12-2011

QUANTUM MECHANICAL AND MOLECULAR MECHANICAL STUDY OF SOLVENT EFFECTS

Dejun Si
University of Nebraska-Lincoln, djsi@unlserve.unl.edu

QUANTUM MECHANICAL AND MOLECULAR MECHANICAL STUDY OF SOLVENT EFFECTS

by

Dejun Si

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Hui Li

Lincoln, Nebraska

December, 2011

QUANTUM MECHANICAL AND MOLECULAR MECHANICAL STUDY OF SOLVENT EFFECTS

Dejun Si, Ph.D.

University of Nebraska, 2011

Adviser: Hui Li

Intermolecular interaction and solvent effects play important roles in determining physical and chemical properties of molecular systems, and must be considered in relevant quantum mechanical (QM) calculations. Due to the high computational cost, full and rigorous QM treatment of both solute and solvent molecules is impractical. Computationally efficient molecular mechanical (MM) methods can be used to describe solvent effects, and combined into QM methods to formulate QM/MM methods. Classical force field method and reaction-field method are the two most popular MM methods. However, the issue of effectively combining MM methods with high-level QM methods remains unsolved.

This thesis presents several novel QM/MM methods. The first is a heterogeneous reaction-field method that can be used to study solute molecules at the interface between two or more phases characterized by different dielectric constants. The second is a second-order perturbation theory/reaction-field method that can be used to obtain accurate QM results in the presence of a reaction-field for both close-shell and open-shell molecules. The third is a time-dependent density functional theory/polarizable force field method that can be used to study solvent effects in electronic transition and excited state molecules.

Acknowledgements

It is an honor for me to thank all the professors, friends, and family members who supported me during the completion of this Ph.D dissertation.

Foremost, I would like to express my deepest appreciation to my advisor Prof. Hui Li for his support and brilliant guidance during my Ph.D study at the University of Nebraska-Lincoln, for his patience, inspiration, encouragement and perceptual knowledge. Research life would not have been so rewarding without his valuable advice.

I would like to show my gratitude to my supervisory committee members: Prof. Joseph J. Barycki, Prof. Mark A. Griep, Prof. Gerard S. Harbison and Prof. Xiao Cheng Zeng for their constructive academic comments and insightful suggestions.

I am grateful to everyone in Prof. Li's group for the interesting discussions on the research and generous help in the daily life. It was very enjoyable working with them and their friendship has been a treasure for me.

Finally I would like to thank my parents and my husband Yongcheng Wang for their unconditional and everlasting love, understanding, and support.

Table of Contents

Chapter 1 Introduction

1.1 General overview

Quantum mechanics (QM) has revolutionized our way of studying chemistry. It provides the laws of motion for both microscopic particles and macroscopic objects. The postulates[1] of QM theory assert that a wave function can fully specify the state of a system and that it evolves in time according to the time dependent Schrödinger equation.[2] For systems in a stationary state, the time independent Schrödinger equation gives a description of the energy of the system and other properties. The Dirac equation[3] further generalized the Schrödinger equation by including relativistic effects. In 1929, Dirac[4] wrote:

> "*The fundamental laws necessary for the mathematical treatment of large parts of physics and the whole of chemistry are thus fully known, and the difficulty lies only in the fact that application of these laws lead to equations that are too complex to be solved*."

In chemistry, the Schrödinger equation is applied in order to describe the motions of electrons and nuclei at low velocities. Since the multi-electron Schrödinger equation cannot be solved analytically, approximate numerical methods have been developed. In 1927, the Heitler-London valence bond[5] treatment for a hydrogen molecule was published. Later, the Hartree-Fock theory[6] utilizing molecular orbitals was established. Further developments include the configuration interaction method (CI)[7], the Møller-Plesset perturbation treatment,[8] the Roothaan-Hall equations,[9,10] and the coupled cluster (CC) approach.[11]

Usually, chemical systems involve a large number of atoms and molecules. A full and rigorous treatment of such a large system using QM methods is impractical. Alternatively, computationally efficient molecular mechanical (MM) methods are often employed to model large molecular systems. In addition, MM methods can be combined with QM methods to formulate QM/MM methods, in which part of the molecular system is described by QM methods, while the remainder is treated by MM methods. Combined QM/MM methods overcome the weaknesses of each method and extend their application to a broader range of chemical problems.

1.2 Dissertation outline

This thesis contains both the application and extension of quantum chemical methods. After this introduction of the fundamental theories for quantum chemistry in the current chapter, Chapters 2, 4 and 5 describe the derivation and implementation of analytical gradients in several QM/MM methods. Chapter 3 describes an application of the method described in Chapter 2.

In Chapter 2, the establishment of a heterogeneous conductor-like solvation model is described. Polarizable continuum models (PCM) have been widely used to study molecules in solutions. Although dielectric polarizable continuum model (DPCM)[12] and integral equation formalism PCM (IEF-PCM)[13,14] consider the anisotropic polarization effects, the analytic nuclear gradients are unavailable. The conductor-like polarizable continuum model (CPCM) is widely used because of its simplicity. In CPCM, the solvent is treated as a homogeneous and isotropic polarizable medium; however, there are cases in which the heterogeneity or anisotropy of the environment surrounding a molecule must be considered. The heterogeneous CPCM model is developed with analytic nuclear

gradients. The main features of this model are: (1) using different local effective dielectrics for different portions of a solute cavity surface following the heterogeneous IEF-PCM developed by Barone *et al.*;[15] (2) assuming that the local effective dielectrics are constants in a geometry optimization process; (3) variationally formulating the solvation operators for Hartree-Fock and Kohn-Sham methods; (4) using the FIXPVA[16] tessellation scheme to obtain continuous and smooth potential energy surfaces.

In Chapter 3, the heterogeneous CPCM method is applied to calculate the Cu^{2+}/Cu^{+} reduction potentials (E^0) of type-1 proteins. In principle, the continuum models cannot describe the structural details of a protein matrix. Previous study shows that if the active site model is relatively large (e.g. ~100 atoms), the protein matrix solvation free energy difference over a series of similar active sites can often be reproduced by a continuum solvation model.[17] Chapter 3 explores the possibility of using continuum models to describe protein matrix and aqueous solvation in reduction potential calculations. Quantum chemical E^0 calculations are usually performed using active site model molecules extracted from X-ray crystal structures. Part of the model molecule is exposed to the bulk solvent and the other part is buried in the protein matrix. Due to the heterogeneous nature of the environment around an active site, using different local effective dielectrics for different portions of the solvent in heterogeneous CPCM method can improve the accuracy of E^0 predictions. The ligand interactions and solvent effects that alter the E^0 are discussed.

Chapter 4 describes the implementation of the analytic nuclear gradients in combined Møller-Plesset perturbation theory and the conductor-like polarizable continuum model (MP2/CPCM). The MP2 methods are accurate quantum mechanical

methods for recovering electron correlation energy, thus, combined MP2/CPCM is a good choice for studying the structures and intermolecular interactions in solutions. Closed shell MP2/CPCM has been derived and implemented by Cammi, Mennucci and Tomasi.[18] PCM methods have not been combined with open shell MP2 methods. The MP2/CPCM methods described in Chapter 4 include both closed shell and open shell cases, including spin-restricted closed shell (RMP2), Z-averaged spin-restricted open shell (ZAPT2), and spin-unrestricted open shell (UMP2) cases. These methods facilitate the computational study on the open shell systems such as T_1 states of acetone, nucleobases and nucleobase pairs.

Chapter 5 presents an establishment of the gradient method for combined time-dependent density functional theory and polarizable force field (TDDFT/MM). Time-dependent density functional theory (TDDFT) method formulated by Casida *et al.*[19] is an efficient and relatively accurate QM method for studying valence and singly excited molecules. Using a polarizable force field in QM/MM calculations is advantageous because the electronic polarization of the MM region can be described, especially when electronic excitation is involved in the QM region. The analytic gradient is implemented for the combined TDDFT/MM method. Chapter 5 illustrates the geometry optimization of a cluster formed by acetone and water using TDDFT/MM method. The frequency and infrared absorption intensity are calculated for carbonyl vibration mode.

The studies in Chapters 2-5 have been published in peer reviewed academic journals.[20,21]

1.3 Theoretical background

1.3.1 Fundamental quantum mechanical theory

Quantum chemistry is intended to describe the electron motion in the field of fixed nuclear point charges. The Schrödinger equation is the foundation of quantum chemistry. It includes differentiation with respect to time t and position $\boldsymbol{r}$ and can be written as:

$$\boldsymbol{H}(\boldsymbol{r},t)\Psi(\boldsymbol{r},t) = i\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} \tag{1-1}$$

where $\boldsymbol{H}$ is the Hamiltonian operator given as a sum of kinetic and potential energy operators for nuclei and the electrons:

$$\boldsymbol{H}(\boldsymbol{r},t) = \boldsymbol{T}(\boldsymbol{r},t) + \boldsymbol{V}(\boldsymbol{r},t) \tag{1-2}$$

The goal is to obtain the approximation solutions $\Psi(\boldsymbol{r},t)$, known as the wave function of the system. If the potential energy operator is independent of time ($\boldsymbol{V}(\boldsymbol{r},t) = \boldsymbol{V}(\boldsymbol{r})$), the time and space variables of the wave function can be separated. The Hamiltonian operator $\boldsymbol{H}$ acting on the wave function yields the energies $E(\boldsymbol{r})$.

$$\boldsymbol{H}(\boldsymbol{r})\Psi(\boldsymbol{r},t) = E(\boldsymbol{r})\Psi(\boldsymbol{r},t) \tag{1-3}$$

Inserting Eq. (1-3) into Eq. (1-2) gives

$$\begin{aligned} &\boldsymbol{H}(\boldsymbol{r})\Psi(\boldsymbol{r},t) = E(\boldsymbol{r})\Psi(\boldsymbol{r},t) = i\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} \\ &\Psi(\boldsymbol{r},t) = \Psi(\boldsymbol{r})e^{-iEt} \end{aligned} \tag{1-4}$$

The Hamiltonian operator represents the total energy. For a system with N_{elec} electrons and N_{nuclei} nuclei, it can be written explicitly as follows

$$\boldsymbol{H} = -\sum_{i=1}^{N_{elec}} \frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N_{nuclei}} \frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N_{elec}}\sum_{A=1}^{N_{nuclei}} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_{elec}}\sum_{j>i}^{N_{elec}} \frac{1}{r_{ij}} + \sum_{A=1}^{N_{nuclei}}\sum_{B>A}^{N_{nuclei}} \frac{Z_A Z_B}{R_{AB}} \quad (1\text{-}5)$$

$$\nabla_i^2 = \frac{\partial}{\partial x_i^2} + \frac{\partial}{\partial y_i^2} + \frac{\partial}{\partial z_i^2}$$

where the Laplacian operator ∇_i^2 and ∇_A^2 contain differentiation with respect to the coordinates of the i^{th} electron and the A^{th} nucleus. M_A is the mass ratio of the nucleus A to an electron i, Z_A is the atomic number of nucleus A, r_{iA} is the distance between electron i and nucleus A, r_{ij} is the distance between electrons i and j, and R_{AB} is the distance between nuclei A and B. The first and second terms in the above equation are the kinetic energy of the electrons and nuclei, respectively; the third term is the Coulombic attraction between electrons and nuclei; the fourth and fifth terms are the repulsion between electrons and between nuclei, respectively.[22-24]

The Born-Oppenheimer (BO) approximation[25] allows for the separation of the electronic and nuclear variables in the Hamiltonian. The electronic Hamiltonian can be written as:

$$\boldsymbol{H}_e = -\frac{1}{2}\sum_{i=1}^{N_{elec}} \nabla_i^2 - \sum_{i=1}^{N_{elec}}\sum_{A=1}^{N_{nuclei}} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_{elec}}\sum_{j>i}^{N_{elec}} \frac{1}{r_{ij}} \quad (1\text{-}6)$$

Now the Schrödinger equation contains the electronic Hamiltonian and electronic wave function only:

$$\boldsymbol{H}_e \Psi_e = E_e \Psi_e \quad (1\text{-}7)$$

Once the electronic wave function has been solved for a particular configuration of nuclei, the nuclear kinetic energy can be recovered using Hessian vibrational analysis. The evaluation of nuclear-nuclear repulsion is trivial. E_e is the electronic energy, and

adding E_e to the constant nuclear repulsion energy yields the total energy for the system E_{tot}.

$$E_{tot} = E_e + \sum_{A=1}^{N_{nuclei}} \sum_{B>A}^{N_{nuclei}} \frac{Z_A Z_B}{R_{AB}} \tag{1-8}$$

In general, the wave function is a CI function in a complex functional space. In the independent particle model, a molecular orbital (MO) is used to represent the wave function of an electron. It is a product of a spatial orbital $\psi_i(\boldsymbol{r})$ and a spin function α or β.

$$\phi_i(\boldsymbol{x}_1) = \begin{cases} \psi_i(\boldsymbol{r})\alpha(\omega) \\ or \\ \psi_i(\boldsymbol{r})\beta(\omega) \end{cases} \tag{1-9}$$

The antisymmetry requirement of the wave function is enforced by building it from Slater Determinants (SD).[26] The general expression for N electrons and N spin orbitals is given as:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{Vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{Vmatrix} \tag{1-10}$$

The factor $1/\sqrt{N!}$ is a normalization factor. The rows of a Slater Determinant are labeled by electrons and the columns are labeled by spin orbitals.

In the Hartree-Fock method,[6] the trial wave function is constructed using a single SD. With the expression of the Hamiltonian and wave function in hand, the electronic energy of the system can be written as the expected value of the Hamiltonian. The variation principle states that the best wave function of this functional form is the one that gives the lowest possible energy. The energy expression is

$$E_0 = \left\langle \Phi_0 \middle| \boldsymbol{H}_e \middle| \Phi_0 \right\rangle \tag{1-11}$$

where $\boldsymbol{H}_e$ is the full electronic Hamiltonian. The operators can be collected as one-electron operator and two-electron operator according to the number of electron indices.

$$\begin{aligned} \boldsymbol{H}_e &= \sum_{i=1}^{N_{elec}} \boldsymbol{h}_i + \sum_{j>i}^{N_{elec}} \boldsymbol{g}_{ij} + \boldsymbol{V}_{ZZ} \\ \boldsymbol{h}_i &= -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N_{nuclei}} \frac{Z_A}{\left|\boldsymbol{R}_A - \boldsymbol{r}_i\right|} \\ \boldsymbol{g}_{ij} &= \frac{1}{\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|} \\ \boldsymbol{V}_{ZZ} &= \sum_{A=1}^{N_{nuclei}} \sum_{B>A}^{N_{nuclei}} \frac{Z_A Z_B}{\left|\boldsymbol{R}_A - \boldsymbol{R}_B\right|} \end{aligned} \tag{1-12}$$

When E_0 is expressed explicitly, most of the terms vanish and only a few terms survive in the expression[22]:

$$\begin{aligned} E_0 &= \sum_{i=1}^{N_{elec}} \left\langle \phi_i(1) \middle| \boldsymbol{h}_1 \middle| \phi_i(1) \right\rangle \\ &\quad + \sum_{i=1}^{N_{elec}} \sum_{j>i}^{N_{elec}} \left(\left\langle \phi_i(1)\phi_j(2) \middle| \boldsymbol{g}_{12} \middle| \phi_i(1)\phi_j(2) \right\rangle - \left\langle \phi_i(1)\phi_j(2) \middle| \boldsymbol{g}_{12} \middle| \phi_j(1)\phi_i(2) \right\rangle \right) + E_{ZZ} \\ &= \sum_{i=1}^{N_{elec}} h_i + \sum_{i=1}^{N_{elec}} \sum_{j>i}^{N_{elec}} \left(J_{ij} - K_{ij} \right) + E_{ZZ} \\ &= \sum_{i=1}^{N_{elec}} \left\langle \phi_i \middle| \boldsymbol{h}_i \middle| \phi_i \right\rangle + \frac{1}{2} \sum_{ij}^{N_{elec}} \left(\left\langle \phi_j \middle| \boldsymbol{J}_i \middle| \phi_j \right\rangle - \left\langle \phi_j \middle| \boldsymbol{K}_i \middle| \phi_j \right\rangle \right) + E_{ZZ} \end{aligned} \tag{1-13}$$

The Coulomb (**J**) and exchange (**K**) operators are defined as

$$\begin{aligned} \boldsymbol{J}_i \left| \phi_j(2) \right\rangle &= \left\langle \phi_i(1) \middle| \boldsymbol{g}_{12} \middle| \phi_i(1)\phi_j(2) \right\rangle \\ \boldsymbol{K}_i \left| \phi_j(2) \right\rangle &= \left\langle \phi_i(1) \middle| \boldsymbol{g}_{12} \middle| \phi_j(1)\phi_i(2) \right\rangle \end{aligned} \tag{1-14}$$

The requirement that MOs must remain orthonormal makes the energy minimization process a constraint optimization. It is solved using Lagrange multipliers. In the variation process, Fock operator $\mathbf{F}_i$ appears:

$$\delta E = \sum_{i}^{N_{elec}} \left(\langle \delta\phi_i \left| \boldsymbol{F}_i \right| \phi_i \rangle + \langle \phi_i \left| \boldsymbol{F}_i \right| \delta\phi_i \rangle \right)$$
$$\boldsymbol{F}_i = \boldsymbol{h}_i + \sum_{j}^{N_{elec}} \left(\boldsymbol{J}_j - \boldsymbol{K}_j \right) \tag{1-15}$$

These nonlinear equations are referred to as Hartree-Fock (HF) or self-consistent field (SCF) equations, and the N-particle problem is reduced to a set of one-particle eigenvalue problems.

$$\boldsymbol{F}_i \, \phi_i = \sum_{ij}^{N_{elec}} \lambda_{ij} \phi_j \tag{1-16}$$

Diagonalizing λ_{ij} will further simplify the equation, and the special set of molecular orbitals ϕ' is called canonical MOs.

$$\boldsymbol{F}_i \, \phi'_i = \varepsilon_i \, \phi'_i \tag{1-17}$$

Since the Fock operator counted the electron repulsion twice, HF theory will correct this before finalizing the total energy.

$$E = \sum_{i=1}^{N_{elec}} \varepsilon_i - \frac{1}{2} \sum_{ij}^{N_{elec}} \left(\boldsymbol{J}_{ij} - \boldsymbol{K}_{ij} \right) + E_{ZZ}$$
$$\varepsilon_i = \langle \phi_i \left| \boldsymbol{F}_i \right| \phi_i \rangle = h_i + \frac{1}{2} \sum_{j}^{N_{elec}} \left(\boldsymbol{J}_{ij} - \boldsymbol{K}_{ij} \right) \tag{1-18}$$

Roothaan[9] introduced the basis set approximation in which each MO can be expanded as a linear combination of atomic orbitals (LCAO). The HF equations is written as

$$\boldsymbol{F}_i \sum_{\nu}^{M_{basis}} c_{\nu i} \chi_\nu = \varepsilon_i \sum_{\nu}^{M_{basis}} c_{\nu i} \chi_\nu$$
$$\phi_i = \sum_{\nu}^{M_{basis}} c_{\nu i} \chi_\nu \tag{1-19}$$

where c_{vi} are the expansion coefficients, M_{basis} is the number of basis functions in the expansion and χ_v are the basis functions. χ_v can be considered as atomic orbitals (s, p, d, etc.) from the rigorous solution of Schrödinger equation for a hydrogen atom. From now on, the problem of solving HF molecular orbitals becomes calculating the set of expansion coefficients c_{vi}. Multiplying by χ_μ^* on the left and then integrating yields the Roothaan-Hall equations (for the closed systems).[9,10]

$$\begin{aligned} &\boldsymbol{FC} = \boldsymbol{SC}\varepsilon \\ &F_{\mu\nu} = \langle \chi_\mu | \boldsymbol{F} | \chi_\nu \rangle \\ &S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \end{aligned} \tag{1-20}$$

$$\begin{aligned} F_{\mu\nu} &= \langle \chi_\mu | \boldsymbol{F} | \chi_\nu \rangle = \langle \chi_\mu | \boldsymbol{h} | \chi_\nu \rangle + \sum_j^{occ} \langle \chi_\mu | \boldsymbol{J}_j - \boldsymbol{K}_j | \chi_\nu \rangle \\ &= \langle \chi_\mu | \boldsymbol{h} | \chi_\nu \rangle + \sum_j^{occ} \sum_{\rho\sigma}^{M_{basis}} c_{\rho j} c_{\sigma j} \left(\langle \chi_\mu \chi_\rho | \boldsymbol{g} | \chi_\nu \chi_\sigma \rangle - \langle \chi_\mu \chi_\rho | \boldsymbol{g} | \chi_\sigma \chi_\nu \rangle \right) \\ P_{\rho\sigma} &= \sum_j^{occ} c_{\rho j} c_{\sigma j} \end{aligned} \tag{1-21}$$

where $P_{\lambda\sigma}$ is defined as a density matrix which is directly related to the expansion coefficients. The sum runs over the occupied orbitals of the molecular system.

The MO coefficients are determined in a self-consistent manner. The HF energy in basis set expression is

$$E_{total} = \sum_{\mu\nu} P_{\mu\nu} \left(T_{\mu\nu} + Z_{\mu\nu} \right) + \frac{1}{2} \sum_{\mu\nu\rho\sigma} P_{\mu\nu} P_{\rho\sigma} \left(\langle \mu\nu | \rho\sigma \rangle - \langle \mu\nu | \sigma\rho \rangle \right) + E_{ZZ} \tag{1-22}$$

There are two types of spin orbitals: restricted spin orbitals, which are constrained to have the same spatial function for α and β spin functions; and unrestricted spin orbitals which may have different spatial functions. The UHF wave function allows different spatial orbitals for the two electrons in an orbital, thus increasing the freedom,

so the UHF energy will in principle be lower than or equal to a corresponding R(O)HF energy. It is well known that UHF wave function has the spin contamination problem, because it is not the eigenfunction of the total spin operator, which means a ground state UHF wave function may also contain contributions from higher states.

The Hartree-Fock method is referred to as a mean-field approximation, implying that it ignores the electron correlation. It uses only one Slater determinant to build the trial wave function. In principle, the exact wave function should be a linear combination of Slater determinants. Therefore, the next logical step is to include more than one Slater determinant into a trial wave function. Configuration interaction (CI)[7,22,23] is a method to variationally determine the expansion coefficients in the multi-determinant wave function. A full CI calculation with a complete basis set gives the best energy. The Møller-Plesset perturbation method[8] selects the Hartree-Fock determinant as the reference. The details will be introduced in Chapter 4. Perturbation methods make corrections to the reference energy and the wave function to a given order, while coupled cluster (CC) methods include all corrections to infinite order.[11] Perturbation methods and CC methods are able to recover the dynamical correlation, which is associated with the instant correlation between electrons. Multi-configuration self-consistent field (MCSCF)[27] methods are mainly used to recover the static part of the correlation, which is also called the near-degeneracy effect. In MCSCF methods, several configurations are selected by considering the computational cost and the properties to be studied. The expansion coefficients of both the determinants and the basis functions are optimized to get the lowest possible energy. Density functional theory (DFT)[22,24,28,29] is a very popular

method because it is able to include electron correlation in a less costly way as compared to the above methods.

1.3.2 Nuclear gradients

Optimization of the molecular geometry is one of the common tasks in computational chemistry. Analytical energy gradients with respect to the nuclear coordinates are important for geometry optimizations and molecular dynamics simulations.

This section provides a brief review of the derivation of the HF energy gradients. Taking the first derivative of the HF energy with respect to a nuclear coordinate x yields

$$\begin{aligned} E^x_{total} &= \sum_{\mu\nu} P_{\mu\nu}\left(T^x_{\mu\nu} + V^x_{\mu\nu,N}\right) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\left(\langle\mu\nu|\rho\sigma\rangle - \langle\mu\nu|\sigma\rho\rangle\right)^x \\ &+ \sum_{\mu\nu} P^x_{\mu\nu}\left(T_{\mu\nu} + V_{\mu\nu,N}\right) + \sum_{\mu\nu\rho\sigma} P^x_{\mu\nu}P_{\rho\sigma}\left(\langle\mu\nu|\rho\sigma\rangle - \langle\mu\nu|\sigma\rho\rangle\right) + E^x_{NN} \end{aligned} \tag{1-23}$$

The third and fourth terms in the above equation contain the change of the density matrix with respect to the nuclear displacement, which does not have an analytic expression. Pulay[22,30] discovered the following connection and solved the problem gracefully.

$$\begin{aligned} E^x_{total} &= \sum_{\mu\nu} P_{\mu\nu}\left(T^x_{\mu\nu} + V^x_{\mu\nu,N}\right) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\left(\langle\mu\nu|\rho\sigma\rangle - \langle\mu\nu|\sigma\rho\rangle\right)^x \\ &+ \sum_{\mu\nu} P^x_{\mu\nu}F_{\mu\nu} + E^x_{NN} \end{aligned} \tag{1-24}$$

$$\begin{aligned} \sum_{\mu\nu} P^x_{\mu\nu}F_{\mu\nu} &= \sum_{\mu\nu}\sum_{i=1}^{N_{elec}} n_i\left(c^x_{\mu i}F_{\mu\nu}c_{\nu i} + c_{\mu i}F_{\mu\nu}c^x_{\nu i}\right) \\ &= \sum_{\mu\nu}\sum_{i=1}^{N_{elec}} n_i\left(c^x_{\mu i}S_{\mu\nu}\varepsilon_i c_{\nu i} + c_{\mu i}S_{\mu\nu}\varepsilon_i c^x_{\nu i}\right) \end{aligned} \tag{1-25}$$

Accepting the condition that MOs are orthonormal yields an important equation and finally solved the problem.

$$\begin{aligned}
&\langle i|j\rangle = \sum_{\mu\nu}^{M_{basis}} c_{\mu i} c_{\nu j} \langle \mu|\nu\rangle = \sum_{\mu\nu}^{M_{basis}} c_{\mu i} c_{\nu j} S_{\mu\nu} = \delta_{ij} \\
&\langle i|j\rangle^x = c^x_{\mu i} S_{\mu\nu} c_{\nu i} + c_{\mu i} S^x_{\mu\nu} c_{\nu i} + c_{\mu i} S_{\mu\nu} c^x_{\nu i} = 0 \\
&2c^x_{\mu i} S_{\mu\nu} c_{\nu i} = -c_{\mu i} S^x_{\mu\nu} c_{\nu i} \\
&W_{\mu\nu} = \sum_{i=1}^{N_{elec}} c_{\mu i} S^x_{\mu\nu} c_{\nu i}
\end{aligned} \tag{1-26}$$

Therefore, the derivative of the density matrix is converted to the derivative of the overlap matrix. The final expression of the HF energy gradient is:

$$\begin{aligned}
E^x_{total} = &\sum_{\mu\nu} P_{\mu\nu} \left(T^x_{\mu\nu} + V^x_{\mu\nu,N} \right) + \frac{1}{2} \sum_{\mu\nu\rho\sigma} P_{\mu\nu} P_{\rho\sigma} \left(\langle \mu\nu|\rho\sigma\rangle - \langle \mu\nu|\sigma\rho\rangle \right)^x \\
&- \sum_{\mu\nu} W_{\mu\nu} S^x_{\mu\nu} + E^x_{NN}
\end{aligned} \tag{1-27}$$

In the following chapters, the gradients will be formulated for HF, MP2, TDDFT energies with the presence of solvent molecules.

1.3.3 Mixed QM/MM models

Water is the most abundant component in living organisms and it widely influences the chemical phenomena in the condensed phases. The importance of modeling solvent effects can hardly be overemphasized. Various models differ by the representation of the molecular structure, the physical interactions during the solvation process, and the formulas used to express the interactions.[24,31,32]

Continuum models use a dielectric medium to represent the solvent molecules, regardless of the atoms and bonds in their structures. Consider a solute molecule represented by a charge distribution, when it is transferred from the gas phase into a solvent, the solvent molecules would rotate to adjust their dipole orientations. The polarized solvent will generate another electric field, in terms of the ‘reaction field’. The presence of a reaction field implies that the Hamiltonian operator should include this

term, therefore the solute wave function will be perturbed and the energy levels will be shifted. The reorientation process costs the free energy. The magnitude of energy lost equals one half of the total interaction energy between the solute and the solvent under the linear response assumption. Furthermore, the solute-solvent interaction also includes dispersion and repulsion, exchange and charge transfer. The details of the polarizable continuum models will be discussed in Chapter 2.

The molecular mechanics (MM), also referred to as force field (FF) methods, divide the potential energy into several terms as functions of nuclear coordinates, and fit the parameters to experimental value or higher-level computational data. In the force field program CHARMM,[33] the typical energy function form is

$$\begin{aligned} G_{MM} = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 \\ & + \sum_{dihedras} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropers} K_\varphi (\varphi - \varphi_0)^2 \\ & + \sum_{nonbond} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \end{aligned} \tag{1-28}$$

where K_b, K_θ, K_{UB}, K_χ, K_φ are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constant, respectively, b, θ, S, χ and φ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero being the equilibrium values for each terms. The first five terms dealing with the specific internal degrees of freedom within molecules are name as bonded energy. Term 1, 2, 3, 5 are the Taylor expansion around an equilibrium distance, terminating at the second order gives the CHARMM function form. The harmonic potential is sufficient for most equilibrium geometries. The way to improve it is to include more terms from the Taylor expansion. Treatment for the bonded

energies is relatively successful due to the short-range nature. However, it is more sophisticated and computationally challenging for the last two non-bonded terms.

The sixth term is the van der Waals energy described using the Lennard-Jones (LJ) potential.[34] It describes the attraction or repulsion between atoms that are not directly bonded. ε is the Lennard-Jones well depth and σ_{ij} is the finite distance at which the inter-particle potential is zero. The van der Waals interaction and electrostatic interaction (the seventh term) are named as non-bonded interaction.

The seventh term is electrostatic interaction due to the positive and negative partial charge of the molecule with only the monopolar potential considered. The partial charges of atoms are determined by empirical rules or fitting to the electrostatic potential calculated using QM methods. The interaction between partial charges is given by the Coulombic potential. MM models vary by the distributions of charges and the mathematical descriptions of the interactions. Using only fixed charges to model electrostatic energy is inaccurate because the electronic charges are actually distributed continuously as electron clouds. Including non-atom-centered charges[35] or higher order moments leads to more reliable models, for example, the distributed multipole analysis (DMA)[36] developed by Stone.

Combined QM/MM methods were originally formulated by Warshel[37] and later developed by many other researchers.[38-45] In a QM/MM/Continuum approach, the system is partitioned into three layers. The inner layer is treated quantum mechanically, the second layer uses a molecular mechanics (MM) method, and the third layer uses a continuum description. In Cui's QM/MM/IEF-PCM model,[46] both QM and MM are embedded in a continuum and MM is represented by fixed partial charges. The effective

fragment potential (EFP)[47,48] method developed by Gordon and coworkers has been interfaced with various branches of PCM methods, such as the Onsager-like model,[48] IEF-PCM,[49] and CPCM.[50] In the EFP method, the solvent molecules are represented by fragments. The charge distribution of the fragments is represented by an arbitrary number of charges, dipoles, quadruples and octupoles. In the combined QM/EFP/CPCM method developed by Li,[51] the linear response and variational treatment are applied to formulate the QM(=HF, DFT)/EFP/CPCM methods. The induced surface charge and induced dipole are iterated self-consistently and the exact analytical nuclear gradients were obtained. Morokuma and coworkers have developed the ONIOM method[52] which applies various levels of theory on various regions of the system. The energies acquired at each level are finally corrected to give the total energy. Instead of performing the high level calculation for a large real system, three less expensive calculations are performed: a low-level calculation for the whole system, the same low-level calculation for a smaller part of the system, and a high-level calculation for the same part of the system. Assuming the corrections are additive, then the energy of the real system at high level can be estimated.

The rigorous way to calculate the solvent effects on molecular properties is to carry out full quantum mechanical calculations on the solution system, and take into account various configurations of the solvent molecules. To date, such a method requires huge computational efforts and is usually impractical. Both continuum models and discrete models have weaknesses. For example, continuum models are not able to accurately describe the solute-solvent interaction, while discrete models are always size-limited and the long-range interaction is problematic.

Chapter 2 Heterogeneous conductor-like solvation model

2.1 Introduction

2.1.1 Polarizable continuum model

Continuum solvation models have been widely used to study molecules in solutions.[32] Rather than describing each solvent molecule explicitly, it employs a dielectric medium to represent the solvent molecules. The dielectric medium is characterized by its dielectric constant, and the value can be taken from the experimental value of the bulk solvent. The solute-solvent interactions can be classified into four terms, including the electrostatic intereaction, cavitation, exchange-repulsion and dispersion. The last three terms have separate functional forms and are treated empirically. This chapter focuses on the electrostatic interaction. The solute molecule is treated as a charge distribution. It polarizes the dielectric continuum and produces induced surface charges, which in turn polarize the solute charge distribution. In the framework of quantum mechanics, the solvation operator is represented by a one-electron operator in the same form as the Coulomb operator. The solvation operator is incorporated into the effective Hamiltonian, so the electronic wave function can be solved self-consistently. It is referred to as the self-consistent reaction-field (SCRF) model. The electrostatic and polarization interactions between the solute molecule and the solvent molecules are described by the classic electrostatic Poisson equation, which expresses the dependence of electrostatic potential on the charge density and the dielectric constant:

$$\nabla^2\phi(\boldsymbol{r}) = -\frac{4\pi\rho(\boldsymbol{r})}{\varepsilon} \tag{2-1}$$

In PCM, the molecular cavity is the portion of space in the medium occupied by the solute. As introduced by Onsager,[53] a cavity should have a physical meaning that it should exclude the solvent while including the largest possible part of the solute charge distribution. The most common way to define a molecular cavity is to use a set of interlocking spheres centered at each nucleus constituting the solute. In some schemes, additional spheres are used to smooth out sharp grooves. The radius of each sphere can use van der Waals radius,[54] or other empirical values, such as the simplified united atomic radii (SUAHF) scaled by 1.2. In general, the PCM system should be a very dilute solution.

The boundary of the molecular cavity is the molecular surface. Since the PCM cavity has a complex shape, the induced surface charges have to be calculated numerically. Different versions of PCM define different formulas to calculate the apparent surface charges (ASCs), namely integral equation formalism PCM (IEF-PCM),[13,14] dielectric polarizable continuum model (DPCM),[12] conductor-like screening models (COSMO[55] or conductor-like polarizable continuum model, CPCM[56,57]), the surface and simulation of volume polarization for electrostatics models (SS(V)PE[58]), the SM*x* models,[59] and etc. The formulas have been tabulated in Tomasi's review[32] in 2005. In CPCM, the boundary conditions are simplified. Given the condition that the potentials caused by the solute and the solvent should cancel out on the surface of a conductor, the ASC is determined by the local value of electrostatic potential instead of its gradient in Eq. (2-1). In practice, the surface is divided into small regions named tesserae. A

differentiable tessellation is defined in order to analytically calculate derivatives with respect to the nuclear displacement.

2.1.2 Heterogeneity of the solution

Although in many cases the solvent can be treated as a homogeneous and isotropic polarizable medium with a dielectric constant, there are cases in which the heterogeneity or anisotropy of the environment surrounding a molecule must be considered. Typical examples include solute molecules at the interface between two phases, proteins and other molecules embedded in lipid bilayers, solvated guest-host complexes, and protein active sites or cofactors in protein matrices.

Using DPCM, Tomasi's group performed pioneering studies on various heterogeneous solvation problems. For example, the energy changes in deformations of long DNA fragments and that in the opening of a DNA double helix,[60] the partial solvation effect in molecular recognition and docking,[61] and polar solutes placed near the surface of two immiscible liquids or at a liquid/vacuum separation.[62] Hoshi, Sakurai, Inoue and Chujo extended the DPCM to treat anisotropic polarization effects in their guest-host complex calculations.[63]

Cances *et al*.[13,14] developed a general integral equation formalism polarizable continuum model (IEF-PCM) for treating anisotropic solvation problems. Frediani *et al*.[64] later extended IEF-PCM to study molecules at diffuse interfaces between two fluid phases by introducing position dependent permittivities. By using different local effective dielectrics for different portions of the solute cavity surface, Iozzi, Cossi, Improta, Rega and Barone further extended IEF-PCM to study the pK_a of a solvent exposed histidine residue in prion protein and a small molecule interacting with a biological membrane.[15] A

detailed discussion on heterogeneous solvation models can be found in a review by Tomasi *et al*.[32]

To date, no analytic nuclear gradients for heterogeneous continuum solvation models have been derived and implemented, and no geometry optimizations have been reported.

In this chapter, a heterogeneous conductor-like solvation model (COSMO and CPCM, high-dielectric versions of the more general IEF-PCM[13,14]) is implemented. Following the heterogeneous IEF-PCM developed by Barone *et al.*,[15] different local effective dielectrics are used for different portions of a solute cavity surface. By variationally formulating the solvation operators for Hartree-Fock and Kohn-Sham methods, using the FIXPVA[16] tessellation scheme, and assuming that the local effective dielectrics are constants in a geometry optimization process, continuous and smooth potential energy surfaces and analytic nuclear gradients can be obtained for the heterogeneous solvation model.

2.2 Theory

2.2.1 Heterogeneous COSMO and CPCM

The conductor-like screening model (COSMO) was originally developed by Klamt and Schüürmann.[55] The conductor-like polarizable continuum model (CPCM), similar to COSMO, was implemented based on the more general IEF-PCM method.[56,57] In the following, COSMO and CPCM are discussed together.

Although previous studies[56,65] have demonstrated that acceptable results could be obtained with COSMO and CPCM for rather low dielectric solvents, it is important to note that both COSMO and CPCM are approximations of IEF-PCM, which is more

rigorous from an electrostatic interaction point of view, and can be used with equal accuracy for both high and low dielectric solvents.[13,14] In this chapter, only COSMO and CPCM are discussed because their simplicity allows for an easier treatment in the development of a heterogeneous model.

It is important to keep in mind that the rigorous variational treatment presented in this chapter for heterogeneous COSMO and CPCM solely means that the mathematical structure of these methods in combination with Hartree-Fock and Kohn-Sham methods is analyzed exactly, not that the models are electrostatically rigorous.

COSMO and CPCM describe the solvent reaction potential with induced surface charges distributed on the solute cavity surface. By using boundary element method, the continuous distribution of the induced surface charge is represented by a set of induced point charges located at the surface tesserae. These point charges, written as a vector **q**, satisfy the following matrix equation:

$$\mathbf{Cq} = -k\mathbf{V} \tag{2-2}$$

where the vector **V** collects the electrostatic potentials created by the solute at the surface tesserae. The elements of symmetric matrix **C** are

$$\mathbf{C}_{ii} = 1.07\sqrt{\frac{4\pi}{a_i}} \tag{2-3}$$

$$\mathbf{C}_{ij} = \frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} \tag{2-4}$$

with a_i being the area and $\mathbf{r}_i$ being the center coordinates of tessera i.

$\boldsymbol{Cq}$ is essentially the potential created by the surface charges at each tessera. If the solvent were an ideal conductor, the corresponding dielectric constant $\varepsilon = \infty$. The boundary condition is that the electric potentials caused by the solute and the solvent on

the cavity surface must cancel out. Since the electric potential **V** caused by the solute is known, it is possible to calculate the surface charge at each tessera. For a solvent with a finite dielectric constant ε, the potentials do not cancel out completely, thus the ASCs are screened by approximately a scaling function $(\varepsilon - 1)/(\varepsilon + k)$ with k being 0.5 or 0.

For homogeneous solvents *k* in Eq. (2-2) is a pure number:

$$k = (\varepsilon - 1)/\varepsilon \tag{2-5}$$

with ε being the dielectric constant of the solvent. Other values for *k* have been used in the literature. For example, Klamt and Schüürmann originally suggested $k = (\varepsilon - 1)/(\varepsilon + 0.5)$.[55]

In this heterogeneous CPCM model, the heterogeneity of the environment surrounding the solute is modeled by using a local effective dielectric ε_i for each surface tessera, similar to that in the heterogeneous IEF-PCM developed by Barone *et al*.[15] Mathematically, this corresponds to replacing the pure number *k* in Eq. (2-2) with a diagonal matrix **K** with the following elements:

$$\mathbf{K}_{ii} = (\varepsilon_i - 1)/\varepsilon_i \tag{2-6}$$

$$\mathbf{K}_{ij} = 0 \tag{2-7}$$

The local effective dielectrics for the tesserae must be determined according to the specific problems, and can be difficult in some cases. For example, it is useful to define the local dielectric for a tessera as a function of its position in the study of biomolecules embedded in lipid membranes. In this heterogeneous CPCM model, local dielectrics are assigned to surface tesserae according to the spheres (i.e. atoms) they belong to: the tesserae on the same sphere have the same local effective dielectric given by the user from the input deck.

2.2.2 Solvation operator in Hartree-Fock and Kohn-Sham methods

In Hartree-Fock and Kohn-Sham calculations, the heterogeneous COSMO and CPCM induced surface charges can be determined separately for solute nuclei and electrons:

$$\mathbf{q}_N = -\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N \tag{2-8}$$

$$\mathbf{q}_e = -\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_e \tag{2-9}$$

where $\mathbf{V}_N$ and $\mathbf{V}_e$ are the nuclear and electronic potential, respectively, at the tesserae. Later it will be shown that actually a more efficient "total charge" scheme can be adapted.

Eq. (2-9) can be equivalently written as the basis set induced surface charges $\mathbf{q}_{\rho\sigma}$ contracted by the density matrix:

$$\mathbf{q}_e(i) = \sum_{\rho\sigma} P_{\rho\sigma}\mathbf{q}_{\rho\sigma}(i) \tag{2-10}$$

where ρ and σ, as well as μ and ν used later, are Gaussian type basis functions; $P_{\rho\sigma}$ is the density matrix; $\mathbf{q}_{\rho\sigma}$ is obtained by solving Eq. (2-11) for the basis set potentials $\mathbf{V}_{\rho\sigma}$ at the tesserae:

$$\mathbf{q}_{\rho\sigma} = -\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_{\rho\sigma} \tag{2-11}$$

with

$$\boldsymbol{V}_{\rho\sigma}(i) = -\left\langle \rho \left| \frac{1}{|\boldsymbol{r}_e - \boldsymbol{r}_i|} \right| \sigma \right\rangle \tag{2-12}$$

where $\boldsymbol{r}_e$ is the electronic coordinate and $\boldsymbol{r}_i$ is the tessera coordinate.

The electrostatic potentials generated by the induced surface charges are then incorporated into the Hartree-Fock or Kohn-Sham equations to variationally determine the total molecular energy E_{total} over a finite basis set:

$$\begin{aligned} E_{total} &= \sum_{\mu\nu} P_{\mu\nu}\left(T_{\mu\nu}+Z_{\mu\nu}\right)+\frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\left(\langle\mu\nu|\rho\sigma\rangle\right)+\sum_{\mu\nu} P_{\mu\nu}\left\langle\mu\middle|V_{xc}(\mathbf{r}_e)\middle|\nu\right\rangle + E_{ZZ} \\ &-\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_i\left\langle\mu\middle|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\middle|\nu\right\rangle-\frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_i\left\langle\mu\middle|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\middle|\nu\right\rangle \\ &+\frac{1}{2}\sum_\alpha\sum_i\frac{Z_\alpha\mathbf{q}_N(i)}{|\mathbf{r}_\alpha-\mathbf{r}_i|}+\frac{1}{2}\sum_{\rho\sigma} P_{\rho\sigma}\sum_\alpha\sum_i\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha-\mathbf{r}_i|} \end{aligned} \tag{2-13}$$

where $T_{\mu\nu}$ and $Z_{\mu\nu}$ are the basis set kinetic energy and electron-nuclei potential energy integrals, respectively; $V_{xc}(\mathbf{r}_e)$ is the Hartree-Fock exchange or the DFT exchange-correlation potential; E_{ZZ} is the nuclear repulsion energy; Z_α and $\mathbf{r}_\alpha$ are the nuclear charge and coordinates of atom α;

The first four terms in Eq. (2-13) have the same forms as those in the gas phase Hartree-Fock and Kohn-Sham methods, but evaluated with the solvent perturbed density matrix. They can be denoted as E_{gas}. The last four terms in Eq. (2-13) are due to induced surface charges, and can be denoted as E_{sol}:

$$\begin{aligned} E_{total} &= E_{gas}+\frac{1}{2}\mathbf{V}_e^T\mathbf{q}_N+\frac{1}{2}\mathbf{V}_e^T\mathbf{q}_e+\frac{1}{2}\mathbf{V}_N^T\mathbf{q}_N+\frac{1}{2}\mathbf{V}_N^T\mathbf{q}_e \\ &= E_{gas}+\frac{1}{2}\mathbf{V}^T\mathbf{q} \\ &= E_{gas}+E_{sol} \end{aligned} \tag{2-14}$$

The solvation operator must be determined based on a variational treatment of Eq. (2-13). The fourth and seventh terms in Eq. (2-13) do not contain electronic coordinates thus are irrelevant. Variational treatments of the first three terms in Eq. (2-13) lead to the gas phase Fock or Kohn-Sham operators. A variational treatment of the fifth term in Eq.

(2-13) leads to an operator representing the potential due to nucleus induced surface charge:

$$\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \Rightarrow \frac{1}{2}\sum_i \frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \tag{2-15}$$

A variational treatment of the sixth term in Eq. (2-13) with respect to the density matrix gives:

$$\begin{aligned}
&\delta\left(-\frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle\right) \\
&= -\frac{1}{2}\sum_{\mu\nu\rho\sigma} \delta P_{\mu\nu}P_{\rho\sigma}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle - \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}\delta P_{\rho\sigma}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \\
&= -\frac{1}{2}\sum_{\mu\nu} \delta P_{\mu\nu}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_e(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle + \frac{1}{2}\sum_{\rho\sigma} \delta P_{\rho\sigma}\sum_i\sum_j \mathbf{V}_e(i)\mathbf{C}_{ij}^{-1}\mathbf{K}_{jj}\mathbf{V}_{\rho\sigma}(j) \\
&= -\frac{1}{2}\sum_{\mu\nu} \delta P_{\mu\nu}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_e(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle - \frac{1}{2}\sum_{\rho\sigma} \delta P_{\rho\sigma}\sum_j \left\langle \rho \left| \frac{\tilde{\mathbf{q}}_e(j)}{|\mathbf{r}_e - \mathbf{r}_j|} \right| \sigma \right\rangle
\end{aligned} \tag{2-16}$$

Therefore, the sixth term leads to an operator representing two sets of potentials due to electron induced surface charges:

$$-\frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_i \left\langle \mu \left| \frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \Rightarrow \frac{1}{2}\frac{\mathbf{q}_e + \tilde{\mathbf{q}}_e}{|\mathbf{r}_e - \mathbf{r}_i|} \tag{2-17}$$

where $\tilde{\mathbf{q}}_e$ is:

$$\tilde{\mathbf{q}}_e = -\mathbf{K}\mathbf{C}^{-1}\mathbf{V}_e \tag{2-18}$$

Because of the following inequality,

$$\mathbf{C}^{-1}\mathbf{K} \neq \mathbf{K}\mathbf{C}^{-1} \tag{2-19}$$

$\tilde{\mathbf{q}}_e$ is different from $\mathbf{q}_e$.

The eighth term in Eq. (2-13) leads to an operator representing a potential due to nucleus induced surface charges:

$$\frac{1}{2}\sum_{\rho\sigma}P_{\rho\sigma}\sum_{\alpha}\sum_{i}\frac{Z_{\alpha}\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_{\alpha}-\mathbf{r}_{i}|}\Rightarrow\frac{1}{2}\sum_{i}\frac{\tilde{\mathbf{q}}_{N}(i)}{|\mathbf{r}_{e}-\mathbf{r}_{i}|} \tag{2-20}$$

where $\tilde{\mathbf{q}}_N$ is:

$$\tilde{\mathbf{q}}_{N}=-\mathbf{K}\mathbf{C}^{-1}\mathbf{V}_{N} \tag{2-21}$$

Therefore, in heterogeneous COSMO and CPCM calculations, the following solvation operator **B** shall be added to the gas phase Fock or Kohn-Sham operators:

$$\mathbf{B}=\frac{1}{2}\frac{\mathbf{q}_{e}+\tilde{\mathbf{q}}_{e}+\mathbf{q}_{N}+\tilde{\mathbf{q}}_{N}}{|\mathbf{r}_{e}-\mathbf{r}_{i}|}=\frac{1}{2}\frac{\mathbf{q}+\tilde{\mathbf{q}}}{|\mathbf{r}_{e}-\mathbf{r}_{i}|} \tag{2-22}$$

Eq. (2-22) shows that the **B** operator can be constructed with the total induced charge $\mathbf{q}=\mathbf{q}_e+\mathbf{q}_N$ and $\tilde{\mathbf{q}}=\tilde{\mathbf{q}}_e+\tilde{\mathbf{q}}_N$. This is more efficient for iterative solution of the surface charges because only one solution is necessary. For homogeneous COSMO and CPCM, because $\tilde{\mathbf{q}}_e=\mathbf{q}_e$ and $\tilde{\mathbf{q}}_N=\mathbf{q}_N$, the **B** operator is simpler:

$$\mathbf{B}=\frac{\mathbf{q}_{e}+\mathbf{q}_{N}}{|\mathbf{r}_{e}-\mathbf{r}_{i}|}=\frac{\mathbf{q}}{|\mathbf{r}_{e}-\mathbf{r}_{i}|} \tag{2-23}$$

Eq. (2-23) shows that for homogeneous COSMO and CPCM it is possible to use only one set of induced charges because the **C** matrix in Eq. (2-2) is symmetric, while for heterogeneous COSMO and CPCM at least two sets of induced charges are necessary because of the inequality shown by Eq. (2-19). It is noted that for DPCM and IEF-PCM, the corresponding matrices are asymmetric and two sets of induced charges shall be used for rigorousness.

2.2.3 Nuclear gradients

A direct differentiation of Eq. (2-13) with respect to a nuclear coordinate *x* produces:

$$
\begin{aligned}
E^x_{total} = & \sum_{\mu\nu} P_{\mu\nu}\left(T^x_{\mu\nu} + V^x_{\mu\nu,N}\right) + \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\langle\mu\nu|\rho\sigma\rangle^x + \sum_{\mu\nu} P_{\mu\nu}\left\langle\mu\left|V_{xc}(\mathbf{r}_e)\right|\nu\right\rangle^x + E^x_{NN} \\
& - \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x - \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P_{\rho\sigma}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^x \\
& + \frac{1}{2}\sum_{\alpha}\sum_i\left(\frac{Z_\alpha\mathbf{q}_N(i)}{|\mathbf{r}_\alpha-\mathbf{r}_i|}\right)^x + \frac{1}{2}\sum_{\rho\sigma}P_{\rho\sigma}\left(\sum_\alpha\sum_i\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha-\mathbf{r}_i|}\right)^x \\
& + \sum_{\mu\nu} P^x_{\mu\nu}\left(T_{\mu\nu} + V_{\mu\nu,N}\right) + \sum_{\mu\nu\rho\sigma} P^x_{\mu\nu}P_{\rho\sigma}\langle\mu\nu|\rho\sigma\rangle + \sum_{\mu\nu} P^x_{\mu\nu}\left\langle\mu\left|V_{xc}(\mathbf{r}_e)\right|\nu\right\rangle \\
& - \frac{1}{2}\sum_{\mu\nu} P^x_{\mu\nu}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_N(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle - \frac{1}{2}\sum_{\mu\nu\rho\sigma} P^x_{\mu\nu}P_{\rho\sigma}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle \\
& - \frac{1}{2}\sum_{\mu\nu\rho\sigma} P_{\mu\nu}P^x_{\rho\sigma}\sum_i\left\langle\mu\left|\frac{\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle + \frac{1}{2}\sum_{\rho\sigma}P^x_{\rho\sigma}\sum_\alpha\sum_i\frac{Z_\alpha\mathbf{q}_{\rho\sigma}(i)}{|\mathbf{r}_\alpha-\mathbf{r}_i|}
\end{aligned}
\tag{2-24}
$$

The last seven terms in Eq. (2-24) contain the derivatives of the density matrix, and can be written as:

$$
\begin{aligned}
& \sum_{\mu\nu} P^x_{\mu\nu}\left(T_{\mu\nu} + V_{\mu\nu,N}\right) + \sum_{\mu\nu\rho\sigma} P^x_{\mu\nu}P_{\rho\sigma}\langle\mu\nu|\rho\sigma\rangle + \sum_{\mu\nu} P^x_{\mu\nu}\left\langle\mu\left|\frac{V_{xc}(\boldsymbol{r})}{|\boldsymbol{r}_e-\boldsymbol{r}|}\right|\nu\right\rangle + \sum_{\mu\nu} P^x_{\mu\nu}\boldsymbol{B}_{\mu\nu} \\
& = -\sum_{\mu\nu} W_{\mu\nu}S^x_{\mu\nu}
\end{aligned}
\tag{2-25}
$$

where W is the energy-weighted density matrix. Clearly, the explicit evaluation of the density matrix derivatives can be avoided, as originally derived by Pulay for the gas phase Hartree-Fock methods.[30]

The first four and the last seven terms in Eq. (2-24) have the same forms as those in the gas phase Hartree-Fock and Kohn-Sham methods, and can be evaluated using the same techniques with the solvent perturbed **P** and **W**. They are denoted as E^x_{gas}. The

remaining terms (fifth, sixth, seventh and eighth) in Eq. (2-24) contain the derivatives of the solvation terms, and are denoted as E_{sol}^{x}:

$$E_{total}^{x} = E_{gas}^{x} + E_{sol}^{x} \tag{2-26}$$

The fifth term in Eq. (2-24) can be written as:

$$\begin{aligned} & -\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu} \sum_{i} \left\langle \mu \left| \frac{\mathbf{q}_N(i)}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle^{x} \\ & = -\frac{1}{2}\sum_{\mu\nu} P_{\mu\nu} \sum_{i} \left\langle \mu \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle^{x} \mathbf{q}_N(i) - \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu} \sum_{i} \left\langle \mu \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \mathbf{q}_N^{x}(i) \end{aligned} \tag{2-27}$$

The derivative of the induced charge $\mathbf{q}^{x}$ in Eq. (2-27) can be avoided by converting the induced surface charge back into potential:

$$\begin{aligned} & -\frac{1}{2}\sum_{\boldsymbol{\mu\nu}} P_{\boldsymbol{\mu\nu}} \sum_{i} \left\langle \boldsymbol{\mu} \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \boldsymbol{\nu} \right\rangle \mathbf{q}_N^{x}(i) \\ & = \frac{1}{2}\sum_{i} \mathbf{V}_e(i)\mathbf{q}_N^{x}(i) = \frac{1}{2}\mathbf{V}_e\mathbf{q}_N^{x} = \frac{1}{2}\mathbf{V}_e\left(-\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N\right)^{x} \\ & = -\frac{1}{2}\mathbf{V}_e\left[\left(\mathbf{C}^{-1}\right)^{x}\mathbf{K}\mathbf{V}_N + \mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N^{x}\right] \\ & = \frac{1}{2}(\mathbf{V}_e\mathbf{C}^{-1})\mathbf{C}^{x}(\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N) - \frac{1}{2}\mathbf{V}_e\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N^{x} \\ & = \frac{1}{2}(\mathbf{V}_e\mathbf{C}^{-1}\mathbf{K})\mathbf{K}^{-1}\mathbf{C}^{x}(\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N) - \frac{1}{2}\mathbf{V}_e\mathbf{C}^{-1}\mathbf{K}\mathbf{V}_N^{x} \\ & = \frac{1}{2}\tilde{\mathbf{q}}_e^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}_N + \frac{1}{2}\tilde{\mathbf{q}}_e^{T}\mathbf{V}_N^{x} \end{aligned} \tag{2-28}$$

Note that **K** is a constant matrix so $\mathbf{K}^{x} = \mathbf{0}$ because the local dielectrics are not subject to change with respect to *x*, as discussed at the end of section 2.2.1. So the fifth term in Eq. (2-24) can be written as:

$$
\begin{aligned}
&-\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_{i}\left\langle \mu\left|\frac{\boldsymbol{q}_N(i)}{|\boldsymbol{r}_e-\boldsymbol{r}_i|}\right|\nu\right\rangle^{x} \\
&=-\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_{i}\left\langle \mu\left|\frac{1}{|\boldsymbol{r}_e-\boldsymbol{r}_i|}\right|\nu\right\rangle^{x}\boldsymbol{q}_N(i)+\frac{1}{2}\tilde{\boldsymbol{q}}_e^{T}\boldsymbol{K}^{-1}\boldsymbol{C}^{x}\boldsymbol{q}_N+\frac{1}{2}\tilde{\boldsymbol{q}}_e^{T}\boldsymbol{V}_N^{x}
\end{aligned}
\tag{2-29}
$$

Similarly, the sixth, seventh and eighth terms can also be simplified, and E_{sol}^{x} can be written as:

$$
\begin{aligned}
E_{sol}^{x} &= -\frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}\sum_{i}\left\langle \mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^{x}\mathbf{q}(i)-\frac{1}{2}\sum_{\rho\sigma}P_{\rho\sigma}\sum_{i}\left\langle \rho\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\sigma\right\rangle^{x}\tilde{\mathbf{q}}(i) \\
&\quad+\frac{1}{2}\tilde{\mathbf{q}}^{T}\mathbf{V}_N^{x}+\frac{1}{2}(\mathbf{V}_N^{x})^{T}\mathbf{q} \\
&\quad+\frac{1}{2}\tilde{\mathbf{q}}_e^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}_N+\frac{1}{2}\tilde{\mathbf{q}}_e^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}_e+\frac{1}{2}\tilde{\mathbf{q}}_N^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}_N+\frac{1}{2}\tilde{\mathbf{q}}_N^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}_e \\
&=-\sum_{\mu\nu}P_{\mu\nu}\sum_{i}\left\langle \mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^{x}\left[\frac{\tilde{\mathbf{q}}(i)+\mathbf{q}(i)}{2}\right]+(\mathbf{V}_N^{x})^{T}\left(\frac{\tilde{\mathbf{q}}+\mathbf{q}}{2}\right)-\frac{1}{2}\tilde{\mathbf{q}}^{T}\mathbf{K}^{-1}\mathbf{C}^{x}\mathbf{q}
\end{aligned}
\tag{2-30}
$$

Eq. (2-30) shows that the derivatives of the solvation terms can be evaluated with the induced surface charges $\mathbf{q}$ and $\tilde{\mathbf{q}}$.

In homogeneous COSMO and CPCM method, $\tilde{\mathbf{q}} = \mathbf{q}$, so E_{sol}^{x} becomes:

$$
E_{sol}^{x} = -\sum_{\mu\nu}P_{\mu\nu}\sum_{i}\left\langle \mu\left|\frac{1}{|\mathbf{r}_e-\mathbf{r}_i|}\right|\nu\right\rangle^{x}\mathbf{q}(i)+(\mathbf{V}_N^{x})^{T}\mathbf{q}+\frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\mathbf{q}^{T}\mathbf{C}^{x}\mathbf{q} \tag{2-31}
$$

The first and second terms in Eq. (2-30) or (2-31) are the electrostatic fields at the surface tesserae due to solute electrons and nuclei, and can be evaluated using standard techniques.

The third term in Eq. (2-30) or (2-31) is the interactions between the induced surface charges and can be evaluated with the derivatives of the **C** matrix, which involves the derivatives of the areas and coordinates of the tesserae with respect to the atomic coordinate *x*.

The third term in Eq. (2-31) can be expanded as follows. The diagonal term is

$$\frac{1}{2}\tilde{\boldsymbol{q}}^T\boldsymbol{K}^{-1}\boldsymbol{C}_{ii}^x\boldsymbol{q} = -\frac{1}{2}\frac{1.07\sqrt{4\pi}}{2\left(a_i\right)^{3/2}}\frac{\partial a_i}{\partial x}\tilde{\boldsymbol{q}}^T\boldsymbol{K}^{-1}\boldsymbol{q} \tag{2-32}$$

and the off-diagonal term is

$$\frac{1}{2}\tilde{\boldsymbol{q}}^T\boldsymbol{K}^{-1}\boldsymbol{C}_{ij}^x\boldsymbol{q} = -\frac{1}{2}\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|^{-3}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)\cdot\left(\frac{\partial \boldsymbol{r}_i}{\partial x} - \frac{\partial \boldsymbol{r}_j}{\partial x}\right)\tilde{\boldsymbol{q}}^T\boldsymbol{K}^{-1}\boldsymbol{q} \tag{2-33}$$

Other terms can be calculated in a straightforward way, so the emphasis will be placed on the evaluation of $\partial a_i / \partial x$. The GEPOL[66,67] scheme has been implemented for DPCM,[12] IEF-PCM[13,14] and CPCM[56,57] for both energy and gradients calculations, and exact analytic gradients, as well as smooth potential energy surfaces,[67] have been obtained. Su and Li implemented a tessellation scheme called Fixed Points with Variable Areas (FIXPVA) for COSMO and CPCM.[16] Since the position of a tessera is defined as a fixed point relative to the center of the associated sphere, the value of $\partial \boldsymbol{r}_i / \partial x$ is either 0 or 1 depending on if the point is on the displaced atom. The tessera areas are chosen to be smooth functions of their distance to neighboring spheres shown in Figure 1 in Ref. [16]

$$\begin{aligned} a_i &= a_0 f_1\, f_2 \\ a_0 &= \frac{4\pi R_A^2}{\mathrm{N_{TS}}}, \quad \mathrm{N_{TS}} = 60,\ 240,\ \text{or } 960 \end{aligned} \tag{2-34}$$

with a_0 being the initial surface area, f_1 and f_2 being the fifth-order polynomial functions:

$$f_1 = \begin{cases} 1, & m > m_2, \\ 10\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^3 - 15\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^4 + 6\left(\dfrac{m^2 - m_1^2}{m_2^2 - m_1^2}\right)^5, & m_2 \geq m > m_1, \\ 0, & m_1 \geq m, \end{cases}$$

$$f_2 = \begin{cases} 1, & n > n_2, \\ 10\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^3 - 15\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^4 + 6\left(\dfrac{n^2 - n_1^2}{n_2^2 - n_1^2}\right)^5, & n_2 \geq n > n_1, \\ 0, & n_1 \geq n, \end{cases} \qquad (2\text{-}35)$$

Because of the use of smooth functions, rigorously continuous and smooth potential energy surfaces, as well as exact analytic gradients can be obtained for COSMO and CPCM calculations with HF, MCSCF and DFT methods.

2.3 Computational methodology

The heterogeneous CPCM code was implemented in the GAMESS[68] package as a new option of the CPCM code previously implemented by Li and Jensen[69] on the basis of the IEF-PCM program originally implemented by Mennucci, Cances, Pomelli and Tomasi.[13,70,71]

In the current implementation, local dielectric constants are assigned to surface tesserae according to the spheres (i.e. atoms in the FIXPVA tessellation scheme) they belong to: the tesserae on the same sphere have the same local dielectric constant. Only the electrostatic interaction was considered; heterogeneous cavitation, dispersion and repulsion terms were not considered.

In the CPCM calculations, spheres with radii of 0, 2.124, 2.016, 1.908, 1.800 and 2.760 Å were used for H, C, N, O, Na and Cl atoms, respectively, to define the molecular cavity; no additional spheres were used. Using zero radii for H atoms means that they do not contribute to form the surface. The tessellation scheme FIXPVA was used with 60

initial tesserae per sphere.[16] The induced surface charges were determined by a semi-iterative DIIS procedure[71,72] with no charge renormalization. Geometry optimization was performed in internal coordinates generated by the automatic delocalized coordinates algorithm.[73]

2.4 Results and discussion

2.4.1 Energy

The simplest way to establish a quantitative picture of the heterogeneous COSMO and CPCM method is probably to use an ion pair. Table 2.1 presents the solvation energies [E_{sol} in Eq. (2-14)] calculated with the heterogeneous CPCM/RHF/aug-cc-pVTZ[74,75] method for NaCl using various dielectrics for the tesserae on the Na sphere and the Cl sphere. The NaCl distance is 2.397 Å as optimized with the gas phase RHF/aug-cc-pVTZ method. Spheres with radii of 1.800 and 2.760 Å were used for Na and Cl, respectively, to define the molecular cavity. It is emphasized that these calculations were performed for illustrative purposes rather than to reproduce any experimental or theoretical results.

Table 2.1 Solvation energy (kcal/mol) calculated with heterogeneous CPCM/RHF/aug-cc-pVTZ for NaCl using different local effective dielectrics for Na and Cl.

	Na (ε=1)	Na (ε=2)	Na (ε=4)	Na (ε=20)	Na (ε=78.39)
Cl (ε =1)	0.00	-10.46	-15.98	-20.53	-21.39
Cl (ε =2)	-3.84	-14.83	-20.62	-25.39	-26.29
Cl (ε =4)	-5.93	-17.20	-23.14	-28.03	-28.95
Cl (ε =20)	-7.68	-19.19	-25.26	-30.25	-31.20
Cl (ε =78.39)	-8.02	-19.58	-25.66	-30.68	-31.63

The diagonal data in Table 2.1 are actually homogeneous CPCM results because the same dielectric was used for Na and Cl. Ongoing from low dielectric 2.00 to high dielectric 78.39, the solvation energy changes from -14.83 to -31.63 kcal/mol, roughly doubled. This rough doubling can be easily explained by the scaling factor $(\varepsilon-1)/\varepsilon$, which changes from 0.5 to ~1.0. However, due to the changes in the polarization of the NaCl electron density ongoing from low to high dielectric solvents, the actual solvation energy is more than doubled. If the Na^+ and Cl^+ are represented by point charges with no polarizability, the solvation energy will be exactly doubled ongoing from $\varepsilon = 2$ to $\varepsilon = \infty$.

Holding $\varepsilon=1$ for Cl but varying ε from 1 to 2, 4, 20 and 78.39 for Na, the solvation energy changes from 0.00 to -10.46, -15.98, -20.53 and -21.39 kcal/mol. Holding $\varepsilon=1$ for Na but varying ε from 1 to 2, 4, 20 and 78.39 for Cl, the solvation energy changes from 0.00, -3.84, -5.93, -7.68 and -8.02 kcal/mol. Clearly, the Na exhibits a greater ε dependence because of its smaller radius (1.8 Å for Na and 2.76 Å for Cl).

2.4.2 Gradients

Table 2.2 presents the analytic and numerical gradients obtained with heterogeneous CPCM/B3LYP[76]/6-31G* methods for acetate in the gas phase HF/6-31G* optimized geometry. The numerical gradients were computed with double displacements (forward and backward) using a step size of 0.001 au for each step. A local effective ε =78.39 was used for the carboxylate group and ε =4.0 was used for the methyl group. Again, this is for illustrative purposes.

The maximum difference between the numerical and analytic gradients is 4.4×10^{-7} au, and the root-mean-square difference is 2.2×10^{-7} au. The default settings in GAMESS produce the gas phase gradients good to about $10^{-7} \sim 10^{-6}$ au. In this sense, the exact

Table 2.2 Gradients (au) calculated for acetate CH_3COO^- using local effective ε=4.0 for the CH_3 group and ε=78.39 for the COO^- group.

Coordinates	Analytic	Numerical
C1X	0.00099739	0.00099745
C1Y	-0.00146286	-0.00146285
C1Z	0.00729240	0.00729245
C2X	-0.00114379	-0.00114410
C2Y	-0.00122523	-0.00122550
C2Z	0.00055169	0.00055125
O3X	-0.00763535	-0.00763505
O3Y	0.00002160	0.00002160
O3Z	-0.00117725	-0.00117720
O4X	0.00231535	0.00231525
O4Y	-0.00663241	-0.00663220
O4Z	-0.00270982	-0.00270960
H5X	-0.01020354	-0.01020405
H5Y	-0.01641703	-0.01641715
H5Z	0.00727599	0.00727605
H6X	-0.02057757	-0.02057760
H6Y	0.02671368	0.02671380
H6Z	-0.01770558	-0.01770560
H7X	0.03624751	0.03624785
H7Y	-0.00099775	-0.00099775
H7Z	0.00647257	0.00647260
Max error		0.00000044
RMS error		0.00000022

gradients are obtained. It is noted that such exact gradients are partially due to the use of the FIXPVA tessellation scheme developed by the Su *et al.*[16] Different tessellation schemes may lead to different solvation energies, total molecular energies, and molecular potential energy surfaces. For example, compared to GEPOL-GB (with no additional spheres), the FIXPVA scheme produces ~10% less surface area and ~1 kcal/mol smaller CPCM solvation energies for some typical solutes such as acetate anion.[16] Such small differences in solvation energies have also been observed for the current heterogeneous CPCM method. The expression of the analytic gradients for the heterogeneous CPCM method [i.e. Eq. (2-30)] is general, and can be implemented by using different tessellation

schemes such as GEPOL-GB as long as the derivatives of the tessera areas and coordinates are available.

2.5 Summary

A heterogeneous conductor-like solvation model that uses different local effective dielectrics for different portions of a solute cavity surface was derived and implemented in the conductor-like polarizable continuum model (CPCM) code in GAMESS. By variationally formulating the solvation operators for Hartree-Fock and Kohn-Sham methods, using the FIXPVA tessellation scheme, and assuming that the local effective dielectrics are constants in a geometry optimization process, continuous and smooth potential energy surfaces and analytic nuclear gradients (accurate to $10^{-7} \sim 10^{-6}$ au) have been obtained for the heterogeneous solvation model. Application of the heterogeneous solvation model to realistic quantum models consisting of ~100 atoms for the type-1 Cu center will be presented in the next chapter.

Chapter 3 Quantum chemical calculation of type-1 Cu reduction potential: ligand interaction and solvation effect

3.1 Introduction

This chapter presents a realistic example, the reduction potential calculation of the type-1 Cu centers, to illustrate the application of the heterogeneous conductor-like polarizable continuum model (Het-CPCM)[21] methods introduced in Chapter 2. Redox active proteins and enzymes containing transition metal ions are essential electron transfer components in biological systems. The reduction potential E^0 is one of the most important quantities that characterize the redox behavior of a metalloprotein. Understanding the structural factors that determine E^0 is of fundamental importance for understanding biological redox chemistry.

Quantum chemical E^0 calculation for metalloproteins remains a highly challenging task as it requires accurate descriptions of the metal-ligand interactions and protein matrix/aqueous solvation to the metal centers. Various methods have been used in the literature. For example, density functional methods have been used to calculate the E^0 of various small iron-sulfur clusters derived from proteins.[77-79] In quantum-electrostatic methods,[78,80-83] the atomic charges derived from quantum chemical calculations are fixed and used in subsequent electrostatic calculations to study the protein modulations on E^0. In combined quantum mechanical and molecular mechanical methods (QM/MM) protein and solvent interactions are usually incorporated into the quantum chemical calculations of the metal centers as a reaction field and/or a force field.[77,84-87]

In principle, protein matrix should be described explicitly with structural details such as electrostatic monopoles, dipoles and polarizabilities. Physical insights into solvation can only be obtained through MM or QM/MM simulations. Continuum models contain no specific interaction terms, and can only empirically predict solvation free energies by using parameterized dielectric constants and molecular cavity sizes. Some authors have successfully parameterized and applied continuum models to describe the protein matrix and bulk aqueous solvation of protein active sites.[15,17,77,83] The protein pK_a calculations show that if the active site model is relatively large (e.g. ~100 atoms), the protein matrix solvation free energy difference over a series of similar active sites can often be well reproduced by a continuum solvation model.[17] In this study, the possibility of using continuum model to describe protein matrix and aqueous solvation in reduction potential calculation is further explored.

Due to the heterogeneous nature of the environment around an active site, the continuum solvation model should also be heterogeneous. Though the idea of using heterogeneous continuum model has a long history, Tapia is probably the first one who developed and applied a quantum chemical heterogeneous continuum model.[88] Using heterogeneous solvation models, Tomasi's group studied the energy changes in deformations of a long DNA fragments and that in the opening of a DNA double helix,[60] the partial solvation effect in molecular recognition and docking,[61] and polar solutes placed near the surface of two immiscible liquids or at a liquid/vacuum separation.[89] Hoshi, Sakurai, Inoue and Chujo extended Tomasi's dielectric polarizable continuum model (DPCM) to treat anisotropic polarization effects in their guest-host complex calculations.[63] Li, Nelson, Peng, Bashford and Noodleman developed and applied a

heterogeneous continuum dielectric model to calculate the reduction potentials of 2Fe2S clusters in ferredoxin and phthalate dioxygenase reductase.[77] By using different local effective dielectrics for different portions of the solute cavity surface, Iozzi, Cossi, Improta, Rega and Barone further extended IEF-PCM[13,14] to study the pK_a of a solvent exposed histidine residue in prion protein and a small molecule interacting with a biological membrane.[15] Mikkelsen and coworkers developed heterogeneous solvation models for studying excited electronic states and optical properties.[90] In addition, the generalized Born model has been extended to treat heterogeneous environments occurred in force field simulation of biological systems.[91]

In this chapter, a heterogeneous conductor-like polarizable continuum model (Het-CPCM)[21] is incorporated in the B3LYP[76,92,93] as a reaction field, and is applied to study the E^0 of five type-1 Cu centers in cucumber stellacyanin, fern dryopteris crassirhizoma (D.c.) plastocyanin, Met148Gln, Met148Leu and wild type *thiobacillus ferrooxidans* (T.f.) rusticyanin, which show a 540 mV range in E^0. Large model molecules consisting of ~100 atoms are extracted from X-ray structures and ~70 atoms are geometrically optimized. A careful examination on the calculated E^0 for 13 plastocyanins (all show E^0~370 mV) suggests that using ~100 atoms and optimizing ~70 atoms can reduce the errors caused by the structural differences in the X-ray structures to below 100 mV. Large triple-zeta basis sets with polarizable and diffuse functions are used so the Cu-ligand interactions, especially polarization, are adequately modeled. Due to the use of the Het-CPCM method, protein matrix and aqueous solvation to the model molecules is treated in an efficient way.

Table 3.1 Thirteen plastocyanin X-ray structures.

PDB	Species	O.S.	E^0 (mV)	E^0 Ref
1KDI[94]	Fern dryopteris crassirhizoma (Adiantum capillus-veneris)	Cu^+	376	95
2BZ7[95]	Fern dryopteris crassirhizoma (Adiantum capillus-veneris), G36P	Cu^{2+}	363	95
2BZC[95]	Fern dryopteris crassirhizoma (Adiantum capillus-veneris), G36P	Cu^+	363	95
5PCY[96]	Poplar (populus nigra)	Cu^+	370	97
1PNC[98]	Poplar (populus nigra)	Cu^{2+}	370	97
1JXG[99]	Poplar (populus nigra), I21C, E25C	Cu^{2+}	348	99
2CJ3[100]	Cyanobacterium (Anabaena variabilis)	Cu^{2+}	360	101
1BXU[102]	Cyanobacterium (Synechococcus sp.) pcc 7942	Cu^{2+}	370	a
1PCS[103]	Cyanobacterium (Synechocystis sp.) pcc 6803, A42D, D47P, A63L	Cu^{2+}	325	103
7PCY[104]	Green alga (Enteromorpha prolifera)	Cu^{2+}	369	97
2PLT[105]	Green alga (Chlamydomonas reinhardtii)	Cu^{2+}	370	a
1AG6[106]	Spinach (Spinacia oleracea), G8D	Cu^{2+}	379	106
1IUZ[107]	Sea lettuce (Ulva pertusa)	Cu^{2+}	363	108

a. The E^0 was not found in the literature and a value of 370 mV is estimated.

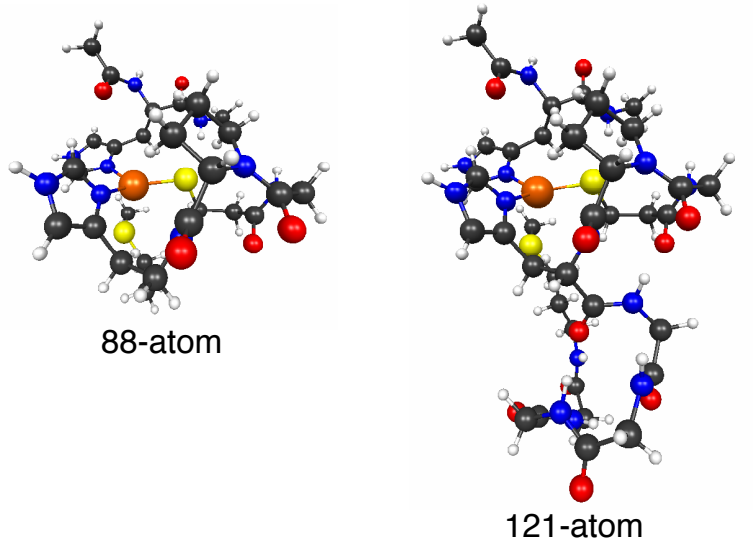


Figure 3.1 88-atom and 121-atom model molecules extracted from the X-ray structure 1KDI (H atoms are added).

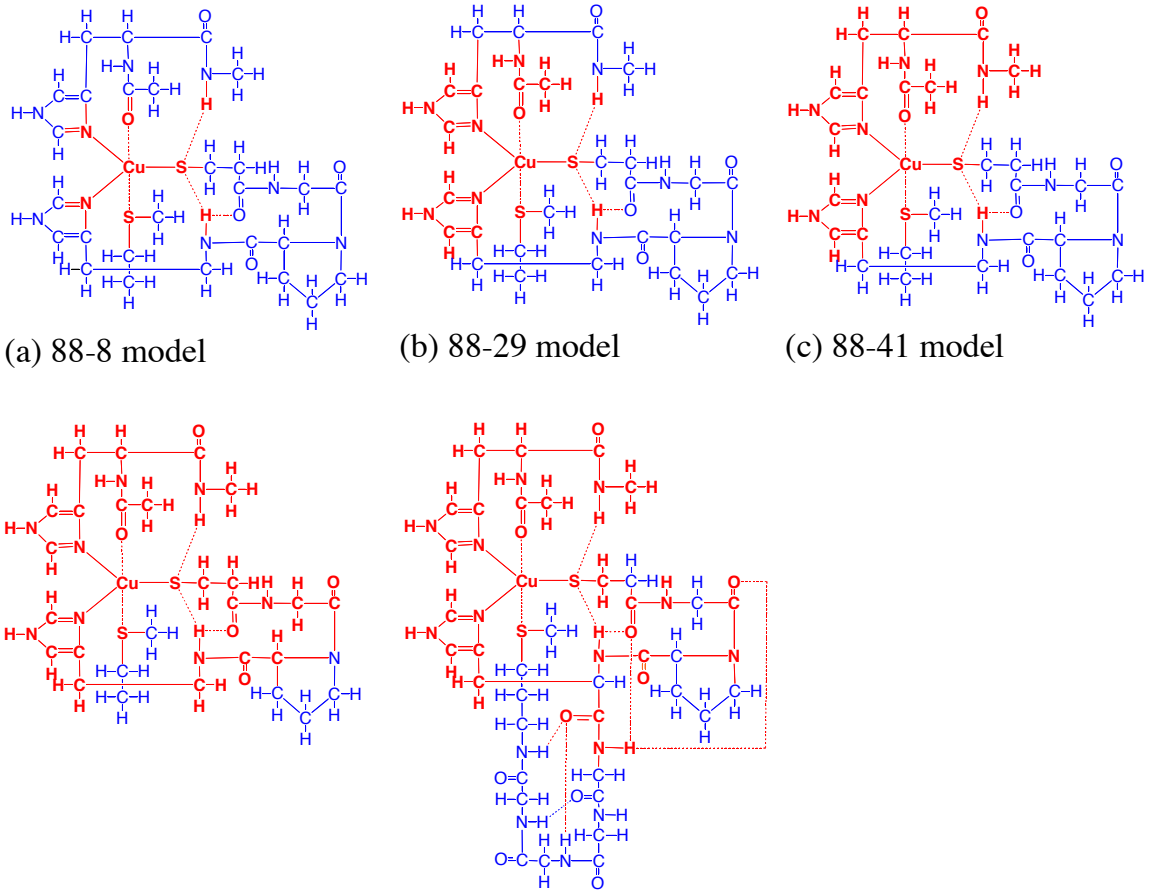
3.2 Computational methodology

All electronic structure calculations were performed with the GAMESS program.[68]

The X-ray structures were obtained from the protein data bank (PDB).[109] Hydrogen atoms were added to the structures using the WHAT IF web interface.[110] Active site model molecules (Figures 3.1, 3.2, 3.3 and 3.4) were extracted from these X-ray structures and edited by manually deleting unwanted atoms and adding new hydrogen atoms to fill the open bonds.

To determine the minimum model size and the minimum number of atoms to be optimized, active site model molecules consisting of 88 atoms and 121 atoms were extracted from 13 X-ray structures for 11 plastocyanins (Table 3.1, Figure 3.1). Plastocyanin is a typical small type-1 Cu protein consisting of ~100 amino acid residues. Despite of the differences in the amino acid sequences, the overall 3D folding of all the 11 plastocyanins considered in this study are similar. The similarity of their type-1 Cu centers is even higher: they possess identical ligands and exhibit almost the same local 3D structures and very similar E^0 (all around 370 mV), implying that the protein matrices have very little effect on the E^0. Chemically identical active sites consisting of up to 130 atoms for these type-1 Cu centers can be isolated from these plastocyanins. This allows one to focus on the local structural factors and short-range interactions without necessarily considering the long-range interactions from the rest of the protein. Geometry optimizations were performed in the gas phase with RHF[9] and ROHF[111] methods, respectively, for the Cu^+ and Cu^{2+} oxidation states of these model molecules. The 6-31G* basis set[112] was used. For the 88-atom models, the coordinates of 0, 8, 29, 41 and 67

atoms were optimized; for the 121-atom models, the coordinates of 0 and 61 atoms were optimized. These model molecules are denoted as 88-0, 88-8, 88-29, 88-41, 88-67, 121-0 and 121-61 models (Table 3.2, Figure 3.2). In Figure 3.2, red atoms are optimized while blue atoms are fixed in their X-ray coordinates.



(a) 88-8 model (b) 88-29 model (c) 88-41 model

(d) 88-67 model (e) 121-61 model

Figure 3.2 88-atom and 121-atom model molecules for plastocyanins. Red atoms are optimized while blue atoms are fixed.

Table 3.2 E^0 (mV) calculated with the gas phase RHF/6-31G* and ROHF/6-31G* methods for plastocyanin using 13 X-ray structures.

Model	88-0	88-8	88-29	88-41	88-67	121-0	121-61	Exp
1KDI	376	376	376	376	376	376	376	376
7PCY	322	308	317	306	414	315	365	369
5PCY	619	216	382	336	409	658	413	370
1AG6	-24	-12	202	220	389	-25	327	379
1PCS	-18	-94	151	171	403	-35	280	325
1PNC	138	197	226	231	407	154	336	370
2BZC	393	271	421	395	402	416	433	363
2BZ7	380	232	398	371	411	372	378	363
1IUZ	288	263				315	349	363
2CJ3	193	184				274	449	360
1JXG	103	87				146	373	348
2PLT	54	18				79	323	370
1BXU	28	-50				66	369	370
Max unsigned error	403	420	177	159	78	404	89	
Mean unsigned error[a]	206	228	93	84	42	193	37	

a. 1KDI excluded.

In the calculation of E^0 for type-1 Cu centers in Cucumber stellacyanin, D.c. plastocyanin, M148Q T.f. rusticyanin, T.f. rusticyanin and M148L T.f. rusticyanin, active site model molecules consisting of ~100 atoms were extracted from the X-ray structures 1JER,[113] 1KDI,[94] 1E30,[114] 2CAK[115] and 1GY2[116] (Figures 3.3 and 3.4). For each model molecule, the coordinates of ~70 atoms were optimized with the CPCM[21,57,117]/B3LYP[76]/6-31G* method, with the R-B3LYP and RO-B3LYP type of wavefunction for Cu^+ and Cu^{2+}, respectively. In order to mimic the forces imposed by the protein, ~30 atoms were fixed in their X-ray coordinates. In Figure 3.4, red atoms are optimized while blue atoms are fixed, and the numbers of total and optimized atoms are indicated after the PDB names.

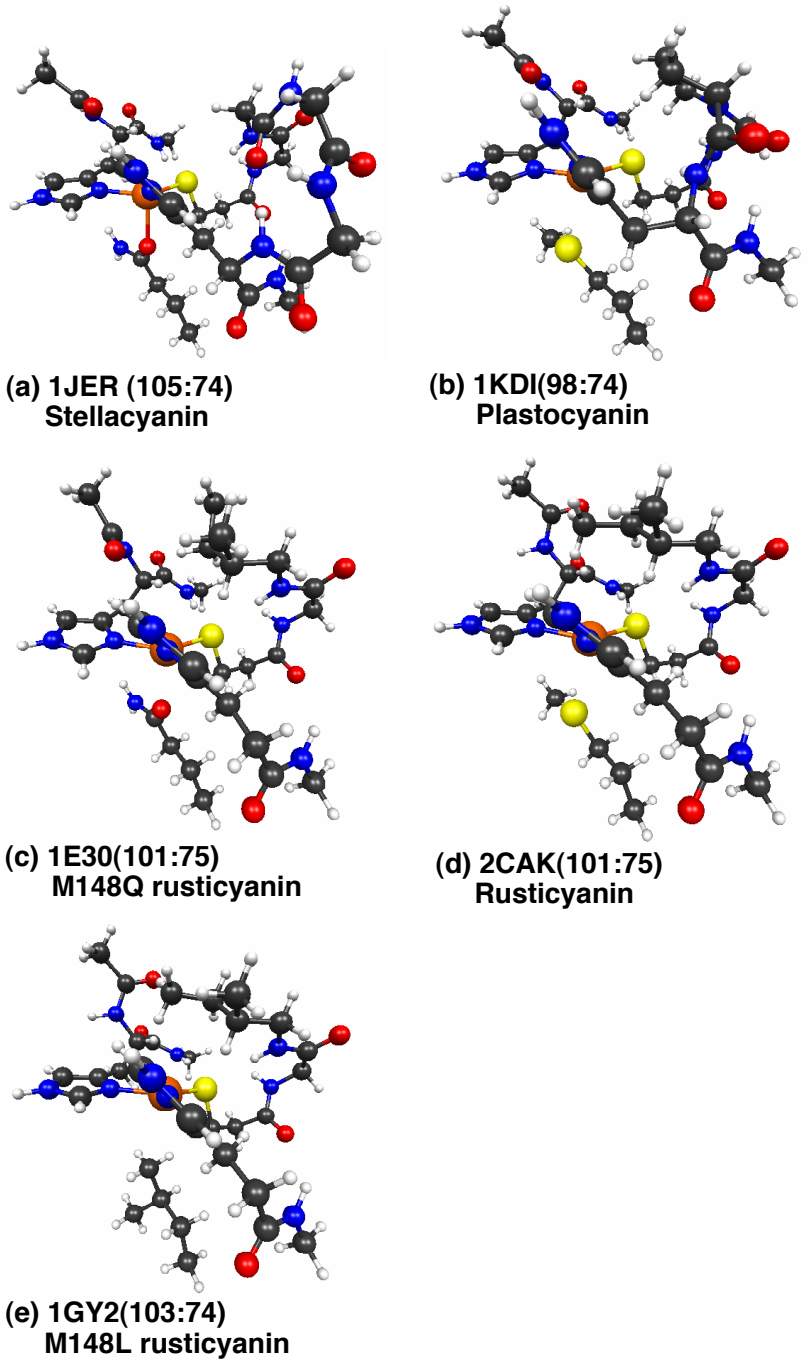


Figure 3.3 Model molecules of five type-1 Cu centers extracted from X-ray crystal structures. The numbers of total and optimized atoms are indicated in the parenthesis.

Two CPCM methods were used. One is the usual homogeneous CPCM (Homo-CPCM,[57,117] an approximation of the more rigorous and accurate IEF-PCM[13,14]) in which only one dielectric constant is used to describe a homogeneous and isotropic solvent, such as bulk water. The other is the heterogeneous CPCM (Het-CPCM[21] described in Chapter 2) in which different local effective dielectric constants can be defined for different surface regions to represent a heterogeneous environment, such as an active site solvated by protein matrix and bulk water. Practically, this is realized by defining different effective dielectric constants for different spheres used to form the solute cavity. In the current implementation, the CPCM boundary elements or tesserae on the same sphere have the same dielectric constant. The Het-CPCM has been implemented for energy and analytic gradient calculation. Both the Homo- and Het-CPCM calculations (energy and gradients) are very efficient, and only add a few percent of computing time to the corresponding gas phase calculations for a molecule consisting of ~100 atoms. The details of the heterogeneous can be found in reference.[21]

In both the Homo-CPCM and Het-CPCM methods, spheres with radii of 2.124, 2.016, 1.908, 2.52 and 2.76 Å were used for C, N, O, S and Cu atoms, respectively, to define the molecular cavity. The FIXPVA tessellation scheme[16] was used with 60 initial tesserae per sphere. No charge renormalization was performed, and only the CPCM electrostatic interaction was calculated as the solvation free energy.

In the Homo-CPCM method, a dielectric constant of 78.39 was used. In the Het-CPCM calculations, spheres associated with solvent-exposed atoms are assigned with a dielectric constant of 78.39. His143 is the solvent-exposed group in the 1E30, 2CAK and 1GY2 model molecules; Most of the atoms in the 1JER and 1KDI model molecules are

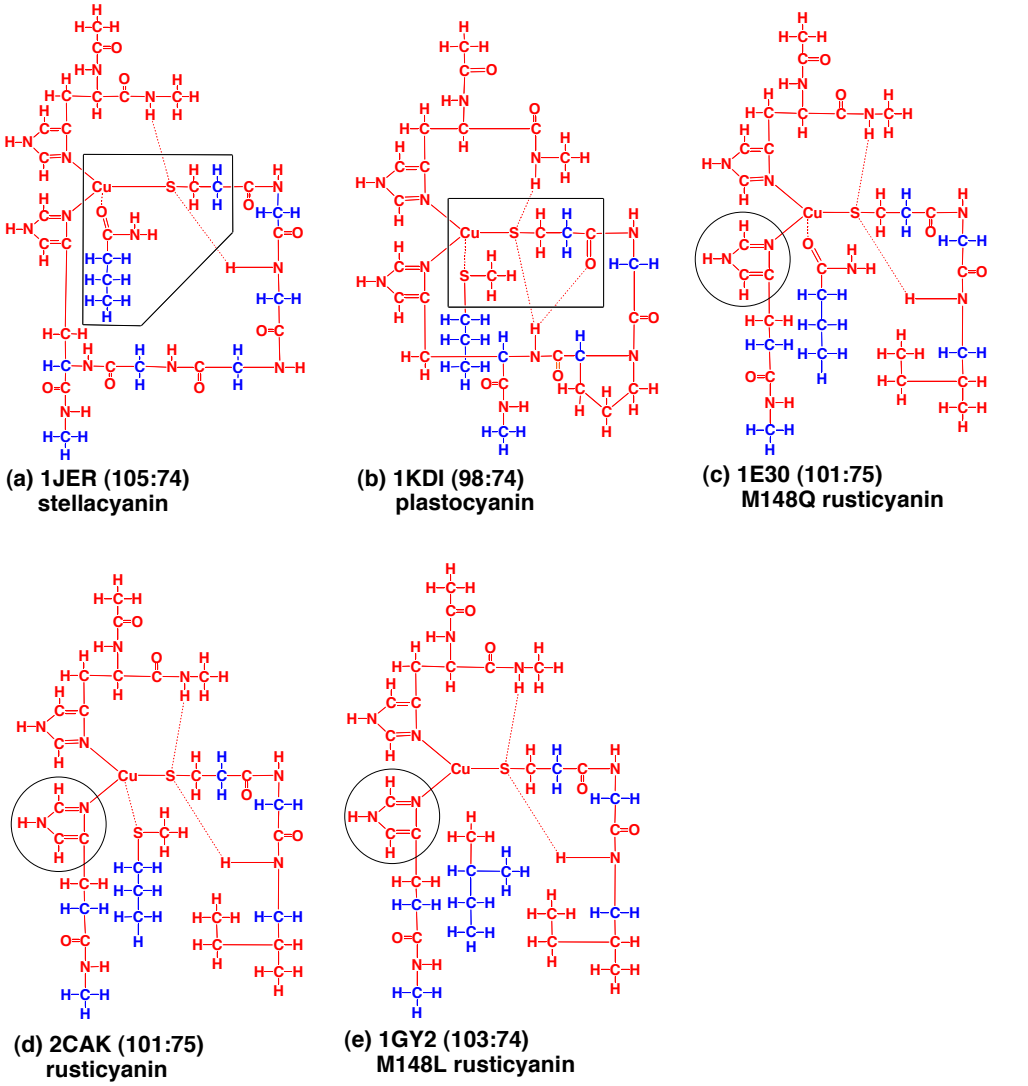


Figure 3.4 Model molecules of five type-1 Cu centers extracted from X-ray crystal structures. Red atoms are optimized while blue atoms are fixed (the numbers of total and optimized atoms are indicated after the PDB names). In Het-CPCM calculations, spheres associated with solvent-exposed atoms (non-cycled atoms for 1JER and 1KDI, cycled atoms for 1E30, 2CAK and 1GY2) are assigned with $\varepsilon = 78.39$ while the spheres associated with protein-buried atoms are assigned with smaller dielectric constants: $\varepsilon = 20$ for 1JER and 1KDI, $\varepsilon = 4$ for others.

solvent-exposed, with 23 and 15 atoms, respectively, being buried (Figure 3.4). The spheres associated with atoms embedded in the protein are assigned with lower dielectric constants: 4 for rusticyanin and 20 for stellacyanin and plastocyanin.

Using dielectric constants between 4 and 20 is reasonable because they have been commonly used for proteins. Experiments[118] show that the dielectric constants of "dry protein powders" are ~1.2, and those of "water-adsorbed protein powders" are typically 2~5, depending on the protein and the weight fraction of adsorbed water. Therefore, many continuum electrostatic studies used 2~4 for protein interiors. However, sophisticated theoretical studies performed by Warshel and coworkers (for example, King *et. al.*[119]) suggest that the dielectric constants of protein interiors are dependent on the bulk solvent, and some protein active sites show higher values such as 10 when the solvent reaction field is considered. Larger values such as 20 could also exist due to solvent effects. In continuum electrostatics pK_a calculations, which are most relevant to the E^0 calculations in the current study, Antosiewicz *et al.* found that the best overall results can be obtained if $\varepsilon = 20$ is used.[120] They also noted that protein interiors should have dielectric constants smaller than 20. Therefore, protein interior dielectric constants are most likely 4~10, and can be as large as 20.

An inspection of the protein environments around the type-1 Cu centers shows that the degree of burial of the type-1 Cu centers is lower for cucumber stellacyanin (1JER) and D.c. plastocyanin (1KDI) as compared to that for rusticyanin. A previous empirical study on protein pK_a (PROPKA[121]) has found that the number of protein C, N, O atoms within ~15.5 Å to a pK_a site can be used to describe the desolvation effect on pK_a shift. In principle, such numbers can also be used to estimate the desolvation effect

on E^0. The numbers of protein C, N, O atoms within 15.5 Å to the Cu ions in the PDB files are 453, 445, 536, 512 and 515, respectively, for 1JER, 1KDI, 1E30, 2CAK and 1GY2. Clearly, the Cu ions in cucumber stellacyanin (1JER) and *D.c.* plastocyanin (1KDI) are less buried (i.e., more solvated) and those in rusticyanin (1E30, 2CAK and 1GY2) are more buried (i.e. less solvated). A molecular dynamics simulation suggests that the protein matrix around the rusticyanin type-1 Cu center is highly hydrophobic and rigid, which corresponds to a low effective dielectric constant.[122] To minimize the arbitrariness, the high-end value $\varepsilon = 20$ was used for 1JER and 1KDI, and the low-end value $\varepsilon = 4$ was used for 1E30, 2CAK and 1GY2. One must keep in mind that the accurate effective dielectric is unknown, and could be significantly different at different portions around the type-1 centers. In general, determining protein interior dielectric is a difficult issue.[83]

Based on the Homo- and Het-CPCM/B3LYP/6-31G* optimized structures, single point energies were calculated using Homo- and Het-CPCM/B3LYP methods with a mixed triple-zeta basis set: the standard 6-311++G(2df,p)[123] for H, C, N, O and S, and cc-pVTZ[75] for Cu. Using such a large basis set is necessary for accurate description of the Cu-ligand interactions in which polarization has significant contributions (see Section 3.3 for discussions). Using the mixed triple-zeta basis set means ~2300 basis functions for each model molecule. In the CPCM calculations with these basis sets, approximately 0.05~0.06 e of electronic charge is distributed outside of the CPCM cavity. Since it is almost a constant in all the model molecules, such charge leaking is unlikely to cause significant differences in the calculated E^0.

Relative method is used to calculate E^0. For a given type-1 Cu protein (Pro) and a reference type-1 Cu protein (Ref) the free energy change for the following electron transfer reaction:

$$\text{ProCu(II)} + \text{RefCu(I)} \xrightarrow{\Delta G} \text{ProCu(I)} + \text{RefCu(II)} \quad (3\text{-}1)$$

is approximated with the electronic energy (including nuclear repulsion and solvation free energy) computed for the model molecules:

$$\Delta G \approx \Delta G^{ele} = \Delta G^{ele}_{\text{ProCu(I)}} + \Delta G^{ele}_{\text{RefCu(II)}} - \Delta G^{ele}_{\text{ProCu(II)}} - \Delta G^{ele}_{\text{RefCu(I)}} \quad (3\text{-}2)$$

The relative method is based on the assumption that the differences in the zero-point-energies, thermal energies and entropies of the model molecules make minor contributions to the relative E^0. This is indeed a good approximation when a series of similar protein active sites are considered. The free energy contributions due to the protein matrix and aqueous solvent are included in the CPCM solvation free energy.

The E^0 of the protein at T=298.15 K is computed by:

$$\mathrm{E}^0_{\text{Pro}} = \mathrm{E}^0_{\text{Ref}} + \Delta G / F \quad (3\text{-}3)$$

where F is the Faraday constant, $\mathrm{E}^0_{\text{Ref}}$ is the experimental E^0 of the reference type-1 Cu protein (relative to standard hydrogen electrode). For the 11 plastocyanins, the fern plastocyanin (1KDI model) is used as the reference. For the five type-1 Cu centers, rusticyanin (2CAK model) is used as the reference.

3.3 Results and discussion

3.3.1 Structural sensitivity

Table 3.2 presents the E^0 calculated with RHF/6-31G* and ROHF/6-31G* methods for the 13 X-ray structures. The experimental E^0 of the plastocyanins are also listed for comparison. Fern plastocyanin[95] (X-ray structure 1KDI[94] in PDB) which has E^0

= 376 mV relative to standard hydrogen electrode, is used as the reference. The E^0 for *cyanobacterium pcc 7942* (1BXU) and green alga (*chlamydomonas reinhardtii*) plastocyanin (2PLT) were not found from the literature and 370 mV were estimated.

With no geometry optimization (88-0 and 121-0 models) or very limited geometry optimization (88-8 models), the relative E^0 calculated with the gas phase RHF/6-31G* and ROHF/6-31G* energies show maximum errors of ~400 mV and mean errors of ~200 mV. Ongoing from the 88-8 models to the 88-29 models, the E^0 are significantly improved, resulting in a maximum unsigned error = 177 mV and a mean unsigned error = 93 mV. Very small changes are observed for all the eight cases ongoing from 88-29 models to 88-41 models: the errors for 7PCY, 5PCY, 2BZC and 2BZ7 are still small (< 63 mV) and the errors for 1AG6, 1PCS and 1PNC are still large (>139 mV), leaving a maximum error = 159 mV and a mean error = 84 mV. Although the calculated E^0 are seemingly good (max error = 78 mV and a mean error = 42 mV), the 88-67 models have intrinsic problems: so many atoms are relaxed that the structures are significantly different from the experimental X-ray structures. Therefore, the 88-67 models are not good models for the proteins, and should not be used. However, they suggest that ~60 or more atoms should be optimized in order to reduce the E^0 errors caused by the structural errors to below 100 mV.

In order to relax the type-1 Cu centers while keeping the structures of the model molecules similar to the protein structures, model molecules with more than 88 atoms shall be used to include more protein interactions and constraints. The 121-atoms model molecules are therefore studied. Figure 3.2(e) shows the 61 atoms optimized in the 121-61 models. As expected, a max error = 89 mV and a mean error = 37 mV are obtained.

To summarize, the calculated relative E^0 are very sensitive to the geometry optimization of the model molecules. Therefore, relatively large (e.g. 100 atoms) model molecules should be used and a sufficient number (e.g. 60) of atoms should be optimized. This guideline is used in the following studies of five type-1 Cu centers.

3.3.2 Cu-ligand distances

The Cu-ligand distances in the Het-CPCM/B3LYP/6-31G* optimized model molecules for cucumber stellacyanin, D.c. plastocyanin, Met148Gln, Met148Leu and wild-type T.f. rusticyanin are listed in Table 3.3. The Homo-CPCM/B3LYP/6-31G* results are very similar, thus not listed.

Table 3.3 Het-CPCM/B3LYP/6-31G* optimized Cu-ligand distances (Å) in the model molecules.

Species	PDB	Cu-S(Cys)		Cu-N(His1)[a]		Cu-N(His2)[b]		Cu-axial[c]	
		Cu^{2+}	Cu^{+}	Cu^{2+}	Cu^{+}	Cu^{2+}	Cu^{+}	Cu^{2+}	Cu^{+}
Cucumber stellacyanin	1JER	2.221	2.288	1.965	2.018	1.947	1.970	2.097	2.181
D.c. plastocyanin	1KDI	2.236	2.290	1.975	1.982	1.964	2.011	2.467	2.613
M148Q T.f. rusticyanin	1E30	2.218	2.285	1.969	1.979	1.943	2.003	2.114	2.204
T.f. rusticyanin	2CAK	2.217	2.326	1.993	2.001	1.962	2.008	2.536	2.499
M148L T.f. rusticyanin	1GY2	2.152	2.254	1.956	1.989	1.942	1.970	2.924	3.012

a. N-terminal His

b. C-terminal His

c. Cu-S(Met) in 1KDI and 2CAK, Cu-O(Gln) in 1JER and 1E30, Cu-C(Leu) in 1GY2.

The Cu-S^-(Cys) and Cu-N(His) bonds are strong and rigid, and show small variations in different model molecules. The optimized Cu^{2+}-S^-(Cys) distances vary from 2.15 to 2.23 Å, and Cu^{+}-S^-(Cys) distances vary from 2.25 to 2.33 Å. The EXAFS data are 2.08 to 2.17 Å for Cu^{2+}-S^-(Cys) distances and 2.15 to 2.22 Å for Cu^{+}-S^-(Cys)

distances.[69,115,124,125] The Cu^{2+}-N(His) distances vary from 1.94 to 1.99 Å, with the C-terminal N(His) distances always shorter than the N-terminal N(His) distances by 0.01~0.03 Å. The Cu^{+}-N(His) distances vary from 1.97 to 2.02 Å, with the C-terminal N(His) distances either shorter or longer than the N-terminal N(His) distances by 0.01~0.05 Å. The EXAFS values are 1.94~2.05 Å for the oxidized forms and 1.92~2.02Å for the reduced forms.[69,115,124,125]

The Cu^{2+}-O(Gln) distances in 1JER and 1E30 models are 2.097 and 2.114 Å, while the Cu^{+}-O(Gln) distances are 2.181 and 2.204 Å, as compared to 2.21 and 2.33 Å in the X-ray structures. The Cu^{+}-O(Gln) distance in an M148Q mutant is 2.30 Å measured using EXAFS.[115]

The Cu^{2+}-S(Met) distances in 1KDI and 2CAK models are 2.467 and 2.536 Å, while the Cu^{+}-S(Met) distances are 2.613 and 2.499 Å, respectively, shorter than the 2.91~2.92 Å in the X-ray structures. The EXAFS data show that the Cu-S(Met) distances are 2.66~2.76 Å and 2.86~2.90 Å for the oxidized and reduced forms of plastocyanin.[69] The corresponding distances are 3.12 and 3.03 Å for the oxidized and reduced forms of T.f. rusticyanin.[115] Geometry optimization of smaller type-1 Cu model molecules performed using second-order perturbation theory method (MP2) gives similarly short distances [~2.6 Å for Cu^{2+}-S(Met) and ~2.3 Å for Cu^{+}-S(Met)], so it is unlikely that the short distances are due to errors in the B3LYP method. Instead, it is an intrinsic property of the $Cu^{2+/+}$- S(Met) bonds. This has been noticed and discussed in the literature for a long time, for example, by Ryde,[126] Solomon[69] and Ando.[127] Recent experimental results[128] from Solomon's group show that the axial Cu-S(Met) distance is not constrained in nitrite reductase so it can take two possible values, 2.4 Å and 4.3 Å. The

constant Cu-S(Met) distances of ~2.9 Å in plastocyanin and rusticyanin crystals are presumably due to interactions from other groups, or packing forces. As shown in Figure 3.4, many atoms of the Met ligand were fixed during geometry optimization but the resulted Cu-S(Met) distances are still too short in the 1KDI and 2CAK models. Therefore, such protein interactions are most likely intermolecular interactions instead of covalent bond forces. The details of these interactions are currently unknown. In subsection 3.3.4, the potential errors in the computed E^0 due to the shortened Cu-S(Met) distances will be discussed.

3.3.3 Solvation effect

Using the Het-CPCM/B3LYP/6-311G(2df,p) method, the calculated relative E^0 for cucumber stellacyanin, D.c. plastocyanin, Met148Gln, Met148Leu and wild-type T.f. rusticyanin are 242, 366, 522, 667 and 825 mV, respectively, compared well to experiment values 260, 376, 563, 667 and 798 mV (Table 3.4). The maximum unsigned error is 41 mV, with a mean unsigned error of 24 mV (2CAK excluded).

Table 3.4 Calculated reduction potential E^0 (mV) for five type-1 Cu centers.

Species	Homo/DZ[a]	Homo/TZ[b]	Het/DZ[c]	Het/TZ[d]	Experiment[e]
Cucumber stellacyanin	333	480	158	242	260
D.c. plastocyanin	554	616	367	366	376
M148Q T.f. rusticyanin	381	501	456	522	563
T.f. rusticyanin	667	667	667	667	667
M148L T.f. rusticyanin	843	849	832	825	798

a. Homo-CPCM/B3LYP/6-31G*

b. Homo-CPCM/B3LYP/6-311++G(2df,p)//Homo-CPCM/B3LYP/6-31G*

c. Het-CPCM/B3LYP/6-31G*

d. Het-CPCM/B3LYP/6-311++G(2df,p)//Het-CPCM/B3LYP/6-31G*

e. For stellacyanin and plastocyanin, pH~7; for rusticyanins, pH=3.2.

The experimental values[95,129] for stellacyanin and plastocyanin were measured at pH ~ 7 and the values[114] for rusticyanins were measured at pH = 3.2. Indeed, the E^0 of rusticyanins changes significantly ongoing from pH = 7 to pH =3.2.[114] According to Giudici-Orticoni *et al.*, the pK_a value of the N^{δ} proton of the solvent-exposed His143 imidazole is ~7 for Cu^{2+} rusticyanin.[130] The pK_a value of the N^{ε} proton of His143 imidazolium should be much lower than 2.[131] So, at pH~3, it is a neutral imidazole coordinating to the $Cu^{2+/1+}$ ions and the E^0 is ~680 mV, while at pH~7, it is an imidazolate anion coordinating to the $Cu^{2+/1+}$ ions and the E^0 is ~550 mV. Such a change in E^0 is apparently due to the charge-charge interaction between the proton and $Cu^{2+/+}$ ions. This study intended to calculate the E^0 at pH 2~3, the natural state of this protein, and used model molecules with neutral imidazoles for the three rusticyanins (Figures 3.3 and 3.4). Therefore, the computed E^0 should be compared to the experimental E^0 at pH ~3.

Using the Homo-CPCM/B3LYP/6-311++G(2df,p) method with ε=78.39, the calculated relative E^0 (480 and 616 mV) for cucumber stellacyanin and *D.c.* plastocyanin are too high as compared to experimental values 260 and 376 mV (at pH ~7).[95,129] This is not surprising because it is intrinsically wrong to use the same dielectric constant to describe the heterogeneous and different environments surrounding the model molecules.

The model molecules for stellacyanin (1JER) and Met148Gln rusticyanin (1E30) both have two H-bonds to S^-(Cys) and the same axial O(Gln) ligand, so they have almost the same E^0, 480 versus 501 mV when the Homo-CPCM method is used. The model molecules for plastocyanin (1KDI) and rusticyanin (2CAK) both have two H-bonds to S^-(Cys) and the same axial S(Met) ligand, so they have similar E^0, 616 versus 667 mV,

with a 51 mV difference mainly caused by the difference in the strength of the H-bonds to the S^-(Cys) ligands (see section 3.3.5).

Clearly, the solvation effects introduced by using ε = 4 for protein buried atoms in rusticyanins (1E30, 2CAK and 1GY2) and ε = 20 for protein buried atoms in stellacyanin (1JER) and plastocyanin (1KDI) can create ~250 mV differences in E^0.

How the bulk water and protein matrix solvation affects the E^0 of the type-1 Cu center can also be understood by examining the absolute E^0 computed for rusticyanin (2CAK model). Using the homogeneous CPCM(ε=78.39)/B3LYP/6-31G* optimized structure and energy, the energy difference is 81.29 kcal/mol or 3525 mV; Using the heterogeneous CPCM/B3LYP/6-31G* optimized structure and energy, the energy difference is 85.79 kcal/mol or 3720 mV. Therefore, in rusticyanin the protein burial can likely raise the reduction potential by ~200 mV. This value is similar to the ~250 mV difference between rusticyanin and stellacyanin or plastocyanin.

In the current model, arbitrariness has not been avoided due to the selection of the effective dielectric constants. Therefore, the results obtained with the Het-CPCM method should be regarded as a semi-quantitative estimation of the desolvation effects on E^0.

Permanent electrostatic interactions from the protein matrix are not considered in this model. A comparison to calculations using electrostatic models will be presented in subsection 3.3.6.

3.3.4 Axial ligands

It is well known that axial ligands can preferentially stabilize Cu^{2+} thus decrease the E^0. For example, the Met148Gln, native and Met148Leu rusticyanins have E^0 of 563, 667 and 798 mV, respectively.[114] The E^0 calculated with the Het-CPCM/B3LYP/6-

311++G(2df,p) method are 522, 667 and 825 mV, respectively, in good agreement with the experimental values 563, 667 and 798 mV (Table 3.4). Homo-CPCM/B3LYP/6-311++G(2df,p) method gives slightly worse results, 501, 667 and 849 mV, respectively (Table 3.4).

A very similar case is the Gln99Met, Gln99Leu and wild type stellacyanin, for which the S(Met) and O(Gln) ligands are able to decrease the E^0 by ~160 and ~320 mV, respectively, as compared to Leu.[125,129]

In the study of $Cu^{2+/+}$-ligand interaction using B3LYP, MP2, CCSD and CCSD(T) methods performed by Su *et al*.,[132] they found that for Cu^+-water, Cu^+-imidazole, Cu^+-$S^-(CH_3)$, Cu^+-$S(CH_3)_2$, all in equilibrium geometries, B3LYP tends to overestimate the interaction energies by ~2, ~4, ~7 and ~5 kcal/mol as compared to coupled cluster singles, doubles with noniterative triples [CCSD(T)] method; for Cu^{2+}-water, Cu^{2+}-imidazole, Cu^{2+}-$S^-(CH_3)$, Cu^{2+}-$S(CH_3)_2$, B3LYP tends to overestimate the interaction energies by ~11, ~18, ~24 and ~25 kcal/mol as compared to coupled cluster singles, doubles (CCSD) method. In the current study, the large errors in the $Cu^{2+/+}$-S^-(Cys) and $Cu^{2+/+}$-N(His) interactions are canceled because all the model molecules have similar Cu-S-N-N core structures. Although it is not clear how much B3LYP will overestimate the axial $Cu^{2+/+}$-O(Gln) and $Cu^{2+/+}$-S(Met) interactions in the model molecules when the Cu ions are already strongly coordinated by the equatorial S^-(Cys) and N(His) ligands, and a continuum solvation model is used, the magnitude of the overestimation should be very smaller, such as ~1 kcal/mol or ~40 mV. Indeed, the Het-CPCM/B3LYP/6-311++G(2df,p) predicted E^0 for Met148Gln and wild type Met148 rusticyanin (relative to Met148Leu rusticyanin) are ~41 and ~27 mV too low as compared to experimental

values. As discussed above, the Cu-S(Met) distance in the 2CAK model is ~2.5 Å as compared to ~2.9 Å in the X-ray structure 2CAK. This shortening may also contribute to the overestimation of the E^0 change ongoing from Met148Leu rusticyanin to wild type Met148 rusticyanin.

The standard 6-31G* basis set is insufficient for modeling electron density polarization, which is crucial in determining the relative coordination strength of Cu to the CH_3(Leu), S(Met) and O(Gln) ligands. For example, for the three rusticyanins, the relative E^0 calculated with the Het-CPCM/B3LYP/6-31G* are 456, 667 and 832 mV, respectively, worse than the 6-311++G(2df,p) results. Similar basis set effects can be seen for cucumber stellacyanin, for which the Het-CPCM/B3LYP calculated E^0 is improved by 84 mV ongoing from 6-31G* to 6-311++G(2df,p). It is also interesting to note that the basis set error is larger when the Homo-CPCM method is used. For example, the E^0 calculated for Met148Gln with the Homo- and Het-CPCM/B3LYP/6-31G* methods are 381 and 456 mV, respectively, as compared to the experimental value 563 mV (Table 3.4). It is well known that triple-zeta quality basis sets can usually converge B3LYP calculated relative energies to within ~1 kcal/mol. Therefore, larger basis sets were not attempted. Using GAMESS, open-shell B3LYP calculations for metal systems are difficult to converge when large basis sets are used.

It must be emphasized that the good agreement between the calculated and experimental axial ligand interactions is contingent on the Het-CPCM solvation effect. For example, the E^0 difference between native and Met148Leu rusticyanins computed with the gas phase B3LYP/6-311++G(2df,p) method is 248 mV, much larger than the 158 mV from the Het-CPCM/B3LYP/6-311++G(2df,p) and the 131 mV from

experiments. This observation suggests that using a continuum solvation model to describe the rusticyanin protein matrix is quite a good approximation. The gas phase data for Met148Gln rusticyanin are not available for comparison because the SCF calculation did not converge. A computational study on His143Met, Met148Gln and wild type rusticyanins show that gas phase results overestimate experimental mutagenesis E^0 changes, while QM/MM methods tend to slightly underestimate.[115]

3.3.5 Hydrogen bonding to S^-(Cys)

Hydrogen bonding to the Cu bound S^-(Cys) ligand can raise the E^0. Experimental mutagenesis of A.f. pseudoazurin show that the Pro80Ala and Pro80Ile variants have one more backbone hydrogen bond to the copper bound S^-(Cys) than the wild type does and show 139 and 180 mV higher E^0, respectively.[81] Similarly, the Pro94Ala, and Pro94Phe mutants of P.d. amicyanin have higher E^0 than the wild type-115 and 150 mV, respectively, due to the creation of a new hydrogen bond to the copper bound S^-(Cys).[133]

The 51 mV difference in the Homo-CPCM/B3LYP/6-311++G(2df,p) calculated E^0 for plastocyanin (1KDI, 616 mV) and rusticyanin (2CAK, 667 mV, the reference) is likely caused by the second hydrogen bond to the S^-(Cys) ligands, which is weaker in 1KDI but stronger in 2CAK. In the optimized oxidized form of the 2CAK model molecule, there are two similar backbone amide hydrogen bonds to the S^-(Cys), with S-H distances of 2.64 and 2.66 Å and S-H-N angles of 168 and 169 degrees, respectively. In the optimized oxidized form of the 1KDI model molecule, the two S-H distances are 2.51 and 2.84 Å, and the two S-H-N angles are 177 and 152 degrees, respectively, suggesting that there are two hydrogen bonds, one is strong and one is weak. This effect cannot be obviously seen in the Het-CPCM results because these two models used very different

dielectric constants for protein buried regions, and the 1KDI model gains an additional E^0 lowering of 250 mV due to the solvation effects.

3.3.6 Comparison to previous calculations

A few computational studies on calculating E^0 of type-1 Cu centers are in the literature. A comparison of the current work with them is presented below.

Olsson and Ryde[84] calculated relative E^0 for type-1 Cu centers in stellacyanin, plastocyanin, azurin, rusticyanin, and ceruloplasmin with B3LYP and the PCM method and model molecules consisting of ~50 atoms. Their main purpose is to elucidate the influence of the axial ligands such as Met and Gln on E^0. They performed full and constrained geometry optimizations and concluded that axial ligand interactions can affect the E^0, similar to the results from this chapter. Compared to their method, this study used larger and more realistic model molecules so some protein interactions are included. In addition, protein matrix and bulk solvation to the active sites were modeled more accurately with the Het-CPCM method while Olsson and Ryde only used Homo-PCM to estimate aqueous solvent effects.

Botuyan *et al.*[82] calculated the relative reduction potentials for French bean plastocyanin and rusticyanin using a continuum electrostatic model. The E^0 differences obtained with NMR structures and X-ray structures are 228 mV and 389 mV, respectively. Their results suggest that the E^0 differences is caused by both hydrophobicity in rusticyanin and some specific charge-charge and charge-dipole interactions. Jimenez *et al.*[122] performed molecular dynamics simulations for rusticyanin and concluded that its high hydrophobicity and rigidity is responsible for its high E^0. Using protein dipole/Langevin dipole and QM/MM frozen density functional free energy

simulation techniques, Olsson, Hong and Warshel[85] predicted similar values for poplar plastocyanin and rusticyanin (300 mV), and suggested that the E^0 differences between plastocyanin and rusticyanin is caused by many small protein dipole interactions with the Cu ions. In this chapter, permanent electrostatic interactions from the protein matrix are not considered. If the hydrophobicity in rusticyanin is not considered, the calculated E^0 for plastocyanin and rusticyanin are similar, 616 versus 667 mV, as shown in Table 3.4. If the hydrophobicity is considered, the calculated E^0 are 366 versus 667 mV (Table 3.4). Therefore, the current study suggests that hydrophobicity is the main cause of the high E^0 of rusticyanin, in accordance with Jimenez *et al.*'s results[122] and in partial agreement with Botuyan *et al.*'s results.[82]

Li *et al.*[86] studied the structural determinants of the E^0 for six type-1 Cu proteins, cucumber stellacyanin, *P.a.* azurin, poplar plastocyanin, *C.c.* laccase, *T.f.* rusticyanin and human ceruloplasmin. Chemical models consisting of ~100 atoms for the type-1 Cu centers were extracted from X-ray structures. Two major structural determinants, Cu ligands and hydrogen bonds to the Cu bound S_{Cys}, were examined by comparing the E^0 of successively simpler models. However, the effect of structure relaxation of the active sites was not fully examined (only a few atoms were optimized) and considerably large errors were in the calculated relative E^0 for different species. In addition, only the 6-31G* basis set was used at that time due to the limit of computing power. Considering geometry relaxation, calculations suggest that solvation effect must be considered in order to explain the large E^0 range of type-1 Cu centers in different species.[134] The major improvements in the current study are the use of Het-CPCM, which is shown to be more

realistic and accurate, and the use of B3LYP method and the 6-311++G(2df,p) basis set, which are more accurate than the HF/6-31G* method.

3.4 Summary

The prediction of reduction potential for proteins is a long-standing issue. Results of this study show that type-1 Cu reduction potential (E^0) calculated with quantum chemical methods are very sensitive to the structures of the model molecules. In order to minimize the errors caused by the differences in the X-ray structures, relatively large (e.g. 100 atoms) model molecules should be used and a sufficient number of atoms (e.g. 60) should be optimized. Using model molecules consisting of ~100 atoms, the E^0 of five type-1 Cu centers in cucumber stellacyanin, D.c. plastocyanin, Met148Gln rusticyanin, wild type rusticyanin and Met148Leu rusticyanin were calculated with a heterogeneous conductor-like polarizable continuum model (Het-CPCM) incorporated in the B3LYP method as a reaction field. The Het-CPCM/B3LYP/6-311++G(2df,p) [cc-pVTZ for Cu] calculated E^0 are 242, 366, 522, 667 and 825 mV, respectively, in good agreement with experimental values 260, 376, 563, 667 and 798 mV (Table 3.4). The very high E^0 (798 mV) for Met148Leu rusticyanin (1GY2) is mainly due to the lack of the axial ligand, and two hydrogen bonds to the Cu bound S^-(Cys) ligand, as well as the hydrophobic and rigid environment around the type-1 Cu center. Compared to Met148Leu rusticyanin, wild type rusticyanin (2CAK) has an axial S(Met) ligand, which brings the E^0 down to 667 mV. The axial O(Gln) ligand in Met148Gln rusticyanin (1E30) is the main cause of its 563 mV E^0. The type-1 Cu center in stellacyanin (1JER) is very similar to that in Met148Gln rusticyanin, but is much more solvated by the aqueous solvent. The type-1 Cu center in plastocyanin (1KDI) is similar to that in wild type rusticyanin, but with a

slightly weaker hydrogen bond to the S^-(Cys) ligand and is much more solvated by the aqueous solvent. According to the Het-CPCM calculations, the difference in the solvation energy can likely create a difference of ~250 mV in the E^0. Mainly due to these reasons, stellacyanin and plastocyanin show much lower E^0 values, 260 and 376 mV. Therefore, ligand interaction (~250 mV) and solvation effect (~250 mV) are the main determinants of the relative E^0 of these five type-1 Cu centers. These factors have been proposed by Malmström, Solomon and Gray.[97,135]

Using QM/CPCM to predict the reduction potential showed both pros and cons. In the future work, the QuanPol program (developed by Li *et al*.,[136] released in August, 2011) can be used to perform QM/MM/Continuum style calculations for Type-1 Cu proteins to gain further insight into this problem.

Chapter 4 Analytic energy gradients in combined second order Møller-Plesset perturbation theory and conductor-like polarizable continuum model calculation

4.1 Introduction

This chapter focuses on second order Møller-Plesset perturbation theory (MP2) methods, including the spin-restricted closed shell RMP2, spin-unrestricted open shell UMP2, and the Z-averaged spin-restricted open shell ZAPT2 methods.[137,138] The MP2 methods are accurate quantum chemical methods for recovering electron correlation energy. Except for a few cases such as aromatic ring stacking, large basis set MP2 calculation is able to give very accurate structures and intermolecular interaction energies for molecules and clusters consisting of H, C, N and O atoms. Pople *et al*. derived and implemented the MP2 analytic gradient.[139] Handy and Schaefer formulated the Z-vector method so MP2 gradient and some other molecular properties can be evaluated efficiently.[140] The ZAPT2 analytic gradient was derived and implemented by Fletcher *et al.*[141] and Aikens *et al.*[142,143]

The polarizable continuum models have been introduced in Chapter 2. The representative continuum solvation models include dielectric solvation models[144] (the earlier DPCM[12] and the more recent IEF-PCM[13]), conductor-like screening models (COSMO[55] and GCOSMO[145] or CPCM[56]), SS(V)PE[58] models and SM*x* models.[59] Cammi, Mennucci and Tomasi[18] established a closed shell RMP2/PCM method that uses RHF/PCM orbitals and their energies to obtain the second order energy correction $E^{(2)}$.

They also derived and implemented a Z-vector method to determine the RMP2/PCM relaxed density and analytic gradient. To the best of my knowledge, PCM methods have not been combined with open shell MP2 methods such as ZAPT2 and UMP2. In this chapter, an extension of Cammi, Mennucci and Tomasi's RMP2/PCM method to ZAPT2 and UMP2 is described. These methods are used to study the S_0 and T_1 states of acetone, nucleobases and nucleobase pairs.

4.2 Theory

4.2.1 Many-body perturbation theory

Since the solution to the Hartree Fock equation is known, the Møller-Plesset perturbation theory[8] takes the sum of one electron Fock operators as the unperturbed Hamiltonian $\boldsymbol{H}_0$, therefore the difference between a real Hamiltonian and $\boldsymbol{H}_0$ is the perturbation $\boldsymbol{H}'$.

$$\begin{aligned}\boldsymbol{H}_0 &= \sum_{i=1}^{N_{elec}} \boldsymbol{F}_i = \sum_{i=1}^{N_{elec}} \left(\boldsymbol{h}_i + \sum_{j=1}^{N_{elec}} \left(\boldsymbol{J}_j - \boldsymbol{K}_j \right) \right) \\ &= \sum_{i=1}^{N_{elec}} \boldsymbol{h}_i + \sum_{i=1}^{N_{elec}} \sum_{j=1}^{N_{elec}} \left\langle \boldsymbol{g}_{ij} \right\rangle = \sum_{i=1}^{N_{elec}} \boldsymbol{h}_i + 2\left\langle \boldsymbol{V}_{ee} \right\rangle \\ \boldsymbol{H}' &= \boldsymbol{H} - \boldsymbol{H}_0 = \sum_{i=1}^{N_{elec}} \sum_{j=1}^{N_{elec}} \boldsymbol{g}_{ij} - \sum_{i=1}^{N_{elec}} \sum_{j=1}^{N_{elec}} \left\langle \boldsymbol{g}_{ij} \right\rangle = \boldsymbol{V}_{ee} - 2\left\langle \boldsymbol{V}_{ee} \right\rangle\end{aligned} \tag{4-1}$$

Since the average electron repulsion has been counted twice in the sum of Fock operator, the perturbation operator should be the exact repulsion minus twice of the average repulsion. The corrections can be made for $\mathbf{H}$ by applying the many-body perturbation theory (MBPT).[22]

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{H}_0 + \lambda \boldsymbol{H}' \\ \boldsymbol{H}_0 \Phi_i &= E_i^{(0)} \Phi_i \quad i = 0,1,2,...,\infty \end{aligned} \tag{4-2}$$

For a time-independent, non-degenerate wave function, the perturbed Schrödinger equation is given by

$$\boldsymbol{H}\Psi = E\Psi \tag{4-3}$$

The energy and wave function can be expanded in a Taylor series:

$$\begin{aligned} E &= \lambda^0 E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots \\ \Psi &= \lambda^0 \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \cdots \end{aligned} \tag{4-4}$$

Expanding the corresponding terms in the Schrödinger equation yields

$$\begin{aligned} &\left(\mathbf{H}_0 + \lambda \mathbf{H}'\right)\left(\lambda^0 \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \cdots\right) \\ &= \left(\lambda^0 E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots\right)\left(\lambda^0 \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \cdots\right) \end{aligned} \tag{4-5}$$

Collecting like powers of λ and equating the coefficients of like powers gives:

$$\begin{aligned} &\lambda^0 : \boldsymbol{H}_0 \Psi^{(0)} = E^{(0)} \Psi^{(0)} \\ &\lambda^1 : \boldsymbol{H}_0 \Psi^{(1)} + \boldsymbol{H}' \Psi^{(0)} = E^{(0)} \Psi^{(1)} + E^{(1)} \Psi^{(0)} \\ &\lambda^2 : \boldsymbol{H}_0 \Psi^{(2)} + \boldsymbol{H}' \Psi^{(1)} = E^{(0)} \Psi^{(2)} + E^{(1)} \Psi^{(1)} + E^{(2)} \Psi^{(0)} \\ &\vdots \\ &\lambda^n : \boldsymbol{H}_0 \Psi^{(n)} + \boldsymbol{H}' \Psi^{(n-1)} = \sum_{i=0}^{n} E^{(i)} \Psi^{(n-i)} \end{aligned} \tag{4-6}$$

The n[th] order energy correction can be calculated by left multiplying Φ_0 and integrating as follows:

$$\begin{aligned} &\left\langle \Phi_0 \middle| \mathbf{H}_0 \middle| \Psi^{(n)} \right\rangle + \left\langle \Phi_0 \middle| \mathbf{H}' \middle| \Psi^{(n-1)} \right\rangle = \sum_{i=0}^{n-1} E^{(i)} \left\langle \Phi_0 \middle| \Psi^{(n-i)} \right\rangle + E^{(n)} \left\langle \Phi_0 \middle| \Psi^{(0)} \right\rangle \\ &E^{(0)} \left\langle \Psi^{(n)} \middle| \Phi_0 \right\rangle + \left\langle \Phi_0 \middle| \mathbf{H}' \middle| \Psi^{(n-1)} \right\rangle = E^{(n)} \left\langle \Phi_0 \middle| \Psi^{(0)} \right\rangle \\ &E^{(n)} = \left\langle \Phi_0 \middle| \mathbf{H}' \middle| \Psi^{(n-1)} \right\rangle \end{aligned} \tag{4-7}$$

The first-order wave function can be expanded using the complete set Φ_i formed by the unperturbed wave functions, known as Rayleigh-Schrödinger perturbation theory.

$$\begin{aligned}\Psi^{(1)} &= \sum_i c_i \Phi_i \\ \lambda^1 &: \boldsymbol{H}_0 \sum_i c_i \Phi_i + \boldsymbol{H}' \Phi_0 = E^{(0)} \sum_i c_i \Phi_i + E^{(1)} \Phi_0\end{aligned} \tag{4-8}$$

Left multiplying Φ_j^* and integrating yields

$$\begin{aligned}&\lambda^1 : \sum_i c_i \left\langle \Phi_j \middle| \boldsymbol{H}_0 \middle| \Phi_i \right\rangle + \left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle - E^{(0)} \sum_i c_i \left\langle \Phi_j \middle| \Phi_i \right\rangle - E^{(1)} \left\langle \Phi_j \middle| \Phi_0 \right\rangle = 0 \\ &\sum_i c_i E_i^{(0)} \left\langle \Phi_j \middle| \Phi_i \right\rangle + \left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle - c_j E^{(0)} = 0 \\ &c_j E_j^{(0)} + \left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle - c_j E^{(0)} = 0 \\ &c_j = \frac{\left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle}{E_0^{(0)} - E_j^{(0)}}\end{aligned} \tag{4-9}$$

Analogous formulas can be generated for the second-order corrections,

$$\Psi^{(2)} = \sum_i d_i \Phi_i \tag{4-10}$$

$$\begin{aligned}&\lambda^2 : \sum_i d_i \left\langle \Phi_j \middle| \boldsymbol{H}_0 \middle| \Phi_i \right\rangle - E^{(0)} \sum_i d_i \left\langle \Phi_j \middle| \Phi_i \right\rangle \\ &+ \sum_i c_i \left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_i \right\rangle - E^{(1)} \sum_i c_i \left\langle \Phi_j \middle| \Phi_i \right\rangle - E^{(2)} \left\langle \Phi_j \middle| \Phi_0 \right\rangle = 0 \\ &d_j = \sum_{i \neq 0} \frac{\left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_i \right\rangle \left\langle \Phi_i \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle}{\left(E_0^{(0)} - E_j^{(0)}\right)\left(E_0^{(0)} - E_i^{(0)}\right)} - \frac{\left\langle \Phi_j \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle \left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle}{\left(E_0^{(0)} - E_j^{(0)}\right)^2}\end{aligned} \tag{4-11}$$

In the Møller-Plesset perturbation theory,[8] the zeroth-order wave function is the HF determinant, and the zeroth-order energy is the sum of MO energies.

$$E^{(0)} = \left\langle \Phi_0 \middle| \boldsymbol{H}_0 \middle| \Phi_0 \right\rangle = \left\langle \Phi_0 \middle| \sum_{i=1}^{N_{elec}} \boldsymbol{F}_i \middle| \Phi_0 \right\rangle = \sum_{i=1}^{N_{elec}} \varepsilon_i \tag{4-12}$$

The first-order energy correction is

$$E^{(1)} = \left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle = \left\langle \boldsymbol{V}_{ee} \right\rangle - 2\left\langle \boldsymbol{V}_{ee} \right\rangle = -\left\langle \boldsymbol{V}_{ee} \right\rangle \tag{4-13}$$

which corrects the double counting of the electron-electron repulsion at zeroth-order. The second-order energy correction can be written as

$$E^{(2)} = \left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Psi^{(1)} \right\rangle = \sum_i c_i \left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Phi_i \right\rangle = \sum_{i \neq 0} \frac{\left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Phi_i \right\rangle \left\langle \Phi_i \middle| \boldsymbol{H}' \middle| \Phi_0 \right\rangle}{E_0 - E_i} \tag{4-14}$$

The matrix elements are the expectation value of the perturbation operator between the HF reference and all possible excited states. Since the perturbation operator is a two-electron operator, all matrix elements containing higher than double excitation terms will vanish. Furthermore, when canonical HF orbitals are used, elements with singly excited states are also zero as follows:

$$\left\langle \Phi_0 \middle| \boldsymbol{H}' \middle| \Phi^a \right\rangle = \left\langle \Phi_0 \middle| \boldsymbol{H}_0 - \sum_{j=1}^{N_{elec}} \boldsymbol{F}_j \middle| \Phi_0^a \right\rangle = \left\langle \Phi_0 \middle| \boldsymbol{H}_0 \middle| \Phi_i^a \right\rangle - \sum_{j=1}^{N_{elec}} \varepsilon_j \left\langle \Phi_0 \middle| \Phi_i^a \right\rangle = 0 \tag{4-15}$$

The first term vanishes because of Brillouin's theorem, and the second term vanishes because the orbitals are orthogonal to each other. Therefore, the second-order correction to the energy only contains a sum over doubly excited determinants, in which the electrons are promoted from occupied orbitals i and j to virtual orbitals a and b. It is expressed as follows in terms of molecular orbitals:

$$E^{(2)} = \sum_{i<j}^{occ}\sum_{a<b}^{vir}\frac{\left\langle\Phi_0\left|\boldsymbol{H}'\right|\Phi_{ij}^{ab}\right\rangle\left\langle\Phi_{ij}^{ab}\left|\boldsymbol{H}'\right|\Phi_0\right\rangle}{E_0 - E_{ij}^{ab}} = \sum_{i<j}^{occ}\sum_{a<b}^{vir}\frac{\left\langle\phi_i\phi_j\left|\phi_a\phi_b\right\rangle - \left\langle\phi_i\phi_j\left|\phi_b\phi_a\right\rangle\right.\right.}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{4-16}$$

4.2.2 MP2/CPCM gradient

The RMP2/PCM method established by Cammi, Mennucci and Tomasi[18] can be generalized to UMP2 and ZAPT2 methods such that the corresponding RHF/PCM, UHF/PCM and ROHF/PCM orbitals and orbital energies are used to obtain the second order energy correction $E^{(2)}$, which is added to the corresponding HF/PCM energy. $E^{(2)}$ is formally identical to the regular MP2 method.

Here the RMP2/CPCM, UMP2/CPCM and ZAPT2/CPCM methods are considered. For all of these three methods, the first derivative (gradient) of the second order energy correction $E^{(2)}$ with respect to a coordinate x can be written in density matrices and basis function integrals,[146]

$$\begin{aligned} E^{(2),x} = &-\sum_{\mu\nu} W_{\mu\nu}^{(2)} S_{\mu\nu}^{x} + \sum_{\mu\nu} P_{\mu\nu}^{(2)} h_{\mu\nu}^{x} \\ &+ \sum_{\mu\nu\rho\sigma} P_{\mu\nu}^{(2)} P_{\rho\sigma} \left\langle\mu\rho\|\nu\sigma\right\rangle^{x} + \sum_{\mu\nu\rho\sigma} \Gamma_{\mu\nu\rho\sigma}^{NS} \left\langle\mu\rho\|\nu\sigma\right\rangle^{x} \\ &+ \sum_{\mu\nu} P_{\mu\nu}^{(2)} \left(\boldsymbol{V}_{\mu\nu}^{T}\boldsymbol{q}_N\right)^{x} + \sum_{\mu\nu\rho\sigma} P_{\mu\nu}^{(2)} P_{\rho\sigma} \left(\boldsymbol{V}_{\mu\nu}^{T}\boldsymbol{q}_{\rho\sigma}\right)^{x} \end{aligned} \tag{4-17}$$

where μ, ν, ρ and σ are spin-orbital basis functions; $\left\langle\mu\rho\|\nu\sigma\right\rangle$ is the standard antisymmetrized two-electron integrals in the $\left\langle 12\|12\right\rangle$ notation; **h** is the one-electron integral; **S** is the overlap integral; $\mathbf{W}^{(2)}$ is the MP2 correction to the HF energy-weighted density matrix; **P** is the HF density matrix; $\mathbf{P}^{(2)}$ is the MP2 correction to the HF density matrix, which can be determined via a modified Z-vector method (to be discussed

below); $\Gamma^{NS}_{\mu\nu\rho\sigma}$ is the nonseparable two-particle density matrix. Here **V** is a set (vector) of electrostatic potentials at the molecular surface points created by the electron density of the basis set product; the superscript T denotes transpose here and hereafter; **q** is a set (vector) of induced surface charges ($\mathbf{q}_N$ due to nuclei and $\mathbf{q}_{\rho\sigma}$ due to electron density of the basis set product). All **q** and **V** satisfy the linear CPCM equation:

$$\mathbf{C}\cdot\mathbf{q} = -\left(1-1/\varepsilon\right)\mathbf{V} \tag{4-18}$$

The elements of the matrix **C** are $\mathbf{C}_{ii} = 1.07\sqrt{4\pi/a_i}$ and $\mathbf{C}_{ij} = 1/\left|\mathbf{r}_i - \mathbf{r}_j\right|$, with a_i being the area and $\boldsymbol{r}_i$ being the center coordinates of surface tessera *i*. ε is the dielectric constant of the solvent. Scaling factors other than $\left(1-1/\varepsilon\right)$ were also suggested.[55,56,65]

The first four terms in Eq. (4-17) have exactly the same forms as those in the regular MP2 energy gradient formula.[140] The last two terms in Eq. (4-17) can be written as:

$$\begin{aligned}
&\sum_{\mu\nu} P^{(2)}_{\mu\nu}\left(\mathbf{V}^T_{\mu\nu}\mathbf{q}_N\right)^x + \sum_{\mu\nu} P^{(2)}_{\mu\nu}\sum_{\rho\sigma} P_{\rho\sigma}\left(\mathbf{V}^T_{\mu\nu}\mathbf{q}_{\rho\sigma}\right)^x \\
&= \left(\mathbf{V}^{(2),x}\right)^T\mathbf{q}_N + \left(\frac{\varepsilon}{\varepsilon-1}\right)\left(\mathbf{q}^{(2)}\right)^T\mathbf{C}^x\mathbf{q}_N + \left(\mathbf{q}^{(2)}\right)^T\mathbf{V}^x_N \\
&+ \left(\mathbf{V}^{(2),x}\right)^T\mathbf{q}_{HF} + \left(\frac{\varepsilon}{\varepsilon-1}\right)\left(\mathbf{q}^{(2)}\right)^T\mathbf{C}^x\mathbf{q}_{HF} + \left(\mathbf{q}^{(2)}\right)^T\mathbf{V}^x_{HF} \\
&= \left(\mathbf{V}^{(2),x}\right)^T\left(\mathbf{q}_N + \mathbf{q}_{HF}\right) + \frac{1}{2}\left(\frac{\varepsilon}{\varepsilon-1}\right)\left(\mathbf{q}^{(2)}\right)^T\mathbf{C}^x\left(\mathbf{q}_N + \mathbf{q}_{HF}\right) + \left(\mathbf{q}^{(2)}\right)^T\left(\mathbf{V}^x_N + \mathbf{V}^x_{HF}\right)
\end{aligned} \tag{4-19}$$

where $\mathbf{V}^{(2),x}$ is the electrostatic field created by $\mathbf{P}^{(2)}$, $\mathbf{q}^{(2)}$ is the CPCM surface charge induced by the electrostatic potential created by $\mathbf{P}^{(2)}$,

$$\mathbf{V}^{(2),x} = \sum_{\mu\nu} P^{(2)}_{\mu\nu}\mathbf{V}^x_{\mu\nu} \tag{4-20}$$

$$\mathbf{q}^{(2)} = \sum_{\mu\nu} P^{(2)}_{\mu\nu}\mathbf{q}_{\mu\nu} \tag{4-21}$$

The first term in the final expression of Eq. (4-19) represents the force between $\mathbf{q}_N+\mathbf{q}_{HF}$ (surface charge induced by nuclei and HF density **P**) and the MP2 relaxed electron density $\mathbf{P}^{(2)}$; The second term represents the force between $\mathbf{q}_N+\mathbf{q}_{HF}$ and $\mathbf{q}^{(2)}$; the third term represents the force between $\mathbf{q}^{(2)}$ and solute nuclei, and the force between $\mathbf{q}^{(2)}$ and the HF electron density **P**.

The evaluation of Eq. (4-19) requires the derivatives of the tessera coordinates and areas with respect to molecular geometric change. In the FIXPVA[16] tessellation scheme, the positions of the surface tesserae are fixed relative to their center atoms so the tessera areas are smooth functions of their distances to neighboring spheres. Therefore, smooth and rigorous analytic derivatives of the tessera positions and areas with respect to atomic coordinates can be obtained.

4.2.3 Z-vector equations

2n+1 rule states that in order to obtain the $(2n)^{th}$ and $(2n+1)^{th}$ energy gradients, it is sufficient to solve the n^{th} order coupled perturbed equations (CPHF) which provide the n^{th} order derivative of the wave function. Handy and Schaefer[140] have proved in their Z-vector method that (2n+1) also applies in nonvariational methods such as perturbation theory. The notations in this chapter follow those used in several literatures: the doubly occupied molecular orbitals (DOCC or D) are denoted as *i*, *j, k* and *l*, the singly occupied (SOCC or S) ones are denoted as *x* and *y,* the virtual (VIRT or V) ones are denoted as *a*, *b*, and *c*, general molecular orbitals (both occupied and virtual) are denoted as *p* and *q*. Superscript x refers to the derivative with respect to nuclear displacement while (x) is the derivative of AO integrals only, not including expansion coefficients. The CPHF

equations[147] are derived by differentiating the Fock matrix equation with respect to nuclear displacement x

$$\begin{aligned} \boldsymbol{A}'\boldsymbol{U}^x &= \boldsymbol{B}^x \\ A'_{aibj} &= \delta_{ab}\delta_{ij}\left(\varepsilon_i - \varepsilon_a\right) - A_{aibj} \end{aligned} \tag{4-22}$$

where A_{aibj} is the orbital Hessian matrix and ε_a and ε_i are the orbital energies.

$$A_{aibj} = \langle ij \| ab \rangle + \langle ib \| aj \rangle \tag{4-23}$$

$$B_{ai}^x = F_{ai}^{(x)} - S_{ai}^{(x)}\varepsilon_i - \frac{1}{2}\sum_{kl}^{occ} S_{kl}^{(x)} A_{ailk} \tag{4-24}$$

$$F_{ai}^{(x)} = H_{ai}^{(x)} + \sum_{k}^{occ}\left[(ai\,|\,kk)^{(x)} - (ak\,|\,ik)^{(x)}\right] \tag{4-25}$$

$$U_{ai}^x = \frac{1}{\left(\varepsilon_i - \varepsilon_a\right)} Q_{ai}^x \tag{4-26}$$

$$Q_{ai}^x \equiv B_{ai}^x + \sum_{c}^{vir}\sum_{k}^{occ} U_{ck}^x A_{aick} \tag{4-27}$$

$$\begin{aligned} L_{ai} = &\sum_{jk}^{occ} P_{jk}^{(2)} A_{aijk} + \sum_{bc}^{vir} P_{bc}^{(2)} A_{aibc} \\ &- \sum_{jk}^{occ}\sum_{b}^{vir} T_{jk}^{ab}(ij\,|\,bk) + \sum_{j}^{occ}\sum_{bc}^{vir} T_{jk}^{bc}(ab\,|\,jc) \end{aligned} \tag{4-28}$$

$$T_{ij}^{ab} = \frac{(ia \,||\, jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{4-29}$$

The CPHF equations Eq. (4-22) become

$$\boldsymbol{U}^x = \left(\boldsymbol{A}'\right)^{-1}\boldsymbol{B}^x \tag{4-30}$$

Define

$$\sum_{ai} U_{ai}^x L_{ai} \equiv \boldsymbol{L}^T\boldsymbol{U}^x \tag{4-31}$$

thus

$$\begin{aligned} \boldsymbol{L}^T\boldsymbol{U}^x &= \boldsymbol{L}^T\left(\boldsymbol{A}'\right)^{-1}\boldsymbol{B}^x = \boldsymbol{Z}^T\boldsymbol{B}^x \\ \boldsymbol{Z}^T &= \boldsymbol{L}^T\left(\boldsymbol{A}'\right)^{-1} \\ \left(\boldsymbol{A}'\right)^T\boldsymbol{Z} &= \boldsymbol{L} \end{aligned} \tag{4-32}$$

Comparing Eq. (4-32) with Eq. (4-22), the number of the equations has been reduced from $3N_{atom}$ to 1. The elements of **Z** are used to define the virtual-occupied block of the density matrix.

$$\sum_{ai} B_{ai}^x Z_{ai} \equiv \sum_{ai} B_{ai}^x P_{ai}^{(2)} \tag{4-33}$$

For closed shell RMP2, the occupied-occupied and virtual-virtual blocks of the $\mathbf{P}^{(2)}$ can be readily evaluated using the RMP2 excitation amplitudes.

$$\begin{aligned} \boldsymbol{P}_{ij}^{(2)} &= -\frac{1}{2}\sum_{kab} a_{ik}^{ab} a_{jk}^{ab} \\ \boldsymbol{P}_{ab}^{(2)} &= -\frac{1}{2}\sum_{ijc} a_{ij}^{ac} a_{ij}^{bc} \end{aligned} \tag{4-34}$$

The double excitation amplitudes are

$$a_{ij}^{ab} = \frac{\langle ij \| ab \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{4-35}$$

The occupied-virtual blocks of $\mathbf{P}^{(2)}$, when written as a vector (Lagrange multiplier), can be determined using the Z-vector equation[140]

$$\sum_{bj} A_{aibj} P_{bj}^{(2)} + \left(\varepsilon_a - \varepsilon_i\right) P_{ai}^{(2)} = L_{ai} \tag{4-36}$$

Here **L** is the Lagrangian.[140]

Cammi, Mennucci and Tomasi showed that by taking into account induced surface charges, the Z-vector method can be used to determine the $\mathbf{P}^{(2)}$ for RMP2/PCM

methods.[18] In this study, the Z-vector method is extended to open shell UMP2/PCM and ZAPT2/PCM methods.

For RMP2/CPCM, induced charge terms should be added to the orbital Hessian matrix **A** and the Lagrangian **L**, and the occupied-occupied block of the $\mathbf{W}^{(2)}$ matrix,

$$A^{pcm}_{aibj} = A_{aibj} + 2\mathbf{V}^T_{ai}\mathbf{q}_{bj} \tag{4-37}$$

$$L^{pcm}_{ai} = L_{ai} + \sum_{jk} P^{(2)}_{jk}\mathbf{V}^T_{ai}\mathbf{q}_{jk} + \sum_{bc} P^{(2)}_{bc}\mathbf{V}^T_{ai}\mathbf{q}_{bc} \tag{4-38}$$

$$W^{(2),\,pcm}_{ij} = W^{(2)}_{ij} - \frac{1}{2}\sum_{pq} P^{(2)}_{pq}\mathbf{V}^T_{ij}\mathbf{q}_{pq} \tag{4-39}$$

For open shell UMP2, two associated Z-vector equations for α and β orbitals should be solved as listed in Ref 143:

$$\begin{aligned}&\sum_{b^\alpha}^{v}\sum_{j^\alpha}^{o}\left\{A_{a^\alpha i^\alpha b^\alpha j^\alpha} + \delta_{a^\alpha b^\alpha}\delta_{i^\alpha j^\alpha}\left(\varepsilon^\alpha_b - \varepsilon^\alpha_j\right)\right\}\boldsymbol{P}^{(2)}_{b^\alpha j^\alpha}\\ &+\sum_{b^\beta}^{v}\sum_{j^\beta}^{o}\left\{A_{a^\alpha i^\alpha b^\beta j^\beta} + \delta_{a^\alpha b^\alpha}\delta_{i^\beta j^\beta}\left(\varepsilon^\beta_b - \varepsilon^\beta_j\right)\right\}\boldsymbol{P}^{(2)}_{b^\beta j^\beta} = -L_{a^\alpha i^\alpha}\end{aligned} \tag{4-40}$$

For UMP2/CPCM, induced charge terms should be added to the orbital Hessian matrix **A** (see Eq. 91, 92, 93, 94, 99, 100 in Ref 143), the Lagrangian **L** (see Eq. 98 in Ref 143), and the occupied-occupied block of the $\mathbf{W}^{(2)}$ matrix (see Eq. 96 in Ref 143).

$$\begin{aligned}\mathbf{A}^{pcm}_{a^\alpha i^\alpha b^\alpha j^\alpha} &= \mathbf{A}_{a^\alpha i^\alpha b^\alpha j^\alpha} + 2\mathbf{V}^T_{b^\alpha j^\alpha}\mathbf{q}_{a^\alpha i^\alpha}\\ \mathbf{A}^{pcm}_{a^\alpha i^\alpha b^\beta j^\beta} &= \mathbf{A}_{a^\alpha i^\alpha b^\beta j^\beta} + 2\mathbf{V}^T_{b^\beta j^\beta}\mathbf{q}_{a^\alpha i^\alpha}\end{aligned} \tag{4-41}$$

$$\begin{aligned}L^{pcm}_{a^\alpha i^\alpha} = L_{a^\alpha i^\alpha} &+ 2\sum_{j^\alpha k^\alpha} P^{(2)}_{j^\alpha k^\alpha}\mathbf{V}^T_{a^\alpha i^\alpha}\mathbf{q}_{j^\alpha k^\alpha} + 2\sum_{j^\beta k^\beta} P^{(2)}_{j^\beta k^\beta}\mathbf{V}^T_{a^\alpha i^\alpha}\mathbf{q}_{j^\beta k^\beta}\\ &+ 2\sum_{b^\alpha c^\alpha} P^{(2)}_{b^\alpha c^\alpha}\mathbf{V}^T_{a^\alpha i^\alpha}\mathbf{q}_{b^\alpha c^\alpha} + 2\sum_{b^\beta c^\beta} P^{(2)}_{b^\beta c^\beta}\mathbf{V}^T_{a^\alpha i^\alpha}\mathbf{q}_{b^\beta c^\beta}\end{aligned} \tag{4-42}$$

$$W^{(2),\,pcm}_{i^\alpha j^\alpha} = W^{(2)}_{i^\alpha j^\alpha} - \sum_{p^\alpha q^\alpha} P^{(2)}_{p^\alpha q^\alpha}\left(\mathbf{V}^T_{p^\alpha q^\alpha}\mathbf{q}_{i^\alpha j^\alpha}\right) - \sum_{p^\beta q^\beta} P^{(2)}_{p^\beta q^\beta}\left(\mathbf{V}^T_{p^\beta q^\beta}\mathbf{q}_{i^\alpha j^\alpha}\right) \tag{4-43}$$

Here only the α equations are given. The β equations can be obtained by switching α and β.

For open shell ZAPT2, a single Z-vector equation can be constructed using a nine-block (but only six independent blocks) orbital Hessian matrix **A** for doubly occupied, singly occupied and virtual orbitals (see Figure 4.3 in Ref 142),

$$\begin{pmatrix} \mathbf{A}_{xi,yj} & \mathbf{A}_{xi,bj} & \mathbf{A}_{xi,by} \\ \mathbf{A}_{ai,yj} & \mathbf{A}_{ai,bj} & \mathbf{A}_{ai,by} \\ \mathbf{A}_{ax,yj} & \mathbf{A}_{ax,bj} & \mathbf{A}_{ax,by} \end{pmatrix} \times \begin{pmatrix} P_{yj} \\ P_{bj} \\ P_{by} \end{pmatrix} = \begin{pmatrix} L_{xi} \\ L_{ai} \\ L_{ax} \end{pmatrix} \tag{4-44}$$

For ZAPT2/CPCM, induced charge terms should be added to the orbital Hessian matrix **A** (see Eq. 81 in Ref 142), the Lagrangian **L** (see Eq. 62-66 in Ref 142) and the double-double, double-single and single-single blocks of the $\mathbf{W}^{(2)}$ matrix (see Eq. 55-61 in Ref 142),

$$\begin{aligned}
\mathbf{A}^{pcm}_{xi,yj} &= \mathbf{A}_{xi,yj} + \mathbf{V}^T_{yj}\mathbf{q}_{xi} \\
\mathbf{A}^{pcm}_{ai,yj} &= \mathbf{A}_{ai,yj} + 2\mathbf{V}^T_{yj}\mathbf{q}_{ai} \\
\mathbf{A}^{pcm}_{ax,yj} &= \mathbf{A}_{ax,yj} + \mathbf{V}^T_{yj}\mathbf{q}_{ax} \\
\mathbf{A}^{pcm}_{ai,bj} &= \mathbf{A}_{ai,bj} + 4\mathbf{V}^T_{bj}\mathbf{q}_{ai} \\
\mathbf{A}^{pcm}_{ax,bj} &= \mathbf{A}_{ax,bj} + 2\mathbf{V}^T_{bj}\mathbf{q}_{ax} \\
\mathbf{A}^{pcm}_{ax,by} &= \mathbf{A}_{ax,by} + \mathbf{V}^T_{by}\mathbf{q}_{ax}
\end{aligned} \tag{4-45}$$

$$\begin{aligned}
L^{pcm}_{xi} &= L_{xi} + 2\sum_{jk} P^{(2)}_{jk}\mathbf{V}^T_{xi}\mathbf{q}_{jk} + 2\sum_{xy} P^{(2)}_{xy}\mathbf{V}^T_{xi}\mathbf{q}_{xy} + 2\sum_{bc} P^{(2)}_{bc}\mathbf{V}^T_{xi}\mathbf{q}_{bc} \\
L^{pcm}_{ai} &= L_{ai} + 4\sum_{jk} P^{(2)}_{jk}\mathbf{V}^T_{ai}\mathbf{q}_{jk} + 4\sum_{xy} P^{(2)}_{xy}\mathbf{V}^T_{ai}\mathbf{q}_{xy} + 4\sum_{bc} P^{(2)}_{bc}\mathbf{V}^T_{ai}\mathbf{q}_{bc} \\
L^{pcm}_{ax} &= L_{ax} + 2\sum_{jk} P^{(2)}_{jk}\mathbf{V}^T_{ax}\mathbf{q}_{jk} + 2\sum_{xy} P^{(2)}_{xy}\mathbf{V}^T_{ax}\mathbf{q}_{xy} + 2\sum_{bc} P^{(2)}_{bc}\mathbf{V}^T_{ax}\mathbf{q}_{bc}
\end{aligned} \tag{4-46}$$

$$W_{ij}^{(2),\,pcm} = W_{ij}^{(2)} - 2\sum_{pq} P_{pq}^{(2)} \mathbf{V}_{ij}^{T} \mathbf{q}_{pq}$$
$$W_{ix}^{(2),\,pcm} = W_{ix}^{(2)} - \sum_{pq} P_{pq}^{(2)} \mathbf{V}_{ix}^{T} \mathbf{q}_{pq} \qquad (4\text{-}47)$$
$$W_{xy}^{(2),\,pcm} = W_{xy}^{(2)} - \sum_{pq} P_{pq}^{(2)} \mathbf{V}_{xy}^{T} \mathbf{q}_{pq}$$

For frozen core MP2 calculations, corresponding modifications can be made following the expressions in Ref 142 and 143.

4.2.4 Solvation energy correction

The MP2/CPCM energy is the sum of the HF/CPCM energy and the second order energy correction $E^{(2)}$. The advantage of using this formulation is that the analytic gradient can be derived. The disadvantage is that the CPCM solvation energy is obtained with the HF density instead of MP2 relaxed density. Here a simple method will be described to obtain the solvation energy using the MP2 relaxed density. In the sense of perturbation theory, the solvation energy obtained using the second order relaxed density is actually at the fourth order.

The solvation energy at the HF level is defined as:

$$G_{sol}^{HF} = \frac{1}{2}\left(\mathbf{V}_N + \mathbf{V}_{HF}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF}\right) \qquad (4\text{-}48)$$

where $\mathbf{V}_N$ and $\mathbf{q}_N$ are the potential and induced charges created by the nuclei, $\mathbf{V}_{HF}$ and $\mathbf{q}_{HF}$ are the potential and induced charges created by the HF electron density.

Similar to Eq. (4-48), a solvation energy G_{sol}^{MP2} can be defined using the MP2 relaxed density,

$$
\begin{aligned}
G_{sol}^{MP2} &= \frac{1}{2}\left(\mathbf{V}_N + \mathbf{V}_{HF} + \mathbf{V}^{(2)}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF} + \mathbf{q}^{(2)}\right) \\
&= \frac{1}{2}\left(\mathbf{V}_N + \mathbf{V}_{HF}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF}\right) + \frac{1}{2}\left(\mathbf{V}_N + \mathbf{V}_{HF}\right)^T \left(\mathbf{q}^{(2)}\right) \\
&+ \frac{1}{2}\left(\mathbf{V}^{(2)}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF}\right) + \frac{1}{2}\left(\mathbf{V}^{(2)}\right)^T \left(\mathbf{q}^{(2)}\right)
\end{aligned}
\tag{4-49}
$$

Here $\mathbf{V}^{(2)}$ and $\mathbf{q}^{(2)}$ are the potential and induced charges created by $\mathbf{P}^{(2)}$. The G_{sol}^{HF} defined in Eq. (4-48), for HF/CPCM calculation is the first term in Eq. (4-49). The second and third terms in Eq. (4-49) are equal to each other and can be combined. The third term (doubled) is used in the following. Therefore, the correction of the solvation energy due to $\mathbf{P}^{(2)}$ is,

$$
\begin{aligned}
G_{sol}^{(2)} &= G_{sol}^{MP2} - G_{sol}^{HF} \\
&= \left(\mathbf{V}^{(2)}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF}\right) + \frac{1}{2}\left(\mathbf{V}^{(2)}\right)^T \left(\mathbf{q}^{(2)}\right) \\
&= \left(\mathbf{V}^{(2)}\right)^T \left(\mathbf{q}_N + \mathbf{q}_{HF} + \frac{1}{2}\mathbf{q}^{(2)}\right)
\end{aligned}
\tag{4-50}
$$

It is difficult to derive the expression of the analytic nuclear gradient for G_{sol}^{MP2} or $G_{sol}^{(2)}$. Therefore, Eq. (4-50) is best used as a single point energy correction.

4.3 Implementation and computational methodology

The MP2/CPCM gradient code was implemented in the GAMESS[68] package based on several gas phase MP2 programs and the CPCM code previously implemented by Li and Jensen[69] on the basis of the IEF-PCM program originally implemented by Mennucci, Cances, Pomelli and Tomasi.[13,70,71] The MP2 programs include the serial RMP2 and UMP2 program,[138] the parallel RMP2 program implemented by Fletcher *et al.*[148], the parallel RMP2 program implemented by Ishimura, Pulay and Nagase,[149] the parallel ZAPT2 program implemented by Fletcher *et al.*[141] and Aikens *et al.*[142], and the parallel UMP2 program implemented by Aikens *et al.*[143,150]. Numerical tests using the

FIXPVA[16] tessellation scheme for CPCM show that the accuracy of the analytic gradients obtained with GAMESS default settings is typically 10^{-6} hartree/bohr. Tightening the default settings leads to better accuracies such as 10^{-7} hartree/bohr, suggesting that the implementations are correct. Using FIXPVA, the MP2/CPCM potential energy surfaces are rigorously continuous and smooth. Eq. (4-50) was also implemented. The $E^{MP2}+G_{sol}^{(2)}$ energy is printed out as "EMP2+EPCM(2)" in the GAMESS output file.

In the CPCM calculations, spheres with radii of 0, 2.124, 2.016 and 1.908 Å (after scaled by 1.20) were used for H, C, N and O atoms, respectively, to define the molecular cavity; no additional spheres were used. Using zero radii for H atoms means that they do not contribute to form the surface. The tessellation scheme FIXPVA was used with 240 initial tesserae per sphere. The induced surface charges were determined by a semi-iterative DIIS procedure[71,72] with no charge renormalization. A dielectric constant = 78.39 was used to represent water solvent. Only the CPCM electrostatic solvation energy was included in these calculations. The aug-cc-pvTZ and aug-cc-pVDZ basis set[74] were used for acetone and nucleobases, respectively. Geometry optimization was performed in internal coordinates generated by the automatic delocalized coordinates algorithm.[73] No symmetry was imposed in the geometry optimization.

The T_1 excited states of acetone, cytosine, thymine and uracil can be well described by an open shell ROHF or UHF wavefunction. Density functional theory (DFT) methods can be used to include some electron correlation. For example, Zhang *et al.* used open shell DFT methods to study interaction of triplet uracil and thymine with water.[151] Consequently, ZAPT2 and UMP2 can be used to describe electron correlation. Thiel *et al.* found that the $S_0 \rightarrow S_1$ excitations of acetone, cytosine, thymine and uracil

can be well described using coupled cluster linear response theory because they are dominated by single excitation of a single determinant RHF reference state.[152] Usually, the $S_0 \rightarrow T_1$ excitations have even higher weights of single excitation as compared to $S_0 \rightarrow S_1$ excitations. Of course, the T_1 state ROHF or UHF wavefunction is not orthogonal to the S_0 state RHF wavefunction because the Fock operators are different in the RHF and ROHF (or UHF) equations. However, this should not be a severe problem if the orthogonality can be approximately satisfied.[153] Compared to costly multireference perturbation theory and coupled cluster linear response theory methods, ZAPT2 and UMP2 methods are much more efficient, and can be used to perform geometry optimization with analytic gradient. Compared to open shell DFT methods, ZAPT2 and UMP2 are able to predict correct intermolecular interactions. So, in many cases, using ZAPT2 or UMP2 methods for T_1 excited states is a viable option. The T_1 excited states of adenine and guanine cannot be accurately described by ROHF or UHF wavefunction, however.

4.4 Results and discussion

4.4.1 Acetone S_0 and T_1 states

Acetone is used as a simple example to illustrate the MP2/CPCM methods.

The S_0 ground state acetone optimized using the gas phase MP2 and solution phase MP2/CPCM has a trigonal planar skeleton. The T_1 triplet state acetone optimized using the gas phase ZAPT2 and UMP2 and the solution phase ZAPT2/CPCM and UMP2/CPCM shows a trigonal pyramidal skeleton. In general the geometries in the gas phase and water are very similar to each other for the same electronic state. The RMP2 optimized carbonyl C=O bond lengths are 1.215 Å in the gas phase and 1.222 Å in water,

respectively. The 0.007 Å lengthening of the C=O bond is probably due to the CPCM stabilization of the partial negative charges on the O atom. The ZAPT2 optimized carbonyl C=O bond lengths are 1.327 Å in the gas phase and 1.329 Å in water. The UMP2 optimized bond lengths are 1.325 Å in the gas phase and 1.327 Å in water. The 0.002 Å lengthening of the C=O bond is smaller in magnitude than that for the S_0 state. The UHF spin contamination is not significant, with $\mathbf{S}^2 = 2.017$ in both the gas phase and the CPCM solution phase.

The energies calculated using HF and MP2 methods for acetone are listed in Table 4.1. As expected, CPCM always leads to some solvation energy for the variational HF methods. This should be true for all molecules. CPCM implicitly affects $E^{(2)}$ by affecting the HF orbitals and orbital energies. The CPCM effects in $E^{(2)}$ can be positive and negative, depending on how the orbitals and orbital energies are affected. Since $E^{(2)}$ typically contains thousands to millions terms, it is very difficult to determine which terms are more important.

Table 4.1 Energies (hartree) of acetone computed using the aug-cc-pVTZ basis set. The gas phase results in each case are based on the MP2/CPCM optimized geometry in the case. CPCM effect is in kcal/mol.

	E^{HF}	$E^{(2)}$	E^{MP2}	$G_{sol}^{(2)}$	$E^{MP2}+G_{sol}^{(2)}$
RMP2 gas phase	-192.035264	-0.825358	-192.860622	0	-192.860622
RMP2/CPCM	-192.048094	-0.822840	-192.870934	-0.002789	-192.873724
CPCM effect	-8.05	+1.58	-6.47	-1.75	-8.22
ZAPT2 gas phase	-191.942713	-0.769660	-192.712373	0	-192.712373
ZAPT2/CPCM	-191.948211	-0.769946	-192.718157	+0.000527	-192.717630
CPCM effect	-3.45	-0.18	-3.63	+0.33	-3.30
UMP2 gas phase	-191.951690	-0.763264	-192.714954	0	-192.714954
UMP2/CPCM	-191.957147	-0.763617	-192.720765	+0.000593	-192.720172
CPCM effect	-3.42	-0.22	-3.65	+0.37	-3.27

$G_{sol}^{(2)}$ can be either positive or negative. For S_0 acetone, $G_{sol}^{(2)}$ is -0.0027895 hartree or -1.75 kcal/mol. RHF and RMP2 methods were used to calculate the dipole moments of gas phase acetone in S_0 state optimized with the RMP2 method. The RHF and RMP2 results are 3.46 and 2.98 Debye, respectively. If only the dipole moment is considered, a larger dipole moment at the RHF level of theory should cause an overestimation of the magnitude of the CPCM solvation energy, and the $G_{sol}^{(2)}$ should be positive. Therefore, the dipole moment along cannot be used to explain the negative $G_{sol}^{(2)}$.

The analytic gradient of $G_{sol}^{(2)}$ is not available. Geometry optimized on the potential energy surface of E^{MP2} [i.e. $E^{HF} + E^{(2)}$ in both the gas phase and CPCM solution phase] is not necessary the minimum point on the potential energy surface of $E^{MP2}+G_{sol}^{(2)}$. Various tests show that $G_{sol}^{(2)}$ is virtually a constant near a minimum point on the potential energy surface of E^{MP2}, so the minimum point is virtually the minimum point on the potential energy surface of $E^{MP2}+G_{sol}^{(2)}$. Table 4.2 lists the E^{MP2} and $E^{MP2}+G_{sol}^{(2)}$ of the last seven steps in the geometry optimization process of S_0 acetone with the RMP2/CPCM method and that of the T_1 acetone with the ZAPT2/CPCM method. Clearly, $G_{sol}^{(2)}$ is very close to a constant. In the RMP2/CPCM case, the minimum on the E^{MP2} surface has an $E^{MP2}+G_{sol}^{(2)}$ energy that is 5.8×10^{-7} hartree (3.7×10^{-4} kcal/mol) higher than the lowest $E^{MP2}+G_{sol}^{(2)}$ energy encountered in the optimization route. In the ZAPT2/CPCM case, the minimum on the E^{MP2} surface indeed gives the lowest $E^{MP2}+G_{sol}^{(2)}$ energy in the optimization route. The experience shows that in approximately 70% geometry optimization cases, the E^{MP2} minima coincide with the $E^{MP2}+G_{sol}^{(2)}$ lowest energies. For the rest cases, the $E^{MP2}+G_{sol}^{(2)}$ energies at the E^{MP2} minima are always within 10^{-5} hartree

(0.006 kcal/mol) to the lowest $E^{MP2}+G_{sol}^{(2)}$ energies. It is recommend the use of E^{MP2} for geometry optimization, and the use of $E^{MP2}+G_{sol}^{(2)}$ for the total energy.

Table 4.2 Energies (hartree) of S_0 and T_1 state acetone in the last 7 steps in MP2/CPCM/aug-cc-pVTZ geometry optimization processes on the E^{MP2} surface. The lowest energy in each series is in **bold**.

step	RMP2/CPCM		ZAPT2/CPCM	
	E^{MP2}	$E^{MP2}+G_{sol}^{(2)}$	E^{MP2}	$E^{MP2}+G_{sol}^{(2)}$
1	-192.8708802330	-192.8736850812	-192.7181239221	-192.7176017801
2	-192.8709276239	-192.8737193511	-192.7181310854	-192.7176045263
3	-192.8709309931	-192.8737211474	-192.7181383579	-192.7176106079
4	-192.8709334255	**-192.8737241137**	-192.7181469876	-192.7176180919
5	-192.8709339006	-192.8737229130	-192.7181553570	-192.7176276519
6	-192.8709340181	-192.8737233000	-192.7181565504	-192.7176292820
7	**-192.8709340430**	-192.8737235289	**-192.7181568035**	**-192.7176298931**

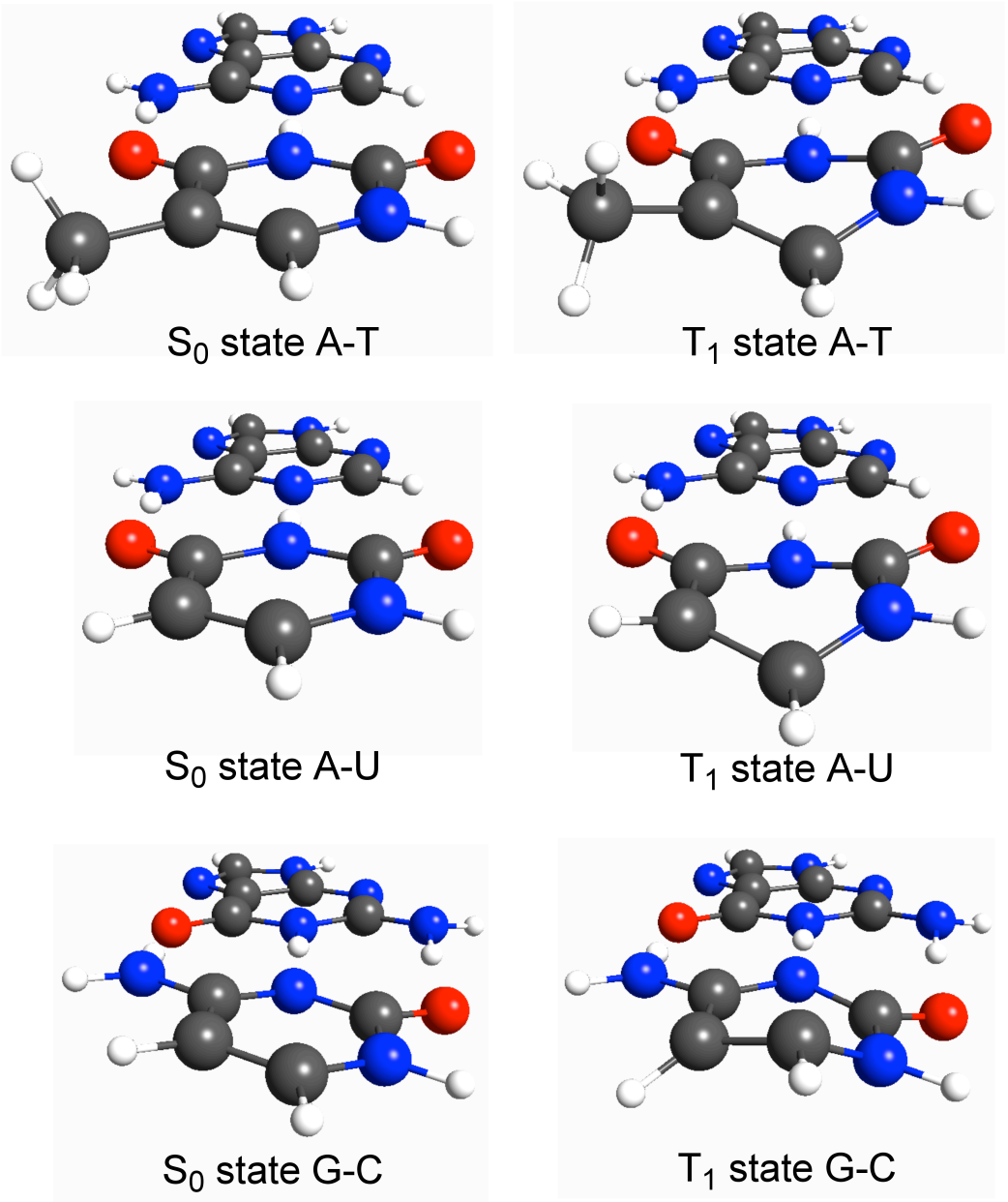


Figure 4.1 Geometries of the S_0 and T_1 states of adenine-thymine (A-T), adenine-uracil (A-U) and guanine-cytosine (G-C) pairs optimized using RMP2/CPCM and ZAPT2/CPCM methods and the aug-cc-pVDZ basis set. The gas phase geometries of these pairs are similar to those in the CPCM solution phase. Free thymine, uracil and cytosine have S_0 and T_1 geometries similar to those in the pairs.

4.4.2 Nucleobases

The S_0 ground state adenine, guanine, cytosine, thymine and uracil all show planar skeletons after gas phase RMP2 optimization (Figure 4.1). Cytosine, thymine and uracil should have C_s symmetry (no symmetry is imposed in the geometry optimization). Adenine and guanine do not have symmetry because their amino groups prefer trigonal pyramidal geometries. In the S_0 ground state, the methyl group of thymine has two H atoms pointing toward the neighboring carbonyl O atom.

After gas phase ZAPT2 and UMP2 optimization, the T_1 excited states of cytosine, thymine and uracil all adapt nonplanar geometries (Figure 4.1). In the T_1 state, the methyl group of thymine has only one H atom pointing toward the neighboring carbonyl O atom. The T_1 excited states of uracil, thymine and cytosine can all be considered as the results of $\pi \rightarrow \pi^*$ excitation of their C=C double bonds, which destroys the C=C double bonds, the aromaticity, and the planarity of these molecules. The changes in the bond lengths clearly indicate that the excitations are localized at the C=C double bonds. In the S_0 states optimized using the gas phase RMP2 method, the C=C double bond length is 1.364, 1.366 and 1.371 Å, respectively, in uracil, thymine and cytosine. In the T_1 states optimized using the gas phase ZAPT2 method these bond lengths are 1.500, 1.505 and 1.492 Å, respectively. Other bond lengths change by at most 0.03 Å.

The solution phase RMP2/CPCM and ZAPT2/CPCM optimized S_0 and T_1 state geometries are very similar to those in the gas phase. CPCM solvation can alter the bond lengths by as much as 0.01 Å. In the gas phase S_0 and T_1 states, the amino group of cytosine has a trigonal pyramidal geometry, while in the CPCM solution phases S_0 and T_1 states, it has a trigonal planar geometry.

Table 4.3 Energy (hartree) of optimized nucleobases in the gas phase (E^{MP2}) and solution phase ($E^{MP2}+G_{sol}^{(2)}$) using the aug-cc-pVDZ basis set. CPCM effect is in kcal/mol.

	adenine	guanine	cytosine	thymine	uracil
RMP2 gas phase	-466.095966	-541.177079	-393.911022	-452.973714	-413.779025
RMP2/CPCM	-466.123432	-541.222339	-393.949308	-453.004176	-413.810450
CPCM effect	-17.24	-28.40	-24.02	-19.12	-19.72
ZAPT2 gas phase			-393.774560	-452.846765	-413.647897
ZAPT2/CPCM			-393.809807	-452.873405	-413.674662
CPCM effect			-22.12	-16.72	-16.80
UMP2 gas phase			-393.776787	-452.846742	-413.648774
UMP2/CPCM			-393.812780	-452.871872	-413.674618
CPCM effect			-22.59	-15.77	-16.22

The CPCM effects in the T_1 state cytosine, thymine and uracil are 1.91, 2.40 and 2.92 kcal/mol (ZAPT2 results in Table 4.3) smaller in magnitude than those for the S_0 states (RMP2 results in Table 4.3). The reduction in the CPCM solvation energy is mainly caused by the changes in the electronic structure, with contributions from geometry relaxation.

The UMP2 results are very similar to the ZAPT2 results. The UHF spin contamination is not significant, with $\mathbf{S}^2$ being approximately 2.05 in all the cases.

4.4.3 Nucleobase pairs

The S_0 ground state adenine-thymine (A-T), adenine-uracil (A-U) and guanine-cytosine (G-C) hydrogen bonding pairs are optimized with the gas phase RMP2 method (Figure 4.1). In all of these three S_0 state pairs, the geometries of adenine, guanine, thymine, uracil and cytosine are very similar to those of their free S_0 states. The A-T pair has a planar skeleton with two H atoms of the thymine methyl group sticking out of the plane; the A-U pair is planar; the G-C pair is not planar because the amino group of

guanine prefers a trigonal pyramidal geometry (the dihedral angle formed by the two aromatic planes is ~5.5 degree).

Table 4.4 Energy (hartree) of optimized nucleobase pairs in the gas phase (E^{MP2}) and solution phase ($E^{MP2}+ G_{sol}^{(2)}$) using the aug-cc-pVDZ basis set. The CPCM effect and formation energy are in kcal/mol.

	adenine-uracil		adenine-thymine		guanine-cytosine	
	absolute	formation	absolute	formation	absolute	formation
RMP2 gas phase	-879.903366	-17.80	-919.097953	-17.74	-935.136344	-30.27
RMP2/CPCM	-879.946358	-7.83	-919.140111	-7.85	-935.188155	-10.36
CPCM effect	-26.98	9.97	-26.45	9.89	-32.51	19.91
ZAPT2 gas phase	-879.770446	-16.68	-918.969335	-16.69	-934.999236	-29.86
ZAPT2/CPCM	-879.809274	-7.01	-919.008226	-7.15	-935.047277	-9.49
CPCM effect	-24.36	9.67	-24.41	9.54	-30.15	20.37

The T_1 excited states of A-T, A-U and G-C pairs can also be described by ROHF, UHF, ZAPT2 and UMP2 methods because the excitations are localized in thymine, uracil and cytosine. Here only the ZAPT2 method is used because the UMP2 method is more costly. After the gas phase ZAPT2 optimization, the T_1 state A-T, A-U and G-C pairs adapt nonplanar geometries (Figure 4.1). In all of these three T_1 state pairs, the geometries of adenine and guanine are very similar to those of their free S_0 states, and the geometries of thymine, uracil and cytosine are very similar to those of their free T_1 states. Therefore, these T_1 excited state pairs can be considered as dimers formed by S_0 state adenine/guanine and T_1 state uracil/thymine/cytosine. In the S_0 state pairs optimized using the gas phase RMP2 method the C=C double bond lengths are 1.365, 1.367 and 1.369 Å, respectively, in uracil, thymine and cytosine. In the T_1 state optimized using the gas phase ZAPT2 method these bond lengths are 1.498, 1.505 and 1.492 Å, respectively.

These values are almost identical to those in the free uracil, thymine and cytosine. In the pairs, other covalent bond lengths change at most by 0.03 Å from S_0 to T_1, while the hydrogen bond lengths change more significantly.

The geometries after solution phase RMP2/CPCM and ZAPT2/CPCM optimization are very similar to those in the gas phase. CPCM solvation can alter covalent bond lengths by as large as 0.03 Å, and hydrogen bond lengths by as large as 0.14 Å. For example, in the S_0 state A-U pair, the middle hydrogen bond is shortened by 0.10 Å, while the other hydrogen bond is lengthened by 0.14 Å, ongoing from the gas phase to CPCM.

The gas phase formation energy of the S_0 ground state G-C pair is -30.27 kcal/mol as calculated from the difference between the gas phase RMP2 optimized energy of the G-C pair (Table 4.4) and the sum of the RMP2 optimized monomer energies (Table 4.3). The solution phase formation energy of the S_0 state G-C pair is -10.36 kcal/mol as calculated from the RMP2/CPCM optimized energies. So, the formation of the S_0 state G-C pair in water is 19.91 kcal/mol less favorable than that in the gas phase because water tends to dissolve the pair. The gas phase formation energy of the T_1 state G-C pair is -29.86 kcal/mol as calculated from the difference between the gas phase ZAPT2 optimized energy of the G-C pair and the sum of the RMP2 optimized energy of S_0 guanine and the ZAPT2 optimized energy of T_1 cytosine. The solution phase formation energy of the T_1 state G-C pair is -9.49 kcal/mol. The values for the A-U and A-T pairs are listed in Table 4.4 for comparison.

Clearly, in both the gas phase and the solution phase the hydrogen bonds in the T_1 state A-U, A-T and G-C pairs is ~1 kcal/mol weaker than those in the S_0 state pairs, and

the CPCM solvation does not appear to significantly affect the weakening of the hydrogen bonding ongoing from S_0 to T_1 state. Aug-cc-pVQZ[74] interaction analysis was performed using the LMO-EDA method[154] for select pairs, and found that the basis set superposition error (BSSE) in the formation energy calculated using the MP2/aug-cc-pVDZ methods is approximately 4 kcal/mol. It is noted that the BSSE do not significantly affect the relative energies discussed above.

The CPCM solvation effects in the T_1 state G-C, A-T and A-U pairs are 2.36, 2.05 and 2.61 kcal/mol smaller in magnitude than those for the S_0 states (Table 4.4). These differences are similar to those (1.91, 2.40 and 2.92 kcal/mol) found for free cytosine, thymine and uracil, implying that they are caused by local changes in the electronic structures and molecular geometries of these molecules.

4.5 Summary

Analytic energy gradients are derived and implemented for MP2/CPCM methods. Very accurate gradient and smooth and continuous potential energy surfaces are obtained by using the fixed points with variable areas (FIXPVA) surface tessellation scheme for CPCM. All closed shell and open shell MP2 methods (RMP2, ZAPT2 and UMP2, serial and parallel programs) in GAMESS are enabled to perform MP2/CPCM gradient calculations. A better CPCM solvation energy can be obtained efficiently by using the MP2 relaxed density. The open shell ZAPT2 and ZAPT2/CPCM methods are used to study the T_1 excited states of three nucleobase pairs (adenine-thymine, adenine-uracil and guanine-cytosine) in the gas phase and aqueous solution phase. It is found that in both the gas phase and solution phase, the hydrogen bonds in the T_1 excited state base pairs are ~1 kcal/mol weaker than those in the S_0 states.

Chapter 5 Analytic energy gradient in combined time-dependent density functional theory and polarizable force field calculation

5.1 Introduction

Besides open shell MP2 methods introduced in Chapter 4, the time-dependent density functional theory (TDDFT) method formulated by Casida *et al.*[19] is another efficient and relatively accurate QM method for studying valence and singly excited states. This chapter formulates the combined quantum mechanical and molecular mechanical (TDDFT/MM) style methods for excited state studies with the presence of solvent. Using a polarizable force field in QM/MM calculations is advantageous because the electronic polarization of the MM region can be described, especially when electronic excitation is involved in the QM region. A conceptually simple and efficient method to incorporate electronic polarization into the MM region is to use dipole polarizability or induced dipoles.[155] Several polarizable QM/MM style excited state methods have been developed by Öhrn and Karlström,[39] Lin and Gao,[40] Kongsted, Osted and Mikkelsen,[41] Nielsen, Christiansen, Mikkelsen and Kongsted,[42] Jensen, van Duijnen, Snijders and Netzel,[43,44] and Yoo, Zahariev, Sok and Gordon.[45] Li *et al*. implemented analytic gradient for a QM/Continuum style TDDFT method that incorporates induced surface charges.[156] The analytic gradient is implemented for a polarizable TDDFT/MM method using a polarizable water model that incorporates induced dipoles, and is used to study a cluster formed by acetone and water.

5.2 Theory

5.2.1 Review of density functional methods

The concept of density functional theory (DFT)[22,24] was originally founded by E. Fermi[28] and L. H. Thomas[29] in the late 1920s. The wide application of DFT became reality after the work of Hohenberg, Kohn and Sham[157] in the 1960s. Modern DFT methods are based on the scheme by Kohn and Sham in 1965 that the kinetic energy functional can be divided into two components: one part can be calculated exactly and treats electrons as non-interacting particles (in the same sense that HF orbitals describing non-interacting electrons). The other part is a small correction term representing electron-electron correlation. The Kohn-Sham (KS) model is a wave function based method (with 3N variables) and shares identical formulas with HF method for the kinetic, electron-nuclear and Coulomb energies. It uses the exchange-correlation term, to represent the correction to the kinetic energy due to the interacting nature of the electrons, and all nonclassic corrections to the electron-electron repulsion energy. The resolution of the Kohn-Sham equation is also processed in a self-consistent fashion. DFT is able to calculate electron correlation in a less costly way as compared to other multi-determinants methods with a similar accuracy.

The functional for exchange-correlation energy can be of three types: The first is local functionals, depending only on the electron density ρ, such as Thomas-Fermi-Dirac method,[158] VWN[159] and PL.[160] The second type is called gradient corrected functionals (GGA), which depends on both ρ and its gradient $\Delta\rho$. For example, the first broadly used GGA exchange functional developed by Becke,[161-164] and later Becke88(B),[92] PW(Perdew-Wang),[165] mPW(modified-Perdew-Wang),[165,166] OptX,[167] X,[168] B86,[161] P

(Perdew86),[169] PBE (Perdew-Burke-Ernzerhof),[170] mPBE.[170,171] The correlation functionals include B88,[172] Perdew 86[169] and PW91 (Perdew-Wang).[173] Another extremely popular GGA correlation function is LYP (Lee-Yang-Parr).[93] The third type is hybrid density functional (H-GGA) methods that mix a certain percentage of Hartree-Fock (or exact) exchange into a conventional GGA method. The weighting factors are determined semiempirically according to experimental atomization energies, ionization potentials, proton affinities, total atomic energies or other data. Some widely used hybrid density functionals include B3LYP,[92,93,164] B3P86,[92,164,165] B3PW91,[92,164,173] B97-1,[174] B97-2,[174] B98,[175] BH&HLYP,[92,93] MPW1K,[166,173,176] mPW3LYP,[93,166,173] O3LYP[93,177] and X3LYP.[92,93,163,168] The general performances of density functionals were summarized in the reviews.[178]

5.2.2 Review of polarizable DFT/MM gradient

In a polarizable QM/MM style DFT (including Hartree-Fock as a special case) calculation, the total energy of the system can be written as

$$E_{DFT/MM} = G_{MM} + E_{DFT} + G_{rep} + G_{disp} + G_{mul} + G_{pol}. \tag{5-1}$$

G_{MM} is the force field energy of the whole QM/MM system, including that within the MM region (such as bond stretching, bending, torsion, electrostatic, repulsion and dispersion), that within the QM region (if any), and that between the QM and MM regions (for example, force field type repulsion and dispersion energies between a QM atom and a MM atom).

E_{DFT} is the DFT electronic (including nuclear repulsion) energy of the QM region,

$$\begin{aligned} E_{DFT} = & \sum_{\mu\nu\sigma} P_{\mu\nu\sigma} h_{\mu\nu} + \sum_{\mu\nu\sigma\lambda\kappa\sigma'} P_{\mu\nu\sigma} P_{\lambda\kappa\sigma'} (\mu\nu|\lambda\kappa) \\ & - c_x \delta_{\sigma\sigma'} \sum_{\mu\nu\sigma\lambda\kappa\sigma'} P_{\mu\nu\sigma} P_{\lambda\kappa\sigma'} (\mu\lambda|\nu\kappa) + \sum_{\mu\nu\sigma\lambda\kappa\sigma'} P_{\mu\nu\sigma} P_{\lambda\kappa\sigma'} f^{XC}_{\mu\nu\sigma,\lambda\kappa\sigma'} + E_{NN} \end{aligned} \tag{5-2}$$

where μ, ν, λ and κ denote basis functions (Mulliken notation is used for two-electron integrals); σ and σ' (and later σ'') are spin indices; **P** is the one-particle density matrix; **h** is the one-electron integral; c_x is the mixing coefficient of the Hartree-Fock exchange energy; $f^{XC}_{\mu\nu\sigma,\lambda\kappa\sigma'}$ is the DFT exchange-correlation kernel; E_{NN} is the nuclear repulsion energy.

G_{rep} is the repulsion energy between QM electrons and MM repulsion points; G_{disp} is the dispersion energy between QM electrons and MM dispersion points; G_{mul} is the electrostatic interaction energy between QM electrons/nuclei and MM electrostatic multipole points (such as charges and dipoles),

$$G_{rep} + G_{disp} + G_{mul} = \sum_{\mu\nu\sigma} P_{\mu\nu\sigma} V_{rep,\mu\nu} + \sum_{\mu\nu\sigma} P_{\mu\nu\sigma} V_{disp,\mu\nu} + \sum_{\mu\nu\sigma} P_{\mu\nu\sigma} V_{mul,\mu\nu} + E_{N,mul} \tag{5-3}$$

where $\mathbf{V}_{rep,\mu\nu}$, $\mathbf{V}_{disp,\mu\nu}$ and $\mathbf{V}_{mul,\mu\nu}$ are the one-electron integrals over the MM repulsion, dispersion and multipole potentials; $E_{N,mul}$ is the electrostatic energy between QM nuclei and MM multipole points.

G_{pol} is the polarization energy of the whole QM/MM system,

$$\begin{aligned} G_{pol} &= -\tfrac{1}{2}\mathbf{F}^T\mathbf{d} \\ &= -\tfrac{1}{2}\left(\mathbf{F}_N + \mathbf{F}_{mul} + \mathbf{F}_e\right)^T\left(\mathbf{d}_N + \mathbf{d}_{mul} + \mathbf{d}_e\right) \\ &= -\tfrac{1}{2}\left(\mathbf{F}_N + \mathbf{F}_{mul}\right)^T\left(\mathbf{d}_N + \mathbf{d}_{mul}\right) - \tfrac{1}{2}\sum_{\mu\nu\sigma\lambda\kappa\sigma'} P_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}\left(\mathbf{F}_{\mu\nu}\right)^T\mathbf{d}_{\lambda\kappa} \\ &\quad -\tfrac{1}{2}\sum_{\mu\nu\sigma} P_{\mu\nu\sigma}\left[\left(\mathbf{F}_N + \mathbf{F}_{mul}\right)^T\mathbf{d}_{\mu\nu} + \left(\mathbf{F}_{\mu\nu}\right)^T\left(\mathbf{d}_N + \mathbf{d}_{mul}\right)\right] \end{aligned} \tag{5-4}$$

where **F** is the electrostatic fields at the dipole polarizability points created by QM nuclei ($\mathbf{F}_N$), electrons ($\mathbf{F}_e$) and MM electric multipoles ($\mathbf{F}_{mul}$); the superscript T denotes transpose here and hereafter; **d** is a set (vector) of induced dipoles, which satisfies the following linear response equation:

$$\mathbf{D}\cdot\mathbf{d}=\mathbf{F} \tag{5-5}$$

The **D** matrix is constructed using the nine-component dipole polarizability tensors and the Cartesian coordinates of the polarizability points.[51,179,180] Due to the linearity of Eq. (5-5), **d** can be determined separately for QM nuclei ($\mathbf{d}_N$,), electrons ($\mathbf{d}_e$) and MM multipoles ($\mathbf{d}_{\mathrm{mul}}$). In addition, the induced dipoles can be determined for the electrostatic field due to a product of two basis functions such as μν, for example, $\mathbf{d}_{\mu\nu}$ and $\mathbf{F}_{\mu\nu}$ in Eq. (5-4).

Once the energy in Eq. (5-1) is variationally minimized by adjusting the density matrix via a standard self-consistent field procedure, Pulay's method[181] can be used to determine the first derivative (gradient) of the total DFT/MM energy with respect to a coordinate *x*:

$$\begin{aligned}
E^x_{DFT/MM} &= G^x_{MM}+E^x_{DFT}+G^x_{rep}+G^x_{disp}+G^x_{mul}+G^x_{pol}\\
&= G^x_{MM}-\sum_{\mu\nu\sigma}W_{\mu\nu\sigma}S^x_{\mu\nu}+\sum_{\mu\nu\sigma}P_{\mu\nu\sigma}h^x_{\mu\nu}+\sum_{\mu\nu\sigma\lambda\kappa\sigma'}P_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}\left(\mu\nu|\lambda\kappa\right)^x\\
&\quad -c_x\delta_{\sigma\sigma'}\sum_{\mu\nu\sigma\lambda\kappa\sigma'}P_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}\left(\mu\lambda|\nu\kappa\right)^x+\sum_{\mu\nu\sigma\lambda\kappa\sigma'}P_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}f^{XC,x}_{\mu\nu\sigma,\lambda\kappa\sigma'}+E^x_{NN}\\
&\quad +\sum_{\mu\nu\sigma}P_{\mu\nu\sigma}V^x_{rep,\mu\nu}+\sum_{\mu\nu\sigma}P_{\mu\nu\sigma}V^x_{disp,\mu\nu}+\sum_{\mu\nu\sigma}P_{\mu\nu\sigma}V^x_{mul,\mu\nu}+E^x_{N,mul}\\
&\quad -\frac{1}{2}\left[\left(\boldsymbol{F}_N+\boldsymbol{F}_{mul}\right)^T\left(\boldsymbol{d}_N+\boldsymbol{d}_{mul}\right)\right]^x-\frac{1}{2}\sum_{\mu\nu\sigma\lambda\kappa\sigma'}P_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}\left[\left(\boldsymbol{F}_{\mu\nu}\right)^T\boldsymbol{d}_{\lambda\kappa}\right]^x\\
&\quad -\frac{1}{2}\sum_{\mu\nu\sigma}P_{\mu\nu\sigma}\left[\left(\boldsymbol{F}_N+\boldsymbol{F}_{mul}\right)^T\boldsymbol{d}_{\mu\nu}+\left(\boldsymbol{F}_{\mu\nu}\right)^T\left(\boldsymbol{d}_N+\boldsymbol{d}_{mul}\right)\right]^x
\end{aligned} \tag{5-6}$$

where **W** is the energy-weighted density matrix, **S** is the overlap matrix.

It is noted that for asymmetric polarization cases where $\left(\mathbf{D}^{-1}\right)^T\neq\mathbf{D}^{-1}$, there is a general expression for the derivative of the product of any electric field $\mathbf{F}_K$ and induced dipoles $\mathbf{d}_L$,

$$\begin{aligned}\left[\left(\mathbf{F}_K\right)^T\mathbf{d}_L\right]^x &= \left(\mathbf{F}_K^x\right)^T\mathbf{d}_L + \mathbf{F}_K\left(\mathbf{D}^{-1}\mathbf{F}_L\right)^x \\ &= \left(\mathbf{F}_K^x\right)^T\mathbf{d}_L + \mathbf{F}_K\mathbf{D}^{-1}\mathbf{D}^x\mathbf{D}^{-1}\mathbf{F}_L + \mathbf{F}_K\mathbf{D}^{-1}\mathbf{F}_L^x \\ &= \left(\mathbf{F}_K^x\right)^T\mathbf{d}_L + \left(\tilde{\mathbf{d}}_K\right)^T\mathbf{D}^x\mathbf{d}_L + \left(\tilde{\mathbf{d}}_K\right)^T\mathbf{F}_L^x\end{aligned} \tag{5-7}$$

with

$$\tilde{\mathbf{d}}_K = \left(\mathbf{D}^{-1}\right)^T\mathbf{F}_K \tag{5-8}$$

So the last three terms in Eq. (5-6) can be written as [see also Eq. (5-4)]

$$-\frac{1}{2}\left[\left(\mathbf{F}^x\right)^T\mathbf{d} + \left(\tilde{\mathbf{d}}\right)^T\mathbf{D}^x\mathbf{d} + \left(\tilde{\mathbf{d}}\right)^T\mathbf{F}^x\right], \tag{5-9}$$

which represents the force and torque imposed on the induced dipoles by the QM nuclei, electrons and MM multipoles, and the force and torque between the induced dipoles. In addition, Li implemented Eq. (5-6) for a polarizable DFT/MM method.[51]

5.2.3 Polarizable TDDFT/MM gradient

In a combined polarizable QM/MM style TDDFT (including time-dependent Hartree-Fock or random phase approximation as a special case) method, such as that implemented by Jensen *et al.*,[43] Nielsen *et al.*[42] and Yoo *et al.*[45], the MM region should affect the TDDFT excitation energy in two ways. First, the MM region, including induced dipoles, affects the ground state DFT spin orbitals and energies, as discussed above. Second, the MM induced dipoles (and only the induced dipoles) directly affect the TDDFT excitation energies, as discussed below.

The central equation in combined QM/MM style TDDFT calculation of the excitation energy is the same as that for regular TDDFT calculation,

$$\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} = \omega\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}, \tag{5-10}$$

where **X** and **Y** are the linear response amplitudes in the excitation and deexcitation processes; ω is the excitation energy, **A** and **B** matrices are the orbital rotation Hessian,

$$\begin{aligned} A_{ia\sigma,jb\sigma'} &= \delta_{ij}\delta_{ab}\delta_{\sigma\sigma'}\left(\varepsilon_{a\sigma}-\varepsilon_{i\sigma}\right)+\left(ia\sigma|bj\sigma'\right)-c_x\delta_{\sigma\sigma'}\left(ib\sigma|aj\sigma'\right) \\ &\quad + f^{XC}_{ia\sigma bj\sigma'} - \frac{1}{2}\left(\mathbf{F}_{ia\sigma}\right)^T\left(\tilde{\mathbf{d}}_{jb\sigma'}+\mathbf{d}_{jb\sigma'}\right) \end{aligned} \tag{5-11}$$

$$B_{ia\sigma,jb\sigma'} = \left(ia\sigma|bj\sigma'\right)-c_x\delta_{\sigma\sigma'}\left(ij\sigma|ab\sigma'\right)+f^{XC}_{ia\sigma bj\sigma'} - \frac{1}{2}\left(\mathbf{F}_{ia\sigma}\right)^T\left(\tilde{\mathbf{d}}_{jb\sigma'}+\mathbf{d}_{jb\sigma'}\right) \tag{5-12}$$

i and *j* represent occupied orbitals, *a* and *b* represent virtual orbitals, ε represents orbital energies. Here both $\mathbf{d}_{jb\sigma'}$ and $\tilde{\mathbf{d}}_{jb\sigma'}$ are used for general asymmetric polarization. Many-body mutual polarization is included in the TDDFT step.[43,182]

The first derivative of the TDDFT excitation energy ω with respect to a coordinate *x* is:

$$\begin{aligned} \omega^x &= \sum_{\mu\nu\sigma} P^{\Delta}_{\mu\nu\sigma} h^x_{\mu\nu} - \sum_{\mu\nu\sigma} W^{\Delta}_{\mu\nu\sigma} S^x_{\mu\nu} + \sum_{\mu\nu\sigma\lambda\kappa\sigma'} \Gamma^{\Delta}_{\mu\nu\sigma\lambda\kappa\sigma'} (\mu\nu|\lambda\kappa)^x \\ &\quad + \sum_{\mu\nu\sigma\lambda\kappa\sigma'} P^{\Delta}_{\mu\nu\sigma} P_{\lambda\kappa\sigma'} \left(f^{XC}_{\mu\nu\sigma,\lambda\kappa\sigma'}\right)^x + \sum_{\mu\nu\sigma\lambda\kappa\sigma'} (X+Y)_{\mu\nu\sigma}(X+Y)_{\lambda\kappa\sigma'}\left(f^{XC}_{\mu\nu\sigma,\lambda\kappa\sigma'}\right)^x \\ &\quad + \sum_{\mu\nu\sigma} P^{\Delta}_{\mu\nu\sigma} V^x_{rep,\mu\nu} + \sum_{\mu\nu\sigma} P^{\Delta}_{\mu\nu\sigma} V^x_{disp,\mu\nu} + \sum_{\mu\nu\sigma} P^{\Delta}_{\mu\nu\sigma} V^x_{mul,\mu\nu} \\ &\quad - \frac{1}{2}\sum_{\mu\nu\sigma\lambda\kappa\sigma'} \left(P^{\Delta}_{\mu\nu\sigma} P_{\lambda\kappa\sigma'} + P_{\mu\nu\sigma} P^{\Delta}_{\lambda\kappa\sigma'}\right)\left[\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^T\left(\tilde{\mathbf{d}}_{\lambda\kappa}+\mathbf{d}_{\lambda\kappa}\right)\right]^x \\ &\quad - \frac{1}{2}\sum_{\mu\nu\sigma} P^{\Delta}_{\mu\nu\sigma}\left[\frac{1}{2}\left(\mathbf{F}_N+\mathbf{F}_{mul}\right)^T\left(\tilde{\mathbf{d}}_{\mu\nu}+\mathbf{d}_{\mu\nu}\right)+\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^T\left(\tilde{\mathbf{d}}_N+\tilde{\mathbf{d}}_{mul}+\mathbf{d}_N+\mathbf{d}_{mul}\right)\right]^x \\ &\quad - \sum_{\mu\nu\sigma\lambda\kappa\sigma'} (X+Y)_{\mu\nu\sigma}(X+Y)_{\lambda\kappa\sigma'}\left[\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^T\left(\tilde{\mathbf{d}}_{\lambda\kappa}+\mathbf{d}_{\lambda\kappa}\right)\right]^x \end{aligned} \tag{5-13}$$

where $\mathbf{P}^{\Delta}$ is the relaxed one-particle difference density matrix; **P** is the ground state one-particle density matrix; $\mathbf{W}^{\Delta}$ is the energy-weighted difference density matrix; $\mathbf{\Gamma}^{\Delta}$ is the two-particle difference density matrix, including the Hartree-Fock exchange contribution in hybrid DFT methods.

The first five terms in Eq. (5-13) have exactly the same forms as those in the regular TDDFT excitation energy gradient formula derived by Furche and Ahlrichs.[183] It is noted that Van Caillie and Amos derived similar formulas.[184] The derivation of Eq. (5-13) is straightforward if one understands that the induced dipole and DFT corrections to the Hartree-Fock energy and gradient are very similar, because both of them are reaction fields that reduce electron-electron repulsion (through polarization and DFT exchange-correlation) and reduce electron-nucleus attraction (only through polarization). Therefore, in Eq. (5-13), $V_{rep,\mu\nu}$ $V_{disp,\mu\nu}$ and $V_{mul,\mu\nu}$ correspond to the one-electron nuclear charge integral in $h_{\mu\nu}$, and $\left(\mathbf{F}_{\mu\nu}\right)^T\left(\tilde{\mathbf{d}}_{\lambda\kappa}+\mathbf{d}_{\lambda\kappa}\right)$ corresponds to $f^{XC}_{\mu\nu\sigma,\lambda\kappa\sigma'}$. A similar comparison between quantum chemical reaction field continuum solvation model and DFT was made.[156]

Here it will be shown that, by adding induced dipoles, Furche and Ahlrichs' formulas[183] for regular TDDFT can be used to determine the $\mathbf{P}^{\Delta}$, $\mathbf{W}^{\Delta}$ and $\mathbf{\Gamma}^{\Delta}$ for QM/MM style TDDFT. Same as for regular TDDFT, $\mathbf{P}^{\Delta}$ in Eq. (5-13) is the sum of the unrelaxed difference density matrix $\mathbf{T}^{\Delta}$ and the $\mathbf{Z}$ density matrix,

$$\mathbf{P}^{\Delta}=\mathbf{T}^{\Delta}+\mathbf{Z}. \tag{5-14}$$

The unrelaxed difference density matrix $\mathbf{T}^{\Delta}$,

$$T^{\Delta}_{ab\sigma}=\frac{1}{2}\sum_{i}\left[\left(X+Y\right)_{ia\sigma}\left(X+Y\right)_{ib\sigma}+\left(X-Y\right)_{ia\sigma}\left(X-Y\right)_{ib\sigma}\right] \tag{5-15}$$

$$T^{\Delta}_{ij\sigma}=-\frac{1}{2}\sum_{a}\left[\left(X+Y\right)_{ia\sigma}\left(X+Y\right)_{ja\sigma}+\left(X-Y\right)_{ia\sigma}\left(X-Y\right)_{ja\sigma}\right] \tag{5-16}$$

$$T^{\Delta}_{ia\sigma}=T^{\Delta}_{ai\sigma}=0 \tag{5-17}$$

is defined with the **X** and **Y** from the QM/MM style TDDFT calculation as shown by Eq. (5-10). The **Z** matrix, when written as a vector (Lagrange multiplier), can be determined using the Z-vector equation[140]

$$\sum_{jb\sigma'}(A+B)_{ia\sigma jb\sigma'}Z_{jb\sigma'} = -R_{ia\sigma}\,, \tag{5-18}$$

where **A** and **B** matrices are the orbital rotation Hessian, which includes induced dipoles, as shown by Eq. (5-11) and (5-12); the Lagrangian **R**,

$$\begin{aligned} R_{ia\sigma} = &\sum_{b}\left[(X+Y)_{ib\sigma}H^{+}_{ab\sigma}[X+Y]+(X-Y)_{ib\sigma}H^{-}_{ab\sigma}[X-Y]\right] \\ &-\sum_{j}\left[(X+Y)_{ja\sigma}H^{+}_{ji\sigma}[X+Y]+(X-Y)_{ja\sigma}H^{-}_{ji\sigma}[X-Y]\right] \\ &+H^{+}_{ia\sigma}[T^{\Delta}]+2\sum_{jb\sigma'kc\sigma''}g^{XC}_{ia\sigma jb\sigma' kc\sigma''}(X+Y)_{jb\sigma'}(X+Y)_{kc\sigma''} \end{aligned} \tag{5-19}$$

is also defined with the **X** and **Y** from the QM/MM style TDDFT calculation as shown by Eq. (5-10), and with $g^{XC}_{ia\sigma jb\sigma' kc\sigma''}$ being a matrix element of the third order derivative of the density functional, which has exactly the same form as in regular TDDFT cases. The transformed vectors $\mathbf{H}^{+}$ and $\mathbf{H}^{-}$ originally defined by Furche and Ahlrichs for arbitrary vector **V** can be written as

$$H^{+}_{pq\sigma}[V] = \sum_{rs\sigma'}\begin{Bmatrix} 2(pq\sigma|rs\sigma') + 2f^{XC}_{pq\sigma rs\sigma'} - (\mathbf{F}_{pq})^{T}(\tilde{\mathbf{d}}_{rs}+\mathbf{d}_{rs}) \\ -c_x\delta_{\sigma\sigma'}\left[(ps\sigma|rq\sigma)+(pr\sigma|sq\sigma)\right] \end{Bmatrix} V_{rs\sigma}\,, \tag{5-20}$$

$$H^{-}_{pq\sigma}[V] = \sum_{rs\sigma'}\left\{c_x\delta_{\sigma\sigma'}\left[(ps\sigma|rq\sigma)-(pr\sigma|sq\sigma)\right]\right\}V_{rs\sigma}\,. \tag{5-21}$$

where *r* and *s* run over all occupied and virtual molecular orbitals, while *p* and *q* run over select occupied and virtual molecular orbitals. Compared to regular TDDFT cases, $\mathbf{H}^{+}[\mathbf{V}]$ now contains induced dipoles.

Once **Z** and $\mathbf{P}^{\Delta}$ are available, $\mathbf{W}^{\Delta}$ and $\mathbf{\Gamma}^{\Delta}$ can be evaluated using Furche and Ahlrichs' formulas,[183]

$$W^{\Delta}_{ij\sigma} = \left(1-\frac{1}{2}\delta_{ij}\right)\left\{\begin{array}{l} +\sum_{a}\omega\left[(X+Y)_{ia\sigma}(X-Y)_{ja\sigma}+(X-Y)_{ia\sigma}(X+Y)_{ja\sigma}\right] \\ -\sum_{a}\varepsilon_{a\sigma}\left[(X+Y)_{ia\sigma}(X+Y)_{ja\sigma}+(X-Y)_{ia\sigma}(X-Y)_{ja\sigma}\right] \\ +H^{+}_{ij\sigma}[P^{\Delta}] \\ +2\sum_{kc\sigma' ld\sigma''} g^{xc}_{ij\sigma kc\sigma' ld\sigma''}(X+Y)_{kc\sigma'}(X+Y)_{ld\sigma''} \end{array}\right\} \tag{5-22}$$

$$W^{\Delta}_{ab\sigma} = \left(1-\frac{1}{2}\delta_{ab}\right)\left\{\begin{array}{l} +\sum_{i}\omega\left[(X+Y)_{ia\sigma}(X-Y)_{ib\sigma}+(X-Y)_{ia\sigma}(X+Y)_{ib\sigma}\right] \\ +\sum_{i}\varepsilon_{i\sigma}\left[(X+Y)_{ia\sigma}(X+Y)_{ib\sigma}+(X-Y)_{ia\sigma}(X-Y)_{ib\sigma}\right] \end{array}\right\} \tag{5-23}$$

$$W^{\Delta}_{ia\sigma} = \sum_{j}\left\{(X+Y)_{ja\sigma}H^{+}_{ji\sigma}[X+Y]+(X-Y)_{ia\sigma}H^{-}_{ji\sigma}[X-Y]\right\}+\varepsilon_{i\sigma}Z_{ia\sigma} \tag{5-24}$$

$$\begin{aligned}\Gamma^{\Delta}_{\mu\nu\sigma\kappa\lambda\sigma'} = {} & \frac{1}{2}\left[2P^{\Delta}_{\mu\nu\sigma}P_{\kappa\lambda\sigma'}+2(X+Y)_{\mu\nu\sigma}(X+Y)_{\kappa\lambda\sigma'}\right] \\ & -c_x\delta_{\sigma\sigma'}\left[\begin{array}{l} +P^{\Delta}_{\mu\lambda\sigma}P_{\kappa\nu\sigma'}+P^{\Delta}_{\mu\kappa\sigma}P_{\lambda\nu\sigma'} \\ +(X+Y)_{\mu\lambda\sigma}(X+Y)_{\kappa\nu\sigma'}+(X+Y)_{\mu\kappa\sigma}(X+Y)_{\lambda\nu\sigma'} \\ -(X-Y)_{\mu\lambda\sigma}(X+Y)_{\kappa\nu\sigma'}+(X-Y)_{\mu\kappa\sigma}(X-Y)_{\lambda\nu\sigma'} \end{array}\right],\end{aligned} \tag{5-25}$$

with the **X** and **Y** from the QM/MM style TDDFT calculation as shown by Eq. (5-10). It is noted again that the transformed vector $\mathbf{H}^{+}$ in Eq. (5-22) and (5-24) contains induced dipoles [see Eq. (5-20)].

The sixth, seventh and eighth terms in Eq. (5-13) represent the forces between the TDDFT difference density and MM repulsion/dispersion/multipole points. These three forces are analogs of the force between the TDDFT difference density and QM nuclei, and can be evaluated using similar techniques.

Using Eq. (5-7), the ninth and tenth terms in Eq. (5-13) can be combined and simplified:

$$
\begin{aligned}
&-\frac{1}{2}\sum_{\mu\nu\sigma\lambda\kappa\sigma'}\left(P^{\Delta}_{\mu\nu\sigma}P_{\lambda\kappa\sigma'}+P_{\mu\nu\sigma}P^{\Delta}_{\lambda\kappa\sigma'}\right)\left[\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^{T}\left(\tilde{\mathbf{d}}_{\lambda\kappa}+\mathbf{d}_{\lambda\kappa}\right)\right]^{x}\\
&-\frac{1}{2}\sum_{\mu\nu\sigma}P^{\Delta}_{\mu\nu\sigma}\left[\frac{1}{2}\left(\mathbf{F}_{N}+\mathbf{F}_{mul}\right)^{T}\left(\tilde{\mathbf{d}}_{\mu\nu}+\mathbf{d}_{\mu\nu}\right)+\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^{T}\left(\tilde{\mathbf{d}}_{N}+\tilde{\mathbf{d}}_{mul}+\mathbf{d}_{N}+\mathbf{d}_{mul}\right)\right]^{x}\\
&=-\frac{1}{2}\left[\left(\mathbf{F}^{\Delta,x}\right)^{T}\mathbf{d}_{e}+\left(\tilde{\mathbf{d}}^{\Delta}\right)^{T}\mathbf{D}^{x}\mathbf{d}_{e}+\left(\tilde{\mathbf{d}}^{\Delta}\right)^{T}\mathbf{F}^{x}_{e}\right]\\
&-\frac{1}{2}\left[\left(\mathbf{F}^{x}_{e}\right)^{T}\mathbf{d}^{\Delta}+\left(\tilde{\mathbf{d}}_{e}\right)^{T}\mathbf{D}^{x}\mathbf{d}^{\Delta}+\left(\tilde{\mathbf{d}}_{e}\right)^{T}\mathbf{F}^{\Delta,x}\right]\\
&-\frac{1}{2}\left[\left(\mathbf{F}^{x}_{N}+\mathbf{F}^{x}_{mul}\right)^{T}\mathbf{d}^{\Delta}+\left(\tilde{\mathbf{d}}_{N}+\tilde{\mathbf{d}}_{mul}\right)^{T}\mathbf{D}^{x}\mathbf{d}^{\Delta}+\left(\tilde{\mathbf{d}}_{N}+\tilde{\mathbf{d}}_{mul}\right)^{T}\mathbf{F}^{\Delta,x}\right]\\
&-\frac{1}{2}\left[\left(\mathbf{F}^{\Delta,x}\right)^{T}\left(\mathbf{d}_{N}+\mathbf{d}_{mul}\right)+\left(\tilde{\mathbf{d}}^{\Delta}\right)^{T}\mathbf{D}^{x}\left(\mathbf{d}_{N}+\mathbf{d}_{mul}\right)+\left(\tilde{\mathbf{d}}^{\Delta}\right)^{T}\left(\mathbf{F}^{x}_{N}+\mathbf{F}^{x}_{mul}\right)\right]\\
&=-\frac{1}{2}\left[\left(\mathbf{F}^{x}\right)^{T}\left(\mathbf{d}^{\Delta}+\tilde{\mathbf{d}}^{\Delta}\right)+\left(\mathbf{F}^{\Delta,x}\right)^{T}\left(\mathbf{d}+\tilde{\mathbf{d}}\right)+\left(\tilde{\mathbf{d}}\right)^{T}\mathbf{D}^{x}\mathbf{d}^{\Delta}+\left(\tilde{\mathbf{d}}^{\Delta}\right)^{T}\mathbf{D}^{x}\mathbf{d}\right]
\end{aligned}
\tag{5-26}
$$

where $\mathbf{F}^{\Delta,x}$ and $\mathbf{d}^{\Delta}$ (and $\tilde{\mathbf{d}}^{\Delta}$) are the electrostatic field gradient and induced dipoles, respectively, due to the TDDFT excitation difference density,

$$\mathbf{F}^{\Delta,x}=\sum_{\mu\nu\sigma}P^{\Delta}_{\mu\nu\sigma}\mathbf{F}^{x}_{\mu\nu} \tag{5-27}$$

$$\mathbf{d}^{\Delta}=-\sum_{\mu\nu\sigma}P^{\Delta}_{\mu\nu\sigma}\mathbf{D}^{-1}\mathbf{F}_{\mu\nu} \tag{5-28}$$

$$\tilde{\mathbf{d}}^{\Delta}=-\sum_{\mu\nu\sigma}P^{\Delta}_{\mu\nu\sigma}\left(\mathbf{D}^{-1}\right)^{T}\mathbf{F}_{\mu\nu} \tag{5-29}$$

The first term in the final expression of Eq. (5-26) represents the electrostatic force and torque imposed on $\mathbf{d}^{\Delta}$ (and $\tilde{\mathbf{d}}^{\Delta}$) by QM nuclei, electrons and MM multipoles; the second term represents the electrostatic force and torque imposed on $\mathbf{d}$ (and $\tilde{\mathbf{d}}$) by the TDDFT difference electron density; the third and forth terms represent the electrostatic force and torque between $\mathbf{d}^{\Delta}$ (and $\tilde{\mathbf{d}}^{\Delta}$) and $\mathbf{d}$ (and $\tilde{\mathbf{d}}$). These forces and torques can be evaluated efficiently using standard electrostatic formulas as if the induced dipoles were permanent dipoles.[51,179,180]

The last term in Eq. (5-13) can be written as

$$\begin{aligned}&\sum_{\mu\nu\sigma\lambda\kappa\sigma'}(X+Y)_{\mu\nu\sigma}(X+Y)_{\lambda\kappa\sigma'}\left[\frac{1}{2}\left(\mathbf{F}_{\mu\nu}\right)^{T}\left(\tilde{\mathbf{d}}_{\lambda\kappa}+\mathbf{d}_{\lambda\kappa}\right)\right]^{x},\\&=\left(\mathbf{F}_{XY}^{x}\right)^{T}\left(\mathbf{d}_{XY}+\tilde{\mathbf{d}}_{XY}\right)+\left(\tilde{\mathbf{d}}_{XY}\right)^{T}\mathbf{D}^{x}\mathbf{d}_{XY}\end{aligned}\tag{5-30}$$

where $\mathbf{F}_{XY}^{x}$ and $\mathbf{d}_{XY}$ are the electrostatic field gradient and induced dipoles, respectively, due to the TDDFT transition state density $(X+Y)_{\mu\nu\sigma}$:

$$\mathbf{F}_{XY}^{x}=\sum_{\mu\nu\sigma}(X+Y)_{\mu\nu\sigma}\mathbf{F}_{\mu\nu}^{x}\tag{5-31}$$

$$\mathbf{d}_{XY}=-\sum_{\mu\nu\sigma}(X+Y)_{\mu\nu\sigma}\mathbf{D}^{-1}\mathbf{F}_{\mu\nu}\tag{5-32}$$

$$\tilde{\mathbf{d}}_{XY}=-\sum_{\mu\nu\sigma}(X+Y)_{\mu\nu\sigma}\left(\mathbf{D}^{-1}\right)^{T}\mathbf{F}_{\mu\nu}\tag{5-33}$$

The first term of the right side of Eq. (5-30) represents the force and torque imposed on the transition state density induced dipoles $\mathbf{d}_{XY}$ (and $\tilde{\mathbf{d}}_{XY}$) by the electrostatic field gradient due to the transition state density $(X+Y)_{\mu\nu\sigma}$; the second term represents the force and torque between $\mathbf{d}_{XY}$ and $\tilde{\mathbf{d}}_{XY}$. Again, these forces and torques can be evaluated efficiently using standard electrostatic formulas as if the transition-state-density-induced dipoles were permanent dipoles.[51,179,180]

Vibration modes, frequencies, infrared and Raman intensities are useful information in experimental study of both ground and excited state molecules. The following formula[185] implemented in GAMESS is used to calculate the total integrated infrared band intensity I_i for the ith normal mode Q_i,

$$I_i=\frac{\pi N_A}{3c^2}\left[\left(\frac{\partial\mu_x}{\partial Q_i}\right)^2+\left(\frac{\partial\mu_y}{\partial Q_i}\right)^2+\left(\frac{\partial\mu_z}{\partial Q_i}\right)^2\right]\tag{5-34}$$

where N_A is the Avogadro's number, *c* is the speed of light, μ is the molecular dipole moment. In full QM calculations, μ is determined with the electron density and the nuclear charges and coordinates. In TDDFT calculations, the total electron density is the sum of the TDDFT difference density $\mathbf{P}^{\Delta}$ and the DFT ground state density **P**. In QM/MM calculations, μ should be the dipole moment of the whole QM/MM system, including the permanent and induced dipoles of the MM part.

5.3 Implementation and numerical results

5.3.1 Simulation details

The formulas derived in Section 5.2 are general and applicable to all polarizable force fields that use induced dipoles. There are many water models available with different potential forms and even more parameters.[186-195] The water may be treated as either rigid or flexible. The most commonly used water models among them are simple point charge (SPC, original[186] and refined[187]), extended simple point charge (SPC/E[188]), transferable intermolecular potential 3 points (TIP3P, original[189] and refined[190]), TIP4P,[191] TIP4P-Ew,[192] TIP5P[193] and TIP5P-E[194]. In Chapter 5, a polarizable QM/MM style TDDFT method was implemented in a local version of the GAMESS[68] package with a polarizable five-point rigid water model (POL5P), and the TDDFT code implemented by Chiba *et al*.[196] All calculations reported in the chapter were performed with the local version of GAMESS.

A 0.8 ns NVE simulation of 512 water molecules in a periodic cubic box (side length = 24.8575 Å, density = 0.997 g/cm^3, with a switching function[179] applied to all types of interactions, including polarization, from 9.000 to 12.428 Å for center-of-mass distances) shows an enthalpy of vaporization of 10.45 kcal/mol at 300 K, close to the

experimental value 10.49 kcal/mol.[197] The center-of-mass RDF peaks at 2.71 Å with a value of 0.325, similar to 2.74 Å and 0.286 from a TIP5P simulation performed using the same settings. TIP5P can accurately reproduce experimental radial distribution functions.[198] The QM-MM repulsion is modeled by a Gaussian-type potential $a \cdot \exp(-bR^2)$ at each site. R is the distance between electron and the repulsion site. The parameters *a* and *b* are chosen to reproduce the correct water dimer binding energy. The optimized binding energy for water dimer is -4.97 kcal/mol when the hydrogen bond donor water is described with B3LYP[76]/aug-cc-pVTZ[74] and the acceptor water is described with POL5P, and is -5.02 kcal/mol when the descriptions are switched and the geometry is reoptimized.

Table 5.1 Comparison between numerical and analytic gradients (au) calculated for S_1 excited state acetone and two water molecules.

	S_1 energy after displacement (hartree)		numerical	analytic	difference
	forward	backward			
water 1 x	-192.9200523247	-192.9200524232	0.0000492	0.0000492	0.0000000
water 1 y	-192.9200526662	-192.9200520790	-0.0002936	-0.0002938	0.0000002
water 1 z	-192.9200519470	-192.9200527853	0.0004191	0.0004190	0.0000002
water 1 rotation x	-192.9200518949	-192.9200528201	0.0004626	0.0004626	0.0000000
water 1 rotation y	-192.9200521646	-192.9200525530	0.0001942	0.0001942	0.0000000
water 1 rotation z	-192.9200524326	-192.9200523150	-0.0000588	-0.0000588	0.0000000
water 2 x	-192.9200520229	-192.9200527254	0.0003513	0.0003513	0.0000000
water 2 y	-192.9200535727	-192.9200511557	-0.0012085	-0.0012098	0.0000013
water 2 z	-192.9200545634	-192.9200501681	-0.0021976	-0.0021979	0.0000003
water 2 rotation x	-192.9200483954	-192.9200563207	0.0039627	0.0039627	0.0000000
water 2 rotation y	-192.9200530695	-192.9200516668	-0.0007014	-0.0007015	0.0000001
water 2 rotation z	-192.9200507497	-192.9200539846	0.0016174	0.0016175	0.0000000
acetone 1 x	-192.9200551828	-192.9200490771	-0.0030529	-0.0030502	-0.0000027
acetone 1 y	-192.9200556830	-192.9200484970	-0.0035930	-0.0035897	-0.0000033
acetone 1 z	-192.9200502772	-192.9200540823	0.0019025	0.0019102	-0.0000077
acetone 2 x	-192.9200459214	-192.9200589202	0.0064994	0.0064991	0.0000003
acetone 2 y	-192.9199828459	-192.9201214474	0.0693007	0.0693002	0.0000005
acetone 2 z	-192.9199670448	-192.9201370867	0.0850210	0.0850165	0.0000045
acetone 3 x	-192.9200595105	-192.9200453025	-0.0071040	-0.0071105	0.0000065
acetone 3 y	-192.9201188533	-192.9199855933	-0.0666300	-0.0666304	0.0000004
acetone 3 z	-192.9201370516	-192.9199671832	-0.0849342	-0.0849345	0.0000003
acetone 4 x	-192.9200491306	-192.9200551262	0.0029978	0.0029952	0.0000026
acetone 4 y	-192.9200529018	-192.9200514729	-0.0007144	-0.0007153	0.0000009
acetone 4 z	-192.9200549074	-192.9200492653	-0.0028210	-0.0028238	0.0000027
acetone 5 x	-192.9200517203	-192.9200529829	0.0006313	0.0006359	-0.0000046
acetone 5 y	-192.9200523806	-192.9200520219	-0.0001793	-0.0001787	-0.0000007
acetone 5 z	-192.9200510353	-192.9200536642	0.0013145	0.0013154	-0.0000010
acetone 6 x	-192.9200558359	-192.9200487529	-0.0035415	-0.0035407	-0.0000008
acetone 6 y	-192.9200525405	-192.9200521118	-0.0002143	-0.0002147	0.0000004
acetone 6 z	-192.9200522317	-192.9200523674	0.0000678	0.0000668	0.0000011
acetone 7 x	-192.9200558075	-192.9200486997	-0.0035539	-0.0035468	-0.0000071
acetone 7 y	-192.9200538540	-192.9200508383	-0.0015079	-0.0015053	-0.0000025
acetone 7 z	-192.9200493545	-192.9200552944	0.0029699	0.0029731	-0.0000032
acetone 8 x	-192.9200488473	-192.9200557030	0.0034279	0.0034340	-0.0000061
acetone 8 y	-192.9200527623	-192.9200518954	-0.0004334	-0.0004315	-0.0000019
acetone 8 z	-192.9200529905	-192.9200516437	-0.0006734	-0.0006777	0.0000043
acetone 9 x	-192.9200529612	-192.9200517354	-0.0006129	-0.0006197	0.0000068
acetone 9 y	-192.9200504144	-192.9200542685	0.0019270	0.0019254	0.0000016
acetone 9 z	-192.9200521600	-192.9200522752	0.0000576	0.0000555	0.0000021
acetone 10 x	-192.9200483296	-192.9200561469	0.0039087	0.0039068	0.0000018
acetone 10 y	-192.9200487946	-192.9200558863	0.0035458	0.0035447	0.0000011
acetone 10 z	-192.9200534740	-192.9200512204	-0.0011268	-0.0011339	0.0000071

Note: The geometry was optimized on the S_0 ground state surface with the B3LYP/aug-cc-pVDZ and the POL5P model. Then the analytic and numerical gradients on the S_1 excited state surface were calculated with the TDB3LYP/aug-cc-pVDZ method and the POL5P model. The numerical gradients were obtained via double displacement with translational step size of ±0.001 bohr and rotational (for the rigid POL5P water model) step size of ±0.001 rad.

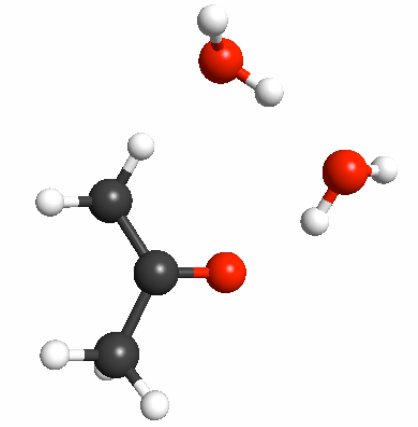


Figure 5.1 A cluster formed by acetone and two water molecules.

5.3.2 Accuracy of the analytic gradient

A cluster formed by acetone and two water molecules (Figure 5.1) is used to demonstrate the TDDFT-POL5P gradient. The geometry was optimized on the S_0 ground state surface with the B3LYP[76]/aug-cc-pVDZ[74] method in which the two water molecules were represented by two pieces of the POL5P model. Then the analytic and numerical gradients on the S_1 excited state surface were calculated with the TDB3LYP-POL5P/aug-cc-pVDZ method. In both the ground state B3LYP and the excited state TDB3LYP calculations, 96 radial and 302 Lebedev angular grid points were used. The numerical gradients were obtained via double displacement with translational step size of ±0.001 bohr and rotational (for the rigid POL5P water model) step size of ±0.001 rad. The results are presented in Table 5.2. For the two POL5P water molecules the root-mean-square difference between the analytic and numerical gradients is 4×10^{-7} au and the largest difference is 13×10^{-7} au. Such a good quality analytic gradient suggests that the formulas and implementation are correct. For acetone atoms, the differences are larger, with a root-

mean-square difference of 37×10^{-7} au and the largest difference of 77×10^{-7} au. This is presumably due to the numerical errors in the DFT grid point integrals because similar errors are seen in ground state DFT calculations. Such errors are not invoked when only the POL5P water molecules are displaced.

Table 5.2 Cartesian coordinates of a cluster formed by acetone and two water molecules optimized on the S_0 ground state surface with the B3LYP/aug-cc-pVDZ and the POL5P method. The two water molecules are modeled by the POL5P potential, which includes non-isotropic polarization.

C	-0.2587930047	0.3774034477	-1.2736224515
C	-0.2195413738	-0.1702741811	0.1355463355
O	-0.1433910823	0.5615436795	1.1110052512
C	-0.2778492811	-1.6750701380	0.2940783705
H	-0.3639287295	1.4666847812	-1.2833456412
H	0.4278555290	-2.1694660223	-0.3895047988
H	0.6767914369	0.1023366576	-1.7873980119
H	-1.0758473659	-0.0897690668	-1.8439509687
H	-0.0664018436	-1.9558496451	1.3311957557
H	-1.2852702468	-2.0269098945	0.0205768243
O	-0.2352506157	4.0120982754	-1.2596020508
H	-0.7279230853	4.8320944912	-1.2263242387
H	-0.1420396468	3.7507922514	-0.3434916611
O	-0.1547969278	3.1457628640	1.3165466519
H	-0.2949959813	2.2026140911	1.2326301058
H	0.6275861195	3.2244504834	1.8623601684

5.3.3 Properties of the acetone-2water cluster

The geometries of acetone and the acetone-2water cluster (Figure 5.1) were optimized on the S_0 and S_1 surfaces using B3LYP, B3LYP-POL5P, TDB3LYP and TDB3LYP-POL5P methods and the aug-cc-pVDZ basis set.[74] Similar initial geometries were purposely used in the geometry optimization so the optimized geometries are

similar to each other, and comparisons among them are possible. The main structural difference is in acetone: on the S_0 surface the three C and one O atoms are virtually coplanar, while on the S_1 surface the O atom is bent out of the plane formed by the three C atoms. Then the Hessian force constants and the derivatives of the molecular dipole moment were obtained via finite differentiation of the analytic gradients and the dipole moment, respectively. Double displacement was used with translational step size of ±0.01 bohr and rotational (for the rigid POL5P water model) step size of ±0.01 rad. All vibration frequencies are real, confirming that the optimized geometries are minima. In the QM/MM calculations, the infrared intensity was determined using Eq. (5-34) and the molecular dipole moment of the whole QM/MM system, including the dipole of the QM part, the permanent and induced dipoles of each POL5P water molecule.

The three-body binding energy between the acetone and two water molecules is calculated as the energy difference between the cluster and the three monomers. The geometries of acetone and the cluster were optimized with the aug-cc-pVDZ basis set, as described above, on the S_0 and S_1 surfaces. A separate geometry optimization for water (ground state) was performed with the B3LYP/aug-cc-pVDZ method. The energies then were refined with the aug-cc-pVTZ basis set to reduce basis set superposition errors in the calculation of binding energies. For the S_0 state, the full B3LYP result is -12.79 kcal/mol, and the B3LYP-POL5P result is -12.82 kcal/mol. For the S_1 state, the full TDB3LYP result is -10.36 kcal/mol, and the TDB3LYP-POL5P result is -10.78 kcal/mol. According to these calculations, upon excitation the acetone-water interaction in the cluster is weakened by ~2 kcal/mol. The POL5P water model reproduces the binding energies very well for both the S_0 and S_1 states.

The vertical $S_0 \rightarrow S_1$ excitation energy calculated with TDB3LYP for B3LYP optimized acetone is 4.36 eV, in good agreement with an experimental maximum absorption value 4.38 eV obtained by Walzl *et al.* using electron scattering method.[199] The TDB3LYP excitation energy is 4.51 eV for the B3LYP optimized acetone-2water cluster, showing a blue shift of 0.15 eV as compared to acetone. Experiment shows that acetone exhibits a blue shift of 0.20 eV[200] ongoing from the gas phase to aqueous solution, in which acetone dynamically forms hydrogen bonds with many water molecules. The TDB3LYP-POL5P excitation energy is 4.47 eV for the B3LYP-POL5P optimized acetone-2water cluster, in good agreement with the full TDB3LYP value 4.51 eV.

The carbonyl C=O bond lengths are 1.217, 1.225 and 1.222 Å, respectively, in the S_0 ground state acetone optimized with B3LYP, acetone-2water cluster optimized with B3LYP, and the acetone-2water cluster optimized with B3LYP-POL5P. The carbonyl C-O bond lengths are 1.313, 1.320 and 1.317 Å, respectively, in the S_1 excited state acetone optimized with TDB3LYP, the acetone-2water cluster optimized with TDB3LYP, and the acetone-2water cluster optimized with TDB3LYP-POL5P. The ~0.1 Å lengthening is reasonable because the $n \rightarrow \pi^*$ valence transition weakens the C-O bond strength. Clearly, the B3LYP-POL5P and TDB3LYP-POL5P methods can accurately reproduce full B3LYP and TDB3LYP results for this bond length.

B3LYP calculation shows that the frequency and infrared spectrum intensity of the carbonyl vibration mode are 1775 cm^{-1} and 4.59 $debye^2amu^{-1}Å^{-2}$ for S_0 acetone, and are 1749 cm^{-1} and 5.94 $debye^2amu^{-1}Å^{-2}$ for the S_0 acetone-2water cluster. B3LYP-POL5P calculation gives 1767 cm^{-1} and 5.31 $debye^2amu^{-1}Å^{-2}$ for the S_0 acetone-2water cluster, in

qualitative agreement with the full B3LYP results (Table 5.3). For the S_0 acetone-2water cluster, full B3LYP predicts a dipole moment of 2.89 D, while B3LYP-POL5P predicts 2.85 D (including the permanent and induced dipoles of the two POL5P water molecules).

For S_1 acetone, TDB3LYP calculation shows that the most infrared-intensive mode is a methyl C-H stretching mode with a frequency of 2923 cm^{-1} and an intensity of 2.42 $debye^2amu^{-1}Å^{-2}$, and the carbonyl mode has a frequency of 1307 cm^{-1} and an intensity of 0.33 $debye^2amu^{-1}Å^{-2}$. So the $n \rightarrow \pi^*$ valence transition results in both C-O bond weakening and dipole moment reduction. For the S_1 acetone-2water cluster, TDB3LYP calculation shows that the most infrared-intensive modes are water O-H stretching, but the acetone methyl C-H stretching mode is still strong, with a frequency of 2932 cm^{-1} and an intensity of 2.20 $debye^2amu^{-1}Å^{-2}$. Like that in S_1 acetone, the carbonyl mode now has a frequency of 1282 cm^{-1} and an intensity of 0.48 $debye^2amu^{-1}Å^{-2}$. For the S_1 acetone-2water cluster, TDB3LYP-POL5P does not predict water O-H stretching because the POL5P is a rigid water model. The acetone methyl C-H stretching mode is also strong, with a frequency of 2931 cm^{-1} and an intensity of 2.08 $debye^2amu^{-1}Å^{-2}$, in qualitative agreement with the full TDB3LYP results. The carbonyl mode has a frequency of 1292 cm^{-1} and an intensity of 0.38 $debye^2amu^{-1}Å^{-2}$, in qualitative agreement with the full TDB3LYP results (Table 5.3). For the S_1 acetone-2water cluster, full TDB3LYP predicts a dipole moment of 1.96 D, while TDB3LYP-POL5P predicts 1.97 D.

Table 5.3 Carbonyl bond length and vibration mode in acetone and an acetone-2water cluster (bond length in Å, frequency in cm^{-1}, infrared intensity in $debye^2amu^{-1}Å^{-2}$).

	S_0 ground state (B3LYP)			S_1 excited state (TDB3LYP)		
	length	frequency	intensity	length	frequency	intensity
acetone	1.217	1775	4.59	1.313	1307	0.33
acetone + 2 water	1.225	1749	5.94	1.320	1282	0.48
acetone + 2 POL5P	1.222	1767	5.31	1.317	1292	0.38

5.4 Summary

The formulas derived in Chapter 5 for evaluating analytic gradient in polarizable QM/MM style TDDFT methods are general, and can be used for all polarizable force fields that employ induced dipoles. Numerical tests show that the formulas are correct and rigorous. In a later work this method has been implemented with a polarizable protein force field in QuanPol program (Quantum chemistry polarizable force field).[136] It is capable of performing TDDFT/MM/Continuum and TDDFT/MM style calculations for various solvents, proteins and DNA/RNA molecules. In the future, this method will be applied to study the excited state molecules in the biological chromophores.

Concluding remarks

This dissertation contains both the application and extension of quantum chemical methods. Several novel QM/MM methods are developed to combine *ab initio* methods with molecular mechanic methods. QM part contains both ground state and excited state methods. MM part includes both implicit and explicit models. Those QM/MM methods can be used to study the structural, electronic and optical properties of biosystems, inorganic materials, and organic solvents. The first part of the thesis, Chapter 1, introduced the theoretical background of this study. The second part of this thesis, Chapters 2 to 5, is the adventures of quantum chemistry in the realm of the condensed phase. In Chapter 2, a heterogeneous conductor-like solvation model that uses different local effective dielectrics for different portions of the solute cavity surface is implemented for quantum chemical Hartree-Fock and Kohn-Sham methods. By variationally formulating the solvation operators for Hartree-Fock and Kohn-Sham methods, continuous and smooth potential energy surfaces and analytic nuclear gradients accurate to $10^{-7} \sim 10^{-6}$ atomic unit have been obtained for the heterogeneous solvation model. In Chapter 3, the heterogeneous solvation model was applied to optimize the structures and calculate reduction potentials for the protein models containing type-1 Cu centers. The active site model molecules consisting of ~100 atoms are partly exposed to the solvent and partly embedded in the protein. The reduction potentials (E^0) of five type-1 Cu centers show a 540 mV range and are in good agreement with experimental values. Chapter 4 presents the implementation of the analytic energy gradient for a combined closed shell RMP2/CPCM and spin-restricted open shell Z-averaged ZAPT2/CPCM and spin-unrestricted open shell UMP2/CPCM methods. It is the first time that rigorously

continuous and smooth potential energy surfaces are obtained. Using these methods, the geometries of the S_0 ground state and the T_1 state of three nucleobase pairs (guanine-cytosine, adenine-thymine and adenine-uracil) in the gas phase and aqueous solution phase are optimized. Chapter 5 formulates the analytic energy gradient for combined time-dependent density functional theory (TDDFT) and polarizable force field methods that incorporate dipole polarizability tensors and linearly induced point dipoles. The analytic gradient of the mutual polarization energy of the force field and the TDDFT excited state can be formulated by using the TDDFT difference density-induced dipoles and the transition state density-induced dipoles. All the forces and torques involving induced dipoles can be efficiently evaluated using standard electrostatic formulas as if the induced dipoles were permanent dipoles. A polarizable five-point water model (POL5P) was used to show that the formulas are general and rigorous. The carbonyl vibration modes and infrared absorption intensities of a cluster formed by an acetone molecule and two water molecules are studied. Combined QM/MM style model advances the QM model and facilitates the application to a broader range of chemical problems. MP2 is one of the most widely used methods for predicting ground state molecular properties and TDDFT is one of the most commonly used methods for predicting excited state properties. Therefore, QM/MM style methods with analytic gradients can be used on many aspects of the theoretical studies of solvent effects. The gradient methods developed in this thesis are general and can be used for all polarizable force fields that employ induced dipoles or induced surface charges.

References

[1] R. J. Hinde, Journal of Chemical Education **77** (12), 1564 (2000).

[2] E. Schrödinger, Ann. Phys. **79**, 361 (1926); E. Schrödinger, Ann. Phys. **79**, 489 (1926); E. Schrödinger, Ann. Phys. **79**, 734 (1926); E. Schrödinger, Ann. Phys. **80**, 437 (1926); E. Schrödinger, Ann. Phys. **81**, 109 (1926); E. Schrödinger, Die Naturwissenschaften **14**, 664 (1926); E. Schrödinger, Phys. Rev. **28**, 1049 (1926).

[3] P. A. M. Dirac, Proceedings of the Royal Society of London. Series A **117** (778), 610 (1928).

[4] P. A. M. Dirac, Proceedings of the Royal Society of London. Series A **123** (792), 714 (1929).

[5] W. Heitler and F. London, Zeitschrift für Physik A Hadrons and Nuclei **44** (6), 455 (1927).

[6] D. R. Hartree, Proc. Cambridge Phil. Soc **24**, 89 (1928); D. R. Hartree, Proc. Cambridge Phil. Soc **24**, 111 (1928); D. R. Hartree, Proc. Cambridge Phil. Soc **24**, 426 (1928); V. Z. Fock, Phys. **61**, 126 (1930); V. Z. Fock, Phys. **62**, 795 (1930).

[7] E. A. Hylleraas, Zeitschrift für Physik A Hadrons and Nuclei **48** (7), 469 (1928).

[8] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[9] C. C. J. Roothaan, Reviews of Modern Physics **23** (2), 69 (1951).

[10] G. G. Hall, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **205** (1083), 541 (1951).

[11] R. J. Bartlett, The Journal of Physical Chemistry **93** (5), 1697 (1989).

[12] S. Miertus, E. Scrocco, and J. Tomasi, Chemical Physics **55** (1), 117 (1981).

[13] E. Cances, B. Mennucci, and J. Tomasi, Journal of Chemical Physics **107** (8), 3032 (1997).

[14] B. Mennucci, E. Cances, and J. Tomasi, The Journal of Physical Chemistry B **101** (49), 10506 (1997).

[15] M. F. Iozzi, M. Cossi, R. Improta, N. Rega, and V. Barone, The Journal of Chemical Physics **124** (18), 184103 (2006).

[16] P. Su and H. Li, The Journal of Chemical Physics **130** (7), 074109 (2009).

[17] H. Li, A. D. Robertson, and J. H. Jensen, Proteins-Structure Function and Bioinformatics **55** (3), 689 (2004).

[18] R. Cammi, B. Mennucci, and J. Tomasi, Journal of Physical Chemistry A **103** (45), 9100 (1999).

[19] M. E. Casida, in *Recent Advances in Density Functional Methods*, edited by D. P. Chong (World Scientific, Singapore, 1995), pp. 155; M. E. Casida, C. Jamorski, K. C. Casida, and D. R. Salahub, Journal of Chemical Physics **108** (11), 4439 (1998).

[20] D. Si and H. Li, J. Chem. Phys. **135** (14), 144107 (2011); D. Si and H. Li, The Journal of Chemical Physics **133** (14), 144112 (2010); D. Si and H. Li, The Journal of Physical Chemistry A **113** (46), 12979 (2009).

[21] D. Si and H. Li, The Journal of Chemical Physics **131** (4), 044123 (2009).

[22] F. Jensen, *Introduction to Computational Chemistry*. (Wiley, 2006).

[23] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. (Dover Publications, 1996).

[24] C. Cramer, *Essentials of Computational Chemistry: Theories and Models*. (Wiley, 2004).
[25] M. Born and J. R. Oppenheimer, Ann. Phys. **84**, 457 (1926).
[26] J. C. Slater, Phys. Rev. **34**, 1293 (1929); J. C. Slater, Phys. Rev. **35**, 509 (1930).
[27] T. L. Gilbert, Journal of Chemical Physics **43** (10), S248 (1965); G. Das and A. C. Wahl, Journal of Chemical Physics **56** (4), 1769 (1972); G. Das and A. C. Wahl, Journal of Chemical Physics **47** (8), 2934 (1967); G. Das and A. C. Wahl, Journal of Chemical Physics **44** (1), 87 (1966); M. W. Schmidt and M. S. Gordon, Annual Review of Physical Chemistry **49**, 233 (1998).
[28] E. Fermi, Rend. Accad. Naz. Lincei **6**, 602 (1927).
[29] L. H. Thomas, Proc. Cambridge Phil. Soc **23** (5) (1927).
[30] P. Pulay, Molecular Physics **17** (2), 197 (1969).
[31] M. Orozco and F. J. Luque, Chemical Reviews **100** (11), 4187 (2000); M. Cossi, N. Rega, I. Soteras, D. Blanco, O. Huertas, A. Bidon-Chanal, F. J. Luque, D. G. Truhlar, J. R. Pliego, B. M. Ladanyi, M. D. Newton, W. Domcke, A. L. Sobolewski, D. Laage, I. Burghardt, J. T. Hynes, M. Persico, G. Granucci, V. M. Huxter, G. D. Scholes, and C. Curutchet, in *Continuum Solvation Models in Chemical Physics* (John Wiley & Sons, Ltd, 2007), pp. 313.
[32] J. Tomasi, B. Mennucci, and R. Cammi, Chemical Reviews **105** (8), 2999 (2005).
[33] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, Journal of Computational Chemistry **4** (2), 187 (1983); A. D. MacKerell, C. L. Brooks, L. Nilsson, B. Roux, Y. Won, M. Karplus, and Schleyer, (John Wiley & Sons: Chichester, 1998), Vol. 1, pp. 271.
[34] J. E. Jones, Proceedings of the Royal Society of London. Series A **106** (738), 463 (1924).
[35] R. W. Dixon and P. A. Kollman, Journal of Computational Chemistry **18** (13), 1632 (1997).
[36] A. J. Stone, Journal of Chemical Theory and Computation **1** (6), 1128 (2005).
[37] A. Warshel and M. Levitt, Journal of Molecular Biology **103** (2), 227 (1976).
[38] U. C. Singh and P. A. Kollman, Journal of Computational Chemistry **7** (6), 718 (1986); M. J. Field, P. A. Bash, and M. Karplus, Journal of Computational Chemistry **11** (6), 700 (1990); J. L. Gao, Accounts of Chemical Research **29** (6), 298 (1996); H. M. Senn and W. Thiel, in *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*, edited by M. Reiher (2007), Vol. 268, pp. 173; H. Lin and D. G. Truhlar, Theoretical Chemistry Accounts **117** (2), 185 (2007).
[39] A. Ohrn and G. Karlstrom, Molecular Physics: An International Journal at the Interface Between Chemistry and Physics **104** (19), 3087 (2006).
[40] Y. Lin and J. Gao, Journal of Chemical Theory and Computation **3** (4), 1484 (2007).
[41] J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, The Journal of Chemical Physics **118** (4), 1620 (2003); J. Kongsted, A. Osted, K. V. Mikkelsen, P.-O. Astrand, and O. Christiansen, The Journal of Chemical Physics **121** (17), 8435 (2004).
[42] C. B. Nielsen, O. Christiansen, K. V. Mikkelsen, and J. Kongsted, Journal of Chemical Physics **126** (15), 154112 (2007).

[43] L. Jensen, P. T. van Duijnen, and J. G. Snijders, Journal of Chemical Physics **119** (7), 3800 (2003).
[44] P. T. Van Duijnen and T. L. Netzel, Journal of Physical Chemistry A **110** (6), 2204 (2006); C. R. Jacob, J. Neugebauer, L. Jensen, and L. Visscher, Physical Chemistry Chemical Physics **8** (20), 2349 (2006).
[45] S. Yoo, F. Zahariev, S. Sok, and M. S. Gordon, Journal of Chemical Physics **129** (14), 144112 (2008).
[46] Q. Cui, J. Chem. Phys. **117** (10), 4720 (2002).
[47] H. Jensen Jan, N. Day Paul, S. Gordon Mark, H. Basch, D. Cohen, R. Garmer David, M. Kraus, and J. Stevens Walter, in *Modeling the Hydrogen Bond* (American Chemical Society, 1994), Vol. 569, pp. 139; *Modeling the Hydrogen Bond*. (American Chemical Society, 1994); W. Chen, J. Chem. Phys. **105** (24), 11081 (1996); M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, The Journal of Physical Chemistry A **105** (2), 293 (2000).
[48] P. Bandyopadhyay, J. Chem. Phys. **113** (3), 1104 (2000).
[49] P. Bandyopadhyay, J. Chem. Phys. **116** (12), 5023 (2002).
[50] H. Li, J. Chem. Phys. **126** (12), 124112 (2007).
[51] H. Li, Journal of Chemical Physics **131** (18), 184103 (2009).
[52] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, Journal of Physical Chemistry **100** (50), 19357 (1996); R. D. J. Froese and K. Morokuma, *Encyclopedia of computational chemistry*. (John Wiley, 1998); S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, Journal of Molecular Structure: THEOCHEM **461-462** (0), 1 (1999); P. B. Karadakov and K. Morokuma, Chemical Physics Letters **317** (6), 589 (2000); T. Vreven and K. Morokuma, Journal of Computational Chemistry **21** (16), 1419 (2000); T. Vreven, J. Chem. Phys. **115** (1), 62 (2001); B. Mennucci, J. M. Martínez, and J. Tomasi, The Journal of Physical Chemistry A **105** (30), 7287 (2001); M. D'Amore, R. Improta, and V. Barone, The Journal of Physical Chemistry A **107** (32), 6264 (2003).
[53] L. Onsager, Journal of the American Chemical Society **58** (8), 1486 (1936); J. Tomasi, Theoretical Chemistry Accounts **103** (3-4), 196 (2000).
[54] A. Bondi, The Journal of Physical Chemistry **68** (3), 441 (1964); R. S. Rowland and R. Taylor, The Journal of Physical Chemistry **100** (18), 7384 (1996); L. Pauling, (1960).
[55] A. Klamt and G. Schuurmann, Journal of the Chemical Society-Perkin Transactions 2 (5), 799 (1993).
[56] V. Barone and M. Cossi, Journal of Physical Chemistry A **102** (11), 1995 (1998).
[57] H. Li and J. H. Jensen, Journal of Computational Chemistry **25** (12), 1449 (2004).
[58] D. M. Chipman, Theoretical Chemistry Accounts **107** (2), 80 (2002); D. M. Chipman and M. Dupuis, Theoretical Chemistry Accounts **107** (2), 90 (2002).
[59] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, and D. G. Truhlar, Journal of Chemical Theory and Computation **3** (6), 2011 (2007).
[60] R. Bonaccorsi, E. Scrocco, and J. Tomasi, International Journal of Quantum Chemistry **29** (4), 717 (1986).

[61] R. Bonaccorsi, M. Hodoscek, and J. Tomasi, Journal of Molecular Structure: Theochem **164** (1-2), 105 (1988).
[62] R. Bonaccorsi, F. Floris, P. Palla, and J. Tomasi, THERMOCHIMICA ACTA **162** (1), 213 (1990).
[63] H. Hoshi, M. Sakurai, Y. Inoue, and R. Chujo, The Journal of Chemical Physics **87** (2), 1107 (1987).
[64] L. Frediani, R. Cammi, S. Corni, and J. Tomasi, The Journal of Chemical Physics **120** (8), 3893 (2004).
[65] M. Cossi, N. Rega, G. Scalmani, and V. Barone, Journal of Computational Chemistry **24** (6), 669 (2003).
[66] J. L. Pascualahuir, E. Silla, and I. Tunon, Journal of Computational Chemistry **15** (10), 1127 (1994); E. Silla, I. Tunon, and J. L. Pascualahuir, Journal of Computational Chemistry **12** (9), 1077 (1991); E. Silla, F. Villar, O. Nilsson, J. L. Pascualahuir, and O. Tapia, Journal of Molecular Graphics **8** (3), 168 (1990); J. L. Pascualahuir and E. Silla, Journal of Computational Chemistry **11** (9), 1047 (1990).
[67] C. S. Pomelli, Journal of Computational Chemistry **25** (12), 1532 (2004).
[68] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Journal of Computational Chemistry **14** (11), 1347 (1993); M. S. Gordon and M. W. Schmidt, in *Theory and applications of computational chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005).
[69] E. I. Solomon, R. K. Szilagyi, S. D. George, and L. Basumallick, Chemical Reviews **104** (2), 419 (2004).
[70] J. Tomasi, B. Mennucci, and E. Cances, Journal of Molecular Structure-Theochem **464** (1-3), 211 (1999).
[71] H. Li, C. S. Pomelli, and J. H. Jensen, Theoretical Chemistry Accounts **109** (2), 71 (2003).
[72] C. S. Pomelli, J. Tomasi, and V. Barone, Theoretical Chemistry Accounts **105** (6), 446 (2001).
[73] J. Baker, A. Kessi, and B. Delley, Journal of Chemical Physics **105** (1), 192 (1996).
[74] T. H. Dunning, Journal of Chemical Physics **90** (2), 1007 (1989).
[75] N. B. Balabanov and K. A. Peterson, The Journal of Chemical Physics **123** (6), 064107 (2005).
[76] R. H. Hertwig and W. Koch, Chemical Physics Letters **268** (5-6), 345 (1997).
[77] J. Li, M. R. Nelson, C. Y. Peng, D. Bashford, and L. Noodleman, Journal of Physical Chemistry A **102** (31), 6311 (1998).
[78] R. A. Torres, T. Lovell, L. Noodleman, and D. A. Case, Journal of the American Chemical Society **125** (7), 1923 (2003).
[79] E. Sigfridsson, M. H. M. Olsson, and U. Ryde, Inorg. Chem. **40** (11), 2509 (2001); M. Sulpizi, S. Raugei, J. VandeVondele, P. Carloni, and M. Sprik, Journal of Physical Chemistry B **111** (15), 3969 (2007).
[80] D. Bashford, M. Karplus, and G. W. Canters, Journal of Molecular Biology **203** (2), 507 (1988); B. Honig and A. Nicholls, Science **268** (5214), 1144 (1995); W.

G. Han, T. Lovell, T. Q. Liu, and L. Noodleman, Inorganic Chemistry **42** (8), 2751 (2003); G. M. Ullmann, L. Noodleman, and D. A. Case, Journal of Biological Inorganic Chemistry **7** (6), 632 (2002); J. J. Mao, K. Hauser, and M. R. Gunner, Biochemistry **42** (33), 9829 (2003); A. Warshel, A. Papazyan, and I. Muegge, Journal of Biological Inorganic Chemistry **2** (1), 143 (1997); P. J. Stephens, D. R. Jollie, and A. Warshel, Chemical Reviews **96** (7), 2491 (1996); G. M. Ullmann and E. W. Knapp, European Biophysics Journal with Biophysics Letters **28** (7), 533 (1999); D. M. Popovic, S. D. Zaric, B. Rabenstein, and E. W. Knapp, Journal of the American Chemical Society **123** (25), 6040 (2001); P. Voigt and E. W. Knapp, Journal of Biological Chemistry **278** (52), 51993 (2003); H. Ishikita and E. W. Knapp, Febs Letters **579** (14), 3190 (2005).

[81] C. A. P. Libeu, M. Kukimoto, M. Nishiyama, S. Horinouchi, and E. T. Adman, Biochemistry **36** (43), 13160 (1997).

[82] M. V. Botuyan, A. ToyPalmer, J. Chung, R. C. Blake, P. Beroza, D. A. Case, and H. J. Dyson, Journal of Molecular Biology **263** (5), 752 (1996).

[83] C. N. Schutz and A. Warshel, Proteins: Structure, Function, and Genetics **44** (4), 400 (2001).

[84] M. H. M. Olsson and U. Ryde, Journal of Biological Inorganic Chemistry **4** (5), 654 (1999).

[85] M. H. M. Olsson, G. Y. Hong, and A. Warshel, Journal of the American Chemical Society **125** (17), 5025 (2003).

[86] H. Li, S. P. Webb, J. Ivanic, and J. H. Jensen, Journal of the American Chemical Society **126** (25), 8010 (2004).

[87] S. N. Datta, J. Sudhamsu, and A. Pandey, J. Phys. Chem. B **108** (23), 8007 (2004); M. Sundararajan, I. H. Hillier, and N. A. Burton, Journal of Physical Chemistry A **110** (2), 785 (2006); K. Paraskevopoulos, M. Sundararajan, R. Surendran, M. A. Hough, R. R. Eady, I. H. Hillier, and S. S. Hasnain, Dalton Transactions (25), 3067 (2006).

[88] O. Tapia and G. Johannin, The Journal of Chemical Physics **75** (7), 3624 (1981).

[89] R. Bonaccorsi, E. Ojalvo, and J. Tomasi, Collection of Czechoslovak Chemical Communications **53**, 2320 (1988); R. Bonaccorsi, E. Ojalvo, P. Palla, and J. Tomasi, Chemical Physics **143** (2), 245 (1990); R. Bonaccorsi, F. Floris, P. Palla, and J. Tomasi, THERMOCHIMICA ACTA **162**, 213 (1990).

[90] S. Jorgensen, M. A. Ratner, and K. V. Mikkelsen, Journal of Chemical Physics **115** (8), 3792 (2001); S. Jorgensen, M. A. Ratner, and K. V. Mikkelsen, Chemical Physics **278** (1), 53 (2002); S. Jorgensen, M. A. Ratner, and K. V. Mikkelsen, Journal of Chemical Physics **116** (24), 10902 (2002); S. Jorgensen, M. A. Ratner, and K. V. Mikkelsen, Journal of Chemical Physics **115** (17), 8185 (2001); M. Sloth, S. Jorgensen, M. Bilde, and K. V. Mikkelsen, Journal of Physical Chemistry A **107** (41), 8623 (2003).

[91] W. Im, M. Feig, and C. L. Brooks, Biophysical Journal **85** (5), 2900 (2003); S. Tanizaki, J. Chem. Phys. **122** (12), 124706 (2005); S. Tanizaki and M. Feig, Journal of Physical Chemistry B **110** (1), 548 (2006); M. Feig and C. L. Brooks, Current Opinion in Structural Biology **14** (2), 217 (2004); L. Moulinier, D. A. Case, and T. Simonson, Acta Crystallographica Section D-Biological Crystallography **59**, 2094 (2003).

[92] A. D. Becke, Physical Review A **38** (6), 3098 (1988).

[93] C. Lee, W. Yang, and R. G. Parr, Physical Review B **37** (2), 785 (1988).

[94] T. Kohzuma, T. Inoue, F. Yoshizaki, Y. Sasakawa, K. Onodera, S. Nagatomo, T. Kitagawa, S. Uzawa, Y. Isobe, Y. Sugimura, M. Gotowda, and Y. Kai, Journal of Biological Chemistry **274** (17), 11817 (1999).

[95] R. Hulsker, A. Mery, E. A. Thomassen, A. Ranieri, M. Sola, M. P. Verbeet, T. Kohzuma, and M. Ubbink, J. Am. Chem. Soc. **129** (14), 4423 (2007).

[96] J. M. Guss, Harrowell, P.R., Murata, M., Norris, V.A., Freeman, H.C. , J.Mol.Biol. **192**, 361 (1986).

[97] H. B. Gray, B. G. Malmstrom, and R. J. P. Williams, Journal of Biological Inorganic Chemistry **5** (5), 551 (2000).

[98] B. A. Fields, H. H. Bartsch, H. D. Bartunik, F. Cordes, J. M. Guss, and H. C. Freeman, Acta Crystallographica Section D-Biological Crystallography **50**, 709 (1994).

[99] M. Milani, L. Andolfi, S. Cannistraro, M. P. Verbeet, and M. Bolognesi, Acta Crystallographica Section D-Biological Crystallography **57**, 1735 (2001).

[100] B. A. Fields, A. P. Duff, K. Govinderaju, M. P. Lee, J. H. W., W. B. Church, J. M. Guss, A. G. Sykes, and H. C. Freeman, (unpublished results).

[101] G. D. Armstrong, S. K. Chapman, M. J. Sisley, A. G. Sykes, A. Aitken, N. Osheroff, and E. Margoliash, Biochemistry **25** (22), 6947 (1986).

[102] T. Inoue, H. Sugawara, S. Hamanaka, H. Tsukui, E. Suzuki, T. Kohzuma, and Y. Kai, Biochemistry **38** (19), 6063 (1999).

[103] A. Romero, B. De la Cerda, P. F. Varela, J. A. Navarro, M. Hervas, and M. A. De la Rosa, Journal of Molecular Biology **275** (2), 327 (1998).

[104] C. A. Collyer, J. M. Guss, Y. Sugimura, F. Yoshizaki, and H. C. Freeman, Journal of Molecular Biology **211** (3), 617 (1990).

[105] M. R. Redinbo, D. Cascio, M. K. Choukair, D. Rice, S. Merchant, and T. O. Yeates, Biochemistry **32** (40), 10560 (1993).

[106] Y. F. Xue, M. Okvist, O. Hansson, and S. Young, Protein Science **7** (10), 2099 (1998).

[107] N. Shibata, T. Inoue, C. Nagano, N. Nishio, T. Kohzuma, K. Onodera, F. Yoshizaki, Y. Sugimura, and Y. Kai, Journal of Biological Chemistry **274** (7), 4225 (1999).

[108] K. Sato, T. Kohzuma, and C. Dennison, Journal of the American Chemical Society **125** (8), 2101 (2003).

[109] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucl. Acids Res. **28** (1), 235 (2000).

[110] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, and G. Vriend, Bioinformatics **14** (6), 523 (1998).

[111] R. McWeeny and G. Diercksen, Journal of Chemical Physics **49** (11), 4852 (1968).

[112] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, and J. A. Pople, Journal of Chemical Physics **77** (7), 3654 (1982); V. A. Rassolov, J. A. Pople, M. A. Ratner, and T. L. Windus, Journal of Chemical Physics **109** (4), 1223 (1998).

[113] P. J. Hart, A. M. Nersissian, R. G. Herrmann, R. M. Nalbandyan, J. S. Valentine, and D. Eisenberg, Protein Science **5** (11), 2175 (1996).
[114] J. F. Hall, L. D. Kanbi, R. W. Strange, and S. S. Hasnain, Biochemistry **38** (39), 12675 (1999).
[115] M. L. Barrett, I. Harvey, M. Sundararajan, R. Surendran, J. F. Hall, M. J. Ellis, M. A. Hough, R. W. Strange, I. H. Hillier, and S. S. Hasnain, Biochemistry **45** (9), 2927 (2006).
[116] L. D. Kanbi, S. Antonyuk, M. A. Hough, J. F. Hall, F. E. Dodd, and S. S. Hasnain, Journal of Molecular Biology **320** (2), 263 (2002).
[117] V. Barone and M. Cossi, The Journal of Physical Chemistry A **102** (11), 1995 (1998).
[118] D. Rosen, Transactions of the Faraday Society **59**, 2178 (1963); S. Bone and R. Pethig, Journal of Molecular Biology **157** (3), 571 (1982); S. Bone and R. Pethig, Journal of Molecular Biology **181** (2), 323 (1985).
[119] G. King, F. S. Lee, and A. Warshel, The Journal of Chemical Physics **95** (6), 4366 (1991).
[120] J. Antosiewicz, J. A. McCammon, and M. K. Gilson, Journal of Molecular Biology **238** (3), 415 (1994).
[121] H. Li, A. D. Robertson, and J. H. Jensen, Proteins-Structure Function and Bioinformatics **61** (4), 704 (2005).
[122] B. Jimenez, M. Piccioli, J. M. Moratal, and A. Donaire, Biochemistry **42** (35), 10396 (2003).
[123] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, The Journal of Chemical Physics **72** (1), 650 (1980).
[124] F. J. Munoz-Lopez, E. F. Beltran, S. Diaz-Moreno, I. Diaz-Moreno, G. Subias, M. A. De la Rosa, and A. Diaz-Quintana, Febs Letters **584** (11), 2346.
[125] S. D. George, L. Basumallick, R. K. Szilagyi, D. W. Randall, M. G. Hill, A. M. Nersissian, J. S. Valentine, B. Hedman, K. O. Hodgson, and E. I. Solomon, Journal of the American Chemical Society **125** (37), 11314 (2003).
[126] U. Ryde, M. H. M. Olsson, K. Pierloot, and B. O. Roos, Journal of Molecular Biology **261** (4), 586 (1996).
[127] K. Ando, Journal of Physical Chemistry B **108** (12), 3940 (2004).
[128] S. Ghosh, X. J. Xie, A. Dey, Y. Sun, C. P. Scholes, and E. I. Solomon, Proceedings of the National Academy of Sciences of the United States of America **106** (13), 4969 (2009).
[129] A. M. Nersissian, C. Immoos, M. G. Hill, P. J. Hart, G. Williams, R. G. Herrmann, and J. S. Valentine, Protein Science **7** (9), 1915 (1998).
[130] M. T. Giudici-Orticoni, F. Guerlesquin, M. Bruschi, and W. Nitschke, Journal of Biological Chemistry **274** (43), 30365 (1999).
[131] A. H. Hunt, A. Toypalmer, N. Assamunt, J. Cavanagh, R. C. Blake, and H. J. Dyson, Journal of Molecular Biology **244** (4), 370 (1994); J. McGinnis, W. J. Ingledew, and A. G. Sykes, Inorg. Chem. **25** (21), 3730 (1986).
[132] P. F. Su and H. Li, Journal of Chemical Physics **131** (1), 014102 (2009).
[133] M. C. Machczynski, H. B. Gray, and J. H. Richards, Journal of Inorganic Biochemistry **88** (3-4), 375 (2002).

[134] J. H. Jensen and H. Li, in *Encyclopedia of Inorganic Chemistry* (John Wiley & Sons, Ltd, 2006).
[135] B. G. Malmstrom and J. Leckner, Current Opinion in Chemical Biology **2** (2), 286 (1998); D. W. Randall, D. R. Gamelin, L. B. LaCroix, and E. I. Solomon, Journal of Biological Inorganic Chemistry **5** (1), 16 (2000).
[136] H. Li, N. M. Thellamurege, and D. Si, mss in preparation, code published in August, 2011 (2011).
[137] J. A. Pople, J. S. Binkley, and R. Seeger, International Journal of Quantum Chemistry **10** (S10), 1 (1976); C. Møller and M. S. Plesset, Physical Review **46** (7), 618 (1934); T. J. Lee and D. Jayatilaka, Chemical Physics Letters **201** (1-4), 1 (1993); J. S. Binkley and J. A. Pople, International Journal of Quantum Chemistry **9** (2), 229 (1975); T. J. Lee, A. P. Rendell, K. G. Dyall, and D. Jayatilaka, The Journal of Chemical Physics **100** (10), 7400 (1994); I. M. B. Nielsen and E. T. Seidl, Journal of Computational Chemistry **16** (10), 1301 (1995).
[138] M. J. Frisch, M. Head-Gordon, and J. A. Pople, Chemical Physics Letters **166** (3), 275 (1990).
[139] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, International Journal of Quantum Chemistry **16** (S13), 225 (1979).
[140] N. C. Handy and H. F. Schaefer III, The Journal of Chemical Physics **81** (11), 5031 (1984).
[141] G. D. Fletcher, M. S. Gordon, and R. S. Bell, Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) **107** (2), 57 (2002).
[142] C. M. Aikens, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, The Journal of Chemical Physics **124** (1), 014107 (2006).
[143] C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, Theoretical Chemistry Accounts **110** (4), 233 (2003).
[144] J. Tomasi, B. Mennucci, and R. Cammi, Chemical Reviews **105** (8), 2999 (2005).
[145] T. N. Truong and E. V. Stefanovich, Chemical Physics Letters **240** (4), 253 (1995).
[146] J. E. Rice and R. D. Amos, Chemical Physics Letters **122** (6), 585 (1985); E. D. Simandiras, R. D. Amos, and N. C. Handy, Chemical Physics **114** (1), 9 (1987).
[147] H. F. Schaefer and Y. Yamaguchi, Theochem-Journal of Molecular Structure **28**, 369 (1986).
[148] G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, in *Advances in Chemical Physics, Vol 110* (1999), Vol. 110, pp. 267.
[149] K. Ishimura, P. Pulay, and S. Nagase, Journal of Computational Chemistry **27** (4), 407 (2006); K. Ishimura, P. Pulay, and S. Nagase, Journal of Computational Chemistry **28** (12), 2034 (2007).
[150] C. M. Aikens and M. S. Gordon, The Journal of Physical Chemistry A **108** (15), 3103 (2004).
[151] R. b. Zhang, T. Zeegers-Huyskens, A. Ceulemans, and M. T. Nguyen, Chemical Physics **316** (1-3), 35 (2005).
[152] M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer, and W. Thiel, The Journal of Chemical Physics **128** (13), 134110 (2008).

[153] A. T. B. Gilbert, N. A. Besley, and P. M. W. Gill, The Journal of Physical Chemistry A **112** (50), 13164 (2008).
[154] P. Su and H. Li, The Journal of Chemical Physics **131** (1), 014102 (2009).
[155] F. J. Vesely, Journal of Computational Physics **24** (4), 361 (1977); M. Neumann, F. J. Vesely, O. Steinhauser, and P. Schuster, Molecular Physics **35** (3), 841 (1978); F. H. Stillinger and C. W. David, Journal of Chemical Physics **69** (4), 1473 (1978); P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, Nature **282** (5738), 459 (1979); A. Warshel, Journal of Physical Chemistry **83** (12), 1640 (1979).
[156] Y. Wang and H. Li, The Journal of Chemical Physics **133** (3), 034108 (2010).
[157] P. Hohenberg and W. Kohn, Physical Review B **136** (3B), B864 (1964); W. Kohn and L. J. Sham, Physical Review **140** (4A), 1133 (1965).
[158] P. A. M. Dirac, Mathematical Proceedings of the Cambridge Philosophical Society **26** (03), 376 (1930).
[159] S. H. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58** (8), 1200 (1980).
[160] J. P. Perdew and A. Zunger, Physical Review B **23** (10), 5048 (1981).
[161] A. D. Becke, The Journal of Chemical Physics **84** (8), 4524 (1986).
[162] A. D. Becke, The Journal of Chemical Physics **96** (3), 2155 (1992); A. D. Becke, The Journal of Chemical Physics **98** (7), 5648 (1993); A. D. Becke, The Journal of Chemical Physics **107** (20), 8554 (1997).
[163] A. D. Becke, The Journal of Chemical Physics **97** (12), 9173 (1992).
[164] A. D. Becke, The Journal of Chemical Physics **104** (3), 1040 (1996).
[165] J. P. Perdew and W. Yue, Physical Review B **33** (12), 8800 (1986).
[166] C. Adamo, J. Chem. Phys. **108** (2), 664 (1998).
[167] N. C. Handy and A. J. Cohen, Molecular Physics **99** (5), 403 (2001).
[168] X. Xu and W. A. Goddard, Proceedings of the National Academy of Sciences of the United States of America **101** (9), 2673 (2004).
[169] J. P. Perdew, Physical Review B **33** (12), 8822 (1986).
[170] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77** (18), 3865 (1996).
[171] C. Adamo, J. Chem. Phys. **116** (14), 5933 (2002).
[172] A. Becke, J. Chem. Phys. **88** (2), 1053 (1988).
[173] J. P. Perdew, presented at the Electronic Structure of Solids '91, 1991 (unpublished).
[174] F. Hamprecht, J. Chem. Phys. **109** (15), 6264 (1998).
[175] H. Schmider, J. Chem. Phys. **108** (23), 9624 (1998).
[176] B. J. Lynch, P. L. Fast, M. Harris, and D. G. Truhlar, The Journal of Physical Chemistry A **104** (21), 4811 (2000); B. J. Lynch, Y. Zhao, and D. G. Truhlar, The Journal of Physical Chemistry A **107** (9), 1384 (2003).
[177] N. E. Schultz, Y. Zhao, and D. G. Truhlar, The Journal of Physical Chemistry A **109** (49), 11127 (2005).
[178] R. G. Parr, Annual Review of Physical Chemistry **34**, 631 (1983); T. Ziegler, Chemical Reviews **91** (5), 651 (1991); R. G. Parr and W. Yang, Annual Review of Physical Chemistry **46** (1), 701 (1995); X. Gonze, Physical Review A **52** (2), 1096 (1995); W. Kohn, A. D. Becke, and R. G. Parr, The Journal of Physical

Chemistry **100** (31), 12974 (1996); E. J. Baerends and O. V. Gritsenko, The Journal of Physical Chemistry A **101** (30), 5383 (1997); P. E. M. Siegbahn and M. R. A. Blomberg, Annual Review of Physical Chemistry **50**, 221 (1999); P. Geerlings, F. De Proft, and W. Langenaeker, Chemical Reviews **103** (5), 1793 (2003); K. Burke, J. Chem. Phys. **123** (6), 062206 (2005); S. r. F. Sousa, P. A. Fernandes, and M. J. o. Ramos, The Journal of Physical Chemistry A **111** (42), 10439 (2007).

179 H. Li, H. M. Netzloff, and M. S. Gordon, Journal of Chemical Physics **125** (19), 194103 (2006).

180 H. Li and M. S. Gordon, Journal of Chemical Physics **126** (12), 124112 (2007).

181 P. Pulay, Molecular Physics **17** (2), 197 (1969).

182 C. B. Nielsen, O. Christiansen, K. V. Mikkelsen, and J. Kongsted, Journal of Chemical Physics **126** (15), 18 (2007).

183 F. Furche and R. Ahlrichs, Journal of Chemical Physics **117** (16), 7433 (2002); F. Furche and R. Ahlrichs, Journal of Chemical Physics **121** (24), 12772 (2004).

184 C. Van Caillie and R. D. Amos, Chemical Physics Letters **308** (3-4), 249 (1999); C. Van Caillie and R. D. Amos, Chemical Physics Letters **317** (1-2), 159 (2000).

185 E. B. Wilson, J. C. Decius, and R. C. Cross, *Molecular Vibrations*. (McGraw-Hill, New York, 1955); A. Komornicki and R. L. Jaffe, The Journal of Chemical Physics **71** (5), 2150 (1979).

186 J. P. M. P. Berendsen H. J. C. , W. F. van Gunsteren, J. Hermans, (1981,).

187 C. D. Berweger, W. F. van Gunsteren, and F. Müller-Plathe, Chemical Physics Letters **232** (5-6), 429 (1995).

188 H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, The Journal of Physical Chemistry **91** (24), 6269 (1987).

189 W. Jorgensen, J. Chem. Phys. **79** (2), 926 (1983).

190 E. Neria, J. Chem. Phys. **105** (5), 1902 (1996).

191 W. L. Jorgensen and J. D. Madura, Molecular Physics **56** (6), 1381 (1985).

192 H. Horn, J. Chem. Phys. **120** (20), 9665 (2004).

193 M. Mahoney, J. Chem. Phys. **112** (20), 8910 (2000).

194 S. Rick, J. Chem. Phys. **120** (13), 6085 (2004).

195 O. Matsuoka, J. Chem. Phys. **64** (4), 1351 (1976); K. Watanabe and M. L. Klein, Chemical Physics **131** (2-3), 157 (1989); Y. Liu and T. Ichiye, The Journal of Physical Chemistry **100** (7), 2723 (1996); V. Buch, P. Sandler, and J. Sadlej, The Journal of Physical Chemistry B **102** (44), 8641 (1998); M. Levitt, M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett, The Journal of Physical Chemistry B **101** (25), 5051 (1997); W. L. Jorgensen and C. Jenson, Journal of Computational Chemistry **19** (10), 1179 (1998); L. X. Dang, The Journal of Physical Chemistry B **102** (3), 620 (1998).

196 M. Chiba, T. Tsuneda, and K. Hirao, The Journal of Chemical Physics **124** (14), 144106 (2006); M. Chiba, T. Tsuneda, and K. Hirao, Chemical Physics Letters **420** (4-6), 391 (2006).

197 *Recommended Reference Materials for the Realization of Physicochemical Properties*, edited by K. N. Marsh (Blackwell Scientific Publications, Oxford, 1987).

198 T. Head-Gordon and G. Hura, Chemical Reviews **102** (8), 2651 (2002).

[199] K. N. Walzl, C. F. Koerting, and A. Kuppermann, The Journal of Chemical Physics **87** (7), 3796 (1987).
[200] N. S. Bayliss and E. G. McRae, The Journal of Physical Chemistry **58** (11), 1006 (1954).