughout this study).

D. Results and discussion

1. FMO accuracy

The FMO accuracy was evaluated in comparison with *ab initio* calculations without fragmentation. First, we optimized the geometry of TEMPO solvated in 8 ethanol molecules using FMO-UDFT and full UDFT. The results are shown in TABLE I. The accuracy is

TABLE I. Differences between FMO-UDFT and UDFT optimized structures for TEMPO solvated in 8 ethanol molecules, in kcal/mol for the energy, and Å for the structure (RMSD). Ethanol is grouped as m molecules per fragment; B3LYP/6-31G(d) is used.

m	RMSD	energy error
1	0.177	-9.193
2	0.133	-3.941

improved when 2 ethanol molecules are grouped into 1 fragment, and a better way is to merge the first solvation shell and the solute into one large fragment but if we do it, there will be only 1 fragment left in this system. Details on the accuracy of the gradient are provided in supplemental materials[129].

The accuracy of FMO-UTDDFT for the polypeptide constructed from the *Pseudomonas aeruginosa* azurin[101] protein is evaluated in comparison with *ab initio*. We provide this

calculation as an example of a system with fragments connected by covalent bonds. The lowest five excited states in azurin are shown in TABLE II. We observe reasonable agreement

TABLE II. Excitation energies (eV) of the lowest states in the polypeptide constructed from azurin for FMO-based and full UTDDFT using 6-31G(d).

		LC-BLYP	
Excited state	FMO1	FMO2	full
1	1.022	1.004	0.990
2	1.610	1.620	1.611
3	2.822	2.853	2.871
4	2.936	2.940	2.948
5	3.057	3.062	3.112
RMSD[a]	0.036	0.025	
		B3LYP	
1	0.121	0.147	0.159
2	0.280	0.271	0.294
3	0.748	0.736	0.760
4	0.892	0.898	0.971
5	1.010	0.977	1.009
RMSD[a]	0.040	0.039	

[a] FMO from ab initio (full).

between *ab initio* and FMO, and the RMSD from *ab initio* results for FMO1 and FMO2 using LC-BLYP (B3LYP) are 0.036 (0.040) and 0.025 (0.039), respectively. We note that LC-BLYP results are strikingly different from B3LYP: the former energies are in the range of 1-3 eV, whereas the latter 0.2-1 eV, i.e., B3LYP predicts very low excitation energies. Nevertheless, FMO does quite well in reproducing both the order of the states and the numerical values of energies, for both of these functionals, with about the same error.

The absorption spectrum of the hydrated 1,4-benzoquinone anion radical is shown in

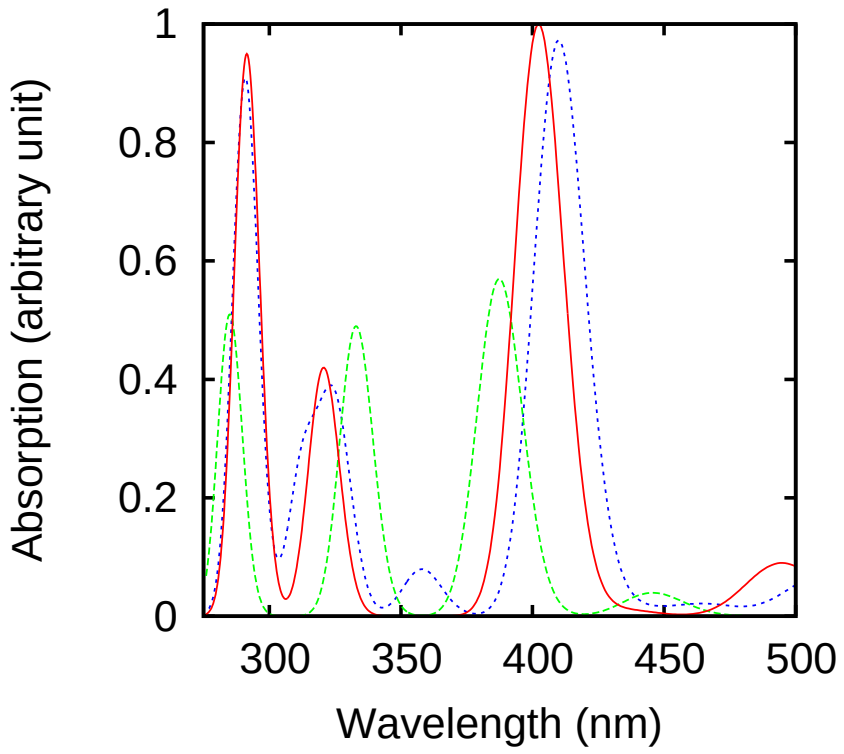


FIG. 2. Absorption spectrum of the 1,4-benzoquinone anion radical, calculated with FMO1-UTDDFT (green dashed line), FMO2-UTDDFT (red solid line), and full UTDDFT (blue dashed line).

FIG. 2, where three prominent peaks are found in both *ab initio* and FMO calculations, corresponding to the excited state 4 (mainly $2b_{3u} \rightarrow 2b_{2g}$ with some $2b_{2g} \rightarrow 3b_{3u}$ contribution), 5 (mainly, $2b_{2g} \rightarrow 1a_u$) and 6 (mainly $2b_{3u} \rightarrow 1a_u$ with some $2b_{2g} \rightarrow 3b_{3u}$ contribution). FMO2 results are much closer to full *ab initio* than FMO1, both in the wavelength and intensity. In this particular system, two-body effects on the excitation energies of absorption are relatively small (0.06-0.26 eV), as summarized in TABLE III.

The largest deviation for FMO1 and FMO2 from full *ab initio* calculations are 0.34 and 0.08 (eV), respectively. Excited states 1, 2, and 3 have noticeable errors at the FMO1 level, mostly corrected with two-body terms, because these excitations are not local to benzo-

TABLE III. Excitation energies (eV) in benzoquinone for FMO-based and full UTDDFT (B3LYP/6-31G(d)). The values in parentheses denote the oscillator strengths.

Excited state	FMO1	FMO2	full
1	2.438(0.000)	2.265(0.000)	2.318(0.000)
2	2.764(0.001)	2.506(0.009)	2.426(0.007)
3	2.788(0.003)	2.845(0.001)	2.813(0.001)
4	3.201(0.057)	3.082(0.100)	3.025(0.097)
5	3.724(0.049)	3.867(0.042)	3.816(0.035)
6	4.351(0.051)	4.255(0.095)	4.266(0.075)

quinone so in order to describe them better, one has to redefine the excited state fragment to include both benzoquinone and the first solvation shell. In the following discussion, we focus on the three states 4, 5 and 6, which have large intensity and which are important for the observed spectrum of benzoquinone.

The details of individual contributions of solvent molecules to the excitation energy are shown in FIG. 3. Before discussing them, it is necessary to explain the physical picture of the FMO model. FMO1 excitation energies are obtained for the solute polarized by all solvent molecules. It is not possible to easily extract the polarization contribution of each water molecule. An important notion is that FMO1 excitation energies already include many-body *electrostatic* effects on the excited states, from all solvent molecules. In FMO2, explicit two-body quantum effects on the excitation energies are evaluated by performing dimer calculations for each solvent-solute pair in the polarizing embedding potential of the whole system. Therefore, these two-body quantum effects are calculated for each solvent molecule explicitly and they are essentially of the exchange-repulsion and charge transfer nature. For RHF, exchange-repulsion contributions can be separated from charge transfer in FMO using the pair interaction energy decomposition analysis (PIEDA)[74].

Because these quantum-effects are usually short-ranged, their magnitude in general decreases with distance, as can be seen in FIG. 3. The largest two-body effects are observed for six water molecules: WAT02, WAT03, WAT04, WAT05, WAT06, and WAT07, which

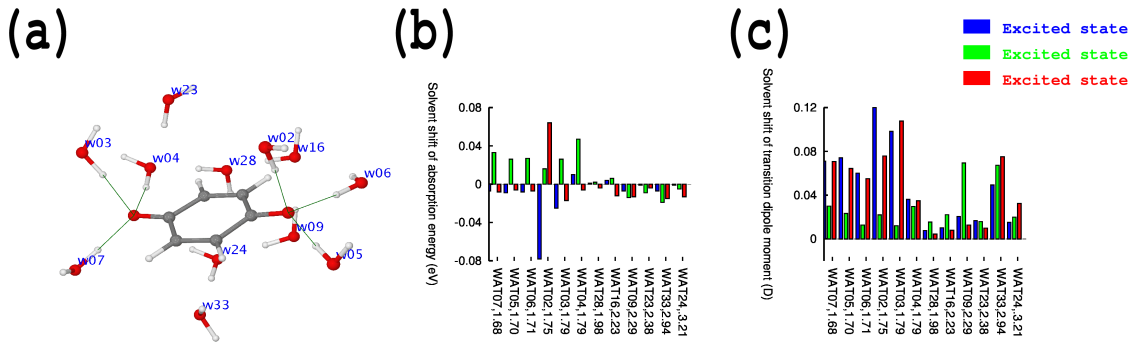


FIG. 3. Quantum-mechanical solvent effects on the excitations in the 1,4-benzoquinone anion radical, for the nearest 12 water molecules. (a) Structure of the cluster showing only the nearest water molecules and their labels. (b) Contributions of individual water molecules to the excitation energies for the excited states 4, 5, and 6. (c) Contributions of individual water molecules to transition dipole moments. The shortest solute-solvent interatomic distances are shown after the fragment label on the x-axis (e.g., WAT07 is separated from the solute by 1.68 Å).

form hydrogen bonds with the solute. However, one can also easily observe, that within the groups of water molecules separated by similar distances, the two-body quantum effects can be very different. For example, WAT05, WAT06, and WAT07 have a large contribution only to the excited state 5, whereas WAT02 has a large effect on the excited states 4 and 6.

These observations can be rationalized by analyzing the orbital contributions (amplitudes) of individual occupied-virtual orbital pairs (see supplemental materials[129] for a rep-

resentative comparison of WAT02 to WAT07). In summary, the following differences are found. The lone pair on WAT02 is directly involved in the excitations, whereas for WAT07 its lone pair makes a small contribution (consequently, excited states 4 and 6 show a large two-body effect for WAT02). An important observation is made from the comparison of monomer and dimer molecular orbitals. Most of them are little affected by the interfragment quantum effects (i.e., monomer and dimer orbitals are similar), whereas others are deformed due to the exchange-repulsion and charge transfer between solute and solvent. When the orbitals, which make large contributions to an excited state, are thus deformed, in general the two-body effects for that particular excited state and the solvent molecule are large.

The role of solvent on the excitation energies (solvatochromic shift) can be of several kinds, from the FMO perspective: a) solvent polarizes the solute and the ground and excited state energies of the solute are shifted differently, which can be rationalized in terms of their dipole moment[130] (this effect is treated by FMO1); b) charge transfer, e.g., from the lone pair on water to solute (FMO2); c) mutual deformation of solute and solvent molecular orbitals, which often explicitly lifts the symmetry of the solute and thus enables otherwise forbidden transitions, as a result of the exchange-repulsion and charge transfer (FMO2). The symmetry of the solute's electronic state is also affected by the electrostatic field of the solvent (FMO1 effect); nevertheless, many molecular orbitals were found to largely retain their symmetry corresponding to D_{2h} of the solute in gas phase. In this particular system, the above effect (b), charge transfer, is the main reason why the excited states 4 and 6 are affected by WAT02, whereas state 5 is not. On the other hand, it is the exchange-repulsion effect (c) distorting the orbital shape (the solute's orbital 26 is strongly affected), which is responsible for state 5 having a large shift due to WAT07.

In contrast to short-ranged two-body quantum effects on the excitation energy, the intensity appears to be more long-ranged. This is due to its physical nature, which is related to the transition dipole derivative. Correspondingly, the long-range nature of the electrostatics shows up as relatively large contributions of the solvent molecules which are not near the solute (FIG. 3): water molecule WAT33 separated by 2.94 Å makes a large contribution.

In this system the separation between the excited states is similar in the chromophore

fragment K and dimers KI (caused by weak two-body quantum effects). This is related to the fact that FMO1 incorporates many-body electrostatic effects, and thus the electrostatic embedding, which shifts the orbital levels, is very similar in FMO1 and FMO2 (although not identical, because the quantum effects in each dimer deform the electron density distribution within the dimer and thus change the electrostatic effect). Due to mainly quantum effects, the orbitals are shifted when comparing monomers and dimers, although usually all levels are shifted by a similar magnitude. Interestingly, in the vicinity of the HOMO-LUMO gap, in almost all cases the orbital shift is downward, i.e., in dimers the orbitals represent more stable levels, for both occupied and virtual orbitals. The only exception is the lone pair of water, which is strongly destabilized in dimers. It can be seen[129] that the lone pair orbital in the dimer with WAT02 mixes very little with the orbitals of the solute, mainly with the lone pair on the nearest oxygen atom in the OH group. Presumably, it is this solute-solvent interaction between the lone pairs which destabilizes the lone pair of water and stabilizes the lone pair of the solute (which does not appear to contribute to the excitation).

At room temperature, solvent molecules move and thus the above effects are dynamic and they should be studied on a representative ensemble of molecular structures[131]. FMO can be used to provide some insights into the solvatochromic shift for a given structure, and, for a realistic picture, this analysis should be statistically averaged for an ensemble.

2. Solvent effect on the absorption spectra of TEMPO

Experimental solvatochromic shifts in TEMPO were reported for hexane, carbon tetrachloride, ethanol methanol, and water molecules.[102,103] We performed geometry optimizations of TEMPO solvated in various solvents. A selection of representative optimized structures of TEMPO solvated in hexane, methanol, and water is shown in FIG. 4. In order to show how the geometry affects the excitation energy, a comparison of several minima for TEMPO solvated in ethanol (6 minima) and methanol (5 minima) is shown in FIG. 5. Some variation on the order of 30 nm is observed. The internal monomer energy of the solute reflects the polarization effect due to the solvent. The stronger is the polarization, the higher (more positive) are the monomer internal energies reflecting the destabilization

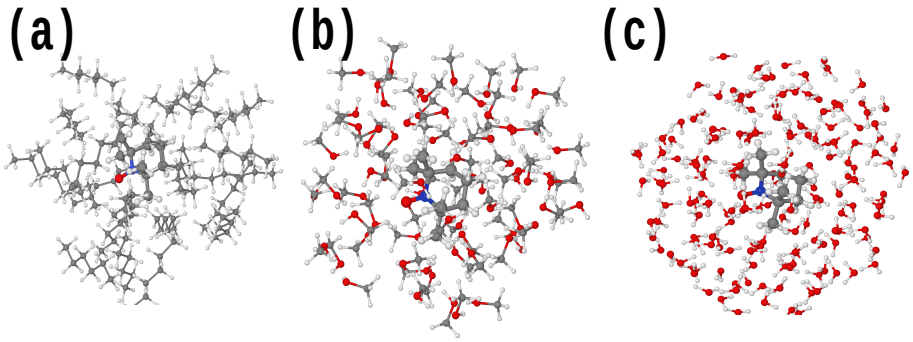


FIG. 4. Optimized geometry of TEMPO solvated in (a) hexane, (b) methanol, and (c) water molecules.

polarization[74]. In general, the structures with larger solute polarization deviate more from experiment.

Another important factor is the hydrogen bonding, whose nature is largely electrostatic contributing to the polarization, and it also involves charge transfer and exchange-repulsion. The latter two terms are not explicitly present in continuum models and can only be taken account of with explicit solvation. The hydrogen bonding summary is given in TABLE IV. Two types of hydrogen bonding are found in this system: {N-O$\cdots$H} –O –H and N-{O$\cdots$H –O} –H, where N-O is in TEMPO (see FIG. 6). The former is stronger and has the shorter bond length than the latter. In general, the stronger the solute-solvent bonding, the lower is the absorption wavelength and the more is the deviation from experiment. The experimentally observed absorption peak positions for methanol and ethanol are 445.2 and 449.7 nm, respectively[102].

We suggest that one reason behind the discrepancy between calculated and experimental values is the thermal fluctuations, namely, the overpolarization of the solute by the solvent in the calculations at 0 K (compared to room temperature experiment). Minima, which have tighter structures, have a stronger polarization (higher internal fragment energy); however, in solution at room temperature the structure is more loose and the polarization is weaker, thus the minima with smaller polarization agree better with experiment: the deviations from the experimental results are reduced from 27 to 2 nm in EtOH, and from 31.7 to 0.9 nm in MeOH (from the most polarized structure to the least polarized structure), for the minima

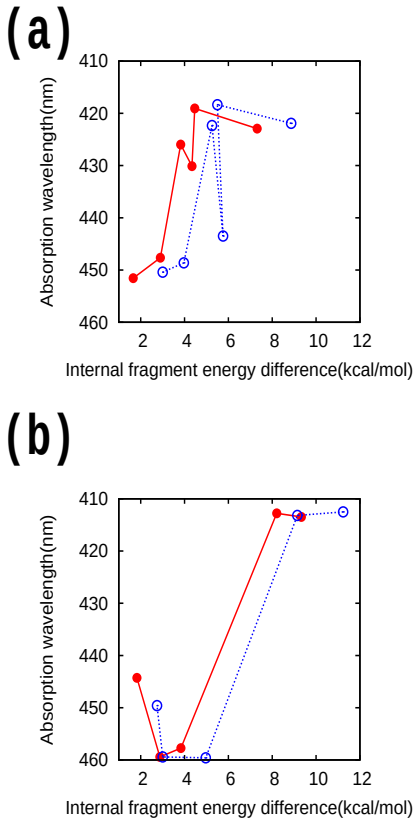


FIG. 5. Relation between the internal energy difference ΔE and the excitation energy for TEMPO solvated in (a) ethanol and (b) methanol, at the FMO2-UTDDFT (B3LYP/6-31G(d)) level (red solid line) and (LC-BLYP/6-31G(d)) level (blue dashed line), where ΔE is the internal energy of TEMPO in solution minus the energy in vacuum.

with the smallest solute polarization.

Because the tightly bound structures at 0 K can be expected to be somewhat affected by the kinetics at room temperature, the relative energies quoted in TABLE IV do not necessarily imply that the structure with the lowest energy at 0 K is the one which most frequently occurs at room temperature. Indeed, we observe that the spectra for the lowest energy structures deviate more from experiment than others with a higher energy, although some deviation is also caused by the level of calculations. For a reliable sampling one should use free energies at room temperature. More tightly bound structures typically have a

TABLE IV. Hydrogen bond lengths R (Å), relative stability E (kcal/mol)[a] and the absorption energy (nm) for solvated TEMPO, at the level of FMO2-UTDDFT/6-31G(d). Hydrogen bond types 1 and 2 are {N-O$\cdots$H} –O –H and N-{O$\cdots$H –O} –H, respectively (see FIG. 6).

		EtOH	
R, type 1	R, type 2	E	absorption
none	none	37.26	451.6
none	1.87	26.15	447.7
1.84	none	103.29	426.0
1.75	none	9.50	419.2
1.86	1.92	0.00	430.1
1.86	1.92	75.77	422.9
		MeOH	
none	none	16.89	444.3
none	1.96	46.58	459.4
none	1.90	0.00	457.7
1.70	none	50.90	413.5
1.70	1.75	45.09	412.8

[a] Calculated with full UDFT/6-31G(d).

smaller entropy and thus are somewhat destabilized at room temperature.

The experimental[102,103] and calculated (FMO-UTDDFT with B3LYP and LC-BLYP) absorption wavelengths are shown in FIG. 7 and TABLE V, where for each system we used the minimum structure with the lowest internal energy of TEMPO (See TABLE 5 for details). The absorption energies obtained with FMO are reasonably close to experiment: the root mean square error (RMSE) is 8.3 (5.9) nm, and the maximum error is 13.2 (9.3) nm for B3LYP (LC-BLYP). Both functionals gave comparable results, with LC-BLYP appearing to perform slightly better perhaps because long-range electrostatic effects in polar solvent are better treated with long-range corrections.

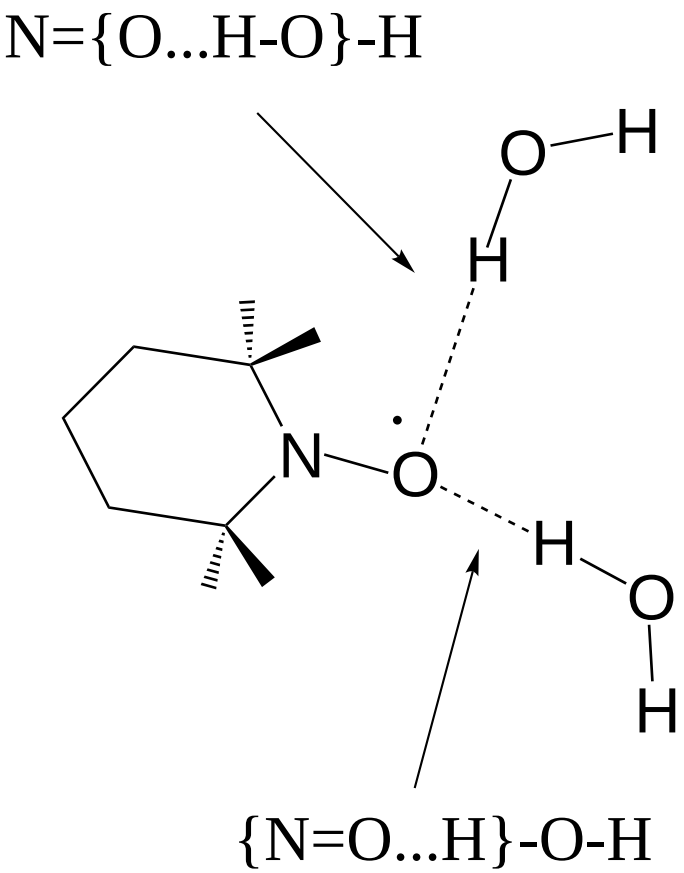


FIG. 6. Schematic illustration of the hydrogen bonding between TEMPO and solvent water molecules.

3. Absorption energy of the blue copper protein

We performed an excited state calculation for the plastocyanin protein using FMO-UTDDFT. The plastocyanin consists of 101 residues, solvated in 196 water molecules (1946 atoms total). The active chromophore contains Cu in the oxidation state +2 (II) and the photosynthesis in plants involves its reduction to +1.

The spectrum calculated with FMO is shown in FIG. 8. We obtained two prominent peaks at 789 (796) nm for the excited state 4, $\sigma \rightarrow \pi*$ (from β HOMO-14 to β LUMO) and 544 (549) nm for the excited state 7, $\pi \rightarrow \pi*$ (from β HOMO-12 to β LUMO) using FMO1 (FMO2). The calculated results are tabulated together with the experimental findings in TABLE VI.

TABLE V. Summary of the observed absorption peak positions (nm) and calculated excitation energies for solvated TEMPO. FMO2-UTDDFT/6-31G* is used.

solvent	Expt.[102]	B3LYP	LC-BLYP
hexane	475.5	466.2	467.9
carbon tetrachloride	469.2	460.5	459.9
ethanol	449.7	451.6	450.4
methanol	445.2	444.3	449.6
water	424.5	411.3	421.1
RMSD from expt.		8.3	5.9

TABLE VI. Summary of observed absorption peak positions (nm) and calculated excitation energies in plastocyanin.

source	method	peak 1	peak 2	reference
spinach	expt	598.84	781.30	[105]
spinach(E. coli)	expt	597.00		[132]
poplar	expt	606.46	749.10	[133]
bean	expt	602.44	738.59	[134]
spinach	calc	598.83	781.30	[106]
spinach	calc	606.00	790.00	[107]
Synechococcus sp.	calc[a]	549.40	795.89	this work

[a] Calculated with FMO2-UTDDFT/6-31G(d).

Although the source of the protein differs and there may be some differences in the structure depending on it, we observe that the calculated and observed results are in the same range. Note that the excited state 6 has a peak at 639 nm (FMO2). However, the nature of this excitation does not agree with other theoretical studies[104,107], and we also assign the excited state 7 rather than 6 as the main transition.

The important molecular orbitals[104,107] shown in FIG. 9, are characterized by the σ and π type mixing interaction of Cu and S in cysteine and also of Cu and N in histidine (FIG. 9-

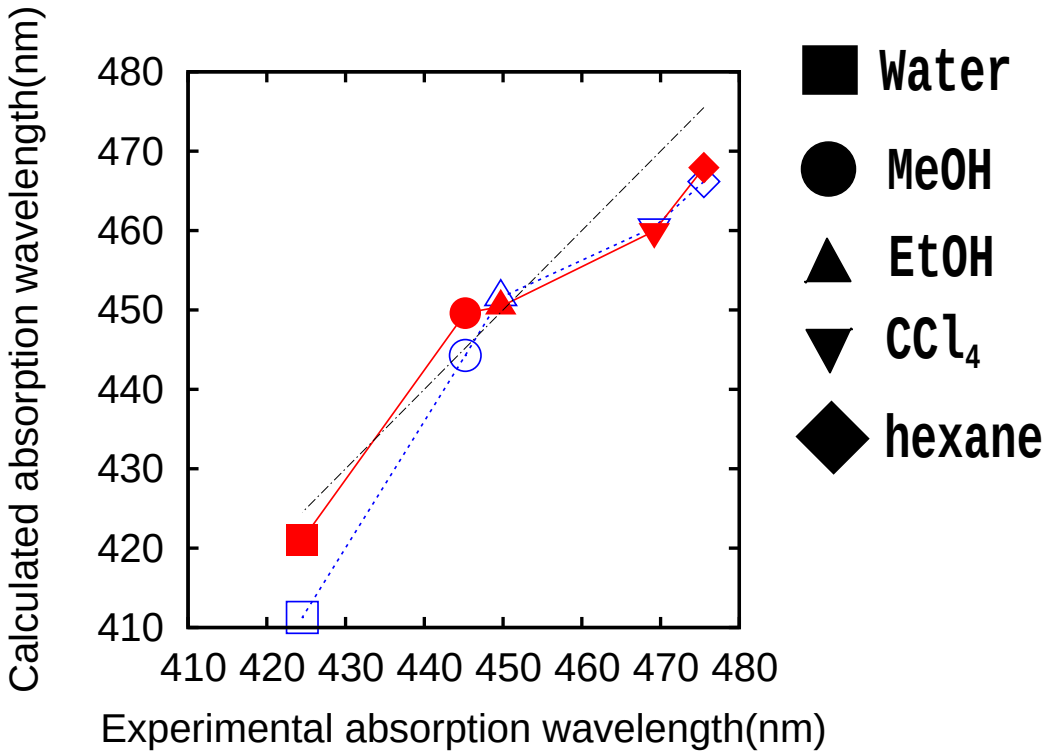


FIG. 7. Experimental and calculated absorption wavelengths of solvated TEMPO. The experimental, LC-BLYP and B3LYP results are shown as dashed black, red solid and blue dotted lines, respectively.

(a) and (b)). The β-LUMO orbital (FIG. 9-(c)) is the π antibonding orbital involving Cu and thiolate[104]. The nature of the excited states 4 and 7 in our study in general agrees to that reported by Monari et. al.[107] However, in the excited state 7, we also found a π-type interaction between methionine and Cu which is not observed in the previous study[107], and methionine is closer to Cu in our optimized structure.

4. Computational timings

The computational timings were measured for TEMPO solvated by water molecules. The results are shown in FIG. 10. As discussed in some detail above, FMO-UTDDFT involves

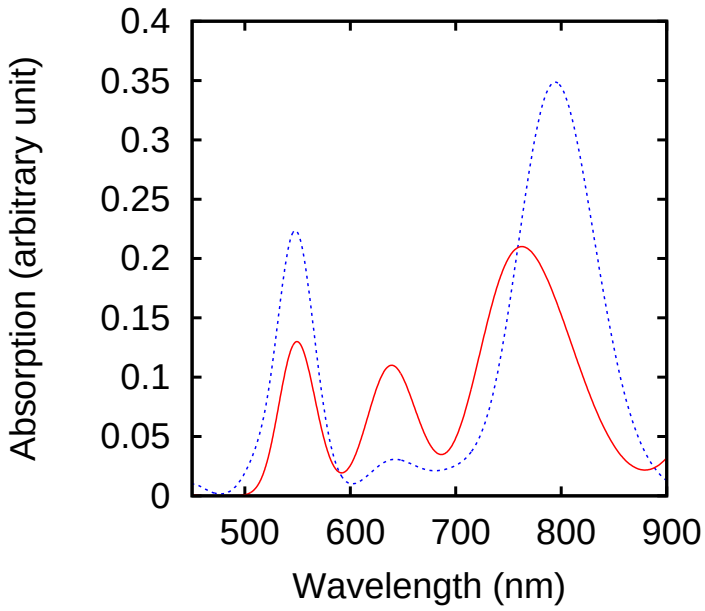


FIG. 8. Calculated absorption spectra of the blue copper protein, plastocyanin computed with FMO1 (blue dashed line) and FMO2 (red solid line) (UTDDFT/6-31G(d)).

several steps with different scaling. As can be seen from the presented timings, the curve becomes essentially flat, i.e., the observed scaling at large N is approximately $O(N^0)$. This is because of the RCORSD approximation: increasing the system size does not increase the number of UTDDFT dimers, which are the main time consuming part of the calculations. However, when the size increases further, the linear scaling part (the ground state) should become apparent and supersede the zero scaling regime.

The linear scaling part in the region of 1-300 atoms does not correspond to this, however. This linear region reflects the linear increase in the number of UTDDFT dimers KI (see above) for small systems, because RCORSD approximation is chosen amply and it is not applied when the interfragment distance is too short. The linear scaling for the sizes between 1-300 atoms comes from the increase in the number of RDFT fragments I (K is the single UDFT fragment). When the size of the system becomes considerable, then the RCORSD

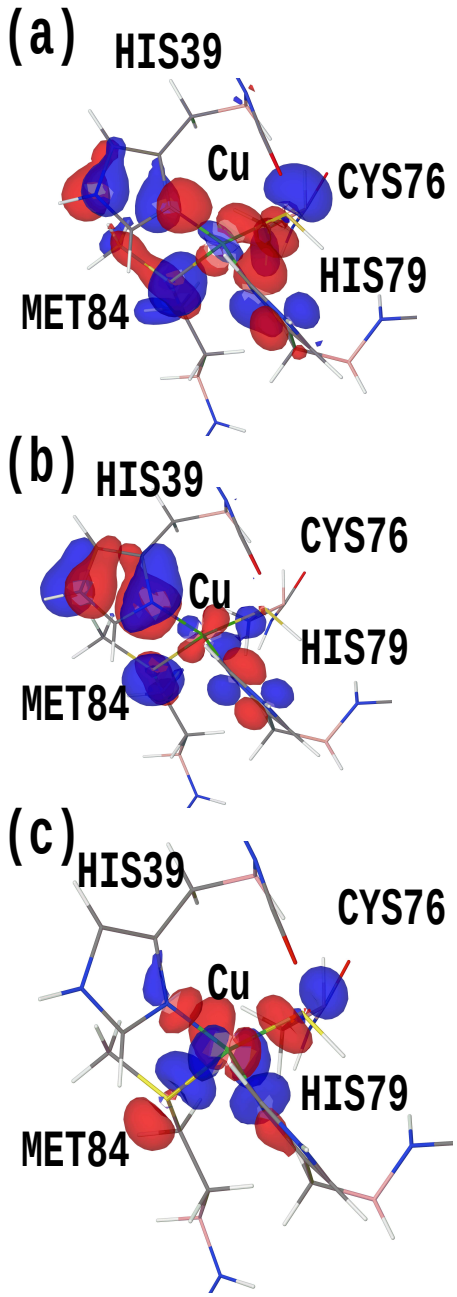


FIG. 9. Molecular orbitals in the plastocyanin: (a) the occupied orbital important for the excited states 4 and 7 (β HOMO-14), (b) the occupied orbital important for the excited state 7 (β HOMO-12) and (c) the virtual orbital important for the excited states 4 and 7 (β LUMO).

approximation screens some dimers and results in only close pairs to be computed (no further increase in timing when more fragments are added).

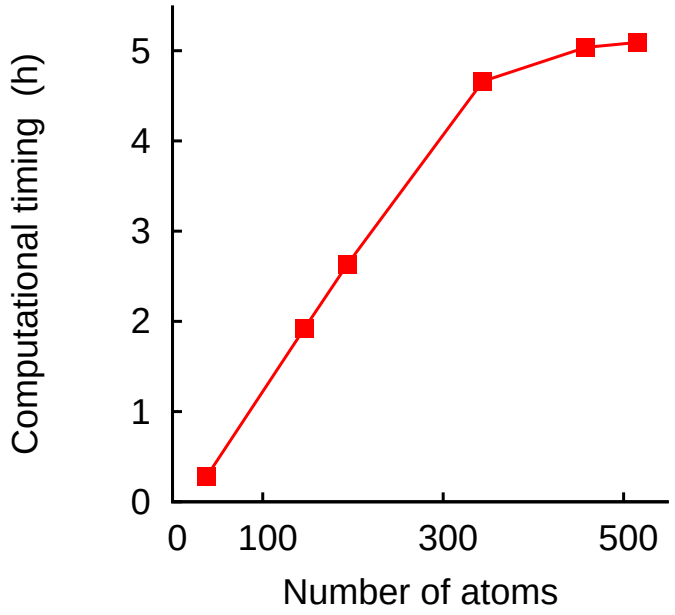


FIG. 10. Computational timing for TEMPO solvated in water, for FMO2-UTDDFT (B3LYP) with 6-31G(d) and RCORSD=2.0, RESDIM = 2.0 and RESPPC=2.5, measured on 16 nodes equipped with 2.93GHz Xeons (8 CPU cores and 12 GB RAM per node).

5. Conclusions

We have developed the unrestricted formalism for density functional theory in the framework of the fragment molecular orbital method. It allows one to optimize the geometry for the ground states and calculate single point excited state energies in the systems with open-shell character.

We have shown that a reasonable agreement can be achieved for the structures and excitation states in solvated organic radicals. In addition, we calculated the spectra of solvated TEMPO radical, and analyzed the nature of solvent effects. The results obtained with FMO optimizations of large solvated systems, followed by single point FMO excited state calculations reproduced experimentally observed solvent effects. FMO provides direct means to separate the electrostatic from quantum effects that solvent exerts upon the excitations, and we have described a case study to demonstrate the analysis. Individual pair corrections of

solvent molecules make a different contribution reflecting how the molecular orbitals of the solute are distorted by solvent. The breaking of the local solute symmetry by solvent has been shown to be the major cause of the change in the nature of the solvatochromic shift.

We have discussed the effect of local minima and the relation between the solute polarization and the excitation energies. The deviation from experiment was attributed to the temperature effect and the general overpolarization of the solute in static (optimized) structures compared to dynamic behavior at room temperature.

In the demonstrative application to the blue copper protein we have shown the potential of the method to biochemistry, as many photochemical processes involve transition metals and unrestricted DFT can be a useful tool to simulate them. Summarizing, we have demonstrated that the developed FMO-based unrestricted DFT can be applied to a variety of systems from solvated organic radicals to proteins, for systems with thousands of atoms. We expect that the developed methods will become a useful tool to study realistic molecular systems.

REFERENCES

[1]T. Helgaker, S. Coriani, P. Jrgensen, K. Kristensen, J. Olsen, and K. Ruud, Chem. Rev. **112**, 543 (2012).

[2]A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).

[3]D. Voet and J. G. Voet, *Biochemistry*, John Wiley and Sons, New York, 2004.

[4]T. J. Wydrzynski and K. Satoh, editors, *Photosystem II, The light-driven water: plastoquinone oxidoreductase*, Springer, Dordrecht, Netherlands, 2005.

[5]J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).

[6]N. Nakatsuji, Chem. Phys. Lett. **59**, 362 (1978).

[7]H. Nakatsuji and K. Hirao, J. Chem. Phys. **68**, 2053 (1978).

[8]J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).

[9]J. Geertsen, M. Rittby, and R. J. Bartlett, Chem. Phys. Lett. **164**, 57 (1989).

[10]H. Werner and E. Reinsch, J. Chem. Phys. **76**, 3144 (1982).

[11]E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).

[12]R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256**, 454 (1996).

[13]R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **109**, 8218 (1998).

[14]S. Hirata and M. Head-Gordon, Chem. Phys. Lett. **314**, 291 (1999).

[15]F. Furche, J. Chem. Phys. **114**, 5982 (2001).

[16]C. Ko, D. K. Malick, D. A. Braden, R. A. Friesner, and T. J. Martínez, J. Chem. Phys. **128**, 104103 (2008).

[17]M. Yoon, Y. Miyamoto, and M. Scheffler, New J. Phys. **13**, 073039 (2011).

[18]M. A. L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, and E. K. U. Gross, *Time-Dependent Density Functional Theory (Lecture Notes in Physics)*, chapter 4, pages 337–353, Springer, 2006.

[19]M. E. Casida, *Time-Dependent Density Functional Response Theory for Molecules*, chapter 5, pages 155–192, World Scientific, Singapore, 1995.

[20]M. Marques and E. Gross, Ann. Rev. Phys. Chem. **55**, 427 (2004).

[21]C. Reichardt, Chem. Rev. **94**, 2319 (1994).

[22]M. Caricato, B. Mennucci, J. Tomasi, F. Ingrosso, R. Cammi, S. Corni, and G. Scalmani, J. Chem. Phys. **124**, 124520 (2006).

[23]R. Improta, G. Scalmani, M. J. Frisch, and V. Barone, J. Chem. Phys. **127**, 074504 (2007).

[24]B. Mennucci, C. Cappelli, C. A. Guido, R. Cammi, and J. Tomasi, J. Phys. Chem. A **113**, 3009 (2009).

[25]V. Luzhkov and A. Warshel, J. Am. Chem. Soc. **113**, 4491 (1991).

[26]A. Broo, G. Pearl, and M. C. Zerner, J. Phys. Chem. A **101**, 2478 (1997).

[27]J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, J. Chem. Phys. **118**, 1620 (2003).

[28]J. Hasegawa, K. J. Fujimoto, and H. Nakatsuji, ChemPhysChem **12**, 3106 (2011).

[29]S. Yoo, F. Zahariev, S. Sok, and M. S. Gordon, J. Chem. Phys. **129**, 144112 (2008).

[30]D. Si and H. Li, J. Chem. Phys. **133**, 144112 (2010).

[31]N. Minezawa, N. D. Silva, F. Zahariev, and M. S. Gordon, J. Chem. Phys. **134**, 054111 (2011).

[32]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632

(2012).

[33]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[34]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[35]X. H. Chen and J. Z. H. Zhang, J. Theor. Comput. Chem. **3**, 277 (2004).

[36]S. Hua, W. Li, and S. Li, Chem. Phys. Chem. **14**, 108 (2013).

[37]M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, and J. A. Boatz, J. Phys. Chem. B **113**, 9646 (2009).

[38]L. Huang and L. Massa, Future Medicinal Chemistry **4**, 1479 (2012).

[39]P. Söderhjelm, J. Kongsted, and U. Ryde, J. Chem. Theory Comput. **6**, 1726 (2010).

[40]N. Sahu, S. D. Yeole, and S. R. Gadre, J. Chem. Phys. **138**, 104101 (2013).

[41]A. Frank, H. M. Möller, and T. E. Exner, J. Chem. Theory Comput. **8**, 1480 (2012).

[42]E. K. Kurbanov, H. R. Leverentz, D. G. Truhlar, and E. A. Amin, J. Chem. Theory Comput. **8**, 1 (2012).

[43]S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, Mol. Phys. **103**, 2255 (2005).

[44]F. Wu, W. Liu, Y. Zhang, and Z. Li, J. Chem. Theory Comput. **7**, 3643 (2011).

[45]C. Yam, Q. Zhang, F. Wang, and G. Chen, Chem. Soc. Rev. **41**, 3821 (2012).

[46]S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, Mol. Phys. **103**, 2255 (2005).

[47]T. Yoshikawa, M. Kobayashi, A. Fujii, and H. Nakai, J. Phys. Chem. B **117**, 5565 (2013).

[48]C. Yam, S. Yokojima, and G. Chen, Phys. Rev. B **68**, 153105 (2003).

[49]F. Wu, W. Liu, Y. Zhang, and Z. Li, J. Chem. Theory Comput. **7**, 3643 (2011).

[50]H. Nakatsuji, T. Miyahara, and R. Fukuda, J. Chem. Phys. **126**, 084104 (2007).

[51]T. Touma, M. Kobayashi, and H. Nakai, Chem. Phys. Lett. **485**, 247 (2010).

[52]M. Miura and Y. Aoki, Mol. Phys. **108**, 205 (2010).

[53]R. A. Mata, H. Stoll, and B. J. C. Cabral, J. Chem. Theory Comput. **5**, 1829 (2009).

[54]Y. Ma and H. Ma, J. Phys. Chem. A **117**, 3655 (2013).

[55]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[56]D. G. Fedorov and K. Kitaura, editors, *The Fragment Molecular Orbital Method: practical*

applications to large molecular systems, CRC press, Boca Raton, FL, 2009.

[57]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A. **111**, 6904 (2007).

[58]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[59]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[60]D. G. Fedorov, Y. Alexeev, and K. Kitaura, J. Phys. Chem. Lett. **2**, 282 (2011).

[61]Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **372**, 342 (2003).

[62]A. W. Lange and G. A. Voth, J. Chem. Theory Comput. **9**, 4018 (2013).

[63]H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, J. Chem. Phys. **138**, 164103 (2013).

[64]H. Sekino, N. Matsumura, and Y. Sengoku, Comput. Lett. **3**, 423 (2007).

[65]Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, Chem. Phys. Lett. **445**, 331 (2007).

[66]Q. Gao, S. Yokojima, D. G. Fedorov, K. Kitaura, M. Sakurai, and S. Nakamura, J. Chem. Theory. Comput. **6**, 1428 (2010).

[67]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[68]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B **111**, 9621 (2007).

[69]A. Devarajan, S. Markutsya, M. H. Lamm, X. Cheng, J. C. Smith, J. Y. Baluyut, Y. Kholod, M. S. Gordon, and T. L. Windus, J. Phys. Chem. B **117**, 10430 (2013).

[70]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B **116**, 503 (2012).

[71]P. V. Avramov, D. G. Fedorov, P. B. Sorokin, S. Sakai, S. Entani, M. Ohtomo, and H. N. Y. Matsumoto, J. Phys. Chem. Lett. **3**, 2003 (2012).

[72]L. Roskop, D. G. Fedorov, and M. S. Gordon, Mol. Phys. **111**, 1622 (2013).

[73]Y. Okiyama, T. Tsukamoto, C. Watanabe, K. Fukuzawa, S. Tanaka, and Y. Mochizuki, Chem. Phys. Lett. **566**, 25 (2013).

[74]D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).

[75]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **116**, 704 (2012).

[76]Y. Mochizuki, K. Fukuzawa, A. Kato, S. Tanaka, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **410**, 247 (2005).

[77]T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, H. Tokiwa, S. Tanaka, and K. Tanaka, Theor. Chem. Acc. **118**, 937 (2007).

[78]M. C. Green, D. G. Fedorov, K. Kitaura, J. S. Francisco, and L. V. Slipchenko, J. Chem. Phys. **138**, 074111 (2013).

[79]S.-I. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, Chem. Phys. Lett. **382**, 611 (2003).

[80]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[81]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).

[82]Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **406**, 283 (2004).

[83]Y. Mochizuki, K. Tanaka, K. Yamashita, T. Ishikawa, T. Nakano, S. Amari, K. Segawa, T. Murase, H. Tokiwa, and M. Sakurai, Theor. Chem. Acc. **117**, 541 (2007).

[84]H. Li, D. G. Fedorov, T. Nagata, K. Kitaura, J. H. Jensen, and M. S. Gordon, J. Comput. Chem. **31**, 778 (2010).

[85]T. Nagata, D. G. Fedorov, and K. Kitaura, Theor. Chem. Acc. **131**, 1136 (2012).

[86]S. R. Pruitt, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Theory Comp. **6**, 1 (2010).

[87]S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, J. Phys. Chem. A **116**, 4965 (2012).

[88]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **122**, 0541081 (2005).

[89]Y. Komeiji, Y. Mochizuki, T. Nakano, and H. Mori, *Recent Advances in Fragment Molecular Orbital-Based Molecular Dynamics(FMO-MD) Simulations.*, InTech, 2012.

[90]H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, J. Chem. Phys. **137**, 044110 (2012).

[91]M. Chiba, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **444**, 346 (2007).

[92]M. Chiba, D. G. Fedorov, and K. Kitaura, J. Chem. Phys. **127**, 104108 (2007).

[93]M. Chiba, D. G. Fedorov, and K. Kitaura, J. Comput. Chem. **29**, 2667 (2008).

[94]M. Chiba, D. G. Fedorov, T. Nagata, and K. Kitaura, Chem. Phys. Lett. **474**, 227 (2009).

[95]M. Chiba and T. Koido, J. Chem. Phys. **133**, 044113 (2010).

[96]H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii, and K. Kitaura, J. Phys. Chem. A **112**,

10887 (2008).

[97]B. F. Milne, M. A. L. Marques, and F. Nogueira, Phys. Chem. Chem. Phys. **12**, 14285 (2010).

[98]Y. Shigemitsu, M. Hagimori, N. Mizuyama, B.-C. Wang, and Y. Tominaga, Dyes Pigment **99**, 940 (2013).

[99]H. Lodish, A. Berk, C. A. Kaiser, M. P. Scott, A. Bretscher, H. Ploegh, and P. Matsudaira, *Molecular Cell Biology Sixth Edition*, Sara Tenney, New York, 2008.

[100]S. Sinnecker, E. Reijerse, F. Neese, and W. Lubitz, J. Am. Chem. Soc. **126**, 3280 (2004).

[101]J. E. Miller, C. Gradinaru, B. R. Crane, A. J. Di Bilio, W. A. Wehbi, S. Un, J. R. Winkler, and H. B. Gray, J. Am. Chem. Soc. **125**, 14220 (2003).

[102]P. Mukerjee, C. Ramachandran, and R. A. Pyter, J. Phys. Chem. **86**, 3189 (1982).

[103]C. Ramachandran, R. A. Pyter, and P. Mukerjee, J. Phys. Chem. **86**, 3198 (1982).

[104]E. I. Solomon, R. K. Szilagyi, S. DeBeer George, and L. Basumallick, Chem. Rev. **104**, 419 (2004).

[105]A. A. Gewirth and E. I. Solomon, J. Am. Chem. Soc. **110**, 3811 (1988).

[106]S. Sinnecker and F. Neese, J. Comput. Chem. **27**, 1463 (2006).

[107]A. Monari, T. Very, J.-L. Rivail, and X. Assfeld, Comput. Theor. Chem. **990**, 119 (2012).

[108]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[109]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[110]Y. Alexeev, M. P. Mazanetz, O. Ichihara, and D. G. Fedorov, Curr. Top. Med. Chem. **12**, 2013 (2012).

[111]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[112]M. S. Gordon and M. W. Schmidt, Advances in electronic structure theory: GAMESS a decade later, in *Theory and Applications of Computational Chemistry, the first forty years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria, pages 1167–1189, Elsevier, Amsterdam, 2005.

[113]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem.

25, 872 (2004).

[114]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[115]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[116]D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. J. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).

[117]M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Ann. Rev. Phys. Chem. **64**, 553 (2013).

[118]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[119]H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).

[120]M. R. Redinbo, D. Cascio, M. K. Choukair, D. Rice, S. Merchant, and T. O. Yeates, Biochemistry **32**, 10560 (1993).

[121]J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[122]D. A. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossvry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, , and P. Kollman, AMBER 12, University of California, San Francisco (2012).

[123]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. A. andJ. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, , and D. J. Fox, Gaussian, Inc., Wallingford CT (2010), Gaussian 09,

Revision C.01.

[124]D. G. Fedorov, T. Ishida, and K. Kitaura, J. Phys. Chem. A **109**, 2638.

[125]C. C. Lovallo and M. Klobukowski, J. Comput. Chem. **24**, 1009 (2003).

[126]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **351**, 475 (2002).

[127]A. Bondi, J. Phys. Chem. **68**, 441 (1964).

[128]C. Steinmann, M. W. Ibsen, A. S. Hansen, and J. H. Jensen, PLOS ONE **7**, e44480 (2012).

[129]See supplementary materials at
ftp : //ftp.aip.org/epaps/journ_chem_phys/E − JCPSA6 − 140 − 034414/ for the details of the gradient accuracy and the important molecular orbitals in solvated benzoquionone.

[130]K. A. Kistler and S. Matsika, J. Phys. Chem. A **113**, 12396 (2009).

[131]Y. Mochizuki, Y. Komeiji, T. Ishikawa, T. Nakano, and H. Yamataka, Chem. Phys. Lett. **437**, 66 (2007).

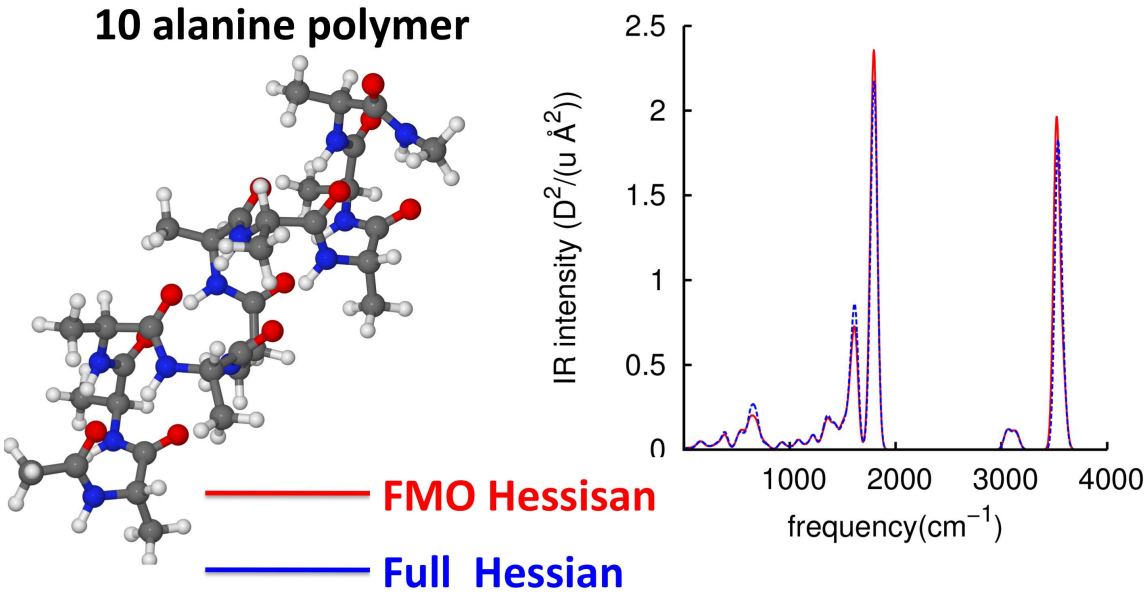
[132]K. Sigfridsson, S. Young, and Ö. Hansson, Biochemistry **35**, 1249 (1996).

[133]K. W. Penfield, R. R. Gay, R. S. Himmelwright, N. C. Eickman, V. A. Norris, H. C. Freeman, and E. I. Solomon, J. Am. Chem. Soc. **103**, 4382 (1981).

[134]E. I. Solomon, J. W. Hare, and H. B. Gray, Proc. Nat. Acad. Sc. U.S.A. **73**, 1379 (1976).

Chapter V

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



H. Nakata, T. Nagata, D. G. Fedorov *et. al.*

J. Chem. Phys. **2013,** *138*, 164103.

DOI :10.1063/1.4800990

View online: http://dx.doi.org/10.1063/1.4800990

V. ANALYTIC SECOND DERIVATIVES OF THE ENERGY IN THE FRAGMENT MOLECULAR ORBITAL METHOD

A. Introduction

Recent years have witnessed a very significant increase in computational hardware, enabling large scale quantum-mechanical calculations of large molecular systems. A number of methods has been developed for this goal, from *ab initio* linear scaling[1–3] and semiempirical methods[4,5] to fragment-based approaches[6–17] recently reviewed[18].

The fragment molecular orbital (FMO) method is a fragment-based approach suggested by Kitaura et al.[19] in 1999 and in the following development[20–22] it has been extended to various kinds of wave functions: second order Møller-Plesset perturbation theory (MP2),[23] coupled cluster,[24] density functional theory (DFT),[25] multiconfigurational self-consistent field,[26] configuration interaction,[27,28] time-dependent DFT[29] and open shell methods.[30,31] The first derivative of the energy with respect to a nuclear coordinate has been formulated for FMO[32–37] and molecular dynamics simulations have been performed.[38–42]

FMO has been applied to protein-ligand binding[43,44] and geometry optimizations.[45–48] Also, FMO has been used to study chemical reactions either using geometries obtained by the quantum mechanics/molecular mechanics (QM/MM) method[49,50] or by performing an expensive series of constrained optimizations along a manually chosen reaction coordinate.[51] However, the need for the analytic second derivatives is apparent. They are necessary to improve the efficiency of geometry optimizations, to provide thermodynamical properties such as zero point energies (ZPE), to search for transition states (TS) in chemical reactions, to evaluate spectroscopic properties such as intensities in infrared (IR) or Raman spectroscopy and for many other purposes.

Considerable efforts have been invested in improving the efficiency of *ab initio* second derivatives of the energy,[52,53] pioneered by Pulay.[54] On the other hand, very few fragment-based methods have analytic second derivatives developed for them[8,9]. Alternatively, analytic second derivatives have been developed for QM/MM[55,56], our own n-layered integrated molecular orbital and molecular mechanics (ONIOM) method[57] and density functional tight

binding[58]. In addition, the partial Hessian vibrational analysis has been developed by Li and Jensen[59] and applied to calculate pKa values.[60] The mobile block Hessian (MBH) model[56,61–63] and the vibrational subsystem analysis (VSA)[64–66] have also been developed. Performance of various partial Hessian approaches has been evaluated by Ghysels et al.[67]

In this paper, we derive analytic second derivatives for the FMO method at the restricted Hartree-Fock (RHF) level, and introduce some approximations for efficiency. The developed method is implemented into GAMESS[68,69] and applied to the calculations of the vibrational frequencies, IR spectrum, ZPE, Gibbs free energy, and the barriers of chemical reactions. The accuracy is evaluated in comparison to conventional *ab initio* results without fragmentation.

B. Theory

1. Summary of the FMO energy gradient

In FMO, one divides the system into fragments and performs *ab initio* calculations of these fragments (monomers) in the embedding electrostatic potential, which depends on the electronic state of each fragment. The monomer calculations are repeated until the embedding potential converges. Consecutively, dimer (pair of fragments) calculations are performed in the embedding potential of monomers fixed at this step. We note that the electronic states of monomers are self-consistent with respect to each other, but the dimer electronic state is not mutually consistent with the embedding potential determined by the monomer densities. More details can be found elsewhere.[22]

The 2-body FMO (FMO2) total energy is given by[70]

$$E = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J}^{N} \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}), \tag{1}$$

where E'_X is the internal energy of fragment X (X=I or IJ, for monomers and dimers, respectively). N is the number of fragments. The dimer density matrix difference $\Delta \mathbf{D}^{IJ}$ in Eq. (1) is defined by

$$\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^{I} \oplus \mathbf{D}^{J}), \tag{2}$$

and $\mathbf{V}^{IJ}$ is the matrix form of the electrostatic potential (ESP) for dimer IJ due to the electron densities and nuclear charges of the fragments other than I and J. The direct sum in Eq. (2) represents a blockwise addition of two monomer matrices into the dimer supermatrix.

$$V_{\mu\nu}^{IJ} = \sum_{K \neq IJ}^{N} \left(u_{\mu\nu}^{K} + v_{\mu\nu}^{K} \right). \tag{3}$$

Here the one-electron u and two-electron v integrals are, respectively,

$$u_{\mu\nu}^{K} = \sum_{A \in K} \langle \mu | \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} | \nu \rangle , \tag{4}$$

$$v_{\mu\nu}^{K} = \sum_{\lambda\sigma \in K} D_{\lambda\sigma}^{K} \left(\mu\nu | \lambda\sigma \right), \tag{5}$$

where $\mathbf{R}_A$ and Z_A are the position and charge of atom A in fragment K, respectively. $D_{\lambda\sigma}^{K}$ is the density matrix element of fragment K and $(\mu\nu|\lambda\sigma)$ is a two-electron integral in the atomic orbital (AO) basis. Throughout this study, the Roman $(ijkl\cdots)$ and Greek $(\mu\nu\rho\sigma\cdots)$ indices denote the molecular orbitals (MOs) and AOs, respectively.

The internal fragment energies in Eq. (1) are

$$E'_X = \sum_{\mu\nu \in X} D_{\mu\nu}^{X} h_{\mu\nu}^{X} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma \in X} \left[D_{\mu\nu}^{X} D_{\lambda\sigma}^{X} - \frac{1}{2} D_{\mu\lambda}^{X} D_{\nu\sigma}^{X} \right] (\mu\nu|\lambda\sigma) + \sum_{\mu\nu \in X} D_{\mu\nu}^{X} P_{\mu\nu}^{X} + E_X^{\mathrm{NR}}, \tag{6}$$

where $h_{\mu\nu}^{X}$ is the one-electron Hamiltonian of X and the nuclear repulsion energy is

$$E_X^{\mathrm{NR}} = \sum_{B \in X} \sum_{A(\in X) > B} \frac{Z_A Z_B}{R_{AB}}, \tag{7}$$

where R_{AB} is the distance between atoms A and B. It is convenient to rewrite Eq. (6) and the ESP contribution in the MO basis:

$$E'_X = \sum_{i \in X}^{\mathrm{occ}} 2h_{ii}^{X} + \sum_{ij \in X}^{\mathrm{occ}} \left[2\left(ii|jj\right) - \left(ij|ij\right) \right] + \sum_{i \in X}^{\mathrm{occ}} 2P_{ii}^{X} + E_X^{\mathrm{NR}}, \tag{8}$$

$$\mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) = 2\left(\sum_{i \in IJ}^{\mathrm{occ}} V_{ii}^{IJ} - \sum_{i \in I}^{\mathrm{occ}} V_{ii}^{IJ} - \sum_{i \in J}^{\mathrm{occ}} V_{ii}^{IJ} \right). \tag{9}$$

In Eq. (9) MOs i of IJ, I and J are used in the three respective terms, arising from Eq. (2). For fragmentation across covalent bonds, the hybrid orbital projection (HOP) contribution

is calculated as

$$\sum_{i\in X}^{\mathrm{occ}} 2P_{ii}^{X} = \sum_{i\in X}^{\mathrm{occ}} 2\left\langle i\right| \widehat{P}^{X} \left|i\right\rangle = \sum_{\mu\nu\in X} D_{\mu\nu}^{X} P_{\mu\nu}^{X}. \tag{10}$$

Here, the HOP operator $\widehat{P}^{X}$ is defined by

$$\widehat{P}^{X} = \sum_{k\in X} B_k \left|\theta_k\right\rangle \left\langle\theta_k\right|, \tag{11}$$

where $|\theta_k\rangle$ is a hybrid orbital and the universal constant B_k is usually set to 10^6 a.u..

The differentiation of E'_X with respect to nuclear coordinate a leads to

$$\begin{aligned}\frac{\partial E'_X}{\partial a} = &\sum_{i\in X}^{\mathrm{occ}} 2h_{ii}^{a,X} + \sum_{ij\in X}^{\mathrm{occ}} \left[2\left(ii|jj\right)^a - \left(ij|ij\right)^a\right] + \sum_{i\in X}^{\mathrm{occ}} 2P_{ii}^{a,X} \\ &+ \sum_{m\in X}^{\mathrm{occ+vir}} \sum_{i\in X}^{\mathrm{occ}} 4U_{mi}^{a,X} F'^{X}_{mi} + \frac{\partial E_X^{\mathrm{NR}}}{\partial a},\end{aligned} \tag{12}$$

where the internal fragment Fock matrix elements F'^{X}_{mi} are given by

$$F'^{X}_{mi} = h_{mi}^{X} + \sum_{k\in X}^{\mathrm{occ}} \left[2(mi|kk) - (mk|ik)\right] + P_{mi}^{X}. \tag{13}$$

In Eq. (12), the superscript a in the integrals denotes AO derivative integrals. The MO-based projection operator matrix P_{ij}^{X} is

$$P_{ij}^{X} = \sum_{\mu\nu\in X} C_{\mu i}^{X*} P_{\mu\nu}^{X} C_{\nu j}^{X}. \tag{14}$$

For the response term (the penultimate term on the right-hand side of Eq. (12)), the following equation defined in the previous study[33] is introduced:

$$\overline{U}^{a,X,Y} = 4\sum_{i\in X}^{\mathrm{occ}} \sum_{r\in X}^{\mathrm{vir}} U_{ri}^{a,X} V_{ri}^{Y}. \tag{15}$$

The response terms $U_{ri}^{a,X}$ associated with ESP V_{ri}^{Y} in the MO basis arise from the expansion of the MO coefficient derivatives in terms of the MO coefficients.[71] There are two types of $\overline{U}^{a,X,Y}$ terms: (a) $\overline{U}^{a,I,I}$ (i.e., $X{=}Y$) arising from the derivative of the monomer terms, and (b) $\overline{U}^{a,X,IJ}$ where X can be I, J or IJ (related to the three $\mathbf{D}$ terms in Eq. (2)). The derivatives of the MO coefficients can be written as

$$\frac{\partial C_{\mu i}^{X}}{\partial a} = \sum_{m\in X}^{\mathrm{occ+vir}} U_{mi}^{a,X} C_{\mu m}^{X}. \tag{16}$$

To obtain the occupied-virtual orbital response $U_{ri}^{a,X}$, one must solve the CPHF equations. This will be discussed in subsequent subsections.

The differentiation of the ESP energy contribution in Eq. (1) with respect to a nuclear coordinate a leads to

$$\begin{aligned}\frac{\partial}{\partial a}\mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) = 2&\left(\sum_{i\in IJ}^{\mathrm{occ}} V_{ii}^{a,IJ} - \sum_{i\in I}^{\mathrm{occ}} V_{ii}^{a,IJ} - \sum_{i\in J}^{\mathrm{occ}} V_{ii}^{a,IJ}\right)\\ &+ 4\left(\sum_{m\in IJ}^{\mathrm{occ+vir}}\sum_{i\in IJ}^{\mathrm{occ}} U_{mi}^{a,IJ}V_{mi}^{IJ} - \sum_{m\in I}^{\mathrm{occ+vir}}\sum_{i\in I}^{\mathrm{occ}} U_{mi}^{a,I}V_{mi}^{IJ} - \sum_{m\in J}^{\mathrm{occ+vir}}\sum_{i\in J}^{\mathrm{occ}} U_{mi}^{a,J}V_{mi}^{IJ}\right)\\ &+ 8\sum_{K\neq IJ}^{\mathrm{occ+vir}}\sum_{m\in K}^{\mathrm{occ}}\sum_{k\in K} U_{mk}^{a,K}\left(\sum_{i\in IJ}^{\mathrm{occ}}(ii|mk) - \sum_{i\in I}^{\mathrm{occ}}(ii|mk) - \sum_{i\in J}^{\mathrm{occ}}(ii|mk)\right).\end{aligned} \tag{17}$$

The collection of all the $\overline{U}^{a,X,Y}$ terms in Eq. (12) and Eq. (17) subject to the differentiation of Eq. (1) can be written as

$$\overline{U}^{a} = -\sum_{I}^{N}\overline{U}^{a,I,I} - \sum_{I>J}^{N}\left(\overline{U}^{a,IJ,IJ} - \overline{U}^{a,I,I} - \overline{U}^{a,J,J}\right) + \sum_{I>J}^{N}\left(\overline{U}^{a,IJ,IJ} - \overline{U}^{a,I,IJ} - \overline{U}^{a,J,IJ}\right). \tag{18}$$

$\overline{U}^{a}$ is zero when no ESP approximations are applied or when all the ESPs are approximated uniformly (e.g., with the ESP-PC approximation).[33] Otherwise, $\overline{U}^{a}$ is only approximately equal to zero. $\overline{U}^{a}$ is the compensation term arising from the use of ESP approximations non-uniformly. If $\overline{U}^{a} = 0$, then in Eq. (17) the dimer-related terms $\overline{U}^{a,IJ,IJ} - \overline{U}^{a,I,IJ} - \overline{U}^{a,J,IJ}$ need not be evaluated, and the only terms that require the solution of CPHF equations come from the monomers, $U_{ri}^{a,I}$. In this work, we do not use ESP approximations and thus $\overline{U}^{a} = 0$.

2. Second derivatives of the FMO total energy

The second derivative of the total FMO energy with respect to nuclear coordinates a and b (the Hessian) is

$$\frac{\partial^2 E}{\partial a\partial b} = \sum_{I}^{N}\frac{\partial^2 E_I'}{\partial a\partial b} + \sum_{I>J}^{N}\left(\frac{\partial^2 E_{IJ}'}{\partial a\partial b} - \frac{\partial^2 E_I'}{\partial a\partial b} - \frac{\partial^2 E_J'}{\partial a\partial b}\right) + \sum_{I>J}^{N}\frac{\partial^2\mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a\partial b}. \tag{19}$$

First, let us consider the internal fragment Hessian contributions appearing in the first and second terms on the right-hand side of Eq. (19):

$$\frac{\partial^2 E'_X}{\partial a \partial b} = \frac{\partial}{\partial b}\left(\frac{\partial E'_X}{\partial a}\right). \tag{20}$$

Inserting the first derivative of the internal energies defined in Eq. (12), one obtains the following equation:

$$\begin{aligned}\frac{\partial^2 E'_X}{\partial a \partial b} = \frac{\partial}{\partial b}\Bigg[&\sum_{i\in X}^{\mathrm{occ}} 2h_{ii}^{a,X} + \sum_{ij\in X}^{\mathrm{occ}} \left[2\,(ii|jj)^a - (ij|ij)^a\right] + \sum_{i\in X}^{\mathrm{occ}} 2P_{ii}^{a,X} \\ &+ \sum_{m\in X}^{\mathrm{occ+vir}} \sum_{i\in X}^{\mathrm{occ}} 4U_{mi}^{a,X} F'^{X}_{mi} + \frac{\partial E_X^{\mathrm{NR}}}{\partial a}\Bigg].\end{aligned} \tag{21}$$

Note that ESP related terms arise in the second derivative of the internal fragment energy. This is because the variation condition results in the diagonal Fock matrix composed of the internal and ESP contributions:

$$F_{ij}^X = F'^{X}_{ij} + V_{ij}^X = \delta_{ij}\epsilon_i^X. \tag{22}$$

Using the relationship, we separate the second derivative of the internal fragment energy in Eq. (21) into the conventional Hessian terms and FMO specific terms. The detailed derivation is provided for completeness elsewhere.[72] The second derivatives of E'_X are

$$\begin{aligned}\frac{\partial^2 E'_X}{\partial a \partial b} =& \sum_{i\in X}^{\mathrm{occ}} \left[h_{ii}^{ab,X} + P_{ii}^{ab,X} + F'^{ab,X}_{ii}\right] \\ &- \sum_{i\in X}^{\mathrm{occ}} 2S_{ii}^{ab,X}\epsilon_{ii}^X + 4\sum_{i\in X}^{\mathrm{occ}}\sum_{j\in X}^{\mathrm{occ}} S_{ji}^{b,X} S_{ij}^{a,X}\epsilon_{ii}^X \\ &+ \sum_{m\in X}^{\mathrm{vir}}\sum_{i\in X}^{\mathrm{occ}} U_{mi}^{b,X}\Big[4F'^{a,X}_{im} - 4S_{mi}^{a,X}\epsilon_{ii}^X \\ &- 2\sum_{j,l\in X}^{\mathrm{occ}} A'_{jl,mi}S_{jl}^{a,X}\Big] \\ &- \sum_{i\in X}^{\mathrm{occ}}\sum_{j\in X}^{\mathrm{occ}} S_{ij}^{b,X}\left[2F'^{a,X}_{ij} - \frac{1}{2}\sum_{k,l\in X}^{\mathrm{occ}} A'_{ij,kl}S_{kl}^{a,X}\right] \\ &- \sum_{i\in X}^{\mathrm{occ}}\sum_{j\in X}^{\mathrm{occ}} S_{ij}^{a,X}\left[2F'^{b,X}_{ij} - \frac{1}{2}\sum_{k,l\in X}^{\mathrm{occ}} A'_{ij,kl}S_{kl}^{b,X}\right] \\ &+ \frac{\partial^2 E_X^{\mathrm{NR}}}{\partial a \partial b} - \overline{U}^{ab,X},\end{aligned} \tag{23}$$

where

$$F'^{ab,X}_{mi} = h^{ab,X}_{mi} + \sum_{j\in X}^{\text{occ}} \left[2\,(mi|jj)^{ab} - (mj|ij)^{ab} \right] + P^{ab,X}_{mi}, \tag{24}$$

$$A'_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk), \tag{25}$$

$$P^{ab,X}_{ii} = \sum_{\mu,\nu} C_{\mu i} C_{\nu i} P^{ab,X}_{\mu\nu}. \tag{26}$$

All the above terms appear in the conventional *ab initio* Hessian calculation except the hybrid projection operators P and the last term $\overline{U}^{ab,X}$, which comes from the ESP:

$$\begin{aligned}\overline{U}^{ab,X} = & -4\sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}}\sum_{m\in X}^{\text{vir}} U^{b,X}_{mi}(V^X_{mj}S^{a,X}_{ij} + V^X_{ij}S^{a,X}_{jm}) \\ & + 4\sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}}\sum_{k\in X}^{\text{occ}} S^{b,X}_{ki}V^X_{ij}S^{a,X}_{jk} - \sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}} 2S^{ab,X}_{ij}V^X_{ij} \\ & + \sum_{m\in X}^{\text{vir}}\sum_{i\in X}^{\text{occ}} 4\frac{\partial(V^X_{mi}U^{a,X}_{mi})}{\partial b}.\end{aligned} \tag{27}$$

Here one can see explicitly, how ESPs contribute to the second derivatives of the internal energies. Because a and b run over the coordinates of the entire system, there are three types of matrix elements: those for which both a and b are in X, those for which only one of them is in X and the other is ESP related, and those for which both a and b are not in X, resulting from the double derivative of ESPs coupled to the electronic state of X (FIG. 1).

3. Second derivative of the electrostatic potential energy contribution

The second derivative of the electrostatic potential energy representing the interfragment density (i.e., charge transfer) matrix $\Delta\mathbf{D}^{IJ}$ coupled with the embedding potential $\mathbf{V}^{IJ}$ is given by

$$\frac{\partial^2}{\partial a\partial b}\mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) = \frac{\partial}{\partial b}\sum_{I>J}^{N}\left(2\sum_{i\in IJ}^{\text{occ}}\frac{\partial V^{IJ}_{ii}}{\partial a} - 2\sum_{i\in I}^{\text{occ}}\frac{\partial V^{IJ}_{ii}}{\partial a} - 2\sum_{i\in J}^{\text{occ}}\frac{\partial V^{IJ}_{ii}}{\partial a}\right). \tag{28}$$

All three terms in Eq. (28) have the same form and can be generalized for compactness using $X = I$, J or IJ. Inserting Eq. (17) into the above equation, one can obtain the following

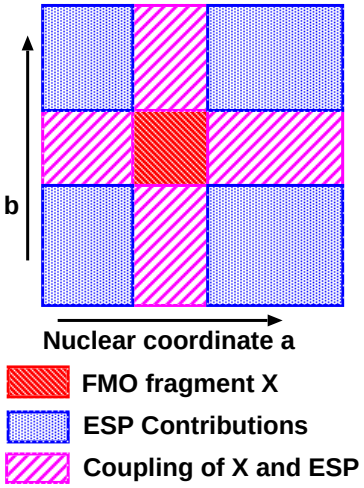


FIG. 1. Structure of the Hessian matrix of second derivatives with respect to nuclear coordinates a and b in FMO, depending on whether 2, 1 or 0 coordinates out of a and b belong to fragment X, whose second derivative is calculated and added to the total Hessian. Because of the electrostatic potential (ESP), there are Hessian contributions for coordinates outside of X.

equation:

$$2\sum_{i\in X}^{\mathrm{occ}}\frac{\partial^2 V_{ii}^{IJ}}{\partial a\partial b}=\frac{\partial}{\partial b}\left\{2\sum_{i\in X}^{\mathrm{occ}}V_{ii}^{a,IJ}+4\sum_{m\in X}^{\mathrm{occ+vir}}\sum_{i\in X}^{\mathrm{occ}}U_{mi}^{a,X}V_{mi}^{IJ}\right.$$
$$\left.+8\sum_{K\neq IJ}\sum_{m\in K}^{\mathrm{occ+vir}}\sum_{k\in K}^{\mathrm{occ}}U_{mk}^{a,K}\sum_{i\in X}^{\mathrm{occ}}(ii|mk)\right\}. \tag{29}$$

The derivation of the second derivative of ESP is not straightforward; there are several ways depending on how we treat the unknown response terms $U_{mi}^{a,X}$. Here, we separate them into ESP terms, its first derivative, its second derivative and the remaining two-electron integrals. Then we separate the response terms $U_{mi}^{a,X}$ and $S_{ij}^{a,X}$. After some algebra,[72] the

final expression is obtained

$$
\begin{aligned}
2\sum_{i\in X}^{\text{occ}}\frac{\partial^2 V_{ii}^{IJ}}{\partial a\partial b} &= 2\sum_{i\in X}^{\text{occ}} V_{ii}^{ab,IJ} - 2\sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}} S_{ij}^{ab,X} V_{ij}^{IJ} \\
&+ 4\sum_{i,j\in X}^{\text{occ}}\sum_{k\in X}^{\text{occ}} S_{ki}^{b,X} S_{ij}^{a,X} V_{jk}^{IJ} \\
&- 2\sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}} S_{ij}^{a,X} V_{ij}^{b,IJ} \\
&- 2\sum_{i\in X}^{\text{occ}}\sum_{j\in X}^{\text{occ}} V_{ij}^{a,IJ} S_{ji}^{b,X} \\
&+ 4\sum_{i\in X}^{\text{occ}}\sum_{m\in X}^{\text{vir}} U_{mi}^{b,X}\Big\{V_{im}^{a,IJ} - \sum_{j}^{\text{occ}}(S_{jm}^{a,X} V_{ij}^{IJ} \\
&+ S_{ij}^{a,X} V_{mj}^{IJ})\Big\} \\
&+ 4\sum_{i\in X}^{\text{occ}}\sum_{m\in X}^{\text{vir}} \frac{\partial(U_{mi}^{a,X} V_{im}^{IJ})}{\partial b} + R^{ab,X}. \qquad (30)
\end{aligned}
$$

Note that $\overline{U}^{ab,X}$ defined in Eq. (27) comes with a minus sign in Eq. (23) and it cancels out with the same terms in Eq. (30). Eq. (30) involves the entire Hessian matrix with a and b

in the whole system. The $R^{ab,X}$ is a response term given by

$$
\begin{aligned}
R^{ab,X} = & 8 \sum_{m\in X}^{\mathrm{vir}} \sum_{i\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{j\in K}^{\mathrm{occ}} U_{mi}^{b,X} \left[\sum_{n\in K}^{\mathrm{vir}} 2(im|jn) U_{nj}^{a,K} - \sum_{k\in K}^{\mathrm{occ}} S_{jk}^{a,K} (im|jk) \right] \\
& + 4 \sum_{i\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{j\in K}^{\mathrm{occ}} \sum_{m\in K}^{\mathrm{vir}} U_{mj}^{a,K} \left[2(ii|jm)^b - \sum_{k\in X}^{\mathrm{occ}} 2 S_{ik}^{b,X} (ik|jm) - \sum_{k\in K}^{\mathrm{occ}} 2 S_{jk}^{b,K} (ii|km) \right. \\
& \left. -2 \sum_{k\in K}^{\mathrm{occ}} S_{km}^{b,K} (ii|jk) + \sum_{n\in K}^{\mathrm{vir}} U_{nj}^{b,K} (ii|nm) - \sum_{k\in K}^{\mathrm{occ}} U_{mk}^{b,K} (ii|jk) \right] \\
& + 4 \sum_{i\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{j\in K}^{\mathrm{occ}} \sum_{m\in K}^{\mathrm{vir}} U_{mj}^{b,K} \left[2(ii|jm)^a - \sum_{k\in X}^{\mathrm{occ}} 2 S_{ki}^{a,X} (ik|jm) - 2 \sum_{k\in K}^{\mathrm{occ}} S_{kj}^{a,K} (ii|km) \right. \\
& \left. -2 \sum_{k\in K}^{\mathrm{occ}} S_{km}^{a,K} (ii|jk) + \sum_{n\in K}^{\mathrm{vir}} U_{nj}^{a,K} (ii|nm) - \sum_{k\in K}^{\mathrm{occ}} U_{mk}^{a,K} (ii|jk) \right] \\
& + 4 \sum_{i\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{k\in K}^{\mathrm{occ}} \sum_{l\in K}^{\mathrm{occ}} S_{kl}^{a,K} \left[\sum_{j\in X}^{\mathrm{occ}} (ij|kl) S_{ij}^{b,X} - (ii|kl)^b + \sum_{j\in K}^{\mathrm{occ}} S_{jl}^{b,K} (ii|jk) \right] \\
& + 4 \sum_{i\in X}^{\mathrm{occ}} \sum_{K\neq IJ} \sum_{k\in K}^{\mathrm{occ}} \sum_{l\in K}^{\mathrm{occ}} S_{kl}^{b,K} \left[\sum_{j\in X}^{\mathrm{occ}} (ij|kl) S_{ij}^{a,X} - (ii|kl)^a + \sum_{j\in K}^{\mathrm{occ}} S_{jl}^{a,K} (ii|jk) \right] \\
& - 4 \sum_{K\neq IJ} \sum_{i\in X}^{\mathrm{occ}} \sum_{j\in K}^{\mathrm{occ}} \sum_{k\in K}^{\mathrm{occ}} S_{jk}^{ab,K} (ii|jk) \\
& - 8 \sum_{K\neq IJ} \sum_{i\in X}^{\mathrm{occ}} \sum_{j\in K}^{\mathrm{occ}} \sum_{m\in K}^{\mathrm{vir}} \sum_{n\in K}^{\mathrm{occ}} U_{mn}^{b,K} U_{nj}^{a,K} (ii|jm) \\
& - 8 \sum_{K\neq IJ} \sum_{i\in X}^{\mathrm{occ}} \sum_{j\in K}^{\mathrm{occ}} \sum_{m\in K}^{\mathrm{vir}} \sum_{n\in K}^{\mathrm{occ}} U_{mn}^{a,K} U_{nj}^{b,K} (ii|jm) \\
& + 8 \sum_{K\neq IJ} \sum_{i\in X}^{\mathrm{occ}} \sum_{j\in K}^{\mathrm{occ}} \sum_{m\in K}^{\mathrm{vir}} U_{mj}^{ab,K} (ii|jm). \qquad (31)
\end{aligned}
$$

Inserting Eq. (23), (27), (30), and (31) in the second derivative of the total FMO energy in Eq. (19), one obtains the fully analytic second derivative. Consequently, it is necessary to calculate three types of unknown response terms $U_{mi}^{a,I}$, $U_{mi}^{a,IJ}$ and $U_{mi}^{ab,I}$. The $U_{mi}^{ab,I}$ is calculated from the second order CPHF equations for monomers.

The derivation of the above equations is fully analytic. However, our current implementation is not fully analytic and includes two approximations for computational efficiency. The approximations reduce the computational cost related to the large dimension of CPHF

equation. First, we omit solving the second order CPHF equation of monomers by neglecting the last three terms in Eq. (31). The second approximation is related to the $U_{mi}^{a,I}$ and $U_{mi}^{a,IJ}$ terms which are obtained by solving the first order CPHF equations of monomers and dimers. The approximation for CPHF equation for monomers and dimers are discussed in the next subsection. The effects are expected to be small, considering that the response terms in the analytic energy gradients are small.[35]

Finally, all fragment contributions are combined in the total Hessian matrix depicted in FIG. 1 representing the whole system (atoms a and b run through all atoms). Because FMO can be applied to the systems containing many thousands of atoms, storing and diagonalizing this matrix can become a serious issue. In this work, we use a straightforward nondistributed diagonalization, which takes relatively little time in comparison to other steps. Nevertheless, it is a cubically scaling step, which in future will have to be improved, for instance, along the lines of partial techniques.[46,47,67] Practically, our present implementation can be used without difficulties for systems containing several thousands of atoms: for example, for 3333 atoms storing the real symmetric matrix and its eigenvectors takes about $(3/2) \times 10^8 \times 8$ bytes of RAM, or 1.2 GB.

4. Approximations in CPHF equations

For the FMO energy gradient at the RHF level[35] the first order CPHF equation is expressed in terms of the orbital responses of monomer fragments, because in Eq. (18) the terms containing dimer responses $U_{mi}^{a,IJ}$ (i.e., $\overline{U}^{a,IJ,IJ}$) cancel out. For the FMO Hessian calculation, however, one should solve both monomer and dimer CPHF equations. The differentiation of the Fock matrix element with respect to a nuclear coordinate a in the MO basis gives

$$\frac{\partial F_{ij}^X}{\partial a} = F_{ij}^{a,X} - \left(\epsilon_j^X - \epsilon_i^X\right) U_{ij}^{a,X} - S_{ij}^{a,X}\epsilon_j^X - \frac{1}{2}\sum_{k\in X}^{\mathrm{occ}}\sum_{l\in X}^{\mathrm{occ}} S_{kl}^{a,X} A_{ij,kl}^{X,X} + \sum_{k\in X}^{\mathrm{vir}}\sum_{l\in X}^{\mathrm{occ}} U_{kl}^{a,X} A_{ij,kl}^{X,X}$$
$$- \frac{1}{2}\sum_{K\neq X}\sum_{k\in K}^{\mathrm{occ}}\sum_{l\in K}^{\mathrm{occ}} S_{kl}^{a,K} A_{ij,kl}^{X,K} + \sum_{K\neq X}\sum_{k\in K}^{\mathrm{vir}}\sum_{l\in K}^{\mathrm{occ}} U_{kl}^{a,K} A_{ij,kl}^{X,K}, \qquad (32)$$

where

$$A_{ij,kl}^{X,X} = 4(ij|kl) - (ik|jl) - (il|jk), \tag{33}$$

$$A_{ij,kl}^{X,K} = 4(ij|kl). \tag{34}$$

The first term on the right-hand side of Eq. (32) describes the Fock derivative term:

$$F_{ij}^{a,X} = h_{ij}^{a,X} + V_{ij}^{a,X} + \sum_{k\in X}^{\text{occ}} [2(ij|kk)^a - (ik|jk)^a] + P_{ij}^{a,X}. \tag{35}$$

The definitions of derivative integral terms in Eq. (35) are given elsewhere.[35,71] From Eq. (32), the following equation can be derived:[73]

$$U_{ij}^{a,X} = \frac{T_{ij}^{a,X}}{\epsilon_j^X - \epsilon_i^X}, \tag{36}$$

where

$$T_{ij}^{a,X} = B_{0,ij}^{a,X} + \sum_{k\in X}^{\text{vir}} \sum_{l\in X}^{\text{occ}} U_{kl}^{a,X} A_{ij,kl}^{X,X} + \sum_{K\neq X} \sum_{k\in K}^{\text{vir}} \sum_{l\in K}^{\text{occ}} U_{kl}^{a,K} A_{ij,kl}^{X,K}. \tag{37}$$

$B_{0,ij}^{a,X}$ in Eq. (37) includes the derivative integral terms in Eq. (32):

$$B_{0,ij}^{a,X} = F_{ij}^{a,X} - S_{ij}^{a,X}\epsilon_j^X - \frac{1}{2}\sum_{k\in X}^{\text{occ}} \sum_{l\in X}^{\text{occ}} S_{kl}^{a,X} A_{ij,kl}^{X,X} - \frac{1}{2}\sum_{K\neq X} \sum_{k\in K}^{\text{occ}} \sum_{l\in K}^{\text{occ}} S_{kl}^{a,K} A_{ij,kl}^{X,K}. \tag{38}$$

Here, consider the CPHF equations for obtaining monomer orbital responses. Both F_{ij}^X and $\partial F_{ij}^X/\partial a$ are zero for $i \neq j$ and the CPHF equations can be constructed as

$$\mathbf{A}'\mathbf{U}^a = \mathbf{B}_0^a, \tag{39}$$

where

$$A_{ij,kl}^{\prime I,I} = \delta_{ik}\delta_{jl}\left(\epsilon_j^I - \epsilon_i^I\right) - A_{ij,kl}^{I,I}, \tag{40}$$

$$A_{ij,kl}^{\prime I,K} = -A_{ij,kl}^{I,K}. \tag{41}$$

Eq. (39) has the dimension of the entire molecular system. In the practical Hessian calculation, Eq. (39) is decoupled into the set of monomer fragment CPHF equations as

$$\mathbf{A}'^{I,I}\mathbf{U}^{a,I} = \mathbf{B}_0'^{a,I}, \tag{42}$$

where we neglected the terms involving the derivatives for $a \notin X$,

$$\begin{aligned} B'^{a,I}_{0,ij} =& B^{a,I}_{0,ij} + \sum_{K\neq I}^{N} \sum_{k\in K}^{\mathrm{vir}} \sum_{l\in K}^{\mathrm{occ}} U^{a,K}_{kl} A^{I,K}_{ij,kl} \\ \approx& B^{a,I}_{0,ij} = F^{a,X}_{ij} - S^{a,X}_{ij}\epsilon^{X}_{j} \\ & -\frac{1}{2}\sum_{k\in X}^{\mathrm{occ}}\sum_{l\in X}^{\mathrm{occ}} S^{a,X}_{kl} A^{X,X}_{ij,kl} - \frac{1}{2}\sum_{K\neq X}\sum_{k\in K}^{\mathrm{occ}}\sum_{l\in K}^{\mathrm{occ}} S^{a,K}_{kl} A^{X,K}_{ij,kl} \\ \approx& F'^{a,I}_{ij} - S^{a,I}_{ij}\epsilon^{I}_{j} - \frac{1}{2}\sum_{k\in I}^{\mathrm{occ}}\sum_{l\in I}^{\mathrm{occ}} S^{a,I}_{kl} A^{I,I}_{ij,kl}. \end{aligned} \tag{43}$$

Therefore, the approximate CPHF equation can be solved independently for each monomer.

The set of CPHF equations for dimer IJ (also arising[74] in the FMO-MP2, but not in the FMO-RHF gradient) is also solved approximately as,

$$\mathbf{A}'^{IJ,IJ}\mathbf{U}^{a,IJ} = \mathbf{B}'^{a,IJ}_{0}, \tag{44}$$

where

$$B'^{a,IJ}_{0,ij} \approx F'^{a,IJ}_{ij} - S^{a,IJ}_{ij}\epsilon^{IJ}_{j} - \frac{1}{2}\sum_{k\in IJ}^{\mathrm{occ}}\sum_{l\in IJ}^{\mathrm{occ}} S^{a,IJ}_{kl} A^{IJ,IJ}_{ij,kl}. \tag{45}$$

Eq. (44) can be derived from Eq. (32). These unknown response terms are calculated and used in Eqs. (23), (27), (30) and (31).

5. Covalent bond fragment boundaries

In this work, we use the hybrid orbital projection (HOP) operator (Eq. (11)) for the fragmentation across covalent bonds, which projects out the redundant orbitals arising from the fact that one atom in each detached bond is placed in two fragments.[34] The orbital energies of the projected-out orbitals are approximately equal to $B = 10^6$ a.u., and these orbitals are effectively removed from the variational space.

In this work, we do not include the projected-out virtual orbitals in the set of CPHF equations. Thus, we separate the response matrix $\mathbf{U}^a$ into the three regions: occupied , virtual and the projected out orbitals (FIG. 2). Then the number of the necessary response terms is the product of the numbers of virtual and occupied molecular orbitals. In addition,

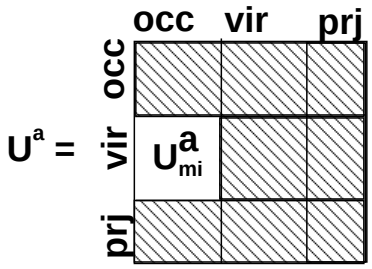


FIG. 2. Orbital response matrix $\mathbf{U}^a$, where occ, vir and prj denote the occupied, virtual and projected out molecular orbitals, respectively.

in our present implementation we did not include HOP terms in the second derivative of internal energies in Eqs. (21) and (23).

Our present derivations apply only to the HOP treatment of fragment boundaries. Some additional work will be required to derive the equations for the adaptive frozen orbital (AFO) bond detachment scheme[75,76].

6. Infrared intensities

In order to define IR intensities I_i, one has to calculate the dipole derivatives μ^a (vectors with x, y and z components).

$$I_i = \left| \sum_a \mu^a Q_{ai} \right|^2 \tag{46}$$

where Q_{ai} is the normal mode i (i.e., Q are eigenvectors of the Hessian). In FMO, we use

$$\mu^a = \sum_I^N \mu^{a,I} + \sum_{I>J}^N \left(\mu^{a,IJ} - \mu^{a,I} - \mu^{a,J} \right), \tag{47}$$

and for each X ($X = I$, J or IJ) we only compute the contributions for $a \in X$, which is an approximation.

C. Computational details

We implemented the analytic second derivatives with several approximations in the FMO code[77,78] in a development version of GAMESS,[68,69] and parallelized them with the general-

ized distributed data interface (GDDI).[79] We did not use ESP approximations or separated dimer approximations in this study. All structures were optimized with the fully analytic energy gradients[35] so that the reported results are for the Hessians computed at the energy minima.

We used a small water cluster $(H_2O)_{10}$ and capped polypeptides MeCO-$(Ala)_{10}$-NHMe polypeptides of 10 alanine residues (112 atoms) forming the α-helix, β-strand, and the extended configuration, denoted as $\alpha(ALA)_{10}$, $\beta(ALA)_{10}$ and $e(ALA)_{10}$, respectively (see FIG. 3). The effects of fragmentation on the accuracy were examined by comparing the results obtained with one and two residues (water molecules) per fragment.

All calculations in this work were performed at the RHF level with the 6-31G(d) basis set with spherical harmonics (ISPHER=1). For both the *ab initio* and FMO Hessians calculations, we used the default CPHF options, solving the CPHF equations with the conjugate gradient method (SOLVER=CONJG) forming the response equations from AO integrals (CPHF=AO). Throughout this work, we used the standard means of statistical thermodynamics to evaluate the Gibbs free energy for a single minimum on the potential energy surface, containing the translational, rotational (rigid rotor) and vibrational (harmonic with unscaled frequencies) components.

Next, we performed a TS search and demonstrate the usefulness of the FMO Hessian for such applications. In this work, we interfaced the standard saddle point driver in GAMESS to FMO (RUNTYP=SADPOINT) and used the default set of options.

We investigated the S_N2 reaction of CH_3X and OH^- in a droplet of 41 water molecules, where X is Cl, Br and F. Then we evaluated the free energy barrier and examine the imaginary mode. The fragmentation of this system was done as follows: three water molecules forming hydrogen bonds with the solute (reactants) were put into one fragment, and the rest of solvent molecules were paired by assigning two neighboring molecules per fragment. The computational time is presented for a cubic water cluster of the linear size from 7 to 10 $\mathring{A}$, containing 51, 93, 132 and 162 atoms and divided into 1 water molecule per fragment.

The initial structure for TS search and water cluster calculations were optimized using Amber force field[80] in NWChem 6.1[81], and then reoptimized with FMO in GAMESS. For making the FMO input files, we used Fragit program.[82]

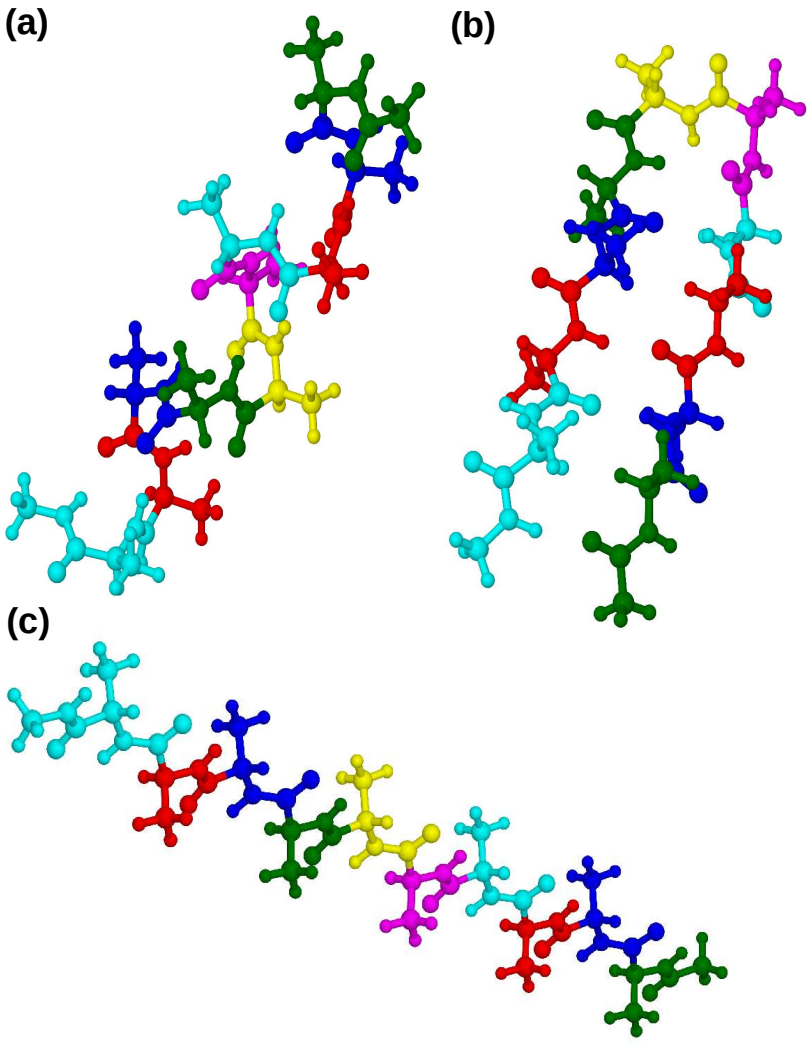


FIG. 3. Polyalanine structures: (a) α-helix $\alpha(\mathrm{ALA})_{10}$ (b) β-turn $\beta(\mathrm{ALA})_{10}$ (c) extended form $\mathrm{e}(\mathrm{ALA})_{10}$.

D. Results and discussion

1. IR spectrum with FMO

The results shown in FIG. 4 and TABLE I demonstrate that IR spectra are well reproduced by FMO. The largest error is observed in the calculation of ten water molecules; the root-mean-square deviation (RMSD) between FMO and *ab initio* frequencies is reduced from 17.9 to 7.4 cm^{-1} by combining two water molecules in fragment, while the maximum

TABLE I. RMSD and the maximum (Max) difference of vibrational frequencies (cm^{-1}) obtained with FMO and *ab initio* Hessians at the RHF/6-31G(d) level. The fragmentation of m water molecules or amino acid residues per fragment is used.

system	m	RMSD	Max
$(H_2O)_{10}$	1	17.9	49.9
$(H_2O)_{10}$	2	7.4	20.2
$\alpha(ALA)_{10}$	1	6.8	30.1
$\alpha(ALA)_{10}$	2	2.7	12.2
$\beta(ALA)_{10}$	1	4.2	16.0
$\beta(ALA)_{10}$	2	1.5	9.4
$e(ALA)_{10}$	1	5.4	34.7
$e(ALA)_{10}$	2	0.5	4.4

difference is reduced from 49.9 to 20.2 cm^{-1}.

The largest errors among the three computed isomers of polyalanine are found in $\alpha(ALA)_{10}$, which has strong many-body effects such as the three-body hydrogen bond coupling (e.g., the coupling of two hydrogen bonds involving three fragments). By merging two adjacent residues into the same fragment, the errors are reduced from 6.8 to 2.7 cm^{-1} (RMSD) and 30.1 to 12.2 cm^{-1} (Max). Other alanine isomers have a similar trend and we obtained RMSD less than 3 cm^{-1} for all three isomers for the division of 2 residues per fragment.

We also note that compared to good accuracy in reproducing vibrational frequencies, FMO shows a somewhat larger error in the IR intensities. Although most of them are in very good agreement with *ab initio* (see FIG. 4), some deviate by 10-30%, but the error is considerably reduced when larger fragments are used. This is because the energy-related frequencies are in general easier to reproduce than intensities (Eq. (46)). The latter are related to eigenvectors of the Hessian matrix. The normal modes can have a considerable delocalization and thus it is more difficult to accurately calculate them in fragment-based calculations. On the other hand, we got a very good accuracy for spatially localized normal

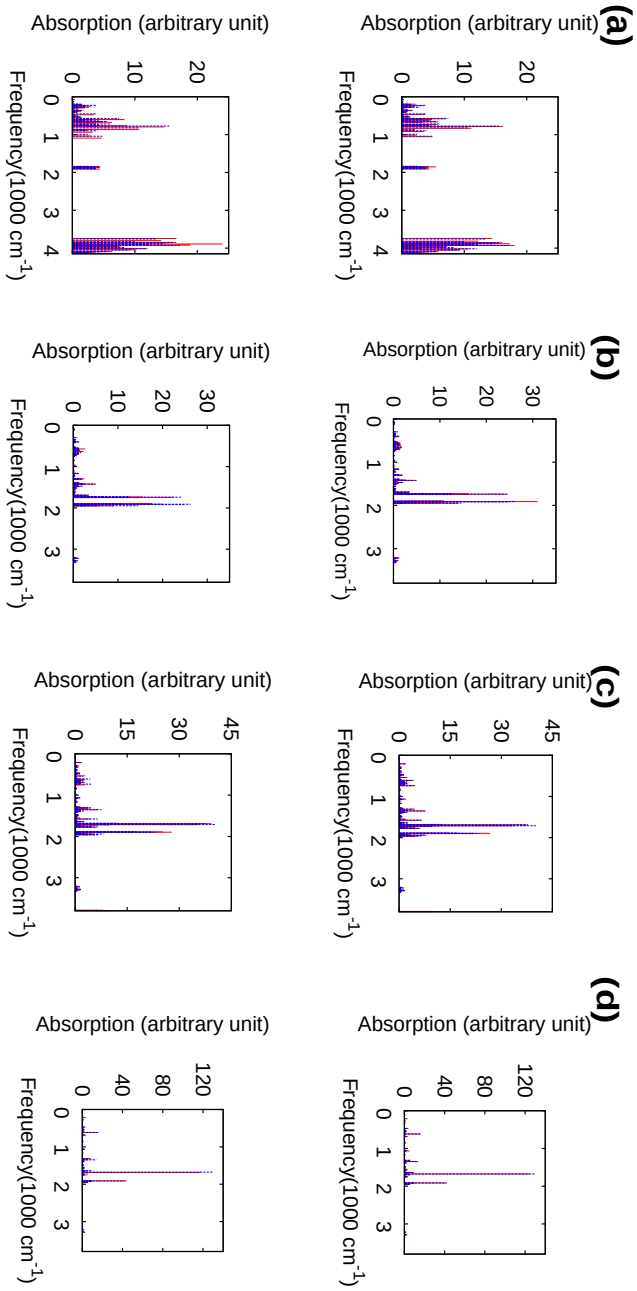


FIG. 4. IR spectra calculated with RHF/6-31G(d) for (a) ten water molecules, (b) $\alpha(\mathrm{ALA})_{10}$, (c) $\beta(\mathrm{ALA})_{10}$ and (d) e$(\mathrm{ALA})_{10}$. The top and bottom spectra show FMO results for 2 and 1 residues per fragment, respectively. FMO and *ab initio* results (using unscaled frequencies for both) are depicted in red solid and blue dotted lines, respectively.

modes, which typically have prominent IR peaks. In this case, the environmental effect is mainly the polarization and FMO can accurately describe it.

The characteristic bands at 1600-1700, 1500-1600, and 1200-1300 cm^{-1} are due to amide I (C=O stretch), amide II (N-H bending coupled with a C-N stretch), and amide III (N-H bending and a C-N stretch) bands, respectively.[83] They are affected by the types of amino acids as well as the protein conformation, and also by the formation of intramolecular hydrogen bonds. These bands are considered to be useful fingerprints to identify dynamic changes during enzymatic reactions.

TABLE II. Calculated infrared peaks for polyalanine at the RHF/6-31G(d) level. The scaling factor of 0.8953 is used for all frequencies.

isomers	Amide I	Amide II	Amide III
	C=O	N-H and C-N	N-H
		FMO (one residue per fragment)	
$\alpha(ALA)_{10}$	1714.1	1561.1	1274.0
$\beta(ALA)_{10}$	1696.1	1527.8	1213.7
$e(ALA)_{10}$	1716.8	1499.9	1203.8
		FMO (two residues per fragment)	
$\alpha(ALA)_{10}$	1711.6	1557.6	1273.6
$\beta(ALA)_{10}$	1696.3	1533.6	1209.6
$e(ALA)_{10}$	1714.6	1500.8	1205.4
		ab initio	
$\alpha(ALA)_{10}$	1711.3	1557.2	1272.1
$\beta(ALA)_{10}$	1696.3	1532.8	1211.5
$e(ALA)_{10}$	1715.7	1500.2	1205.1

The positions of the three prominent peaks (FIG. 5) are summarized in TABLE II, where we used the scaling factor[84] of 0.8953 in order to relate them to observed values. The calculated peaks are within the expected ranges listed above. As found in TABLE II, the frequencies obtained by FMO accurately reproduced those by *ab initio* calculations. One

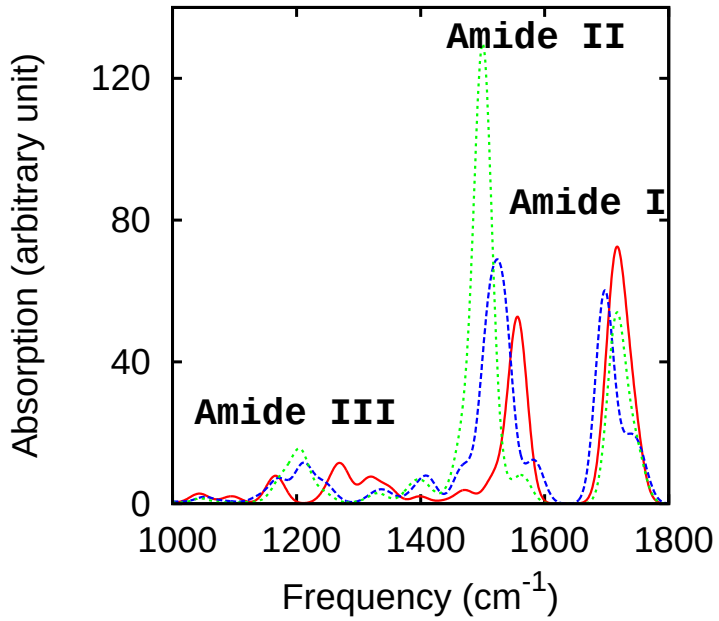


FIG. 5. Superposition of the IR spectra obtained by *ab initio* RHF/6-31G(d) calculations of $\alpha(\mathrm{ALA})_{10}$ (red solid line), $\beta(\mathrm{ALA})_{10}$ (blue dashed line) and $\mathrm{e}(\mathrm{ALA})_{10}$ (green dotted line). The frequencies are scaled[84] by 0.8953.

can see that the fragmentation of 1 residue per fragment has satisfactory accuracy with the largest error of 5.0 cm^{-1}, reduced to only 1.9 cm^{-1} by assigning two residues per fragment.

It is interesting to observe the effect of the secondary structure on the peak positions (TABLE II). The amide I peak for the α-helix and the extended form coincides within 4.4 cm^{-1}, and the values for both differs from the β-turn by about 15-19 cm^{-1} (at the *ab initio* level). The amide II peak of 1500.2 cm^{-1} in $\mathrm{e}(\mathrm{ALA})_{10}$ is shifted up by 57.0 and 32.6 cm^{-1} in $\alpha(\mathrm{ALA})_{10}$ and $\beta(\mathrm{ALA})_{10}$, respectively. The amide III peak and other smaller peaks in the range of 1000-1450 cm^{-1} show a large variation between the three isomers: the amide III peak for the β-turn and the extended form coincide within 5.6 cm^{-1}, and differ from the α-helix by about 61-67 cm^{-1}. It may be possible to identify the secondary structure of polypeptides by their IR spectra, although in moieties made from various amino acids the

peak structure is likely to be more complicated.

The environment not only shifts the peaks but it also affects the intensities. One can see from FIG. 5, that the three isomers have very different intensities, determined by the gradient of the dipole moment. In particular, the extended form of polyalanine has a very large intensity at 1500.2 cm^{-1} possibly related to the alternating dipole moments of adjacent residues.[85]

2. ZPE and Gibbs free energy with FMO

The summary of zero point energies is shown in the TABLE III. The maximum FMO error vs *ab initio* (about 2.0 kcal/mol) is observed in $\alpha(ALA)_{10}$, for one residue per fragment, and for two residues per fragment the largest ZPE error is 0.3 kcal/mol for the same isomer, whereas for the other two isomers FMO results agree to ab initio exactly as shown. The reason why the α-helix has a larger error is that in general it has larger many-body effects leading to a larger error in FMO.[77]

TABLE III. Zero point energy (kcal/mol) at the RHF/6-31G(d) level.

system	*ab initio*	FMO 1frg[a]	FMO 2frg[a]
$\alpha(ALA)_{10}$	639.1	641.1	639.4
$\beta(ALA)_{10}$	638.6	639.2	638.6
$e(ALA)_{10}$	636.4	637.1	636.4

[a] 1frg and 2frg stand for the fragmentation into 1 and 2 amino acid residues per fragment, respectively.

It is interesting to compare the values of ZPE for the three isomers. ZPE decreases in the order of the α-helix, β-turn and the extended form. This is related to somewhat larger frequencies in the more tightly bound isomers. Li et al.[86] discussed the factors determining the relative stability of the three isomers in solution, in which ZPE was not considered. Now we see that at 0 K the ZPE destabilizes the α-helix and β-turn relatively to the extended form by 2.7 and 2.2 kcal/mol, respectively. The thermal correction at 298.15 K (the Gibbs free energy minus ZPE) of the α-helix and β-turn relatively to the extended form is 8.1 and

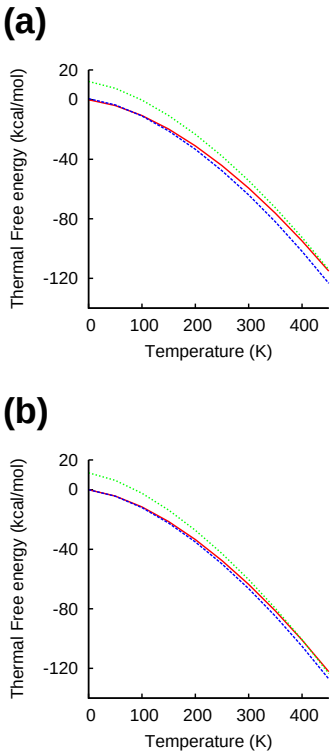


FIG. 6. Gibbs free energy for the three polyalanine isomers: $\alpha(\mathrm{ALA})_{10}$ (red solid line), $\beta(\mathrm{ALA})_{10}$ (blue dashed line) e$(\mathrm{ALA})_{10}$ (green dotted line). The energy of $\alpha(\mathrm{ALA})_{10}$ at 0 K is set as the zero level. (a) FMO Hessian (b) *ab initio*. RHF/6-31G(d) is used.

5.2 kcal/mol, respectively In other words, both ZPE and the thermal correction destabilize these two isomers versus the extended form.

Next, we plot the Gibbs free energy for the three polyalanine isomers in gas phase as a function of temperature (FIG. 6), relative to the Gibbs free energy at 0 K of $\alpha(\mathrm{ALA})_{10}$, which has the lowest energy at 0 K. The difference in the Gibbs free energies at 0 K comes from two factors: quantum-mechanical energy and ZPE. As discussed above, ZPE contributes 2.2-2.7 kcal/mol and the rest comes from the relative RHF energies. The α-helix and β-turn have a nearly identical energy at 0K, while the extended form is much less stable than the α-helix (by 12.3 and 11.4 kcal/mol in FMO and *ab initio*, respectively).

As the temperature increases, the less rigid extended form gains more free energy, mainly via low frequency vibrations, while the β-turn becomes more stable than the α-helix at high

temperatures. Small frequencies ν contribute more to the vibrational partition function $Q(\nu, T)$ in statistical thermodynamics[87],

$$Q(\nu, T) = 1/\left(1 - e^{-\frac{h\nu}{kT}}\right), \tag{48}$$

where k is the Boltzmann constant, h is the Planck constant and T is temperature. One can see that $Q(\nu, T) > 1$ for a finite frequency and temperature. The larger the frequency, the closer is $Q(\nu, T)$ to 1, and smaller frequencies have larger $Q(\nu, T)$. The free energy contribution is $-kT\ln Q(\nu, T)$. For ZPE the trend is opposite, that is, larger frequencies contribute more to ZPE and free energy.

We observe that FMO closely reproduced the *ab initio* temperature dependence of the relative free energies, although some deviations at high temperature are seen, related to the accuracy in reproducing small frequencies.

It is very interesting to observe that the large energy gap between the three isomers is very much reduced when the thermal corrections are included. It was reported earlier that in solution at the MP2/PCM/6-31G(d) level[86], the α-helix and β-turn are more stable than the extended form by 20.6 and 14.1 kcal/mol, respectively (in this work at the gas phase RHF/6-31G(d) level, the values are 14.1 and 13.6 kcal/mol, respectively). ZPE and the thermal correction at 298.15 K added together reduce the difference to the extended form considerably: by 8.1+2.7=10.8 and 5.2+2.2=7.4 (kcal/mol) for the α-helix and β-turn, respectively.

This can be seen as the justification why all three isomer forms provide building blocks for biological systems, because based on the earlier findings without considering ZPE and thermal corrections it would seem that the extended form is too unstable. The solvent effects, electron correlation and the ZPE+thermal corrections all make major contributions to the relative stabilities, and thus for a meaningful comparison of relative stabilities of biological systems it is indispensable to consider all of these factors.

3. Transition state search in FMO

We studied the S_N2 reaction of CH_3X and OH^- (X is Cl, Br and F) by comparing the free energy barrier and the vibrational mode corresponding to the imaginary frequency obtained

with FMO and *ab initio* calculations. The results are shown in TABLE IV. The activation

TABLE IV. Free energy barrier ΔG (kcal/mol) at 298.15 K for the S_N2 reaction between CH_3X (X=F, Cl and Br) and OH^-, and the corresponding imaginary mode ω (cm^{-1}), at the RHF/6-31G(d) level.

method		F	Cl	Br
FMO	ΔG	17.1	13.3	10.2
ab initio	ΔG	16.5	13.0	10.2
FMO	ω	585.8	470.3	428.5
ab initio	ω	590.7	472.1	427.6

free energies are accurately described; the difference between FMO and *ab initio* is 0.55, 0.30 and 0.03 kcal/mol for F, Cl and Br, respectively. The vibrational mode with the imaginary frequency shows a good agreement between FMO and *ab initio*; the difference is within 5 cm^{-1}.

A comparison of TS geometries is presented in TABLE V and FIG. 7. One can see that the transition state geometries are reproduced by FMO with good accuracy: the RMSDs of the solute between FMO and *ab initio* are 0.004-0.023 Å. The geometry of the whole solvated system has a good accuracy for Cl and Br and a somewhat larger deviation for F. This is due to the fact that there are numerous minima on the energy surface representing the solvent degrees of freedom, and the transition state search in FMO apparently found another local minimum. This can easily happen for the outer solvent shell which is relatively loosely bound. In addition to RMSD, we also show in TABLE V that the important bond length for this reaction is well reproduced by FMO.

The imaginary mode for Cl obtained with FMO is shown in FIG. 8 (a). The vibrational mode is spatially localized around the solute, consistent with *ab initio* calculations shown in FIG. 8 (b). Among F, Cl and Br, the largest imaginary frequency (590 cm^{-1}) is observed for F and the frequencies monotonously decrease from F to Br. The free energy barriers also

TABLE V. RMSD (Å) between the transition state structures obtained with *ab initio* and FMO, and the C-X bond length R_{CX} (Å) in the TS state of the chemical reaction between CH_3X (X=F, Cl and Br) and OH^-. RHF/6-31G(d) is used.

property	F	Cl	Br
RMSD[a]	0.468	0.143	0.121
RMSD[b]	0.023	0.008	0.004
R_{CX}(FMO)	1.792	2.198	2.325
R_{CX}(*ab initio*)	1.792	2.200	2.326

[a] Entire system including 41 water molecules. [b] Solute only.

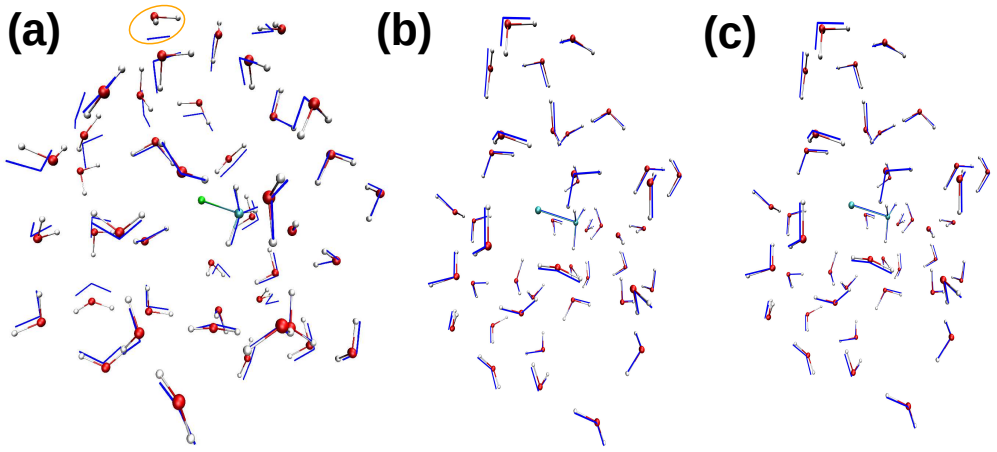


FIG. 7. Superposition of FMO (balls and sticks) and *ab initio* (blue wire frame) transition state structures at the RHF/6-31G(d) level for solvated (a) OH-CH_3-F, (b) OH-CH_3-Cl and (c) OH-CH_3-Br. The ellipse shows a pair of water molecules in the outer solvent shell for which the deviation is considerable.

decrease from F to Br, possibly related to the C-X bond weakening with increasing atomic number. These results are consistent with what is known in organic chemistry.[88]

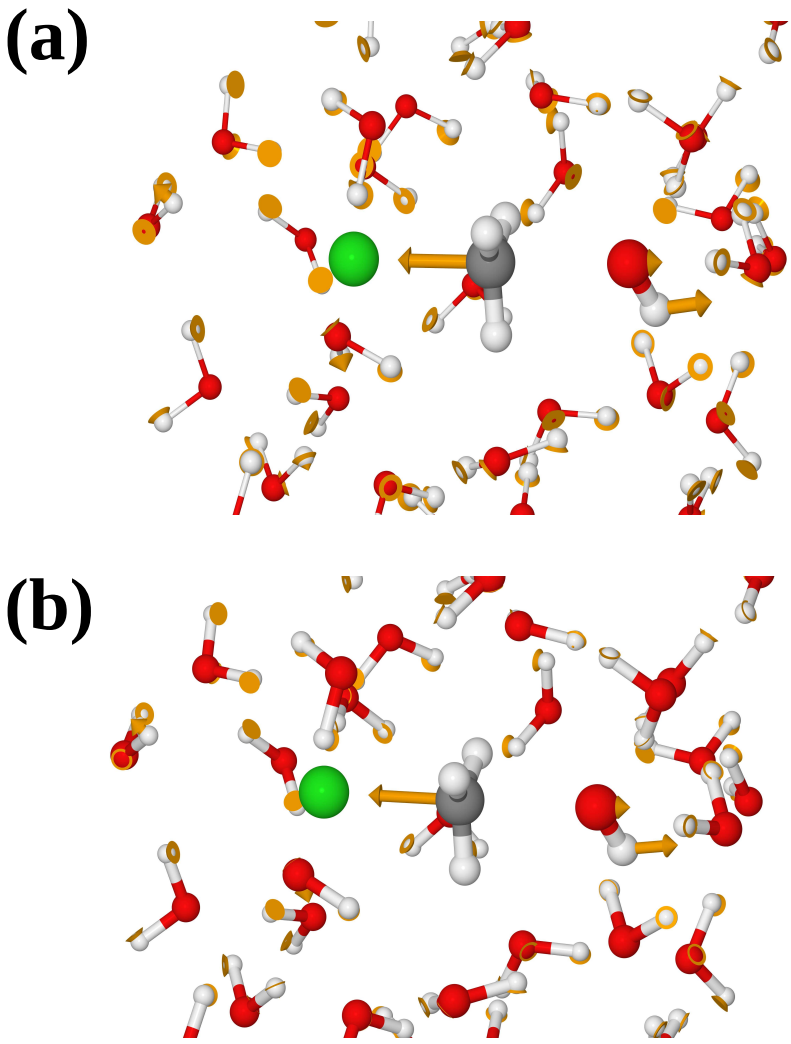


FIG. 8. The imaginary mode for the transition state of OH-CH_3-Cl solvated in 41 water molecules: (a) FMO and (b) *ab initio*.

4. Computational time

Computational time and the memory requirements are shown in FIG. 9 for water clusters consisting of 51 to 162 atoms (i.e. the water box sizes are 7 , 8 , 9 and 10 $\mathring{A}$). The timings are measured on four nodes equipped with dual 2.93 GHz Xeons (6 cores per CPU, 48 cores total) and 64 GB RAM per node.

FMO is faster than *ab initio* (FIG. 9 a). We observe also that the scaling of the FMO

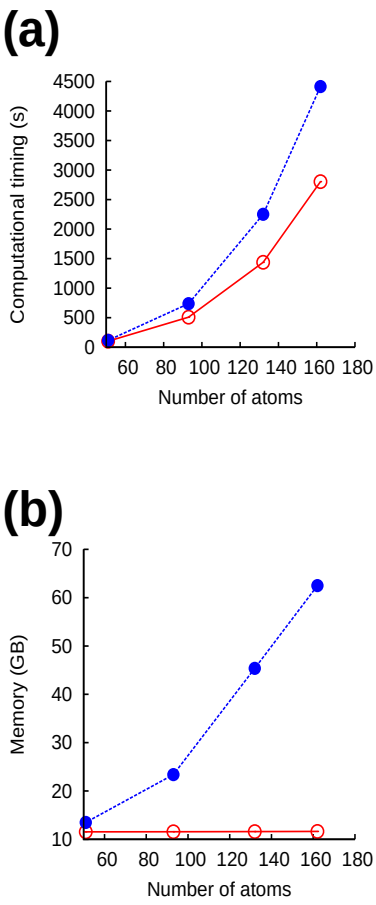


FIG. 9. Resources for Hessian calculations at the RHF/6-31G(d) level: (a) computational time (second) and (b) required RAM (GB). FMO and *ab initio* results are shown in red solid and blue doted lines, respectively.

Hessian is lower than that of *ab initio* and that it is not linear to the system size we used here. One can expect a better scaling of FMO with larger system sizes. The highest scaling steps in *ab initio* and FMO are different: in the former, it is solving CPHF for the entire system, in the latter it is the Hessian diagonalization for the whole system (cubic scaling). In this work we did not derive an analytic Hessian for ESP and electrostatic dimer approximations in FMO, and those can reduce the timings even further. For the system, whose results are shown in FIG. 9, the main cost for FMO is not the Hessian matrix diagonalization but the cost to solve CPHF and obtain the first and second derivatives of fragments and dimers.

Ultimately, this cost should be linear when all approximations are properly derived in future, and then the remaining cost will be one Hessian matrix diagonalization.

Ab initio Hessian requires very large memory (about 64 GB of RAM for the largest system), as shown in FIG. 9 b. This memory amount is necessary to solve CPHF equations and it may be reduced with alternative ways of solving CPHF equations (CPHF=AODDI), which may be more time consuming. In FMO one has to provide a memory amount scaling quadratically with the system size to store the total Hessian, but at least for the water clusters we used, the main memory requirement in FMO is not to store the Hessian, but to solve CPHF of monomers or dimers. Thus, the required memory only depends on the largest dimer size, not on the number of fragments (not on the system size). Indeed, one observes that the required memory in FMO does not depend on the system size and it is only 80 MB for all systems in FIG. 9 (b).

E. Conclusions

We have derived the analytic second derivative of the FMO energy, introduced some approximations to it and implemented the approximate method into GAMESS. Other fragment-based methods using electrostatic embedding may benefit from our work. We have also implemented the transition state search using FMO, opening a new exciting field of applications to chemical reactions. In addition, our Hessian can be used to locate transition states in *ab initio* methods, when memory requirements are too prohibitive to calculate the Hessian but one still desires to use *ab initio* methods for the electronic state in the TS search with a cheaper Hessian from FMO. Some further improvement may be obtained by calculating the Hessian for a subset of fragments, analogously to partial Hessian approaches[67] or in the spirit of the partial energy gradient[46] or domain-based FMO.[47]

We have demonstrated that the FMO Hessian allows us to accurately simulate IR spectra, evaluate the Gibbs free energy and zero point energy and do transition state search for some representative systems including water clusters, polypeptides and a S_N2 chemical reaction.

In the application of the Hessian to IR spectra of polyalanine we have provided some insights on the effect of the secondary structure upon these properties. We have also discussed

the thermochemistry of polyalanines including the zero point energy and Gibbs free energy. We found that the zero point energy stabilizes the extended form in comparison to the α-helix and β-turn by 2.7 and 2.2 kcal/mol, respectively. In gas phase at higher temperature the energetic differences among the three structures decrease and the β-turn becomes more stable than the α-helix. FMO Hessian should be a useful tool for future studies of large molecular systems.

REFERENCES

[1]G. E. Scuseria, J. Phys. Chem. A **103**, 4782 (1999).

[2]R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski, editors, *Linear-Scaling Techniques in Computational Chemistry and Physics*, Springer, Berlin, 2011.

[3]J. R. Reimers, editor, *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, Wiley, New York, 2011.

[4]E. Nikitina, V. Sulimov, V. Zayets, and N. Zaitseva, Int. J. Quant. Chem. **97**, 747 (2004).

[5]J. J. P. Stewart, J. Mol. Model. **15**, 765 (2009).

[6]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[7]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[8]S. Sakai and S. Morita, J. Phys. Chem. A **109**, 8424 (2005).

[9]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101 (2008).

[10]Y. Tong, Y. Mei, J. Z. H. Zhang, L. L. Duan, and Q. G. Zhang, J. Theor. Comput. Chem. **8**, 1265 (2009).

[11]P. Söderhjelm and U. Ryde, J. Phys. Chem. A **113**, 617 (2009).

[12]M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, and J. A. Boatz, J. Phys. Chem. B **113**, 9646 (2009).

[13]X. He and K. M. Merz, J. Chem. Theory Comput. **6**, 405 (2010).

[14]M. Kobayashi, T. Kunisada, T. Akama, D. Sakura, and H. Nakai, J. Chem. Phys. **134**, 034105 (2011).

[15]L. Huang, H. J. Bohorquez, C. F. Matta, and L. Massa, Int. J. Quant. Chem. **111**, 4150 (2011).

[16]M. A. Collins, Phys. Chem. Chem. Phys. **14**, 7744 (2012).

[17]Y. Aoki and F. L. Gu, Phys. Chem. Chem. Phys. **14**, 7640 (2012).

[18]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[19]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999),

[20]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007),

[21]D. G. Fedorov and K. Kitaura, editors, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC Press, Boca Raton, FL, 2009.

[22]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[23]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[24]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).

[25]D. G. Fedorov and K. Kitaura, Chem. Phys. Lett. **389**, 129 (2004),

[26]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **122**, 054108 (2005),

[27]Y. Mochizuki, S. Koikegami, S. Amari, K. Segaea, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **406**, 283 (2005),

[28]Y. Mochizuki, K. Tanaka, K. Yamashita, T. Ishikawa, T. Nakano, S. Amari, and M. Sakurai, Theor. Chem. Acc. **117**, 541 (2007),

[29]M. Chiba, D. G. Fedorov, and K. Kitaura, J. Chem. Phys. **127**, 104108 (2007).

[30]S. R. Pruitt, D. G. Fedorov, and M. S. Gordon, J. Phys. Chem. A **116**, 4965 (2012).

[31]H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, J. Chem. Phys. **137**, 044110 (2012).

[32]K. Kitaura, S. I. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, Chem. Phys. Lett. **336**, 163 (2001).

[33]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **475**, 124 (2009).

[34]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).

[35]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[36]T. Nagata, D. G. Fedorov, H. Li, and K. Kitaura, J. Chem. Phys. **136**, 204112 (2012).

[37]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **544**, 87 (2012).

[38]Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **372**, 342 (2003).

[39]Y. Komeiji, Y. Mochizuki, T. Nakano, and D. G. Fedorov, J. Mol. Str. (THEOCHEM) **898**, 2 (2009).

[40]T. Nagata, D. G. Fedorov, and K. Kitaura, Theor. Chem. Acc. **131**, 1136 (2012).

[41]M. Sato, H. Yamataka, Y. Komeiji, and Y. Mochizuki, Chem. Eur. J. **18**, 9714 (2012).

[42]K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 5008 (2012).

[43]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[44]K. Ohno, K. Mori, M. Orita, and M. Takeuchi, Curr. Med. Chem. **18**, 220 (2011).

[45]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007),

[46]T. Ishikawa, N. Yamamoto, and K. Kuwata, Chem. Phys. Lett. **500**, 149 (2010).

[47]D. G. Fedorov, Y. Alexeev, and K. Kitaura, J. Phys. Chem. Lett. **2**, 282 (2011).

[48]T. Tsukamoto, Y. Mochizuki, N. Watanabe, K. Fukuzawa, and T. Nakano, Chem. Phys. Lett. **535**, 157 (2012).

[49]T. Ishida, D. G. Fedorov, and K. Kitaura, J. Phys. Chem. B **110**, 1457 (2006).

[50]I. V. Polyakov, B. L. Grigorenko, A. A. Moskovsky, V. M. Pentkovski, and A. V. Nemukhin, Chem. Phys. Lett. **556**, 251 (2013).

[51]C. Steinmann, D. G. Fedorov, and J. H. Jensen, submitted .

[52]P. Deglmann, F. Furche, and R. Ahlrichs, Chem. Phys. Lett. **362**, 511 (2002).

[53]Y. Alexeev, M. W. Schmidt, T. L. Windus, and M. S. Gordon, J. Comput. Chem. **28**, 1685 (2007).

[54]P. Pulay, Mol. Phys. **17**, 197 (1969).

[55]Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).

[56]A. Ghysels, H. L. Woodcock III, J. D. Larkin, B. T. Miller, Y. Shao, J. Kong, D. V. Neck, V. V. Speybroeck, M. Waroquier, B. R. Brooks, J. Chem. Theory Comput. **7**, 496 (2011).

[57]S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Str.: THEOCHEM **461**, 1 (1999).

[58]H. A. Witek, S. Irle, and K. Morokuma, J. Chem. Phys. **121**, 5163 (2004).

[59]H. Li and J. H. Jensen, Theor. Chem. Acc. **107**, 211 (2002).

[60]J. H. Jensen, H. Li, A. D. Robertson, and P. A. Molina, J. Phys. Chem. A **109**, 6634 (2005).

[61]A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, J. Chem. Phys. **126**, 224102 (2007).

[62]A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B. R. Brooks, and M. Waroquier, J. Chem. Theory Comput. **5**, 1203 (2009).

[63]A. Ghysels, D. Van Neck, V. Van Speybroeck, B. R. Brooks, and M. Waroquier, J. Chem. Phys. **130**, 084107 (2009).

[64]W. J. Zheng and B. R. Brooks, Biophys. J. **89**, 167 (2005).

[65]H. L. Woodcock, W. J. Zheng, A. Ghysels, Y. H. Shao, J. Kong, and B. R. Brooks, J. Chem. Phys. **129**, 214109 (2008).

[66]J. Hafner and W. Zheng, J. Chem. Phys. **130**, 194111 (2009).

[67]A. Ghysels, V. V. Speybroeck, E. Pauwels, S. Catak, B. R. Brooks, D. V. Neck, and M. Waroquier, J. Comput. Chem. **31**, 994 (2010).

[68]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, , J. Comput. Chem. **14**, 1347 (1993).

[69]M. S. Gordon and M. W. Schmidt, Advances in electronic structure theory: Gamess a decade later, in *Theory and Applications of Computational Chemistry, the first forty years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria, pages 1167–1189, Elsevier, Amsterdam, 2005.

[70]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **351**, 475 (2002),

[71]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, and J. Goddard, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press, New York, 1994, CAN 123:66488 65-3 General Physical Chemistry USA. Book written in English.

[72]See supplementary material at

ftp : //ftp.aip.org/epaps/journ_chem_phys/E − JCPSA6 − 138 − 031315/ for the details

of the derivations.

[73]C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S Gordon, Theor. Chem. Acc. **110**, 233 (2003).

[74]T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, J. Chem. Phys. **135**, 044110 (2011).

[75]D. G. Fedorov, J. H. Jensen, R. C. Deka, and K. Kitaura, J. Phys. Chem. A **112**, 11808 (2008).

[76]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).

[77]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004),

[78]Y. Alexeev, M. P. Mazanetz, O. Ichihara, and D. G. Fedorov, Curr. Top. Med. Chem. **12**, 2013 (2012).

[79]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004),

[80]J. Wang, P. Cieplak, and P. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[81]M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, Comput. Phys. Commun. **181**, 1477 (2010).

[82]C. Steinmann, M. W. Ibsen, A. S. Hansen, and J. H. Jensen, PLOS ONE **7**, e44480 (2012).

[83]J. Kong and S. Yu, Acta Biochimica et Biophysica Sinica **39**, 549 (2007).

[84]A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).

[85]D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007),

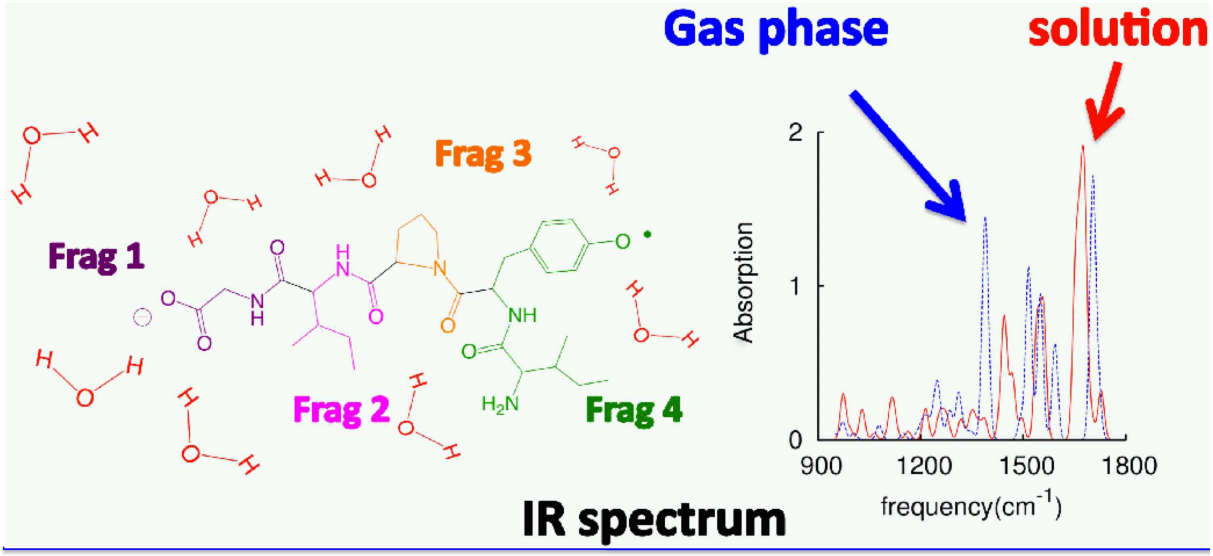
[86]H. Li, D. G. Fedorov, T. Nagata, K. Kitaura, J. H. Jensen, and M. S. Gordon, J. Comput. Chem. **31**, 778 (2010).

[87]P. Atkins and J. de Paula, *Physical Chemistry*, W. H. Freeman and Company, New York, 2002.

[88]S. Warren, J. Clayden, N. Greeves, and P. Wothers, *Organic Chemistry*, Oxford University Press, New York, 2008.

Chapter VI

Efficient vibrational analysis for unrestricted Hartree-Fock based on the fragment molecular orbital method



H. Nakata, D. G. Fedorov *et. al.*

Chem. Phys. Lett. **2014**, *603*, 67-74.

DOI : 10.1016/j.cplett.2014.04.028

View online: http://www.sciencedirect.com/science/article/pii/S0009261414003066

VI. EFFICIENT VIBRATIONAL ANALYSIS FOR UNRESTRICTED HARTREE-FOCK BASED ON THE FRAGMENT MOLECULAR ORBITAL METHOD

A. Introduction

Vibrational frequency calculations are very useful to simulate vibrational infrared (IR) and Raman spectra, estimate thermodynamic properties such as entropy and free energy, and perform stationary point search. In order to obtain vibrational frequencies, one has to calculate the matrix of the second derivatives of the energy and diagonalize it[1–3].

For large systems Hessian calculations are very expensive even at the level of Hartree-Fock. One can apply the combined quantum-mechanical and molecular-mechanical (QM/MM) method[4–6], our own n-layered integrated molecular orbital and molecular mechanics (ONIOM)[7], and density functional tight binding method[8]. There are also various partial approaches where the Hessian is computed usually numerically only for a subset of atoms[9–11].

Alternatively, there is a variety of fragment-based methods[12–15], for some of which the analytic second derivative has been developed for closed shell molecules: molecular tailoring approach[16], generalized energy-based fragmentation method[17], integrated multi-center molecular orbital (MO) approach[18], and the fragment molecular orbital (FMO) method[19]. There is a need for a normal mode analysis applicable to large radical systems.

In FMO[20–22], one divides the system into fragments (also called monomers), and calculates each of them in the embedding electrostatic potential (ESP) due to the remaining fragments. After the fragment electronic states converge with respect to ESP, fragment pair calculations are performed. FMO has been applied to proteins[23,24], DNA[25], and inorganic systems[26,27]. Geometry optimizations[28,29] and molecular dynamic simulations[30,31] can be conducted using the fully analytic energy gradients in both restricted[32] and unrestricted[33] formulations. An important by-product of fragment-based calculations is the pair interaction energies (PIEs), which can be decomposed into physically meaningful components[34,35].

In this Letter, we develop the analytic second derivative of the energy for unrestricted Hartree-Fock (UHF) based on FMO (FMO-UHF), using coupled perturbed Hartree-Fock

(CPHF)[36] to obtain the response terms arising from the coupling of the electronic state of fragment pairs and the embedding ESP. The FMO Hessian as originally formulated for restricted Hartree-Fock (RHF)[19] is very expensive, because a quadratic number of Hartree-Fock calculations for fragment pairs has to be done. To address this problem, in this Letter we develop the second derivative for the electrostatic dimer approximation (ES-DIM)[37], where the number of Hartree-Fock dimers scales linearly with the system size, while the rest is treated with a very fast and accurate approximation. The accuracy and efficiency of FMO-UHF Hessian are evaluated in comparison with the *ab initio* calculation and experimental results.

B. Theory

1. Second derivative of the FMO-UHF energy

In FMO-UHF[33] some fragments are calculated with UHF, and the rest with RHF. A dimer calculation is performed with UHF, only when at least one fragment in the dimer is UHF, otherwise RHF dimers are computed. Although we derive the equations for any number of UHF fragments, our current implementation of FMO-UHF is limited to one UHF fragment.

The second derivative of the total energy E in FMO-UHF with respect to nuclear coordinates a and b in the two-body expansion is

$$\begin{aligned}\frac{\partial^2 E}{\partial a \partial b} = &\sum_{I=1}^{N^{\mathrm{RHF}}} \frac{\partial^2 E'_I}{\partial a \partial b} + \sum_{K=1}^{N^{\mathrm{UHF}}} \frac{\partial^2 E'_K}{\partial a \partial b} \\ &+ \sum_{\substack{I>J \\ R_{IJ}<R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{RHF}}} \frac{\partial^2 \Delta E_{IJ}}{\partial a \partial b} + \sum_{\substack{K>L \\ R_{KL}<R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{UHF}}} \frac{\partial^2 \Delta E_{KL}}{\partial a \partial b} \\ &+ \sum_{K=1}^{N^{\mathrm{UHF}}} \sum_{\substack{I=1 \\ R_{KI}<R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{RHF}}} \frac{\partial^2 \Delta E_{KI}}{\partial a \partial b} + \frac{\partial^2 \Delta E_{\mathrm{ES}}}{\partial a \partial b}. \end{aligned} \tag{1}$$

where the pair interaction energy ΔE_{IJ} is

$$\Delta E_{IJ} = E'_{IJ} - E'_I - E'_J + \mathrm{Tr}\left(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}\right). \tag{2}$$

E'_X is the internal energy of monomers ($X = I$) or dimers ($X = IJ$). N^{RHF} and N^{UHF} are the numbers of the respective fragments. I and J run over RHF fragments, whereas K and L denote UHF fragments. $\Delta\mathbf{D}^{IJ}$ is the difference between the density matrix of dimer $\mathbf{D}^{IJ}$ and the sum of monomer densities $\mathbf{D}^I$ and $\mathbf{D}^J$, $\Delta\mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^I \oplus \mathbf{D}^J)$. A UHF density matrix without a spin label denotes the sum of the densities for α and β orbitals.

If two fragments I and J are sufficiently separated from each other (the interfragment distance R_{IJ} is larger than the threshold $R_{\mathrm{ES-DIM}}$) then the quantum effects such as charge transfer are small and can be neglected. The energy of such separated dimers is calculated as the electrostatic (ES) interaction; ΔE_{ES} is the sum of all such contributions, defined later.

The electrostatic embedding potential $\mathbf{V}^{IJ}$ is

$$V^{IJ}_{\mu\nu} = \sum_{K\neq I,J}^{N_{\mathrm{RHF}}+N_{\mathrm{UHF}}} (u^K_{\mu\nu} + v^K_{\mu\nu}), \tag{3}$$

where $u^K_{\mu\nu}$ and $v^K_{\mu\nu}$ are the one and two-electron Coulomb potentials, respectively:

$$u^K_{\mu\nu} = \sum_{A\in K} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle, \tag{4}$$

$$v^K_{\mu\nu} = \sum_{\lambda,\rho\in K} D^K_{\lambda\rho}(\mu\nu|\lambda\rho). \tag{5}$$

$(\mu\nu|\rho\lambda)$ is the two-electron integral, and A labels atoms with the charge Z_A and coordinates $\mathbf{R}_A$. Greek and small Roman indices denote atomic and molecular orbitals, respectively. i,j,k and l run over occupied MOs, whereas m denotes virtual MOs.

For RHF fragments, the derivatives of the internal energy E'_X and $\mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})$ are given in[19]. Here, we derive the equations for UHF fragments. The second derivative of the internal energy for UHF fragment X is as follows (see Supplementary materials for more

details).

$$
\begin{aligned}
\frac{\partial^2 E'_X}{\partial a \partial b} = \sum_\sigma \Bigg[& \frac{1}{2} \sum_{i^\sigma}^{\text{occ}} \left(h^{ab,X}_{i^\sigma i^\sigma} + F'^{ab,X}_{i^\sigma i^\sigma} \right) - \sum_{i^\sigma}^{\text{occ}} S^{ab,X}_{i^\sigma i^\sigma} \epsilon^X_{i^\sigma} \\
& + 2 \sum_{i^\sigma}^{\text{occ}} \sum_{j^\sigma}^{\text{occ}} S^{a,X}_{i^\sigma j^\sigma} S^{b,X}_{i^\sigma j^\sigma} \epsilon^X_{i^\sigma} \\
& + 2 \sum_{i^\sigma}^{\text{occ}} \sum_{m^\sigma}^{\text{vir}} U^{b,X}_{m^\sigma i^\sigma} \Bigg\{ F'^{a,X}_{m^\sigma i^\sigma} - S^{a,X}_{m^\sigma i^\sigma} \epsilon^X_{i^\sigma i^\sigma} \\
& - \frac{1}{2} \sum_{j^\sigma}^{\text{occ}} \sum_{k^\sigma}^{\text{occ}} S^{a,X}_{k^\sigma j^\sigma} A'_{k^\sigma j^\sigma i^\sigma m^\sigma} - \frac{1}{2} \sum_{j^{\overline{\sigma}}}^{\text{occ}} \sum_{k^{\overline{\sigma}}}^{\text{occ}} S^{a,X}_{k^{\overline{\sigma}} j^{\overline{\sigma}}} A'_{k^{\overline{\sigma}} j^{\overline{\sigma}} m^\sigma i^\sigma} \Bigg\} \\
& - \sum_{i^\sigma}^{\text{occ}} \sum_{j^\sigma}^{\text{occ}} S^{a,X}_{i^\sigma j^\sigma} \Bigg\{ F'^{b,X}_{i^\sigma j^\sigma} - \frac{1}{4} \sum_{k^\sigma}^{\text{occ}} \sum_{l^\sigma}^{\text{occ}} S^{b,X}_{k^\sigma l^\sigma} A'_{i^\sigma j^\sigma k^\sigma l^\sigma} \\
& - \frac{1}{4} \sum_{k^{\overline{\sigma}}}^{\text{occ}} \sum_{l^{\overline{\sigma}}}^{\text{occ}} S^{b,X}_{k^{\overline{\sigma}} l^{\overline{\sigma}}} A'_{i^\sigma j^\sigma k^{\overline{\sigma}} l^{\overline{\sigma}}} \Bigg\} \\
& - \sum_{i^\sigma}^{\text{occ}} \sum_{j^\sigma}^{\text{occ}} S^{b,X}_{i^\sigma j^\sigma} \Bigg\{ F'^{a,X}_{i^\sigma j^\sigma} - \frac{1}{4} \sum_{k^\sigma}^{\text{occ}} \sum_{l^\sigma}^{\text{occ}} S^{a,X}_{k^\sigma l^\sigma} A'_{i^\sigma j^\sigma k^\sigma l^\sigma} \\
& - \frac{1}{4} \sum_{k^{\overline{\sigma}}}^{\text{occ}} \sum_{l^{\overline{\sigma}}}^{\text{occ}} S^{a,X}_{k^{\overline{\sigma}} l^{\overline{\sigma}}} A'_{i^\sigma j^\sigma k^{\overline{\sigma}} l^{\overline{\sigma}}} \Bigg\} \Bigg] \\
& - \overline{U}^{ab,X,X} + \frac{\partial^2 E^{\text{NR}}_X}{\partial a \partial b}.
\end{aligned} \tag{6}
$$

σ is the spin label (α or β). $\overline{\sigma}$ means the spin opposite to σ, i.e., if $\sigma = \alpha$, then $\overline{\sigma} = \beta$. ϵ are orbital energies. The superscripts a (or b) and ab denote the first and second derivatives, respectively. $h^{a,X}_{ij}$ ($h^{ab,X}_{ij}$), $F'^{a,X}_{ij}$ ($F'^{ab,X}_{ij}$) and $S^{a,X}_{ij}$ ($S^{ab,X}_{ij}$) are the derivatives of the core Hamiltonian, the internal fragment Fock matrix and overlap integrals, respectively (see the Supplemental materials for the definitions). The relation between the conventional $\mathbf{F}$ and internal $\mathbf{F}'$ Fock matrices is $\mathbf{F} = \mathbf{F}' + \mathbf{V}$, i.e., they differ by the ESP. All of these three derivative quantities ($\mathbf{h}$, $\mathbf{F}'$ and $\mathbf{S}$) are MO-transformed derivative integrals in AO basis, e.g.,

$$
h^{a,X}_{ij} = \sum_{\mu,\nu \in X} C^{X*}_{\mu i} C^X_{\nu j} h^{a,X}_{\mu\nu}, \tag{7}
$$

where $C^X_{\mu i}$ are the MO coefficients.

As in the RHF Hessian[19], we neglected the contribution of the second derivative of the hybrid projection operator to $F'^{ab,X}_{i^\sigma j^\sigma}$. U^a is the orbital response[32] for a change in a nuclear coordinate a and $\overline{U}^{ab,X,Y}$ is the contribution to the Hessian arising from the second response contributions (see the Supplemental materials). The orbital Hessian is

$$A'_{i^\sigma j^\sigma, k^\sigma l^\sigma} = 2(i^\sigma j^\sigma | k^\sigma l^\sigma) - (i^\sigma l^\sigma | j^\sigma k^\sigma) \tag{8}$$

$$A'_{i^\sigma j^\sigma, k^{\overline{\sigma}} l^{\overline{\sigma}}} = 2(i^\sigma j^\sigma | k^{\overline{\sigma}} l^{\overline{\sigma}}). \tag{9}$$

The second derivative of the embedded density transfer term for UHF dimers KI is

$$\frac{\partial^2 \mathrm{Tr}(\Delta \mathbf{D}^{KI} \mathbf{V}^{KI})}{\partial a \partial b} = \frac{\partial^2 \mathrm{Tr}(\mathbf{D}^{KI} \mathbf{V}^{KI})}{\partial a \partial b} - \frac{\partial^2 \mathrm{Tr}(\mathbf{D}^{K} \mathbf{V}^{KI})}{\partial a \partial b} - \frac{\partial^2 \mathrm{Tr}(\mathbf{D}^{I} \mathbf{V}^{KI})}{\partial a \partial b}, \tag{10}$$

where for $X = I, J$ or IJ,

$$\begin{aligned}
\frac{\partial^2 \mathrm{Tr}(\mathbf{D}^{X} \mathbf{V}^{KI})}{\partial a \partial b} = \sum_\sigma \Bigg[& \sum_{i^\sigma \in X}^{\mathrm{occ}} V^{ab,KI}_{i^\sigma i^\sigma} \\
& - \sum_{i^\sigma \in X}^{\mathrm{occ}} \sum_{j^\sigma \in X}^{\mathrm{occ}} (S^{a,X}_{i^\sigma j^\sigma} V^{b,KI}_{i^\sigma j^\sigma} + S^{b,X}_{j^\sigma i^\sigma} V^{a,KI}_{j^\sigma i^\sigma}) \\
& + 2 \sum_{i^\sigma \in X}^{\mathrm{occ}} \sum_{m^\sigma \in X}^{\mathrm{vir}} U^{b,X}_{m^\sigma i^\sigma} V^{a,KI}_{m^\sigma i^\sigma} \Bigg] \\
& + R^{ab,X} + \overline{U}^{ab,X,KI},
\end{aligned} \tag{11}$$

where $V^{a,KI}_{i^\sigma j^\sigma}$ and $V^{ab,KI}_{i^\sigma j^\sigma}$ are the first and second derivatives of ESP, respectively. $R^{ab,KI}$ are the two-electron integral response terms (see the Supplementary materials for the definition):

$$R^{ab,KI} = \sum_{Y \neq K,I}^{N^{\mathrm{RHF}}} R^{ab,KI(Y)} + \sum_{Z \neq K,I}^{N^{\mathrm{UHF}}} R^{ab,KI(Z)} \tag{12}$$

where Y and Z denote fragment contributions to the ESP of dimers.

The analytic second derivative of the energy requires solving the first order CPHF equations for the whole system. For practical efficiency, we introduce the approximation decoupling these equations and solve them for individual fragments.

2. The second derivative of the energy of separated dimers

The second derivative of the energy of separated dimers is

$$\frac{\partial^2 \Delta E_{\mathrm{ES}}}{\partial a \partial b} = \sum_{\substack{I>J \\ R_{IJ} \geq R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{RHF}}} \frac{\partial^2 E_{IJ}^{\mathrm{ES}}}{\partial a \partial b} + \sum_{\substack{K>L \\ R_{KL} \geq R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{UHF}}} \frac{\partial^2 E_{KL}^{\mathrm{ES}}}{\partial a \partial b} + \sum_{K=1}^{N^{\mathrm{UHF}}} \sum_{\substack{I=1 \\ R_{KI} \geq R_{\mathrm{ES-DIM}}}}^{N^{\mathrm{RHF}}} \frac{\partial^2 E_{KI}^{\mathrm{ES}}}{\partial a \partial b}, \tag{13}$$

In this Letter we derived the analytic derivatives for both RHF (see the Supplementary materials) and UHF, as described below. For a dimer composed of one UHF fragment K and one RHF fragment I (UHF-UHF dimers can be derived analogously).

$$\frac{\partial^2 E_{KI}^{\mathrm{ES}}}{\partial a \partial b} = \overline{E}_{\mathrm{ES}}^{ab,K,K(I)} + \overline{E}_{\mathrm{ES}}^{ab,I,I(K)} + \overline{U}^{ab,K,K(I)} + \overline{U}^{ab,I,I(K)} + R_{\mathrm{ES}}^{ab,KI}, \tag{14}$$

where $\overline{U}^{ab,K,L(I)}$ is the contribution of only fragment I in the ESP acting on fragment L contracted with the density of fragment K (see Eq. (66) in Supplementary materials for an explicit definition), whereas $\overline{U}^{ab,K,L}$ is the sum over all fragments I corresponding to the full ESP in which fragment L is embedded.

$$\overline{U}^{ab,K,L} = \sum_{I \neq L} \overline{U}^{ab,K,L(I)}, \tag{15}$$

and

$$\overline{E}_{\mathrm{ES}}^{ab,K,K(I)} = \sum_{\sigma} \left[- \sum_{i^\sigma \in K}^{\mathrm{occ}} \sum_{j^\sigma \in K}^{\mathrm{occ}} (S_{i^\sigma j^\sigma}^{b,K} V_{i^\sigma j^\sigma}^{a,K(I)} + S_{i^\sigma j^\sigma}^{a,K} V_{i^\sigma j^\sigma}^{b,K(I)}) + 2 \sum_{i^\sigma \in K}^{\mathrm{occ}} \sum_{m^\sigma \in K}^{\mathrm{vir}} U_{m^\sigma i^\sigma}^{b,K} V_{m^\sigma i^\sigma}^{a,K(I)} + \sum_{i^\sigma \in K}^{\mathrm{occ}} V_{i^\sigma i^\sigma}^{ab,K(I)} \right], \tag{16}$$

$$
\begin{aligned}
R_{\rm ES}^{ab,KI} = \sum_{\sigma} \Bigg[& -4 \sum_{i\in I}^{\rm occ} \sum_{j\in I}^{\rm occ} \sum_{k^\sigma\in K}^{\rm occ} \sum_{m^\sigma\in K}^{\rm vir} S_{ij}^{a,I} U_{m^\sigma k^\sigma}^{b,K} (ij|k^\sigma m^\sigma) \\
& - 4 \sum_{i^\sigma\in K}^{\rm occ} \sum_{j^\sigma\in K}^{\rm occ} \sum_{k\in I}^{\rm occ} \sum_{m\in I}^{\rm vir} S_{i^\sigma j^\sigma}^{a,K} U_{mk}^{b,I} (i^\sigma j^\sigma|km) \\
& + 2 \sum_{i\in I}^{\rm occ} \sum_{j\in I}^{\rm occ} \sum_{k^\sigma\in K}^{\rm occ} \sum_{l^\sigma\in K}^{\rm occ} (S_{ij}^{a,I} S_{k^\sigma l^\sigma}^{b,K} + S_{ij}^{b,I} S_{k^\sigma l^\sigma}^{a,K}) (ij|k^\sigma l^\sigma) \\
& - 2 \sum_{i\in I}^{\rm occ} \sum_{j^\sigma\in K}^{\rm occ} (ii|j^\sigma j^\sigma)^{ab} \Bigg],
\end{aligned} \tag{17}
$$

and the contribution of fragment I to the ESP acting on fragment K is

$$
V_{\mu\nu}^{K(I)} = \sum_{A\in I} \langle \mu \left| \frac{-Z_A}{|\mathbf{r}-\mathbf{R}_A|} \right| \nu \rangle + \sum_{\lambda,\sigma\in I} D_{\lambda\sigma}^{I} (\mu\nu|\sigma\lambda). \tag{18}
$$

Collecting all response terms in Eqs. (6), (11), and (13), one obtains

$$
\begin{aligned}
\overline{U}^{ab} = & \sum_{I=1}^{N^{\rm RHF}} \overline{U}^{ab,I} + \sum_{K=1}^{N^{\rm UHF}} \overline{U}^{ab,K} \\
& + \sum_{\substack{I>J \\ R_{IJ}<R_{\rm ES-DIM}}}^{N^{\rm RHF}} \left(-\overline{U}^{ab,I,IJ} - \overline{U}^{ab,J,IJ} \right) \\
& + \sum_{\substack{K>L \\ R_{KL}<R_{\rm ES-DIM}}}^{N^{\rm UHF}} \left(-\overline{U}^{ab,K,KL} - \overline{U}^{ab,L,KL} \right) \\
& + \sum_{K=1}^{N^{\rm UHF}} \sum_{\substack{I=1 \\ R_{KI}<R_{\rm ES-DIM}}}^{N^{\rm RHF}} \left(-\overline{U}^{ab,K,KI} - \overline{U}^{ab,I,KI} \right) \\
& + \overline{U}^{ab,{\rm ES-DIM}},
\end{aligned} \tag{19}
$$

where the $\overline{U}^{ab,{\rm ES-DIM}}$ is the sum of all individual terms ($\overline{U}^{ab,K,K(I)} + \overline{U}^{ab,I,I(K)}$) in Eq. (14). For the FMO gradient[38] $\overline{U}^{a} = 0$, and likewise for the FMO Hessian $\overline{U}^{ab} = 0$, provided that no point charge approximation is used in ESP. Therefore, one does not need to compute various $\overline{U}^{ab}$ terms in Eqs. (6), (11), and (14), because their sum is zero.

3. *The first order CPHF equation in UHF*

To obtain the analytic second derivative in FMO-UHF, one has to solve the first order CPHF equations for each monomer and dimer. CPHF equations of monomers are coupled to each other, but they can be approximately decoupled as in RHF[19]. The CPHF equations for fragment X computed with UHF are

$$\mathbf{A}^{X,X}\mathbf{U}^{a,X} = \mathbf{B}_0^{a,X}, \tag{20}$$

where we neglected the terms involving the derivatives for $a \notin X$ and used the following definitions for the same and opposite spin, respectively:

$$A^{X,X}_{i^\sigma j^\sigma, r^\sigma k^\sigma} = \delta_{i^\sigma r^\sigma}\delta_{j^\sigma k^\sigma}\left(\epsilon^X_{j^\sigma} - \epsilon^X_{i^\sigma}\right) - A'_{i^\sigma j^\sigma, r^\sigma k^\sigma} \tag{21}$$

$$A^{X,X}_{i^\sigma j^\sigma r^{\overline{\sigma}} k^{\overline{\sigma}}} = -A'_{i^\sigma j^\sigma, r^{\overline{\sigma}} k^{\overline{\sigma}}} \tag{22}$$

The matrix elements of $\mathbf{B}^{a,X}$ are

$$\begin{aligned} B^{a,X}_{0,i^\sigma j^\sigma} \approx\ & F'^{a,X}_{i^\sigma j^\sigma} + V^{a,X}_{i^\sigma j^\sigma} - S^{a,X}_{i^\sigma j^\sigma}\epsilon^X_{j^\sigma} \\ & - \frac{1}{2}\sum_{k^\sigma \in X}^{\mathrm{occ}}\sum_{l^\sigma \in X}^{\mathrm{occ}} S^{a,K}_{k^\sigma l^\sigma} A'_{i^\sigma j^\sigma, k^\sigma l^\sigma}. \\ & - \frac{1}{2}\sum_{k^\gamma \in X}^{\mathrm{occ}}\sum_{l^\gamma \in X}^{\mathrm{occ}} S^{a,K}_{k^\gamma l^\gamma} A'_{i^\sigma j^\sigma, k^{\overline{\sigma}} l^{\overline{\sigma}}}. \end{aligned} \tag{23}$$

Note that $V^{a,X}_{i^\sigma j^\sigma}$ is from ESP, and we only consider contributions for atoms $a \in X$.

Finally, we summarize some important points about the FMO Hessian. In the FMO gradient, one has to solve self-consistent Z-vector (SCZV)[32] equations for monomers, in order to obtain the gradient contribution due to responses U^a via the Z-vector (without obtaining individual values of U^a). For FMO-Hessians one cannot use the Z-vector formulation, because individual values of U^a are needed. Therefore, we solve Eq. (20) directly, and this has to be done for each monomer and RHF or UHF dimer, but not for separated dimers. This equation is solved once without self-consistency, as indicated above. The second order responses U^{ab} are not needed if ESP approximations are not used.

C. Computational details

Because production version of GAMESS[39,40] did not have the analytic *ab initio* UHF Hessian, we implemented both conventional and FMO-based UHF Hessians. The latter was parallelized with the generalized distributed data interface (GDDI)[41].

The accuracy of the FMO-UHF Hessian is evaluated in comparison to *ab initio* for stable organic radicals[42] (see the structures in Figure S1 in the Supplementary materials) and a small polypeptide NH_2-IYPIG-COO, which we denote in short as IYPIG (IYPIG is the sequence of 5 amino acids). One amino acid residue was assigned per fragment, except that the N-terminus (isoleucine) and the tyrosine are merged into one fragment Ile1Tyr-2. We note that we distinguish residue fragments in FMO method from conventional residues by adding a hyphen after the name of the residue. They differ by the CO group because fragmentation is done at Cα atoms[21].

For the oxidation of tyrosine (Y) producing a radical doublet, the IR spectra were measured in experiment[43], and *ab initio* calculations were also done for this system[44]. In this Letter, we performed Hessian calculations for several structural isomers of this polypeptide. Computational timings were evaluated for triplet oxygen solvated in water clusters.

The interfragment distance R_{IJ} is defined as the shortest interatomic distance between the two fragments, divided by the sum of the van-der-Waals radii of the two atoms; consequently, both R_{IJ} and $R_{\mathrm{ES-DIM}}$ are unitless. We used the 6-31G(d) basis set with spherical harmonics (ISPHER=1) in all calculations. Geometry optimizations were performed with the threshold OPTTOL set to 10^{-4} a.u./bohr. Fully analytic energy gradients in FMO-UHF, and the Hessian calculations were performed at the equilibrium geometries. CPHF equations were solved in AO basis with the conjugate gradient method. The initial geometries of solvated systems were generated with NWChem 6.1[45]. For making FMO input files, we used Fragit program[46].

D. Results and discussion

1. Accuracy of the FMO-UHF Hessian

We evaluated the accuracy of FMO-UHF vibrational frequencies and IR intensities for solvated stable organic radical molecules, 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) and dimethyl-amino-nitronyl-nitroxide (DMANN). Because the many-body effects in water are significant, DMANN, solvated in water, shows somewhat larger errors than TEMPO, solvated in DMF.

The results are shown in FIG. 1 and TABLE I. *Ab initio* and FMO frequencies are very similar, while the IR intensities show some deviations. The root-mean-square deviation (RMSD) of frequencies between FMO and *ab initio* is about 1.0 and 5.5 cm^{-1} for TEMPO and DMANN, respectively. The maximum deviations are 4.0 and 36.0 cm^{-1} in TEMPO and DMANN, respectively. The RMSD for intensities were about 0.3 and 1.6 for TEMPO and DMANN, respectively, while the maximum intensity deviations were about 2.7 and 8.8. The largest intensity was about 30 and 24 for TEMPO and DMANN, respectively. We observe that the ES-DIM approximation is very accurate: the errors are essentially zero even for the threshold factor $R_{\mathrm{ES-DIM}} = 1.5$, which corresponds to an aggressive use of the approximation.

Because intensities are affected by the delocalization of normal modes, they are more difficult to evaluate accurately. In TEMPO, the largest deviation of frequencies from *ab initio* is found in the low frequency region, in which the vibrational modes are delocalized significantly. Similarly, in DMANN, the maximum errors are for the vibrational modes delocalized over several water molecules. The errors in DMANN are larger than in TEMPO due to hydrogen bonding between water molecules.

The summary of the thermodynamical properties is given in TABLE II. The zero point energies (ZPE), enthalpies H and Gibbs free energies G, obtained with FMO, differ from *ab initio* by at most 0.8, 0.4 and 2.2 kcal/mol, respectively. As we discussed earlier[19], the latter property typically has the largest errors because of low frequencies, which contribute significantly to the entropy. For a better accuracy, larger fragments can be used.

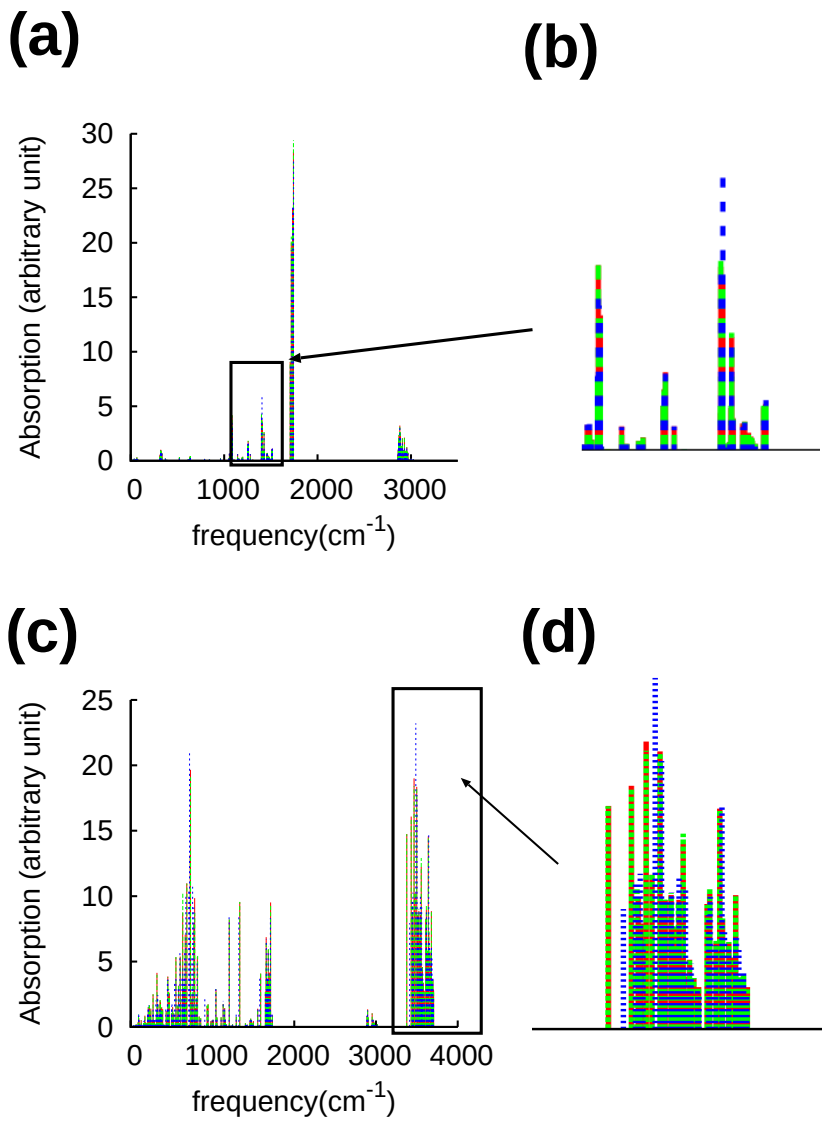


FIG. 1. IR spectra calculated using FMO-UHF with and without the ES-DIM approximation (green dashed and red solid lines, respectively) are compared with *ab initio* UHF (blue dotted line) for TEMPO solvated in explicit DMF molecules: the full spectrum (a) and its magnified part (b); DMANN solvated in explicit water molecules: the full spectrum (c) and its magnified part (d). The ES-DIM threshold is 2.0.

TABLE I. RMSD and the maximum (Max) difference of vibrational frequencies (cm^{-1}) obtained with FMO and *ab initio* Hessians at the UHF/6-31G(d) level for TEMPO solvated in DMF, and DMANN solvated in water. R is the threshold for the electrostatic dimer approximation (ES-DIM).

system	R	frequency		IR intensity	
		RMSD	Max	RMSD	Max
TEMPO	not used	0.6	4.0	0.3	2.5
TEMPO	2.0	0.7	4.0	0.3	2.8
TEMPO	1.5	0.7	4.0	0.3	2.8
DMANN	not used	5.3	36.0	1.5	8.6
DMANN	2.0	5.4	36.0	1.6	8.6
DMANN	1.5	5.5	36.4	1.6	8.7

TABLE II. The thermal correction to zero-point energies (ZPE), enthalpies H and Gibbs free energies G (kcal/mol), computed at the UHF/6-31G(d) level. R is the threshold for the electrostatic dimer approximation in FMO.

method	R	ZPE	H	G
		TEMPO solvated in DMF		
ab initio		526.70	559.92	458.73
FMO	not used	526.74	559.92	459.27
FMO	2.0	526.74	559.92	459.31
FMO	1.5	526.74	559.92	459.31
		DMANN solvated in water		
ab initio		459.74	497.14	399.45
FMO	not used	460.50	497.56	401.16
FMO	2.0	460.52	497.57	401.63
FMO	1.5	460.55	497.59	401.64

2. *IR spectra of the IYPIG polypeptide in gas phase*

We calculated the IYPIG polypeptide in gas phase, representing the active center in the photosynthetic (PS) II D1 subunit (PDB code: 3ARC)[47]. The tyrosine residue can be oxidized, forming a free radical (FIG. 2). Difference spectra between the reactants and products in the oxidation reaction were measured by FTIR[43] and vibrational frequencies of the tyrosine radical were theoretically investigated[44].

We calculated three isomers of IYPIG, to identify which of them has the spectrum most closely resembling the experimental one. The initial structures were the α-helix, β-turn, and extended forms, denoted by **A**, **B**, and **E**, respectively. After geometry optimization was performed, these structures were deformed, in particular **E**, while the secondary structure of the α-helix and β-turn can be recognized in **A** and **B**, respectively.

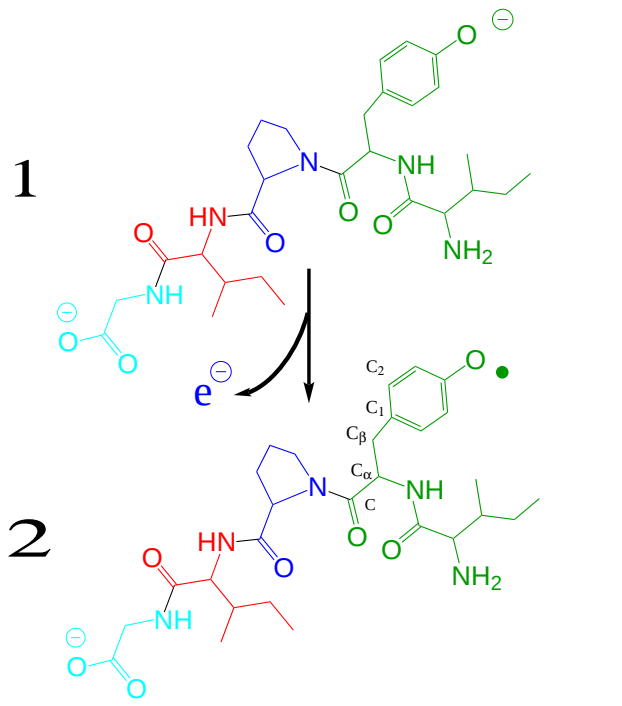
The sum of the electronic and thermal Gibbs free energy differences at the experimental temperature of 77 K between the structural isomers are shown in FIG. 2. We see that for the reactants, there is very little difference and all three isomers are close in energy, **A** being the most stable isomer. On the other hand, upon oxidation **B** is clearly the most stable. For completeness, we evaluated IR spectra for all structural isomers.

TABLE III. Important vibrational modes ν (cm^{-1}) in IYPIG isomers **A**, **B** and **E** in the singlet (S=0) and doublet (S=1/2) states in gas phase for the oxidation reaction, calculated at the UHF/6-31G(d) level and scaled[49] by the factor of 0.8953 for comparison with experimental frequencies in water[43].

mode	S	**A**	**B**	**E**	expt.
ν_{8a}	0	1606.3*	1607.0	1605.6	1602-1605
ν_{19a}	0	1522.9*	1550.0	1538.4	1499-1502
ν_{8a}	1/2	1539.5	1550.0*	1535.3	1550-1560
ν_{7a}	1/2	1519.3	1516.1*	1525.3	1514-1516

* The most stable isomer in terms of Gibbs free energies at 77 K.

The spectra are presented in FIG. 3 and TABLE III. The prominent peaks related to tyrosine vibrational modes are assigned as ν_{8a} and ν_{19a} for singlet, and ν_{8a} and ν_{7a} for doublet



	ΔG (kcal/mol)		
1	0.00	3.65	1.59
2	0.00	−14.42	7.25

FIG. 2. Schematic illustration for the oxidation reaction of the IYPIG polypeptide, where the reactant and product are denoted as 1 and 2, respectively (upper panel). The backbone conformations of the optimized structures **A**, **B**, and **E** are shown in the middle panel. The relative Gibbs free energies for each isomer at 77 K are shown in the lower panel.

(see FIG. 3-(a)) following the experimental study[43] (note that each of them is actually not

a single peak but rather a band in the spectrum).

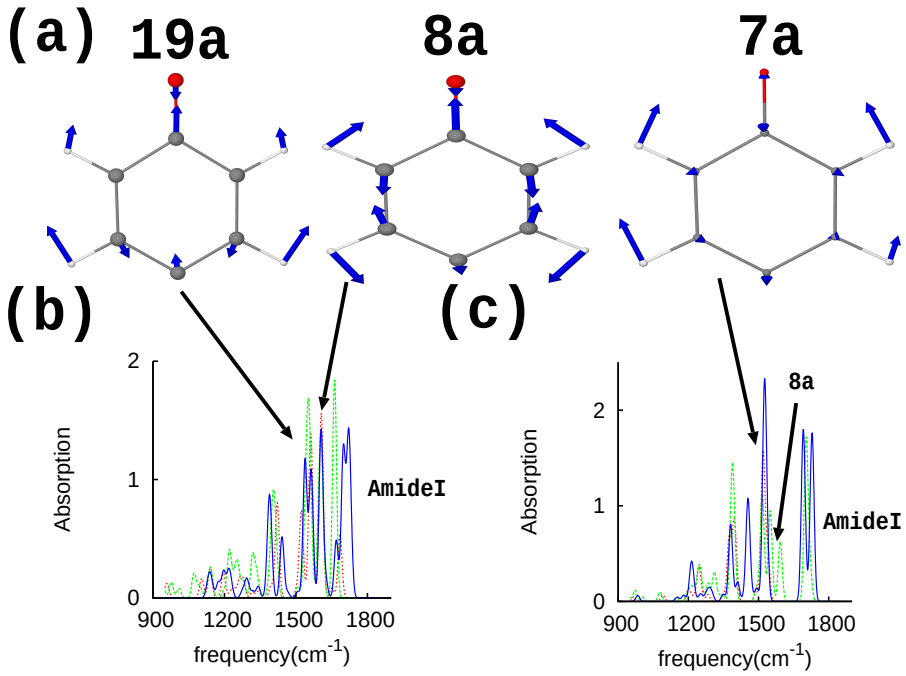


FIG. 3. Vibrational modes of the IYPIG polypeptide in gas phase, calculated with FMO-UHF/6-31G(d): (a) schematic representation, (b) the singlet reactant spectrum and (c) the doublet product spectrum. The spectra of **A**, **B** and **E** isomers are shown as red, green, and blue lines, respectively. We use the notation of[50] for the assignment of the vibrational modes. The frequencies are scaled[49] by the factor of 0.8953.

For the reactant (singlet state), the RMSD between FMO and experimental frequencies is 27.1 cm^{-1}. The most intense peak of ν_{19a} is for the vibrational mode of tyrosine coupled with the Amide II band. Other ν_{19a} peaks more localized on tyrosine have smaller intensities. For the product in the doublet spin state, the RMSD between FMO and experimental frequencies is 11.4 cm^{-1}. The vibrational mode of ν_{7a} is coupled with the Amide II band, but in the doublet this mode ν_{7a} is much more localized on the tyrosine than in ν_{19a}, and the coupling to the Amide II band is weak, which may explain its better accuracy.

For the oxidized polypeptide, the **B** isomer is not only most stable, but its calculated spectrum most closely resembles the experiment. The dihedral angles C-C_α-C_β-C_1, and C_α-

C_β-C_1-C_2 (FIG. 2) are 85 and 179 degrees, respectively, similar to the angles in the most stable conformation of tyrosine (85 and 165 degrees, respectively) from the earlier theoretical study[44]. Summarizing, we conclude that the **B** isomer is the most stable isomer of oxidized IYPIG, whereas for IYPIG itself all three isomers have comparable energy with the **A** isomer being marginally more stable than the other two.

3. Solvent effects on the spectrum of IYPIG

Because IYPIG is negatively charged, it is interesting to investigate the effects of solvent upon its IR spectrum. We optimized geometry and computed the Hessian for the doublet state of the **B** isomer solvated in the 3 Å shell of explicit water (FIG. 4-(a)). The spectrum is shown in FIG. 4-(b). The tyrosine vibrational modes are coupled with those of COO (symmetric), Amide I and II.

The vibrational frequencies of phenol radical in gas phase are in the range from 1450 to 1500 cm^{-1}. The positions of these peaks are not substantially affected by the interaction of phenol with other residues and solvent. The ν_{8a} peak of phenol radical in gas phase is 1492.5 cm^{-1}, whereas for IYPIG solvated in water the value is 1495.2 cm^{-1}, with a solvent shift of 2.7 cm^{-1}. On the other hand, the tyrosine vibrational modes coupled with other amino acid residues are sensitive to the environment. The tyrosine vibrational modes in IYPIG with a large intensity are coupled with the COO symmetric mode (around 1400 cm^{-1}), Amide II (1500-1600 cm^{-1}), and Amide I (1600-1700 cm^{-1}) vibrational modes.

The modes for the three prominent peaks, COO (symmetric), Amide I and II, are coupled with the water bending mode (around 1595 cm^{-1}). Because the COO (symmetric) and Amide II vibrational frequencies are smaller than that of the water bending mode, these frequencies are shifted up. The opposite is the case for Amide I. The solvent shifts of the COO symmetric mode, Amide I (C=O in peptides) and Amide II (N-H coupled to C-N in peptides) are 58.5, -29.7 and 8.0 cm^{-1}, respectively. This trend in solvent shifts is reasonable because it can be expected that the interaction of these functional groups with water decreases in this order.

We can analyze the interaction of solvent with COO^-, by looking at the fragment which

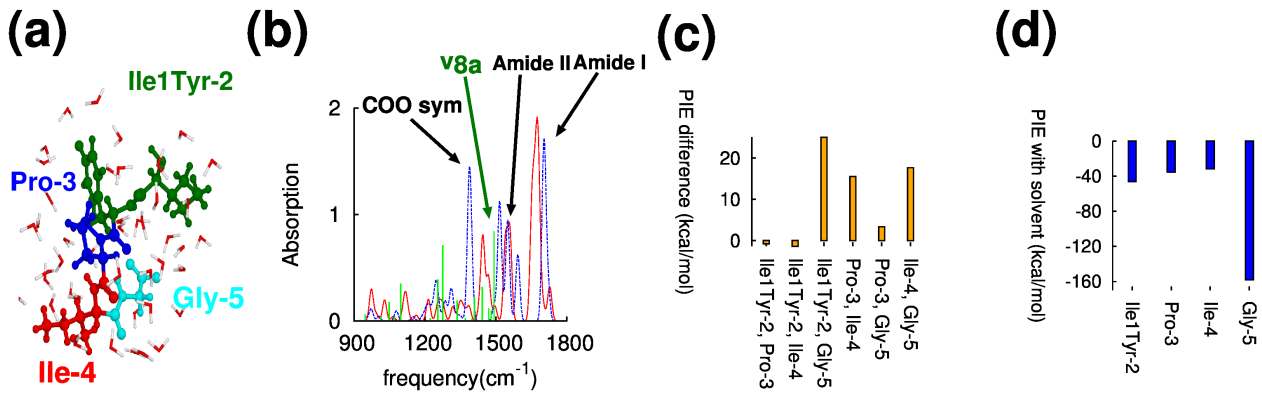


FIG. 4. Solvent effects upon the IR spectrum of IYPIG (the **B** isomer of the doublet oxidized form). (a) IYPIG solvated in water is divided into 4 fragments. (b) Calculated IR spectrum: in gas phase (blue dashed line) and in water (red solid line). For comparison, the spectrum of phenol radical is also shown as green solid line. (c) Pair interaction energy (PIE) difference between solution and gas phase, for fragment pairs in IYPIG. (d) The sum of the PIEs between a fragment in IYPIG and all water molecules.

includes it (Gly-5), assuming that this group is the main contributor. As found in other studies[48], the addition of solvent in general destabilizes the solute by making its intramolecular interactions less attractive. From FIG. 4-(c), one can see that the addition of solvent

destabilized the pair interactions (Eq. 2) between solute fragments by as much as about 25 kcal/mol (Ile1Tyr-2 and Gly-5). On the other hand, there is a strong attractive interaction between the solute and solvent (FIG. 4-(d)) as large as 160 kcal/mol for Gly-5 (the C-terminus containing COO^-), compensating the destabilization, according to the polarization picture in FMO[34]. This large solute-solvent interaction correlates with the large solvent shift of 58.5 cm^{-1}.

4. Computational timing

The computational timings (FIG. 5) are evaluated for a triplet oxygen molecule solvated in cubic boxes of water molecules with linear sizes of 5, 7, 9, 10, 12, 15 and 18 Å containing 20, 44, 125, 143, 230, 434 and 704 atoms, respectively. The timing calculations were performed on 16 nodes equipped with 2.93 GHz Xeons (8 CPU cores and 12 GB RAM per node). By using the electrostatic dimer approximation, which is shown above to be very accurate, the Hessian is speed up the factor of 9 for the largest system.

Ab initio calculations took 21 and 248 seconds for the smallest two systems, compared to 10 and 28 seconds in FMO. *Ab initio* Hessian requires large memory: for instance, for 44 atoms the required amount is 2.65 GB per core, whereas the calculation of 125 atoms requested more than 10 GB, and the calculation was not possible. On the other hand, the largest FMO Hessian calculation for the system of 704 atoms used only 1.12 GB RAM per core. FMO-UHF Hessian is about 9 times faster than full *ab initio* calculations for the largest calculation we could perform with *ab initio* due to the limitation of our computer resources. Because the ES-DIM approximation was developed in this Letter for FMO-UHF Hessian as well, we also did timing tests for FMO-RHF (Figure S2 in the Supplementary materials).

E. Conclusions

We have derived the analytic second derivative of the energy for open-shell systems using the FMO-based UHF method and parallelized in GAMESS. In addition, we have derived

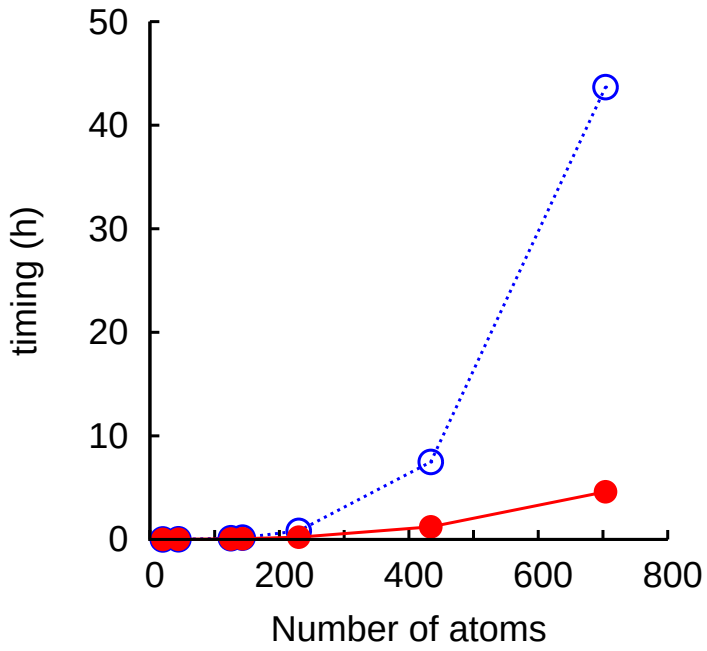


FIG. 5. Computational timing of FMO-UHF/6-31G(d) Hessian calculations of triplet oxygen molecule solvated in explicit water. Red solid and blue dashed lines denote the timings with ($R_{\mathrm{ES-DIM}}$=2.0) and without the ES-DIM approximation, respectively.

the second derivative of the energy for the separated dimer approximation (ES-DIM) for both restricted and unrestricted FMO methods, which is crucial to reduce the cost of FMO Hessians. The accuracy of vibrational frequencies and thermodynamical properties is demonstrated for solvated organic radicals and the method is applied to simulate the IR spectra of a polypeptide related to photosynthesis.

We have shown that the vibrational analysis obtained with FMO-UHF is accurate in comparison to full *ab initio*, although for IR intensities a deviation as much as about 30% is observed for some delocalized vibrational modes. The frequencies are, however, quite accurately reproduced.

We have found that for the neutral singlet IYPIG polypeptide all three isomers have similar energies, whereas for the oxidized anionic doublet species the distorted β-turn is

the most stable isomer, in agreement with experimental findings[43]. We have identified the trend that the solvent shift of vibrational frequencies is correlated with the solute-solvent interactions.

FMO-UHF Hessian is efficiently implemented as shown by the timings tests; the calculation of a molecular system consisting of 704 atoms took about 4.6 hours on 128 CPU cores (2.93 GHz Xeon CPUs). We expect that the FMO-based Hessian will become a useful tool to simulate IR spectra and also to obtain thermodynamic properties such as free energies in large systems containing free radicals, transition metals and other open-shell systems.

Appendix A. Supplementary data

Supplementary data associated with this Chapter can be found, in the online version, at http : //www.sciencedirect.com/science/article/pii/S0009261414003066.

REFERENCES

[1]P. Deglmann, F. Furche, and R. Ahlrichs, Chem. Phys. Lett. **362**, 511 (2002).

[2]Y. Alexeev, M. W. Schmidt, T. L. Windus, and M. S. Gordon, J. Comput. Chem. **28**, 1685 (2007).

[3]P. Pulay, Mol. Phys. **17**, 197 (1969).

[4]A. Warshel and M. Karplus, J. Am. Chem. Soc. **94**, 5612 (1972).

[5]Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).

[6]M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Ann. Rev. Phys. Chem. **64**, 553 (2013), PMID: 23561011.

[7]S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Str.: THEOCHEM **461**, 1 (1999).

[8]H. A. Witek, S. Irle, and K. Morokuma, J. Chem. Phys. **121**, 5163 (2004).

[9]W. J. Zheng and B. R. Brooks, Biophys. J. **89**, 167 (2005).

[10]H. Li and J. H. Jensen, Theor. Chem. Acc. **107**, 211 (2002).

[11]A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, J. Chem. Phys. **126**, 224102 (2007).

[12]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[13]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[14]W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[15]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[16]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101 (2008).

[17]W. Hua, T. Fang, W. Li, J.-G. Yu, and S. Li, J. Phys. Chem. A **112**, 10864 (2008), PMID: 18837491.

[18]S. Sakai and S. Morita, J. Phys. Chem. A **109**, 8424 (2005).

[19]H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, J. Chem. Phys. **138**, 164103 (2013).

[20]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[21]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A. **111**, 6904 (2007).

[22]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[23]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[24]Y. Alexeev, M. P. Mazanetz, O. Ichihara, and D. G. Fedorov, Curr. Top. Med. Chem. **12**, 2013 (2012).

[25]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B **111**, 9621 (2007).

[26]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B **116**, 503 (2012).

[27]P. V. Avramov, D. G. Fedorov, P. B. Sorokin, S. Sakai, S. Entani, M. Ohtomo, and H. N. Y. Matsumoto, J. Phys. Chem. Lett. **3**, 2003 (2012).

[28]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[29]D. G. Fedorov, Y. Alexeev, and K. Kitaura, J. Phys. Chem. Lett. **2**, 282 (2011).

[30]Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **372**, 342 (2003).

[31]Y. Komeiji, T. Ishikawa, Y. Mochizuki, H. Yamataka, and T. Nakano, J. Comput. Chem. **30**, 40 (2009).

[32]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[33]H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, J. Chem. Phys. **137**, 044110 (2012).

[34]D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).

[35]M. C. Green, D. G. Fedorov, K. Kitaura, J. S. Francisco, and L. V. Slipchenko, J. Chem. Phys. **138**, 074111 (2013).

[36]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, and J. Goddard, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press, New York, 1994.

[37]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **351**, 475 (2002).

[38]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **475**, 124 (2009).

[39]N. W. Schmidt, K. K. Baldridge, J. A. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[40]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[41]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004).

[42]J. Clayden, N. Greeves, S. Warren, and P. Wothers, *Organic Chemistry*, Oxford university press, New York, 2001.

[43]I. R. Vassiliev, A. R. Offenbacher, and B. A. Barry, J. Phys. Chem. B **109**, 23077 (2005).

[44]K. Range, I. Ayala, D. York, and B. A. Barry, J. Phys. Chem. B **110**, 10970 (2006), PMID: 16771350.

[45]M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, Comput. Phys.

Commun. **181**, 1477 (2010).

[46]C. Steinmann, M. W. Ibsen, A. S. Hansen, and J. H. Jensen, PLOS ONE **7**, e44480 (2012).

[47]Y. Umena, K. Kawakami, J.-R. Shen, and N. Kamiya, Nature **473**, 55 (2011).

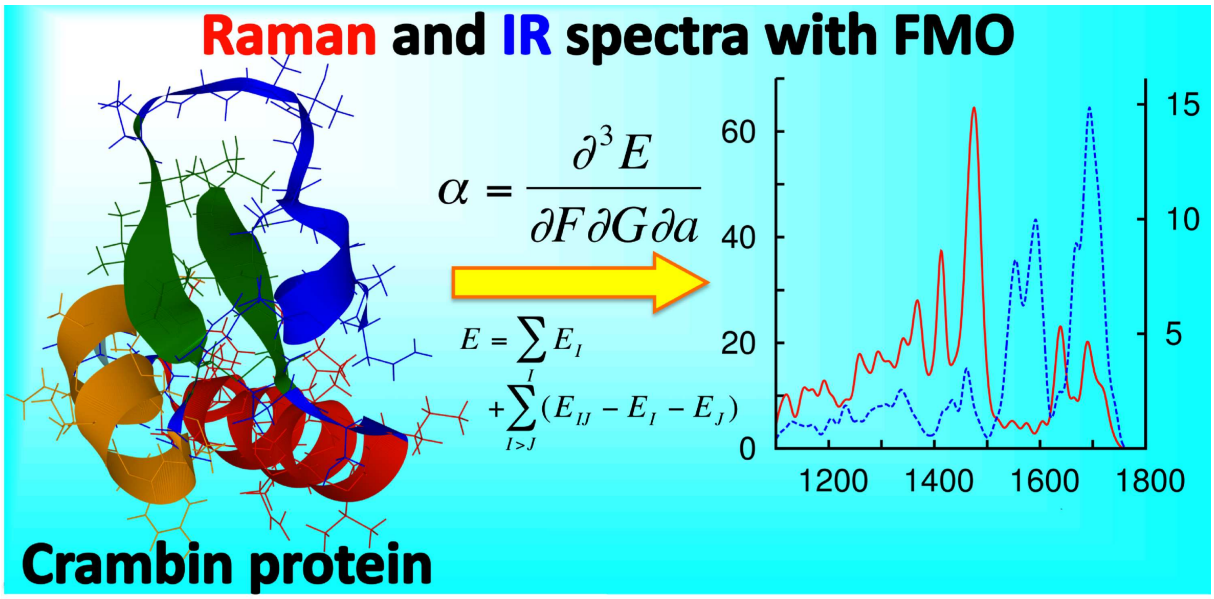
[48]Y. Komeiji, T. Ishida, D. G. Fedorov, K. Kitaura, and N. Kamiya, J. Comp. Chem. **28**, 1750 (2007).

[49]A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).

[50]G. N. R. Tripathi and R. H. Schuler, J. Chem. Phys, **81**, 113 (1984).

Chapter VII

Simulations of Raman spectra using the fragment molecular orbital method



H. Nakata, D. G. Fedorov *et. al.*

J. Chem. Theor. Comput. **2014**, *10*, 3689-3698.

DOI : 10.1021/ct5003829

View online: http://pubs.acs.org/doi/abs/10.1021/ct5003829

VII. SIMULATIONS OF RAMAN SPECTRA USING THE FRAGMENT MOLECULAR ORBITAL METHOD

A. Introduction

Vibrational infrared (IR)[1] and Raman[2,3] spectroscopies play an important role in chemistry. Quantum-mechanical (QM) simulations of infrared and Raman spectra require the second derivative of the energy (the Hessian) with respect to nuclear coordinates.[4–6] Reducing high computational cost, various alternatives to full *ab initio* calculations of the Hessian are available such as combined QM and molecular mechanics (QM/MM),[7,8] our own n-layered integrated molecular orbital and molecular mechanics (ONIOM) method,[9] and density functional based tight binding (DFTB).[10] It is also possible to evaluate the Hessian numerically, usually using analytic energy gradients, which can also be done for a subset of atoms.[11–13]

A calculation of IR intensities or Raman activities requires derivatives of both dipole moment and polarizability tensor with respect to normal vibrational coordinates. The latter derivative is the third derivative of the energy, and it can be calculated for Hartree-Fock,[14,15] and density functional theory.[16,17] *Ab initio* Raman calculations have been conducted for polypeptides representing biological systems.[18–20] There are various efficient quantum-mechanical methods such as semiempirical approaches,[21,22] linear-scaling algorithms,[23,24] and fragment-based methods.[25–37] However, analytic second derivatives are available only for a few of these methods.[38–42]

For the fragment molecular orbital (FMO) method[43–47] the analytic first[48,49] and second[42,50] derivatives of the energy have been developed. Dynamic polarizabilities can also be calculated.[51] FMO has been applied to a variety of molecules including proteins,[52,53] DNA[54] and inorganic systems.[55–59]

In this work, we develop a method to simulate Raman spectra with FMO using both restricted[43] and unrestricted[60,61] Hartree-Fock (HF). We use the notation of HF throughout this work when we mean either RHF or UHF. Because the dispersion interaction is very important in chemistry, we employ the Grimme's empirical dispersion energy.[62,63] To the

best of our knowledge, the analytic second derivative for this dispersion model has not been published and we derive it in this work. We also derive the coupled perturbed Hartree-Fock (CPHF) equations[64] under the electric field perturbation in the framework of FMO, and use them to obtain the third derivatives of the energy necessary for Raman activities.

The accuracy of Raman spectra computed with FMO is evaluated for vitamin C[65] and α-tocopherol[66] (vitamin E) for closed-shell restricted HF (RHF), and 2,2,6,6-tetramethylpiperidine 1-oxyl (TEMPO) for open-shell unrestricted HF (UHF). The computational efficiency is demonstrated for water clusters. We apply our developed method to a polystyrene oligomer and crambin (PDB: 1CRN),[67] and discuss insights into the Raman spectra of large systems provided by FMO.

B. Theory

1. FMO summary

In FMO, the system is divided into fragments (also called monomers), which are calculated in the embedding electrostatic potential (ESP), obtained from the density of all fragments. These monomer calculations are repeated iteratively until convergence, followed by the calculation of fragment pairs (dimers) in the presence of ESP. More details can be found elsewhere.[42] The total energy in FMO is expanded in the many-body series[68,69] and in this work we use the two-body expansion (FMO2).

The total energy E in FMO2 is

$$E^{\mathrm{FMO2}} = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J}^{N} \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}), \tag{1}$$

where E'_X is the internal fragment energy of fragment X (X=I or IJ). N is the number of fragments. The internal energies are obtained from self-consistent field (SCF) energies E_X after SCF converges by subtracting the interaction with the embedding potential (ESP) $\mathbf{V}^X$ as

$$E'_X = E_X - Tr(\mathbf{D}^X \mathbf{V}^X), \tag{2}$$

where $\mathbf{D}^X$ is the density matrix of X and

$$V^X_{\mu\nu} = \sum_{K\neq X}^{N} \left(u^K_{\mu\nu} + v^K_{\mu\nu}\right). \tag{3}$$

The one-electron u and two-electron v integrals are

$$u^K_{\mu\nu} = \sum_{A\in K} \langle\mu| \frac{-Z_A}{|\mathbf{r}-\mathbf{R}_A|} |\nu\rangle, \tag{4}$$

$$v^K_{\mu\nu} = \sum_{\lambda\sigma\in K} D^K_{\lambda\sigma} (\mu\nu|\lambda\sigma), \tag{5}$$

where $\mathbf{R}_A$ and Z_A are the coordinates and nuclear charge of atom A, respectively. A, B, and C are used to label atoms, while Greek indices are used for atomic orbitals. $(\mu\nu|\lambda\sigma)$ is the two-electron integral in atomic orbital basis. $\Delta\mathbf{D}^{IJ}$ is the difference between the electron density matrix of dimer IJ and the sum of the densities of monomers I and J.

The second derivative of the energy in eq 1 with respect to the nuclear coordinates a and b is (more details are given elsewhere[42]),

$$\frac{\partial^2 E^{\mathrm{FMO2}}}{\partial a\partial b} = \sum_{I}^{N} \frac{\partial^2 E'_I}{\partial a\partial b} + \sum_{I>J}^{N} \left(\frac{\partial^2 E'_{IJ}}{\partial a\partial b} - \frac{\partial^2 E'_I}{\partial a\partial b} - \frac{\partial^2 E'_J}{\partial a\partial b}\right) + \sum_{I>J}^{N} \frac{\partial^2 \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a\partial b}, \tag{6}$$

The second derivative of the internal energy E'_X is

$$\frac{\partial^2 E'_X}{\partial a\partial b} = \frac{\partial^2 E'^{\mathrm{HF}}_X}{\partial a\partial b} + \frac{\partial^2 E^{\mathrm{dis}}_X}{\partial a\partial b} - \overline{U}^{ab,X}, \tag{7}$$

where E^{HF}_X and E^{dis}_X are the HF and dispersion energies, respectively, and $\overline{U}^{ab,X}$ is the Hessian contribution arising from the orbital responses related to the derivative of MO coefficients in ESP. These responses are FMO-specific terms, whereas HF Hessians are calculated for fragments and their pairs as in regular RHF (UHF) calculations. In order to obtain the response terms, CPHF equations have to be solved.[42] The second derivatives of the dispersion terms in eq 7 are derived in this work (see below).

2. Raman activity calculation

The normal Raman activity[14,15] of a vibrational mode i can is

$$J_i = 45\alpha'^2_i + 7\gamma'^2_i. \tag{8}$$

The symmetric contribution α_i' is defined as

$$\alpha_i' = \frac{1}{3}\left[\left(\frac{\partial \alpha_{xx}}{\partial Q_i}\right) + \left(\frac{\partial \alpha_{yy}}{\partial Q_i}\right) + \left(\frac{\partial \alpha_{zz}}{\partial Q_i}\right)\right], \tag{9}$$

The polarizability tensor α_{xy} is the second derivative of the energy with respect to the external electric fields $\mathcal{F}_x$ and $\mathcal{F}_y$

$$\alpha_{xy} = \frac{\partial^2 E}{\partial \mathcal{F}_x \partial \mathcal{F}_y}, \tag{10}$$

and $\partial \alpha_{xy}/\partial Q_i$ is the derivative of the polarizability tensor with respect to a normal coordinate Q_i, which is a linear combination of atomic coordinates.

$$Q_i = \sum_{a=1}^{3M} l_{ai} a, \tag{11}$$

where a is a Cartesian coordinate of an atom, M is the number of atoms, and l_{ai} is the eigenvector of the Hessian for the normal mode i. The polarizability tensor derivatives $\partial \alpha_{xy}/\partial Q_i$ are obtained from the Cartesian derivative $\partial \alpha_{xy}/\partial a$ as

$$\frac{\partial \alpha_{xy}}{\partial Q_i} = \sum_{a=1}^{3M} l_{ai} \frac{\partial \alpha_{xy}}{\partial a}. \tag{12}$$

The anisotropic derivative contribution γ_i' is defined as

$$\begin{aligned}\gamma_i'^2 = \frac{1}{2}\Bigg[&\left(\frac{\partial \alpha_{xx}}{\partial Q_i} - \frac{\partial \alpha_{yy}}{\partial Q_i}\right)^2 + \left(\frac{\partial \alpha_{yy}}{\partial Q_i} - \frac{\partial \alpha_{zz}}{\partial Q_i}\right)^2 + \left(\frac{\partial \alpha_{zz}}{\partial Q_i} - \frac{\partial \alpha_{xx}}{\partial Q_i}\right)^2 \\ &+6\left\{\left(\frac{\partial \alpha_{xy}}{\partial Q_i}\right)^2 + \left(\frac{\partial \alpha_{yz}}{\partial Q_i}\right)^2 + \left(\frac{\partial \alpha_{zx}}{\partial Q_i}\right)^2\right\}\Bigg].\end{aligned} \tag{13}$$

In order to obtain normal coordinates associated with vibrational frequencies, it is necessary to calculate the second derivatives of the energy with respect to nuclear coordinates a and b and transform them into mass-weighted Cartesian coordinates using atomic masses m_a,

$$H'_{ab} = \frac{\partial^2 E}{\partial a \partial b}/\sqrt{m_a m_b}. \tag{14}$$

By diagonalizing the mass-weighted Hessian matrix $\mathbf{H}'$, one obtains the vibrational frequencies and the normal coordinates as its eigenvalues and eigenvectors, respectively. To calculate the Raman activity, one has to obtain both the Hessian matrix in eq 14 and the third derivative of the energy with respect to pairs of the electric field components $\mathcal{F}$ and a nuclear coordinate a.

3. Second derivative of the dispersion energy

There are the original[62] and modified[63] formulations of the dispersion (dis) energy, which we denote by D2 and D3, respectively. Both of them are completely independent of the electronic state and thus their addition is not coupled to it in any way. The first and second derivative of the dispersion energy are added to the corresonding electronic contributions.

$$E_{\rm dis}^{\rm D2} = - s_6 \sum_{A,B} \frac{C_6^{AB}}{R_{AB}^6} f_{d,6}^{\rm D2}(R_{AB}), \tag{15}$$

$$E_{\rm dis}^{\rm D3} = - \sum_{n=6,8,10} \sum_{A,B} s_n \frac{C_n^{AB}}{R_{AB}^n} f_{d,n}^{\rm D3}(R_{AB}), \\ - \sum_{A,B,C} f_{d,3}^{\rm D3}(\overline{R}_{ABC}) E^{ABC}, \tag{16}$$

where C_n^{AB} denotes the averaged n-order dispersion coefficient for an atomic pair AB. R_{AB} and $\overline{R}_{ABC}$ are the pairwise and averaged (over three pairs) internuclear distances, respectively. s_n is a scale factor (a parameter). $f_{d,6}^{\rm D2}(R_{AB})$ and $f_{d,n}^{\rm D3}(R_{AB})$ are the damping functions

$$f_{d,6}^{\rm D2}(R_{AB}) = \frac{1}{1 + \exp\left(-dR_{AB}/(s_R R_{AB}^0) - 1\right)}, \tag{17}$$

$$f_{d,n}^{\rm D3}(R_{AB}) = \frac{1}{1 + 6\left(R_{AB}/(s_{R,n} R_{AB}^0)\right)^{-\alpha_n}}, \tag{18}$$

where R_{AB}^0 is the sum of the atomic van-der-Waals radii derived from *ab initio* calculations[70] and scaled by the factor of 1.1. s_R and $s_{R,n}$ are the multipliers for the van-der-Waals radii. d is a damping factor. α_n are integer parameters: $\alpha_3 = 16$, $\alpha_6 = 14$, and for $n > 6$, $\alpha_{n+2} = \alpha_n + 2$. The third-order energy for three atoms ABC is

$$E^{ABC} = \frac{C_9^{ABC}(3\cos\theta_a\cos\theta_b\cos\theta_c + 1)}{(R_{AB}R_{BC}R_{CA})^3}, \tag{19}$$

where θ_a, θ_b, and θ_c are the angles in the triangle formed by the three atoms A, B and C. C_9^{ABC} is the triple-dipole constant.

The analytic second derivative of $E_{\mathrm{dis}}^{\mathrm{D2}}$ is

$$\begin{aligned}
\frac{\partial^2 E_{\mathrm{dis}}^{\mathrm{D2}}}{\partial a \partial b} = & -\sum_A \sum_B s_6 C_6^{AB} \left[\frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \frac{\partial \Delta R_{AB,\{b\}}}{\partial b} \right. \\
& \left(\frac{d}{s_R R_{AB}^0} (f_{d,6}^{\mathrm{D2}})^2 \exp\left[-\frac{dR_{AB}}{s_R R_{AB}^0} - 1 \right] \frac{1}{R_{AB}^7} - 6 \frac{f_{d,6}^{\mathrm{D2}}}{R_{AB}^8} \right) \\
& + \frac{2d}{s_R R_{AB}^0} \exp\left[-\frac{dR_{AB}}{s_R R_{AB}^0} - 1 \right] \frac{f_{d,6}^{\mathrm{D2}}}{R_{AB}^7} \frac{\partial f_{d,6}^{\mathrm{D2}}}{\partial b} \Delta R_{AB,\{a\}} \frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \\
& + \frac{d}{s_R R_{AB}^0} \frac{\partial \left(\exp\left[-\frac{dR_{AB}}{s_R R_{AB}^0} - 1 \right] \right)}{\partial b} \frac{(f_{d,6}^{\mathrm{D2}})^2}{R_{AB}^7} \Delta R_{AB,\{a\}} \frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \\
& - 7 \frac{d}{s_R R_{AB}^0} (f_{d,6}^{\mathrm{D2}})^2 \exp\left[-\frac{dR_{AB}}{s_R R_{AB}^0} - 1 \right] \frac{1}{R_{AB}^9} \\
& \times \Delta R_{AB,\{a\}} \Delta R_{AB,\{b\}} \frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \frac{\partial \Delta R_{AB,\{b\}}}{\partial b} - 6 \frac{\partial f_{d,6}^{\mathrm{D2}}}{\partial b} \frac{\Delta R_{AB,\{a\}}}{R_{AB}^8} \frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \\
& \left. + 48 \frac{f_{d,6}^{\mathrm{D2}}}{R_{AB}^{10}} \Delta R_{AB,\{a\}} \Delta R_{AB,\{b\}} \frac{\partial \Delta R_{AB,\{a\}}}{\partial a} \frac{\partial \Delta R_{AB,\{b\}}}{\partial b} \right], \qquad (20)
\end{aligned}$$

where $\Delta R_{AB,\{a\}}$ denotes the difference in a specific Cartesian coordinate (not the interatomic distance). An atomic coordinate a (from 1 to $3M$) corresponds to atom C_a (from 1 to M) with the Cartesian coordinate component p_a (where $p_a = x, y, z$).

$$\Delta R_{AB,\{a\}} = R_{A,p_a} - R_{B,p_a}, \qquad (21)$$

where R_{A,p_a} is the p_a Cartesian component of the coordinate of atom A. It is convenient to calculate the derivatives of $f_{d,6}^{\mathrm{D2}}$ and $\exp(-dR_{AB}/(s_R R_{AB}^0) - 1)$ first and then evaluate the second derivative of $E_{\mathrm{dis}}^{\mathrm{D2}}$. See supporting information for more details on the derivation of eq 20.

We verified that the analytic second derivative of $E_{\mathrm{dis}}^{\mathrm{D2}}$ is correct by comparing the results to numeric derivatives. We did not develop the analytic second derivative for $E_{\mathrm{dis}}^{\mathrm{D3}}$ in this study. Instead, we calculate numeric second derivatives for it, using analytic first derivatives. Such an implementation is general and can be used for any future modifications of the dispersion energy.

4. Derivative of the polarizability tensor

The derivative of the polarizability tensor α is the third derivative of the energy with respect to a coordinate a and two components of the electric field $\mathcal{F}_i$ ($i = x, y$ and z). We obtain it by calculating the energy gradient in the presence of the electric field and differentiating this gradient numerically twice (double differencing[14,15] for $\mathcal{F}_x\mathcal{F}_x$, $\mathcal{F}_y\mathcal{F}_y$, $\mathcal{F}_z\mathcal{F}_z$, $\mathcal{F}_x\mathcal{F}_y$, $\mathcal{F}_y\mathcal{F}_z$, and $\mathcal{F}_z\mathcal{F}_x$). In total, 19 single point energy gradient calculations are necessary, varying these fields (1 for no field, 6 for the symmetric terms such as $\mathcal{F}_x\mathcal{F}_x$ and 12 for the asymmetric terms such as $\mathcal{F}_x\mathcal{F}_y$). All fragment calculations in this subsection are performed with the field included, so that for instance, $\mathbf{D}^{IJ}$ is the density matrix of fragment IJ, calculated with the field. Also, all explicit derivations below are provide for RHF, and we omit the corresponding UHF equations.

In FMO, the energy gradient in the presence of the electric field $\mathcal{F}$ is

$$\frac{\partial E}{\partial a} = \sum_I^N \frac{\partial E'_I}{\partial a} + \sum_{I>J}^N \left(\frac{\partial E'_{IJ}}{\partial a} - \frac{\partial E'_I}{\partial a} - \frac{\partial E'_J}{\partial a} \right) + \sum_{I>J}^N \frac{\partial \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ})}{\partial a}, \tag{22}$$

and the derivative of the internal fragment energy is

$$\begin{aligned} \frac{\partial E'_X}{\partial a} = & \sum_{\mu\nu} D^X_{\mu\nu} h^{a,X}_{\mu\nu} + \sum_{\mu\nu\in X} D^X_{\mu\nu} P^{a,X}_{\mu\nu} \\ & + \frac{1}{2}\left[D^X_{\mu\nu} D^X_{\lambda\sigma} - \frac{1}{2} D^X_{\mu\lambda} D^X_{\nu\sigma} \right] (\mu\nu|\lambda\sigma)^a \\ & + \sum_{\mu\nu} D^X_{\mu\nu} \mathcal{F} \cdot (\mu|\hat{\mu}^0|\nu)^a \\ & - 2 \sum_{ij\in X}^{\mathrm{occ}} S^{a,X}_{ij} F'^X_{ji} - \overline{U}^{a,X,X} + \frac{\partial E^{\mathrm{NR}}_X}{\partial a}, \end{aligned} \tag{23}$$

where F'^X_{ji}, $S^{a,X}_{ij}$ and E^{NR}_X are the internal Fock matrix elements (related to the full Fock matrix as $F^X_{ji} = F'^X_{ji} + V^X_{ji}$), the derivative overlap integrals, and the nuclear repulsion energy of fragment X, respectively.

The $\hat{\mu}^0$ is the permanent dipole moment operator, and $\overline{U}^{a,X,Y}$ is defined as

$$\overline{U}^{a,X,Y} = 4 \sum_{i\in X}^{\mathrm{occ}} \sum_{r\in Y}^{\mathrm{vir}} U^{a,X}_{ri} V^Y_{ri}, \tag{24}$$

where $U_{ri}^{a,X}$ are the response terms related to the change in the MO coefficients with respect to a change in nuclear coordinates. The ESP of X in MO basis

$$V_{ri}^{X} = \sum_{\mu\nu \in X} C_{\mu r}^{X*} C_{\nu i}^{X} V_{\mu\nu}^{X}, \tag{25}$$

where $C_{\mu r}^{X}$ are the expansion coefficients of MO i in the basis of AOs μ, in fragment X. Throughout, i, j, k, and l denote occupied molecular orbitals, while r is a virtual molecular orbital. h denotes the one-electron core Hamiltonian, whereas P is the projection operator used to divide the AO space for atoms at the fragment border.[71]

The differentiation of the embedded density difference term in eq 22 gives

$$\begin{aligned} \frac{\partial \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ})}{\partial a} =& \sum_{\mu\nu \in IJ} \frac{\partial D_{\mu\nu}^{IJ} V_{\mu\nu}^{IJ}}{\partial a} - \sum_{\mu\nu \in I} \frac{\partial D_{\mu\nu}^{I} V_{\mu\nu}^{IJ}}{\partial a} \\ & - \sum_{\mu\nu \in J} \frac{\partial D_{\mu\nu}^{J} V_{\mu\nu}^{IJ}}{\partial a} \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{\partial D_{\mu\nu}^{X} V_{\mu\nu}^{IJ}}{\partial a} =& 2 \sum_{i \in X}^{\mathrm{occ}} V_{ii}^{a,IJ} + 4 \sum_{r \in X}^{\mathrm{vir}} \sum_{i \in X}^{\mathrm{occ}} U_{ri}^{a,X} V_{ri}^{IJ} \\ & + 8 \sum_{k \in IJ} \sum_{K \neq I,J} \sum_{r \in K}^{\mathrm{vir}} \sum_{i \in K}^{\mathrm{occ}} U_{ri}^{a,K} (kk|ri), \end{aligned} \tag{27}$$

where the derivative ESP in MO basis is

$$V_{ri}^{a,IJ} = \sum_{\mu\nu \in IJ} C_{\mu r}^{IJ*} C_{\nu i}^{IJ} V_{\mu\nu}^{a,IJ}. \tag{28}$$

The derivative ESP in AO basis is

$$V_{\mu\nu}^{a,IJ} = \sum_{K \neq I,J}^{N} \left[\sum_{A \in K} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle^{a} + \sum_{\lambda,\sigma \in K} D_{\lambda\sigma}^{K} (\mu\nu|\lambda\sigma)^{a} \right]. \tag{29}$$

Inserting eqs 23 and 26 into eq 22, one can obtain the analytic energy gradient in the presence of the electric field $\mathcal{F}$, for which one has to calculate the response terms $U_{ri}^{a,K}$. Therefore, solving CPHF equations is necessary with the inclusion of the external electric field $\mathcal{F}$.

5. CPHF equations for the electric field

For the FMO energy gradient in both RHF and UHF, the first order CPHF equations involve only monomer response terms (i.e., no dimer responses).[48,60] The CPHF equations are obtained from the derivative of the Fock matrix in the presence of electric field $\mathcal{F}$ with respect to a nuclear coordinate a.

$$\begin{aligned}\frac{\partial F_{ij}^{I}}{\partial a} =& F_{ij}^{a,I} - (\epsilon_j^I - \epsilon_i^I) U_{ij}^{a,I} \\ & - \frac{1}{2}\sum_{k\in I}^{\mathrm{occ}}\sum_{l\in I}^{\mathrm{occ}} S_{kl}^{a,I} A'_{ij,kl} + \sum_{r\in I}^{\mathrm{vir}}\sum_{k\in I}^{\mathrm{occ}} U_{rk}^{a,I} A'_{ij,rk} \\ & + \frac{1}{2}\sum_{K\neq I}\sum_{k\in K}^{\mathrm{occ}}\sum_{l\in K}^{\mathrm{occ}} S_{kl}^{a,K} A_{ij,kl}^{I,K} - \sum_{K\neq I}\sum_{r\in K}^{\mathrm{vir}}\sum_{k\in K}^{\mathrm{occ}} U_{rk}^{a,K} A_{ij,rk}^{I,K},\end{aligned} \tag{30}$$

where ϵ_i^X is the energy of MO i in fragment X, $S_{kl}^{a,K}$ are the derivatives of overlap integrals. The off-diagonal part of the orbital Hessian $A_{ij,kl}^{I,K}$ is

$$A_{ij,kl}^{I,K} = -4(ij|kl), \tag{31}$$

where $(ij|kl)$ is a two-electron integral in MO basis and the integral derivative contribution to the Fock matrix gradient is

$$\begin{aligned}F_{ij}^{a,I} =& h_{ij}^{a,I} + V_{ij}^{a,I} + P_{ij}^{a,I} + \mathcal{F}\cdot(i|\hat{\mu}^0|j)^a \\ & + \sum_{k\in I}^{\mathrm{occ}}[2(ij|kk)^a - (ik|jk)^a].\end{aligned} \tag{32}$$

Using the relationship of $\partial F_{ij}^I/\partial a = 0$ for $i \neq j$ because of the diagonality, one obtains the CPHF equation in the presence of the electric field $\mathcal{F}$,

$$\mathbf{B}_0^a = \mathbf{AU}, \tag{33}$$

where the diagonal block of the matrix $\mathbf{A}$ for fragment I is

$$A_{ij,kl}^{I,I} = \delta_{ik}\delta_{jl}(\epsilon_j^I - \epsilon_i^I) - A'_{ij,kl} \tag{34}$$

$$A'_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk), \tag{35}$$

The components of $\mathbf{B}_0^a$ are

$$B_{0,ij}^{a,I} = F_{ij}^{a,I} - S_{ij}^{a,I}\epsilon_j^I - \sum_{k\in I}^{\text{occ}}\sum_{l\in I}^{\text{occ}} S_{kl}^{a,I}[2(ij|kl) - (ik|jl)] - \sum_{K\neq I}\sum_{k\in K}^{\text{occ}}\sum_{l\in K}^{\text{occ}} 2S_{kl}^{a,K}(ij|kl). \quad (36)$$

Compared to the FMO gradient without the field, the necessary modification of the CPHF equations is the addition of the derivative of the dipole moment, $\mathcal{F}\cdot(i|\hat{\mu}_x^0|j)^a$ in eq 32. The CPHF equations are solved using the Z-vector method.[64]

Summarizing, for the normal Raman activity in eq 8, one has to perform geometry optimization, calculate the Hessian and do normal vibrational analysis and then evaluate the polarizability tensor α according to eqs 9 and 13. In this work we used the most recent formulation of the Grimme's dispersion energy[63] available in GAMESS and because of the triatomic terms (see eq 20) we numerically differentiated analytic derivatives of the dispersion energy, although we also implemented the analytic second derivative for the earlier version of the dispersion energy[62] in eq 16. The third derivative of the polarizability tensor α is numerically obtained by performing 19 energy gradient calculations in the presence of the electric field, for which it is necessary to use eqs 23, 32, and 36. The CPHF equations are solved using the conjugate gradient method in AO basis.

C. Computational details

We implemented the calculation of normal Raman activities for both RHF and UHF formulations of FMO in GAMESS[72,73] and parallelized it with the generalized distributed data interface (GDDI).[74] FragIt program was used for making FMO input files.[75] In order to analyze Raman spectra, intense Raman active normal modes were localized by the unitary transformation method,[76] and the respective transformed vibrational modes and their frequencies were used as for the peak assignment. The unitary transformations were performed with Movipac program.[77]

The accuracy of FMO-HF is evaluated in comparison with full HF calculations. For this purpose, the following three systems were used: vitamin C solvated in 30 water molecules

as a test of FMO-RHF, TEMPO solvated in N,N-dimethylformamide (DMF) as a test of FMO-UHF and α-tocopherol (vitamin E) as a test of FMO-RHF with fragments connected by covalent bonds (See FIG. 1 for the fragmentation details). The computational efficiency is demonstrated on box-shaped water clusters, with the linear box sizes from 5 to 17 Å (from 24 to 495 atoms). For applications, we calculated a polystyrene oligomer and crambin (PDB code:1CRN),[67] with 223 and 642 atoms, respectively.

We used the default values of various thresholds. For the numeric second derivative of the dispersion energy, we did double differencing with the offset of 10^{-6} a.u., while for the numeric derivatives of the gradient with respect to the electric field we used double differencing with the offset of 0.002 a.u. For the separated dimer approximation[78] we used the threshold of 2.0 (it is unitless, because it is applied to unitless distances computed as the usual values divided by the sum of the atomic van-der-Waals radii). All geometries were optimized with the geometry convergence threshold OPTTOL of 10^{-4} hartree/bohr. For all calculations we used the 6-31G(d) basis set with spherical harmonics (ISPHER=1), and the Grimme's dispersion model[63] was employed to consider the dispersion contributions. The geometry optimization of crambin was performed with the polarizable continuum model (PCM)[79,80] using the default PCM settings at the FMO-PCM$\langle 1 \rangle$ level.[80]

TEMPO solvated in explicit DMF molecules is fragmented as one molecule per fragment, while 2 water molecules are merged into one fragment for vitamin C solvated in explicit water. This was done according to the pair interaction energies (PIE) in FMO[81], by merging the pair with the largest PIE first, then the second largest and so on. The fragmentation pattern of vitamin C is shown in FIG. 1.

We used the Gaussian broadening in the IR and Raman spectra simulations. The contribution of mode i to the IR intensity I (and likewise Raman activity) is calculated as

$$I(\omega) = \frac{I_i}{\eta} \exp\left[-\left(\frac{\omega - \omega_i}{\eta} \right)^2 \right], \tag{37}$$

where I_i and ω_i are the IR intensity and frequency for the normal vibrational mode i, η is the broadening parameter for the linewidth (10 cm^{-1} is used throughout this study).

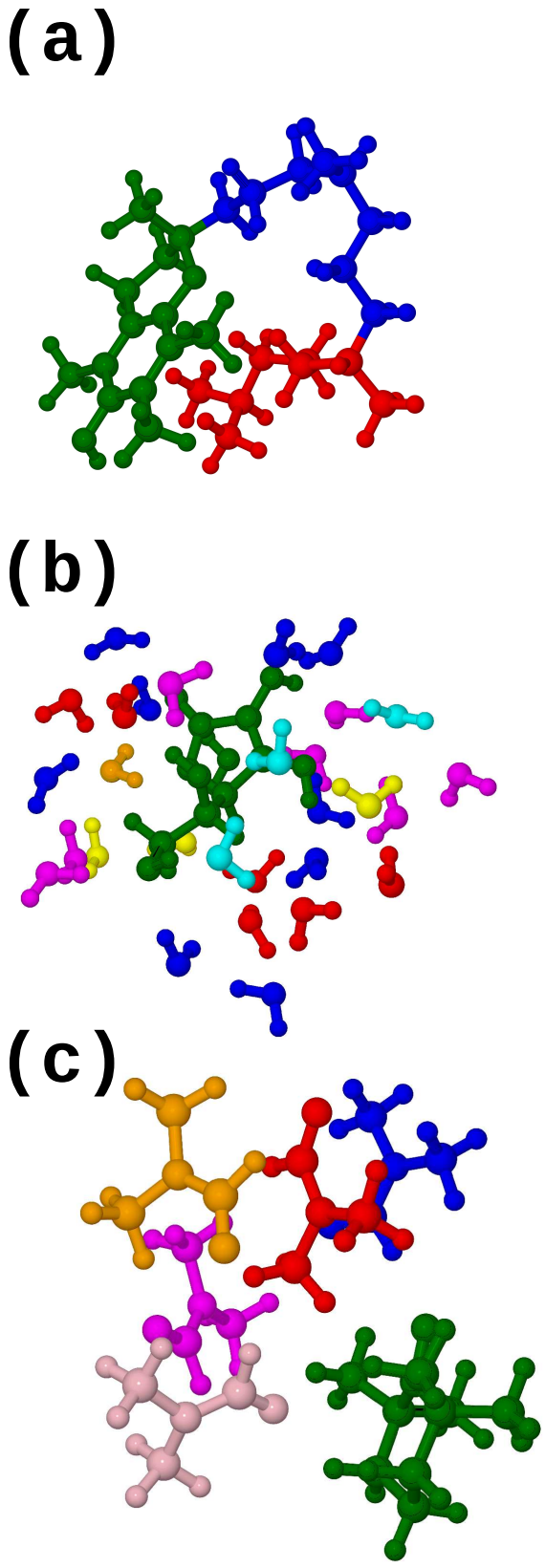


FIG. 1. Fragmentation of the molecular systems showing fragments in different colors. (a) vitamin E, (b) vitamin C solvated in water, and (c) TEMPO solvated in DMF.

D. Results and discussion

1. FMO accuracy

The results shown in FIG. 2 demonstrate the FMO-HF accuracy in comparison with full HF calculations for vitamin C solvated in water molecules, α-tocopherol, and TEMPO solvated in DMF. Among all normal modes, the maximum FMO-HF deviation from full HF for vitamin C, α-tocopherol and TEMPO is 27.1, 6.2 and 7.1 cm^{-1}, respectively, while rmsd is 8.9, 1.3 and 1.1 cm^{-1}, respectively.

For vitamin C the hydrogen bonding brings in many-body quantum effects (e.g., the three-body charge transfer, exchange-repulsion and their coupling for two hydrogen bonds involving three fragments), which are not accounted for in FMO2. These effects tend to increase the errors, which are especially large for vibrational modes delocalized over several water molecules, coupled with OH in vitamin C. Likewise, the maximum errors in α-tocopherol and TEMPO are found in the delocalized vibrational modes.

TABLE I. Frequencies (cm^{-1}) and Raman activities ($Å^4$/u) for prominent peaks of the solute in vitamin C, α-tocopherol, and TEMPO, computed with FMO-HF and full HF using the 6-31G(d) basis set.

system	vibration		Raman activity	
	FMO-HF	HF	FMO-HF	HF
vitamin C[a]	3884.5	3886.6	150.4	144.9
vitamin E[a]	3203.8	3204.0	403.2	369.9
TEMPO[b]	3237.2	3237.4	130.6	124.0

[a] singlet with RHF, [b] doublet with UHF.

A summary of the intense Raman peaks is given in TABLE I. The errors in frequencies are 2.1, 0.2 and 0.2 cm^{-1} in vitamin C, vitamin E, and TEMPO, respectively, while the errors in the Raman activities are 5.5, 33.3 and 6.6 $Å^4$/u, respectively.

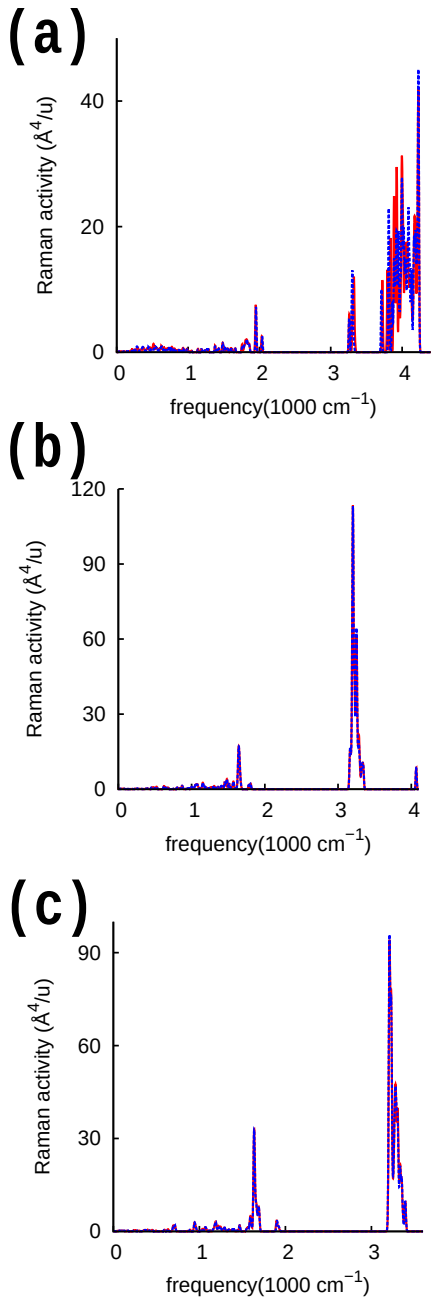


FIG. 2. Raman spectra calculated with FMO-HF and full HF, shown with red solid and blue dashed lines, respectively. (a) vitamin C solvated in water, (b) vitamin E, and (c) TEMPO solvated in DMF. See SI for detailed numeric results.

2. *Analysis of the Raman spectrum of a polystyrene oligomer*

Raman and IR spectra of a polystyrene oligomer, divided into two units per fragment, are shown in FIG. 3. In order to analyze these spectra, we localized the vibrational modes using the unitary transformation method[76] for the polystyrene oligomer divided into the head (reaction center), body (middle part) or tail (capped with Phe-COO) domains. The Raman active vibrational peaks can be divided into the low (450-1600 cm^{-1}) and high frequency (2864-3044 cm^{-1}) regions. The localization analysis clearly shows that these two regions are delocalized and localized, respectively.

The vibrational modes in the low frequency region (450-1000 cm^{-1}) describe C-C backbone stretching, and C-H bending of the backbone CH_2 groups. They are delocalized over the entire system and it is not meaningful to decompose these modes into the three domains.

On the other hand, the high frequency region (2864-3044 cm^{-1}), corresponding to the C-H stretching in phenyl groups or the backbone, can be analyzed. The results of the analysis are for a selection of normal modes representing several important peaks are shown in FIG. 3-(c,d). The vibrational modes for the C-H backbone stretching are very clearly divided between the terminal domains (head and tail) and the body domain. The vibrational modes localized in the terminal domains have the vibrational frequencies in the middle range (from 2872 to 2913 cm^{-1}), whereas the body domain has the frequencies in the ranges of 2870-2872 cm^{-1} and 2913-2918 cm^{-1}. Comparing the terminal domains, more active modes are localized in the head.

For the C-H stretching in phenyl groups one can also observe a distinctly different localization on either the body or terminal domains, with the exception of some modes with comparable localization factors. Also, in general many active vibrational modes of phenyl C-H vibrations are localized on the body domain with small or negligible contributions from the terminal domains. The coupling of vibrations of the body and terminal domains observed for some modes may be important for analyzing measured polymer spectra.

We note that the localization analysis in general can only provide a qualitative assignment. Most vibrational modes do have contributions from all three domains; however, for the C-H stretching it is often possible to identify the main contributor. Summarizing, the main

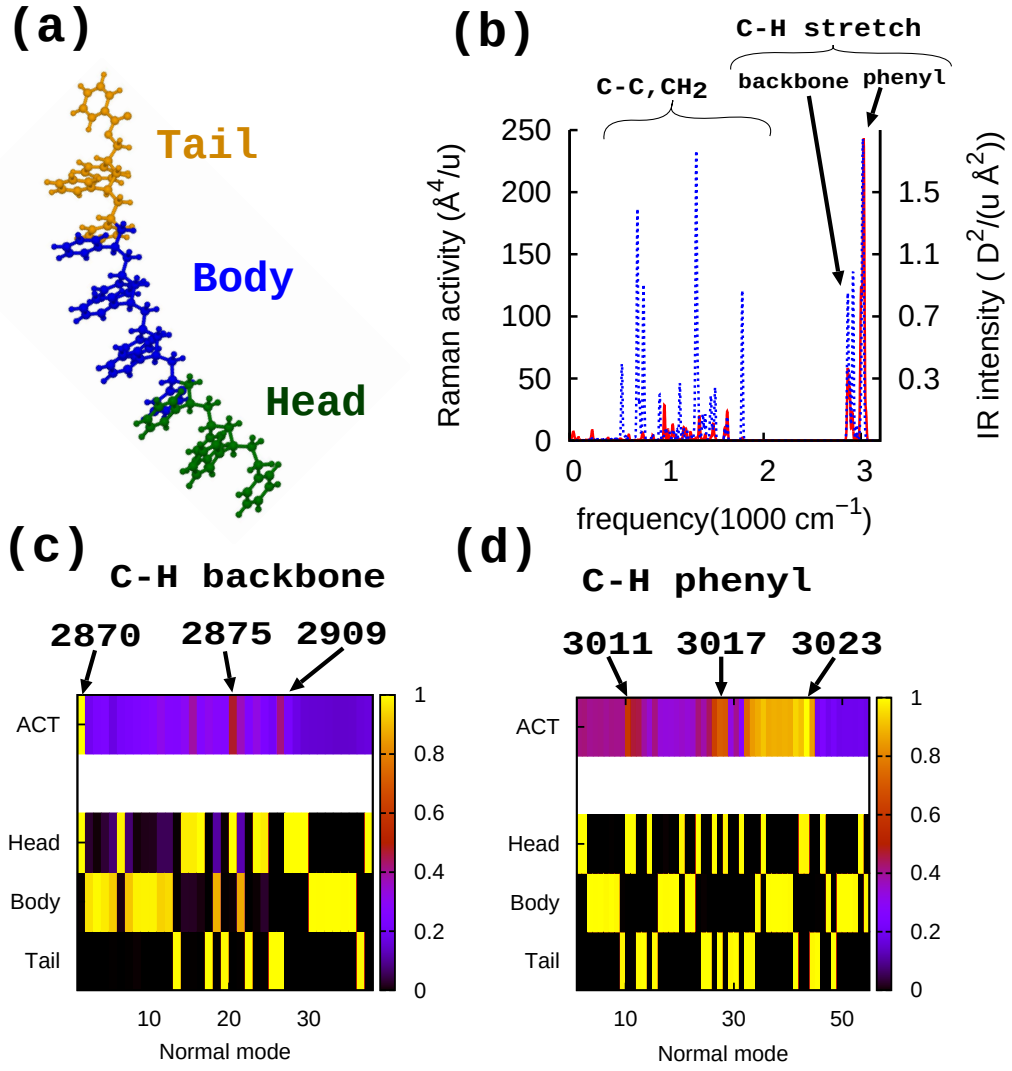


FIG. 3. (a) Fragmentation of the polystyrene oligomer. (b) Normal Raman spectrum (red solid line), and IR spectrum (blue dashed line), calculated with FMO-RHF/6-31G(d). Analysis of some selected localized vibrational modes for C-H stretching in (c) the backbone and (d) phenyl groups. Normal modes are numbered in the order of increasing frequencies, some of which in cm^{-1} are shown above the activity bar. The color bar on the top (act) represents relative Raman activity (the fraction of 1.0, shown as yellow, is the highest peak; other peaks are shown as fractions of the highest peak). The color bar on the right (with numbers from 0 to 1) shows the color scheme used both for activities (act) and fractions of the domains to the given vibrational mode. For instance the smallest shown frequency in (c) is colored yellow for "Head" and black for the rest, which means that the head is the main component for that vibration, whereas the activity on the top is yellow, which indicates that this is the highest peak (the largest activity). Frequencies are scaled[91] by 0.8953 for comparison with experiment.

benefit of the localization analysis is the ability to label individual peaks by the geometric domain (head, body or tail) for localized modes.

3. IR and Raman spectra of crambin

Crambin is composed of two α-helices and one β-turn; this protein is relatively rigid due to three disulfide bonds and it was studied with various theoretical methods.[82–88] Some information about the secondary structure of proteins can be obtained from the amide I band in Raman spectra.[89] We note that because of peak broadening, a single peak corresponds to a number of vibrational modes in the nearest vicinity of the main frequency, which has the highest intensity.

Using FMO/PCM, we optimized the structure of crambin solvated in water (all atoms). The optimized geometry superimposed on the X-ray (PDB: 1CRN) crystal structure in FIG. 4 shows their reasonable agreement, with the rmsd from experiment of 0.5249 Å.

For the optimized structure we calculated both IR and Raman spectra of crambin, shown in FIG. 5-(a). In the IR spectrum, both Amide I and Amide II (N-H bending coupled with C-N stretch) bands have high intensities. In the Raman spectrum, the CH_3 stretching mode has prominent peaks in addition to the less intense Amide I band, while the Raman activity of the Amide II band is very small. A comparison of the calculated and experimental Raman peaks is given in TABLE II while IR peaks are summarized in TABLE III. The CH_3 stretching peaks from FMO calculations are 1412, and 1474 cm^{-1}, whereas the Amide I peaks are 1645, and 1687 cm^{-1}. The experimentally obtained[90] peaks are 1410, 1480, 1610, and 1660 cm^{-1}, in good agreement with our calculations.

TABLE II. Prominent Raman peaks (cm^{-1}) in crambin calculated with FMO-RHF/6-31(d) and measured in experiment.[90] The calculated frequencies are scaled[91] by 0.8953.

	CH_3		Amide I	
FMO	1412	1475	1639	1690
experiment	1410	1480	1610	1660

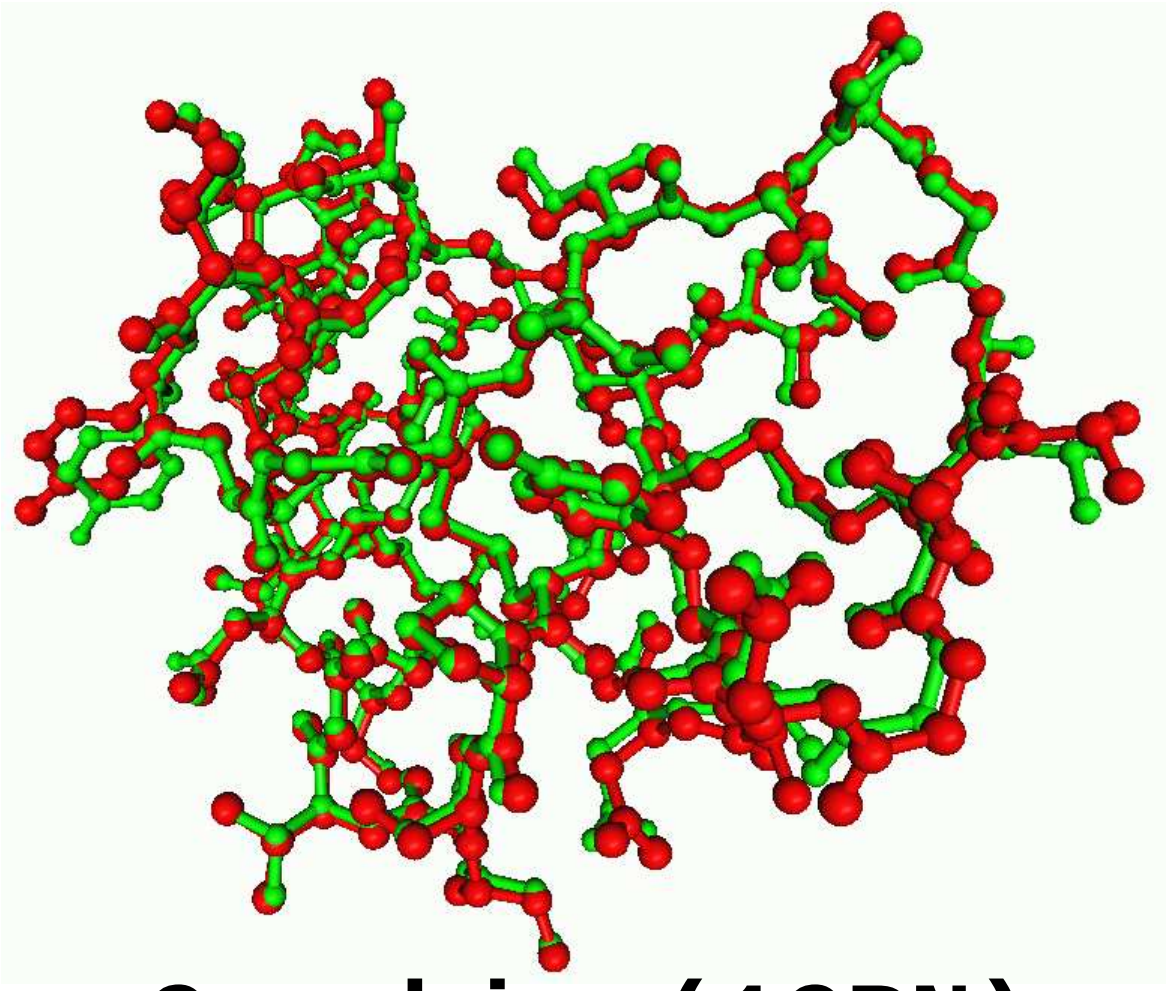


FIG. 4. Superimposed crambin structures obtained with FMO-RHF/D/PCM/6-31G(d) (red) on the X-ray experiment[67] (green). The rmsd between them is 0.5249 Å.

To assign domains in FIG. 5-(b) to observed Raman peaks, we applied the localization analysis. The results for selected CH_3 and Amide I bands are shown in FIG. 5-(c) and (d), respectively. The CH_3 stretching peak at 1412 cm^{-1} (not shown) is attributed to the α-Helices A and B, and the β-sheet, whereas the most intense CH_3 stretching peak at 1474 cm^{-1} is attributed mainly to the α-Helix A with a small contribution from the α-Helix B. For the Amide I band, the 1645 cm^{-1} peak is delocalized over the whole protein, whereas

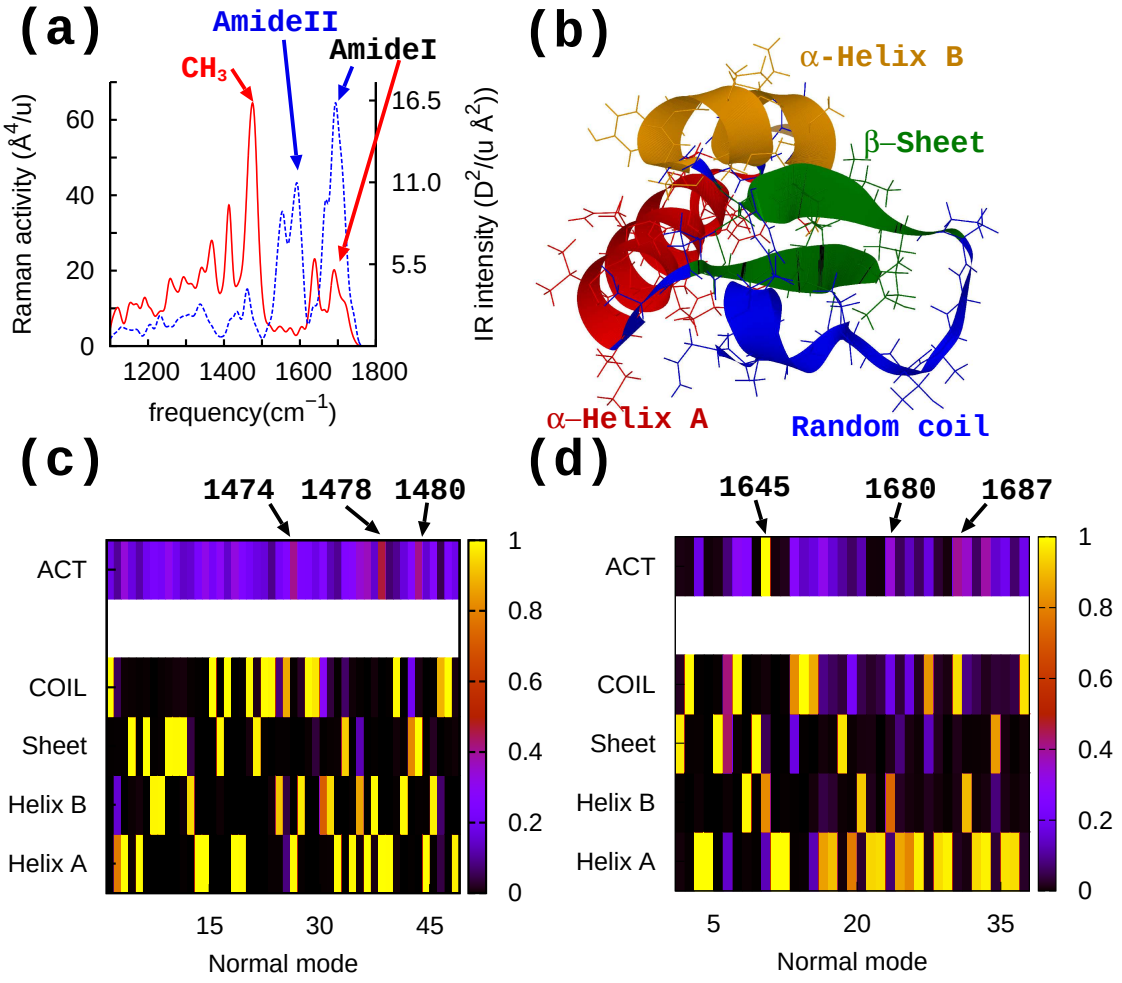


FIG. 5. (a) Calculated Raman (red solid line) and IR (blue dashed line) spectra of crambin. The frequency is scaled by 0.8953 for better comparison to experiment. (b) Schematic illustration of crambin showing the domains. Domain contributions to Raman activities for (c) Amide I and (d) Amide II bands. The color bar on the top (act) represents relative Raman activity (the fraction of 1.0, shown as yellow, is the highest peak; other peaks are shown as fractions of the highest peak). Normal modes are numbered in the order of increasing frequencies, some of which in cm^{-1} are shown above the activity bar. The color bar on the right (with numbers from 0 to 1) shows the color scheme used both for activities (act) and fractions of the domains to the given vibrational mode. For instance the smallest shown frequency in (c) is colored yellow for "Coil" and black for the rest, which means that the coil is the main component for that vibration, whereas the activity on the top is light blue, which indicates a relatively small value. Helix, Sheet and COIL denote the α-Helix, β-sheet and Random coil, respectively.

the 1687 cm^{-1} peak is mainly due to the α-Helix A.

TABLE III. Prominent IR peaks (cm^{-1}) in crambin calculated with FMO-RHF/6-31(d) The frequencies are scaled[91] by 0.8953.

Type	Frequency	Intensity
Amide II-1	1557.9	20.9
Amide II-2	1594.7	36.0
Amide I-1	1659.5	22.5
Amide I-2	1682.5	38.2

4. Computational timing

The computational efficiency is shown for the box-shaped cluster of water molecules, with the linear box sizes of 5.0, 8.0, 13.5, 16.0, and 17.0 Å containing 24, 93, 249, 408 and 495 atoms, respectively. The timings were measured on 16 nodes equipped with 2.93 GHz Xeons (8 CPU cores and 12 GB memory per node). The computational timings for calculating Raman activities are shown in FIG. 6 (these numbers do not include the Hessian timings).

It can be seen that the scaling of FMO is considerably lower than that of full RHF. For the largest system containing 495 atoms, FMO-RHF and full RHF took 2.1 and 6.4 hours, respectively.

We note that the most expensive step for Raman simulations is not the calculation of the derivatives of the polarizability tensor α but the Hessian calculation to obtain the normal coordinates and vibrational frequencies.[50] For crambin the Hessian and Raman activity calculations took 53 and 26 hours, respectively (using 25 Xeon nodes, 228 CPU cores in total). The FMO methods significantly reduces both memory requirements and computational cost and enables large scale simulations of IR and Raman spectra.

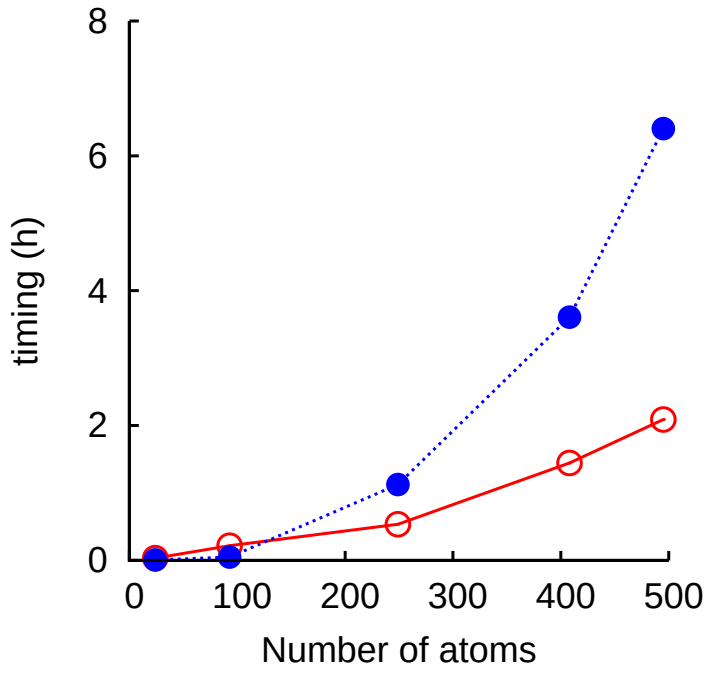


FIG. 6. Timings of Raman activity calculations for box-shaped clusters of water molecules, depicted for FMO-RHF and RHF with red solid and blue dotted lines, respectively (measured on 16 nodes with 2.93 GHz Xeons, equipped with 8 CPU cores and 12 GB RAM per node). 6-31G(d) is used.

E. Conclusions

We have developed a method to calculate Raman activities in the framework of the fragment molecular orbital, implemented in GAMESS and parallelized using GDDI. In addition, we have also developed analytic and numeric second derivatives for the Grimme's dispersion models.[62,63] We have shown that Raman activities calculated with FMO are in reasonable agreement with full HF results, for a number of representative closed-shell and radical systems.

As an application of the developed method, we have calculated IR and Raman spectra of a polystyrene oligomer and a small protein, crambin. We have also optimized the structure of crambin using FMO and reproduced the experimental X-ray structure with rmsd of 0.5249 Å. The experimental Raman spectrum of crambin is in good agreement with our calculated

results. By the application of the vibrational mode localization analysis we have assigned calculated peaks to particular geometric domains, which is possible for many but not all modes, because some of them are delocalized over the whole system. The delocalization effects in FMO2 are accounted for with pairwise corrections from fragment dimers, however some limitations to the accuracy are imposed by the treatment of delocalized modes in this way.

FMO calculations of Raman spectra are considerably faster than full HF calculations, and they also require much less memory. The Raman activity calculation of crambin (642 atoms) took 115 MB of memory per core and 26.5 hours using 25 Xeon nodes (228 CPU cores in total). The Hessian and IR calculation of crambin took 2204 MB of memory per core and 53 hours on the same cluster. FMO can be applied to simulate Raman and IR spectra of realistic polymers with defects, irregular and amorphic structures, and the observed peaks can be assigned based on their normal coordinates.

Supporting Information

Detailed derivations of the second derivatives for the dispersion model, Tables of Raman peaks computed with FMO-HF and HF for vitamins C and E, and TEMPO, as well as IR and Raman peaks computed with FMO for the polystyrene oligomer are provided in the Supporting Information. This material is available free of charge via the Internet at http : //pubs.acs.org/doi/suppl/10.1021/ct5003829/suppl_file/ct5003829_si_001.pdf

REFERENCES

[1]H. Kim and M. Cho, Chem. Rev. **113**, 5817 (2013).

[2]E. Vass, M. Hollósi, F. Besson, and R. Buchet, Chem. Rev. **103**, 1917 (2003).

[3]J. R. Ferraro and K. Nakamoto, editors, *Introductory Rraman Spectroscopy*, Academic Press, London, 1994.

[4]P. Deglmann, F. Furche, and R. Ahlrichs, Chem. Phys. Lett. **362**, 511 (2002).

[5]Y. Alexeev, M. W. Schmidt, T. L. Windus, and M. S. Gordon, J. Comput. Chem. **28**, 1685 (2007).

[6]P. Pulay, Mol. Phys. **17**, 197 (1969).

[7]A. Warshel and M. Karplus, J. Am. Soc. Chem. **94**, 5612 (1972).

[8]Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).

[9]S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Str.: THEOCHEM **461**, 1 (1999).

[10]H. A. Witek, S. Irle, and K. Morokuma, J. Chem. Phys. **121**, 5163 (2004).

[11]W. Zheng and B. R. Brooks, Biophys. J. **89**, 167 (2005).

[12]H. Li and J. H. Jensen, Theor. Chem. Acc. **107**, 211 (2002).

[13]A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, J. Chem. Phys. **126**, 224102 (2007).

[14]A. Komornicki and J. W. McIver, J. Chem. Phys. **70**, 2014 (1979).

[15]G. B. Bacskay, S. Saebø, and P. R. Taylor, Chem. Phys. **90**, 215 (1984).

[16]B. G. Johnson and J. Florian, Chem. Phys. Lett. **247**, 120 (1995).

[17]A. Stirling, J. Chem. Phys. **104**, 1254 (1996).

[18]K. Jalkanen, I. Degtyarenko, R. Nieminen, X. Cao, L. Nafie, F. Zhu, and L. Barron, Theor. Chem. Acc. **119**, 191 (2008).

[19]K. Ramnarayan, H. Bohr, and K. Jalkanen, Theor. Chem. Acc. **119**, 265 (2008).

[20]V. W. Jügensen and K. Jalkanen, Phys. Biol. **3**, S63 (2006).

[21]E. Nikitina, V. Sulimov, V. Zayets, and N. Zaitseva, Int. J. Quant. Chem. **97**, 747 (2004).

[22]J. J. P. Stewart, J. Mol. Model. **15**, 765 (2009).

[23]S. Goedecker and G. E. Scuseria, Comp. Sci. Eng. **5**, 14 (2003).

[24]P. G. Mezey and J. Leszczynski, editors, *Linear-Scaling Techniques in Computational Chemistry and Physics.*, Springer, New York, 2011.

[25]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[26]P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).

[27]J. L. Gao, J. Phys. Chem. B **101**, 657 (1997).

[28]M. Kobayashi and H. Nakai, J. Chem. Phys. **138**, 044102 (2013).

[29]X. He and K. M. Merz, J. Chem. Theory Comput. **6**, 405 (2010).

[30]H. Yu, H. R. Leverentz, P. Bai, J. I. Siepmann, and D. G. Truhlar, J. Phys. Chem. Lett. **5**, 660 (2014).

[31]Y. Tong, Y. Mei, J. Z. H. Zhang, L. L. Duan, and Q. G. Zhang, J. Theor. Comp. Chem. **8**, 1265 (2009).

[32]P. Söderhjelm, J. Kongsted, and U. Ryde, J. Chem. Theory Comput. **6**, 1726 (2010).

[33]J. Gao and Y. Wang, J. Chem. Phys. **136**, 071101 (2012).

[34]M. A. Collins, Phys. Chem. Chem. Phys. **14**, 7744 (2012).

[35]A. Frank, H. M. Möller, and T. E. Exner, J. Chem. Theory Comput. **8**, 1480 (2012).

[36]M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Ann. Rev. Phys. Chem. **64**, 553 (2013).

[37]K. Kiewisch, C. R. Jacob, and J. Visscher, J. Chem. Theory Comput. **9**, 2425 (2013).

[38]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101 (2008).

[39]W. Hua, T. Fang, W. Li, J.-G. Yu, and S. Li, J. Phys. Chem. A **112**, 10864 (2008).

[40]S. Sakai and S. Morita, J. Phys. Chem. A **109**, 8424 (2005).

[41]J. C. Howard and G. S. Tschumper, J. Chem. Phys. **139**, 184113 (2013).

[42]H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, J. Chem. Phys. **138**, 164103 (2013).

[43]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[44]D. G. Fedorov and K. Kitaura, editors, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC press, Boca Raton, FL, 2009.

[45]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A. **111**, 6904 (2007).

[46]D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[47]S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, and K. Fukuzawa, Phys. Chem. Chem. Phys. **16**, 10310 (2014).

[48]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).

[49]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **544**, 87 (2012).

[50]H. Nakata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, Chem. Phys. Lett. **603**, 67 (2014).

[51]Y. Mochizuki, T. Ishikawa, K. Tanaka, H. Tokiwa, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **418**, 418 (2006).

[52]M. P. Mazanetz, O. Ichihara, R. J. Law, and M. Whittaker, J. Cheminf. **3**, 2 (2011).

[53]T. Sawada, D. G. Fedorov, and K. Kitaura, J. Am. Chem. Soc. **132**, 16862 (2010).

[54]T. Watanabe, Y. Inadomi, K. Fukuzawa, T. Nakano, S. Tanaka, L. Nilsson, and U. Nagashima, J. Phys. Chem. B **111**, 9621 (2007).

[55]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).

[56]P. J. Carlson, S. Bose, D. W. Armstrong, T. Hawkins, M. S. Gordon, and J. W. Petrich, J. Phys. Chem. B **116**, 503 (2012).

[57]P. V. Avramov, D. G. Fedorov, P. B. Sorokin, S. Sakai, S. Entani, M. Ohtomo, and H. N. Y. Matsumoto, J. Phys. Chem. Lett. **3**, 2003 (2012).

[58]Y. Okiyama, T. Tsukamoto, C. Watanabe, K. Fukuzawa, S. Tanaka, and Y. Mochizuki, Chem. Phys. Lett. **566**, 25 (2013).

[59]L. Roskop, D. G. Fedorov, and M. S. Gordon, Mol. Phys. **111**, 1622 (2013).

[60]H. Nakata, D. G. Fedorov, T. Nagata, S. Yokojima, K. Ogata, K. Kitaura, and S. Nakamura, J. Chem. Phys. **137**, 044110 (2012).

[61]H. Nakata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, Theor. Chem. Acc. **1477**, 133 (2014).

[62]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[63]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[64]Y. Yamaguchi, H. F. Schaefer III, Y. Osamura, and J. Goddard, editors, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press, New York, 1994.

[65]Y. Wang and Y. Tominaga, J. Phys. Soc. Jap. **62**, 4198 (1993).

[66]T. Cai, H. Gu, F. Liu, and M. Xie, Appl. Spectrosc. **66**, 114 (2012).

[67]M. M. Teeter, Proc. Nat. Acad. Sc. U.S.A. **81**, 6014 (1984).

[68]D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[69]T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, and S. Tanaka, Chem. Phys. Lett. **523**, 128 (2012).

[70]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[71]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).

[72]N. W. Schmidt, K. K. Baldridge, J. A. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[73]M. S. Gordon and M. W. Schmidt, Advances in electronic structure theory: Gamess a decade later, in *Theory and Applications of Computational Chemistry, the first forty years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria, pages 1167–1189, Elsevier, Amsterdam, 2005.

[74]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004).

[75]C. Steinmann, M. W. Ibsen, A. S. Hansen, and J. H. Jensen, PLOS ONE **7**, e44480 (2012).

[76]C. R. Jacob, S. Luber, and M. Reiher, J. Phys. Chem. B **113**, 6558 (2009).

[77]T. Weymuth, M. P. Haag, K. Kiewisch, S. Luber, S. Schenk, C. R. Jacob, C. Herrmann, J. Neugebauer, and M. Reiher, J. Comput. Chem. **33**, 2186 (2012).

[78]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[79]H. Li, D. G. Fedorov, T. Nagata, K. Kitaura, J. H. Jensen, and M. S. Gordon, J. Comput. Chem. **31**, 778 (2010).

[80]T. Nagata, D. G. Fedorov, H. Li, and K. Kitaura, J. Chem. Phys. **136**, 204112 (2012).

[81]D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).

[82]C. Van Alsenoy, C.-H. Yu, A. Peeters, J. M. L. Martin, and L. Schäfer, J. Phys. Chem. A **102**, 2246 (1998).

[83]J. J. P. Stewart, J. Mol. Str.: THEOCHEM **401**, 195 (1997).

[84]T. E. Exner and P. G. Mezey, Phys. Chem. Chem. Phys. **7**, 4061 (2005).

[85]T. Ogawa, N. Kurita, H. Sekino, O. Kitao, and S. Tanaka, Chem. Phys. Lett. **397**, 382 (2004).

[86]A. van der Vaart, D. Suarez, and K. M. Merz, J. Chem. Phys. **113**, 10512 (2000).

[87]J. P. Ritchie, Chem. Phys. Lett. **387**, 243 (2004).

[88]J. Kästner, S. Thiel, H. M. Senn, P. Sherwood, and W. Thiel, J. Chem. Theory Comput. **3**, 1064 (2007).

[89]R. Tuma, J. Raman Spectr. **36**, 307 (2005).

[90]R. W. Williams and M. M. Teeter, Biochemistry **23**, 6796 (1984).

[91]A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).

Chapter VIII:

Conclusion:
The future beyond this thesis

$$E = \sum_{I}^{N} E'^{I} + \sum_{\substack{I>J \\ IJ \in RHF}}^{N} (E'_{IJ} - E'_{I} - E'_{J}) + \sum_{\substack{I>J \\ IJ \in RHF}}^{N} Tr(\Delta D^{IJ} V^{IJ})$$

I,J : RHF Fragment

$$+ \sum_{K}^{N} E'^{K} + \sum_{K>L}^{N} (E'_{KL} - E'_{K} - E'_{L}) + \sum_{K>L}^{N} Tr(\Delta D^{\alpha,KL} V^{KL} + \Delta D^{\beta,KL} V^{KL})$$

K,L: UHF Fragment

Definition

Density matrix difference	$\Delta D^{IJ} = D^{IJ} - (D^{I} \oplus D^{J})$
Internal fragment energy	**E'_I, E'_J, E'_K, E'_L etc.**
ESP potential	**V_I, V_J, V_{IJ}, V_{KL} etc.**

VIII. CONCLUSION: THE FUTURE BEYOND THIS THESIS

Recently, the applications in quantum chemistry is becoming more and more larger systems due to the rapid increase of computer resources, and the realistic computational modeling become possible. The fundamental theory and practical methodology for various kind of property evaluations have been developed over the past decades. However, the methodology is limited due to its large amount of computational timing, and excess memory requirement, thus any biological vibration analysis in entire protein is difficult. New theoretical idea and technique are needed to open the application territory of quantum chemistry for large biological systems.

The force behind this study is the inevitable requirement of the simple, easy to handle, free, and efficient program package that can be used for gaining the interpretation of complex phenomena in large biological systems. As a consequence of these desires, in this thesis, the analytic derivative techniques of quantum chemical calculation are extended to fragment molecular orbital (FMO) method, and largely six new FMO method have been developed;
(1) The Unrestricted Hartree-Fock (UHF) calculation based on the FMO (FMO-UHF)
(2) The FMO-UHF is accelerated by the point charge electrostatic potential (ESP-PC) approximation.
(3) DFT is interfaced with FMO for open-shell calculation.
(4) The analytic energy second order derivative of energy (Hessian) is developed.
(5) The Hessian is accelerated to calculate more than hundreds atoms in the system
(6) The FMO Raman calculation predict the full protein Raman spectrum accurately.
The approach toward the biological vibration spectroscopy such as Raman scattering for the entire protein systems is the new challenge in this study, and the extension of FMO method to open-shell system enable us to investigate various spectra for oxidation reactions, protein metal enzyme, and many other material sciences.

By these new theory and method development in this thesis, quite a large biological reaction analyses and various biological spectra become possible, opening the new application territory for the computational theoretical chemistry. The all quantum chemistry program used in this study have been implemented in GAMESS quantum chemical package, and it

is available free of charge via the internet (http://www.msg.ameslab.gov/gamess), and we expect that the developed method will be widely used in the field of biological computational chemistry.

The remaining problems in quantum chemistry method are the difficulty to apply the method to calculate the solid state physics, where the systems possess the strongly polarized covalent bonds across the fragments. In this case, the hybrid projected approach dose not work, and new theoretical idea and algorithm are necessary for the separation of these polarized covalent bonds without losing accuracy of energy, gradient, and Hessian. The achievement of the new scheme for the fragment separation across the covalent bonds would be also open the new application territory of quantum chemistry, and our effort to develop the quantum chemical methodology would be continued.

ACKNOWLEDGMENT

As for the basis of this reserch, the Fragment molecular orbital method is developed, and coded by Dr. Fedorov G. Dmitri and Prof. Dr. Kitaura Kazuo, and the pionner work for fully analytic energy gradient have been reported by Dr. Nagata. I sincerely appreciate to Dr. Fedorov, Prof. Dr. Kitaura, and Dr. Nagata for the significant help of computational programing, manuscript modifications, and sharing the informations, which are essential for this thesis.

I sincerely thank for Prof. Nakamura Shinichiro, Prof. Yokojima Satoshi, Prof. Sakurai minoru, Prof. Gordon S. Mark, Dr. Schmidt W. Michael, Dr. Zahariev Federico, Dr. Ogata Koji, and Dr. Brorsen R. Kurt for helping this project and the publications for this study. I also thanks Dr. Gao Qi for kindly suggesting and discussing this research project.

This study have been mostly performed in RIKEN, Research Cluster for Innovation, Nakamura Lab, and Center for Biological Resources and Informatics, Tokyo Institute of Technology, Sakurai Lab, I thank for the kindly help of all the members. This work in part have been done in Ames Laboratory, US-DOE and Department of Chemistry, Iowa State University, Ames, Gordon Lab, and Department of Chemistry, Purdue University, Slipchenko lab., and I thank for the kindly help of the members.

This work was partly supported by JSPS KAKENHI Grant Number 262235. The main part of numerical calculations was carried out on TSUBAME2.0 at Global Scientific Information and Computing Center of Tokyo Institute of Technology, and RIKEN Integrated Cluster of Clusters (RICC) at RIKEN for the computer resources.

PUBLICATION LIST

1 **Chapter II**

Hiroya Nakata, Dmitri G. Fedorov, Takeshi Nagata, Satoshi Yokojima, Koji Ogata, Kazuo Kitaura, and Shinichiro Nakamura, Unrestricted Hartree-Fock Based on the Fragment Molecular Orbital Method: Energy and Its Analytic Gradient, *J. Chem. Phys.* **2012**, *137*, 044110.

2 **Chapter III**

Hiroya Nakata, Dmitri G. Fedorov, Satoshi Yokojima, Kazuo Kitaura, and Shinichiro Nakamura, Derivatives of the approximated electrostatic potentials in unrestricted HartreeFock based on the fragment molecular orbital method and an application to polymer radicals, *Theor. Chem. Acc.* **2014**, *133*, 1477.

3 **Chapter IV**

Hiroya Nakata, Dmitri G. Fedorov, Satoshi Yokojima, Minoru Sakurai, Kazuo Kitaura, Shinichiro Nakamura, Unrestricted density functional theory based on the fragment molecular orbital method for the ground and excited state calculations of large systems. *J. Chem. Phys.* **2014**, *140*, 144101.

4 **Chapter V**

Hiroya Nakata, Takeshi Nagata, Dmitri G. Fedorov, Satoshi Yokojima, Kazuo Kitaura, Shinichiro Nakamura, Analytic second derivatives of the energy in the fragment molecular orbital method, *J. Chem. Phys.* **2013**, *138*, 164103.

5 **Chapter VI**

Hiroya Nakata, Dmitri G. Fedorov, Satoshi Yokojima, Kazuo Kitaura, Shinichiro Nakamura, Efficient vibrational analysis for unrestricted HartreeFock based on the fragment molecular orbital method, *Chem. Phys. Lett.* **2014**, *603*, 67-74.

6 **Chapter VII**

Hiroya Nakata, Dmitri G. Fedorov, Satoshi Yokojima, Kazuo Kitaura, Shinichiro Nakamura, Simulations of Raman Spectra Using the Fragment Molecular Orbital Method,

J. Chem. Theor. Comput **2014**, *10*, 3689-3698.

- Other related publications

7 Hiroya Nakata, Michael W. Schmidt, Dmitri G. Fedorov, Kazuo Kitaura, Shinichiro Nakamura, and Mark S. Gordon, Efficient Molecular Dynamics Simulations of Multiple Radical Center Systems Based on the Fragment Molecular Orbital Method, *J. Phys. Chem. A* **2014**, *118*, 9762-9771.

8 David Simoncini, Hiroya Nakata, Koji Ogata, Shinichiro Nakamura, and Kam YJ Zhang, Quality assessment of predicted protein models using energies calculated by the fragment molecular orbital method, *Molecular Informatics, impress.*

9 Kurt R. Brorsen, Federico Zahariev, Hiroya Nakata, Dmitri G. Fedorov, and Mark S. Gordon, Analytic Gradient for Density Functional Theory Based on the Fragment Molecular Orbital Method, *J. Chem. Theor. Comput, impress.*