

Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations

Qiang Cui

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Martin Karplus

*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138 and
Laboratoire de Chimie Biophysique, ISIS Université Louis Pasteur, 67000 Strasbourg, France*

(Received 13 July 1999; accepted 30 September 1999)

Analytical second derivatives for combined QM/MM calculations have been formulated and implemented in the CHARMM program interfaced with the *ab initio* quantum mechanical GAMESS and CADPAC programs. This makes possible evaluation of vibrational frequencies and infrared intensities in large systems that cannot be treated effectively by QM or MM alone; examples are polarizable molecules in solution and substrates or transition states in enzymes. Test calculations on a number of systems, including formamide in water, butanol, a model transition state structure for triosephosphate isomerase and the active site model of myoglobin, show that the MM description of the environment can capture much of its polarization effects on the QM region. Thus the implementation of analytical second derivatives within the QM/MM framework has considerable potential for the study of large systems. © 2000 American Institute of Physics. [S0021-9606(99)30548-1]

I. INTRODUCTION

Dramatic progress has been made in the field of computational chemistry in recent years. Although highly accurate (“chemical accuracy”) calculations are still limited to systems of less than ten heavy atoms,¹ useful studies can now be made for significantly larger systems. One example is the application of linear scaling methods² in a semiempirical framework to proteins and nucleic acids.³ However, in most cases of interest, only part of the system needs to be treated quantum mechanically and that part may require high level density functional or *ab initio* methods. For such problems (e.g., reactions in solution or in enzymes), combined quantum mechanical and classical mechanical (QM/MM) methods,⁴ are well suited.⁵ For a QM region of limited size (<50 atoms), the QM/MM can easily be applied to systems containing thousands of atoms. Both with semiempirical and *ab initio* quantum mechanics, QM/MM methods have been shown to provide fundamental insights into the mechanisms of chemical and enzymatic reactions.^{4,5} Most of the studies have concentrated on the potential energy surface, though solvation effects on pK_a values,⁶ and solvent-induced spectral shifts in the UV absorption of organic molecules⁷ and charge polarization have also been studied.⁸ However, quantum mechanical calculations can provide additional information. It is of considerable interest, therefore, to develop methods for evaluating other properties. Such an extension of the methodology corresponds to that introduced for purely QM calculations of small molecules, the study of so-called “weak interaction,” in the early 1960.⁹ As a first step in that direction, we have implemented methods to calculate analytical second derivatives within the QM/MM framework. Emphasis is given to *ab initio* and density functional QM

methodology. A subsequent paper describes the formulation for chemical shift calculations. Extensions to other properties are planned.

Normal mode analyses are of great importance for many applications. The frequencies obtained through such calculations can be directly related to experimental infrared and/or Raman measurements, as well as thermodynamic properties. The derived normal modes can be used in characterizing the dynamic behavior of fluids¹⁰ and biomolecules.¹¹ Also, normal mode analysis is important for evaluating the rates of reactions within the framework of variational transition state theories¹² and the reaction path Hamiltonian method.¹³

With the state of *ab initio* techniques, meaningful normal mode analysis can be carried out accurately for rather complex molecules, including nucleic acid bases¹⁴ and porphine,¹⁵ for example. Such studies have been of considerable help in spectral assignments. The high cost of *ab initio* and density functional methods prohibits calculations for macromolecules. With molecular mechanics (MM), one can readily perform normal mode calculations for proteins and nucleic acids, particularly with the recent progress in diagonalization algorithms.¹⁶ However, there are many cases where local polarization effects are important and/or bond making, bond breaking transition states are involved so that standard MM force fields are not applicable. Also, MM methods are difficult to apply to systems involving transition metals¹⁷ or excited states. It is attractive, therefore, to introduce normal mode techniques into QM/MM approaches, so that one can treat the active part of a large system with a quantum mechanical method, and the environment in a less sophisticated fashion with molecular mechanics.

In Sec. II we outline the theory of analytical second derivatives for QM/MM formulations and describe its imple-

mentation in the CHARMM program interfaced with the GAMESS and CADPAC quantum mechanical programs. Section III presents some test cases to illustrate the effectiveness of QM/MM normal mode analysis. The choice of the QM/MM partitioning and the use of link atoms are considered. A concluding discussion is presented in Sec. IV.

II. THEORY AND IMPLEMENTATION

A. Analytical Hessian matrix element

In the present work, the QM/MM Hamiltonian is written in the familiar form:

$$H = H^{\text{QM}} + H^{\text{QM/MM}} + H^{\text{MM}}, \quad (1)$$

where

$$H^{\text{QM/MM}} = \sum_d^N \sum_c^M \frac{-Q_c e}{|\mathbf{R}_c - \mathbf{r}_d|} + V_{\text{Nucl}}^{\text{QM/MM}} + V_{\text{van}}^{\text{QM/MM}}. \quad (2)$$

Here N is the number of electrons in the QM part, and M is the number of MM atoms. The last two terms in Eq. (2) are the nuclear–nuclear repulsion and van der Waals interactions, respectively, between QM and MM atoms. Correspondingly, the total energy of the system can be written as

$$E = \langle \Phi | H^{\text{QM}} + \sum_d^N \sum_c^M \frac{-Q_c e}{|\mathbf{R}_c - \mathbf{r}_d|} | \Phi \rangle + E_{\text{Nucl}}^{\text{QM/MM}} + E_{\text{van}}^{\text{QM/MM}} + E^{\text{MM}} \quad (3a)$$

$$= E_{el}^{\text{QM}} + E_{el}^{\text{QM/MM}} + E_{\text{Nucl}}^{\text{QM/MM}} + E_{\text{van}}^{\text{QM/MM}} + E^{\text{MM}}, \quad (3b)$$

where Φ is the wave function of the system.

For a normal mode analysis, the second derivative of the energy with respect to nuclear displacements, the Hessian matrix, is required. The Hessian matrix elements for the last three terms in Eq. (3b) are straightforward and will not be discussed further. For the second derivatives of the first two terms in Eq. (3b), we illustrate the method in detail using Hartree–Fock (HF) QM level. Since only one-electron operators are introduced in the present QM/MM formulation, extensions to DFT is strictly parallel and is summarized briefly.

1. Hartree–Fock/MM case

In general, for electronic structure theories with nonexact wave functions, the Hellmann–Feynman theorem¹⁸ is not satisfied and the derivatives of the electronic energy involve both the derivatives of the Hamiltonian and that of the wave function. Since the latter is most often expressed in terms of molecular orbitals which in turn depend on atom-centered basis functions, the energy derivatives contain contributions from the derivatives of molecular integrals as well from the derivatives of the molecular orbital (MO) coefficients. For configuration interaction methods, the derivatives of the CI coefficients also make contributions, in general. Many simplifications can be achieved if the method is variational, where only the n th derivative of the variational parameters are needed for the $(2n+1)$ th order derivative of energy due to the Wigner theorem.^{19(b)} For example, only the first order derivatives of the molecular orbital coefficients are required

to compute the second derivative of the energy in the Hartree–Fock theory since the MO coefficients are variationally optimized.

For a closed-shell RHF wave function without MM perturbations, the total electronic energy is written in the usual form:

$$E_{el}^{\text{HF}} = 2 \sum_i^{\text{d.o.}} h_{ii} + \sum_{i,j}^{\text{d.o.}} [2(ii|jj) - (ij|ij)], \quad (4)$$

where the h_{ii} and $(ij|kl)$ are one- and two-electron integrals and the notation “d.o.” indicates that the summation is restricted to doubly occupied MO’s. The variational condition in terms of the Fock matrix is

$$F_{ij} = h_{ij} + \sum_k [2(ij|kk) - (ik|jk)] = \varepsilon_i \delta_{ij}, \quad (5)$$

with ε_i being the molecular orbital eigenvalue. The second derivative of Eq. (4) with respect to the nuclear displacement a and b can be written¹⁹

$$\begin{aligned} \frac{\partial^2 E_{el}}{\partial a \partial b} = & 2 \sum_i^{\text{d.o.}} h_{ii}^{ab} + \sum_{i,j}^{\text{d.o.}} \{2(ii|jj)^{ab} - (ij|ij)^{ab}\} \\ & - 2 \sum_i^{\text{d.o.}} S_{ii}^{ab} \varepsilon_i - 2 \sum_i^{\text{d.o.}} \eta_{ii}^{ab} \varepsilon_i \\ & + 4 \sum_i^{\text{all}} \sum_j^{\text{d.o.}} (U_{ij}^b F_{ij}^a + U_{ij}^a F_{ij}^b) \\ & + 4 \sum_i^{\text{al}} \sum_j^{\text{d.o.}} U_{ij}^a U_{ij}^b \varepsilon_i + 4 \sum_{i,k}^{\text{all}} \sum_{jl}^{\text{d.o.}} U_{ij}^a U_{kl}^b A_{ij,kl}, \end{aligned} \quad (6)$$

where

$$\eta_{ij}^{ab} = \sum_m^{\text{all}} (U_{im}^a U_{jm}^b + U_{im}^b U_{jm}^a - S_{im}^a S_{jm}^b - S_{im}^b S_{jm}^a), \quad (7)$$

and

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk) \quad (8)$$

by making use of the variational condition [Eq. (5)] and the orthonormality condition on the MO’s.

In Eq. (6), “all” indicates summation over occupied and unoccupied orbitals and the h_{ij}^{ab} , $(ij|kl)^{ab}$, S_{ij}^{ab} , and F_{ij}^a are the partial derivatives of one-electron matrix elements, two-electron integrals, overlap matrix elements and Fock matrix elements, respectively, these terms involve only the derivatives of the one- or two-electron operator and basis functions.^{19(b)}

To obtain the derivatives of the MO coefficients, the U^a matrix is introduced. It is defined in terms of the molecular orbital coefficients ($C_{i\mu}$).

$$\frac{\partial C_{i\mu}}{\partial a} = \sum_j^{\text{all}} U_{ij}^a C_{j\mu}, \quad (9)$$

where the μ denotes atomic orbitals. The U^a matrix is obtained from the coupled perturbed HF (CPHF) equations, which are considered at the end of this section. As mentioned

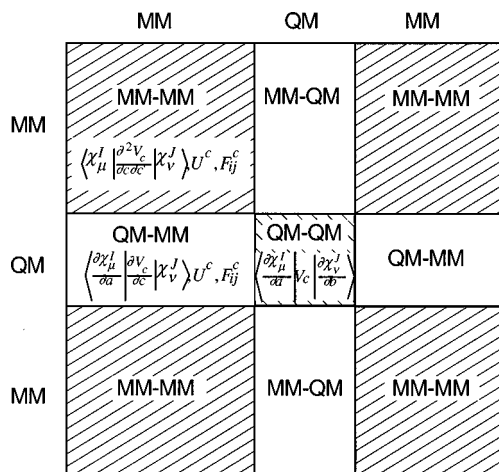


FIG. 1. The structure of the Hessian matrix for the QM/MM electronic energy and explicit MM contributions (see text). The V_c is the one-electronic operator corresponding to the QM/MM electronic interaction, $V_c = -eQ_c/|\mathbf{R}_c - \mathbf{r}|$. The U^c, F_{ij}^c are the CPHF solutions and Fock matrix integral-derivatives corresponding to the MM degrees of freedom, respectively. The χ_μ^I is the μ th atomic basis function on QM nucleus I .

earlier, only the first derivatives of the $C_{i\mu}$ appear in the second derivative of RHF energy due to the Wigner theorem.^{19(b)}

In the QM/MM method, additional one-electron operators are introduced corresponding to the interactions between the electrons and the partial charges on MM atoms, as in Eq. (2). Consequently, the second derivative of E_{el} has three types of matrix elements: the QM–QM, QM–MM (MM–QM) and MM–MM blocks, as illustrated schematically in Fig. 1. The expressions for the matrix elements in each block can be derived from Eq. (6). There are two general types of contributions that originate from the MM atoms. They are the derivatives of the QM/MM one-electron integrals and the terms involving the U^a matrix elements corresponding to the MM degrees of freedom. The first contribution arises from the direct electrostatic interaction between the QM electrons and the MM point charges, and the second contribution originates from the variation of the MO coefficients due to the MM atomic displacement. The required MM quantities for each block are summarized in Fig. 1.

For the QM–QM block, the only term in Eq. (6) that explicitly involves new contributions is h_{ij}^{ab} , which is augmented with the following integral derivative:

$$\langle \partial \chi_\mu^I / \partial a | \frac{-eQ_c}{|\mathbf{R}_c - \mathbf{r}|} | \partial \chi_\nu^J / \partial b \rangle \delta_{aI} \delta_{bJ},$$

$$a \subseteq \text{QM}, \quad b \subseteq \text{QM}, \quad c \subseteq \text{MM}. \quad (10)$$

The $\chi_\mu^I(\chi_\nu^J)$ denotes atomic basis function $\mu(\nu)$ centered on atom $I(J)$ and the symbol δ_{aI} should be understood as “the derivative involving atom I in the direction (X, Y, Z) specified by a .”

For the QM–MM (MM–QM) block and the MM–MM block, the terms involving the overlap matrix derivatives and the two-electron integral derivatives in Eq. (6) do not appear because there are no basis functions on the MM atoms. The h_{ij}^{ab} are left with terms of the form:

$$\langle \partial \chi_\mu^I / \partial a | \frac{\partial}{\partial b} \left(\frac{-eQ_c}{|\mathbf{R}_c - \mathbf{r}|} \right) | \chi_\nu^J \rangle \delta_{aI} \delta_{bC}, \quad a \subseteq \text{QM},$$

$$b \subseteq \text{MM}, \quad c \subseteq \text{MM}, \quad (11)$$

$$\langle \chi_\mu^I | \frac{\partial^2}{\partial a \partial b} \left(\frac{-eQ_c}{|\mathbf{R}_c - \mathbf{r}|} \right) | \chi_\nu^J \rangle \delta_{bC} \delta_{aC}, \quad a \subseteq \text{MM},$$

$$b \subseteq \text{MM}, \quad c \subseteq \text{MM}. \quad (12)$$

In most quantum chemical programs, the above terms are evaluated using translational invariance.²⁰ Defining $V_c = -eQ_c/|\mathbf{R}_c - \mathbf{r}|$, we have

$$\langle \partial \chi_\mu^I / \partial a | \frac{\partial V_c}{\partial b} | \chi_\nu^J \rangle \delta_{aI} \delta_{bC}$$

$$= -\langle \partial^2 \chi_\mu^I / \partial a \partial b | V_c | \chi_\nu^J \rangle \delta_{aI} \delta_{bI}$$

$$- \langle \partial \chi_\mu^I / \partial a | V_c | \partial \chi_\nu^J / \partial b \rangle \delta_{aI} \delta_{bJ}, \quad (13)$$

and

$$\langle \chi_\mu^I | \frac{\partial^2 V_c}{\partial a \partial b} | \chi_\nu^J \rangle \delta_{aC} \delta_{bC}$$

$$= \langle \partial^2 \chi_\mu^I / \partial a \partial b | V_c | \chi_\nu^J \rangle \delta_{aI} \delta_{bI}$$

$$+ \langle \partial \chi_\mu^I / \partial a | V_c | \partial \chi_\nu^J / \partial b \rangle \delta_{aI} \delta_{bJ}$$

$$+ \langle \partial \chi_\mu^I / \partial b | V_c | \partial \chi_\nu^J / \partial a \rangle \delta_{aJ} \delta_{bI}$$

$$+ \langle \chi_\mu^I | V_c | \partial^2 \chi_\nu^J / \partial a \partial b \rangle \delta_{aJ} \delta_{bJ}. \quad (14)$$

For these two blocks, the last four terms in Eq. (6) are nonvanishing and involve the U^a matrices and Fock derivative matrix, where now a, b can be either QM or MM for the QM–MM block, and only MM for the MM–MM block.

The slow step of the analytical Hessian calculation is the determination of the U^a matrix elements from the coupled perturbed Hartree–Fock (CPHF) equations.¹⁹ They have the form

$$(\varepsilon_j - \varepsilon_i) U_{ij}^a - \sum_k^{\text{vir}} \sum_l^{\text{d.o.}} A_{ij,kl} U_{kl}^a = B_{ij}^a, \quad (15)$$

where

$$B_{ij}^a = F_{ij}^a - S_{ij}^a \varepsilon_j - \sum_{kl}^{\text{d.o.}} S_{kl}^a [2(ij|kl) - (ik|jl)], \quad (16)$$

and “vir” refers to the unoccupied orbitals. It should be noted that Eq. (15) only needs to be solved for the independent occupied–virtual block, while the redundant occupied–occupied block that also contributes to Eqs. (6) and (7) is determined from the derivative of the orthonormality condition of the MO’s; that is,

$$U_{ij}^a + U_{ji}^a + S_{ij}^a = 0 \Rightarrow U_{ij}^a = U_{ji}^a = -\frac{1}{2} S_{ij}^a (i, j \in \text{d.o.}). \quad (17)$$

For the MM degrees of freedom, the same CPHF equations are used but there are two major simplifications due to the fact that there are no basis functions on the MM atoms. First, the inhomogeneity is simplified, i.e.,

$$B_{ij}^a = F_{ij}^a = h_{ij}^a = \langle i | \frac{\partial}{\partial a} \left(\frac{-eQ_a}{|\mathbf{R}_a - \mathbf{r}|} \right) | j \rangle, \quad a \subseteq \text{MM}, \quad (18)$$

since $S_{ij}^{\text{MM}} = 0$. For the same reason, the redundant block of U^a is zero according to Eq. (17) so only the coupling between the occupied and unoccupied orbitals has to be considered.

2. DFT/MM case

In density functional theories, the two-electron exchange integrals in the Hartree–Fock theory are replaced with the exchange–correlation functional, so that Eq. (4) for the electronic energy becomes

$$E_{\text{el}} = 2 \sum_i^{\text{d.o.}} h_{ii} + 2 \sum_{ij}^{\text{d.o.}} (ij|jj) + E_{\text{xc}}(\rho). \quad (19)$$

The variational condition is given as

$$F_{ij}^{\text{KS}} = h_{ij} + 2 \sum_k (ij|kk) + \langle i | \delta E_{\text{xc}} / \delta \rho | j \rangle = \varepsilon_i \delta_{ij}, \quad (20)$$

where $\delta E_{\text{xc}} / \delta \rho$ is the exchange–correlation potential defined as the functional derivative of E_{xc} .

Taking the second derivative of Eq. (19), and making manipulations similar to those used in the RHF case, one obtains an expression for the Hessian matrix element that is very similar to Eq. (6) except for the contributions from the exchange–correlation potential to the Fock matrix derivatives and the $A_{ij,kl}$ terms. The same applies to the coupled-perturbed Kohn–Sham (CPKS) equations in comparison to the CPHF equations. We refer the reader to Ref. 21 for detailed expressions.

To summarize, we have described in detail the algorithm for RHF/MM and DFT/MM analytical second derivative calculations based on the results for the corresponding purely QM methods. Two types of extra contributions need to be computed. They are the “direct” terms originating from the explicit derivatives of the QM/MM electrostatic interaction, and the “indirect” terms arising from the dependence of the MO coefficients on the MM atomic positions. In the similar spirit, it is straightforward to compute the QM/MM Hessian for correlated QM methods, such as Møller–Plesset perturbation theory,²² by modifying existing algorithms for the pure QM second derivative evaluation.²³

B. Normal mode vectors and Infrared (IR) intensities

After obtaining the QM/MM second derivative matrix, vibrational frequencies and normal modes are obtained through diagonalization of the mass-weighted Hessian matrix \mathbf{H} ,²⁴

$$\mathbf{H}\mathbf{L}_k = [\mathbf{M}^{-(1/2)}(\nabla^2 E)\mathbf{M}^{-(1/2)}]\mathbf{L}_k = \lambda_k \mathbf{L}_k, \quad (21)$$

where \mathbf{M} is the diagonal mass matrix, and \mathbf{L}_k and λ_k are the k th eigenvector and eigenvalue, respectively.

In addition to vibrational frequencies, IR intensities of the modes are also of interest. The IR intensity for mode I_k is proportional to the projection of the molecular dipole derivative on to the corresponding eigenvector \mathbf{L}_k ; ²⁵ that is,

$$I_k = \frac{N\pi}{3c^2} |\nabla \bar{\mu} \cdot \mathbf{L}_k|^2, \quad (22)$$

where the N is Avogadro’s number, c is the speed of light.

The nuclear contribution from the QM and MM atoms to the dipole moment derivative is trivial to calculate and will not be discussed further. The electronic component of the dipole derivative can be obtained by taking the nuclear derivative of the electronic dipole moment; that is^{25(b)}

$$\begin{aligned} \frac{\partial \mu_f^{\text{elec}}}{\partial a} &= - \frac{\partial^2 E_{\text{elec}}}{\partial a \partial f} \\ &= -4 \sum_i^{\text{virt.}} \sum_j^{\text{d.o.}} U_{ij}^a h_{ij}^f + 2 \sum_{ij}^{\text{d.o.}} S_{ij}^a h_{ij}^f - 2 \sum_i^{\text{d.o.}} h_{ii}^f, \end{aligned} \quad (23)$$

where f denotes the external electric field perturbation, h_{ij}^f and h_{ii}^f are the dipole integrals and the dipole derivative integrals, respectively, defined by²⁶

$$h_{ij}^f = \sum_{\mu\nu} C_{i\mu} C_{j\nu} \frac{\partial h_{\mu\nu}}{\partial f} = -e \sum_{\mu\nu} C_{i\mu} C_{j\nu} \langle \chi_\mu | r_f | \chi_\nu \rangle, \quad (24)$$

$$\begin{aligned} h_{ij}^{af} &= \sum_{\mu\nu} C_{i\mu} C_{j\nu} \frac{\partial^2 h_{\mu\nu}}{\partial a \partial f} \\ &= -e \sum_{\mu\nu} C_{i\mu} C_{j\nu} [\langle \partial \chi_\mu / \partial a | r_f | \chi_\nu \rangle + \langle \chi_\mu | r_f | \partial \chi_\nu / \partial a \rangle]. \end{aligned} \quad (25)$$

The e in Eqs. (24) and (25) stands for the absolute value of electronic charge.

For the dipole derivative with respect to the QM atoms, there is no explicit MM contribution. The effect of MM atoms is contained implicitly in the U^a matrix [see Eq. (15)]. For the dipole derivative with respect to the MM atoms, which do not have any basis functions, the last two terms of Eq. (23) vanish, and the remaining (first) term is readily calculated with the CPHF solutions for the MM degrees of freedom [i.e., Eq. (15)].

Correspondingly, one can obtain the MM contribution to the Raman intensity of vibrational modes, which is related to the nuclear derivative of polarizability tensors. This is not considered in the current work.

C. MM polarization kernel from the U^a matrix

It is interesting to investigate further the physical meaning of the U^a matrix for the MM degrees of freedom. It represents the extent to which MM atomic *displacements* polarize the wave function of the QM region, and consequently should be useful in the analysis of QM/MM interactions. In the literature of density functional theory, much attention has been paid to the so-called “polarization kernel” which represents the redistribution of the electron density, $\rho(\mathbf{r})$, with respect to the change in the external potential $v(\mathbf{r}')$,²⁷

$$\left(\frac{\partial \rho(\mathbf{r})}{\partial v(\mathbf{r}')} \right)_N = \frac{\partial^2 E}{\partial v(\mathbf{r}) \partial v(\mathbf{r}')}, \quad (26)$$

where the subscript N indicates conservation of the total number of electrons. The polarization kernel and related quantities can be derived using the principle of electronegativity or chemical potential equalization,²⁸ which forms the basis of charge sensitivity analysis.²⁹ In several recent studies, attempts have been made to describe polarization effects in molecular dynamics simulations within such a framework, in most cases in a semiempirical fashion.³⁰ Morita and Kato introduced a coarse grained version of the polarization kernel, $(\partial Q_A/\partial V_B)$, based on an atomic site representation in *ab initio* MO theory, where the Q_A and V_B are the charge population and electrostatic potential at atom A and B , respectively. It has been shown that this quantity is very useful in the analysis of charge fluctuation and treatment of solute polarization in molecular dynamics simulations.³¹

Within the QM/MM framework, it is of interest to examine quantities of similar nature, such as the change of charge population on the QM atoms as a function of MM atomic displacement. If one defines the charge population based on the Mulliken approximation:

$$Q_A = Z_A - \sum_{\mu \in A} \sum_{\nu} P_{\mu\nu} S_{\nu\mu}, \quad (27)$$

the expression for its derivative, $\partial Q_A/\partial a$, is

$$\begin{aligned} \partial Q_A/\partial a &= - \sum_{\mu \in A} \sum_{\nu} S_{\nu\mu}^a P_{\mu\nu} - \sum_{\mu \in A} \sum_{\nu} S_{\nu\mu} \frac{\partial P_{\mu\nu}}{\partial a} \quad (28a) \\ &= - \sum_{\mu \in A} \sum_{\nu} S_{\nu\mu}^a P_{\mu\nu} \\ &\quad - \sum_{\mu \in A} \sum_{\nu} S_{\nu\mu} \left[\sum_i^{\text{d.o.}} \sum_m^{\text{vir.}} U_{im}^a (C_{m\mu} C_{i\nu} + C_{i\mu} C_{m\nu}) \right], \quad (28b) \end{aligned}$$

where μ goes over the atomic bases on QM atom A and ν over all the basis functions, and a can refer to the displacement of a QM or MM atom. For the displacement a referring to the MM atoms, the first term in Eq. (28b) vanishes due to the zero derivatives of the overlap integrals. One can also define another set of polarization kernels, $\partial Q_A/\partial Q_C$, where Q_C represents the MM partial charge if one wishes to find the susceptibility of the charge on a QM atom to the charge on the MM atom. In this case, one has to solve the CPHF equation for the MM degrees of freedom with the MM atomic charge as the external perturbation, and substitute the resultant U^Q matrix into Eq. (28).

In short, the $\partial Q_A/\partial \xi$ matrices characterize the effect of MM atom motion ($\xi \sim$ MM displacement a) and composition ($\xi \sim$ MM charge Q_C) on the electron distribution on the QM atoms. Such information can be of great interest in the analysis of QM/MM interactions, such as for probing environmental effects on the active sites of enzymes.³²

D. Implementation

The algorithms to calculate the analytic Hessian matrix, IR intensity and MM polarization kernel for QM/MM methods have been implemented in CHARMM³³ which is interfaced³⁴ with two *ab initio* packages GAMESS³⁵ and

CADPAC.³⁶ The current implementation allows analytical second derivative calculations at the level of RHF/MM with CHARMM/GAMESS and RDFT/MM with CHARMM/CADPAC.

Before presenting test cases, we discuss a few technical issues that are unique to QM/MM methods.

1. Approximation to the CPHF equations

The CPHF equations, Eq. (15), have to be solved for each nuclear degree of freedom. Although this is true for QM calculations as well, for large system involving many MM degrees of freedom, the large number of linear equations would make the QM/MM Hessian calculation prohibitive. However, the inhomogeneity of Eq. (15), B_{ij}^a , given for MM atoms in Eq. (18), should be small in magnitude for MM atoms far from the QM region. The simplest approximation is to omit the CPHF equations for MM atoms far from the QM region; i.e., a cutoff can be introduced. Alternatively, one may obtain the approximate solution of the CPHF equations by using the simpler uncoupled expression in which $U_{ij}^a \approx F_{ij}^a/(\epsilon_i - \epsilon_j)$. In the current implementation, the user has the freedom of selecting any portion of the MM atoms in a macromolecule to perform either of these two levels of approximations. As is shown later, the uncoupled approximation yields improved results over complete neglect of the CPHF equations.

2. Treatment of link atoms

Link atoms are commonly introduced to saturate the valence of the QM part when the QM/MM partition is performed across bonds. In general, the link atoms do not have a strong influence on the results, provided that the partition is not performed across a polar or unsaturated chemical bond and that the partition is far enough from the region of interest. However, as has been pointed out, in each application the effect of the link atom should be checked by test calculations to determine the best partition scheme and verify that the error introduced is small.^{4(b),37,38}

For QM/MM Hessian calculations, the link atom introduces additional degrees of freedom so that their contribution to the Hessian needs to be projected out before the normal mode analysis. This can be done by using the following projection scheme,

$$\mathbf{H}' = (\mathbf{I} - \mathbf{L}_{\text{link}} \mathbf{L}_{\text{link}}^T) \mathbf{H} (\mathbf{I} - \mathbf{L}_{\text{link}} \mathbf{L}_{\text{link}}^T) \quad (29)$$

where the \mathbf{H} and \mathbf{H}' denotes the unprojected and projected Hessian, respectively; \mathbf{I} is the identity matrix, and \mathbf{L}_{link} is a vector containing 1 for the degrees of freedom corresponding to the link atoms and 0 otherwise.

Even with the projection, the link atom contaminates the low vibrational frequencies since projecting out the link atom degrees of freedom is equivalent to performing the normal mode analysis with a fixed link atom. Therefore, the overall translations and rotations of the actual molecule (i.e., the molecule without the link atom) are no longer in free space and they now correspond to vibrations in the molecule including the link atom. As a result, low vibrational frequencies are contaminated (mixed) with the overall translation

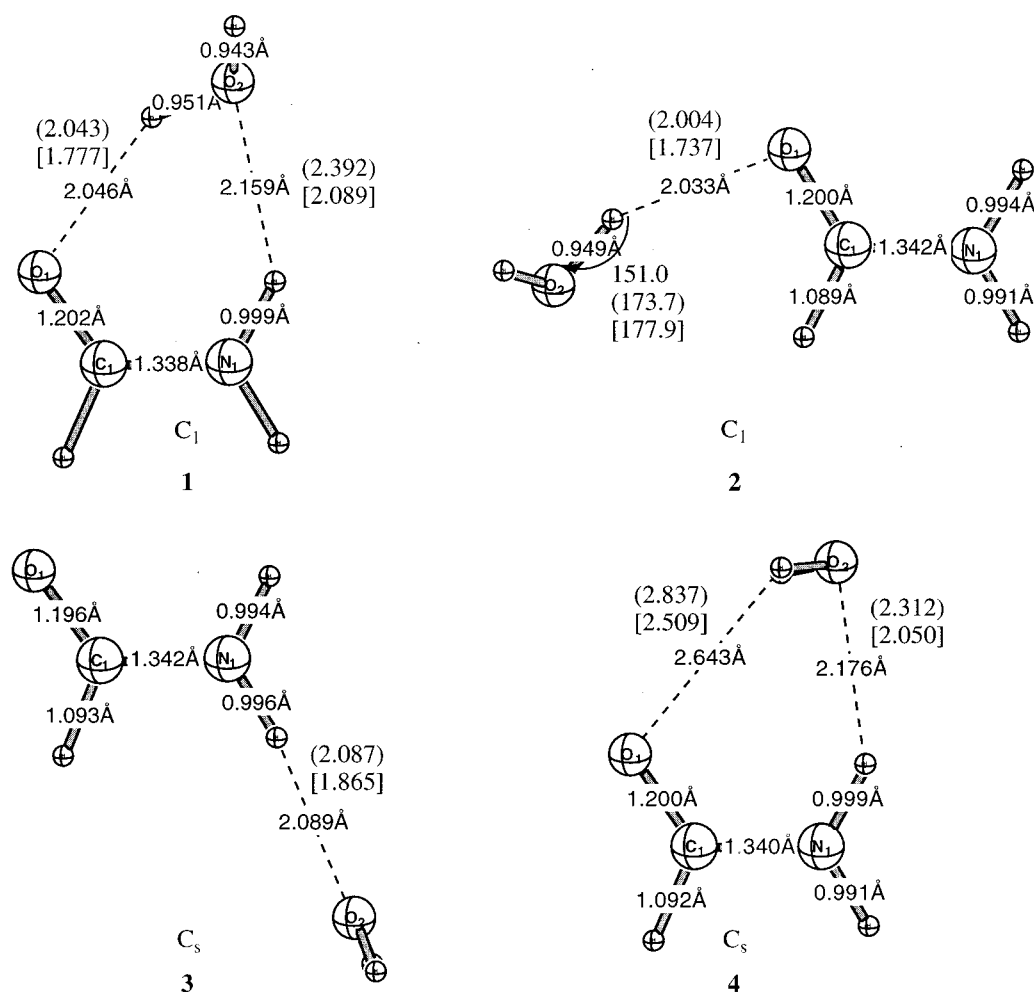


FIG. 2. Optimized structures for formamide–water complexes. The numbers without parentheses are obtained at the HF/6-31G(d,p) level. Numbers in parentheses are obtained at the HF/6-31G(d,p)/MM level with a modified TIP3P description for the water molecule. Numbers in brackets are the MM results using the CHARMM22 force field.

and rotation modes of the actual molecule. This can be resolved with a further projection step, in which the overall translation/rotation of the actual molecule is further projected out,

$$\mathbf{H}'' = \left(\mathbf{I} - \sum_i^{T,R} \mathbf{L}_i \mathbf{L}_i^T \right) \mathbf{H}' \left(\mathbf{I} - \sum_i^{T,R} \mathbf{L}_i \mathbf{L}_i^T \right), \quad (30)$$

where the \mathbf{L}_i are the eigenvectors correspond to the infinitesimal translation and rotation of the actual molecule,³⁹ Schmidt orthogonalized to the link atom motion \mathbf{L}_{link} . The resultant Hessian \mathbf{H}'' is used in the normal mode analysis. It should be understood that projection will only get rid of the *explicit* contribution from the link atom. The existence of the link atom still has implicit influence on the Hessian matrix elements of the boundary QM atoms. Such an effect is illustrated in the next section.

III. TEST CALCULATIONS

In this section, we use several examples to illustrate the usefulness of QM/MM frequency and intensity calculations and the effects of the approximate CPHF solution and the link atom.

A. Formamide–water

The shift of the vibrational frequency for a specific functional group in a molecular system is often of interest; an example is the frequency of substrate carbonyl in the enzyme triosephosphate isomerase, as compared with the corresponding frequency in solution.⁴⁰ From such measurements, information about the structural changes during a reaction or upon environmental alteration (such as mutation of a residue in the active site of a protein) can be derived. We have chosen the formamide–water system as an example. The purpose is to examine if the MM description of the water can account for the shift in the vibrational spectrum of formamide in different binding configurations. For this purpose, we compare the full QM results with QM/MM calculations. Geometry optimization and normal mode analysis have been carried out at both the full QM level (HF/6-31G(d,p)⁴¹) and the QM/MM level. In the QM/MM calculations, the formamide is treated with HF/6-31G(d,p) while the water molecule is described with a modified TIP3P model⁴² that has a flexible structure.^{33,43} For comparison, pure MM calculations have also been performed with the CHARMM22 force field.⁴⁴ To obtain a good QM/MM geometry, a specific set of van der

TABLE I. Selected properties for the formamide–water system.^a

Properties ^b	Formamide		Complex 1			Complex 2			Complex 3			Complex 4		
	QM	MM	QM	QM/MM	MM	QM	QM/MM	MM	QM	QM/MM	MM	QM	QM/MM	MM
Bind energy			9.6	8.0	10.3	6.8	6.2	7.3	5.6	5.9	7.0	7.9	6.0	7.9
C=O, ν	1784.1	1778.3	1751.6	1760.5	1775.7	1752.6	1773.6	1765.6	1773.4	1773.9	1739.8	1761.6	1768.3	1728.3
Inten.	509.9	185.7	540.1	557.8	204.7	557.2	553.3	199.8	542.0	516.7	199.8	558.2	566.2	225.3
N–H s-str, ν	3433.3	3412.9	3377.5	3383.5	3406.9	3431.1	3426.6	3413.3	3401.0	3351.8	3400.0	3382.8	3381.3	3406.3
Inten.	58.0	75.1	128.3	101.3	105.0	69.8	64.0	66.3	188.5	117.4	66.3	135.0	97.4	105.1
N–H as-str, ν	3563.4	3537.9	3541.3	3544.2	3529.8	3560.5	3556.0	3537.6	3534.7	3516.5	3528.9	3543.2	3543.1	3529.8
Inten.	65.8	219.1	102.2	79.9	208.1	72.6	71.2	220.1	158.4	70.2	220.1	108.8	77.5	215.3
RMS error ^c		64.2		35.2	57.3		40.1	60.2		25.9	64.0		42.0	55.0
RMS CPHF				7.0/20.1			4.2/20.5			3.6/12.2			4.7/16.5	
Error ^d				(2.0/4.5)			(1.3/6.0)			(1.4/2.3)			(1.7/2.9)	

^aQM here is HF/6-31G(d,p), MM is the CHARMM22 force field. In the QM/MM calculations, the formamide is treated with HF/6-31G(d,p), and the water is described with a modified flexible TIP3P model.

^bBinding energies of water to formamide are measured in kcal/mol and do not include zero point energy corrections. The vibrational frequencies are in wave number (cm^{-1}), and the IR intensities are in km/mol . The values for all the full QM vibrational frequencies and some of the QM/MM (see text) frequencies have been scaled by 0.8929.

^cThe RMS error in the QM/MM and MM vibrational frequencies compared to the full QM results, given in cm^{-1} .

^dThe RMS error in the QM/MM vibrational frequencies and IR intensities due to approximation for the CPHF contribution from the MM degrees of freedom, compared to the exact QM/MM results. The numbers without parentheses are obtained without the MM CPHF components, and these with parentheses are obtained with the uncoupled approximation (see text).

Waals parameters has been used for the oxygen, nitrogen, and the polar hydrogen atoms of the formamide molecule.⁴⁵ The well depth for O, N, and polar H is -0.16 , -0.17 , and -0.09 kcal/mol, respectively, and R_{\min} are 3.98, 4.30, and 0.60 Å, respectively.

(a) Formamide–water complex. The optimized structures at the full QM and QM/MM levels are shown in Fig. 2, and binding energies, selected vibrational frequencies and intensities are shown in Table I. Consistent with previous QM/MM study with semiempirical QM,^{4(b)} four different formamide–water stationary configurations were found. However, as pointed out by Gordon *et al.*,⁴⁶ only three are local minima and the fourth is actually a saddle point connecting complex 1 and its mirror image. As seen in Fig. 2 and Table I both the geometries and binding energies from the QM/MM calculations are in rather good agreement with the full QM results. The MM binding energies are also satisfactory, but the hydrogen bond distances are shorter by ~ 0.2 – 0.3 Å. This shortening is introduced to obtain correct liquid properties and accounts for the absence of dispersion terms in the HF calculations; for a discussion, see Ref. 44.

Since the water molecule is interacting with the formamide molecule in different ways in the four structures (Fig. 2), one expects to see the signature of these structural differences in the IR spectrum. In Fig. 3, we have shown the predicted IR stick spectrum of the four complexes at the QM/MM level, along with the full QM and MM results. Since the water molecule is interacting most strongly with the C=O group and the $-\text{NH}_2$ group, it is expected that the largest changes occur in the C=O stretch, and the N–H stretch modes, which are summarized in Table I. To make the comparison between full QM and MM results easier, we have scaled down the QM vibrational frequencies according to the well established factor, 0.89, for Hartree–Fock calculations with double-zeta polarization quality basis set.⁴⁷ For the QM/MM vibrational frequencies, we have performed the

same scaling to all the modes except the water O–H stretch and H–O–H bending modes.

In terms of frequencies, the CHARMM22 results are in approximate agreement with the scaled QM values, with the RMS error around 60 cm^{-1} for both the free formamide molecule and the four formamide–water complexes. The IR intensities from the pure MM calculations, however, are quite different from the QM results, as seen clearly in Fig. 3. The largest discrepancy lies in the IR intensity of the C=O vibration, which has been shown to be overestimated by around a factor of two at the Hartree–Fock level in the case of H_2CO .^{25(b)} With this taken into account, the differences between pure QM and MM IR intensities are still significant, with MM producing the wrong trend in some cases, e.g., for the antisymmetric N–H stretch shown in Table I. This is not entirely surprising since dipole derivatives are not considered during the development of the CHARMM force field. In the work of Krimm *et al.*,⁴⁸ vibrational spectra of large organic molecules are obtained by fitting MM atomic charge and internal coordinate dipole parameters against *ab initio* dipole derivatives to improve the MM IR intensities. Rather good agreement with full *ab initio* spectrum has been obtained in this way.

The QM/MM results, shown in Fig. 3 and Table I, demonstrate that the polarization effect from the water on the formamide have been reproduced approximately in most cases. This is particularly clear from the water-induced vibrational frequency shifts summarized in Table I. The RMS errors compared to the full QM results are around 30 cm^{-1} at the QM/MM levels, a notable improvement over the pure MM calculations. The overall good agreement between the full QM and QM/MM IR spectrum shown in Fig. 3 is not surprising, since the major part of the system (formamide) is treated at the QM level. As shown in Table I, the QM/MM method predicts not only the shift in the frequencies of the three “sensitive” modes but also their intensities. For com-

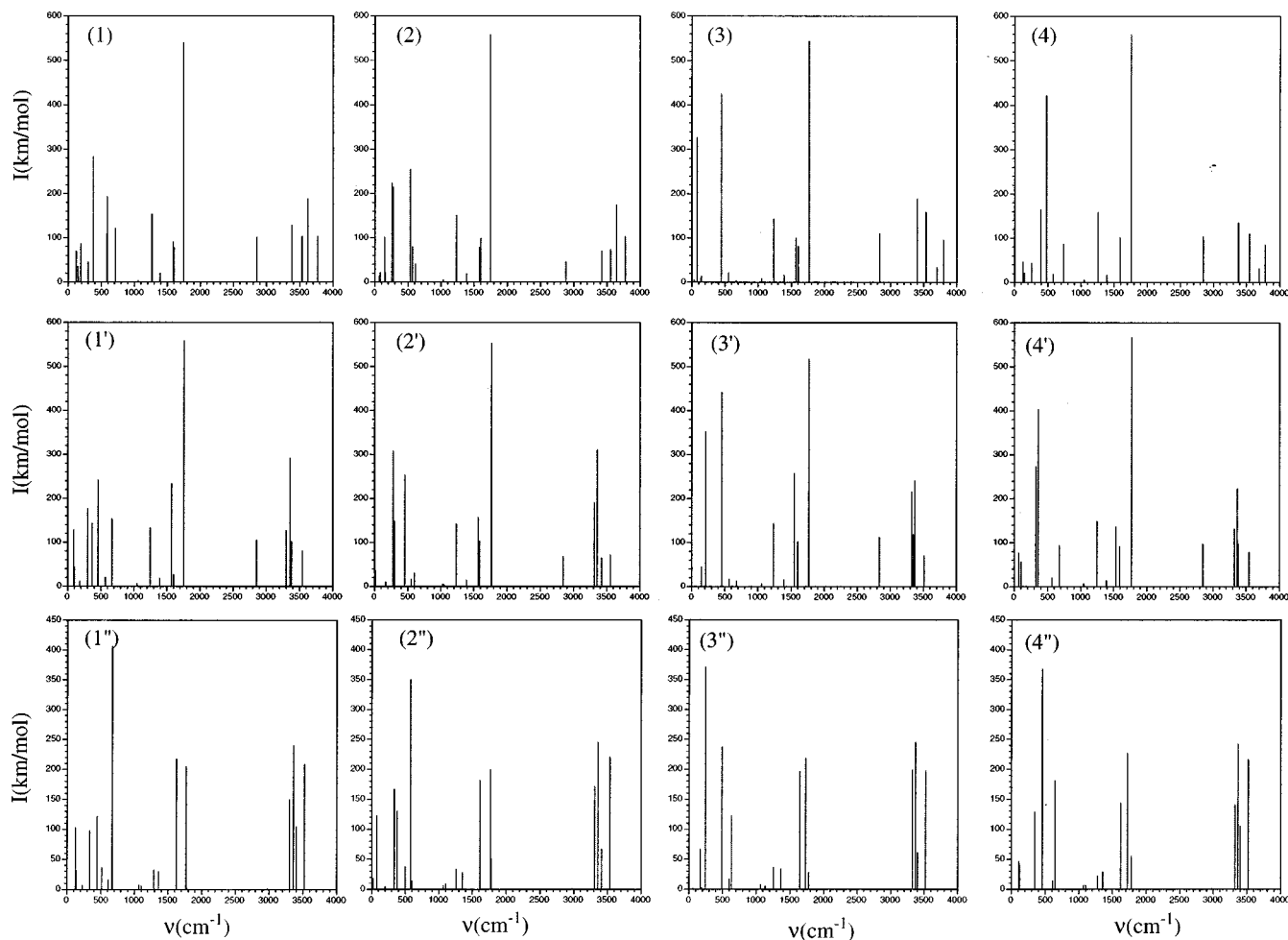


FIG. 3. Theoretically predicted IR spectrum for the formamide-water complexes in Fig. 2. Frequencies are in wave numbers, and intensities are in km/mol. (1)–(4) are from HF/6-31G(d,p) calculations; (1')–(4') are obtained at the QM/MM level with modified TIP3P water. (1'')–(4'') are pure MM results with the CHARMM22 force field. All the vibrational frequencies from full QM calculations are scaled by 0.8929, and some of the frequencies from the QM/MM calculations are scaled as well (see text). The frequencies and intensities for the C=O stretch, symmetric and asymmetric N–H stretches are also summarized in Table I.

plex 1, as the example, the predicted frequency shifts of the C=O stretch, symmetric and asymmetric N–H stretches at the QM/MM level are -23.6 , -49.8 , and -19.2 cm^{-1} , respectively, rather close to the values from full QM calculations (-32.5 , -55.8 , and -22.1 cm^{-1} , respectively). For comparison, the corresponding MM values are -2.6 , -6.0 , and -8.1 cm^{-1} , respectively, which are too small in magnitude. For the shifts in the IR intensities, similar trend applies. The QM/MM predicted shifts are 48.0, 43.3, and 14.1 km/mol for these three modes, respectively, and the full QM results are 30.2, 70.3, and 36.4 km/mol, respectively. The corresponding values from full MM calculations are 19.0, 29.9, and -11.0 km/mol, respectively, with the wrong sign for the asymmetric N–H stretch. Compared to the method of Krimm *et al.*,⁴⁸ the QM/MM approach requires less *ad hoc* fitting and is more generally applicable. The current QM/MM approach, which describes the polarization effect from the MM region in an explicit manner may also be preferable to the IMOMO method of Morokuma *et al.*, for which analytical Hessian has also been implemented.⁴⁹ Since the QM region is not polarized explicitly by the MM region in IMOMO, it is expected that the water induced shift in the

vibrational frequencies and intensities cannot be reproduced as well. This issue should be a less severe problem in the IMOMO method where the environmental polarization effect can be described at a lower QM level such as the semiempirical method.

As mentioned in Sec. II, one may choose to make approximations in solving the CPHF equations for the MM degrees of freedom, taking advantage of the relatively small magnitude of the effects. To illustrate this, we have carried out a set of normal mode analyses for the four formamide-water complexes with different approximations for the MM CPHF contributions. The CPHF equations correspond to the MM degrees of freedom are either completely omitted, or the uncoupled solutions are used. The RMS errors compared to exact QM/MM results are summarized in Table I. Clearly, the errors in the vibrational frequencies are small, on the order of several wave numbers, with either level of approximation. The error in the IR intensity, however, is more notable, and is much smaller if the uncoupled CPHF solution is used. There are two possible sources of errors for the IR intensity due to the approximation for the MM CPHF contribution. First, the MM CPHF solutions modify the dipole

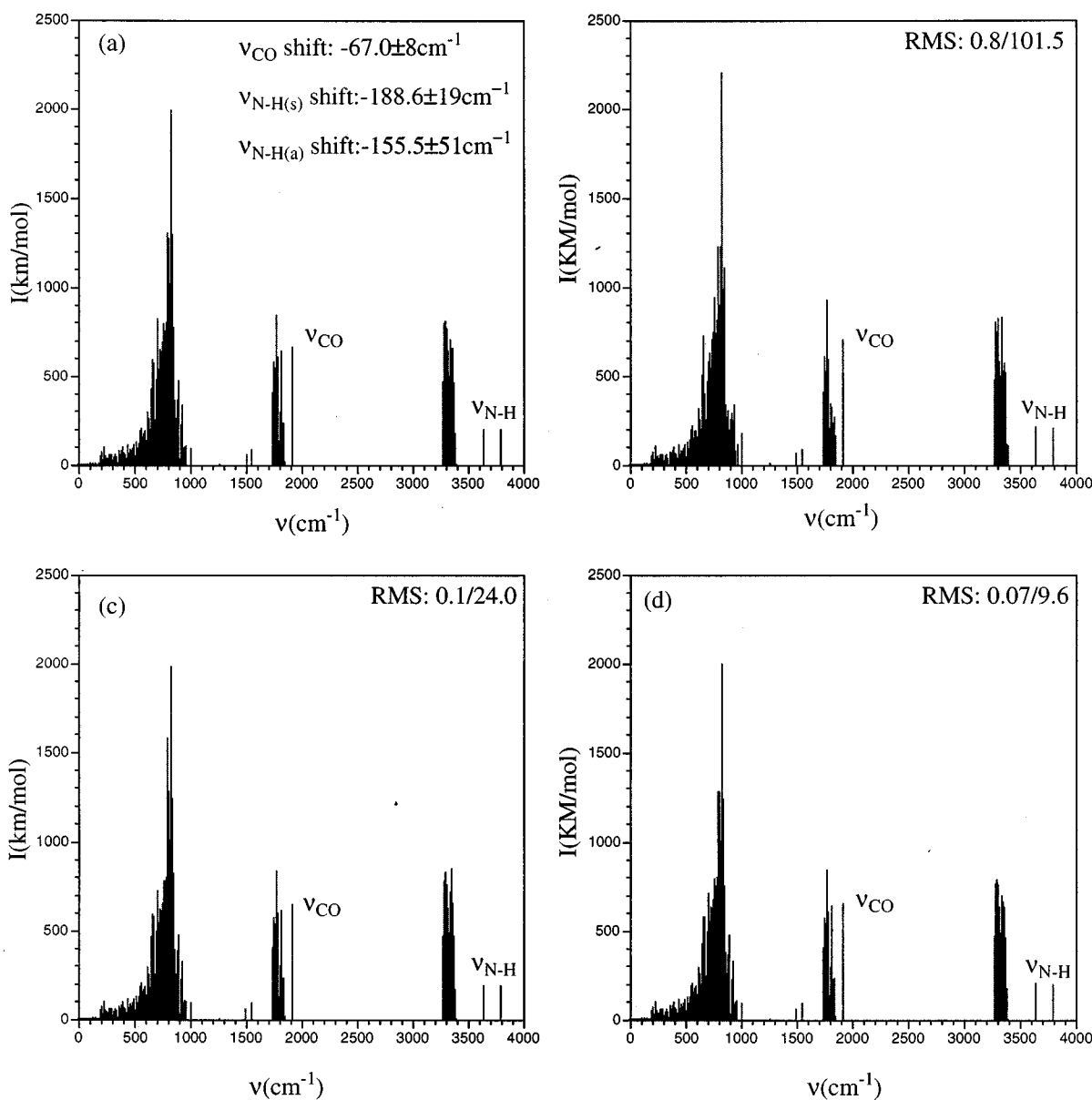


FIG. 4. Theoretically predicted IR spectrum for a formamide molecule in a 8 Å radius water sphere. The formamide molecule is treated with HF/6-31G(d,p), and the water molecules are described with the modified TIP3P model. The frequencies are not scaled in this case. (a) Exact QM/MM results. (b) The QM/MM results without the CPHF contribution from the MM degrees of freedom. (c) The QM/MM results with the uncoupled approximation for all the MM CPHF contribution. (d) The QM/MM results with the decoupled approximation for the water molecules 4 Å away from the oxygen atom in formamide. The RMS errors in frequencies and intensities, before and after the slash, respectively, are measured relative to the exact QM/MM results. The shift in the formamide vibrational frequencies are obtained from ten normal mode analyses with minimized configurations generated along a AM1/MM trajectory (see text).

derivatives with respect to the MM atomic displacements. Although generally small in magnitude ($\sim 10^{-1}$ – 10^{-2} a.u.), they can produce notable effects in IR intensities for modes involving mainly the MM atoms. Secondly, the MM CPHF solutions modify the Hessian matrix elements for the QM/MM and MM/MM blocks, and therefore affect the normal mode eigenvectors that also determine the IR intensity according to Eq. (22). For the formamide–water complexes, we found that the first source is mainly responsible for error in the high frequency modes corresponding to the water motion, while the second source is more severe in the cases of low frequency intermolecular modes.

(b) Solvated formamide molecule. To provide an addi-

tional test of the QM/MM second derivative method, we have carried out calculations for one formamide molecule solvated in a 8 Å water sphere. Ten configurations have been generated from a 40 ps molecular dynamics run at the AM1/MM level where the formamide is treated with AM1. Subsequently, HF/6-31G(d,p)/MM normal mode calculations are carried out following energy minimization for these configurations. A typical predicted IR spectrum is shown in Fig. 4(a), in which the vibrational frequencies are *not* scaled. The mean solvent induced shift in the C=O stretch, and symmetric and antisymmetric N–H stretches are also given. Clearly, the uncertainties of the frequency shift are still rather large particularly for the N–H stretches, and more points are needed to obtain fully converged results. Compared to the

simple water–formamide complexes discussed above, the shifts in the frequencies and intensities are larger in magnitude in the water sphere. For instance, the shifts of the C=O stretches, symmetric and asymmetric N–H stretches in the molecular complexes are less than 30, 60, and 30 cm^{-1} , respectively, while the values in the water sphere are around 60, 188, and 155 cm^{-1} , respectively. Evidently, this is due to the fact that the formamide molecule is polarized by larger number of solvent molecules in the water sphere. The current protocol is a rather efficient procedure to determine solvent induced shifts in vibrational frequencies and intensities. It is an alternative to the traditional method where the vibrational spectrum is obtained by the Fourier transform (FT) of the dipole–dipole correlation function obtained from molecular dynamics calculations.⁵⁰ The current QM/MM normal mode approach is flexible in the sense that a computationally economic method (such as AM1/MM) can be used to efficiently sample configurations and that only limited number of energy minimization and normal mode calculations have to be performed at a more reliable level.

We also examine the CPHF contribution from the MM degrees of freedom. In Fig. 4(b), the results are obtained with no MM CPHF contributions. It is seen that the error in the vibrational frequencies is rather small, while that for the IR intensity is large. Using the uncoupled CPHF solutions improves the situation significantly, as shown in Fig. 4(c). If the CPHF corresponding to the solvent molecules within 4 Å around the oxygen atom in the formamide are solved exactly while the rest, 76 out of 89 water molecules, are treated with the uncoupled solution, the results become very satisfactory. For comparison, the four calculations in Fig. 4 take 10.3, 4.5, 4.7, and 5.5 minutes, respectively, on a Linux Alpha machine.

To summarize, we have demonstrated in this section that the QM/MM method is computationally efficient to predict environmental induced frequency shift. It can also give reasonable IR intensity information, which is a significant advantage over the pure MM methods using standard force field which do not consider dipole derivatives during parametrization. We have also shown that the MM CPHF contribute little to the vibrational frequencies but significantly to the IR intensities. The uncoupled CPHF approximation is usually very satisfactory, and gives much better IR intensities with little computational cost compared to completely neglecting the MM CPHF.

B. Butanol: The effect of QM/MM partition and link atom on normal modes

As the next example, we consider the IR spectrum of butanol, where the partition between the QM and MM regions is across a covalent bond. The purpose of this example is to illustrate the effect of QM/MM partition and link atom on the normal modes.

Optimization and normal mode analyses have been performed for the *trans*-butanol at five different levels: full QM, full MM, and three QM/MM calculations with different QM/MM partitions. The QM level here is taken to be HF/6-31G(d,p), and MM is the CHARMM22 force field.⁴⁴ As illustrated in Fig. 5, the three QM/MM calculations take the

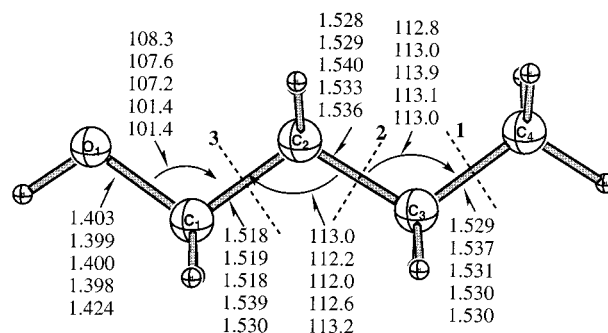


FIG. 5. Optimized structure of butanol at different levels, showing heavy atom bond lengths and angles. The numbers are shown in the following order from top: full HF/6-31G(d,p); HF/6-31G(d,p)/MM with three different QM/MM partitions (as indicated with dotted lines); pure MM with the CHARMM22 force field. Distances are in Angstrom and angles are in degrees.

–C₃H₆(OH), –C₂H₄(OH), and –CH₂(OH) group as the QM part, respectively. As described previously,^{4(b)} a link hydrogen atom is added to the boundary QM carbon to saturate its valence. The position of the link atom is optimized during energy minimization, and it does not interact with any MM atoms explicitly. It should be noted that all the bonded terms involving both QM and MM atoms are computed in the QM/MM calculations.

The optimized structures are presented in Fig. 5, and the predicted IR spectrums at varieties of levels are shown in Fig. 6. Unlike the formamide–water case, the QM and MM regions are directly bonded and are strongly mixed in the normal modes. Therefore, it is not straightforward to select modes for frequency scaling, and unscaled results are shown in Fig. 6.

The structures from different calculations are all quite similar; the largest differences occur for the O–C–C angle, where the full MM and QM/MM with the third partitioning scheme (see Fig. 5) give values $\sim 7^\circ$ smaller than the full QM or QM/MM calculations with larger QM regions. Similar to the case of formamide–water, the full MM frequencies are in good agreement with the scaled full QM results (not shown), with a RMS error of 48 cm^{-1} . The IR intensities, on the other hand, are rather different at the two extreme levels, as illustrated by Figs. 6(a) and 6(e). The overall agreement between the predicted QM/MM and the full QM spectrum is quite encouraging, even for the QM/MM calculation with the smallest QM part. It is interesting to notice that even with one –CH₂(OH) group described at the QM level, the overall spectrum shows rather significant improvement over the pure MM result, especially in terms of the IR intensities. By carefully inspecting the vibrational frequencies, we found that for every QM/MM calculations there is a mode with $\sim 1800 \text{ cm}^{-1}$ which involves mainly the –CH₂ unit at the QM/MM boundary. This value is much higher than that for the corresponding modes predicted by the full QM and pure MM, $\sim 1600 \text{ cm}^{-1}$ and $\sim 1400 \text{ cm}^{-1}$, respectively. Therefore, it is clearly an artifact due to the presence of the link atom. This mode is hardly noticeable in the spectrum shown in Fig. 6 due to its very small IR intensity.

To further compare QM/MM and full QM results and

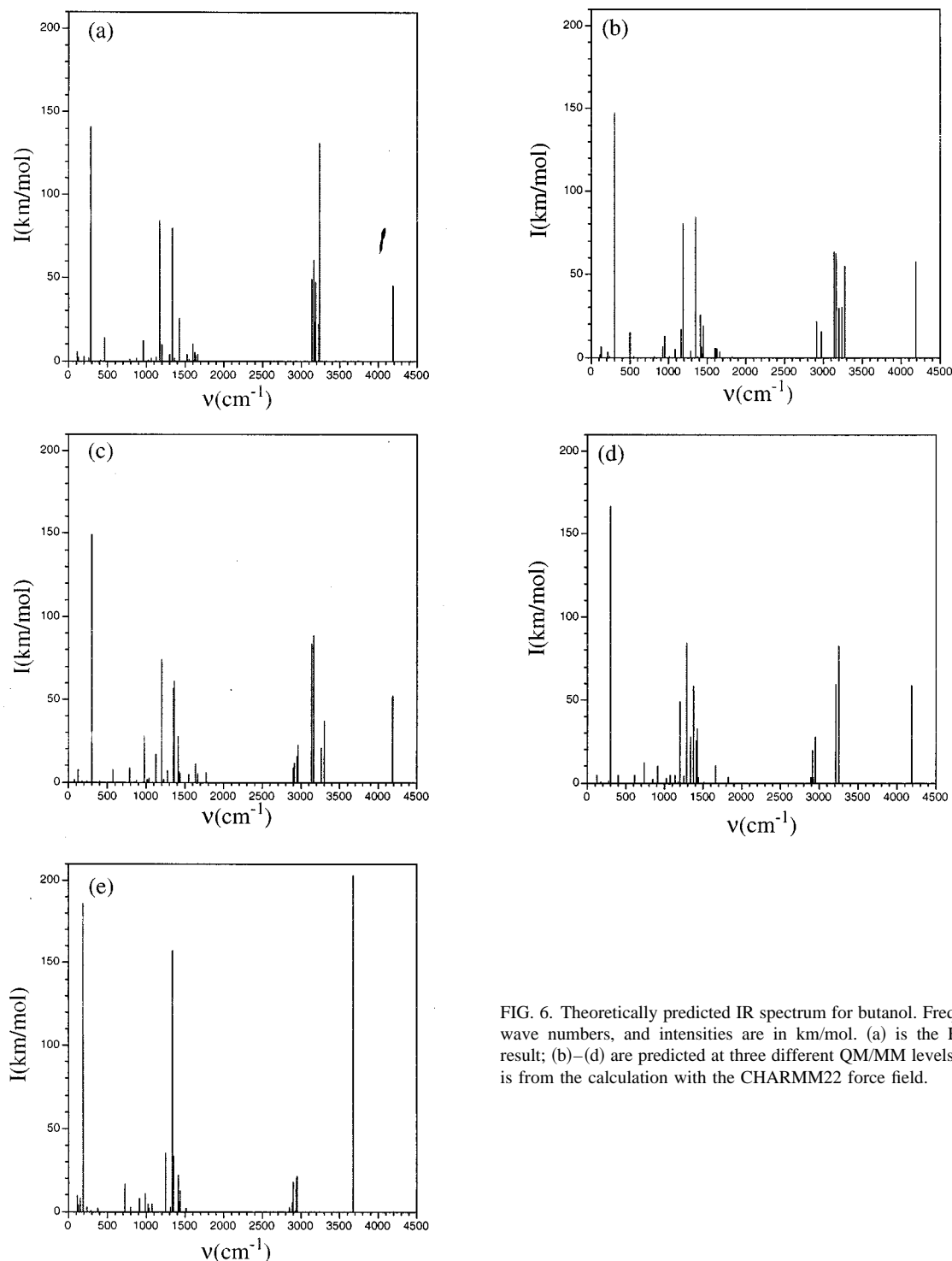


FIG. 6. Theoretically predicted IR spectrum for butanol. Frequencies are in wave numbers, and intensities are in km/mol. (a) is the HF/6-31G(d,p) result; (b)–(d) are predicted at three different QM/MM levels (see text). (e) is from the calculation with the CHARMM22 force field.

illustrate the link atom effect, we have calculated the overlap between the eigenvectors from these two sets of calculations. The largest overlap values for each normal mode at different QM/MM levels are summarized in Figs. 7(a)–7(c). Most of the overlap between the QM/MM and full QM results are larger than 0.75, suggesting that the motion predicted by the QM/MM method is consistent with the full QM calculation. There are several regions where the overlap values are small, ~ 0.5 . This suggests that the QM/MM normal modes correspond to a linear combination of several modes predicted at the QM level. This mixing effect is most significant for the

abnormally high frequency mode $\sim 1800 \text{ cm}^{-1}$ mentioned above. The full QM eigenvector that has the largest overlap with this QM/MM mode involves the motion of more than one (CH_2) unit, as illustrated by the bottom structures in Fig. 7. At the QM/MM level, however, this mode becomes more localized, as shown by the structures in the middle of Fig. 7. This is due to the existence of the link atom, which still influences the Hessian matrix elements of the boundary QM atoms implicitly, despite the fact that its explicit contribution to the Hessian has been projected out according to Eq. (29).

The effect of the second projection step specified by Eq.

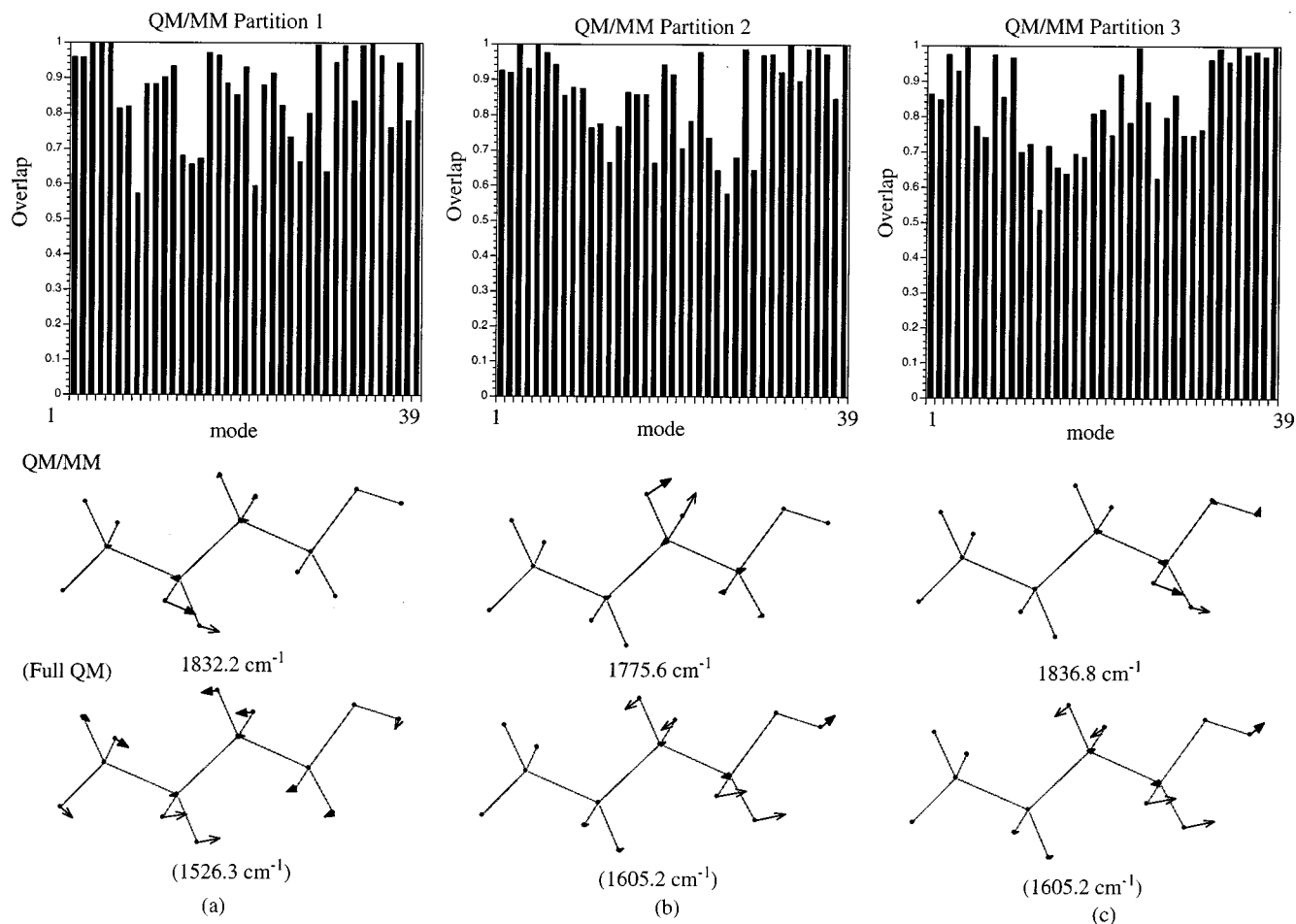


FIG. 7. Largest overlap of QM/MM (see text) normal mode eigenvectors and the full QM results for the three QM/MM partitions for butanol. Results in (a), (b), and (c) are obtained with the QM/MM partition 1, 2, and 3, respectively, indicated in Fig. 5. Also shown are the eigenvectors (atomic displacements are indicated by arrows) for the abnormally high frequency mode due to the presence of link atom in the QM/MM calculation, along with the mode that has the largest overlap from the full QM calculation (bottom structure). The number with and without parentheses are the full QM and QM/MM frequencies, respectively.

(30) is illustrated in Fig. 8 with the first QM/MM partition as the example. Figure 8(a) shows the largest overlaps (in percent) between the QM/MM normal modes without the second projection and the overall translational/rotational eigenvectors of the actual molecule without the link atom. Only the low frequency vibrational modes are “contaminated” to a significant extent. In Fig. 8(b), the relative shifts in the vibrational frequencies (in percent) due to the projection defined by Eq. (30) are plotted. Although only the low frequency modes are affected significantly, the effect is also observable in the high frequency region. The effect of the second projection on the eigenvectors is insignificant, with the overlaps between unprojected and projected results very close to unity for nearly all the modes (not shown).

For butanol, the overall QM/MM predicted frequencies and eigenvectors are in reasonable agreement with the full QM calculations in the presence of the link atom, with most modes having a large overlap with the full QM eigenvectors. Nevertheless, the link atom does produce artifacts in the normal modes, e.g., certain modes become more localized compared to the full QM results. This observation once again highlights the importance of carefully examining the link

atom effect when applying QM/MM methods to large systems.

C. A model transition state structure in triose phosphate isomerase (TIM)

One important area for applying the QM/MM frequency calculation is in modeling transition states of enzymes. First, normal mode calculations are useful during the optimization of transition states with the Quasi-Newton based algorithms.⁵¹ Second, the vibrational frequencies and eigenvectors at the saddle point are essential for the reaction path following algorithms as well as rate constant calculations.

As a simple example, we consider a model transition state structure for the triosephosphate isomerase (TIM) catalyzed reaction.^{5(a)} As shown in Fig. 9, we include the singly protonated imidazole and a model enediolate substrate. The transition state corresponds to the intramolecular proton transfer within the substrate. As found in more extensive calculations,⁵² this is not the rate limiting step of the TIM catalyzed reaction. It is included here as an example to demonstrate that a QM/MM description for the active site is sat-

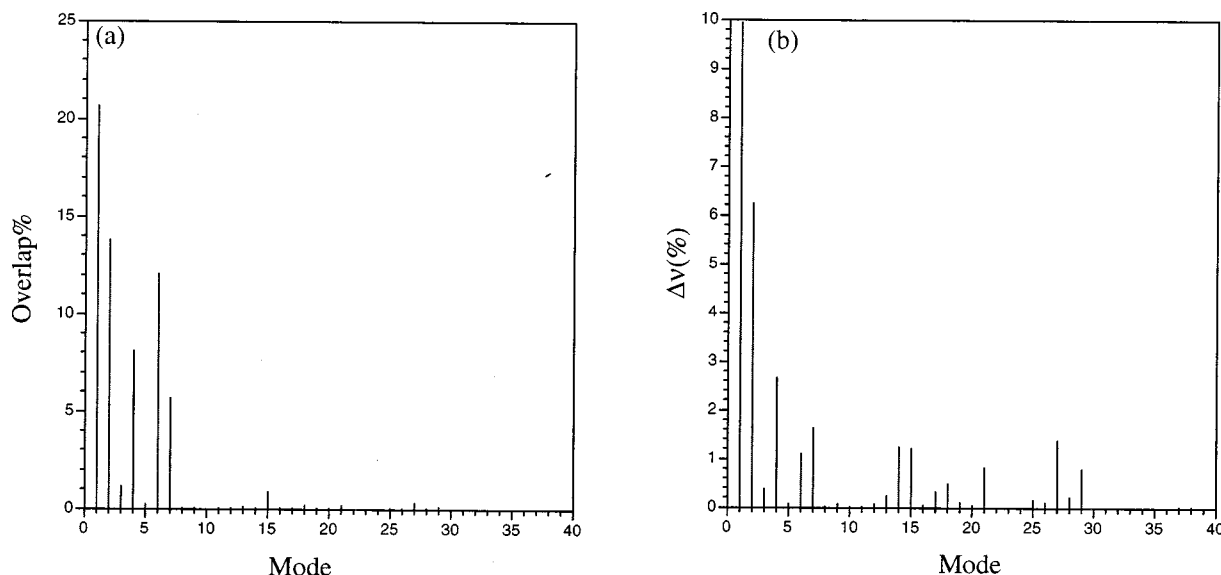


FIG. 8. Effect of the projection defined by Eq. (30) on the normal modes of butanol with the first QM/MM partition. (a) Values of the largest overlaps (in percent) between unprojected QM/MM normal modes and the overall translational/rotational eigenvectors of butanol without the link atom. (b) Relative shifts (in percent) in the vibrational frequencies due to the projection Eq. (30).

isfactory, as compared to full QM. Since full QM calculation has to be carried out for comparison, only one histidine residue is included in the current model, which is described with the CHARMM22 force field in the QM/MM calculation. For comparison, the corresponding transition state of the isolated substrate in the gas phase is also included. The QM level here is B3LYP/6-31+G(d,p).⁵³ The gas phase structure is optimized with GAUSSIAN94,⁵⁴ while the model enzymatic transition state is obtained using the TRAVEL module in CHARMM with the conjugate peak refinement algorithm.⁵⁵ To avoid large distortion of the structure from that in the enzyme, the position of the C₁ in the substrate and C₃, C₄ in the model histidine were fixed during the optimization.

As shown in Fig. 9, the imidazole polarizes the oxygen atom in the substrate such that the transition state comes later, if we imagine the proton being transferred from the left to the right. The imaginary vibrational frequency is lowered by nearly 200 cm⁻¹ compared to that in the gas phase structure. This would significantly reduce the tunneling effect in enzyme compared to the gas phase. At the QM/MM level, the polarization effect is weaker judging from the longer histidine–oxygen distance, although the transition state structure is very similar to the full QM result. The shift in the imaginary frequency is smaller, ~50 cm⁻¹ than that from the full QM calculation. Thus, the QM/MM description of the active site, although approximate, is satisfactory.

D. A simple model for the active site of myoglobin

In the previous examples, all the calculations have been carried out for main group elements. In this section, we shall illustrate the application of DFT/MM method to systems containing transition metal elements that are more difficult to treat with molecular mechanics.¹⁷ Specifically, we present a study of a simplified model for the active site of myoglobin.

For myoglobin and hemoglobin, numerous studies have been concerned with effect of electric field in the heme active site generated by the protein environment on the stretching frequency of CO bound to the heme iron.⁵⁶ To examine if an MM treatment of the environment can predict the frequency shift of the C=O stretch, we have constructed a simplified model for the active site of myoglobin. Three complexes have been calculated, whose structures are shown in Fig. 10. The first complex (**com_i**) is a simple model for isolated CO-heme system, which has been used in several previous studies.⁵⁷ The second complex (**com_h1**) includes an additional imidazole molecule, which mimics the distal histidine in myoglobin. In the third complex (**com_h2**), the orientation of the histidine is altered from that in **com_h1**, so as to orient the NH proton of the imidazole away from the CO. Such different distal histidine orientations have been considered in previous studies of the CO stretching frequency.⁵⁸ The QM level in this set of calculations is B3LYP,⁵³ and the basis set is the all electron set of double-zeta quality from Ahlrichs and co-workers.⁵⁹ Due to the size of the systems, no polarization functions have been added in the geometry optimization or normal mode analysis. In the QM/MM calculations, the heme–CO complex is the QM part, and the imidazoles in **com_h1** and **com_h2** are treated with the CHARMM22 force field. To avoid large changes in the relative orientation of the imidazole and model heme in **com_h2**, the position of the Fe, and that of the CE1, HE1 atoms of imidazole are fixed during geometry optimization. The optimized structures are shown in Fig. 10, and selected normal modes are shown in Fig. 11.

We begin by examining the trends from the pure DFT calculations. First, it is noted that the optimized Fe–CO distance of 1.76 Å in **com_i** is in good agreement with experimental measurement of a model heme complex (1.77 Å),⁶⁰ as well as with theoretical calculations (1.720 Å) with more

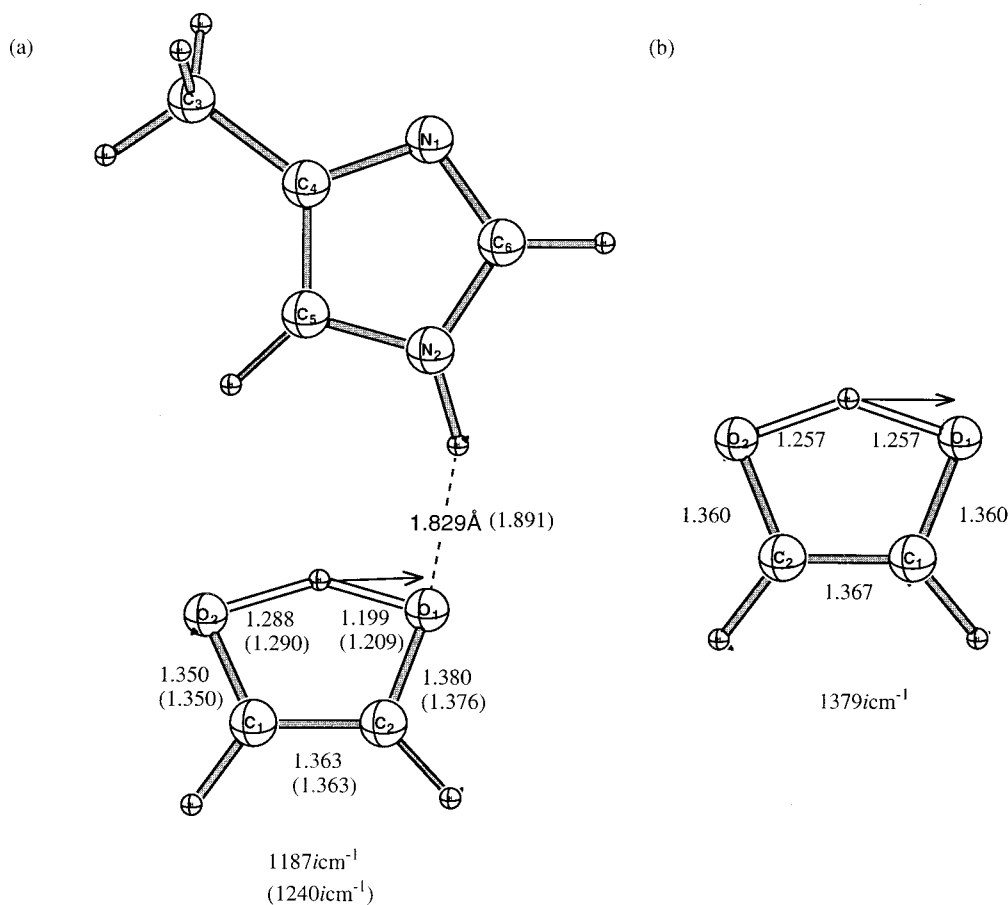


FIG. 9. (a) A model transition state structure in the triosephosphate isomerase (TIM) catalyzed reactions. The number without parentheses are full B3LYP/6-31+G(d,p) results, and these in the parentheses are QM/MM results with the CHARMM22 model histidine. (b) The corresponding structure in the gas phase. The arrows indicates the eigenvector correspond to the imaginary mode. The distances are given in Angstroms (see text).

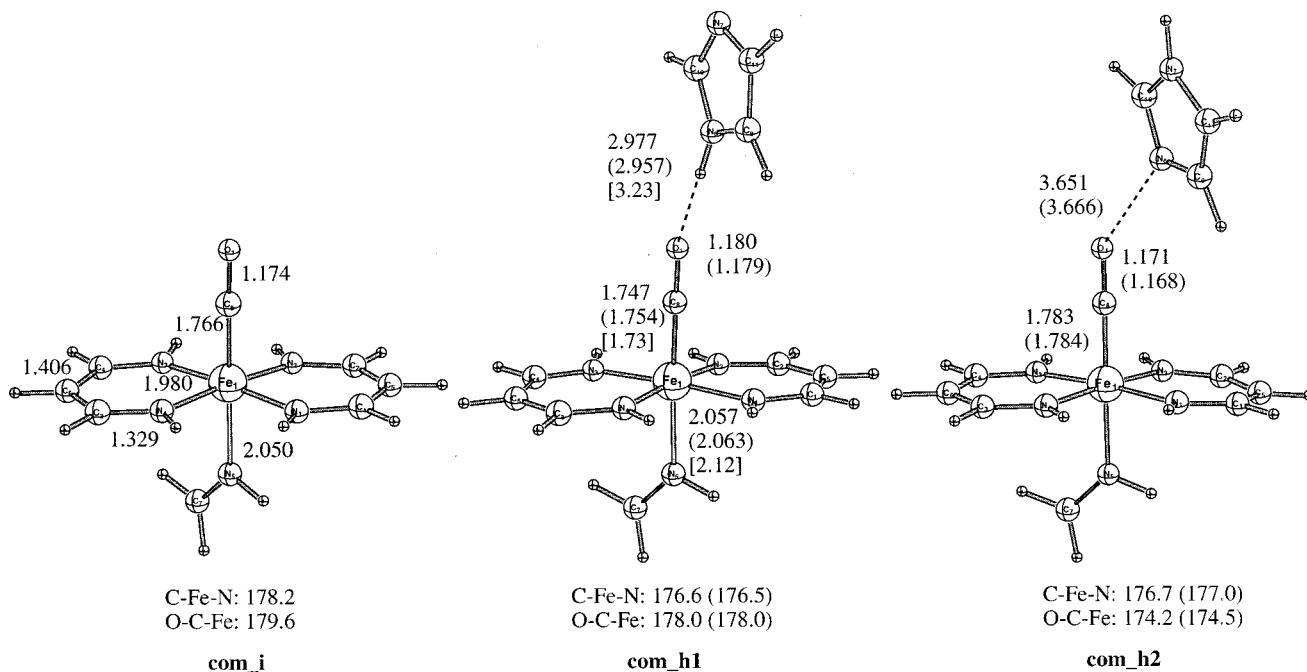


FIG. 10. The optimized structures for a simple model for the active site of myoglobin. Numbers without parentheses are obtained with full QM calculations (see text for detail). Numbers in parentheses are obtained at the QM/MM level with the model histidine treated with CHARMM22 force field. Numbers in brackets are from recent x-ray study in Ref. 62. Distances are given in Angstroms.

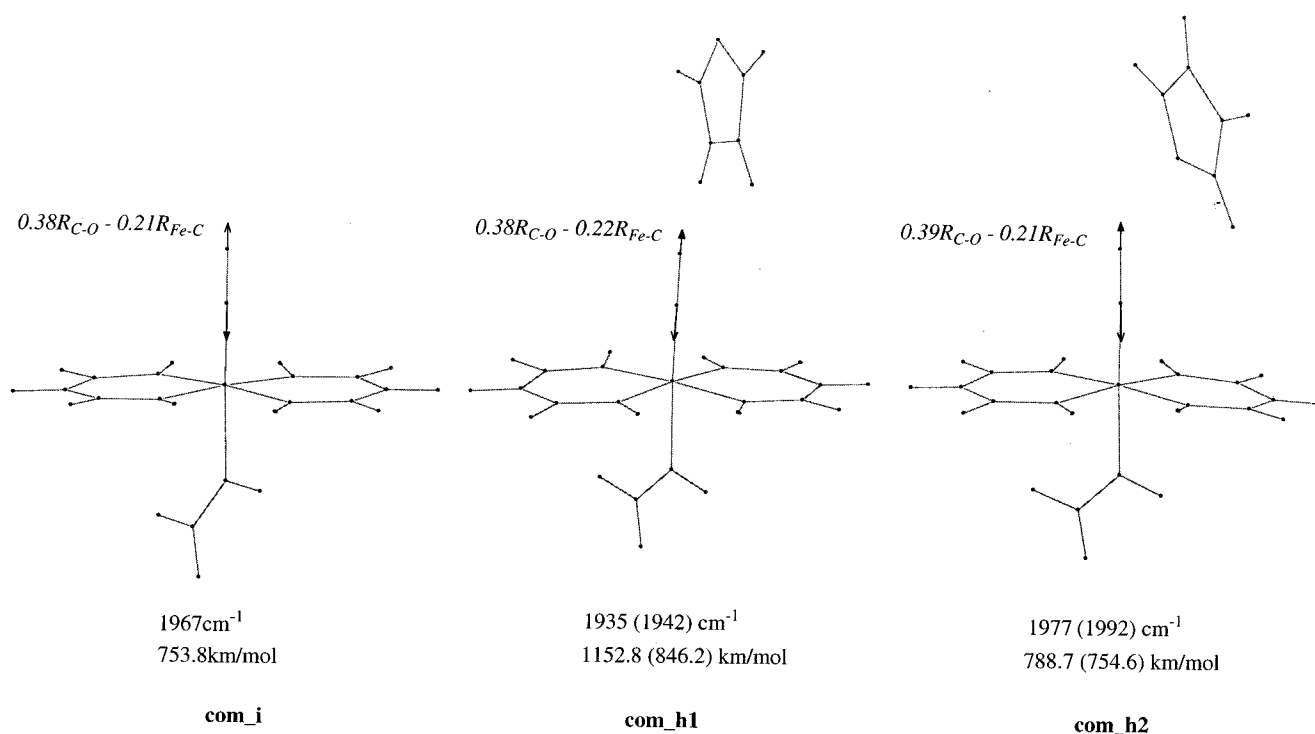


FIG. 11. The normal modes correspond to the C–O stretch for the active site model of myoglobin. Numbers without parentheses are obtained with full QM calculations. Numbers in parentheses are obtained at the QM/MM level (see text and Fig. 10 caption). Frequencies are in wave numbers, and intensities are in km/mol. Also shown in italics are the displacement vectors in terms of Fe–C and C–O stretches correspond to the normal modes.

sophisticated models of the heme.⁶¹ The FeCO moiety is linear and perpendicular to the “heme” with a CO bond length of 1.174 Å. Upon adding the distal group in **com_h1**, the Fe–CO distance gets shorter while C–O becomes more stretched by 0.006 Å due to the polarization from the distal group. The Fe–CO distance of 1.747 Å is in good agreement with the latest x-ray study⁶² as well as with previous theoretical calculations.⁶¹ Interestingly, the effect of the distal histidine in **com_h2** is to slightly increase the Fe–C distance and shortens the C–O bond compared to the case of **com_i**, indicates that the interaction between bound CO and the histidine in the orientation of **com_h2** is unfavorable. In parallel to the structural changes, the vibrational frequencies involving CO vary rather significantly in the three complexes. In Fig. 11, we show that the normal mode corresponds to the CO stretch. The vibrational frequency drops 32 cm⁻¹ for **com_h1**, and increases by 10 cm⁻¹ for **com_h2**. In other words, we predict that the CO stretching frequency in **com_h1** is lower than that in **com_h2** by ~40 cm⁻¹. It is interesting to note that this trend is in disagreement to the calculations of Oldfield *et al.* based on electric field consideration.⁵⁸ From the electric field generated by the dipole moment of histidine [see Fig. 1 in Ref. 58(b)], they propose that the CO stretching frequency in a conformational substate similar to **com_h1** would be higher than that in conformational substate of type **com_h2**. The discrepancy between the current study and their model might be due to the geometrical limitation of our model, and will be studied further with calculations in the full protein environment.⁶³ However, it appears reasonable to observe a lower CO stretching frequency in **com_h1** compared to **com_h2**, since

the CO is polarized via a weak hydrogen bond by the imidazole in **com_h1**. Also included in Fig. 11 are the normal mode displacements in terms of internal coordinates, obtained from the Cartesian displacements with the Wilson B matrix.³⁹ Clearly, the so-called CO stretch mode involves significant amount, ~25%, of Fe–C stretch. In other words, the Fe–C σ bond character is also altered in this mode, in addition to the Fe–CO π backbonding framework as commented on in the literature.⁵⁶ Furthermore, it is seen that the CO stretch mode involves little heme motion, and is mostly perpendicular to the heme plane. This observation has implications for the interpretation of recent photoselective spectroscopic measurements using CO stretch as the probe.⁶⁴ Calculations with a larger heme model and protein environment are in progress.⁶³

The QM/MM geometries are very close to the full QM results as shown in Fig. 10. The trend in the vibrational frequency shift is well reproduced at the QM/M level. The error compared to the full QM result is 8 cm⁻¹ and 15 cm⁻¹ for **com_h1** and **com_h2**, respectively. As to the IR intensity, the results are qualitatively correct in terms of the sign for the effect of the distal group, but quantitatively too small in the magnitude.

The last two examples show that the polarization effect from the residues in the active site of proteins can be captured to a reasonable degree with the QM/MM treatment. This observation further justifies the use of QM/MM methods in probing the influence of the protein environment on both the electronic and dynamic properties of its active site.

IV. CONCLUSION

In *ab initio* quantum chemistry, analytical derivative theories have made possible the calculations of many important molecular properties. Despite improvements of numerical techniques and computation facilities, such calculations are still prohibitive for large systems, including enzymes. In the present work, we begin the development of analytical derivative theories in the combined QM/MM framework, which should be efficient in predicting properties of a specific group in a large system. In particular, the method for analytical Hessian calculations in the combined QM/MM framework has been implemented in the GAMESS/CHARMM and CADPAC/CHARMM programs to allow one to evaluate frequencies and infrared intensities of specific groups in biological systems. Several examples illustrate that MM description of the environment can capture the polarization effect on the QM region to a satisfactory extent, as judged from the vibrational frequencies and IR intensities compared with full QM calculations. The problem created by a link atom is examined and it is shown that it can give rise to artifacts in the modes involving the boundary atoms. This suggests that one has to carefully examine the effect of link atom when applying the QM/MM method. An alternative is to use methods that do not introduce extra atoms when the QM/MM boundary is within a molecule.⁶⁵

In the near future, we envision the extension of analytical derivative theories within the QM/MM framework into other areas. We have implemented a QM/MM method for predicting NMR chemical shift, which is expected to be a useful tool for deriving structural information and interpreting data for biological systems. Other interesting properties include Raman intensities, vibrational and electronic circular dichroism, and spin-orbit coupling. By implementing these quantities in the QM/MM methodology, it will become possible to extend the study of "weak interactions" to very large systems.

ACKNOWLEDGMENTS

We thank H. Guo for helpful discussions on myoglobin. The work was supported in part by a grant from the Department of Energy. Part of the calculations were performed on the CRAY J90 at NERSC, and the IBM SP2 at CCST of the Argonne National Laboratory.

¹(a) L. A. Curtiss, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **103**, 4192 (1995); (b) G. A. Petersson *et al.*, *ibid.* **109**, 10570 (1998).

²See, for example (a) C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, *Chem. Phys. Lett.* **230**, 8 (1994); (b) M. C. Strain, G. E. Scuseria, and M. J. Frisch, *Science* **271**, 51 (1996); (c) W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991).

³(a) D. M. York, T. S. Lee, and W. T. Yang, *Phys. Rev. Lett.* **80**, 5011 (1998); (b) T. S. Lee, D. M. York, and W. T. Yang, *J. Chem. Phys.* **105**, 2744 (1996); (c) A. D. Daniels, J. M. Millam, and G. E. Scuseria, *ibid.* **107**, 425 (1997).

⁴(a) A. Warshel and M. Karplus, *J. Am. Chem. Soc.* **94**, 5612 (1972); (b) M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990); (c) F. Maseras and K. Morokuma, **16**, 1170 (1995); (d) T. Matsubara, F. Maseras, N. Koga, and K. Morokuma, *J. Phys. Chem.* **100**, 2573 (1996); (e) J. Gao, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1996), Vol. 7, p. 119; (f) U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1996); (g) T. N. Truong, T.-T. T. Truong, and E. V. Stefanovich, *J. Chem. Phys.*

107, 1881 (1997); (h) D. Bakowies and W. Thiel, *J. Phys. Chem.* **100**, 10580 (1996).

⁵See, for example (a) P. A. Bash *et al.*, *Biochemistry* **30**, 5826 (1991); (b) P. A. Bash, M. J. Field, and M. Karplus, *J. Am. Chem. Soc.* **109**, 8092 (1987); (c) J. Gao, *Acc. Chem. Res.* **29**, 298 (1996); (d) M. A. Thompson and G. K. Schenter, *J. Phys. Chem.* **99**, 6374 (1995); (e) R. D. Froese, S. Humbel, M. Svensson, and K. Morokuma, *J. Phys. Chem. A* **101**, 227 (1997); (f) S. Humbel, S. Sieber, and K. Morokuma, *J. Chem. Phys.* **105**, 1959 (1996).

⁶(a) P. A. Bash *et al.*, *Proc. Natl. Acad. Sci. USA* **93**, 3698 (1996); (b) J. Li *et al.*, *Inorg. Chem.* **35**, 4694 (1996).

⁷J. L. Gao, *J. Am. Chem. Soc.* **116**, 9324 (1994).

⁸D. M. York, T. S. Lee, and W. T. Yang, *J. Am. Chem. Soc.* **118**, 10940 (1996).

⁹M. Karplus, *Rev. Mod. Phys.* **32**, 455 (1960).

¹⁰See, for example, (a) R. M. Stratt, *Acc. Chem. Res.* **28**, 201 (1995); (b) F. H. Stillinger and T. A. Weber, *J. Phys. Chem.* **87**, 2833 (1983); (c) F. H. Stillinger and T. A. Weber, *Science* **225**, 983 (1984); (d) S. Sastry, P. G. Debenedetti, and F. H. Stillinger, *Nature (London)* **393**, 554 (1998); (e) I. Ohmine and H. Tanaka, *Chem. Rev.* **93**, 2545 (1993).

¹¹See, for example, (a) D. A. Case, *Curr. Opin. Struct. Biol.* **4**, 285 (1994); (b) K. Hensen, *Proteins: Struct., Funct., Genet.* **33**, 417 (1998) (c) J. Ma and M. Karplus, *Proc. Natl. Acad. Sci. USA* **95**, 8502 (1998).

¹²See, for example (a) D. G. Truhlar, A. D. Isaacson, and B. C. Garrett, in *The Theory of Chemical Reaction Dynamics*, edited by M. Baer (CRC, Boca Raton, FL, 1985), Vol. 4, pp. 65(b); D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, *J. Phys. Chem.* **100**, 12771 (1996).

¹³W. H. Miller, N. C. Handy, and J. E. Adams, *J. Chem. Phys.* **72**, 99 (1980).

¹⁴J. Wiorkiewicz-Kuczera and M. Karplus, *J. Am. Chem. Soc.* **112**, 5324 (1990).

¹⁵(a) P. M. Kozlowski, A. A. Jarzecki, and P. Pulay, *J. Phys. Chem.* **100**, 7007 (1996); (b) P. M. Kozlowski, A. A. Jarzecki, and P. Pulay, *X. Y. Li, ibid.* **100**, 13985 (1996).

¹⁶(a) B. R. Brooks, D. Janežič, and M. Karplus, *J. Comput. Chem.* **16**, 1522 (1996); (b) L. Mouawad and D. Perahia, *Biopolymer* **33**, 599 (1993); (c) H. W. T. van Vlijmen, Ph.D. Thesis, Harvard University, 1997.

¹⁷A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, *J. Am. Chem. Soc.* **114**, 10024 (1992).

¹⁸(a) H. Hellmann, *Einführung in die Quantenchemie* (Franz Deuticke, Leipzig, 1937) (b) R. P. Feynman, *Phys. Rev.* **41**, 721 (1939); (c) A. C. Hurley, *Proc. R. Soc. London, Ser. A* **226**, 170 (1954).

¹⁹(a) J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkely, *Int. J. Quantum Chem.* **S13**, 255 (1979); (b) Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry, Analytical Derivative Methods in ab initio Molecular Electronic Structure Theory* (Oxford University Press, Oxford, UK, 1994).

²⁰(a) K. Kitaura, S. Obara, K. Morokuma, *Chem. Phys. Lett.* **77**, 452 (1981); (b) Q. Cui, D. G. Musaev, M. Svensson, and K. Morokuma, *J. Phys. Chem.* **100**, 10936 (1996).

²¹N. C. Handy, D. J. Tozer, G. J. Laming, C. W. Murray, and R. D. Amos, *Isr. J. Chem.* **33**, 331 (1993).

²²See, for example, A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory* (Dover, New York, 1996).

²³See, for example (a) R. J. Harrison, G. B. Fitzgerald, W. D. Laidig, and R. J. Bartlett, *Chem. Phys. Lett.* **124**, 291 (1986); (b) E. A. Salter, G. W. Trucks, and R. J. Bartlett, *J. Chem. Phys.* **90**, 1767 (1989).

²⁴See, for example, B. Brooks, D. Janežič, and M. Karplus, *J. Comput. Chem.* **16**, 1522 (1995).

²⁵(a) A. Komornicki and R. L. Jaffe, *J. Chem. Phys.* **71**, 2150 (1979); (b) Y. Yamaguchi *et al.*, *J. Chem. Phys.* **84**, 2262 (1986).

²⁶D. P. Craig and T. Thirunamachandran, *Molecular Quantum Electrodynamics* (Academic, New York, 1984).

²⁷R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).

²⁸(a) R. T. Sanderson, *Science* **114**, 670 (1951); (b) W. J. Mortier, K. van Genechten, and J. Gasteiger, *J. Am. Chem. Soc.* **107**, 829 (1985); (c) W. J. Mortier, S. Ghosh, and S. Shankar, *J. Chem. Phys.* **86**, 5063 (1987); (d) M. Berkowitz and R. G. Parr *ibid.* **88**, 2554 (1988).

²⁹See, for example (a) B. G. Baekelandt *et al.*, *J. Phys. Chem.* **99**, 9784 (1995); (b) B. G. Baekelandt, W. J. Mortier, J. L. Lievens, and R. A. Schoonheydt, *J. Am. Chem. Soc.* **113**, 6730 (1991); (c) T. K. Ghanty, and S. K. Ghosh, *J. Phys. Chem.* **95**, 6512 (1991); **98**, 1840 (1994).

- ³⁰(a) A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.* **95**, 3358 (1991); (b) S. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.* **101**, 6141 (1994); (c) D. M. York and W. Yang, *ibid.* **104**, 159 (1996).
- ³¹(a) A. Morita and S. Kato, *J. Chem. Phys.* **108**, 6809 (1998); (b) **109**, 5511 (1998).
- ³²Q. Cui and M. Karplus (work in progress).
- ³³B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
- ³⁴P. Lyne, M. Hodoscek, and M. Karplus, *J. Phys. Chem.* **103**, 3462 (1999).
- ³⁵GAMESS, M. W. Schmidt, K. K. Baldrige, 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. S. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).
- ³⁶CADPAC5.0, R. D. Amos, University of Cambridge.
- ³⁷See, for example, K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, *Int. J. Quantum Chem.* **60**, 1189 (1996).
- ³⁸N. Reuther, A. Dejaegere, B. Maigret, and M. Karplus (unpublished).
- ³⁹E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
- ⁴⁰(a) J. G. Belasco and J. R. Knowles, *Biochemistry* **19**, 472 (1980); (b) E. A. Komives *et al.*, *ibid.* **30**, 3011 (1991); (c) M. Karplus *et al.*, *Faraday Discuss.* **93**, 239 (1992).
- ⁴¹(a) R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.* **54**, 724 (1971); (b) W. J. Hehre, R. Ditchfield, and J. A. Pople, *ibid.* **56**, 2257 (1972); (c) P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973).
- ⁴²W. L. Jorgensen, J. Chandrasekhar, and J. P. Madura, *J. Chem. Phys.* **79**, 926 (1983).
- ⁴³E. Neria, S. Fisher, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
- ⁴⁴A. MacKerell *et al.* *J. Phys. Chem. B* **102**, 3586 (1998).
- ⁴⁵R. Yelle (private communication).
- ⁴⁶P. N. Day *et al.*, *J. Chem. Phys.* **105**, 1968 (1996).
- ⁴⁷J. A. Pople, A. P. Scott, M. W. Wong, and L. Radom, *Isr. J. Chem.* **33**, 345 (1993).
- ⁴⁸K. Palmo and S. Krimm, *J. Comput. Chem.* **19**, 754 (1998).
- ⁴⁹S. Dapprich *et al.*, *J. Mol. Struct.: THEOCHEM* **461**, 1 (1999).
- ⁵⁰See, for example, M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
- ⁵¹See, for example, H. B. Schlegel, *Adv. Chem. Phys.* **67**, 249 (1987).
- ⁵²Q. Cui and M. Karplus (unpublished).
- ⁵³(a) A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988); (b) C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988); (c) A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ⁵⁴GAUSSIAN94 revision D.3, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
- ⁵⁵S. Fisher and M. Karplus, *Chem. Phys. Lett.* **194**, 252 (1992).
- ⁵⁶See, for example (a) J. R. Hill *et al.*, *J. Phys. Chem.* **100**, 18023 (1996); (b) S. Hu, K. M. Vogel, and T. G. Spiro, *J. Am. Chem. Soc.* **116**, 11187 (1994).
- ⁵⁷See, for example (a) M. T. McMahon *et al.*, *J. Am. Chem. Soc.* **120**, 4784 (1998); (b) A. Strich and A. Veillard, *Theor. Chim. Acta* **60**, 379 (1981); (c) P. Jewsbury *et al.*, *J. Am. Chem. Soc.* **116**, 11586 (1994).
- ⁵⁸(a) J. D. Augspurger, C. E. Dykstra, and E. Oldfield, *J. Am. Chem. Soc.* **113**, 2446 (1991); (b) E. Oldfield, K. Guo, J. D. Augspurger, and C. E. Dykstra *ibid.* **113**, 7537 (1991).
- ⁵⁹A. Schafer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).
- ⁶⁰S. M. Peng and J. A. Ibers, *J. Am. Chem. Soc.* **98**, 8032 (1976).
- ⁶¹(a) C. Rovira *et al.*, *J. Phys. Chem.* **101**, 8914 (1997); (b) E. Sigfridsson, U. Ryde, *J. Inorg. Biol. Chem.* **4**, 99 (1999).
- ⁶²G. S. Kachalova, A. N. Popov, and J. D. Bartunik, *Science* **284**, 473 (1999).
- ⁶³Q. Cui, H. Guo and M. Karplus (unpublished).
- ⁶⁴M. Lim, T. A. Jackson, and P. A. Anfinrud, *Science* **269**, 962 (1995).
- ⁶⁵(a) J. Gao, P. Amara, C. Alhambra, and M. J. Field, *J. Phys. Chem.* **102**, 4714 (1998); (b) Y. K. Zhang, T. S. Lee, and W. T. Yang, *J. Chem. Phys.* **110**, 46 (1999); (c) V. Thery *et al.*, *J. Comput. Chem.* **15**, 269 (1994).