2578 *J. Phys. Chem. A* **2003,** *107,* 2578−2588

Coupled Cluster/Molecular Mechanics Method: Implementation and Application to Liquid Water

Jacob Kongsted,* Anders Osted,† and Kurt V. Mikkelsen‡

Department of Chemistry, H. C. Ørsted Institute, University of Copenhagen,
DK-2100 Copenhagen Ø, Denmark

Ove Christiansen§

Department of Chemistry, University of Århus, Langelandsgade 140, DK-8000 Århus C, Denmark

Received: September 30, 2002; In Final Form: December 9, 2002

In this paper, we present theory and implementation of the first method which combines coupled cluster and molecular mechanics (CC/MM) theory. By introducing the interactions between the solute (QM) and the solvent (MM) molecules into a coupled cluster Lagrangian, we calculate interaction and solvation energies of condensed phase molecules. Also, we derive a hierarchy in the coupling between the two subsystems and thereby study the importance of including the solvent polarization directly into the optimization conditions of the solute wave function for calculation of energies. The method is tested on the water dimer. Furthermore, calculations of the interaction and solvation energies of liquid water are presented.

I. Introduction

In recent years, the use of combined quantum mechanics and molecular mechanics force fields (QM/MM)[1−6] has been shown to be very successful for the description of condensed phase problems in chemistry. The success of this method is due to the fact that the QM/MM method allows for a detailed description of the interesting part of the system (QM), whereas the solvent or surrounding medium (MM) may be described using a much coarser method. In this way we may describe large systems in an effective manner. This is important because of the increasing interest in calculation of properties in biological systems. Another reason for the importance of such methods is the fact that most chemistry takes place in condensed phases and the QM/MM method matches conceptually our way of thinking of the system as a solute embedded in a solvent.

In the QM/MM approach, the interactions between the two subsystems are usually modeled using classical expressions for the interactions between charges and induced charges or induced dipole moments. Furthermore, to account for dispersion and short-range exchange-repulsion effects, a van der Waals term is included in the interaction operator. The interactions between the two subsystems may be treated using the commonly used mean-field approach, but direct-field interaction approaches have also been considered.[7−10]

Many of the presented QM/MM models combine semiempirical electronic structure methods and molecular mechanics[4,5,11,12] but the use of ab initio correlated and uncorrelated descriptions of the QM system has also been presented[13−26] together with density functional theory (DFT).[27,28] Also, quantum mechanics have been combined with interatomic potential functions.[29−31] Other hybrid methods include the use of ab initio quantum chemical calculations combined with the extended reference interaction site model (RISM) for the liquid state[32,33] and the effective potential method by Day et al.[34]

In this paper, we present theory and implementation of the first QM/MM method which combines coupled cluster (CC) and molecular mechanics, the CC/MM model.

The coupled cluster method has previously been shown to be among the most successful electronic structure methods for molecules in a vacuum.[35,36] For example, very accurate calculations of molecular equilibrium structures have been reported using the coupled cluster method.[37] Furthermore, the CC method allows for a systematic increase in accuracy using a hierarchy of CC methods, CCS, CC2, CCSD, CC3,[38,39] etc. (or SCF, MP2, CCSD, CCSD(T) if only ground-state static properties are studied). Thus, the combination of coupled cluster theory and molecular mechanics allows us to extend this hierarchy of CC methods from vacuum towards the condensed phase.

Recently, theory and implementation of the CC method coupled to a dielectric continuum has been reported.[40,41] However, in this model, the short-range interactions are poorly described and the discrete nature of the surrounding medium is completely neglected. The incorporation of short-range effects may be obtained using a semi-continuum model,[42−45] in which some of the surrounding molecules are included in the quantum system. However, the computational complexity is greatly increased compared to the continuum model.

By using the CC/MM model we may, at relatively low additional computational cost, introduce an improved molecular coupling between the two subsystems and thereby avoid the great computational expense associated with a semi-continuum calculation. Thereby, we also avoid the difficulty of splitting the total wave function into a solute and a solvent part in order to calculate nonlocalized properties such as the polarizability and hyperpolarizability tensors.[46]

The introduction of polarization effects in QM/MM theory has previously been shown to be of great importance.[11,12,17,47−49] To perform a systematic analysis of the importance of including polarization effects directly into the optimization of the wave

---

* To whom correspondence should be addressed. E-mail: kongsted@theory.ki.ku.dk.

† E-mail: osted@theory.ki.ku.dk.

‡ E-mail: kmi@theory.ki.ku.dk.

§ E-mail: ove@chem.au.dk.

10.1021/jp022171r CCC: $25.00

function, we introduce a hierarchy of four QM/MM models. In each model, we describe the QM system using a correlated CC wave function but the coupling to the MM environment is treated at different levels of theory. To extract the importance of electron correlation in the QM system, we also perform HF/MM[17] calculations of the interaction and solvation energies. Therefore, we can combine the different hierarchies; a hierarchy of CC models for the correlated description of the QM system, a hierarchy for the one-electron basis set, and another hierarchy concerning the coupling between the two subsystems. We note that a hierarchy concerning the coupling between the two subsystems has been proposed before but at the semiempirical level of theory.[50] Furthermore, the hierarchy described in this work differs from the one in ref 50 because we are more interested in the description of the perturbation of the wave function due to a polarizable MM region.

The general aspects of the theory for CC/MM mean-field and direct-field methods in relation to derivation of energies and response functions has been discussed in ref 10. The focus of the present manuscript is on the implementation and significance of a Lagrangian approach when developing a coupled cluster molecular mechanics method. We report theory and the first implementation and test calculations of ground-state energies for liquid water at the CCSD level of theory. Knowing the present performance of coupled cluster methods, it is clear that the numerical results obtained from a coupled cluster molecular mechanics method are benchmark results at the present stage.

In section II, we present theory for the QM/MM model, and the equations for the calculation of the polarization and interaction energies are considered. Furthermore, we introduce coupled cluster theory, and the optimization conditions for the CC/MM wave function are derived. In section III, we outline the computational details concerning the implementation of the CCSD/MM model. Section IV contains results for a study of the water dimer, and in section V, we present calculations of liquid water. Finally, section VI contains a summary and a discussion of future work using the CCSD/MM model.

II. Theory

This section presents the theoretical background for the CC/MM method and is divided into three parts. In section A, we present the QM/MM model including polarization effects, in B the theory of the coupled cluster wave function in a vacuum, and in C the theory of the CC/MM method.

A. QM/MM with Polarization Effects. In the QM/MM approach, the total Hamiltonian is decomposed into three contributions[2−4,11]

$$\hat{H} = \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}} + \hat{H}_{\mathrm{MM}} \tag{1}$$

where $\hat{H}_{\mathrm{QM}}$ is the usual many-body vacuum Hamiltonian, $\hat{H}_{\mathrm{QM/MM}}$ represents the interaction Hamiltonian, and $\hat{H}_{\mathrm{MM}}$ describes the classically treated part of the total system, which is represented by moleculr mechanics (MM). In this paper, we are mainly interested in the representation and calculation of the interaction term. Employing a mean-field description[10] of this interaction, we may decompose the interaction Hamiltonian into the following contributions:

$$\hat{H}_{\mathrm{QM/MM}} = \hat{H}^{\mathrm{el}} + \hat{H}^{\mathrm{vdw}} + \hat{H}^{\mathrm{pol}} \tag{2}$$

The first term, $\hat{H}^{\mathrm{el}}$, represents the electrostatic interactions between the electrons and the nuclei in the QM system and the partial charges in the MM system. This term is described using quantum mechanics. Thus, $\hat{H}^{\mathrm{el}}$ is modeled as a simple Coulomb interaction

$$\hat{H}^{\mathrm{el}} = -\sum_{s=1}^{S} \hat{N}_s + E^{\mathrm{el,nuc}}_{\mathrm{S,N}} \tag{3}$$

where the electronic contribution is written as

$$\hat{N}_s = \sum_{pq} \left\langle \phi_p \left| \frac{q_s}{|\vec{R}_s - \vec{r}_{\mathrm{i}}|} \right| \phi_q \right\rangle \hat{E}_{pq} \equiv \sum_{pq} n^s_{pq} \hat{E}_{pq} \tag{4}$$

and the energy due to the interaction of the MM partial charges and the QM nuclei is written as

$$E^{\mathrm{el,nuc}}_{\mathrm{S,N}} = \sum_{s=1}^{S} \sum_{n=1}^{N} \frac{q_s Z_n}{|\vec{R}_s - \vec{R}_n|} \tag{5}$$

The index s runs over all of the sites in the MM system which is usually all of the MM atoms and q_s is the charge at site s. The vector $\vec{R}_n$ is the position vector for the nth (QM) nucleus, and Z_n is the corresponding nuclear charge. The position vector for site s is labeled $\vec{R}_s$, and the position vectors for the (QM) electrons are denoted $\vec{r}_i$. The set $\{|\phi_q\rangle\}$ represents the molecular orbitals. The quantity $\hat{E}_{pq}$ is a one-electron excitation operator written in terms of creation ($\hat{a}^{\dagger}_{p\sigma}$) and annihilation ($\hat{a}_{p\sigma}$) operators

$$\hat{E}_{pq} = \sum_{\sigma} \hat{a}^{\dagger}_{p\sigma} \hat{a}_{q\sigma} \tag{6}$$

where σ refers to the projected spin.

To model dispersion and short-range exchange repulsion effects, a van der Waals term, $\hat{H}^{\mathrm{vdw}}$, is included in the interaction operator. Here, we choose to make use of a 6−12 type Lennard-Jones potential

$$\hat{H}^{\mathrm{vdw}} = \sum_{a=1}^{A} \sum_{m:\mathrm{center}} \left[\frac{A_{ma}}{|\vec{R}_m - \vec{R}_a|^{12}} - \frac{B_{ma}}{|\vec{R}_m - \vec{R}_a|^{6}} \right] \tag{7}$$

In eq 7 the index a (m) refers to the center of mass of each MM (QM) molecule. These sums may easily be extended to summations over sites in the MM and QM molecules. However, usually when describing solvation processes the quantum system only consists of one QM molecule and in this case the sum over m drops out of eq 7.

Finally, to model the MM polarization of the QM system and vice versa, we use a semiclassical description of polarization where a polarization term, $\hat{H}^{\mathrm{pol}}$, is included in the interaction Hamiltonian

$$\hat{H}^{\mathrm{pol}} = -\frac{1}{2} \sum_{a=1}^{A} \boldsymbol{\mu}^{\mathrm{ind}}_a \cdot (\hat{\mathbf{R}}\mathbf{r}_a + \mathbf{E}^n(\vec{R}_a)) \tag{8}$$

In eq 8 the operator $\hat{\mathbf{R}}\mathbf{r}_a$ is the QM electronic electric field operator

$$\hat{\mathbf{R}}\mathbf{r}_a = \sum_{pq} \left\langle \phi_p \left| \frac{\vec{r}_i - \vec{R}_a}{|\vec{r}_i - \vec{R}_a|^3} \right| \phi_q \right\rangle \hat{E}_{pq} \equiv \sum_{pq} \mathbf{t}^a_{pq} \hat{E}_{pq} \tag{9}$$

and $\mathbf{E}^{\mathrm{n}}(\vec{R}_{\mathrm{a}})$ is the electric field, due to the QM nuclei, at the center of mass a of each MM molecule

$$\mathbf{E}^{n}(\vec{R}_a) = \sum_{n=1}^{N} \frac{Z_n(\vec{R}_a - \vec{R}_n)}{|\vec{R}_a - \vec{R}_n|^3} \tag{10}$$

The vector, $\boldsymbol{\mu}_a^{\mathrm{ind}}$, is the induced dipole moment at the center of mass a of each MM molecule. In a linear approximation, the induced dipole moment is related to the total electric field, $\boldsymbol{E}_a^{\mathrm{total}}$

$$\boldsymbol{\mu}_a^{\mathrm{ind}} = \boldsymbol{\alpha}_{\boldsymbol{a}} \cdot \boldsymbol{E}_a^{\mathrm{total}} \tag{11}$$

where $\boldsymbol{\alpha}_a$ is the polarizability tensor at the center of mass of each MM molecule. The total electric field has four contributions

$$\mathbf{E}_a^{\mathrm{total}} = \langle \hat{\mathbf{R}}\mathbf{r}_a \rangle + \mathbf{E}^{n}(\vec{R}_a) + \mathbf{E}^{s}(\vec{R}_a) + \mathbf{E}^{\mathrm{ind}}(\vec{R}_a) \tag{12}$$

where $\mathbf{E}^{s}(\vec{R}_a)$ is the electric field due to the MM partial charges and the last term represents the electric field due to the other induced dipole moments. Note that eq 8 is slightly different from the corresponding equation in our previous work.[10] This is due to the fact that we have chosen to include the interaction energy between the MM induced dipoles and the MM partial charges in the MM energy. Hence, the MM energy may be decomposed into a intramolecular term, $E_{\mathrm{MM}}^{\mathrm{intra}}$, and an intermolecular contribution, $E_{\mathrm{MM/MM}}$. The intermolecular MM Hamiltonian is written as

$$H_{\mathrm{MM/MM}}[|\Psi\rangle] = \frac{1}{2}\sum_{s,\, s'(s\neq s')}^{S} \frac{q_s q_{s'}}{|\vec{R}_s - \vec{R}_{s'}|} - \frac{1}{2}\sum_{a=1}^{A} \boldsymbol{\mu}_a^{\mathrm{ind}} \mathbf{E}^{s}(\vec{R}_a) + E_{\mathrm{MM/MM}}^{\mathrm{vdw}} \tag{13}$$

where the term $E_{\mathrm{MM/MM}}^{\mathrm{vdw}}$ is the van der Waals MM/MM energy. Note that the dependence of $H_{\mathrm{MM/MM}}$ on the QM system through the induced dipole is illustrated by writing $H_{\mathrm{MM/MM}}$ as a functional of the wave function, $|\Psi\rangle$. Even though this energy contribution is included in the MM energy, we explicit consider the effect of the last term in eq 13 in the optimization of the wave function. The important point to note here is that if the interaction energy between the MM induced dipoles and the MM partial charges is included in the MM energy we do not need to consider the induced dipole moments in the situation where the QM system is replaced by a classical molecule (called $\mathbf{p}_a^{\mathrm{ind}}$ in ref 10) in order to decompose the interaction energy into physical recognizable energy contributions.

In a mean-field approximation, the induced dipole moment is, in each iteration of the optimization of the wave function, calculated according to eqs 11 and 12. This means, that the induced dipole moment becomes independent of the electronic degrees of freedom. Therefore, the induced dipole moment entering eq 8 becomes an ordinary vector and *not* a vector of electronic operators.

By taking the expectation value of eq 2 and making use of eq 11, we obtain the following expression for the total interaction energy

$$E_{\mathrm{QM/MM}} = E^{\mathrm{vdw}} + E^{\mathrm{el}} + E^{\mathrm{pol}} = E^{\mathrm{vdw}} + E_{\mathrm{S,N}}^{\mathrm{el,nuc}} - \sum_{s=1}^{S}\langle \hat{N}_s \rangle - \frac{1}{2}\sum_{a=1}^{A}\langle \hat{\mathbf{R}}\mathbf{r}_a \rangle^{\mathrm{T}} \boldsymbol{\alpha}_a \{\langle \hat{\mathbf{R}}\mathbf{r}_a \rangle + \mathbf{O}_a^{\mathrm{ns}}(\vec{R}_a)\} + O_{\mathrm{ind}}^{\mathrm{ns}} \tag{14}$$

where the vector $\mathbf{E}_a^{\mathrm{ns}}(\vec{R}_a)$ and the energy term $E_{\mathrm{ind}}^{\mathrm{ns}}$ are calculated according to

$$\mathbf{O}_a^{\mathrm{ns}}(\vec{R}_a) = 2\mathbf{E}^{n}(\vec{R}_a) + \mathbf{E}^{s}(\vec{R}_a) + \mathbf{E}^{\mathrm{ind}}(\vec{R}_a) \tag{15}$$

and

$$O_{\mathrm{ind}}^{\mathrm{ns}} = -\frac{1}{2}\sum_{a=1}^{A}[(\mathbf{E}^{n\prime}(\vec{R}_a))^{\mathrm{T}} \boldsymbol{\alpha}^{a}\{\mathbf{E}^{n}(\vec{R}_a) + \mathbf{E}^{s}(\vec{R}_a) + \mathbf{E}^{\mathrm{ind}}(\vec{R}_a)\}] \tag{16}$$

Here, $\mathbf{E}^{s}(\vec{R}_a)$ represents the electric field at the center of mass a of each MM molecule due to the partial MM charges

$$\mathbf{E}^{s}(\vec{R}_a) = \sum_{s \notin a} \frac{q_s(\vec{R}_a - \vec{R}_s)}{|\vec{R}_a - \vec{R}_s|^3} \tag{17}$$

and the field due to the induced dipole moments, $\mathbf{E}^{\mathrm{ind}}(\vec{R}_a)$, is written in terms of the dipole tensor, $\mathbf{T}_{aa'}$

$$\mathbf{E}^{\mathrm{ind}}(\vec{R}_a) = \sum_{a' \neq a} \mathbf{T}_{aa'} \boldsymbol{\mu}_{a'}^{\mathrm{ind}} \tag{18}$$

where

$$\mathbf{T}_{aa'} = \frac{1}{|\vec{R}_a - \vec{R}_{a'}|^3}\left[\frac{3(\vec{R}_a - \vec{R}_{a'})(\vec{R}_a - \vec{R}_{a'})^{\mathrm{T}}}{|\vec{R}_a - \vec{R}_{a'}|^2} - \mathbf{1}\right] \tag{19}$$

The summation restriction $s \notin a$ used in eq 17 is due to the fact that only sites s *not* found in the molecule with center of mass a is to be included in the summation. The polarization contribution to the total interaction energy is given as

$$E_{\mathrm{pol}} = -\frac{1}{2}\sum_{a=1}^{A}\langle \hat{\mathbf{R}}\mathbf{r}_a \rangle^{\mathrm{T}} \boldsymbol{\alpha}_a \{\langle \hat{\mathbf{R}}\mathbf{r}_a \rangle + \mathbf{O}_a^{\mathrm{ns}}(\vec{R}_a)\} + O_{\mathrm{ind}}^{\mathrm{ns}} \tag{20}$$

We note that the total induction energy of the system consists of E_{pol} in eq 20 together with interaction between the induced moments and the field due to the MM partial charges (last term in eq 13), whereas the dispersion energy is included in the van der Waals term.

In this paper, we set up two hierarchies of QM/MM models; one concerning the electronic description of the QM system and one concerning the coupling between the QM and MM system. For the hierarchy concerning the coupling of the two subsystems, we introduce the following four models:

In model A, we neglect the MM polarization; that is, we keep only $\hat{H}^{\mathrm{el}}$ and $\hat{H}^{\mathrm{vdw}}$ in the expression for the interaction Hamiltonian, $\hat{H}_{\mathrm{QM/MM}}$.

In model B, we neglect the MM polarization in the optimization of the QM/MM wave function, but we calculate, using this wave function, the polarization energy as the *expectation value* of eq 8; that is, we use the wave function obtained without the MM polarization to calculate the polarization energy as well as the induction contribution to the MM energy, eq 13. In this approach, we also neglect the contribution to the electric field from the induced dipole moments (the last term in eq 12). Therefore, the induced dipole moments are determined in a noniterative manner.

In model C, we proceed as in model B and use the wave function obtained without MM polarization, but in this case, we include the full solution to the set of equations for the induced dipole moments; that is, we use an iterative procedure to calculate the induced dipole moments and obtain for the polarization energy

$$E_{\mathrm{pol}}^{\mathrm{C}} = -\frac{1}{2}\sum_{a=1}^{A} \boldsymbol{\mu}_a^{\mathrm{ind}} \cdot (\langle \hat{\mathbf{R}}\mathbf{r}_a \rangle + \mathbf{E}^{n}(\vec{R}_a)) \tag{21}$$

Again we note that there is an additional induction energy contribution because of the MM system through eq 13. Finally, in model D, we introduce the solvent polarization in the optimization of the CC/MM wave function.

Using these four models, we derive a hierarchy of CC/MM models, and we may study the importance of including the MM polarization in the calculation of interaction energies.

B. Coupled Cluster Theory for States in a Vacuum. In this subsection, CC theory for the vacuum case is described.

The CC wave function ansatz is given by[51,52]

$$|CC\rangle = \exp(\hat{T})|\text{HF}\rangle \tag{22}$$

with

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_n = \sum_{i=1}^{n}\sum_{\mu_i} t_{\mu_i}\hat{\tau}_{\mu_i} \tag{23}$$

The t_{μ_i} parameters are the excitation amplitudes and $\hat{\tau}_{\mu_i}$ are the corresponding *i*-electron excitation operators. The state, $|\text{HF}\rangle$, is the Hartree−Fock reference wave function. The CC energy may be calculated as

$$E_{\text{CC}} = \langle \text{HF}|\, \hat{H} \exp(\hat{T})|\text{HF}\rangle \tag{24}$$

and the amplitude equations are given as

$$e_{\mu_i} = \langle \mu_i|\, \exp(-\hat{T})\hat{H}\exp(\hat{T})|\text{HF}\rangle = 0 \tag{25}$$

where $\langle \mu_\text{i}| = \langle \text{HF}|\hat{\tau}^{\dagger}_{\mu_i}$, $\langle \mu_i|\nu_j\rangle = \delta_{\mu,\nu}\delta_{i,j}$.

This parametrization of the CC wave function possesses the flexibility to represent the full configuration interaction (FCI) wave function when the cluster expansion in eq 23 includes all possible electronic excitations in the molecular system. Truncating the expansion leads to a series of approximate CC models. Truncating the cluster expansion after the second term, i.e., $\hat{T} = \hat{T}_1 + \hat{T}_2$, defines the coupled cluster singles and doubles (CCSD)[53] model.

Because the energy and wave function parameters are determined by projection, the CC method is nonvariational. However, it is possible to introduce a variational Lagrangian[54−57]

$$L_{\text{CC}}(\boldsymbol{t}, \bar{\boldsymbol{t}}) = E_{\text{CC}}(\boldsymbol{t}) + \sum_{i,\mu_i} \bar{t}_{\mu_i} e_{\mu_i} = E_{\text{CC}}(\boldsymbol{t}) + \bar{\boldsymbol{t}}\boldsymbol{e}(\boldsymbol{t}) \tag{26}$$

where the vector $\bar{\boldsymbol{t}}$ contains the Lagrangian multipliers and the vector $\boldsymbol{e}(\boldsymbol{t})$ is the amplitude equations. We require that the Lagrangian is simultaneously stationary with respect to $\boldsymbol{t}$ and $\bar{\boldsymbol{t}}$ and we obtain

$$0 = \frac{\partial L_{\text{CC}}(\boldsymbol{t}, \bar{\boldsymbol{t}})}{\partial \bar{t}_{\mu_i}} = e_{\mu_i}(\boldsymbol{t}) = \langle \mu_i|\, \exp(-\hat{T})\hat{H}\exp(\hat{T})|\text{HF}\rangle \tag{27}$$

$$0 = \frac{\partial L_{\text{CC}}(\boldsymbol{t}, \bar{\boldsymbol{t}})}{\partial t_{\nu_j}} = \frac{\partial E_{\text{CC}}(\boldsymbol{t})}{\partial t_{\nu j}} + \sum_{i,\mu_i} \bar{t}_{\mu_i} \frac{\text{d}e_{\mu_i}(\boldsymbol{t})}{\text{d}t_{\nu_j}} = \langle \Lambda|[\hat{H}, \hat{\tau}_{\nu_\text{j}}]|\, \text{CC}\rangle \tag{28}$$

with $\langle \Lambda|$ defined as

$$\langle \Lambda| = (\langle \text{HF}| + \sum_{i,\mu_i} \bar{t}_{\mu_i}\langle \mu_i|)\exp(-\hat{T}) \tag{29}$$

When these equations are fulfilled, the Lagrangian gives the CC energy. By using this Lagrangian technique, we may show that in CC theory an expectation value for a real operator is evaluated according to the asymmetric expression

$$\langle \hat{X}\rangle = \langle \Lambda|\hat{X}|\, \text{CC}\rangle \tag{30}$$

C. Coupled Cluster/Molecular Mechanics Wave-Function. In the previous subsections, we have described QM/MM energy contributions and the CC theory for molecules in a vacuum. The problem is how to combine these equations. This is complicated by the fact that the CC methodology is not variational. For a variational method one would straightforwardly optimize the energy with the additional QM/MM energy contributions. We solve this nontrivial problem by using the concept of a variational Lagrangian. This has the advantage of ensuring the correct limit when the cluster expansion is not truncated. This limit is for example not guaranteed if simply an additional effective QM/MM one electron operator is introduced into the CC equations.

To derive the optimization conditions for the CC/MM wave function, we extend the vacuum CC Lagrangian by augmenting it with the interaction term in a way similar to the approach where the CC method is coupled to a dielectric continuum.[40,41] This is simply done by adding the vacuum Lagrangian and the CC/MM interaction energy but in a form where we have introduced the CC expectation values (eq 30). The CC/MM Lagrangian becomes[10]

$$\begin{aligned} L_{\text{CCMM}}(\boldsymbol{t}, \bar{\boldsymbol{t}}) = {} & \\ & \langle \Lambda|\, \hat{H}_{\text{QM}}|\text{CC}\rangle - \sum_{s=1}^{S}\langle \hat{N}_s\rangle - \frac{1}{2}\sum_{a=1}^{A}\langle \hat{\mathbf{R}}\mathbf{r}_a\rangle^{\text{T}}\boldsymbol{\alpha}_a[\langle \hat{\mathbf{R}}\mathbf{r}_a\rangle + \\ & \mathbf{O}_a^{\text{ns}}(\vec{R}_a)] + E^{\text{vdw}} + E_{\text{S,N}}^{\text{el,nuc}} + O_{\text{ind}}^{\text{ns}} + E_{\text{MM}}(\boldsymbol{t}, \bar{\boldsymbol{t}}) \end{aligned} \tag{31}$$

Note, that this Lagrangian is nonlinear in both the $\boldsymbol{t}$ and $\bar{\boldsymbol{t}}$ parameters.

As in the vacuum case, we require that this Lagrangian is stationary with respect to both the $\boldsymbol{t}$ and $\bar{\boldsymbol{t}}$ parameters and by introducing the one-electron interaction operator, $\hat{T}^{\text{g}}$

$$\hat{T}^{\text{g}} = -\sum_{s=1}^{S}\hat{N}_{\text{s}} - \sum_{a=1}^{A}\left[\langle \Lambda|\, \hat{\mathbf{R}}\mathbf{r}_{\mathbf{a}}|CC\rangle + \frac{1}{2}\mathbf{E}_{\text{a}}^{\text{ns}}(\vec{R}_{\text{a}})\right]^{\text{T}}\boldsymbol{\alpha}_{\text{a}}\hat{\mathbf{R}}\mathbf{r}_{\mathbf{a}} \tag{32}$$

where

$$\mathbf{E}_a^{\text{ns}}(\vec{R}_a) = 2\mathbf{E}^{n}(\vec{R}_a) + 2\mathbf{E}^{s}(\vec{R}_a) + \mathbf{E}^{\text{ind}}(\vec{R}_a) \tag{33}$$

we obtain the optimization conditions for the CC/MM wave function

$$\frac{\partial L_{\text{CCMM}}(\boldsymbol{t}, \bar{\boldsymbol{t}})}{\partial \bar{t}_{\mu_i}} = \langle \mu_i|\, \exp(-\hat{T})[\hat{H}_{\text{QM}} + \hat{T}^{\text{g}}]\exp(\hat{T})|\, \text{HF}\rangle = 0 \tag{34}$$

and

$$\frac{\partial L_{\text{CCMM}}(\boldsymbol{t}, \bar{\boldsymbol{t}})}{\partial t_{\nu_i}} = \langle \Lambda|[\hat{H}_{\text{QM}} + \hat{T}^{\text{g}}, \hat{\tau}_{\nu_i}]|\, \text{CC}\rangle = 0 \tag{35}$$

Note that because the $\hat{T}^{\text{g}}$ operator depends on both the $\boldsymbol{t}$ and $\bar{\boldsymbol{t}}$ parameters we find that eqs 34 and 35 are coupled. Clearly, this represents an additional complication compared to the corresponding optimization conditions for a molecule in a vacuum. It is also important to note that because the CC/MM method is nonvariational the CC/MM total energy is generally not bounded from below by the exact total energy. Here, exact total energy refers to the corresponding total energy calculated using a full configuration interaction/molecular mechanics (FCI/MM) wave function for a given one-electron basis set. Because the lack of such a lower bound has not been shown to be a problem for vacuum CC calculations, we do not expect this to be a problem in the case of a CC/MM wave function. In the limit of a full cluster expansion, the CC/MM method has the flexibility to represent the FCI/MM results. Thus, the CCSD/MM method becomes exact (for a given one-electron basis set) for a two-electron QM system coupled to a force field.

In the above derivations of the CC/MM wave function optimization conditions, we used the *unrelaxed* approach. By introducing orbital relaxation, a corresponding orbital relaxed CC/MM wave function approach could be derived. However, because the relaxed approach has been shown to spoil the pole structure of the CC response functions,[57] we prefer to work within the unrelaxed formulation of the CC/MM model in the treatment of the polarization term.

III. Implementation

The CC/MM method is implemented in a local version of the Dalton program package[58] at the CCSD level of theory. As noted before, the CCSD/MM optimization equations for $\mathbf{t}$ and $\bar{\mathbf{t}}$ are coupled because of the introduction of the $\hat{T}^g$ operator. Different approaches can be adopted for their solution. On one hand, one could iterate the solution of $\mathbf{t}$ and $\bar{\mathbf{t}}$ simultaneously. On the other hand, one could for a given $\hat{T}^g$ operator solve them as were they completely decoupled and then perform some "outer" iterations on the T^g operator (to be detailed now). Both strategies should of course converge to the same solution but potentially at different costs.

Our procedure for solving the equations is most in line with the second of the above strategies, but with some additional features. This choice was made initially because it agreed best with the existing program structure and required fewer changes to the existing modularity. Below, we also discuss its efficiency. The basic procedure is illustrated in Figure 1. From a set of $\mathbf{t}$ and $\bar{\mathbf{t}}$ parameters, we construct the $\hat{T}^g$ operator, and treating this as fixed, we may solve the $\mathbf{t}$ equations and thereafter the $\bar{\mathbf{t}}$ equations (depending on the $\boldsymbol{t}$ parameters) as in the vacuum case but with modified one-electron integrals. This gives a new set of $\mathbf{t}$ and $\bar{\mathbf{t}}$ parameters, and the $\hat{T}^g$ operator can then be updated. Restarting with the new $\mathbf{t}$ and $\bar{\mathbf{t}}$ parameters, this procedure may be continued until convergence is obtained in these so-called "outer" iterations. As input for the $\boldsymbol{t}$ and $\bar{\boldsymbol{t}}$ parameters, we may use amplitudes from a previous vacuum, dielectric solvent, or another QM/MM calculation with the same QM configuration, or we may simply start from vacuum MP2 amplitudes. In each iteration, we also need to update the induced dipole moments, $\boldsymbol{\mu}_a^{\mathrm{ind}}$, which can be calculated using eq 11. However, because the induced dipole moments depend on the induced electric field, this equation has to be solved iteratively. The computational cost of solving the induced moment equations will typically be small compared to the cost of solving the CC equations.

When solving the $\mathbf{t}$ and $\bar{\mathbf{t}}$ equations iteratively through a set of outer iterations, we may choose between different strategies. We may continue the iterations in the $\mathbf{t}$ equations until convergence is obtained and then use this set of $\mathbf{t}$ parameters in the solution of the $\bar{\mathbf{t}}$ equations which in turn also are iterated until convergence is obtained. However, from a computational point of view, it seems more favorable only to perform a few iterations (n_{itt}) in the $\mathbf{t}$ equations and then use this set of $\mathbf{t}$ parameters in the solution of the $\bar{\mathbf{t}}$ equations. Also, in the solution of the $\bar{\mathbf{t}}$ equations, we only perform a few iterations (n_{ittbar}). With these sets of (un)converged $\mathbf{t}$ and $\bar{\mathbf{t}}$ parameters, we now update the $\hat{T}^g$ operator and repeat the solution of the $\mathbf{t}$ and $\bar{\mathbf{t}}$ parameters until convergence is obtained in both the $\boldsymbol{t}$ and the $\bar{\boldsymbol{t}}$ parameters and of course with respect to the outer iterations. The number of iterations can be limited in various ways: we have simply introduced the numbers n_{itt} and n_{ittbar} as input values, allowing us to experiment and test this functionality as will be described now.

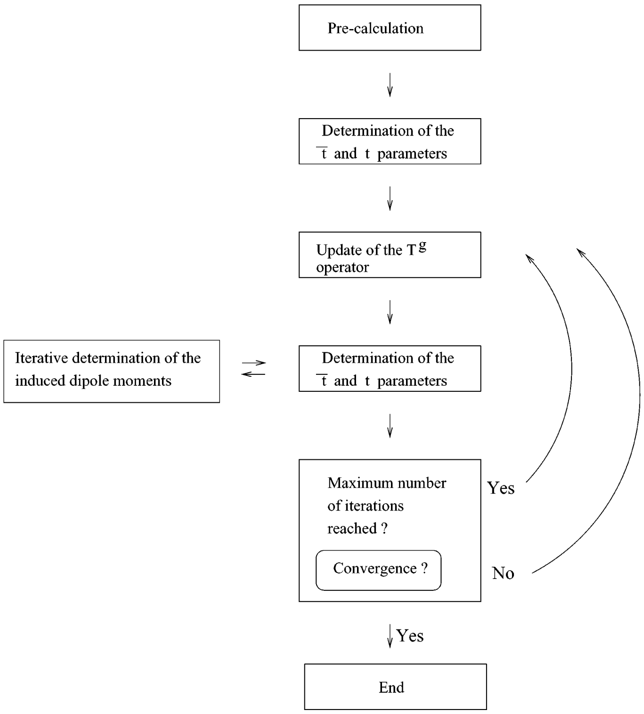

Figure 1. Illustration of the CC/MM wave function optimization procedure.

In Table 1, we have shown the results for a series of calculations of a water molecule (the QM system) surrounded by 127 other water molecules (the MM system). The MM parameters are as given in section IV and the basis set used is the aug-cc-pVTZ basis set.[67] In Table 1, n_{itt} is the maximum number of iterations in the solution of the $\mathbf{t}$ equations and n_{ittbar} is the corresponding maximum number of iterations in the solution of the $\bar{\mathbf{t}}$ equations. In the last column, ∞ means that the $\boldsymbol{t}$ equations are iterated to convergence before the $\bar{\mathbf{t}}$ equations are solved. Also, these equations are iterated to convergence before the update of the $\hat{T}^g$ operator. The symbol N_{outer} is the total number of *macro* iterations, i.e., updates of the $\hat{T}^g$ operator. The term N_{it} is the total number of "inner iterations" which is used as a measure of the cost − the number of times we have taken the most time-consuming steps.

The nonlinear $\mathbf{t}$ equations are solved by a DIIS algorithm (direct inversion in the iterative subspace) where in each iteration we have to evaluate the vector function in eq 34 with the given trial $\mathbf{t}$. The number of iterations taken is added to N_{it}. After the solution of the $\mathbf{t}$ equations, some intermediates are calculated as a necessary precursor for the efficient solution of the $\bar{\mathbf{t}}$ equations. The cost of the calculation of all intermediates are the same order of magnitude as one iteration in the $\mathbf{t}$ equations, and an additional one is added to N_{it} to account for this. Then the linear $\bar{\mathbf{t}}$ equations are solved by a linear equation solver requiring the transformation of the Jacobian in each iteration. The cost of one Jacobian transformation is, as the previously

TABLE 1: Total Number of Iterations (N_{it}) as a Function of the Maximum Number of Iterations in the Solution of the t (n_{itt}) and $\bar{t}$ (n_{ittbar}) Equations[a]

n_{itt}	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	6	∞
n_{ittbar}	2	2	3	4	5	3	4	5	6	3	4	5	6	3	4	5	6	7	∞
N_{outer}	25	25	9	6	5	9	6	5	5	9	6	5	5	9	6	5	5	5	5
N_{it}[b]	112	119	58	49	50	61	52	53	58	64	55	56	61	67	59	59	64	69	94

[a] The QM system consists of a water molecule and 127 other water molecules represent the MM part of the system. The basis set used is the aug-cc-pVTZ basis set. [b] Thresholds used in the calculations are 10^{-7} for the change in the "vacuum" energy term in two iterations for the **t** equations and 10^{-7} for the norm of the residual in the $\bar{\mathbf{t}}$ equations. The change in energy, norm of **t** and norm of $\bar{\mathbf{t}}$ in two subsequent outer iterations are all required to be smaller than 10^{-6} for convergence.

calculated intermediates, roughly of the same order of magnitude as one iteration in the ***t*** equations. The number of iterations taken is added to N_{it}. On convergence, N_{it} is roughly proportional to the cost of the calculation.

Compared to a vacuum CC calculation, the cost of a CC/MM calculation may be more expensive with the same QM system because of two primary effects: (i) more iterations are required and (ii) each iteration takes longer time. The remaining contribution is, at least for the cases we have studied so far, of less importance. In each iteration in CCSD, we have a number of terms which we for brevity denote as N^6 scaling terms where N is the number of orbitals. All of the CC/MM contributions scales according to N^5 so their computational scaling is more modest. However, one should note that they are repeated a number of times depending on the number of sites and center of masses for the polarizability. The calculation of the QM/MM vector function for this example is about 50% more expensive than for a vacuum calculation.

As seen from Table 1, the total number of iterations is minimized if the maximum number of iterations in the solution of the **t** equations are 3 and the $\bar{\mathbf{t}}$ equations are 4. Decreasing the number of maximum iterations increases the total number of iterations and in the situation where $n_{\mathrm{ittbar}} = 2$ and $n_{\mathrm{itt}} = 2$ or 3 the total number of iterations exceeds that in the case where $n_{\mathrm{itt}} = n_{\mathrm{ittbar}} = \infty$.

The total number of iterations necessary for calculating the energy for a water molecule in a vacuum is found to be 11. In order also to calculate the dipole moment, the total number of iterations (N_{it} including **t**, $\bar{\mathbf{t}}$ and intermediates) increases to 28 (with similar thresholds and accuracy). In the QM/MM case, we need, to obtain the energy, to solve for both the **t** and $\bar{\mathbf{t}}$ parameters because of the introduction $\hat{T}^g$ operator. This also means that in order to calculate first order properties, like the dipole moment, we do not need to solve any extra equations. This is different from the vacuum case where the energy may be calculated with only knowledge of the **t** parameters. Hence, the vacuum calculation to be compared with the numbers in Table 1 is that for the dipole moment. Thus, we find that the introduction of the MM environment causes the total number of iterations to increase about a factor of 2 or less (from 28 to 49) compared to the vacuum case when choosing the maximum number of iterations properly.

Let us finally comment on this situation compared to an algorithm for the simultaneous solution of ***t*** and $\bar{\boldsymbol{t}}$ (a set of coupled nonlinear equations). To evaluate eqs 34 and 35 in each iteration, we require 3 units according to the counting done with N_{it} (1 for **t**, 1 for preparing intermediates with the particular ***t***, and finally 1 for the $\bar{\mathbf{t}}$ part). In our example, 11 **t** and 16 $\bar{\mathbf{t}}$ were needed. Taking 16 as a rough guess of the number of iterations required in such an algorithm, we end up with a count of 48 almost as the best choice in the implemented algorithm. So even though potentially more efficient, to gain important speed up compared to our algorithm, reconsideration of the algorithm with respect to construction of intermediates and other aspects must also be taken into account. Note that in this case we have searched for the optimal combination for solving the equation. Naturally, one would not do that in general. However, the choice appears not to be very difficult and sensitive. Thereby, reasonable good convergence is obtained using $n_{\mathrm{itt}} = n_{\mathrm{ittbar}} = 3$, 4, or 5.

The QM/MM integrals are calculated using the integral code in Dalton.[58] We need to calculate both potential energy integrals

$$n^s_{pq} = \left\langle \phi_p \left| \frac{q_s}{|\vec{R}_s - \vec{r}_i|} \right| \phi_q \right\rangle \tag{36}$$

and electric field integrals

$$\mathbf{t}^a_{pq} = \left\langle \phi_p \left| \frac{\vec{r}_i - \vec{R}_a}{|\vec{r}_i - \vec{R}_a|^3} \right| \phi_q \right\rangle \tag{37}$$

The only extension of the already existing code is that $\vec{R}_s$ and $\vec{R}_a$ refer to sites and center of masses, respectively, outside the QM region. However, this problem is not difficult to solve since, as seen from eqs 36 and 37, the only input we need is the value of q_s and the relative distances between the partial charges or induced dipole moments and the QM electrons. This has been presented in ref 17. In the optimization of the CC/MM wave function, the energy can be calculated, according to eq 31, within each macro-iteration, i.e., every time the $\hat{T}^g$ operator has been updated.

IV. The Water Dimer

In this section, we illustrate the CCSD/MM formalism on the water dimer. Here, we calculate the interaction energy as a function of the O−O distance and compare the CCSD/MM results with the corresponding ab initio results where both water molecules are treated at the CCSD level of theory. The water dimer has been used in a number of previous benchmark studies of QM/MM methods.[9,59,60] We emphasize that the MM parameters used are not optimized for use in dimer calculations and/or QM/MM calculations so the model is not fitted to the particular application. It should also be noted that we consider the water dimer not because it is easy to describe by QM/MM methods but rather because it is a really challenging benchmark system for QM/MM methods and of fundamental interest.

A. Computational Details. In the calculations, we have used the CCSD(T) intermolecular optimized geometry from ref 62 where the water monomer geometry has been frozen at the experimental monomer geometry.[63] The parameters used for the MM molecule are given as[64] $A_{\mathrm{ma}} = 2.083 \times 10^6$ a.u., $B_{\mathrm{ma}} = 45.41$ au, $\bar{\alpha} = 9.718$ au, $q_{\mathrm{H}} = 0.3345$, and $q_{\mathrm{O}} = -0.669$. A one-dimensional potential energy surface scan has been calculated as single-point energies by displacing the molecules, with all angles fixed, along the axis through the oxygen atoms. This is illustrated in Figure 2. In Figure 2, the notation of donor and acceptor used through out this paper is illustrated. The term

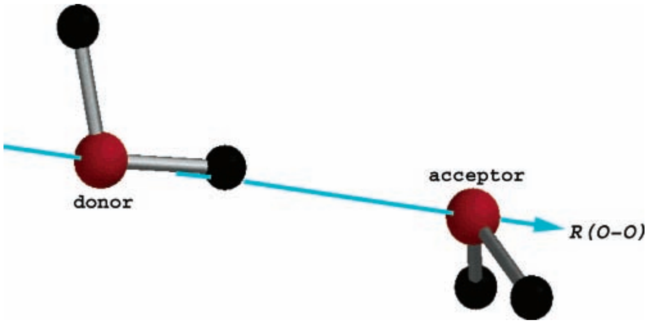

Figure 2. The water dimer. The water molecule denoted "donor" donates a proton, whereas the molecule denoted "acceptor" accepts a proton in the formation of the hydrogen bond between the two molecules. In calculating the interaction energy, the water molecules are displaced along the axis connecting the two oxygen atoms.

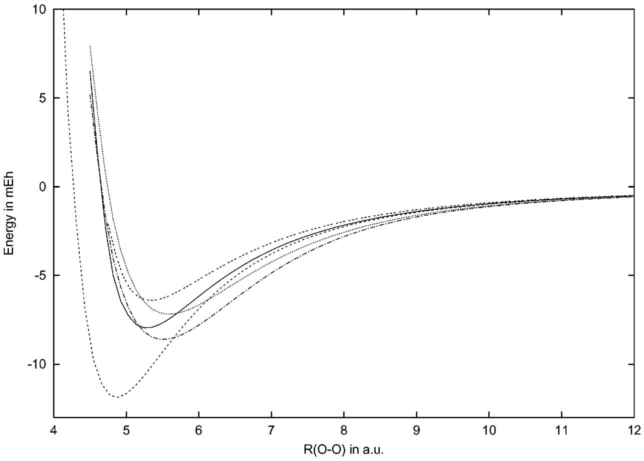

Figure 3. Comparison of the total interaction energy of the water dimer calculated at different levels of theory. The QM and QM/MM results are obtained using the aug-cc-pVTZ basis set. (-•-) CCSD calculated interaction energy. (•••) CCSD calculated interaction energy with "counterpoise correction". (- - -) $E_{QM/MM}$ with the proton donor as the QM molecule. (—) $E_{QM/MM}$ with the proton acceptor as the QM molecule. (- - - -) MM calculated interaction energy.

donor is used for the molecule donating a proton to the hydrogen bond, whereas the term acceptor is used for the molecule accepting a proton in the hydrogen bond.

B. Results and Discussion. Calculations have been performed at the CCSD/MM level of theory employing the cc-pVDZ and the aug-cc-pVTZ basis sets. This was done to investigate the effects of including diffuse functions and more polarization functions in the basis set. Either the proton acceptor or the proton donor molecule has been represented as the QM system. The CCSD/MM energy is shown in Figures 3 and 4 and the different energy contributions are addressed in Figures 5−8. For comparison, a full quantum mechanical calculation of the dimer at the CCSD level of theory employing the aug-cc-pVTZ basis set and a full molecular mechanical (MM) calculation were performed. The corresponding interaction energies are shown in Figure 3 and for the quantum mechanical results with and without the "counterpoise correction" (CPC).[65]

In Figure 3, it is seen that the $E_{QM/MM}$ curve with the acceptor as the QM system reproduces the potential energy surface of the quantum mechanical calculations fairly well but overestimates the attractive part of the energy between the two molecules. This leads to a shortening of the equilibrium distance R_{O-O} from 5.58 to 5.28 au (5%) compared to the CCSD(CPC) results. The MM curve has about the same equilibrium distance as the $E_{QM/MM}$ curve of the QM acceptor, but the interaction energy is underestimated compared to the quantum mechanical curves. The CC/MM curve, with the acceptor as the QM system, gives a better description of the interaction energy, producing a curve lying between the two quantum mechanical curves. The shortening of the equilibrium distance is expected to be partially

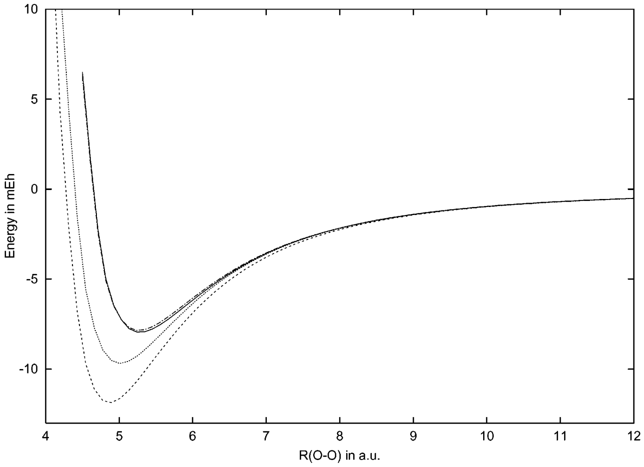

Figure 4. Comparison of $E_{QM/MM}$ calculated with different basis sets. (•••) The proton donor as the QM molecule with the cc-pVDZ basis set. (- - -) The proton donor as the QM molecule with the aug-cc-pVTZ basis set. (-•-) The proton acceptor as the QM molecule with the cc-pVDZ basis set. (—) The proton acceptor as the QM molecule with the aug-cc-pVTZ basis set.

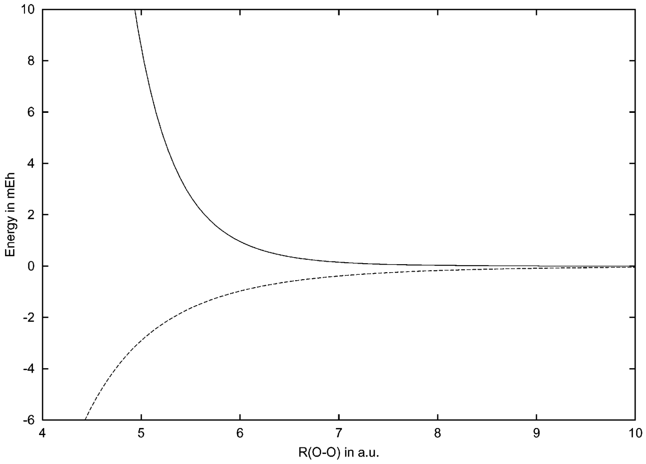

Figure 5. Two contributions to the van der Waals term. (—) The repulsive component of the potential *A/R*.[12] (- - -) The attractive component of the potential (the dispersion term) *B/R*.[6]

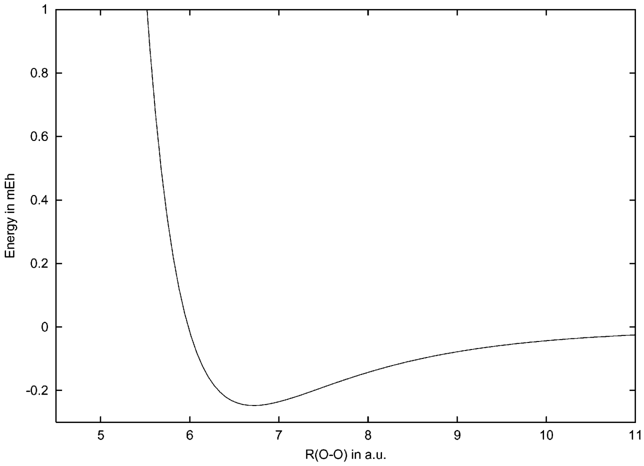

Figure 6. (—) van der Waals energy.

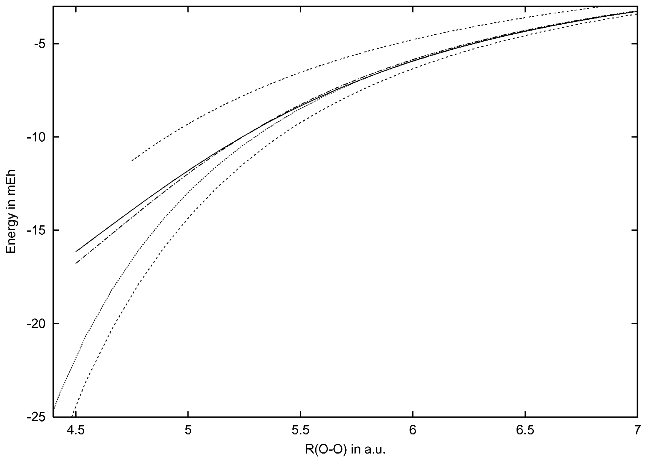

Figure 7. Electrostatic contribution to the interaction energy, compared for different basis sets. Also, the MM calculated values are shown. (···) The proton donor as the QM molecule with the cc-pVDZ basis set. (- - -) The proton donor as the QM molecule with the aug-cc-pVTZ basis set. (-·-) The proton acceptor as the QM molecule with the cc-pVDZ basis set. (—) The proton acceptor as the QM molecule with the aug-cc-pVTZ basis set. (- - - -) MM calculated electrostatic interaction energy.

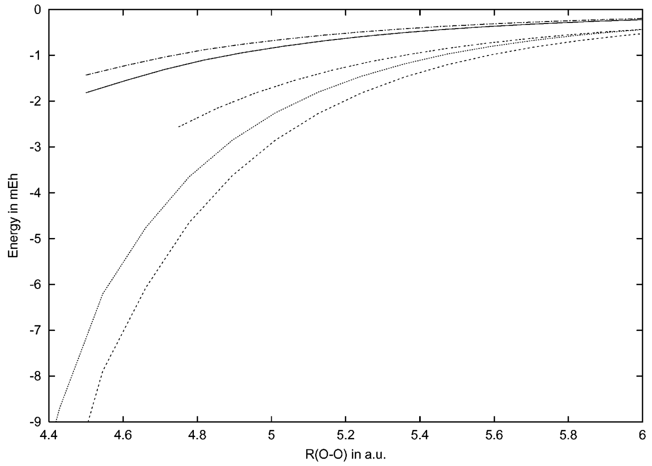

Figure 8. Polarization contribution to the interaction energy, compared for different basis sets. Also the MM calculated values are shown. (···) The proton donor as the QM molecule with the cc-pVDZ basis set. (- - -) The proton donor as the QM molecule with the aug-cc-pVTZ basis set. (— · —). The proton acceptor as the QM molecule with the cc-pVDZ basis set. (—) The proton acceptor as the QM molecule with the aug-cc-pVTZ basis set. (- - - -) MM calculated polarization energy.

because of the use of parameters found for liquid water which has been shown to overestimate the dimer interaction.[59] When the donor molecule is represented by quantum mechanics, the curve is lowered by 3.8 mE_h at the minimum of the energy curve. This corresponds to a lowering of 50% in the interaction energy compared to the QM acceptor case. The corresponding equilibrium distance is shortened to 4.87 au or by 13% compared to the CCSD(CPC) distance.

In Figure 4, it is seen that the $E_{QM/MM}$ curve of the QM acceptor molecule is almost unaffected when diffuse functions are included in the basis set, whereas it strongly effects the QM donor curve. Probably, this is due to the lack of short-range repulsion from the MM system in the wave function. In the CC/MM model, the dispersion and the short-range repulsion energies are modeled by the Lennard-Jones potential shown in Figures 5 and 6, but no short range repulsion effects enter the optimization conditions of the electronic wave function. Therefore, the electrons are not prevented from entering regions where they actually should be repelled by electrons of the solvent molecules.

In the QM acceptor case, the electrons are allowed to be located closer to the MM donor when diffuse functions are included in the basis set. This results in a small lift in the electrostatic energy which is seen in Figure 7. From Figure 8, we note that the polarization energy is lowered because the electrons interact stronger with the induced dipole moment at the center of mass of the donor. The changes are of the same size and cancel each other leaving $E_{QM/MM}$ almost unaltered as seen in Figure 3.

When describing the donor molecule by quantum mechanics, the electrostatic potential acting on the electrons is different from the case of the QM acceptor above. When diffuse functions are included in the basis set, the electrons are pulled toward the protons, because no short-range repulsion effects are included in the wave function optimization. This results in a substantial lowering of the electrostatic energy as seen in Figure 7. Also, the polarization energy is lowered as the electrons are drawn closer to the center of mass of the acceptor. This is seen in Figure 8. The overall $E_{QM/MM}$ is lowered as an effect of including diffuse functions in the basis set as seen in Figure 3.

The above discussion is supported by investigations performed by studying the dependence of the dipole moment and the electronic second-order moments. Other groups have found this change in energy when representing the donor or acceptor as the QM system in QM/MM models.[9,59,60] The calculated energy differences are smaller than what is found in this work using the aug-cc-pVTZ basis set. Probably, this is due to the lack of diffuse functions in the basis sets used in refs 9 and 59. However, as shown in ref 60 and discussed in ref 61, the permutational asymmetry in the interaction energy of the water dimer may be corrected by including explicitly the short-range repulsion into the wave function. However, a better representation of the MM molecules incorporating higher order multicenter multipole contributions to the electrostatic potential and the use of distributed polarizabilities (including anisotropies) would probably also reduce the asymmetric character in the interaction energy. Also, one could use partial charges not restricted to be located at the atomic sites and chosen such as to reproduce higher order electric moments correctly.

In conclusion, our results clearly show that caution should be taken in the construction of QM/MM calculations, as the QM/MM interface is not free of introducing artifacts, the size of which depends critically on details of the calculation. Using the cc-pVDZ basis set without diffuse functions and neglecting the MM polarizability in the CC/MM model the energy difference between the QM donor and QM acceptor is reduced to about 0.1 mE_h which is only one-third of the difference calculated in ref 59 using similar sized basis sets and without polarization. This only serves to illustrate that our methodology is not inherently worse than the common standard in this respect. Of course, using a smaller and more inaccurate basis set and neglecting important interactions is not a solution to the problem in general.

V. Energetic Study of Liquid Water

Having discussed the water dimer, we now turn our attention to a much larger sample consisting of 128 water molecules, where one of them is treated using quantum mechanics and the other 127 water molecules represent the MM part of the system.

A. Computational Details. All of the calculations are performed at the experimental gas-phase geometry of the water monomer.[64]

TABLE 2: Ground State QM/MM Energy Contributions for Liquid Water Calculated Using Model D (in au)

theory	basis set	$E_{\mathrm{QM/MM}}$	E^{vdw}	E^{el}	E^{pol}	$\Delta E_{\mathrm{MM/MM}}$	E_{sol}
HF/MM	aug-cc-pVDZ	−0.041697	0.009455	−0.042204	−0.008948	−0.000349	−0.034752
CCSD/MM	aug-cc-pVDZ	−0.040156	0.009455	−0.040969	−0.008642	−0.000288	−0.032703
HF/MM	aug-cc-pVTZ	−0.043164	0.009455	−0.043287	−0.009332	−0.000435	−0.035362
CCSD/MM	aug-cc-pVTZ	−0.042160	0.009455	−0.042474	−0.009141	−0.000401	−0.033776
HF/MM	aug-cc-pVQZ	−0.043546	0.009455	−0.043565	−0.009436	−0.000462	−0.035497
CCSD/MM	aug-cc-pVQZ	−0.042798	0.009455	−0.042948	−0.009305	−0.000443	−0.034164

In this work, we define an average geometry for the water cluster, and using this average geometry, we perform one QM/MM calculation.

The average geometry is obtained from molecular dynamical (MD) simulations performed for a box containing 128 water molecules utilizing periodic boundary conditions together with a spherical cutoff distance of 10.0 Å and a neighbor list technique with automatic check of the update interval. The cutoff distance is not used in the QM/MM calculations. The temperature and the pressure were kept constant using 298 K and 0.103 MPa as external values utilizing a scaling procedure. After equilibration of the sample, the simulation time for each trajectory was 20 ps. The total number of trajectories is 8000, starting from different initial velocity distributions.[66] The average geometry of the water sample is obtained from a Boltzmann sampling involving the 8000 trajectories.

The parameters are taken from ref 64 and are the same as for the water dimer. All of the QM/MM calculations employ the correlation consistent (aug)-cc-pVXZ (X = D,T,Q) basis set.[67]

The solvation energy is obtained according to

$$E_{\mathrm{Sol}} = \langle \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}} \rangle_{\mathrm{QM/MM}} - \langle \hat{H}_{\mathrm{QM}} \rangle_{\mathrm{Vac}} + \Delta E_{\mathrm{MM/MM}} \tag{38}$$

where the subscripts QM/MM and Vac refer to the QM/MM and vacuum optimized wave functions, respectively, and where the term $\Delta E_{\mathrm{MM/MM}}$ is defined as

$$\Delta E_{\mathrm{MM/MM}} = -\frac{1}{2}\sum_{a=1}^{A} (\boldsymbol{\mu}_a^{\mathrm{ind}} - \mathbf{p}_a^{\mathrm{ind}})\mathbf{E}^s(\vec{R}_a) \tag{39}$$

As in ref 10, $\mathbf{p}_a^{\mathrm{ind}}$ is the induced dipole moments in the case where the QM system is absent and replaced by a classical molecule. Thus, $\Delta E_{\mathrm{MM/MM}}$ represents the QM induced interaction energy in the MM system. When calculating the sum $\Sigma_{a=1}^{A}$ $\boldsymbol{\mu}_a^{\mathrm{ind}}\mathbf{E}^s(\vec{R}_a)$, the contribution from the molecule previously described using quantum mechanics is excluded. Thus, in this sense, this particular classical molecule may be viewed as a spectator molecule. However, in order not to treat relaxation effects in the water cluster directly, it is important to include the effect of this classical molecule in the determination of the classically induced dipole moments, $\mathbf{p}_a^{\mathrm{ind}}$.

To investigate the importance of inclusion of polarization effects in the MM system, we perform, using the CCSD/MM model, four series of calculations using model A, B, C, and D. Furthermore, to show the importance of electron correlation in the QM system, we compare, using model D, the HF/MM[17] and CCSD/MM results.

Finally, to compare the QM/MM approach to other models, we have also calculated the solvation energy of liquid water using the dielectric continuum model.[40,68] In these calculations, a water molecule is embedded in a spherical cavity with a radius equal to 4.0 au and surrounded by the dielectric continuum characterized by a static dielectric constant equal to 78.54 au.[68] Furthermore, in the multipole expansion of the solute charge distribution, we include moments up to $L_{\max} = 10$.

B. Results. Table 2 contains a comparison of the correlated CCSD/MM and the uncorrelated HF/MM results for the interaction and solvation energies. Furthermore, we have calculated and listed each contribution to the interaction energy separately. From this table, it is observed that the largest effect of solvation is found using a uncorrelated description. This is also known from the corresponding continuum calculations[40,68] and from earlier work using the MCSCF/MM wave function description of a solvated water molecule[17] where overestimation of the solvation and interaction energy is interpreted to be due to the large ionic component in the HF wave function. From Table 2, we find that the largest deviations are found using the smallest basis set and that the solvation energy is most sensitive to correlation effects. This deviation is found to be at least 6%.

Table 3 contains the results for the interaction energy of water in water. The interaction energy has been split up into three contributions, the van der Waals part, E^{vdw}, the electrostatic contribution, E^{el}, and the polarization contribution, E^{pol}. Furthermore, we have, using eq 38, calculated the solvation energy and, using eq 39, the QM induced interaction energy in the MM system.

From Table 3, we first note that each energy contribution (except the van der Waals term) is quite sensitive to the inclusion of diffuse functions in the basis set. Probably this appears because the diffuse functions tend to lower the QM electron−electron repulsion energy and that the inclusion of diffuse functions in the basis set allow the QM electrons to be located closer to the MM system with the possibility of lowering the corresponding interaction energy.

When studying the hierarchy of QM/MM models (model A, B, C, and D), we find that model A underestimates all the energy contributions, compared to model D (except the van der Waals term which does not depend on electronic wave function parameters), which results in an overall underestimation of the solvation energy. This is clearly due to the total neglect of polarization effects in the MM part of the system.

In models B and C, we also find this underestimation of the interaction energies ($E_{\mathrm{QM/MM}}$); that is, the wave function obtained without the MM polarization cannot account for all polarization effects, but the results are greatly improved compared to model A. Thus, we find that the calculation of polarization energy using even the quite simplified treatments improves the results significantly.

From Table 3, we also note that the results obtained using model B do not differ much from the corresponding results using model C. However, including the iterative determination of the induced dipole moments (model C) slightly lowers the polarization energy resulting in a overall lowering of the corresponding interaction energy of about 0.5%. Thus, we find that in the approach where the wave function is obtained without the MM polarization the electric field due to the induced dipole moments, and hence an iterative determination of the corresponding induced dipole moments, may be neglected.

TABLE 3: Ground State CCSD/MM(X) (X = A, B, C) Energy Contributions for Liquid Water (in au) and Dielectric Continuum (DC) Solvation Energies.

model	basis set	$E_{\mathrm{QM/MM}}$	E^{vdw}	E^{el}	E^{pol}	$\Delta E_{\mathrm{MM/MM}}$	E_{sol}
A	cc-pVDZ	−0.026777	0.009455	−0.036233		0.000000	−0.023941
B	cc-pVDZ	−0.033233	0.009455	−0.036233	−0.006456	−0.000536	−0.030933
C	cc-pVDZ	−0.033409	0.009455	−0.036233	−0.006631	0.000209	−0.030364
D	cc-pVDZ	−0.035868	0.009455	−0.038092	−0.007231	0.000060	−0.030753
DC	cc-pVDZ						−0.006985
A	aug-cc-pVDZ	−0.028560	0.009455	−0.038015		0.000000	−0.024414
B	aug-cc-pVDZ	−0.035958	0.009455	−0.038015	−0.007399	−0.000729	−0.032540
C	aug-cc-pVDZ	−0.036156	0.009455	−0.038015	−0.007596	−0.000023	−0.032033
D	aug-cc-pVDZ	−0.040156	0.009455	−0.040969	−0.008642	−0.000288	−0.032703
DC	aug-cc-pVDZ						−0.007681
A	aug-cc-pVTZ	−0.029594	0.009455	−0.039050		0.000000	−0.024983
B	aug-cc-pVTZ	−0.037312	0.009455	−0.039050	−0.007717	−0.000793	−0.033494
C	aug-cc-pVTZ	−0.037517	0.009455	−0.039050	−0.007923	−0.000096	−0.033002
D	aug-cc-pVTZ	−0.042160	0.009455	−0.042474	−0.009141	−0.000401	−0.033776
DC	aug-cc-pVTZ						−0.007778
A	aug-cc-pVQZ	−0.029935	0.009455	−0.039390		0.000000	−0.025214
B	aug-cc-pVQZ	−0.037741	0.009455	−0.039390	−0.007806	−0.000813	−0.033833
C	aug-cc-pVQZ	−0.037949	0.009455	−0.039390	−0.008014	−0.000118	−0.033347
D	aug-cc-pVQZ	−0.042798	0.009455	−0.042948	−0.009305	−0.000443	−0.034164
DC	aug-cc-pVQZ						−0.007858

The polarization energy is, in all the calculations, negative which reflects that the polarization effects tend to stabilize the system. Even though the electrostatic contribution is dominating, the polarization energy contributes as much as up to 25% of the total interaction energy. Thus, the introduction of polarization effects is crucial for accurate calculations. Additionally, polarization effects are important when calculating molecular properties.

From Table 3, we observe that $\Delta E_{\mathrm{MM/MM}}$ calculated using model B ($\Delta E_{\mathrm{MM/MM}}{}^{\mathrm{B}}$) always is less than $\Delta E_{\mathrm{MM/MM}}$ calculated using either model A, C, or D. Thus, even though the interaction energy decreases monotonically when going from model A to model D, the solvation energy need not to behave the same way. Actually, using the smallest basis set, the solvation energy calculated using model B ($E_{\mathrm{sol}}^{\mathrm{B}}$) is less than the solvation energy calculated using model D. Generally, we find that $E_{\mathrm{sol}}^{\mathrm{B}} < E_{\mathrm{sol}}^{\mathrm{C}}$ which is a consequence of the noniterative nature of the B model. In models B, C, and D, we also find that the van der Waals energy is of similar magnitude, but of opposite sign, as the polarization energy. Thus, these two contributions tend to cancel each other. This cancellation has also been observed in other works.[13,17,47]

Finally, when comparing the solvation energy calculated using the QM/MM and continuum model, we find that the continuum results are underestimated about a factor of 4 compared to the results obtained using model D. A large discrepancy is expected, especially for a molecule as polar as water, because of the neglect of specific intermolecular interactions in the continuum model.[68]

VI. Discussion and Summary

In this paper, we have derived and implemented the combined CC/MM wave function approach at the CCSD level of theory, the CCSD/MM method.

The use of CC methods in QM/MM calculations is certainly an improvement in accuracy on the QM part compared to standard QM/MM methods. Traditionally, an improvement in accuracy of the wave function model also warrants the use of a larger one-electron basis set for obtaining a balanced calculation with respect to accuracy and cost. However, the first test calculation on the water dimer shows clearly that larger basis sets may increase the magnitude of some undesired artifacts of the QM/MM link. This is a general problem using simple MM methods, and it is not unique to our CC/MM method, but the problem may become more obvious and its solution more important for accurate QM methods such as CCSD. In any case, the improvement of the effective operators describing the QM/MM link is an important subject of further research.

We have derived a hierarchy of CC/MM models in which the coupling terms are treated more elaborately going from level A to D. Finally, we have shown using a test sample consisting of 128 water molecules that both polarization and correlation effects are important in an accurate calculation of the interaction and solvation energies. For the calculations of polarization energies, we found that model B and C, in which the obtained wave function excludes the MM polarization, actually gives quite good results when compared to model D where the MM polarization effects are directly included in the optimization of the wave function. However, as we will discuss in a future paper, the introduction of the MM polarization directly into the wave function optimization condition is important for obtaining accurate results for many molecular response properties calculated within the QM/MM approach.[72]

In this paper, we have used a static approach for the calculation of interaction and solvation energies; that is, we have used the final averaged MM configuration from the MD simulation to perform *one* QM/MM calculation. Alternatively, we could perform a QM/MM calculation for each configuration in the MD simulation and then obtain the final results by averaging over all the QM/MM calculations. Even though the last method is quite time-consuming, it has the very important property that the effect of the MM molecules is treated in a dynamical way. A more pragmatic intermediate approach would be to perform the MD simulation using the MM force field and only sample a limited number of configurations.

The derived CC/MM method is quite general. For example, it allows for a construction of a semi-CC/MM model, in which a first solvation shell is included in the QM calculation while the rest of the solvent molecules are treated classically. This approach would be more time-consuming, but it allows for an even better description in particular of dispersion and short-range exchange-repulsion effects compared to the CC/MM method. In future work, we will consider possibilities to include short-range repulsion effects in the optimization condition of the wave function. Finally, the CC/MM model allows for the construction of a continuum-CC/MM model where the QM/MM system is embedded in a dielectric continuum.[69−71]

Acknowledgment. O.C. acknowledges support by the Swedish Natural Science Research Council (NFR) and the Danish National Science Research Council (SNF). K.V.M. thanks Statens Naturvidenskabelige Forskningsråd, Statens Tekniske Videnskabelige Forskningsråd, Danish Center for Scientific Computing (DCSC) and the EU-network MOLPROP for support.

References and Notes

(1) Noell, J.; Morokuma, K. *Chem. Phys. Lett.* **1975**, *36*, 465.
(2) Warshel, A.; Levit, M. *J. Mol. Biol.* **1976**, *103*, 227.
(3) Singh, U. C.; Kollman, P. A. *J. Comput.Chem.* **1986**, *7*, 718.
(4) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(5) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492.
(6) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516.
(7) Thole, B. T.; Van Duijnen, P. Th. *Chem. Phys.* **1982**, 71, 211.
(8) DeVries, A. H.; Van Duijnen, P.Th.; Juffer, A. H.; Rullmann, J. A. C.; Dijkman, J. P.; Merenga, H.; Thole, B. T. *J. Comput. Chem.* **1995**, *16*, 37.
(9) Van Duijnen, P. Th.; DeVries, A. H. *Int. J. Quantum. Chem.* **1996**, *60*, 1111.
(10) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *Mol. Phys* **2002**, *100*, 1813.
(11) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(12) Gao, J. *J. Phys. Chem.* **1992**, *96*, 6432.
(13) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182.
(14) Martín, M. E.; Sánchez, M. L.; Del Valle, F. J. O.; Aguilar, M. A. *J. Chem. Phys.* **2000**, *113*, 6308.
(15) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, *112*, 1133.
(16) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 3721.
(17) Poulsen, T. D.; Kongsted, J.; Osted, A.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys* **2001**, *115*, 2393.
(18) Poulsen, T. D.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2001**, *115*, 7843.
(19) Poulsen, T. D.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2001**, *116*, 3730.
(20) Sánchez, M. L.; Aguilar, M. A.; Del Valle, F. J. O. *J. Comput. Chem.* **1997**, *18*, 313.
(21) Sánchez, M. L.; Aguilar, M. A.; Del Valle, F. J. O. *J. Mol. Struct. (THEOCHEM)* **1998**, *426*, 181.
(22) Martín, M. E.; Sánchez, M. L.; Aguilar, M. A.; Del Valle, F. J. O. *J. Mol. Struct. (THEOCHEM)* **2001**, *537*, 213.
(23) Martín, M. E.; Aguilar, M. A.; Chalmet, S.; Ruiz-López, M. *Chem. Phys. Lett.* **2001**, *344*, 107.
(24) Kawashima, Y.; Dupuis, M.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 248.
(25) Dupuis, M.; Kawashima, Y.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 1256.
(26) Dupuis, M.; Aida, M.; Kawashima, Y.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 1242.
(27) Stanton, R. V.; Hartsough, D. S.; Merz, K. M. *J. Phys. Chem.* **1993**, *97*, 11868.
(28) Tuñón, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F.; Rivail, J. L. *J. Comp. Chem.* **1996**, *17*, 19.
(29) Brandle, M.; Sauer, J.; Dovesi, R.; Harrison, N. M. *J. Chem. Phys.* **1998**, *109*, 10379.
(30) Sierka, M.; Sauer, J. *J. Chem. Phys.* **2000**, *112*, 6983.
(31) Sauer, J.; Sierka, M. *J. Comput. Chem.* **2000**, *21*, 1470.
(32) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993**, *214*, 391.
(33) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994**, *100*, 7443.
(34) Day, P. N.; Jensen, J. H.; Gordon, M. S., Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
(35) Gauss, J. *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998.
(36) Paldus, J. *Methods in Computational Physics*; Plenum: New York, 1992.
(37) Helgaker, T.; Gauss, J.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430.
(38) Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409.
(39) Christiansen, O.; Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1995**, *103*, 7429.
(40) Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys* **1999**, *110*, 1365.
(41) Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys.* **1999**, *110*, 8348.
(42) Newton, M. D. *J. Phys. Chem.* **1975**, *79*, 2795.
(43) Warshel, A. *Chem. Phys. Lett.* **1978**, *55*, 454.
(44) Sánchez Marcos, E.; Terryn, B.; Rivail, J. L. *J. Phys. Chem.* **1985**, *87*, 4695.
(45) Ågren, H.; Medina-Llanos, C.; Mikkelsen, K. V. *Chem. Phys.* **1987**, *115*, 43.
(46) Mikkelsen, K. V.; Luo, Y.; Ågren, H.; Jørgensen, P. *J. Chem. Phys.* **1995**, *102*, 9362.
(47) Gao, J. *J. Comp. Chem.* **1997**, *18*, 1061.
(48) Gao, J.; Pavelites, J. J. *J. Am. Chem. Soc.* **1992**, *114*, 1912.
(49) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975.
(50) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(51) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256.
(52) Lee, T. J.; Scuseria, G. E. *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Kluwer Academic: Dordrecht, The Nethrelands, 1995.
(53) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.
(54) Arponen, J. *Ann. Phys.* **1983**, *151*, 311.
(55) Helgaker, T.; Jørgensen, P. *Theor. Chim. Acta.* **1989**, *75*, 111.
(56) Koch, H.; Jensen, H. J. Aa.; Jørgensen, P.; Helgaker, T.; Scuseria, G. E.; Schaefer, H. F. *J. Chem. Phys.* **1990**, *92*, 4924.
(57) Christiansen, O.; Jørgensen, P.; Hättig, C. *Int. J. Quantum. Chem.* **1998**, *68*, 1.
(58) Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Auer, A. A.; Bak, K. L.; Bakken, V.; Christiansen, O.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Hättig, C.; Hald, K.; Halkier, A.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Pedersen, T. B.; Ruden, T. A.; Sanchez, A. Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O. *Dalton, an ab initio electronic structure program, Release 1.2*; 2001; see http://www.kjemi.uio.no/software/dalton/dalton.html.
(59) Tu, Y.; Laaksonen, A. *J. Chem. Phys.* **1999**, *111*, 7519.
(60) Ben-Nun, M.; Martínez, T. J. *Chem. Phys. Lett.* **1998**, *290*, 289.
(61) Chalmet, S.; Ruiz-López, M. F. *Chem. Phys. Lett.* **2000**, *329*, 154.
(62) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theo. Chem. Acc.* **1997**, *97*, 150.
(63) Hoy, A. R.; Mills, I. M.; Strey, G. *Mol. Phys.* **1972**, *24*, 1265.
(64) Ahlström, P.; Wallqvist, A.; Engström, A.; Jönsson, B. *Mol. Phys.* **1989**, *68*, 563.
(65) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(66) Billing, G. D.; Mikkelsen, K. V. *Chem. Phys.* **1994**, *182*, 249.
(67) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(68) Mikkelsen, K. V.; Ågren, H.; Jensen, H. J. Aa.; Helgaker, T. *J. Chem. Phys.* **1988**, *89*, 3086.
(69) Chalmet, S.; Rinaldi, D.; Ruiz-López, M. F. *Int. J. Quantum. Chem.* **2001**, *84*, 559.
(70) Moriarty, N. W.; Karlström, G. *J. Chem. Phys.* **1997**, *106*, 6470.
(71) Sánchez, M. L.; Martín, M. E.; Aguilar, M. A.; Del Valle, F. J. O. *J. Comput. Chem.* **2000**, *21*, 705.
(72) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Chem. Phys.* **2003**, *118*, 1620.