Development and application of QM/MM methods to study the solvation effects and surfaces

by

Pooja Arora Dibya

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Physical Chemistry

Program of Study Committee:
Mark S. Gordon, Major Professor
James W. Evans
Jacob W. Petrich
Klaus Schmidt-Rohr
Theresa L. Windus

Iowa State University

Ames, Iowa

2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am deeply thankful to many people in my life for all my accomplishments. I will start with my family: my grandmother Vidyawati, my parents Mohinder Pal and Reeta, my brother Saksham and my sister Neha, for all the love, care and support they have given me.

It is difficult to overstate my gratitude to my advisor, Prof. Mark S. Gordon, for mentoring me throughout my graduate studies. In addition, the opportunities provided in Gordon group trained me in all the aspects of graduate career such as development, application of GAMESS, presenting scientific results in regional and national meetings, writing manuscripts. Your constant encouragement and support guided me well throughout the graduate school career. Thank you for all this.

I would also like to thank all present and past Gordon group members for their insightful comments during practice talks, group meetings and the manuscript editing. I would like to thank my M.S advisor, Prof. Kurur, who taught me chemistry to a point that encouraged me to pursue higher studies.

My husband, I cannot thank you enough for everything, I wouldn't have done it without your support.

CHAPTER 1. GENERAL INTRODUCTION

I. General Overview

Quantum mechanical (QM) calculations have the advantage of attaining high-level accuracy, however QM calculations become computationally inefficient as the size of the system grows. Solving complex molecular problems on large systems and ensembles by using quantum mechanics still poses a challenge in terms of the computational cost. Methods that are based on classical mechanics are an inexpensive alternative, but they lack accuracy. A good trade off between accuracy and efficiency is achieved by combining QM methods with molecular mechanics (MM) methods to use the robustness of the QM methods in terms of accuracy and the MM methods to minimize the computational cost. Two types of QM combined with MM (QM/MM) methods are the main focus of the present dissertation: the application and development of QM/MM methods for solvation studies and reactions on the Si(100) surface.

The solvation studies were performed using a discreet solvation model that is largely based on first principles called the effective fragment potential method (EFP).[1,2] The main idea of combining the EFP method with quantum mechanics is to accurately treat the solute-solvent and solvent-solvent interactions, such as electrostatic, polarization, dispersion and charge transfer, that are important in correctly calculating solvent effects on systems of interest. A second QM/MM method called SIMOMM (surface integrated molecular orbital molecular mechanics) is a hybrid QM/MM embedded cluster model that mimics the real surface.[3] This method was employed to calculate the potential energy surfaces for reactions

of atomic O on the Si(100) surface. The hybrid QM/MM method is a computationally inexpensive approach for studying reactions on larger surfaces in a reasonably accurate and efficient manner.

II. Dissertation Organization

This thesis is comprised of four chapters: Chapter 1 describes the general overview and motivation of the dissertation and gives a broad background of the computational methods that have been employed in this work. Chapter 2 illustrates the methodology of the interface of the EFP method with the configuration interaction with single excitations (CIS) method to study solvent effects in excited states. Chapter 3 discusses the study of the adiabatic electron affinity of the hydroxyl radical in aqueous solution and in micro-solvated clusters using a QM/EFP method. Chapter 4 describes the study of etching and diffusion of oxygen atom on a reconstructed Si(100)-2X1 surface using a hybrid QM/MM embedded cluster model (SIMOMM). Chapter 4 elucidates the application of the EFP method towards the understanding of the aqueous ionization potential of Na atom. Finally, a general conclusion of this dissertation work and prospective future direction are presented in Chapter 6.

III. Theoretical Background

The present chapter gives an overview of the QM/MM methods that have been employed in this dissertation, namely QM/EFP and an embedded QM/MM cluster model called

SIMOMM. Also presented here is a brief discussion of the various QM methods that have been employed in this work. The basis of most QM methods is the Schrodinger equation. The time-independent Schrodinger equation[4-10] can be written as:

$$H\Psi = E\Psi \tag{1}$$

Ψ is the wavefunction which is the eigenfunction of the Hamiltonian with eigenvalue E. The Hamiltonian for the n electrons and N nuclei is given as:

$$\hat{H} = -\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{n}\sum_{A=1}^{N}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} + \sum_{A=1}^{N}\sum_{B>A}^{N}\frac{Z_A Z_B}{R_{AB}} \tag{2}$$

where ∇^2 is the Laplacian operator and Z_A is the atomic number of nucleus A. r_{iA} is the distance between electron i and nucleus A. r_{ij} is the distance between electrons i and j and R_{ab} is the distance between nuclei A and B. The first term in Eq. (2) represents the kinetic energy of the electrons. The second term represents the kinetic energy of the nuclei, the third term denotes the Coulombic attraction between the electrons and nuclei. The fourth and fifth terms are the repulsion terms between the electrons and the nuclei, respectively.

Due to the differences between the masses of the electrons and the nuclei, the Schrodinger equation can be separated into equations that govern the electrons and nuclei, according to the Born Oppenheimer approximation.[11] This means that for the electronic part of the system, the second term in Eq. (2) can be neglected and the fifth term can be considered to be constant. The Hamiltonian can be written as the electronic Hamiltonian:

$$H_{el} = -\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{A=1}^{N}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} \tag{3}$$

and the Schrodinger equation can be written as:

$$H_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{4}$$

where Ψ_{el} is the electronic wavefunction and H_{el} is the electronic Hamiltonian.

The total energy can be determined by adding the electronic energy and the nuclei-nuclei repulsion term.

Approximations must be introduced in order to solve the Schrodinger equation. The simplest approximation is the Hartree-Fock (HF) approach[12-14]. In this approach, a product of one-electron wave functions, called orbitals, is used to approximate the electronic wavefunction. For a two particle system, the wavefunction can be written in the form of a Hartree product[12-14] which is given as:

$$\Psi_{el}(e_1,e_2) \approx \psi_1(e_1)\ \psi_2(e_2) \tag{5}$$

A linear combination of Hartree product functions form the Slater determinant[15,16] that satisfies the antisymmetry principle. The electronic wavefunction is then represented as an antisymmetrized product of one-electron wavefunctions (spin orbitals, $\psi(e)$). The wavefunction in the Slater determinant form for n electrons is given below

$$\Psi(e_1,e_2,.....,e_n) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(e_1) & \psi_2(e_2) & .... & \psi_n(e_n) \\ \psi_1(e_1) & \psi_2(e_2) & .... & \psi_n(e_n) \\ \vdots & \vdots & & \vdots \\ \psi_1(e_1) & \psi_2(e_2) & .... & \psi_n(e_n) \end{vmatrix} \tag{6}$$

where $\psi_i(e_i)$ is the ith molecular spin orbital and e_i is the ith electron.

The energy is determined by using the variational principle, which leads to the HF equations,

$$\hat{F}\psi_i = \varepsilon_i\psi_i \tag{7}$$

Where $\hat{F}$ is the Fock operator, ε_i is the energy of the molecular orbital ψ_i. Since the Fock operator, $\hat{F}$, depends on molecular orbitals, ψ_i, the initial orbitals are guessed and then refined iteratively. Therefore, the HF equations (Eq. 7) must be solved iteratively using self-

consistent field (SCF) method. First, an initial guess for the wavefunction is made, followed by the formation of the Fock operator, which is then used to generate the improved wavefunction. This procedure is iterated until self-consistency is reached. The wavefunction is expressed in the form of a linear combination of atomic orbitals (LCAO) as given in Eq. (8).

$$\psi_i = \sum_{\mu} C_{\mu i} \chi_{\mu} \tag{8}$$

$C_{\mu i}$ are the MO expansion coefficients, N is the number of basis functions and χ_{μ} are the basis functions. This is the basis set approximation and as the basis set becomes larger and more complete, the HF energy reaches a complete basis set limit.

The main drawback of the HF method is that it does not account for the instantaneous electron-electron correlation (explicit electron-electron interaction) because the electrons are in the average field of all the other (N-1) electrons and the nuclei. Due to the neglect of the electron correlation, the HF method gives an upper limit to the exact energy and the difference between the exact and the HF energy is called the electron correlation energy.

In order to recover the electron correlation several methods have emerged that are termed post HF methods. such as configuration interaction singles (CIS), coupled-cluster methods (CC) and perturbation theory (PT), are summarized below.

III.A. Configuration interaction singles (CIS)

CIS is a truncated configuration interaction (CI) method[17], which considers only the single excitations from a Hartree-Fock ground state reference. The CIS wavefunction and the total energy are written as follows.

$$E_{CIS} = E_{HF} + \sum_{ia} a_{ia}^2 (\varepsilon_a - \varepsilon_i) - \sum_{ijab} a_{ia} a_{jb} (ja \| ib) \tag{9}$$

$$|\Psi_{CIS}\rangle = |\Phi_{HF}\rangle + \sum_{ia} a_{ia} |\phi_{ia}\rangle \tag{10}$$

where i and j represent the occupied molecular orbitals in the ground state; a and b represent the unoccupied molecular orbitals. The virtual and occupied orbital energies are given by ε_a and ε_i, respectively; E_{HF} and $|\Phi_{HF}\rangle$ denote the ground state Hartree Fock energy and wavefunction respectively. The $|\phi_{ia}\rangle$ represent the singly excited determinants formed by replacing orbital i in $|\Phi_{HF}\rangle$ by orbital a. a_{ia} and a_{jb} are the CIS coefficients which are unknown and are deduced as normalized eigenvectors of the Hamiltonian matrix given below.

$$H_{iajb} = \langle \phi_{ia} | H | \phi_{jb} \rangle = \left[E_{HF} + \varepsilon_a - \varepsilon_i \right] \delta_{ij} \delta_{ab} - \left(ja \| ib \right) \tag{11}$$

The diagonalization of the above matrix is accomplished by using the Davidson procedure[18] in which the Hamiltonian matrix is multiplied by a set of trial CIS vectors. The CIS coefficients (eigenvectors) and the associated CIS energies (eigenvalues), E_{CIS}, for the desired roots are then obtained in an iterative process.

III.B. Coupled-cluster methods

Coupled-cluster (CC)[19-21] methods are one of the most accurate approaches for recovering the

electron correlation energy. The CC expression for the wavefunction is given by:

$$\Psi_{CC} = e^{T}\Psi_{o} \tag{12}$$

Where Ψ_o is a HF wavefunction and $\hat{T}$ is

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + ... + \hat{T}_n \tag{13}$$

where $\hat{T}_1$, $\hat{T}_2$ ... are sums of one-particle, two-particle….excitation operators that act on the HF wavefunction. The exponential operator (e^T) is defined by the Taylor series expansion given as:

$$e^{T} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} ..... = \frac{\hat{T}^k}{k!} \tag{14}$$

The one-particle and two-particle excitation operators are defined as:t

$$\hat{T}_1 \Psi_o = \sum_{i}^{occ}\sum_{a}^{virt} t_i^a \Psi_i^a \qquad\qquad \hat{T}_2 \Psi_o = \sum_{i>j}^{occ}\sum_{a>b}^{virt} t_{ij}^{ab} \Psi_{ij}^{ab} \tag{15}$$

where, t_i^a and t_{ij}^{ab} are the coefficients (amplitudes) for various determinants.

To apply coupled-cluster theory, one approximates the operator $\hat{T}$ by truncating it or including only some of the operators in Eq. (13). For example, inclusion of only $\hat{T}_1$ and $\hat{T}_2$ in Eq. (13) gives the CCSD method, inclusion of only $\hat{T}_1$, $\hat{T}_2$ and $\hat{T}_3$ gives the CCSDT method, and so on. Although the CCSDT method generally gives very accurate results, it is very computationally demanding. Therefore one uses an approximate form of the CCSDT method called CCSD(T).[22] It is the most commonly used CC method, which includes single and double excitations and non-iterative quasi-perturbative triples that are evaluated using perturbation theory and added to the CCSD results. The CCSD(T) method generally gives very accurate energies for molecules that are near equilibrium geometries and have small

configurational mixing. CCSD(T) gives very accurate results for the correlation energy with a reasonable computational cost.

III.C. Perturbation theory

Perturbation theory is a post HF method in which the electron correlation energy is treated as a small perturbation to the total Hamiltonian of the system. The Hamiltonian of the system is

$$H = H^{(0)} + H' \tag{16}$$

$H^{(0)}$ is the zeroth order part of the Hamiltonian with known eigenvalues and eigenfunctions and H' represents a small perturbation to the system.

The basic idea of perturbation theory is to correct the solution of the unperturbed problem (for which the eigenfunctions and eigenvalues are known) that is under the influence of a small perturbation. The main assumption in perturbation theory is that H' is much smaller than $H^{(0)}$. The energy and the wavefunction can be expanded in a Taylor series and truncated after nth order, giving the nth order perturbation correction to the energy.

$$E^{MPn} = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} + ... + E_0^{(n)} \tag{17}$$

The subscript zero denotes the ground state of the system. $E^{MPn} = E_0^{(0)} + E_0^{(1)} + E_0^{(2)}$ gives the energy corrected to second order in perturbation theory. Because Møller and Plesset first showed how to formulate second order perturbation using HF orbitals, the method is often called second order Møller-Plesset perturbation theory (MP2)[23].

Perturbation theory can also be built on the MCSCF (multi-configurational self-consistent field) reference wavefunction. Then, the method is called multi-reference second order

perturbation theory (MRPT2). Implementations include second order multi-reference Møller-Plesset perturbation theory (MRMP2)[24-26], multi-configurational quasidegenerate perturbation theory (MCQDPT)[27,28] and complete active space second order perturbation theory (CASPT2)[29].

III.D. Density Functional Theory (DFT)

The Hohenberg-Kohn[30] theorem states that the ground state electronic energy of a system and other electronic properties can be uniquely determined by the electron density of the system (ρ). In wavefunction-based ab initio methods, the wavefunction of an N-electron system includes 3N variables whereas the electron density, ρ, is a function of only three coordinates (x,y,z) of electrons ($\rho(x,y,z)$). This is often taken to be an advantage of density-based methods. The energy is then a function of a function, which means the energy is a functional of the electron density ($F[\rho(x,y,z)]$). If the ground state electron density is known, one can calculate the ground state energy (E_{DFT}) and the ground state molecular properties.

$$E_{DFT} = F[\rho(x,y,z)] \tag{18}$$

The main challenge in the DFT method is that the exact forms of the density, and therefore of the kinetic energy functional and the exchange and correlation functionals are not known. In general, the approximate functionals used in chemistry are of four types: LDA (local density approximation) which assumes the density of the molecule to be uniform throughout the molecule and in which the energy depends only on the electron density. The second type is called GGA (generalized gradient approximation) method, which accounts for the non-uniformity of the electron density in the molecule by including both the electron density and

its gradient in the functionals. The third type of DFT functional, called a hybrid functional, incorporates some Hartree-Fock exchange into a GGA. An example is the popular B3LYP (Becke three-parameter Lee-Yang-Parr)[31,32] functional. The fourth type of functional, the meta-GGA, incorporates the second derivative of the density indirectly through the kinetic energy density.[33,34]

III.E. Effective Fragment Potential (EFP)

The effective fragment potential (EFP)[1] is a discrete method that accounts for solvent molecules explicitly and enables one to capture the solute-solvent and solvent-solvent interactions. There are two EFP methods: EFP1 and EFP2. The EFP1 method was developed specifically for water.[1,2,35,36] There are three interaction terms in EFP1 that are added to the solute Hamiltonian: (a) Coulomb or electrostatic interactions between solute-solvent (QM-EFP1) and solvent-solvent (EFP1-EFP1) are represented by a distributed multipole analysis (DMA)[37] expanded through octopoles. A damping term is used to account for the overlapping charge densities when the interacting solute-solvent or solvent-solvent molecules are close to each other. (b) The polarization interaction term is treated using a finite field dipole-induced dipole model interaction; in which the interaction is iterated to self-consistency. (c) The third term is called the charge transfer/exchange repulsion that is the remainder term fitted to a set of HF water dimer calculations and accounts for the interactions that are not included in the first two terms. The formulation of the EFP1 method for a water molecule μ and a QM coordinate s is:

$$E_{\text{int}eraction} = E_{Coulomb} + E_{polarization} + E_{exchange\ repulsion/charge\ tranfer} \tag{19}$$

$$E_{\text{int}eraction} = \sum_{k=1}^{K} V_k^{Elec}(\mu,s) + \sum_{l=1}^{L} V_l^{Pol}(\mu,s) + \sum_{m=1}^{M} V_m^{\text{Re}\,p}(\mu,s) \tag{20}$$

The EFP1 method described above has been implemented for HF and DFT levels of theory[38].

A more general EFP method, that has no empirically fitted parameters, called EFP2[2] may be generated for any system. The EFP2-EFP2 interaction terms include Coulomb, polarization, exchange repulsion, charge transfer [33]and dispersion.[36]

$$E_{\text{int}eraction} = E_{Coulomb} + E_{polarization} + E_{exchange\ repulsion} + E_{\text{dispersion}} + E_{charge\ tranfer} \tag{21}$$

Both single-point energy and gradient calculations for EFP2-EFP2 interaction terms have been implemented in GAMESS. However, the EFP2-QM exchange repulsion gradient implementation and the EFP2-QM dispersion terms are still under construction. So, at present, the EFP2-QM is not generally useful and has not been used in this dissertation.

III.F. Hybrid QM/MM method SIMOMM

The surface integrated molecular orbital molecular mechanics (SIMOMM)[3] method is an embedded cluster hybrid quantum mechanics (QM) and molecular mechanics (MM) approach to represent surfaces. In this approach, the system is divided into two parts; one part is a small chemically active site (where the bond-breaking or formation occurs), which can be treated with any appropriate level of QM. The remainder of the system is a larger, relatively inactive site called the bulk region, which is represented by a MM method**.** When the active site is cut out of the bulk region, hydrogen atoms are used as placeholders in the

interfacial region (boundary region) between the QM and MM regions to cap the dangling bonds. These are usually called "capping" hydrogens. During the MM part of the calculation, these capping hydrogens are replaced by the actual MM atoms. In the SIMOMM energy calculations, first the MM energy of the bulk region, E_{MM}, is calculated and the interactions of the MM part with the active region are not included. The second step is to calculate the energy of the active region, E_{QM}, using an appropriate level of quantum mechanics, with the capping hydrogens in place. The long-range interactions between the QM and MM parts are accounted for by the MM calculations. The SIMOMM total energy, E_T, is given in Eq. (22)

$$E_T = E_{QM} + E_{MM} + E_{QM\text{-}MM} \tag{22}$$

The total gradient is calculated using the following formula.

$$\frac{dE_T}{d\vec{R}_{active}} = \frac{\partial E_{QM}}{d\vec{R}_{active}} + \frac{\partial E_{MM}}{d\vec{R}_{active}} \tag{23}$$

$$\frac{dE_T}{d\vec{R}_{boundary}} = \frac{\partial E_{QM}}{d\vec{R}_{boundary}} \tag{24}$$

$$\frac{dE_T}{d\vec{R}_{bulk}} = \frac{\partial E_{MM}}{d\vec{R}_{bulk}} \tag{25}$$

R_{active}, $R_{boundary}$, R_{bulk} refer to the coordinates of the active, boundary and bulk region atoms respectively.

References

(1) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *Journal of Chemical Physics* **1996**, *105*, 1968.
(2) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *Journal of Physical Chemistry A* **2001**, *105*, 293.
(3) Shoemaker, J. R.; Burggraf, L. W.; Gordon, M. S. *Journal of Physical Chemistry A* **1999**, *103*, 3245.
(4) Schrodinger, E. *Annalen Der Physik* **1926**, *81*.
(5) Schrodinger, E. *Annalen Der Physik* **1926**, *79*, 361.

(6) Schrodinger, E. *Annalen Der Physik* **1926**, *79*, 734.
(7) Schrodinger, E. *Annalen Der Physik* **1926**, *79*, 489.
(8) Schrodinger, E. *Annalen Der Physik* **1926**, *80*, 437.
(9) Schrodinger, E. *Naturwissenschaften* **1926**, *14*, 664.
(10) Schrodinger, E. *Physical Review* **1926**, *28*, 1049.
(11) Born, M.; Oppenheimer, R. *Annalen Der Physik* **1927**, *84*, 0457.
(12) Hartree, D. R. *Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 426.
(13) Hartree, D. R. *Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 111.
(14) Hartree, D. R. *Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 89.
(15) Slater, J. C. *Physical Review* **1930**, *35*, 0509.
(16) Slater, J. C. *Physical Review* **1929**, *34*, 1293.
(17) Foresman, J. B.; Headgordon, M.; Pople, J. A.; Frisch, M. J. *Journal of Physical Chemistry* **1992**, *96*, 135.
(18) Davidson, E. R. *Journal of Computational Physics* **1975**, *17*, 87.
(19) Paldus, J. *NATO ASI Series, Series B: Physics* **1992**, *293*, 99.
(20) Barlett, R. J. *Advanced Series in Physical Chemistry* **1995**, *2*, 1047.
(21) Lee, T. J.; Scuseria, G. E. *Understanding Chemical Reactivity* **1995**, *13*, 47.
(22) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Computer Physics Communications* **2002**, *149*, 71.
(23) Moller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 0618.
(24) Hirao, K. *Chemical Physics Letters* **1992**, *190*, 374.
(25) Hirao, K. *Chemical Physics Letters* **1992**, *196*, 397.
(26) Hirao, K. *International Journal of Quantum Chemistry* **1992**, 517.
(27) Nakano, H. *Journal of Chemical Physics* **1993**, *99*, 7983.
(28) Nakano, H. *Chemical Physics Letters* **1993**, *207*, 372.
(29) Andersson, K.; Malmqvist, P. A.; Roos, B. O. *Journal of Chemical Physics* **1992**, *96*, 1218.
(30) Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *136*, B864.
(31) Becke, A. D. *Physical Review A* **1988**, *38*, 3098.
(32) LEE, C. T.; Yang, W. T.; Parr, R. G. *Physical Review B* **1988**, *37*, 785.
(33) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Physical Review Letters* **1999**, *82*, 2544.
(34) Van Voorhis, T.; Scuseria, G. E. *Journal of Chemical Physics* **1998**, *109*, 400.
(35) Jensen, J. H.; Gordon, M. S. *Journal of Chemical Physics* **1998**, *108*, 4772.
(36) Adamovic, I.; Gordon, M. S. *Molecular Physics* **2005**, *103*, 379.
(37) Stone, A. J. *Chemical Physics Letters* **1981**, *83*, 233.
(38) Adamovic, I.; Freitag, M. A.; Gordon, M. S. *Journal of Chemical Physics* **2003**, *118*, 6725.

CHAPTER 2: SOLVENT INDUCED FREQUENCY SHIFTS: CONFIGURATION INTERACTION SINGLES COMBINED WITH THE EFFECTIVE FRAGMENT POTENTIAL METHOD

A paper submitted to the Journal of Physical Chemistry

Pooja Arora, Lyudmila V. Slipchenko, Simon P. Webb, Albert DeFusco, and Mark S. Gordon

Abstract

The simplest variational method for treating electronic excited states, configuration interaction with single excitations (CIS), has been interfaced with the effective fragment potential (EFP) method to provide an effective and computationally efficient approach for studying the qualitative effects of solvents on the electronic spectra of molecules. Three different approaches for interfacing a non-self-consistent field (SCF) excited state quantum mechanics (QM) method and the EFP method are discussed. The most sophisticated and complex approach (termed "Fully Self Consistent") calculates the excited state electron density with fully self-consistent accounting for the polarization (induction) energy of effective fragments. The simplest approach ("method 1") includes a strategy that indirectly adds the EFP perturbation to the CIS wavefunction and energy via modified Hartree-Fock molecular orbitals, so that there is no direct EFP interaction with the excited state density. An intermediate approach ("method 2") accomplishes the latter in a non-iterative perturbative manner. Theoretical descriptions of the three approaches are presented, and test results of solvent induced shifts using methods 1 and 2 are compared with fully ab initio values. These

comparisons illustrate that, at least for the test cases examined here, modification of the ground state Hartree-Fock orbitals is the largest and most important factor in the calculated solvent-induced shifts. Method 1 is then employed to study the aqueous solvation of coumarin 151 and compared with experimental measurements.

I. Introduction

The ability to predict and understand solvent induced shifts in electronic spectra is of great importance in chemical, biological and medicinal sciences[1-4]. Solvent induced shifts in electronic spectra result from two sources: (a) intrinsic differences in the solute due to interactions with the field produced by solvent molecules and (b) differences between ground and excited state solute-solvent interactions due to modifications of the solute electron density by the surrounding solvent molecules. It has been demonstrated that excited state properties of some solutes, such as coumarins, are very sensitive to interactions with surrounding solvent molecules.[5-8] Therefore, coumarins are widely used as a tool to investigate the solute-solvent interactions and solvation dynamics.[4,9-12]. For example, recently, coumarin 153 has been used as a probe to study solvent dynamics in proteins by time-dependent fluorescence Stokes shift[13] measurements[14]. In order to understand the effects of polarity and the H-bonding of the solvents on the electronic spectrum of coumarin 120, a study by Zhao et al using TDDFT[15,16] (time dependent density functional theory) with the PCM[17] (polarizable continuum model) for solvents was performed[8]. By including the explicit solvent molecules that are hydrogen bonded to the solute, these authors predicted that

the intermolecular solute-solvent and solvent-solvent interactions at the microscopic level affect the transition energies of coumarin 120 in aqueous solution.

A study by Karelson et al on solvent induced shifts indicates that two explicit water molecules forming H-bonds with pyrimidine are necessary to accurately predict its solvent shift.[18] Similar conclusions were reached by Cave et al who, using TDDFT combined with a dielectric continuum solvent model, showed that the excitation energies of coumarin 120 and coumarin 151 are overestimated compared to the experiments due to the lack of explicit descriptions of solute-solvent interactions in the solvent model.[19] Therefore, it is important to reliably model the solvent molecules and their impact on electronic spectra in order to capture the correct solvent induced shifts in the absorption spectra.

Theoretical investigations of absorption spectra in the condensed phase are limited by difficulties in accurately incorporating the solvent environment into the quantum treatment of a solute system of interest. The treatment of the solvent molecules using ab initio methods would capture the solvent effects most accurately, but such treatments are limited by their computational demands. There have been some methodological advances for the study of condensed phase electronic spectroscopy, especially using dielectric continuum methods to represent the solvents[18,20-27]. While continuum methods are computationally inexpensive, they cannot describe explicit solute-solvent interactions such as hydrogen bonding. In other words, the microscopic structure around the solute molecules is not adequately described by the implicit solvent methods. On the other hand, discrete solvent methods treat each solvent molecule explicitly and the bulk behavior can also be described by using molecular dynamics and Monte Carlo simulation techniques. The disadvantage of the explicit solvent models is that they are dependent on the quality of the model potential and on the sampling of the

configurational space. The latter is usually required to be extensive and therefore computationally demanding.

The present work introduces a discrete approach for analyzing solvent effects on electronic spectra, in which the singly excited configuration interaction (CIS) method is combined with the effective fragment potential (EFP)[28,29] method. The EFP, a model potential that is largely based on first principles, has been demonstrated to accurately reproduce the effects of solvents on electronically excited states. Yoo et al combined the TDDFT method for excited states with EFP to study the optical properties of molecules in the condensed phase[30]. This TDDFT/EFP investigation successfully reproduced the experimentally observed solvent induced shifts of acetone. While the present work focuses on the simple CIS method, the strategies that are presented here are relevant for most ab initio excited state methods.

Several theoretical methods are routinely employed for determining excited state properties in the gas and condensed phases. These include CIS[31-33], TDDFT[15,16,32], complete active space self-consistent field (CASSCF)[34-37], configuration interaction with single and double excitations (CISD)[38], symmetry adapted cluster - configuration interaction (SAC-CI)[39,40], equation-of-motion couple-cluster (EOM-CC)[41-43], multi-reference perturbation theory (MRPT)[44,45] and multi-reference CI (MRCI)[46,47]. CIS is the simplest and least computationally demanding of these methods. It provides a qualitatively correct characterization of excited states that are dominated by single excitations.

There are two versions of the EFP method. EFP1 is specific to water and has been implemented for Hartree-Fock (HF) and density functional theory (DFT). EFP2 is a more general method that has not yet been fully interfaced with ab initio methods. The focus of the

present work is to combine CIS with the EFP1/HF method to calculate solvent-induced shifts in the UV spectra of solute molecules. Three approaches are considered, all of which are relevant as well to most of the excited state methods mentioned in the previous paragraph and to EFP1/DFT as well. Only minor modifications will be required for the more general EFP2 method[28,48,49], once that method has been fully interfaced with ab initio methodology.

An important contribution to solvent-induced spectral shifts is the induction (polarization) energy, because the electron density of the solute changes upon excitation, and the polarization of the solvent will respond to the altered electron density in its excited state. Some studies have considered the importance of mutual solute-solvent polarization between solute and solvent molecules on the excitation energy [46,50,51]. Xu et al examined the effects of solute polarization of the n-π* transition of formaldehyde in the condensed phase using a QM/MM method that combines MRCI and molecular dynamics simulations using a classical force field[46]. They found that the solute polarizability is an important component of solvent induced shifts of formaldehyde, contributing about 35% of the shift in the calculated excitation energy. Aidas et al have found that the inclusion of explicit polarization due to solvent molecules in combined quantum mechanics/molecular mechanics (QM/MM) calculations of excited states slightly lowers the excitation energy[50]. Studies by Kongsted et al that combine the coupled-cluster method with an MM method also predicted that neglecting the polarization effects overestimates the excitation energies.[51]

Several studies on electronic spectroscopy in the condensed phase have combined a QM method with explicit solvent models that incorporate polarization effects. Luzhkov et al developed a hybrid QM/MM method to study the solvent effects on electronic spectra using a Langevin dipole[52] solvent model.[53] Another study by Thompson et al described a QM

method combined with a polarizable MM method to study excited states.[54] Gao et al implemented a combined QM - polarizable MM potential approach for excited states to examine the solvent effects on pyrimidine.[55] They used a semi-empirical method for the solute and a classical model for the solvent molecules. Karelson and coworkers successfully extended the SCRF (self-consistent reaction field)[56] implicit model to study solvent effects on excited states including solvent polarization[24]. As outlined in the next section, a key feature of the EFP method is that it includes a solvent polarization term that is iterated to self-consistency within the quantum mechanical part of the calculation. The EFP solvent model poses an advantage over continuum models by being an explicit and polarizable solvent model that can describe the instantaneous electronic response of the solvent molecules for electronic excitations.

In the present study, three approaches to the CIS/EFP1 QM/MM method are presented. In the most sophisticated and complex approach (termed "Fully Self Consistent") the polarization (induction) term of EFP is fully iterated to be self-consistent with the excited state wavefunction. This method provides an excited state that is fully consistent with the environment, by calculating the response of the environment according to the electron density of the excited state, within the CIS iterative (Davidson diagonalization[57]) procedure. The second approach ("method 1") is the simplest way to indirectly add the polarization perturbation to the excited state via modified HF orbitals. The third approach ("method 2") is a compromise between the first two, in which the calculation of the excited state solvent response within the CIS iterative procedure is avoided by employing a one-time perturbative correction term that estimates the solvent response for the excited state density.

The outline of this paper is as follows. In Section II, the methodological details of the three approaches for interfacing the CIS and EFP1/HF methods are described. Section III describes the computational details. Section IV benchmarks and illustrates the accuracy of the CIS/EFP1 schemes for several small molecules. Applications of the CIS/EFP1 interface to acetone and coumarin 151 using a molecular dynamics simulation are also presented. Conclusions are drawn in Section V.

II. Theory

A. Summary of the EFP method. The effective fragment potential method has been described in many previous papers,[28,29] so it is only briefly summarized here. The version of the EFP method used in the present work is based in part on the Hartree-Fock (HF) method and is referred to as EFP1/HF. However, the entire discussion is equally applicable to the analogous method that was derived from DFT, EFP1/DFT.[58] EFP1/HF contains three terms that describe solute-solvent and solvent-solvent Columbic, induction and exchange repulsion interactions. Coulomb (electrostatic) interactions in the EFP method are represented by a distributed multipole analysis (DMA) up through octopoles [59]. The electrostatic EFP contribution to the QM Hamiltonian is[28]

$$V_k^{es}(m,s) = \sum_{k=1}^{N_k}\left[-\frac{q_k}{r_k} - \sum_{a}^{x,y,z}\mu_a^k \hat{F}_a(r_k) - \frac{1}{3}\sum_{ab}^{x,y,z}\Theta_{ab}^k \hat{F}_{ab}(r_k) - \frac{1}{15}\sum_{abc}^{x,y,z}\Omega_{abc}^k \hat{F}_{abc}(r_k)\right] \quad (1)$$

where μ, Θ and Ω are the EFP dipole, quadrupole and octopole moments respectively. $\hat{F}_a$, $\hat{F}_{ab}$ and $\hat{F}_{abc}$ are the solute electric field, field gradient and second derivative field operators respectively. N_k is the total number of EFP multipole expansion points. A damping term is

used to account for the overlapping charge densities when the solute-solvent or solvent-solvent molecules are very close to each other.[60]

The polarization (induction) term is treated using a finite field dipole-induced dipole model; in which the interaction is iterated to self-consistency. In the EFP approach, the polarizability tensor is expressed in terms of individual localized molecular orbital (LMO) tensors for each LMO in the molecule; for example, two bond LMOs and two lone pair LMOs in water. The polarization/induction contribution to the QM Hamiltonian is given in Eq. (2). F is the field due to the ab initio part of the system and $\tilde{\alpha}^{l}_{ab}(m)$ is the polarizability component of the mth fragment in the lth localized orbital; a, b run over the x, y, z coordinates.

$$\mathrm{V}^{\mathrm{pol}}_{l} = -\frac{1}{2}\sum_{a,b}^{x,y,z} F_a(r_1)\tilde{\alpha}^{l}_{ab}(m)\left\langle F_b(r_1)\right\rangle \tag{2}$$

The third EFP term is a remainder term that accounts for all interactions that are not accounted for in the first two terms. At the Hartee-Fock (HF) level of theory, these are the exchange repulsion + charge transfer interactions. For QM-EFP interactions, the remainder term is expanded in terms of Gaussian functions:

$$\mathrm{V}^{\mathrm{rem}}_{m} = \sum_{j}\beta_{m,j}\exp(-\alpha_{m,j}r_m^2) \tag{3}$$

where m refers to a fragment center for the exchange repulsion potential. The expansion points are the atomic centers and the center of mass. The parameter β is generally set equal to unity, and the expansion includes only one term. The parameters α are obtained by evaluating the HF water dimer potential at many points, subtracting the Coulomb and induction interactions, and fitting the QM-EFP remainder term to the remaining HF

interaction[28]. The corresponding EFP-EFP remainder interaction is obtained in a similar manner, except that exponential functions are used, rather than Gaussians.

B. EFP-QM Interface. In the present study, the solvent effects are treated using the effective fragment potential (EFP) method. The aim is to calculate the vertical excitations, with and without the presence of solvent molecules, at the optimized ground state geometry, in order to assess the affects of the solvent on the calculated excitation energies. Because of the dependence of the EFP induction interaction on the solute electron density, the changing electron density of the solute upon electronic excitation must be accounted for by iterating the dipole-induced dipole interaction to self-consistency. In order to accomplish this, three approaches have been developed, as summarized in the following paragraphs.

Fully Self Consistent method: This method is the most rigorous approach for combining the EFP1/HF method with a QM method for excited states. It involves a coupled iterative procedure that solves both the solute wavefunction (represented by CIS) and the solvent induced dipoles (represented by EFP1/HF) to obtain an excited state that is fully consistent with the environment. Since this method adds the polarization perturbation in a self-consistent manner within the CIS Davidson diagonalization iterative procedure, it is the most accurate and complete description of the inclusion of polarization perturbation in the excited state energy.

The total Hamiltonian of the excited state system can be written as:

$$H^{\mathrm{EX}} = H_o^{\mathrm{EX}} + H_{pol}^{\mathrm{EX}} \tag{4}$$

The superscript EX represents the excited state.

H_{pol}^{EX} in Eq. (4) is the EFP1 polarization interaction term. H_{o}^{EX} is

$$H_{o}^{\mathrm{EX}} = H_{oo}^{\mathrm{EX}} + H_{es}^{\mathrm{EX}} + H_{rem}^{\mathrm{EX}} \tag{5}$$

H_{oo}^{EX} in Eq. (5) is the gas-phase time-independent QM Hamiltonian of the system. H_{es}^{EX} represents the electrostatic interaction term and H_{rem}^{EX} is the exchange repulsion + charge transfer EFP interaction term.

The total excited state energy of the system is given as follows:

$$E^{EX} = \langle \psi_{CIS} | H_{o}^{\mathrm{EX}} + H_{pol}^{\mathrm{EX}} | \psi_{CIS} \rangle \tag{6}$$

where ψ_{CIS} is the CIS wavefunction.

The polarization/induction interaction must be iterated to self-consistency. In order to derive the polarization contribution to the excited state energy, consider the polarization energy expression in terms of induced dipoles that is analogous to the ground state expression obtained by Day et al, [29]

$$E_{pol}^{EX} = -\frac{1}{2}\sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{\mathrm{EX}} - \vec{\mu}_{\mathrm{i}}^{\mathrm{EX}'}) \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} \tag{7}$$

The quantities in Eq. (7) are defined as follows: $\vec{\mu}_{\mathrm{i}}^{EX}$ is the total induced dipole moment vector at the polarizable point i in the EFP. The polarizable points are taken to be the centroids of the localized molecular orbitals in the effective fragment. The induced dipole moments may be written in terms of the polarizability $\tilde{\alpha}_i$ at polarizable point i and the total field $\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}}$:

$$\vec{\mu}_{i}^{EX} = \tilde{\alpha}_i \, \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}}$$

$\vec{\mu}_i^{EX'}$ in Eq. (7) is the dipole moment induced by the field of the induced dipoles and is written as

$$\mu_i^{EX'} = \tilde{\alpha}_i^{\mathrm{T}}\ \mathrm{F}_{\mathrm{i}}^{\mu,\mathrm{EX}}$$

where $\tilde{\alpha}_i^{\mathrm{T}}$ is the transpose of the polarizability tensor. The total field $\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}}$ in Eq. (7) at the polarizable point i contains four components and is written as

$$\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} = \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{nuc}} + \langle \psi_{CIS} | \vec{f}_i^{el} | \psi_{CIS} \rangle + \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{efp}} + \vec{\mathrm{F}}_{\mathrm{i}}^{\mu.EX} \tag{8}$$

where, $\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{nuc}}$ and $\langle \psi_{CIS} | \vec{f}_i^{el} | \psi_{CIS} \rangle$ are the fields from the QM nuclei and from the electrons, respectively

$\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{efp}}$ represents the field due to the static multipoles on the EFP fragments and $\vec{\mathrm{F}}_{\mathrm{i}}^{\mu.EX}$ is the field vector from induced dipoles on the EFP fragments.

The contribution from the polarization energy in the excited state can be obtained using the variational method. The functional can be formed as:

$$\mathrm{L} = E^{EX} - \mathrm{W}^{\mathrm{EX}}(\langle \psi_{CIS} | \psi_{CIS} \rangle - 1) \tag{9}$$

where W^{EX} is the Lagrange multiplier due to the normalization constraint. Also, W^{EX} represents the energy that is obtained directly from quantum mechanics and contains the polarization contribution.

Eqs. (6) - (9) lead to,

$$\mathrm{L} = E_o^{EX} - \frac{1}{2}\sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_i^{EX'}) \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{totEX}} - \mathrm{W}^{\mathrm{EX}}(\langle \psi_{CIS} | \psi_{CIS} \rangle - 1) \tag{10}$$

where $E_o^{EX} = \langle \psi_{CIS} | H_o^{\mathrm{EX}} | \psi_{CIS} \rangle$

$$\mathrm{L} = \mathrm{E}_{\mathrm{o}}^{\mathrm{EX}} - \frac{1}{2}\sum_{\mathrm{i}}(\tilde{\alpha}_i \ \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} + \frac{1}{2}\sum_{\mathrm{i}}(\tilde{\alpha}_i^{\mathrm{T}} \ \vec{\mathrm{F}}_{\mathrm{i}}^{\mu,\mathrm{EX}})\ \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} - \mathrm{W}^{\mathrm{EX}}(\langle\psi_{CIS}|\psi_{CIS}\rangle \qquad (11)$$

Recall that H_o^{EX} contains the contributions from the EFP Coulomb and remainder interactions. Variation of Eq. (11) with respect to the wavefunction parameters gives the following:

$$\delta\mathrm{L} = \delta\mathrm{E}_{\mathrm{o}}^{\mathrm{EX}} - \frac{1}{2}\sum_{\mathrm{i}}[(\tilde{\alpha}_i + \tilde{\alpha}_i^{\mathrm{T}})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} - \tilde{\alpha}_i^{\mathrm{T}} \ \vec{\mathrm{F}}_{\mathrm{i}}^{\mu,\mathrm{EX}}]\ \delta\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle - \mathrm{W}^{\mathrm{EX}}\ \delta\langle\psi_{CIS}|\psi_{CIS}\rangle$$

(12)

Applying the condition $\delta\mathrm{L} = 0$ to Eq. (12) gives:

$$\langle\delta\psi_{CIS}|H_o^{\mathrm{EX}}|\psi_{CIS}\rangle - \frac{1}{2}\sum_{\mathrm{i}}[(\vec{\mu}_{\mathrm{i}}^{EX} + \vec{\mu}_{\mathrm{i}}^{EX^{\pm}} - \vec{\mu}_i^{EX'})]\langle\delta\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle - \mathrm{W}^{\mathrm{EX}}(\langle\delta\psi_{CIS}|\psi_{CIS}\rangle) + \mathrm{cc} = 0 \quad (13)$$

where cc stands for complex conjugate and $\vec{\mu}_i^{EX^{\pm}} = \tilde{\alpha}_{\mathrm{i}}^{\mathrm{T}} \ \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}}$

$\vec{\mu}_{\mathrm{i}}^{EX}$ and $\vec{\mu}_{\mathrm{i}}^{EX^{\pm}}$ in Eq. (13) are equal if the polarizability tensor is symmetric ($\tilde{\alpha}_i^{\mathrm{T}} = \tilde{\alpha}_i$).

W^{EX} is:

$$\mathrm{W}^{\mathrm{EX}} = \langle\psi_{CIS}|H_o^{\mathrm{EX}} - \frac{1}{2}\sum_{\mathrm{i}}([\vec{\mu}_{\mathrm{i}}^{EX} + \vec{\mu}_{\mathrm{i}}^{EX^{\pm}}]\vec{f}_i^{\,el}) + \frac{1}{2}\sum\vec{\mu}_i^{EX'}\vec{f}_i^{\,el}|\psi_{CIS}\rangle \qquad (14)$$

$$\mathrm{W}^{\mathrm{EX}} = E_o^{\mathrm{EX}} - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} + \vec{\mu}_{\mathrm{i}}^{EX^{\pm}} - \vec{\mu}_i^{EX'})\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle \qquad (15)$$

Eq. (14) represents the Hamiltonian matrix containing the contribution from the polarization perturbation in the form of induced dipoles. In order to obtain the final converged CIS eigenvalues and eigenvectors, one of two approaches could be employed: (a) An ideal way to add the polarization perturbation in the CIS Hamiltonian would be to use the relaxed excited

state density (expectation value density + non-Hellman Feynman contribution)[61]. This would require solving multiple coupled-perturbed Hartree-Fock (CPHF) equations[62,63] to iterate on the relaxed excited state density and the corresponding induced dipoles. (b) Alternatively, one could form the non-relaxed excited state density (ignoring the non-Hellman Feynman term) and the corresponding induced dipoles and iterate only within the Davidson diagonalization procedure to obtain the final eigenvectors and eigenvalues. Obtaining the excited state that is fully consistent with the environment using either of these two approaches is likely to be computationally demanding.[64]

The variational procedure does not produce the complete energy; therefore, the total energy is:

$$E^{\mathrm{EX}} = E_o^{\mathrm{EX}} + E_{pol}^{\mathrm{EX}} \tag{16}$$

Substituting E_o^{EX} from Eq. (15) and E_{pol}^{EX} from Eq. (7) to form E^{EX}:

$$E^{\mathrm{EX}} = \mathrm{W}^{\mathrm{EX}} + \left(\frac{1}{2} \sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{EX} + \vec{\mu}_{\mathrm{i}}^{EX^{\pm}} - \vec{\mu}_i^{EX'}) \langle \psi_{CIS} | \vec{f}_i^{el} | \psi_{CIS} \rangle - \frac{1}{2} \sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_i^{EX'}) \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot.EX}} \right) \tag{17}$$

The second term in Eq. (17) is the correction term to the quantum mechanical energy, W^{EX}, that is necessary in order to obtain the correct total energy.

The ground state energy obtained using the same procedure as in Eqs. (6) - (17) is

$$E^{\mathrm{G}} = \mathrm{W}^{\mathrm{G}} + \left(\frac{1}{2} \sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{G} + \vec{\mu}_{\mathrm{i}}^{G^{\pm}} - \vec{\mu}_i^{G'}) \langle \psi | \vec{f}_i^{el} | \psi \rangle - \frac{1}{2} \sum_{\mathrm{i}} (\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_i^{G'}) \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot.G}} \right) \tag{18}$$

where the superscript G represents the ground state and $\vec{\mu}_{\mathrm{i}}^{G}$, $\vec{\mu}_{\mathrm{i}}^{G^{\pm}}$ and $\vec{\mu}_i^{G'}$ are the ground state induced dipoles. ψ is the ground state wavefunction. W^{G} is the quantum mechanical

energy of the ground state that contains the polarization contribution from EFP and is obtained via the variational procedure.

The transition energy for the Fully Self Consistent method can be obtained as

$$\Delta E = E^{\mathrm{EX}} - E^{\mathrm{G}} \tag{19}$$

$$\begin{aligned}\Delta E = \mathrm{W}^{\mathrm{EX}} + \left(\frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} + \vec{\mu}_{\mathrm{i}}^{EX^{\pm}} - \vec{\mu}_{i}^{EX'})\langle\psi_{CIS}|\vec{f}_{i}^{\,el}|\psi_{CIS}\rangle - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_{i}^{EX'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot.EX}}\right) - \\ \mathrm{W}^{\mathrm{G}} - \left(\frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} + \vec{\mu}_{\mathrm{i}}^{G^{\pm}} - \vec{\mu}_{i}^{G'})\langle\psi|\vec{f}_{i}^{\,el}|\psi\rangle - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{i}^{G'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot.G}}\right)\end{aligned} \tag{20}$$

Eq. (17) is the final equation for the excited state energy for the Fully Self Consistent method that includes iterating excited state induced dipoles to self-consistency within the CIS iterative procedure. This procedure, while possibly tractable for CIS, would be very time consuming for complex methods such as multi-reference CI, multi-reference perturbation theory, CISD, or EOM-CC[41]. Now, consider two approximations to the Fully Self Consistent method.

Method 1: The simplest approach is to include the polarization effect due to the EFP solvent molecules only in the ground state HF orbitals. The excited states are then altered because the HF orbitals have been modified and these modified MOs in turn alter the CI coefficients. There is no direct modification of the CI coefficients.

The total excited state energy for method 1 including the EFP1 perturbation is given as

$$E_{1}^{EX} = \langle\psi_{CIS}|H_{\mathrm{o}}^{\mathrm{EX}} + H_{\mathrm{pol}}^{\mathrm{G}}|\psi_{CIS}\rangle \tag{21}$$

The numerical subscript in Eq. (21) and in equations hereafter, indicates the method number (1 or 2).

$H_{\mathrm{o}}^{\mathrm{EX}}$ is the Hamiltonian operator for the ab initio part plus the contribution from the EFP1 exchange repulsion + charge transfer and electrostatic terms.

Similar to the Fully Self Consistent method, the variational procedure is applied to method 1, and the functional is:

$$\mathrm{L} = E_1^{EX} - W_1^{EX}(\langle\psi_{CIS}|\psi_{CIS}\rangle - 1) \tag{22}$$

Following the same procedure as in Eqs. (6) - (17), the QM energy containing the contribution from the polarization perturbation can be obtained as:

$$W_1^{EX} = \langle\psi_{CIS}|(H_o^{EX} - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_i^{G'}).\vec{f}_i^{\,el})|\psi_{CIS}\rangle \tag{23}$$

$$W_1^{EX} = E_o^{EX} - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{\mathrm{G}} - \vec{\mu}_i^{G'})\ \langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle \tag{24}$$

The total excited state energy is then:

$$E_1^{EX} = E_o^{EX} + E_{pol}^{G} \tag{25}$$

where the polarization energy, E_{pol}^{G} , as derived for the ground state, is:

$$E_{pol}^{G} = -\frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{\mathrm{G}} - \vec{\mu}_{\mathrm{i}}^{\mathrm{G}'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,G}} \tag{26}$$

$$E_1^{EX} = W_1^{EX} + \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{\mathrm{G}} - \vec{\mu}_i^{G'})\ \langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{\mathrm{G}} - \vec{\mu}_i^{G'})\ \vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,G}} \tag{27}$$

Eq. (27) is the final working equation for total excited state energy for method 1.

The transition energy is:

$$\Delta E = E_1^{EX} - E^G$$

(28)

Using Eq. (27) and Eq. (18), one obtains the following for the excitation energy:

$$\Delta E = \left(W_1^{EX} + \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} - \vec{\mu}_i^{\,G'}) \ \langle \psi_{CIS} | \vec{f}_i^{\,el} | \psi_{CIS} \rangle - \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} - \vec{\mu}_i^{\,G'}) \ \vec{\mathrm{F}}_i^{\mathrm{tot,G}} \right) - \left(\mathrm{W}^{\mathrm{G}} + \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} + \vec{\mu}_i^{\,G^{\pm}} - \vec{\mu}_i^{\,G'}) \ \langle \psi | \vec{f}_i^{\,el} | \psi \rangle - \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} - \vec{\mu}_i^{\,G'}) \ \vec{\mathrm{F}}_i^{\mathrm{tot,G}} \right) \quad (29)$$

$$\Delta E = \left(W_1^{EX} + \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} - \vec{\mu}_i^{\,G'}) \ \langle \psi_{CIS} | \vec{f}_i^{\,el} | \psi_{CIS} \rangle \right) - \left(\mathrm{W}^{\mathrm{G}} + \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} + \vec{\mu}_i^{\,G^{\pm}} - \vec{\mu}_i^{\,G'}) \ \langle \psi | \vec{f}_i^{\,el} | \psi \rangle \right) \quad (30)$$

Method 2: In method 2, an excited state energy is obtained by including the ground state solvent-induced dipole terms as described in the formulation of method 1; then, once the iterative process is complete, a one-time correction term is added to account for the solvent response for the excited state density. This is accomplished by re-calculating the induced dipoles corresponding to the excited state density.

The method 1 Eqs. (21) - (24) also apply to method 2. They lead to the method 2 quantum mechanical energy in analogy with Eq. (24):

$$\mathrm{W}_2^{\mathrm{EX}} = E_o^{\mathrm{EX}} - \frac{1}{2}\sum_i (\vec{\mu}_i^{\,G} - \vec{\mu}_i^{\,G'}) \ \langle \psi_{CIS} | \vec{f}_i^{\,el} | \psi_{CIS} \rangle \quad (31)$$

where $\mathrm{W}_2^{\mathrm{EX}}$ is the method 2 quantum mechanical energy.

Since the application of the variational method does not provide the complete polarization energy contribution, the total energy of the hybrid QM/MM system is (see Eq. (25).

$$E_2^{\mathrm{EX}} = E_o^{\mathrm{EX}} + E_{pol}^{EX} \tag{32}$$

where E_{pol}^{EX} is the polarization energy for the excited state (Eq. (7)). So (see Eq. (27)),

$$E_2^{\mathrm{EX}} = \mathrm{W}_2^{\mathrm{EX}} + \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{\mathrm{G}} - \vec{\mu}_i^{G'})\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_i^{EX'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} \tag{33}$$

Eq. (33) contains the explicit polarization response to the excited state density unlike method 1 in which no explicit polarization response is added to the excited state.

The solvated QM excited state energy obtained using method 1 contains the ground state induced dipole terms (Eq. (31)) and, in principle, the excited state energy could be formulated using Eq. (6). Therefore, a correction term that accounts for the interaction of the ground state dipoles and the excited state wavefunction, must be added to $\mathrm{W}_2^{\mathrm{EX}}$ in method 2. This correction term is:

$$\langle\psi_{CIS}|H_o^{\mathrm{EX}} + H_{pol}^{EX}|\psi_{CIS}\rangle - \langle\psi_{CIS}|H_o^{\mathrm{EX}} + H_{pol}^{G}|\psi_{CIS}\rangle \tag{34}$$

The second term in Eq. (34) corresponds to the second term in Eq. (31), which is the ground state induced dipole term. The first term in Eq. (34) adds the electron polarization response to the excited state dipoles.

The leading correction term in Eq. (34) can be simplified to:

$$= -\left(\frac{1}{2}\sum_{\mathrm{i}}[(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_{\mathrm{i}}^{EX'}) - (\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{\mathrm{i}}^{G'}](\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle)\right) \tag{35}$$

Adding the correction term (Eq. (35)) to Eq. (33),

$$\begin{aligned} E_2^{\mathrm{EX}} = \mathrm{W}_2^{\mathrm{EX}} &+ \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_i^{G'})\,\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_i^{EX'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot,EX}} \\ &- \left(\frac{1}{2}\sum_{\mathrm{i}}[(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_i^{EX'}) - (\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_i^{G'})](\langle\psi_{CIS}|\vec{f}_i^{\,el}|\psi_{CIS}\rangle)\right) \end{aligned} \tag{36}$$

Eq. (36) can be written in the form of Eq. (27) as:

$$\mathrm{W}_1^{\mathrm{EX}} + \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{i}^{G'})\left\langle \psi_{CIS}\left|\vec{f}_{i}^{\,el}\right|\psi_{CIS}\right\rangle = E_1^{\mathrm{EX}} + \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{i}^{G'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot},G} \tag{37}$$

Using $\mathrm{W}_1^{\mathrm{EX}} = \mathrm{W}_2^{\mathrm{EX}}$ and Eq. (37), Eq. (36) becomes

$$\begin{aligned} E_2^{\mathrm{EX}} = E_1^{\mathrm{EX}} &+ \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{i}^{G'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot},G} - \frac{1}{2}\sum_{\mathrm{i}}(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_{i}^{EX'})\vec{\mathrm{F}}_{\mathrm{i}}^{\mathrm{tot},\mathrm{EX}} - \\ &\left(\frac{1}{2}\sum_{\mathrm{i}}[(\vec{\mu}_{\mathrm{i}}^{EX} - \vec{\mu}_{i}^{EX'}) - (\vec{\mu}_{\mathrm{i}}^{G} - \vec{\mu}_{i}^{G'})](\langle \psi_{CIS}\left|\vec{f}_{i}^{\,el}\right|\psi_{CIS}\rangle)\right) \end{aligned} \tag{38}$$

Eq. (38) is the final working equation for method 2. Note that the response to the excited state density is not fully iterated to self-consistency.

The three methods described here are general approaches and can be applied to other QM methods as well.

III. Computational methods

Solvent induced shifts are calculated by taking the difference between the gas phase and aqueous phase vertical excitation energies of the solute. Several examples are used to test the CIS/EFP1 method: a water dimer ($H_2O(H_2O)$), formaldehyde with one water ($HCHO(H_2O)$), dimethyl sulphoxide in four waters ($CH_3SOCH_3(H_2O)_4$) and dimethyl sulphoxide in fourteen waters ($CH_3SOCH_3(H_2O)_{14}$). The solute molecules in these examples are treated using the CIS method, and the water molecules are described with the EFP1/HF method. The ground state structures are optimized using RHF/6-31G+(d,p), and the vertical excitation energies are calculated using the CIS/EFP methods 1 and 2 described in the previous section. The solvent-induced shifts calculated in this manner are compared with full ab initio calculations

in which both the solute and the solvent molecules are treated using the CIS/6-31G+(d) level of theory at the same geometry as for CIS/EFP1.

In addition to the test examples discussed in the previous paragraph, the CIS/EFP1 method 1 was also used to calculate the solvent induced shifts of solute molecules in bulk solution. Two systems, acetone in 100 EFP1 water molecules and coumarin 151 in 150 EFP1 waters, are chosen to study the solvent induced shifts in the condensed phase. For both systems, the molecular dynamics (MD) method was employed to generate several configurations. In the MD simulations, an isolated system consisted of 100 EFP1 waters with one acetone molecule and 150 EFP1 water molecules with one coumarin 151 molecule. For both systems, acetone and coumarin151 belong to the QM region and the EFP1 water molecules belong to the MM region. A Nosé-Hoover chain method[65] was employed to perform the canonical ensemble (NVT) simulation at 300K temperature, and a 1 fs time step was chosen. Snapshots were taken at every 10 fs and at each snapshot, the vertical excitation energy was calculated using CIS/EFP1 method 1. All calculations were done using electronic structure code GAMESS.[66] The structures shown in the figures are visualized with a graphical interface to GAMESS called MacMolPlot[67].

IV. Results and discussion

Tables 1-4 show the vertical excitation energies and the solvent induced shifts calculated using CIS/EFP1 methods 1 and 2. The solvent induced shifts of formaldehyde + 1 water are examined in Table 1. The shifts calculated using both methods 1 and 2 are in very good agreement with the full ab initio values. The same is true for the water dimer shown in Table 2, dimethyl sulphoxide$(H_2O)_4$, Table 3, and $CH_3SOCH_3(H_2O)_{14}$, Table 4. In all examples, the

errors are on the order of a few hundredths of eV. The accuracy of method 1 relative to method 2 and to the full QM results for both absolute excitation energies and solvent shifts, suggests that at least for these four test molecules, the indirect effect of the perturbation of the ground state molecular orbitals by the EFP1 potential makes the overwhelming contribution to the solvent-induced shifts. In comparison, the perturbation of the excited state wavefunction has only a very minor effect. This may not always be the case, and many more examples must be tested with CIS and more sophisticated excited state methods. However, a tentative conclusion is that the fully consistent method is not expected to be required. Consequently, in the next two examples, only method 1 is considered. Method 1 is now applied to study acetone in 100 EFP1 water molecules and coumarin 151 in 150 EFP1 waters.

Acetone in 100 EFP1: The absorption of acetone has been extensively studied in previous theoretical calculations[30,50,68,69] as well as in experimental measurements[70-72]. The S_0 to S_1 excitation in acetone molecule is the n→π* excitation that occurs from an O lone pair into the π* orbital of the carbonyl double bond. The highest occupied (n) and lowest unoccupied (π*) molecular orbitals are shown in Figure 1.

MD simulations were used to generate several configurations of acetone in 100 EFP1 water molecules at 300 K. A snapshot of acetone in 100 EFP1 waters during a MD simulation is shown in Figure 2. A simulated absorption spectrum was generated, by calculating the vertical excitation energy at each snapshot obtained from the MD simulation. Simulated absorption spectra of the acetone molecule in both the gas and aqueous phases are shown in Figure 3. The curve on the left illustrates the gas phase n→π* absorption spectrum, while the one on the right shows the absorption spectrum for solvated acetone. The calculated acetone

spectrum shows a solvent-induced blue shift for the n→π*; this is qualitatively consistent with the experimental results[72]. The averaged gas phase and aqueous phase acetone n→π* vertical excitation energies and solvent-induced shifts are compared in Table 5 with previous experimental measurements[70,71] and theoretical calculations.[30] The CIS/EFP1 vertical excitation energies are overestimated, reflecting the approximate nature of the CIS method; however, the solvent-induced shifts in the absorption spectrum predicted by CIS/EFP1 are in qualitative agreement with the previous results.

Coumarin 151 in 150 EFP waters: Coumarins have been studied extensively because they are known to exhibit desirable anti-cancer and antibiotic properties.[73,74] They also exhibit interesting solvent dynamics[4,9,75,76] There have been several experimental and theoretical studies of the excited state properties of coumarin 151.[19,69,77-79] The π→π* transition in the gas and condensed phases of coumarin 151 molecule is studied here. The HOMO (π) and LUMO (π*) coumarin 151 orbitals are shown in Figure 4. Simulated π→π* absorption spectra in the gas and condensed phases were generated using several configurations obtained from MD simulations at 300 K (see Figure 5). A snapshot of coumarin 151 in 150 EFP waters during the MD simulation is shown in Figure 6. The averaged π→π* vertical excitation energies in the gas phase and aqueous phases are compared with experimental values in Table 6. The CIS/EFP1 method overestimates the vertical excitation energies relative to the experiments, as one would expect. However, a red shift is correctly predicted for the π→π* excitation of coumarin 151, as the medium is changed from gas to aqueous phase; this is qualitatively consistent with previous experimental and theoretical calculations.[13,43,47,55,68-75]

V. Conclusions

In order to study the solvent induced shifts in the electronic spectra, the work herein discusses three approaches that have been formulated to interface a QM method for excited states (CIS) with an explicit solvent method called EFP. The main question that has been addressed in this study is how to incorporate the polarization perturbation due to the solvent molecules in the excited state energy.

The fully self-consistent method is the most sophisticated and complex approach that calculates the solvent response within the CIS iterative procedure to obtain an excited state that is fully consistent with the environment. The second approach, method 1, is the simplest approach that indirectly alters the excited states via HF orbitals that are modified due to their interactions with the EFP solvent molecules. The third approach (method 2) is an approximation to the fully self-consistent method and adds a one-time perturbative correction term that includes the solvent response for the excited state density. Methods 1 and 2 have been implemented and successfully tested in GAMESS for micro-solvated solute molecules. The test examples examined here show that both methods 1 and 2 predict vertical excitation energies and solvent shifts that are in good agreement with the full ab initio results. The accuracy of method 1 relative to the full QM results leads to the conclusion that the indirect effect of the perturbation of the ground state molecular orbitals by the EFP1 potential makes by far the most significant contribution to the solvent-induced shifts. Indeed, the modifications in method 2 relative to method 1 have a very small impact on the solvent-induced shifts. The CIS/EFP1 method 1, as applied to the prediction of bulk solvent effects on the vertical excitation energies of acetone and coumarin 151, exhibits qualitative

agreement with the experimental measurements. Therefore, method 1 is a simple method that can be employed to semi-quantitatively study solvent effects in large systems.

It is important to recognize that the formulations described here to combine the excited state method, CIS, with the explicit solvent model, EFP1, are general and can therefore be extended to more sophisticated excited state methods, such as EOM-CC, multi-reference CI, multi-reference perturbation theory and CISD to accurately capture the quantitative solvent effects on the excited states.

Acknowledgements

This research was supported by a grant (to MSG) from the US Department Of Energy, Office of Science, Basic Energy Sciences, administered by the Ames Laboratory, Iowa State University, and by support (to LVS) from Purdue University, the Petroleum Research fund (49271-DNI6) and the National Science Foundation (CHE-0955419). The authors gratefully acknowledge helpful discussions with Drs. Michael W. Schmidt and Paul N. Day.

References

(1) Chapman, C. F.; Maroncelli, M. *Journal of Physical Chemistry* **1991**, *95*, 9095.
(2) Chowdhury, P. K.; Halder, M.; Sanders, L.; Arnold, R. A.; Liu, Y.; Armstrong, D. W.; Kundu, S.; Hargrove, M. S.; Song, X.; Petrich, J. W. *Photochem Photobiol* **2004**, *79*, 440.
(3) Chapman, C. F.; Fee, R. S.; Maroncelli, M. *Journal of Physical Chemistry* **1990**, *94*, 4929.
(4) Jimenez, R.; Fleming, G. R.; Kumar, P. V.; Maroncelli, M. *Nature* **1994**, *369*, 471.
(5) Rechthaler, K.; Kohler, G. *Chemical Physics* **1994**, *189*, 99.
(6) Jones, G.; Feng, Z. M.; Bergmark, W. R. *Journal of Physical Chemistry* **1994**, *98*, 4511.
(7) Jones, G.; Jackson, W. R.; Choi, C.; Bergmark, W. R. *Journal of Physical Chemistry* **1985**, *89*, 294.
(8) Zhao, W. W.; Pan, L.; Bian, W. S.; Wang, J. P. *Chemphyschem* **2008**, *9*, 1593.

(9) Horng, M. L.; Gardecki, J. A.; Papazyan, A.; Maroncelli, M. *Journal of Physical Chemistry* **1995**, *99*, 17311.
(10) Adhikary, R.; Barnes, C. A.; Petrich, J. W. *Journal of Physical Chemistry B* **2009**, *113*, 11999.
(11) Chowdhury, P. K.; Halder, M.; Sanders, L.; Calhoun, T.; Anderson, J. L.; Armstrong, D. W.; Song, X.; Petrich, J. W. *Journal of Physical Chemistry B* **2004**, *108*, 10245.
(12) Halder, M.; Headley, L. S.; Mukherjee, P.; Song, X.; Petrich, J. W. *Journal of Physical Chemistry A* **2006**, *110*, 8623.
(13) Pal, S. K.; Zewail, A. H. *Chemical Reviews* **2004**, *104*, 2099.
(14) Bose, S.; Adhikary, R.; Mukherjee, P.; Song, X. Y.; Petrich, J. W. *Journal of Physical Chemistry B* **2009**, *113*, 11061.
(15) Bauernschmitt, R.; Ahlrichs, R. *Chemical Physics Letters* **1996**, *256*, 454.
(16) Bauernschmitt, R.; Haser, M.; Treutler, O.; Ahlrichs, R. *Chemical Physics Letters* **1997**, *264*, 573.
(17) Miertus, S.; Scrocco, E.; Tomasi, J. *Chemical Physics* **1981**, *55*, 117.
(18) Karelson, M.; Zerner, M. C. *Journal of the American Chemical Society* **1990**, *112*, 9405.
(19) Cave, R. J.; Burke, K.; Castner, E. W. *Journal of Physical Chemistry A* **2002**, *106*, 9294.
(20) Karelson, M. M.; Katritzky, A. R.; Zerner, M. C. *International Journal of Quantum Chemistry* **1986**, 521.
(21) Chipman, D. M. *Theoretical Chemistry Accounts* **2002**, *107*, 80.
(22) Tomasi, J.; Persico, M. *Chemical Reviews* **1994**, *94*, 2027.
(23) Tomasi, J.; Mennucci, B.; Cammi, R. *Chemical Reviews* **2005**, *105*, 2999.
(24) Karelson, M. M.; Zerner, M. C. *Journal of Physical Chemistry* **1992**, *96*, 6949.
(25) Li, J.; Cramer, C. J.; Truhlar, D. G. *International Journal of Quantum Chemistry* **2000**, *77*, 264.
(26) Cossi, M.; Barone, V. *Journal of Chemical Physics* **2000**, *112*, 2427.
(27) Minezawa, N.; Kato, S. *Journal of Chemical Physics* **2007**, *126*.
(28) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *Journal of Physical Chemistry A* **2001**, *105*, 293.
(29) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *Journal of Chemical Physics* **1996**, *105*, 1968.
(30) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. *Journal of Chemical Physics* **2008**, *129*.
(31) Foresman, J. B.; Headgordon, M.; Pople, J. A.; Frisch, M. J. *Journal of Physical Chemistry* **1992**, *96*, 135.
(32) Tuck, P. O.; Mawhinney, R. C.; Rappon, M. *Physical Chemistry Chemical Physics* **2009**, *11*, 4471.
(33) Sanchez, M. L.; Aguilar, M. A.; Delvalle, F. J. O. *Journal of Physical Chemistry* **1995**, *99*, 15758.
(34) Kina, D.; Arora, P.; Nakayama, A.; Noro, T.; Gordon, M. S.; Taketsugu, T. *International Journal of Quantum Chemistry* **2009**, *109*, 2308.
(35) Krauss, M.; Webb, S. P. *Journal of Chemical Physics* **1997**, *107*, 5771.
(36) Kina, D.; Nakayama, A.; Noro, T.; Taketsugu, T.; Gordon, M. S. *Journal of Physical Chemistry A* **2008**, *112*, 9675.

(37) Besley, N. A.; Hirst, J. D. *Journal of the American Chemical Society* **1999**, *121*, 8559.
(38) Parusel, A. B. J.; Rettig, W.; Sudholt, W. *Journal of Physical Chemistry A* **2002**, *106*, 804.
(39) Nakatsuji, H.; Hirao, K. *Chemical Physics Letters* **1977**, *47*, 569.
(40) Miyahara, T.; Nakatsuji, H.; Hasegawa, J.; Osuka, A.; Aratani, N.; Tsuda, A. *Journal of Chemical Physics* **2002**, *117*, 11196.
(41) Geertsen, J.; Rittby, M.; Bartlett, R. J. *Chemical Physics Letters* **1989**, *164*, 57.
(42) Fan, P. D.; Valiev, M.; Kowalski, K. *Chemical Physics Letters* **2008**, *458*, 205.
(43) Kowalski, K.; Piecuch, P. *Journal of Chemical Physics* **2004**, *120*, 1715.
(44) Kawashima, Y.; Hashimoto, T.; Nakano, H.; Hirao, K. *Theoretical Chemistry Accounts* **1999**, *102*, 49.
(45) Nakayama, K.; Nakano, H.; Hirao, K. *International Journal of Quantum Chemistry* **1998**, *66*, 157.
(46) Xu, Z. R.; Matsika, S. *Journal of Physical Chemistry A* **2006**, *110*, 12035.
(47) Kistler, K. A.; Matsika, S. *Journal of Physical Chemistry A* **2009**, *113*, 12396.
(48) Jensen, J. H.; Gordon, M. S. *Journal of Chemical Physics* **1998**, *108*, 4772.
(49) Adamovic, I.; Gordon, M. S. *Molecular Physics* **2005**, *103*, 379.
(50) Aidas, K.; Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *Journal of Physical Chemistry A* **2005**, *109*, 8001.
(51) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *Journal of Chemical Physics* **2003**, *118*, 1620.
(52) Warshel, A. *Journal of Physical Chemistry* **1979**, *83*, 1640.
(53) Luzhkov, V.; Warshel, A. *Journal of the American Chemical Society* **1991**, *113*, 4491.
(54) Thompson, M. A.; Schenter, G. K. *Journal of Physical Chemistry* **1995**, *99*, 6374.
(55) Gao, J. L.; Byun, K. *Theoretical Chemistry Accounts* **1997**, *96*, 151.
(56) Tapia, O.; Goscinski, O. *Molecular Physics* **1975**, *29*, 1653.
(57) Davidson, E. R. *Journal of Computational Physics* **1975**, *17*, 87.
(58) Adamovic, I.; Freitag, M. A.; Gordon, M. S. *Journal of Chemical Physics* **2003**, *118*, 6725.
(59) Stone, A. J. *Chemical Physics Letters* **1981**, *83*, 233.
(60) Slipchenko, L. V.; Gordon, M. S. *Molecular Physics* **2009**, *107*, 999.
(61) Feynman, R. P. *Physical Review* **1939**, *56*, 340.
(62) Gerratt, J.; Mills, I. M. *Journal of Chemical Physics* **1968**, *49*, 1730.
(63) Gerratt, J.; Mills, I. M. *Journal of Chemical Physics* **1968**, *49*, 1719.
(64) Defusco, A.; Schmidt, M. W.; Ivanic, J.; Gordon, M. S. *in preparation.*
(65) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Molecular Physics* **1996**, *87*, 1117.
(66) Gordon, M. S.; Schmidt, M. W. *In Theory and Applications of Computational Chemistry: The First Forty Years*, Dykstra, C. E., Frenking, G., Kinm, K. S., Scuseria, G. E.; Eds. (Elsevier, Amsterdam, 2005).
(67) Bode, B. M.; Gordon, M. S. *Journal of Molecular Graphics & Modelling* **1998**, *16*, 133.

(68) Rohrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. *Chemphyschem* **2003**, *4*, 1177.
(69) Sulpizi, M.; Rohrig, U. F.; Hutter, J.; Rothlisberger, U. *International Journal of Quantum Chemistry* **2005**, *101*, 671.
(70) Bayliss, N. S.; Mcrae, E. G. *Journal of Physical Chemistry* **1954**, *58*, 1006.
(71) Bayliss, N. S.; Willsjoh.G. *Spectrochimica Acta Part a-Molecular Spectroscopy* **1968**, *A 24*, 551.
(72) Hayes, W. P.; Timmons, C. J. *Spectrochimica Acta* **1965**, *21*, 529.
(73) Singer, K. D.; Lalama, S. L.; Sohn, J. E.; Small, R. D. *Nonlinear Optical Properties of Organic Molecules and Crystals*; Academic Press: Orlando, FL, 1987, Chap. II-8.
(74) Nicoud, J. F.; Twieg, R. J. *Nonlinear Optical Properties of Organic Molecules and Crystals*; Academic Press: Orlando, FL, 2002: Chap. II-3.
(75) Chowdhury, P. K.; Halder, M.; Sanders, L.; Arnold, R. A.; Liu, Y.; Armstrong, D. W.; Kundu, S.; Hargrove, M. S.; Song, X.; Petrich, J. W. *Photochemistry and Photobiology* **2004**, *79*, 440.
(76) Halder, M.; Mukherjee, P.; Bose, S.; Hargrove, M. S.; Song, X. Y.; Petrich, J. W. *Journal of Chemical Physics* **2007**, *127*.
(77) Pryor, B. A.; Palmer, P. M.; Andrews, P. M.; Berger, M. B.; Topp, M. R. *Journal of Physical Chemistry A* **1998**, *102*, 3284.
(78) Sulpizi, M.; Carloni, P.; Hutter, J.; Rothlisberger, U. *Physical Chemistry Chemical Physics* **2003**, *5*, 4798.
(79) Neugebauer, J.; Jacob, C. R.; Wesolowski, A. T.; Baerends, E. J. *Journal of Physical Chemistry* **2005**, *109*, 7805.

Table 1: Vertical excitation energies (n→π*) and solvent induced shifts (eV) for HCHO(H_2O) calculated using CIS/EFP methods 1 and 2 and with full ab initio CIS.

	Method 1	Method 2	Full ab initio
Vertical Excitation Energy (ΔE)	4.87	4.86	4.87
Solvent shift[a]	0.18	0.17	0.18

[a]Gas phase vertical excitation energy is 4.69 eV.

Table 2: Vertical excitation energies (n→σ*) and solvent induced shifts (eV) for $H_2O(H_2O)$ calculated using CIS/EFP1 methods 1 and 2 and with full ab initio CIS.

	Method 1	Method 2	Full ab initio
Vertical Excitation Energy (ΔE)	9.75	9.72	9.71
Solvent shift[a]	0.42	0.39	0.38

[a]Gas phase vertical excitation energy is 9.33 eV.

Table 3: Vertical excitation energies (n→σ*) and solvent induced shifts (eV) for $CH_3SOCH_3(H_2O)_4$ calculated using CIS/EFP1 methods 1 and 2 and with full ab initio CIS.

	Method 1	Method 2	Full ab initio
Vertical Excitation Energy (ΔE)	7.06	7.05	7.06
Solvent shift[a]	0.16	0.16	0.17

[a]Gas phase vertical excitation energy is 6.89 eV.

Table 4: Vertical excitation energies (n→σ*) and solvent induced shifts (eV) for $CH_3SOCH_3(H_2O)_{14}$ calculated using CIS/EFP1 methods 1 and 2 and with full ab initio CIS**.**

	Method 1	Method 2	Full ab initio
Vertical Excitation Energy (ΔE)	7.29	7.28	7.30
Solvent Shift[a]	0.50	0.49	0.51

[a]Gas phase vertical excitation energy is 6.79 eV.

Table 5. Comparison of the calculated average vertical excitation energies (n→π*) and solvent induced shifts (eV) of acetone in gas and aqueous phase with previous work.

	Gas phase		
	Expt[56,57,58]	TDDFT[24]	CIS
Vertical Excitation Energy (ΔE)	4.48	4.38	5.16
	Aqueous phase		
	Expt[56,57,58]	TDDFT/EFP1[24]	Method 1
Vertical Excitation Energy (ΔE)	4.67-4.69	4.59	5.52
Solvent Shift	0.19-0.21	0.21	0.36

Table 6. Comparison of the average vertical excitation energy ($\pi \rightarrow \pi^*$) and solvent shifts (eV) of coumarin *151* in gas and aqueous phase with previous experimental work.

	Expt[61] (gas)	Expt[64] (aqueous)	CIS (gas)	Method 1
Vertical Excitation Energy (ΔE)	3.55	3.48	4.99	4.91
Solvent Shift		0.08		0.07

FIGURE CAPTIONS

Figure 1. The highest occupied (a) and lowest unoccupied (b) molecular orbitals of acetone (CH_3COCH_3)

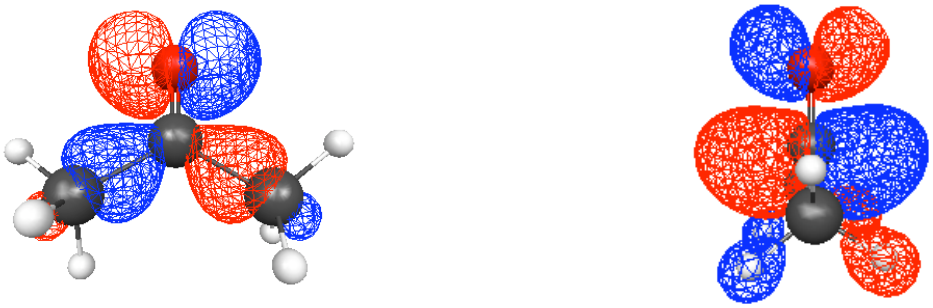
Figure 2. A snapshot of acetone in 100 EFP1 water molecules during the molecular dynamics simulation.

Figure 3. A simulated spectrum for the n→π* vertical excitation energy of acetone. The curve on the left is for gas phase acetone and the one on the right is for solvated acetone.

Figure 4. The highest occupied (π) and lowest unoccupied (π*) molecular orbitals of coumarin 151 (7-amino-4-trifluoromethyl-1,2-benzopyrone).

Figure 5. A simulated spectrum for a π→π* vertical excitation energy of coumarin 151. The right curve is for the gas phase coumarin 151 and the left curve represents the solvated coumarin 151.

Figure 6. A snapshot of coumarin 151 in 150 EFP1 water molecules during a molecular dynamics simulation.



(a) HOMO (n) (b) LUMO (π*)

Figure 1. The highest occupied (a) and lowest unoccupied (b) molecular orbitals of acetone (CH_3COCH_3)

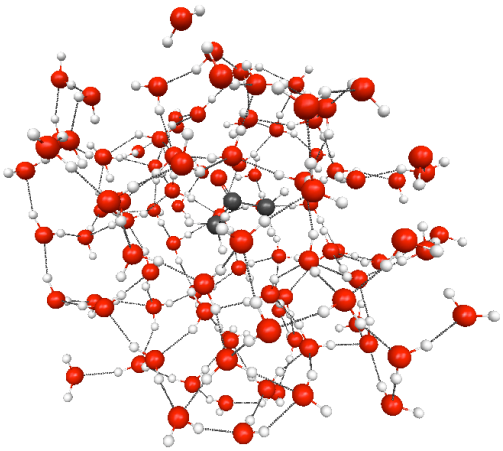


Figure 2. A snapshot of acetone in 100 EFP1 water molecules during the molecular dynamics simulation.

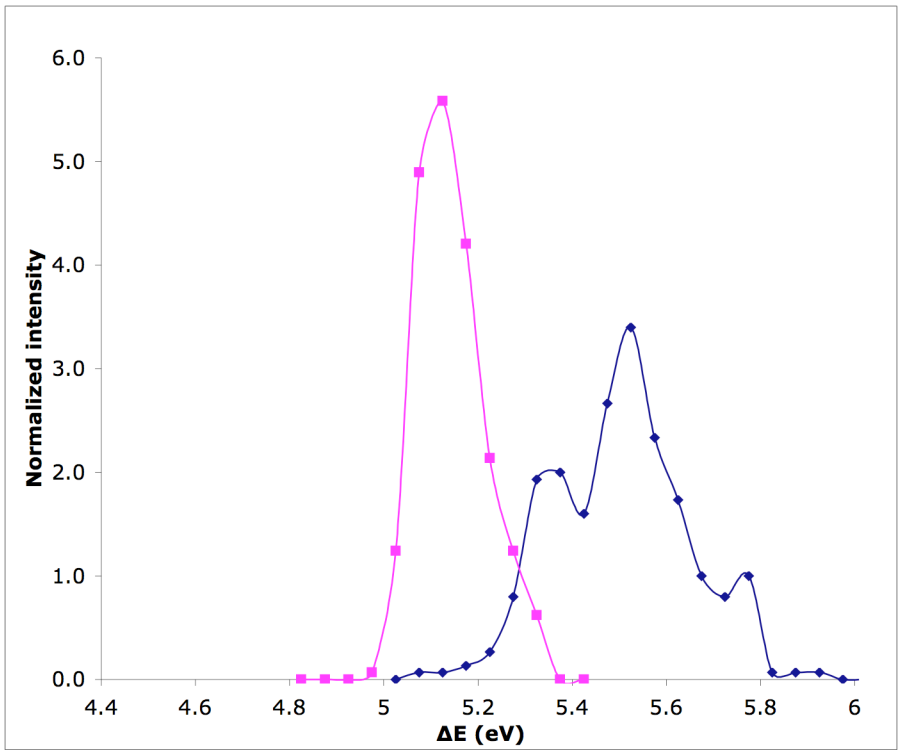
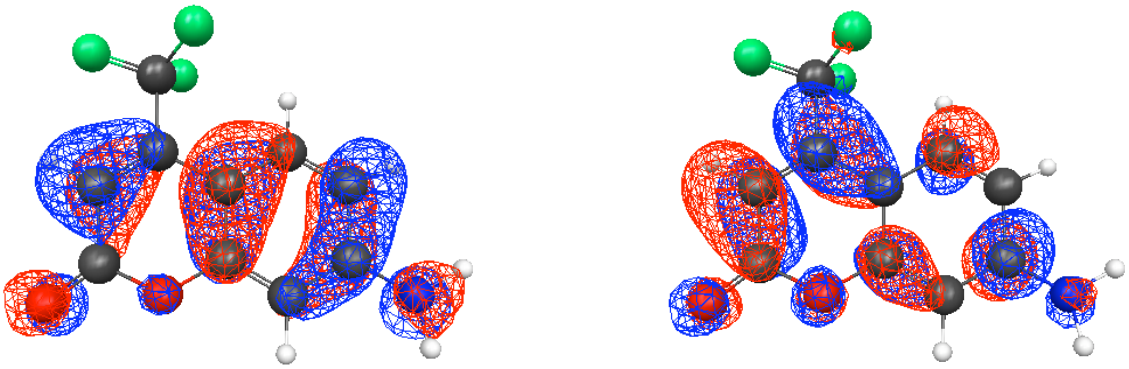


Figure 3. A simulated spectrum for the n→π* vertical excitation energy of acetone. The curve on the left is for gas phase acetone and the one on the right is for solvated acetone.



(a) HOMO (π) (b) LUMO (π*)

Figure 4. The highest occupied (π) and lowest unoccupied (π*) molecular orbitals of coumarin *151* (7-amino-4-trifluoromethyl-1,2-benzopyrone).

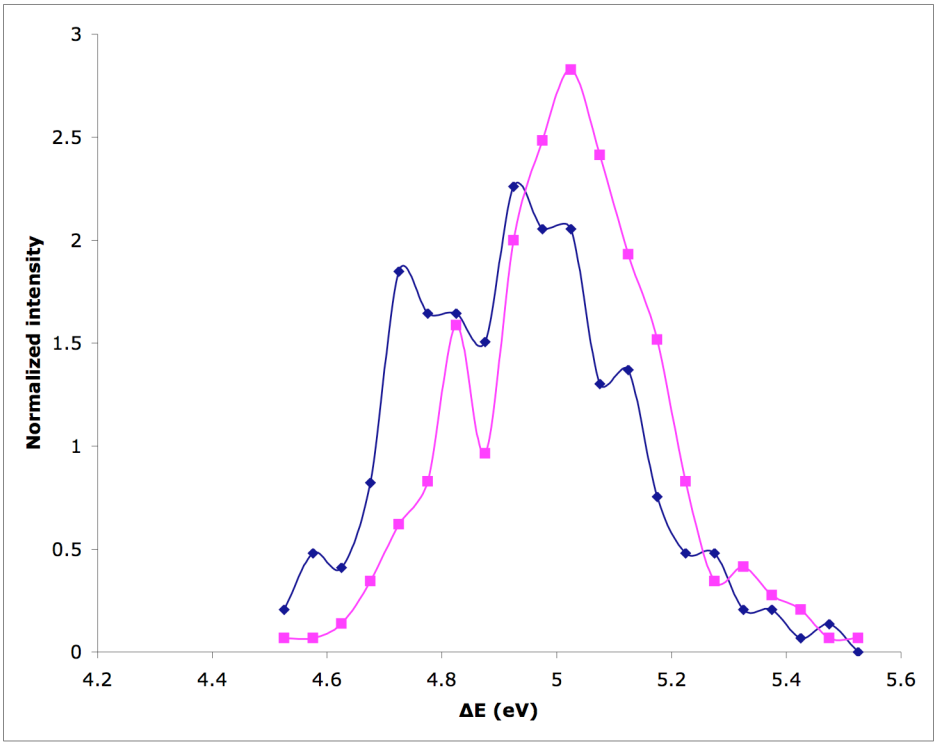


Figure 5. A simulated spectrum for a π→π* vertical excitation energy of coumarin *151*. The right curve is for the gas phase coumarin *151* and the left curve represents the solvated coumarin *151*.

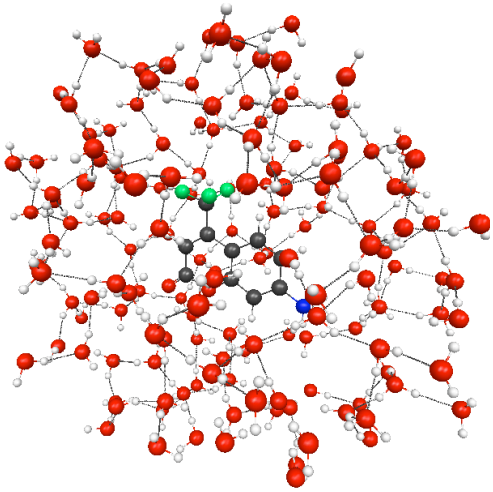


Figure 6. A snapshot of coumarin *151* in 150 EFP1 water molecules during a molecular dynamics simulation.

CHAPTER 3: THEORETICAL STUDY OF THE ELECTRON AFFINITY OF OH IN WATER

To be submitted to the Journal of Physical Chemistry

Pooja Arora and Mark S. Gordon

Abstract

The adiabatic electron affinity of the hydroxyl radical in aqueous clusters has been studied by a systematic series of calculations on the hydration of a hydroxyl radical ($OH\cdot$) and a hydroxide anion (OH^-). Ab initio methods have been used for the solute molecule, and the water molecules are represented using the effective fragment potential (EFP) method. Minimum energy configurations for the $OH\cdot(H_2O)_n$ and $OH^-(H_2O)_n$ clusters were determined using Monte Carlo/simulated annealing calculations. The adiabatic electron affinity of the hydroxyl radical in aqueous solution is predicted to be 101.5 kcal/mol, based on second order perturbation theory (MP2) calculations at the combined Hartree-Fock (HF)/EFP geometries. Surrounding the 15-water clusters with a continuum solvent illustrates that the value of the aqueous electron affinity has not converged at 15 water molecules. The global minimum searches for $OH^-(H_2O)_n$ clusters suggest that the global minimum structures are surface species for up to 14 waters. For n = 15, the global minimum has the OH^- completely surrounded by waters; that is, $OH^-(H_2O)_{15}$ is an interior global minimum structure. This is consistent with previous studies on F^- and Cl^-. In contrast, the global minima of $OH\cdot(H_2O)_n$ clusters predict that

the OH radical has a propensity to stay at the surface even up to 15 waters. There exists an interesting similarity between the $OH\cdot(H_2O)_n$ global minimum structures and those of the corresponding water clusters.

I. Introduction

There has been great interest in the study of free radicals like hydroxyl (OH·) because of its importance in atmospheric science, radiation chemistry, and biological processes. For example, the hydroxyl radicals formed in the earth's atmosphere[1,2] from the dissociation of ozone molecules by ultraviolet radiation are known to reduce greenhouse gases like carbon monoxide and methane.[3] The photolysis of nitrates[4] dissolved in water and radiolysis of liquid water are additional sources of hydroxyl radical formation. In biological systems, the Fenton reaction[5], in which hydrogen peroxide oxidizes Fe(II) to Fe(III) to form hydroxyl radical is an important source of hydroxyl formation. Hydroxyl radical is also believed to be associated with the ageing[6] process and diseases like Parkinson's.[7] In the critical review by Dorfman and Adam[8], hydroxyl radical chemistry in aqueous solution is analyzed in detail.

Due to the high gas phase electron affinity of OH· (42.0 kcal/mol)[9], the hydroxyl radicals formed in biological systems are known to cause damage to biomolecules[10,11] like amino acids, enzymes, carbohydrates and most importantly DNA. For instance, the hydroxyl radical can cause oxidative damage to DNA base pairs like guanine and adenine.[12-14] Mundy et al. studied the interaction of guanine with the hydroxyl radical in the gas and aqueous phases using a Car-Parrinello molecular dynamics (CPMD) approach and a dielectric continuum solvation model.[15] The presence of an unpaired

electron makes the hydroxyl radical highly unstable and short-lived, rendering it very difficult in some instances to study experimentally. Therefore, theoretical studies can be very useful.

Because biological systems largely exist in aqueous solution, the present study is focused on the behavior and properties of $OH\cdot$ in water. Since OH^- is an important component in water, especially when electrolytes are present, the electron affinity of $OH\cdot$ in water is of particular interest. The importance of the OH radical has stimulated several experimental studies including the self-recombination of the OH radical to form hydrogen peroxide and the absorption spectrum of the hydroxyl radical in aqueous solution by spectroscopic methods.[16] It was observed that with an increase in temperature, the hydrogen bonded OH radical peak (250 nm) decreases and the absorption band for hydrogen peroxide intensifies. The decrease of the 250 nm peak with an increase in temperature was attributed to the decrease in the population of hydrogen-bonded hydroxyl radical. Another study[17] on the vertical excited states of OH radical in $(H_2O)_n$ clusters, for n=1,7 and 16, using time dependent density functional theory (TDDFT) revealed the presence of a solvent → solute charge transfer peak for $OH\cdot(H_2O)_{n=16}$. The structures in that study were optimized using the DFT method with the MPW1K functional[18], and the excitations were determined using TDDFT within the Tamm-Dancoff approximation.[19] Several studies have also been performed on the microsolvation of OH radical[20-28] using ab initio electronic structure methods to study the structural details of the OH radical in water clusters. These ab initio studies show that the hydroxyl radical participates in H-bonding with water as a hydrogen acceptor through the oxygen atom and as a hydrogen donor via the hydrogen atom. Hamad et al

studied OH· - water complexes for 1-5 waters using DFT/BLYP, MP2 and DFT/BHLYP[22]. They found that the two DFT methods do not predict the correct OH· - water interaction. Vassilev et al studied the OH radical in liquid water using molecular dynamics (MD) simulations with the BLYP functional. These authors predicted that the OH radical is bound to three water molecules via H–bonds and has a hemi bond (three-electron two-center hydrogen bond) to a fourth water molecule[27]. CPMD studies by Khalack et al. on the OH radical in water exhibit the similarity of the structural and dynamical properties of OH radical – water clusters and pure water clusters[23]. Couto et al. analyzed the structure, vibrational spectrum and energetics of $OH\cdot(H_2O)_6$ and water clusters using DFT with the MPW1PW91[29] functional and the aug-cc-pVDZ basis set[21]. They also studied the OH radical in bulk water using Monte Carlo simulations and found that the hydration enthalpies of the OH radical and water differ by less than 2.4 kcal/mol.

Fewer studies have focused on the calculation of the electron affinity of the hydroxyl radical in aqueous solution. Recently, DFT MD calculations combined with experiments were performed by Adriaanse et al. to estimate the vertical attachment energy of the hydroxyl radical from the detachment energy. These authors reported the adiabatic ionization potential of OH^- in aqueous solution to be 145.2 kcal/mol[28]. Mundy et al. calculated the electron affinity of the hydroxyl radical in aqueous solution to be 142.3 kcal/mol[15] using DFT with the Becke three-parameter Lee-Yang-Parr (B3LYP) functional method and the COntinuum Solvation MOdel (COSMO) for the solvent.

The present paper describes an investigation of the electron affinity of hydrated hydroxyl radical by microsolvating the hydroxyl radical (OH·) and hydroxide anion

(OH^-) with up to 15 water molecules. The effective fragment potential (EFP)[30,31] method has been used to treat the water molecules. An EFP is a discrete model potential method for treating intermolecular interactions including solvent effects. HF/EFP calculations are compared in this work with correlated electronic structure methods to establish the reliability of EFPs for this type of application. The global and low-energy local minima structures of $OH^-(H_2O)_n$ and $OH\cdot(H_2O)_n$ have been studied as a function of n, in order to understand the convergence of the electron affinity as n increases. Also, the relative energies of surface and interior structures of $OH^-(H_2O)_n$ and $OH\cdot(H_2O)_n$ are analyzed, since an interior structure can be interpreted as a "fully solvated" species. Finally, the binding energies of the solute to the solvent molecules are calculated for the $OH^-(H_2O)_n$ and $OH\cdot(H_2O)_n$ clusters as a function of the number of water molecules.

II. Computational Methods

The $OH\cdot(H_2O)_n$ and $OH^-(H_2O)_n$ clusters were initially constructed by a stepwise addition of water molecules to the solute with n ranging from 0 to 15 waters. Global minimum energy structures of the $OH\cdot(H_2O)_{n(1\text{-}15)}$ and $OH^-(H_2O)_{n(1\text{-}16)}$ clusters were determined using a Monte Carlo[32]/simulated annealing (MC/SA) method[33,34]. The initial temperature used in the MC/SA calculations was 600 K. The system was systematically cooled to 300K, over 14 steps. Geometry optimizations were performed after every 10 steps in the simulation. Initial structures for the MC/SA simulations were determined by placing water molecules in various positions relative to the structures with one less water molecule. Initial structures were also obtained by removing one water molecule from

various systematically chosen positions of the n-water global minimum structure and then optimizing to obtain the structure with n-1 waters. In order to find a global minimum for each n, ~1400 structures were sampled in a MC/SA run. For the MC/SA simulations, the solute molecule was treated with Hartree-Fock (HF) and the 6-31++G(d,p)[35,36] basis set. The water molecules were represented using the Hartree Fock based effective fragment potential (EFP1/HF) method. Full ab initio geometry optimizations were then performed starting from the HF/EFP1 global minimum structures, for each value of n, using second order perturbation theory (MP2) with the same basis set. In addition, singles + doubles coupled cluster calculations with perturbative triples, CCSD(T), were performed at the MP2 optimized geometries for n up to 5 and then extrapolated to n=8. The extrapolations of the CCSD(T) values were obtained using the following equation.

$$\Delta E_{EST}(CCSD(T))_{aug\text{-}cc\text{-}pVDZ} = \Delta E(MP2)_{aug\text{-}cc\text{-}pVDZ} + |\Delta E(CCSD(T) - \Delta E(MP2)|_{6\text{-}31+G(d)} \quad (1)$$

Stationary points for the optimized structures were characterized by calculating and diagonalizing the Hessian (matrix of energy second derivatives) using the Hartree-Fock method for both the solute and solvent molecules. A positive definite Hessian implies a local minimum. Zero point energy (ZPE) corrections scaled by an empirical factor of 0.89[37] were included in all of the reported energy differences.

The electron affinity of $OH\cdot(H_2O)_{n(0\text{-}15)}$ was calculated using both the HF/EFP1 and MP2 methods for each value of n. The total and differential binding energies of the solute to the water molecules were calculated for $OH\cdot(H_2O)_{n(1\text{-}15)}$ and $OH^{-}(H_2O)_{n(1\text{-}15)}$. All calculations were performed using the General Atomic and Molecular Electronic

Structure System (GAMESS)[38] except the CCSD(T) calculations for which ACESII[39] was used.

In order to account for the additional effect of the bulk solvent, beyond that of 15 water molecules, on the prediction of the aqueous adiabatic electron affinity, calculations were also performed for both OH radical and OH^- in 15 explicit solvent molecules embedded in a continuum using the conductor-like polarizable continuum model (C-PCM)[40], referred to here as PCM. The optimized $OH\cdot(H_2O)_{15}$ and $OH^-(H_2O)_{15}$ structures were found using the HF/EFP1+PCM method.[41-43] The solute molecules ($OH\cdot$ and OH^-) were treated using the HF method and the 15 water molecules were represented by the EFP1 method. A single point MP2+PCM calculation was performed, in which the OH and 15 water molecules were treated using the MP2 method in a PCM continuum.

III. Results and discussion

A. $OH^-(H_2O)_{n=1-16}$: Figures 1-4 illustrate the $OH^-(H_2O)_n$ local and global minimum energy structures for each value of n. Beneath each structure is a unique name (nI), where 'n' represents the number of water molecules present in the cluster and 'I' is a letter. The relative energies between the global and local minimum energy structures at the (HF/EFP1) and [MP2] levels of theory are also shown underneath each structure. The label (X+Y+Z) gives the number of water molecules in the first (X), second (Y) and third (Z) solvent shells. Molecules (X) in the first solvent shell are those that participate in direct H-bonding with the solute. Molecules (Y) in the second solvent shell H-bond with the molecules in the first shell. Molecules (Z) in the third solvent shell H-bond with the molecules in the second shell, and so on. The molecules in each solvent shell are

divided into two sets depending upon the H-bonding distance from the solute or other solvent molecules. So, for example, the designation (3+1+0) for structure 4B in Figure 1 means that by these criteria there are three water molecules in the first solvent shell and one in the second solvent shell. In addition, some structures have a sub-label to indicate how many molecules are outside the range of 1.5 Å -2.5 Å. So, the notation (6+6{2}+0) for 12c in Figure 4 means that of the six water molecules in the second solvent shell, two are outside this range. In Figures 1-4, up to three lowest energy structures of $OH^-(H_2O)_n$ are presented for each value of n. Energies are quoted relative to the zero energy structure, indicated by (0) or [0] for HF/EFP1 and MP2, respectively.

There are small differences in the relative energies predicted by HF/EFP1 and MP2, up to ~2 kcal/mol. The two methods usually predict the same global minimum for a given value of n, and when they differ (e.g., for n = 3, 4, 6, 8), the energy differences are small.

For n=1–14 in $OH^-(H_2O)_n$, all of the global minima are found to be surface structures. That is, no interior global minimum structure was found until the 15th water molecule was added; the OH^- appears to be completely solvated when n = 15, as may be seen in the global minimum, 15A. A second, local minimum structure that is completely solvated is also observed at n=15 (15C in Figure 4). MP2 predicts that this local minimum (15C) is 5.2 kcal/mol higher in energy than the global minimum structure. In order to further confirm the complete solvation, 16 water molecules were added to OH^-. As for n = 15, the two lowest-energy minima for n = 16 have the OH^- fully embedded in the water cluster (See 16A and 16B in Figure 4).

For n=1-4 in $OH^-(H_2O)_n$, all of the water molecules in the HF/EFP1 global minimum structures have a direct hydrogen bond (within the H-bonding distance of 2.5 Å) with the hydroxide ion. The number of these hydrogen bonds is taken to be the coordination number of the OH^- in the first solvent shell. The fifth water in $OH^-(H_2O)_5$ does not coordinate with the hydroxide ion in the structures 5A, 5B and 5C in Figure 2. Similarly, in structure 6A, the sixth water does not directly H-bond with the solute. So, one may conclude that these outer water molecules begin to form a second solvent shell. Apparently, for n=4-7 in $OH^-(H_2O)_n$, the anion generally prefers a coordination number of four, although there are some structures that have a coordination number of five (e.g., 6A and 7C). For n=8 and greater, generally the preferred coordination number remains five except at n=15 where the coordination number increases to seven.

It is interesting to consider the origin of the change in the preferred OH^- coordination number as the number of water molecules, n, increases. The HF/EFP1 global minimum structure for $OH^-(H_2O)_4$ (4A) has a coordination number of 4, whereas the MP2 global minimum structure (4B) has a coordination number of 3. However, the difference in energy between the two MP2 structures is only 0.1 kcal/mol. For n=6 (Figure 2), the MP2 global minimum structure (6C) has 4 waters in the first solvent shell and is 2.3 kcal/mol lower in energy than 6A which is coordinated to 5 waters. The energy order of these two species is reversed at the HF/EFP1 level of theory, with 6A lower in energy by 0.5 kcal/mol. For n=7, both HF/EFP1 and MP2 predict the same global minimum structure (7A) with a coordination number of 4. A local minimum structure (7C), that has a coordination number of 5, is 1.2 kcal/mol and 4.2 kcal/mol higher in energy than 7A at the HF/EFP1 and MP2 levels of theory, respectively (See

Figure 2). Upon adding the 8th water (see Figure 3), the preferred coordination number of OH^- increases from 4 to 5. As shown in Figures 3 and 4, the preferred coordination number remains five until n=14. The preferred coordination number further increases to 7 when the 15th water is added. So, it appears that the lowest energy coordination number for OH^- increases as the number of water molecules increases.

The average H-bond distances between the solute and the first shell solvent water molecules are listed in Table 1. As the number of solvent molecules increases, the average H-bond distance generally (with a few exceptions) increases, suggesting a concomitant decrease in the H-bond strength. Although the EFP distances are larger than those predicted by MP2, the trends are the same. The increase in the coordination number of the hydroxide ion can be probed further by analyzing the Mulliken charges[45-47] on the anion and the neighboring water molecules. Table 1 lists the average MP2 Mulliken charges on the solute (OH^-) and the first shell water molecules of $OH^-(H_2O)_n$ at each n. Compared to the gas phase hydroxide ion, the net charge on the anion in $OH^-(H_2O)_n$ (Table 1, column 6), initially decreases. This is accompanied by a small increase in the negative charge on the water molecules in the first solvent shell (column 10, Table 1). This indicates that there is a charge transfer from the solute (OH^-) to the solvent water molecules. As n increases further, from 8-15, the hydroxide ion charge increases to an absolute value greater than 1.0, with an especially large negative charge (-1.5) on the hydroxide O atom. This suggests a largely Coulombic attraction at n = 15, with a larger coordination number that will be limited largely by the crowding of water molecules about the anion. As expected, the hydrogen atom pointing towards the anion (forming the H-bond with the anion) is more positive than the one that is pointing away.

Examination of the local and global minimum energy structures of $OH^-(H_2O)_n$ clusters for n = 1-12 suggests that the hydroxide ion tends to preferentially form hydrogen bonds through its oxygen atom, rather than its hydrogen atom (See Figures 1-4), until full solvation is achieved at n = 15. This is apparently due to the weaker H-bond propensity for the OH^- hydrogen than for the OH^- oxygen. This observation is consistent with several experimental studies[48,49]. Evidence for this is the long H-bond distance of 2.58 Å for the hydroxide hydrogen, suggesting a very weak H-bond. This observation is reminiscent of previous studies of amino acids[50-53] in which the solvent water molecules congregate near the carboxyl group until this region is saturated, before solvating other regions that form weaker hydrogen bonds.

B. $OH\cdot(H_2O)_{n=1-15}$: The lowest energy structures of $OH\cdot(H_2O)_{n=1-15}$ are shown in Figures 5-8 and the labeling beneath the structures is the same as that for $OH^-(H_2O)_n$ in Figures 1-4. In general, the neutral system does not appear to form as many or as strong hydrogen bonds to water molecules as does the anion. The global minimum structures of $OH\cdot(H_2O)_{n=1-15}$ predicted by Monte Carlo simulations show that the hydroxyl radical is not fully solvated at any value of n that is considered here. Instead, $OH\cdot$ resides on the surface of each water cluster. An attempt to form an internal isomer for $OH\cdot(H_2O)_n$ at n=15 only finds an MP2 isomer (15I) that is 17 kcal/mol higher in energy than the surface global minimum structure (15A). (See Figure 8)

The global minimum structures of $OH\cdot(H_2O)_{n=1-15}$ also suggest that the hydroxyl radical is directly hydrogen bonded to only 2–3 water molecules as n increases from 1 to 15. This means that the first solvent shell of $OH\cdot(H_2O)_n$ generally consists of only 2-3

waters for n=1-15, unlike the hydroxide ion in $OH^-(H_2O)_n$ where the coordination number increases upon adding more water molecules. Of course, OH^- is a strong H-bond acceptor due to the negative charge on O, whereas $OH\cdot$ is a strong H–bond donor. As shown in Figure 9, the H-bond distances are consistently longer when OH is a proton acceptor than when it is a proton donor. This suggests weaker H-bonds when OH is a proton acceptor. It is also clear that when OH is a proton donor, the hydrogen bonding is similar to that in water clusters, as is evident from the comparable CCSD(T) binding energies values of the water dimer and the OH-water dimer when OH is a proton donor (see Table 2). In contrast, the OH-water interaction appears to be much weaker than the water-water interaction when OH is a proton acceptor. This is also illustrated in Table 2, where it is shown that the CCSD(T) OH-water binding energy is much smaller when OH is a proton acceptor. This latter phenomenon contributes to the smaller coordination number for $OH\cdot$ vs. OH^- and to the reticence of $OH\cdot$ to act as an internal solute for the relatively small clusters examined here, since water-water interactions are overall stronger than OH-water interactions. For example, an attempt to coordinate the hydroxyl radical with more than 3 water molecules results in the very high energy structure 15I (See Figure 8), 17.2 kcal/mol higher in energy than the surface species 15A. The ability of $OH\cdot$ to function as both a proton donor and proton acceptor in hydrogen bonds to water molecules is similar to the behavior of water molecules in water clusters. Hartree-Fock calculations by Engdahl et al.[54] on the water-OH complex exhibit trends in the bond lengths that are similar to those shown in Figure 9. This suggests that the HF/EFP1 method captures the character of this hydrogen bonding interaction correctly. Thus, the

OH radical in $OH\cdot(H_2O)_n$ behaves at least qualitatively like another water molecule with one less hydrogen atom. As discussed in Section D below, the similarity between the two clusters is also apparent in the comparable total binding energies of the solute to the water cluster in $OH\cdot(H_2O)_n$ and $(H_2O)_n$.

C. Adiabatic Electron Affinity of $OH\cdot(H_2O)_{n=0-15}$: The adiabatic electron affinity (AEA) is the energy difference between the neutral system $[OH\cdot(H_2O)_n]$ and the corresponding anion $[OH^-(H_2O)_n]$ at their respective equilibrium geometries. The electron affinity for the $OH\cdot(H_2O)_n$ is calculated with Eq. (2), using the energies of the geometry optimized global minima for each value of n.

$$E[OH\cdot(H_2O)_n] - E[OH^-(H_2O)_n] \qquad (2)$$

The gas phase AEA predicted by HF/EFP1, MP2 and CCSD(T) are compared with the experimental values[55,56] in Table 3. The HF/EFP1 method using the 6-31++G(d,p) basis set predicts the gas phase AEA of the hydroxyl radical to be +3.3 kcal/mol, in poor agreement with the experimental value of -42.0 kcal/mol[9]. Indeed, HF incorrectly predicts the additional electron to be unbound. However, when electron correlation is taken into account using the MP2 and CCSD(T) levels of theory with reliable basis sets, the predicted AEA is within 1 kcal/mol of the experimental value. MP2 and CCSD(T) are in good agreement with each other. Chipman[57] also found that a high level of correlation and a large basis set is required to predict the correct value of the gas phase electron affinity of the hydroxyl radical.

The HF/EFP1 and MP2 AEAs for $OH\cdot(H_2O)_n$ are shown in Table 4 for n = 0 to n = 15. The CCSD(T) method has also been used to calculate the AEA of $OH\cdot(H_2O)_n$ for n=1-5 and extrapolated to n=8 using Eq. (1). The CCSD(T) values suggest that the binding of the extra electron is over-estimated by MP2 and greatly under-estimated by HF. Figure 10 plots the HF/EFP1, MP2 and CCSD(T) absolute AEA values as a function of the number of water molecules n to illustrate the change in the electron affinity as n increases. The HF and MP2 AEA appear to be converging slowly to a constant value, whereas this is not apparent for the CCSD(T) curve. Table 4 and Figure 10 again illustrate the importance of electron correlation in predicting the electron affinity of OH. As shown in Table 4, there is a significant CCSD(T) basis set effect for the $OH\cdot(H_2O)_n$ AEA values, similar to the effect on the gas phase values illustrated in Table 3. In addition, the CCSD(T)/aug-cc-pVTZ AEA appears to be converging more slowly than that for the smaller basis set. Although calculating the CCSD(T)/aug-cc-pVTZ AEA values for the larger values of n would be beyond the computational capabilities, the results for n = 0, 1, 2 suggest a slower convergence than that illustrated in Figure 10.

As shown in Table 4, at n=15, the MP2 AEA is 101.5 kcal/mol, whereas the HF/EFP1 value is 56.0 kcal/mol. As noted in the Introduction, previously calculated DFT + COSMO values for the $OH\cdot$ electron affinity are in the range 142-145 kcal/mol[15, 28]. This is much larger than the 101.5 kcal/mol MP2 value for 15 water molecules. However, although the HF and MP2 electron affinities appear to be converging to a value that is much smaller than the 142-145 kcal/mol predicted by DFT, the CCSD(T)

vales may be converging more slowly. Among the possible reasons for the discrepancy between the solvated electron affinities predicted previously and those predicted in the present work are (a) the higher levels of theory (MP2 and CCSD(T)) employed in the present work and (b) the possible inability of small clusters to sufficiently reproduce the bulk solvent effects. In order to explore the latter possibility, additional calculations were performed on the n = 15 clusters using the PCM method. The results are given in parentheses in Table 4. At the MP2 level of theory, the predicted AEA is -130.6 kcal/mol. This result illustrates that 15 explicit water molecules are insufficient to account for bulk solvation. On the other hand, the coupled cluster AEA values are consistently smaller than those predicted by MP2. So, assuming that PCM adequately accounts for the bulk effects and that PCM will not have a much larger effect on the CCSD(T) clusters than on the MP2 clusters, it would seem likely that the n = 15 MP2 + PCM AEA is an upper limit to the AEA in aqueous solution.

D. Binding energies: The differential binding energies (DBE) and the total binding energies (TBE) to bind the solvent water molecules to the solute molecule have been calculated for both $OH\cdot(H_2O)_{n(=1\text{-}15)}$ and $OH^-(H_2O)_{n(1\text{-}15)}$ using Eqs. (3) and (4) respectively, at the HF/EFP1 and MP2 levels of theory.

Differential binding energy (DBE): $E[A(H_2O)_n] - E[A(H_2O)_{n-1}] - E(H_2O)$ (3)

Total binding energy (TBE): $E(A(H_2O)_n) - \{E(A) + nE(H_2O)\}$ (4)

In Eqs (3) and (4), 'A' represents the solute (hydroxyl radical or hydroxide anion) and $E[A(H_2O)_n]$ and $E[A(H_2O)_{n-1}]$ are the Boltzmann averaged energies calculated using the following equation.

$$E_n = \frac{\sum_i X_i Exp(-\Delta E_i / RT)}{\sum_i Exp(-\Delta E_i / RT)} \qquad (5)$$

Where X_i is the energy of the i^{th} structure including the zero point energy correction. ΔE_i is the energy difference between the i^{th} and the global minimum structure at a particular value of n (T=298K).

Table 5 illustrates the DBE values of $OH^-(H_2O)_n$ calculated at the HF/EFP1, MP2, and CCSD(T) levels of theory. The calculated and experimental DBE[9,58,59] values decrease as the number of water molecules increases up to n~5, after which the values oscillate. The CCSD(T) DBE are in good agreement with experiment. This agreement would very likely improve with the use of better basis sets. The MP2 DBE values agree qualitatively with the experimental values, with percentage errors in the range of ~3% – 17%. The percentage errors for the HF/EFP1 DBE values relative to the experimental values are ~3.0% – 30%. All three levels of theory are in qualitative agreement with the experimentally observed trend. Beyond n~5, the MP2 DBE oscillates around the experimental value of -11 to -12 kcal/mol for the larger values of n.

The absolute values for the calculated and experimental TBE[9,59] values for $OH^-(H_2O)_n$, listed in Table 6, increase with n. As one would expect, the MP2 TBE are in better agreement with experiment than those based on HF, and the CCSD(T) values are in the best agreement with experiment, even with the modest basis set that was used here. The MP2 clusters are over-bound, as one would expect.

Table 7 lists the HF/EFP1, MP2 and CCSD(T) DBE and TBE values for $OH\cdot(H_2O)_n$ clusters. The HF/EFP1 and MP2 DBE values increase until n~3-4 and then fluctuate as n

increases further. The CCSD(T) DBEs increase with the cluster size, with some fluctuation observed at n=3. Smaller $OH\cdot(H_2O)_n$ DBE values compared with those for $OH^-(H_2O)_n$ clusters are consistent with the favorable clustering of the water molecules around the anion and the reticent behavior of the OH radical to H-bond with the water molecules. The TBE values of $OH\cdot(H_2O)_n$ clusters are also shown in Table 7, and the HF/EFP1, MP2 and CCSD(T) values increase as the number of water molecules increases. HF/EFP1 underestimates the binding energies compared to MP2, but the trend is qualitatively correct. The $OH\cdot(H_2O)_n$ HF/EFP1 TBE values are similar to those of the analogous water clusters, suggesting a structural similarity between the two systems. The agreement between the TBE values of the two systems is within 1-3 kcal/mol. (See Figure 9 and Section B for the structural comparison between the two systems)

IV. Conclusions

A systematic study of the adiabatic electron affinity (AEA) of the micro-solvated hydroxyl radical ($OH\cdot(H_2O)_n$) is presented at the HF/EFP1 and correlated levels of theory (MP2 and CCSD(T)). Up to 15 explicit water molecules are added to $OH\cdot$ using the EFP explicit solvent model. The results indicate that with increasing number of water molecules, the HF/EFP1 AEA values increase at a faster rate initially and then appear to slowly converge to -56.0 kcal/mol at n=15. MP2 AEA values follow the same trend as HF/EFP1 method. However, the MP2 AEA converges at a much higher absolute value (-101.5 kcal/mol at n=15) than HF/EFP1, illustrating the importance of the electron correlation in the AEA calculations. When the continuum solvent model

(PCM) is included in the calculations with 15 explicit MP2 water molecules, the AEA exhibits a further large increase to -130.6 kcal/mol. This large increase illustrates the fact that 15 water molecules is not sufficient to converge the calculated AEA to the value one would observe in bulk aqueous solution. Since CCSD(T) predicts smaller AEA for all values of n for which this level of theory is available, the AEA in bulk aqueous solution is likely to be smaller (in an absolute sense) than the MP2 value of -130.6 kcal/mol.

As noted in the Introduction, Mundy et al[15] reported the OH electron affinity to be -142 kcal/mol, based on DFT combined with a continuum solvent model. The present study, using MP2 for $OH\cdot(H_2O)_{15}$ and $OH^-(H_2O)_{15}$ clusters combined with the PCM continuum method predicts an AEA value of -130 kcal/mol. It is likely that the incorporation of explicit solvent molecules to account for the solute-solvent hydrogen bonding contributes to the difference between these the two studies.

The analysis of $OH^-(H_2O)_n$ indicate that the first global minimum that is completely solvated occurs at n = 15. Both HF/EFP1 and MP2 methods predict that the hydroxide ion accepts up to 6 hydrogen bonds from the surrounding water molecules and donates none until the 15th water molecule is added. The observation of a hyper-coordination at the oxygen site of the anion (OH^-) and weak H-bond formation via the hydrogen site of the anion is consistent with previous experimental studies.

The structural analysis of $OH\cdot(H_2O)_n$ clusters suggests that all of the global minimum structures calculated using HF/EFP1 and MP2 are surface structures; no interior global minimum structures were found. The structural analysis also illustrates

considerable similarity between $OH\cdot(H_2O)_n$ and $(H_2O)_n$ global minimum structures. Furthermore, the two systems also have comparable binding energies. This is consistent with several previous theoretical studies, including that of Khalack et al[23] in which a similarity between the radial distribution functions of $OH\cdot(H_2O)_n$ and $(H_2O)_n$ was observed. The structural comparison of $OH\cdot(H_2O)_n$ and $(H_2O)_n$ clusters illustrates that the OH radical is a weak H-bond acceptor and a stronger H–bond donor than the water molecule, in agreement with the previous DFT studies by Couto et al.[21] The systematic study presented in this work also shows that a computationally cost effective and an explicit solvent model, EFP, is able to capture the correct structural features of the micro-solvated hydroxide ion and micro-solvated hydroxyl radical.

Acknowledgements

This research was supported by a grant from the US Department Of Energy, Office of Science, Basic Energy Sciences, administered by the Ames Laboratory, Iowa State University.

References

(1) Turco, R.; Plumb, A.; Condon, E. *Geophysical Research Letters* **1990**, *17*, 313.
(2) Rosswall, T. *Environmental Science & Technology* **1991**, *25*, 567.
(3) Wennberg, P. O. *Nature* **2006**, *442*, 145.
(4) Zafiriou, O. C.; McFarland, M.; Bromund, R. H. *Science* **1980**, *207*, 637.
(5) Jurva, U.; Wikstrom, H. V.; Bruins, A. P. *Rapid Commun Mass Spectrom* **2002**, *16*, 1934.
(6) Mattson, M. P. *advances in cell ageing and gerontoloy*; New York: Elsevier, 2001; Vol. 4.
(7) Linert, W.; Jameson, G. N. L. *Journal of Inorganic Biochemistry* **2000**, *79*, 319.
(8) Dorfman, L. M.; Adams, G. E. *Reactivity of Hydroxyl Radical in Aqueous Solutions*, 1973.

(9) Arshadi, M.; Kebarle, P. *Journal of Physical Chemistry* **1970**, *74*, 1483.
(10) Halliwell. B, K. J. H. C. G. *Free Radicals in Biology and Medicine*; Oxford Univrsity Press. Oxford, 1989.
(11) Koppenol, W. H. *Free Radical Biology and Medicine* **1991**, *10*, 85.
(12) Cadet, J.; Delatour, T.; Douki, T.; Gasparutto, D.; Pouget, J. P.; Ravanat, J. L.; Sauvaigo, S. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **1999**, *424*, 9.
(13) Bamatraf, M. M. M.; O'Neill, P.; Rao, B. S. M. *Journal of Physical Chemistry B* **2000**, *104*, 636.
(14) Douki, T.; Spinelli, S.; Ravanat, J. L.; Cadet, J. *Journal of the Chemical Society-Perkin Transactions 2* **1999**, 1875.
(15) Mundy, C. J.; Colvin, M. E.; Quong, A. A. *Journal of Physical Chemistry A* **2002**, *106*, 10063.
(16) Janik, I.; Bartels, D. M.; Jonah, C. D. *Journal of Physical Chemistry A* **2007**, *111*, 1835.
(17) Tsai, M. K.; Kowalski, K.; Valiev, M.; Dupuis, M. *Journal of Physical Chemistry A* **2007**, *111*, 10478.
(18) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *Journal of Physical Chemistry A* **2000**, *104*, 4811.
(19) Hirata, S.; Head-Gordon, M. *Chemical Physics Letters* **1999**, *314*, 291.
(20) KIM, K. S.; Kim, H. S.; Jang, J. H.; Kim, H. S.; Mhin, B. J.; Xie, Y. M.; Schaefer, H. F. *Journal of Chemical Physics* **1991**, *94*, 2057.
(21) do Couto, P. C.; Guedes, R. C.; Cabral, B. J. C.; Simoes, J. A. M. *Journal of Chemical Physics* **2003**, *119*, 7344.
(22) Hamad, S.; Lago, S.; Mejias, J. A. *Journal of Physical Chemistry A* **2002**, *106*, 9104.
(23) Khalack, J. M.; Lyubartsev, A. P. *Journal of Physical Chemistry A* **2005**, *109*, 378.
(24) XIE, Y. M.; Schaefer, H. F. *Journal of Chemical Physics* **1993**, *98*, 8829.
(25) Wang, B. S.; Hou, H.; Gu, Y. S. *Chemical Physics Letters* **1999**, *303*, 96.
(26) Zhou, Z. Y.; Qu, Y. H.; Fu, A. P.; Du, B. N.; He, F. X.; Gao, H. W. *International Journal of Quantum Chemistry* **2002**, *89*, 550.
(27) Vassilev, P.; Louwerse, M. J.; Baerends, E. J. *Journal of Physical Chemistry B* **2005**, *109*, 23605.
(28) Adriaanse, C.; Sulpizi, M.; VandeVondele, J.; Sprik, M. *Journal of the American Chemical Society* **2009**, *131*, 6046.
(29) Adamo, C.; Barone, V. *Chemical Physics Letters* **1997**, *274*, 242.
(30) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *Journal of Chemical Physics* **1996**, *105*, 1968.
(31) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *Journal of Physical Chemistry A* **2001**, *105*, 293.
(32) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* **1953**, *21*, 1087.
(33) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *Journal of Chemical Physics* **2000**, *112*, 2063.
(34) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(35) Hehre, W. J.; Ditchfie.R; Pople, J. A. *Journal of Chemical Physics* **1972**, *56*, 2257.

(36) Harihara.Pc; Pople, J. A. *Theoretica Chimica Acta* **1973**, *28*, 213.
(37) Pople, J. A. *Isr J Chem* **1993**, *33*, 345.
(38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *Journal of Computational Chemistry* **1993**, *14*, 1347.
(39) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *International Journal of Quantum Chemistry* **1992**, 879.
(40) Barone, V.; Cossi, M. *Journal of Physical Chemistry A* **1998**, *102*, 1995.
(41) Miertus, S.; Scrocco, E.; Tomasi, J. *Chemical Physics* **1981**, *55*, 117.
(42) Cances, E.; Mennucci, B.; Tomasi, J. *Journal of Chemical Physics* **1997**, *107*, 3032.
(43) Li, H.; Jensen, J. H. *Journal of Computational Chemistry* **2004**, *25*, 1449.
(44) Mulliken, R. S. *Journal of Chemical Physics* **1955**, *23*, 2343.
(45) Mulliken, R. S. *Journal of Chemical Physics* **1955**, *23*, 2338.
(46) Mulliken, R. S. *Journal of Chemical Physics* **1955**, *23*, 1833.
(47) Mulliken, R. S. *Journal of Chemical Physics* **1955**, *23*, 1841.
(48) Botti, A.; Bruni, F.; Imberti, S.; Ricci, M. A.; Soper, A. K. *Journal of Chemical Physics* **2003**, *119*, 5001.
(49) Aziz, E. F.; Ottosson, N.; Faubel, M.; Hertel, I. V.; Winter, B. *Nature* **2008**, *455*, 89.
(50) Aikens, C. M.; Gordon, M. S. *Journal of the American Chemical Society* **2006**, *128*, 12835.
(51) Wood, G. P. F.; Gordon, M. S.; Radom, L.; Smith, D. M. *Journal of Chemical Theory and Computation* **2008**, *4*, 1788.
(52) Mullin, J. M.; Gordon, M. S. *Journal of Physical Chemistry B* **2009**, *113*, 14413.
(53) Mullin, J. M.; Gordon, M. S. *Journal of Physical Chemistry B* **2009**, *113*, 8657.
(54) Engdahl, A.; Karlstrom, G.; Nelander, B. *Journal of Chemical Physics* **2003**, *118*, 7797.
(55) Smith, J. R.; Kim, J. B.; Lineberger, W. C. *Physical Review A* **1997**, *55*, 2036.
(56) Schulz, P. A.; Mead, R. D.; Jones, P. L.; Lineberger, W. C. *Journal of Chemical Physics* **1982**, *77*, 1153.
(57) Chipman, D. M. *Journal of Chemical Physics* **1986**, *84*, 1677.
(58) Merrill, G. N.; Webb, S. P. *Journal of Physical Chemistry A* **2003**, *107*, 7852.
(59) Meotner, M.; Speller, C. V. *Journal of Physical Chemistry* **1986**, *90*, 6616.

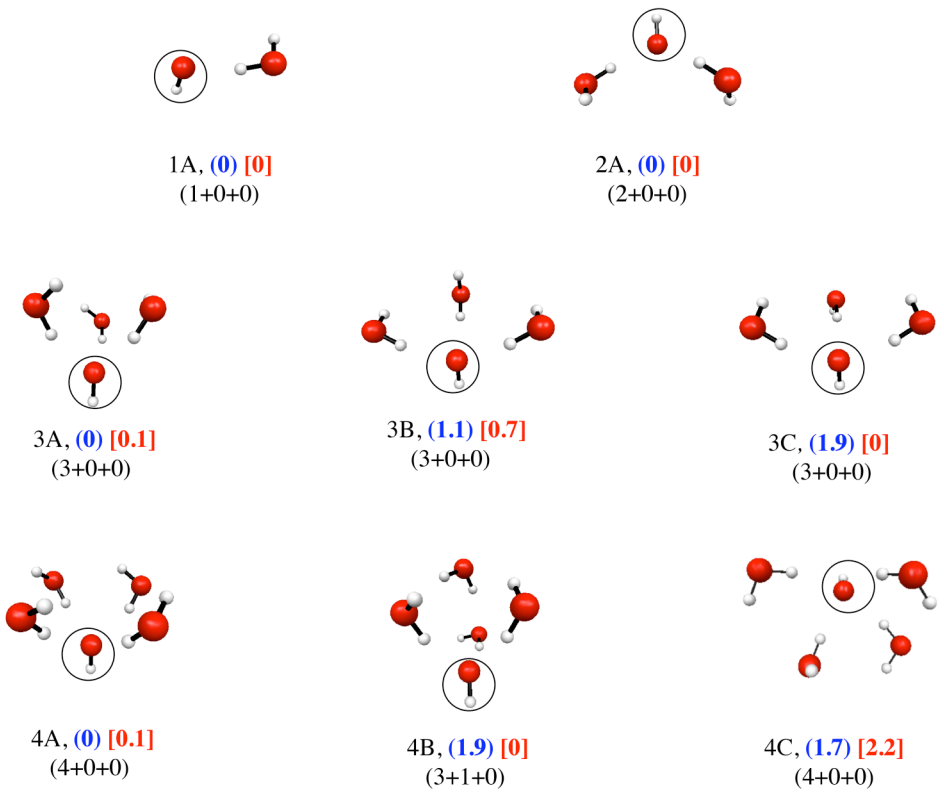


Figure 1: Lowest energy structures of $OH^-(H_2O)_n$ for n=1-4. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y) and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH anion is circled in all of the structures.

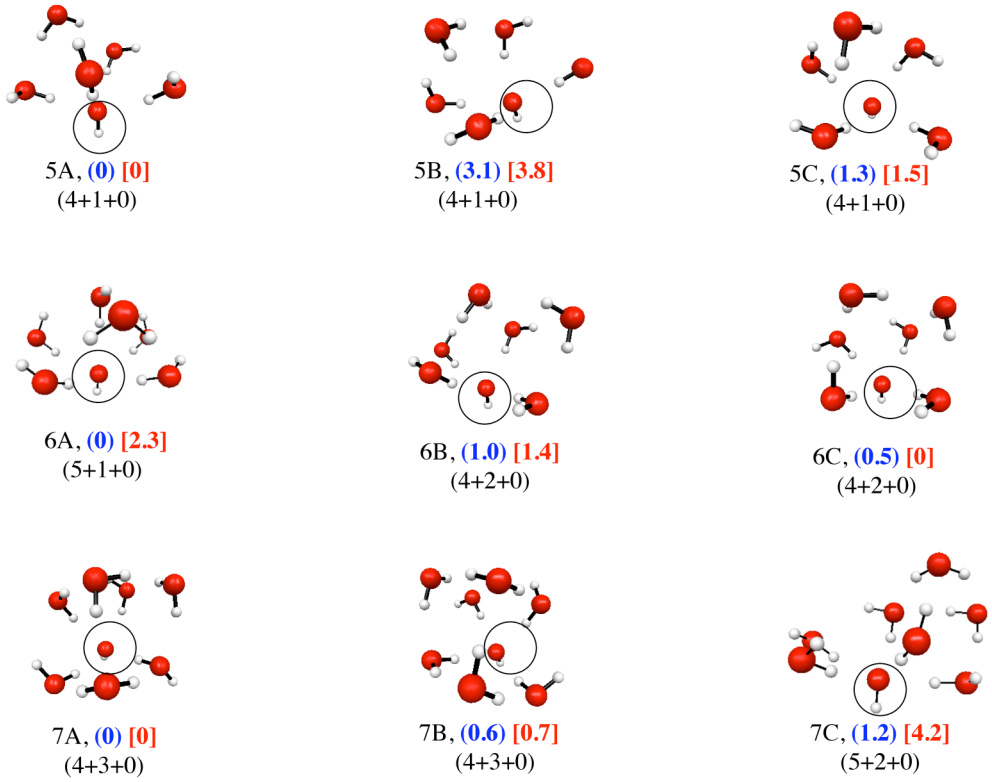


Figure 2: Lowest energy structures of $OH^-(H_2O)_n$ for n=5-7. (X+Y+Z) indicates the number of solvent molecules in the first shell (X), second shell (Y) and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH anion is circled in all of the structures.

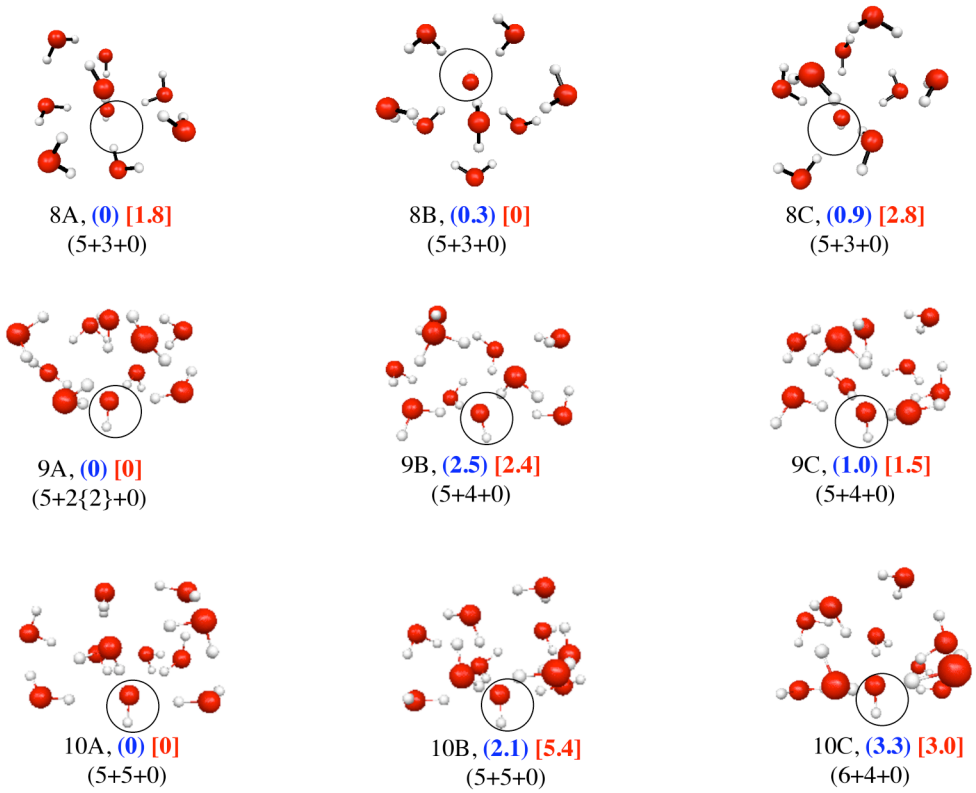


Figure 3: Lowest energy structures of $OH^-(H_2O)_n$ for n=8-10. (X+Y+Z) indicates the number of solvent molecules in the first shell (X), second shell (Y) and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH anion is circled in all of the structures.

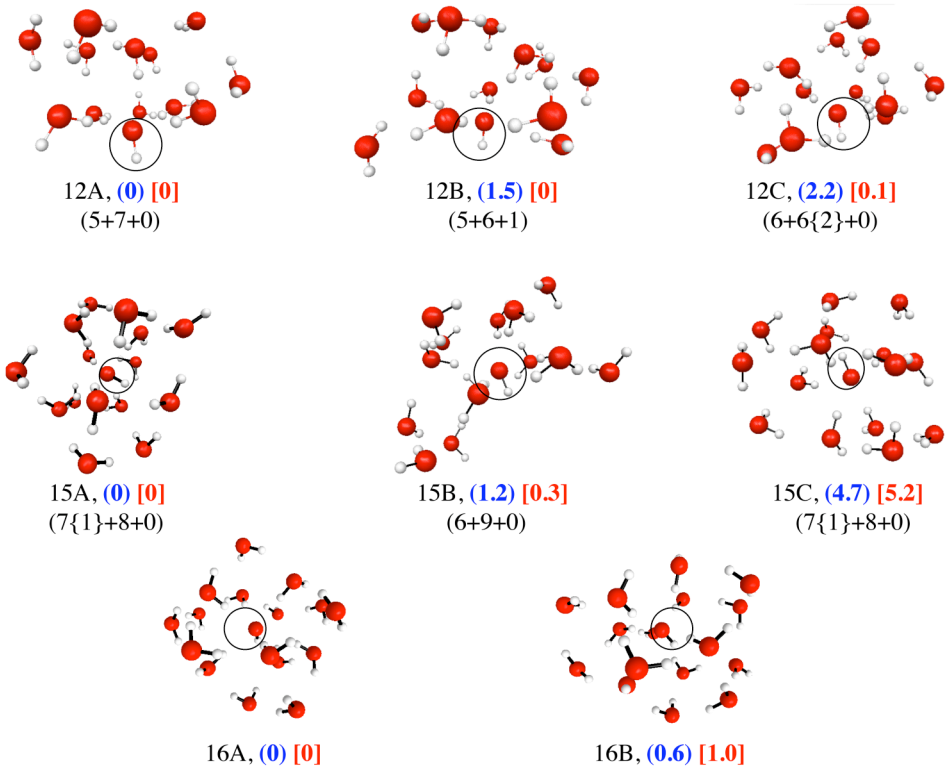


Figure 4: Lowest energy structures of $OH^-(H_2O)_n$ for n=12-16. (X+Y+Z) indicates the number of solvent molecules in the first shell (X), second shell (Y) and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH anion is circled in all of the structures.

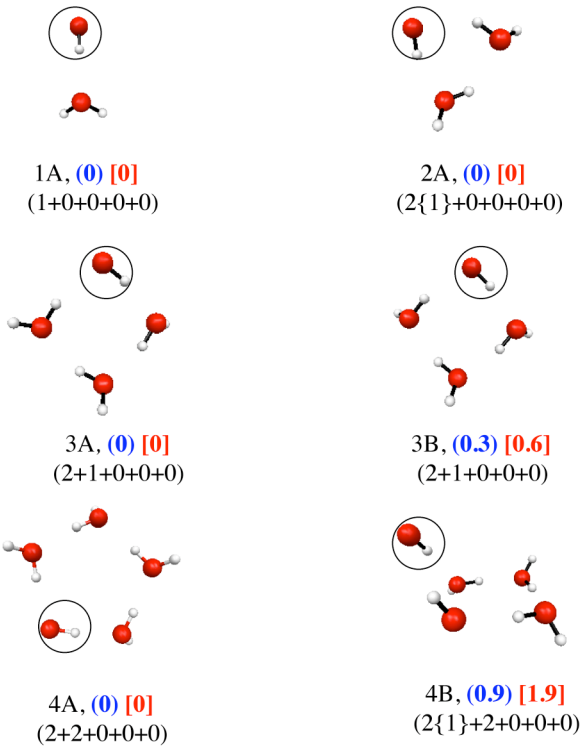


Figure 5: Lowest energy structures of $OH^-(H_2O)_n$ for n=1-4. (X+Y+Z+W+V) indicates the number of solvent molecules in first shell (X), second shell (Y), third shell (Z), fourth shell (W) and fifth shell (V). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH radical is circled in all of the structures.

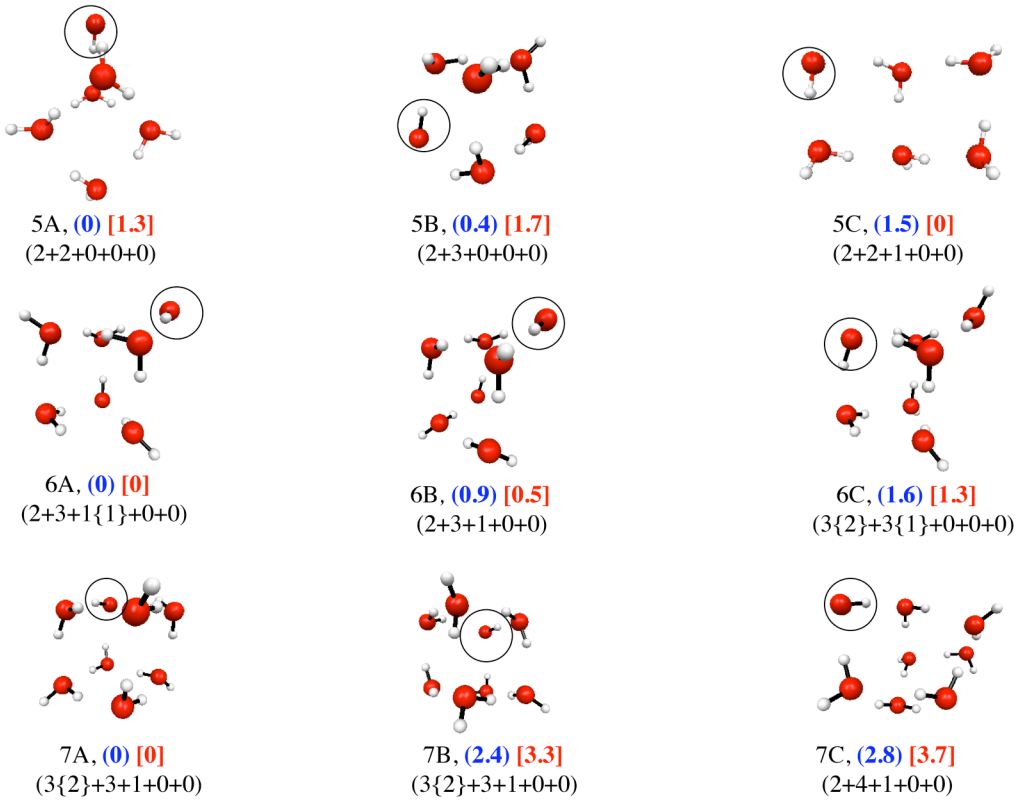


Figure 6: Lowest energy structures of $OH^-(H_2O)_n$ for n=5-7. (X+Y+Z+W+V) indicates the number of solvent molecules in first shell (X), second shell (Y), third shell (Z), fourth shell (W) and fifth shell (V). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH radical is circled in all of the structures.

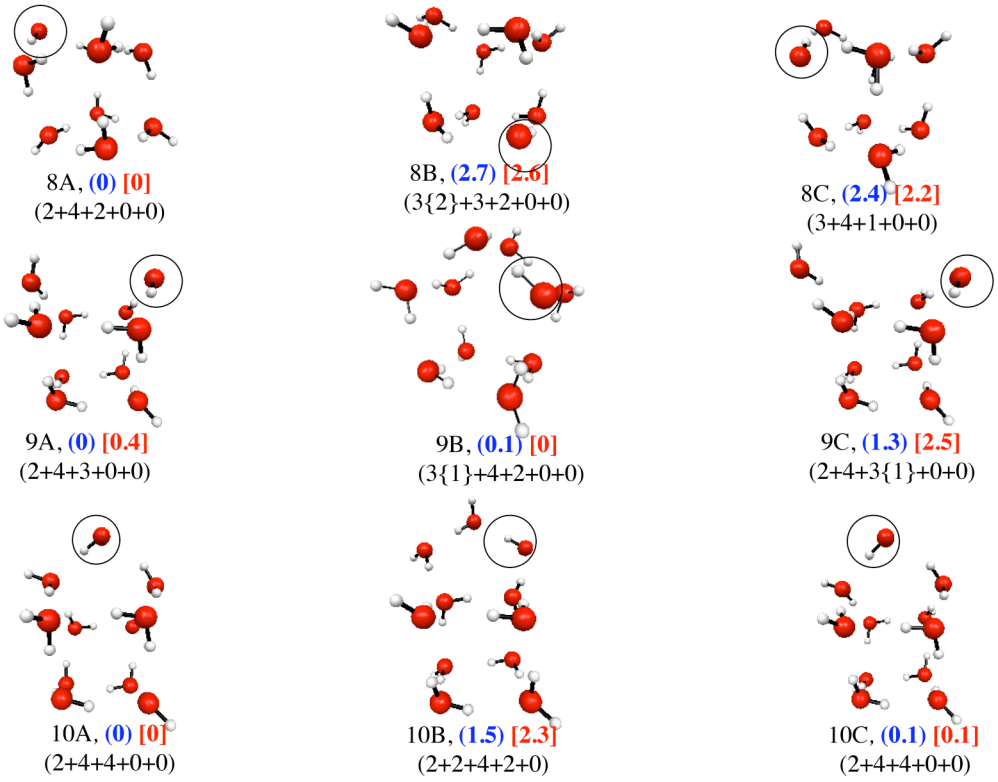


Figure 7: Lowest energy structures of $OH^-(H_2O)_n$ for n=8-10. (X+Y+Z+W) indicates the number of solvent molecules in first shell (X), second shell (Y), third shell (Z), fourth shell (W) and fifth shell (V). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH radical is circled in all of the structures.

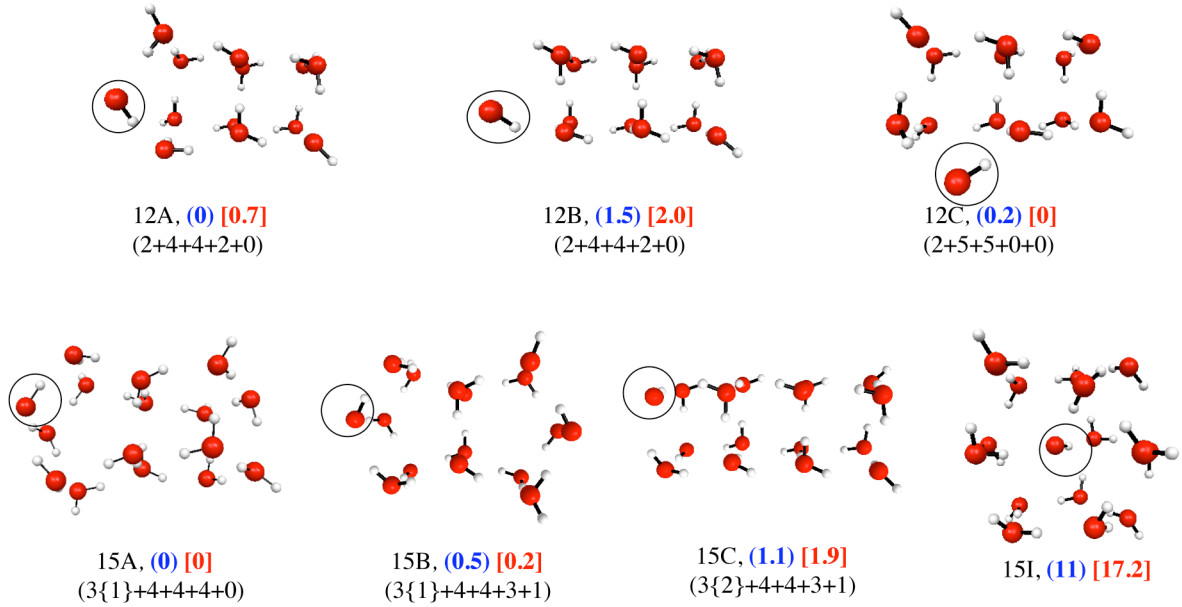


Figure 8: Lowest energy structures of $OH^-(H_2O)_n$ for n=12-15. (X+Y+Z+W) indicates the number of solvent molecules in first shell (X), second shell (Y) third shell (Z), fourth shell (W) and fifth shell (V). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) or [0] relative energy suggests the global minimum structure. The OH radical is circled in all of the structures.

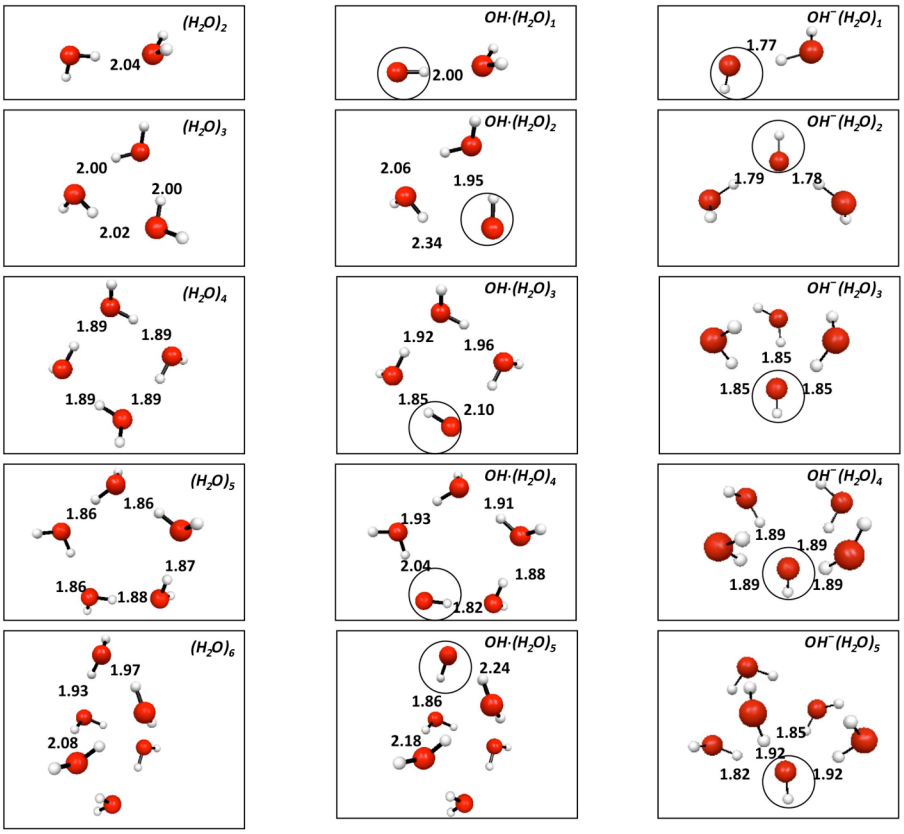


Figure 9: Structures and bond lengths of $(H_2O)_n$, $OH\cdot(H_2O)_n$ and $OH^-(H_2O)_n$ clusters. The geometries were obtained using HF/EFP1 level of theory. The OH radical/OH anion is circled in all of the structures.

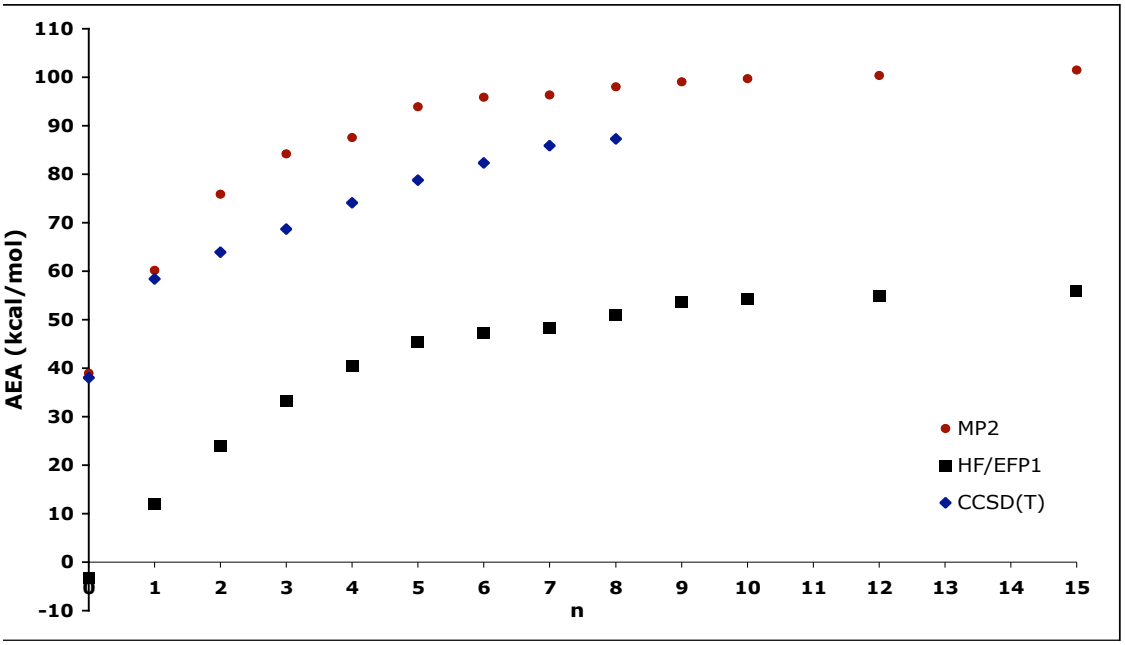


Figure 10: Plot of adiabatic electron affinity (AEA) of $OH{\cdot}(H_2O)_n$ with increasing number of water molecules (n). Squares indicate HF/EFP1 values. Circles represent the full ab initio MP2 values. Diamonds are for the CCSD(T) values. The n=6-8 AEA values at CCSD(T) level of theory are extrapolated using Eq. (1) (see text).

Table 1: The average MP2 Mulliken charges and H-bond distances for the global minimum structures of $OH^-(H_2O)_n$.

number of water molecules	R(Å)		Mulliken charges (a.u)						
	HF/EFP1	MP2	(OH^-)			(H_2O)			
			q(O)	q(H)	q(OH^-)	q(O)	q(H_A)	q(H_B)	q(H_2O)
0			-1.21	0.21	-1.00	-0.71	0.35	0.35	0.00
1	1.77	1.40	-1.12	0.25	-0.87	-0.90	0.50	0.27	-0.05
2	1.79	1.55	-1.19	0.29	-0.91	-0.84	0.49	0.30	-0.05
3	1.85	1.67	-1.28	0.32	-0.96	-0.82	0.48	0.33	-0.01
4	1.89	1.76	-1.27	0.33	-0.94	-0.81	0.44	0.35	-0.02
5	1.88	1.75	-1.28	0.33	-0.95	-0.84	0.46	0.37	-0.20
6	1.94	1.83	-1.32	0.35	-0.97	-0.82	0.44	0.38	-0.12
7	1.85	1.73	-1.25	0.34	-0.90	-0.88	0.47	0.38	-0.01
8	1.91	1.81	-1.39	0.37	-1.03	-0.88	0.48	0.40	0.12
9	1.92	1.82	-1.40	0.36	-1.04	-0.92	0.47	0.43	0.01
10	1.94	1.83	-1.30	0.36	-0.94	-0.91	0.46	0.43	-0.01
12	1.89	1.80	-1.28	0.37	-0.91	-0.95	0.46	0.44	-0.01
15	2.02	1.88	-1.51	0.37	-1.15	-0.90	0.41	0.47	0.01

R is the average H-bond distance between the hydroxide ion and the first shell solvent molecules. q(O) is the Mulliken charge on oxygen atom of OH^-. q(OH^-) is the average Mulliken charge on hydroxide ion. q(H_A) and q(H_B) represent the average Mulliken charges on hydrogen atoms, of H_2O, that are pointing towards the ion and away from the ion respectively. q(H_2O) is the average Mulliken charge on a water molecule.

Table 2: The calculated binding energies of water dimer and OH-water dimer. All the values are in kcal/mol.

	CCSD(T)	
	aug-cc-pVTZ	aug-cc-pVQZ
water-dimer	-5.22	-5.17
OH (donor)-water	-5.78	-5.30
OH (acceptor)-water	-1.50	-1.05

Table 3: Adiabatic electron affinity of gas phase hydroxyl radical at different levels of theory. All values are in kcal/mol

HF/EFP1	MP2		CCSD(T)			Expt[a]
6-31++G(d,p)	6-31++G(d,p)	6-311++G(2d,2p)	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	
+3.3	-39.0	-42.0	-38.1	-40.0	-41.4	-42.0

[a]Data taken from ref. 55 and 56.

Table 4: Adiabatic electron affinity (kcal/mol) of $OH^{\bullet}(H_2O)_n$ for n=0-15. The CCSD(T) values for n=6-8 are the extrapolated values using equation 1 (see text). The values in parentheses are obtained by adding bulk solvent effects using the PCM solvent model. The previously calculated value of AEA of hydroxyl in aqueous phase using a continuum model is 142.3 kcal/mol (Ref 15).

number of water molecules	HF/EFP1	MP2	CCSD(T)	
	6-31G++(d,p)		aug-cc-pVDZ	aug-cc-pVTZ
0	+3.3	-39.0	-38.1	-40.4
1	-12.0	-60.2	-58.4	-61.8
2	-24.0	-75.9	-63.9	-70.6
3	-33.3	-84.3	-68.7	
4	-40.4	-87.6	-74.1	
5	-45.5	-93.9	-78.9	
6	-47.2	-95.9	-82.3	
7	-48.3	-96.3	-85.9	
8	-51.0	-98.1	-87.3	
9	-53.7	-99.1		
10	-54.3	-99.8		
12	-55.0	-101.4		
15	-56.0 (-86.7)	-101.5 (-130.6)		

Table 5: HF/EFP1, MP2 and CCSD(T) differential binding energies for $OH^-(H_2O)_n$

number of water molecules	Expt[a]	HF/EFP1		MP2		CCSD(T) aug-cc-pvDZ	
	binding energy (kcal/mol)	binding energy (kcal/mol)	%error[b]	binding energy (kcal/mol)	%error[b]	binding energy (kcal/mol)	%error[b]
1	-26.5+1.0	-18.4	-30.7	-25.6	3.0	-26.3	0.8
2	-17.6+1.0	-17.1	-2.8	-20.6	17.2	-18.1	2.8
3	-16.2+1.0	-15.4	-5.1	-16.7	3.5	-15.7	-2.8
4	-12.3+1.0	-13.1	9.3	-14.0	13.8	-14.1	14.7
5	-11.5+1.0	-11.2	-2.7	-12.5	9.2	-12.1	5.2
6	-11.2+1.0	-8.1	-28.1	-12.6	12.7		
7		-12.4		-12.8			
8		-9.9		-10.3			
9		-8.5		-11.9			
10		-7.3		-6.9			
12		-8.5		-10.7			
15		-7.3		-11.0			

[a]Data taken from ref 59. [b]The percentage errors are calculated by taking the difference of the calculated value and the experimental value at each value of n. All the values are in kcal/mol.

Table 6: HF/EFP1, MP2 and CCSD(T) total binding energies for $OH^-(H_2O)_n$.

number of water molecules	Expt[a]	HF/EFP1		MP2		CCSD(T) aug-cc-pVDZ	
	binding energy (kcal/mol)	binding energy (kcal/mol)	%error[b]	binding energy (kcal/mol)	%error[b]	binding energy (kcal/mol)	%error[b]
1	-26.5+1.0	-18.4	-30.7	-25.6	3.0	-26.3	0.8
2	-44.1+1.4	-35.5	-19.6	-46.3	5.0	-44.4	-0.6
3	-60.3+1.7	-50.8	-15.7	-63.0	4.6	-60.1	0.3
4	-72.3+2.0	-63.9	-11.5	-77.1	6.6	-74.2	-2.7
5	-83.8+2.2	-75.1	-10.3	-89.6	7.0	-86.3	-3.0
6	-95.0+2.4	-83.2	-12.4	-102.3	7.7		
7		-95.6		-115.0			
8		-105.4		-125.4			
9		-113.9		-137.3			
10		-121.2		-144.1			
12		-138.1		-165.6			
15		-160.1		-198.6			

[a]Data taken from ref 59. [b]The percentage errors are calculated by taking the difference of the calculated value and the experimental value at each value of n.

Table 7: HF/EFP1, MP2 and CCSD(T) binding energies (kcal/mol) for $OH^{\bullet}(H_2O)_n$

number of water molecules	Differential binding energies $OH^{\bullet}(H_2O)_n$			Total binding Energies $OH^{\bullet}(H_2O)_n$			$(H_2O)_n$
	HF/EFP1	MP2	CCSD(T) aug-cc-pVDZ	HF/EFP1	MP2	CCSD(T) aug-cc-pVDZ	HF/EFP1
1	-4.0	-4.5	-4.2	-4.0	-4.5	-4.2	-2.4
2	-5.1	-4.9	-5.0	-9.0	-9.4	-9.1	-8.4
3	-7.6	-8.2	-10.3	-16.5	-17.6	-19.4	-16.4
4	-5.0	-9.5	-7.6	-21.5	-27.1	-27.0	-22.6
5	-6.9	-8.3	-8.6	-28.5	-35.4	-35.6	-28.2
6	-6.4	-10.5		-34.8	-45.9		-36.0
7	-6.8	-8.5		-41.6	-54.4		-45.5
8	-9.7	-11.7		-51.3	-66.2		-52.1
9	-6.5	-9.9		-57.8	-76.1		-60.4
10	-6.0	-8.8		-63.9	-84.9		-65.3
12	-8.1	-10.9		-80.1	-106.7		-80.9
15	-7.3	-9.8		-102.0	-135.9		-104.0

CHAPTER 4: DIFFUSION OF ATOMIC OXYGEN ON THE Si(100) SURFACE

A paper to be submitted to The Journal of Physical Chemistry

Pooja Arora, Wei Li, Piotr Piecuch, James W. Evans, Marvin Albao, and Mark S. Gordon

Abstract

The processes of etching and diffusion of atomic oxygen on the reconstructed Si(100) – 2 x 1 surface are investigated using an embedded cluster QM/MM (Quantum Mechanics/Molecular Mechanics) method, called SIMOMM (Surface Integrated Molecular Orbital Molecular Mechanics). Hopping of an oxygen atom along the silicon dimer rows on a $Si_{15}H_{16}$ cluster embedded in an $OSi_{136}H_{92}$ MM cluster model is studied using the SIMOMM/UB3LYP (unrestricted density functional theory (UDFT) method with the Becke three-parameter Lee-Yang-Parr (B3LYP) hybrid functional), the Hay-Wadt effective core potential and its associated double zeta plus polarization basis set. The relative energies at stationary points on the diffusion potential energy surface were also obtained with three coupled-cluster (CC) methods, including the canonical CC approach with singles, doubles, and non-iterative quasi-perturbative triples (CCSD(T)), the canonical left-eigenstate completely renormalized (CR) analog of CCSD(T), termed CR-CC(2,3), and the linear scaling variant of CR-CC(2,3) employing the cluster-in-molecule (CIM) local correlation ansatz, abbreviated as CIM-CR-CC(2,3). The pathway and energetics for the diffusion of oxygen from one dimer to another are presented, with the activation energy estimated to be 71.9 and 79.5 kcal/mol at the canonical CR-CC(2,3)/6-31G(d) and CIM-CR-CC(2,3)/6-

311G(d) levels of theory, respectively. The canonical and CIM CR-CC(2,3)/6-31G(d) barrier heights (excluding zero point vibrational energy contributions) for the etching process are both 87.3 kcal/mol.

I. Introduction

The silicon atoms of the topmost layer of the Si(100) surface can form covalent bonds with adjacent surface atoms to form pairs called silicon dimers. The surface reconstructs by forming silicon dimer rows as shown in Figure 1. Even after the reconstruction, the dimer Si atoms are bonded to only three other atoms. Since Si does not readily form π bonds, these surface Si atoms are highly reactive. A common terminology is to say that the surface Si atoms have "dangling bonds".[1]

The Si(100)-2x1 surface oxidizes to form silicon dioxide (SiO_2) by thermal oxidation. Silicon dioxide is known to be an insulator[2,3] which has technological importance in the fabrication/doping of micro electronic devices.[4-7] The silicon dioxide film formed on the silicon surface due to thermal oxidation acts to block a dopant from reaching the silicon surface.

Several studies have been reported on the oxidation of the silicon surface.[8-12] Depending upon the surface temperature and oxygen pressure, active oxidation/etching by oxygen or passive oxidation/oxide formation can occur.[13,14] At low-T or high-P, passive oxidation is observed, resulting in the formation of an oxide film on the surface. At high-T or low-P, removal of Si by desorption of volatile SiO (etching) is observed. Controlling the process of oxidation and etching by varying the conditions is very important to attain uniformity and

precision in the size and shape of micro-electronic devices, including transistors and capacitors that are made of semiconductor material such as silicon.[7,15,16] Therefore, understanding the mechanisms of adsorption, diffusion, and desorption of oxygen on the silicon surface at the atomic level is crucial.

The main goal of this paper is to study the detailed mechanism of the diffusion of atomic oxygen on the Si(100)-2x1 surface from one Si dimer to an adjacent dimer. Repeated hopping between adjacent dimers leads to long-range diffusion of oxygen along dimer rows, a key process for process for oxide island formation. To study this mechanism, a quantum mechanics/molecular mechanics (QM/MM) hybrid approach, referred to as the surface integrated molecular orbital MM (SIMOMM)[17] model, has been used. SIMOMM is a computationally less expensive method than full quantum mechanical calculations on a given size system and can account for the chemistry of the surface atoms as well as (to some degree) the bulk effects in large clusters with reasonable accuracy. The SIMOMM method has been used in several studies on Si(100)[18-20] and SiC(100)[21] surfaces.

An additional goal is to explore the SiO desorption/etching mechanism at higher levels of theory than those that were used earlier, to assess the accuracy of previously reported barrier heights.[22] This involves calculations at three different levels of coupled-cluster (CC) theory, namely, the canonical CC approach with singles (S), doubles (D), and non-iterative quasi-perturbative triples (T), i.e., CCSD(T),[23] the left-eigenstate completely renormalized (CR) analog of canonical CCSD(T), termed CR-CC(2,3),[24] and the linear scaling extension of CR-CC(2,3) employing a suitably modified variant of the local correlation cluster-in-molecule (CIM) ansatz of Refs. 25 and 26, developed in Refs. 27 and 28, termed CIM-CR-CC(2,3).

The two processes of interest here, diffusion and etching are not entirely disconnected, since one key structure (referred to below as the back-bond species) plays a central role in both processes.

Some of the earlier studies of the reaction mechanisms for the oxidation and etching processes have been described[22,29-31] using QM/MM and kinetic Monte Carlo (KMC)[32] simulation methods. The QM/MM method is described above. KMC[32] is a method to analyze the evolution of lattice-gas models describing the configuration of oxygen adatoms and oxide islands on the Si(100) surface, given that rates for all relevant processes are specified. Choi et al.[22] studied the mechanism for the SiO desorption process using the SIMOMM model. The geometries were obtained using the complete active space self-consistent field (CASSCF) method and the effective core potential (ECP) HW(d) basis set. This was augmented by multi-reference second-order perturbation (MRMP2) theory with a mixed basis set: HW(d) for the silicon atoms and the 6-311G(d) basis set for the oxygen atom. These authors calculated the overall barrier for the SiO desorption process to be 89.8 kcal/mol. This is a little higher than the 83.0-87.6 kcal/mol barrier predicted by Uchiyama et al.[33] using plane wave density functional theory (DFT) within the generalized gradient approximation (GGA)[34]. Another KMC simulation predicted the SiO desorption barrier to be in the range 73.8-80.7 kcal/mol,[30] but the modeling incorporated assumptions regarding O diffusion and oxide formation. An experimental value (79.3 kcal/mol) was determined by employing X-Ray photoelectron spectroscopy (XPS) and supersonic molecular beam techniques.[13] In another experiment the etching barrier, as measured by TPD (temperature programmed desorption) studies,[8,14] is in the range of 80-90 kcal/mol. So, the experiments

have reported an approximately 10 kcal/mol range for this barrier height. The experimental uncertainties increase this range.

Figure 2 shows the SiO desorption stationary points (minima and transition states (TSs)) for the etching mechanism proposed by Choi et al.[22] Structure a represents the on-top structure, in which the oxygen atom is on top of one of the silicon dimer atoms; structure c is the back-bond structure, a minimum on the potential energy surface (PES), in which the oxygen atom is bonded to a Si atom of the surface and to a silicon atom in the next layer; b is the TS between the on-top and the back-bond species. Structure g is formed just before the final etched product of the SiO desorption mechanism is formed. The final etched product is also referred to as SOLA[21] (Si surface with One Less silicon Atom) and is formed after one silicon atom has been removed by an oxygen atom.

Long range diffusion of the oxygen atom between surface dimers was previously studied using the KMC simulation method by Pelz and co-workers.[37] They employed a 55.3-57.7 kcal/mol diffusion activation energy. Another KMC study by Esteve et al. assumed the activation energy to be 57.6 kcal/mol.[38] A plane wave DFT/GGA study by Hemeryck et al. most recently predicted the activation energy for the diffusion process to be 42.8 kcal/mol.[10] These authors used a pseudo-potential,[39,40] and modeled the Si(100) surface as a periodic slab to simulate the bulk. So, there is a 15 kcal/mol range in the theoretical predictions. Experimental scanning tunneling microscopy (STM) measurements were used by Pelz and coworkers[41,42] to study the oxidation of Si(100). They predicted the activation energy for the diffusion process to be 56.3 kcal/mol. Several secondary ion mass spectrometry (SIMS) experiments have also been performed to study the diffusion of oxygen atoms at the Si/SiO2 interface[43,44,45]. The values predicted by SIMS experiments are in the range of 41.5-76.1

kcal/mol. So, the experiments on the diffusion process do not appear to be definitive, nor are the previous DFT and KMC calculations. The central result that arises from most of the previous studies is that an important structure in the diffusion mechanism is the back-bond species. So, the oxygen diffusion process apparently involves a back-bond species.

An important requirement of studies of surface phenomena is to create a model that adequately treats the bulk effects and that accurately takes into account the chemistry of the surface atoms. In the present work, this is accomplished by employing the embedded cluster SIMOMM[17] QM/MM approach to simulate the surface.

II. Computational methods

The SIMOMM model employed here is composed of a QM $Si_{15}H_{16}$ (two-dimer) quantum region embedded in a larger $Si_{136}H_{92}$ MM cluster, as shown in Figure 3. The $Si_{15}H_{16}$ cluster is shown in Figure 4. The QM/MM system is a much bigger cluster containing 9 Si–Si dimers and is 9 layers deep, as shown in Figure 3.

To study the diffusion process, the QM/MM ($OSi_{136}H_{92}$) clusters were optimized using unrestricted density functional theory (UDFT)/SIMOMM with the Becke three-parameter Lee-Yang-Parr (B3LYP) hybrid functional. The HW ECP basis set[46] augmented with one set of d polarization functions was used for the O and Si atoms, and the 3-21G basis set was employed for the H atoms in the QM cluster. The 6-31G(d) basis set[47,48] was also employed to assess basis set effects. The most accurate relative energies characterizing the diffusion pathway were then obtained with the CC methods. As mentioned in the Introduction, the two canonical CC methods used in this study were CCSD(T)[23] and CR-CC(2,3),[24,49,50] both with the 6-31G(d) basis set. The CR-CC(2,3) approach was used because it can account for the

significant diradical character that has been observed in previous work on this surface.[20,51,52] The CC relative energies were calculated at the SIMOMM:UB3LYP/6-31G(d) optimized geometries, using the SIMOMM:UB3LYP/6-31G(d) ZPE corrections as well. Note that the CC calculations were performed only on the QM clusters, since the main effect of the bulk contained in the embedded cluster model will be on the predicted structure.[53]

Recently, the Piecuch group developed two different types of linear scaling extensions of CCSD, CCSD(T), and CR-CC(2,3), based on refined versions of the CIM local correlation ansatz, abbreviated CIM-CCSD, CIM-CCSD(T), and CIM-CR-CC(2,3)[25-28,54]. The CIM-CC methods are characterized by a natural linear scaling of CPU time with the system size, replacing the N^6 and N^7 steps of the canonical CCSD and CCSD(T)/CR-CC(2,3) approaches by steps that scale linearly with the size of the system once the system becomes large enough. The design of the orbital subsystems in the CIM-CC approach[28] makes use of only one threshold parameter for selecting orbital environments for the groups of localized occupied orbitals, called in the CIM-CC methodology "central orbitals". This threshold parameter is designated by ζ and the CIM-CR-CC(2,3) calculation with a given ζ value is abbreviated CIM(ζ)-CR-CC(2,3). For example, CIM(0.005)-CR-CC(2,3) refers to the CIM-CR-CC(2,3) approach with $\zeta = 0.005$. The basis sets used in the mixed CIM-CR-CC(2,3) calculations for the stationary points along the diffusion pathway were 6-31G(d) and 6-311G(d).

To study the SiO desorption/etching process, the QM/MM ($OSi_{136}H_{92}$) structures were optimized using the SIMOMM:CASSCF(12,11) method, where the notation (12,11) means that 12 electrons are distributed in all possible ways among the 11 orbitals in the CASSCF active space. The CASSCF active orbitals for the SiO desorption process are shown

in Figure 5. The active space includes the π and π* SiO orbitals and electrons (a (4,4) space), the bonding and antibonding orbitals of the newly formed Si-Si σ bond (a (2,2) space), the two dangling bonds on silicon (a (2,2) space), and the σ and σ* SiO orbitals and electrons and the 3s orbital on the Si in SiO (a (4,3) space). As noted above, the HW ECP basis set was used for the O and Si atoms, and the 3-21G basis set was used for the H atoms in the QM cluster. In addition to the HW(d) basis set, the 6-31G(d) basis set was also used to optimize the geometry of the QM/MM cluster. In order to incorporate dynamic correlation, single point energy calculations were subsequently performed with the SIMOMM:MRMP2(12,11)/6-31G(d) method at the SIMOMM:CASSCF(12,11)/6-31G(d) geometries. The etching barrier was also calculated using the canonical CCSD and CR-CC(2,3) approaches, and the pure and mixed CIM-CR-CC(2,3) methods, all using the 6-31G(d) basis set at the SIMOMM:CASSCF(12,11)/6-31G(d) geometries. As noted above, the CC calculations were performed only on the QM cluster.

All geometries were fully optimized in both the QM and MM regions. All of the stationary points, including minima and TS structures, were characterized by computing and diagonalizing the Hessian. A positive definite Hessian indicates that a local minimum has been found, while one negative eigenvalue suggests that a saddle point (TS) has been located. The zero point energy corrections for the diffusion stationary points were calculated using the UB3LYP method. For the etching part of the calculations, zero point vibrational energies were calculated using CASSCF level of theory. Intrinsic reaction coordinate (IRC)[55] calculations were used to connect the TS structures with the reactants and the products. The IRC calculations were performed using the second-order method developed by Gonzalez and

Schlegel (GS2)[56,57] using a step size of 0.3 $(amu)^{1/2}$bohr. All of the calculations were performed without imposing any symmetry constraints on the structures.

The GAMESS (General Atomic and Molecular Electronic Structure System) program[57] was used in all of the computations except the CIM-CC calculations. The CIM-CC calculations were performed using the suite of computer programs described in Refs. 27, 28, and 50, which are interfaced with the GAMESS Hartree-Fock, orbital localization, and integral transformation routines. The CIM-CC calculations rely on the design of the CIM orbital subsystems introduced in Ref. 28. MM3[58-60] parameters were used for the MM part of the calculations. The QM/MM (SIMOMM) calculations were carried out using the GAMESS/Tinker interface.[61,62]

III. Results and Discussion

A. Comparison of the UB3LYP and CASSCF geometries

SIMOMM:CASSCF(12,11)/HW(d) geometry optimizations were carried out for the structures including TS (b) and on-top (a) on the diffusion PES (see Figure 6) to compare the predicted geometries with those predicted by the SIMOMM:UB3LYP/HW(d) level of theory. Figure 6 shows the structures and bond lengths of the diffusion stationary points (a)-(e). Structure c in Figure 6 is also a stationary point on the etching PES, as is structure f (etched products). The UB3LYP Si-Si dimer bond lengths are similar to the CASSCF bond lengths with deviations on the order of ~0.1Å. Most of the bond distances shown in Figure 6 are in even better agreement.

In addition, the two most important structures on the etching PES (back-bond and the etched surface/SOLA) have been optimized using the SIMOMM:CASSCF(12,11)/HW(d)

method to compare the geometrical differences with the SIMOMM:UB3LYP level of theory (see Figure 6). The UB3LYP bond lengths are in very good agreement with the CASSCF geometries for these structures. Table 1 shows all of the bond lengths that were obtained at both the CASSCF and the UB3LYP levels of theory, as well as the angles on the diffusion PES. The UB3LYP method does a reasonable job of reproducing the CASSCF bond lengths and angles. Since the UB3LYP method is computationally less expensive than CASSCF and has fewer convergence problems, UB3LYP has been used for the generation of all other structures on the etching PES.

B. Evaluation of the CIM-CC methods

The CR-CC(2,3) method is a single-reference CC approach which can provide high quality relative energies in diradical regions of a PES that result, for example, from single bond breaking.[24,49,50,63-71] This suggests that CR-CC(2,3) is a viable method to describe the dangling bonds on the Si(100) surface (see Figure 1). The canonical CR-CC(2,3) approach, with its intrinsic N^7 scaling of the CPU time, is generally limited to molecules of small to moderate size. To reduce the computer costs associated with the canonical CR-CC(2,3) calculations, the local correlation CIM-CR-CC(2,3) approach in the pure and mixed forms mentioned in Section 2[27,28,54], which replaces the N^7 CPU steps by steps that scale linearly with the system size, is employed in this study.

Table 2 compares the relative energies predicted by the canonical and CIM CC methods for two stationary points on the etching PES, the back-bond structure c and the etched products f. The threshold ζ used in the CIM-CC calculations summarized in Table 2 was set at 0.003. The error in the relative energy resulting from the pure CIM-CR-CC(2,3) calculations,

relative to the canonical CR-CC(2,3) method, is ~0.4 kcal/mol. Much of this difference between the canonical and CIM-CR-CC(2,3) relative energies originates from the error in the CIM-CCSD calculations, which produce a relative energy that differs from the corresponding canonical CCSD energy by 0.5 kcal/mol (see Table 2). When the mixed CIM-CR-CC(2,3) approach, in which one adds the local triples correction of CIM-CR-CC(2,3) to the canonical CCSD energy, is used, the accuracy improves and the error relative to canonical CR-CC(2,3) decreases to 0.04 kcal/mol. This is because the bulk of the correlation energy is in the CCSD part, so the relative accuracy of CIM-CR-CC(2,3) vs. canonical CR-CC(2,3) is defined almost entirely by the accuracy of the preceding CCSD calculation. The CPU times and memory requirements characterizing the CIM-CC approaches, compared with the corresponding canonical CC calculations for the etching stationary points, are shown in Table 3. Despite the use of a rather tight threshold ($\zeta = 0.003$), the CIM-CC methodology offers noticeable savings in the computer effort, by a factor of ~2 in the triples correction. These savings become more significant when larger basis sets and less strict threshold values are employed, as illustrated below for the diffusion pathway.

Table 4 compares the relative energies of the stationary points along the diffusion pathway obtained using various CC methods and basis sets. The mixed CIM-CR-CC(2,3) calculations for the 6-31G(d) basis set were performed at two different z values, 0.005 and 0.01. The mixed CIM-CR-CC(2,3) calculations for the larger 6-311G(d) basis set were performed using $\zeta = 0.01$ to conserve computer time. As shown in Table 4, the use of the less tight threshold in the mixed CIM-CR-CC(2,3) calculations changes the relative energies of the stationary points along the diffusion pathway by ~3 kcal/mol or less compared to the tighter $\zeta = 0.005$ value when the smaller basis set is used. The general agreement between the mixed

CIM(0.01)-CR-CC(2,3) and canonical CR-CC(2,3) relative energies for the 6-31G(d) basis set is excellent, with errors on the order of 1 kcal/mol or less, except for TS (d) for which the error is ~5 kcal/mol. Also shown in Table 4 are the mixed CIM(0..5)-CR-CC(2,3)/6-311G(d) relative energies that were estimated assuming additivity of corrections for the level of theory and basis set. These relative energies are all within 1 kcal/mol or less of those obtained with the canonical CR-CC(2,3)/6-31G(d) calculations. As shown in Table 5, the use of the less strict threshold results in significant savings in the CPU time and memory relative to ζ = 0.005.

C: Etching of the silicon surface (Si(100)) by oxygen atom

The etching/desorption process corresponds to the removal of a silicon atom from the Si(100) surface by an oxygen atom. The stationary points (minima and TS structures) for the etching mechanism proposed by Choi et al.[22] are summarized in Figure 2. It was found in the previous work that the back-bond structure (c) is a key minimum on the PES. The product structure corresponding to SiO + the etched surface (SOLA) is the highest energy structure on the PES. Choi et al. calculated the corresponding overall MRMP2 activation barrier by taking the energy difference between the back-bond structure and that of SOLA+ SiO, obtaining ~90 kcal/mol.

In this work the etching barrier is defined in the same manner; that is, by calculating the energy difference between the back-bond/lowest energy structure (c) and the SOLA/etched structure + SiO (f) shown in Figure 6. The SIMOMM:MRMP2(12,11)/6-31G(d) etching barrier is 81.4 kcal/mol, approximately 8 kcal/mol lower than the value predicted by Choi et al. using the mixed basis set (see Table 6). As shown in Table 6, the etching barrier

calculated using the canonical CR-CC(2,3) and mixed CIM-CR-CC(2,3) methods employing the 6-31G(d) basis set, of 87.3 kcal/mol, is approximately 2 kcal/mol lower than that calculated by Choi et al. All of these barriers, predicted by high levels of theory, are within the experimental range[8,11,14] of 80-90 kcal/mol, suggesting that the previous KMC predictions[30] are slightly low.

D. Diffusion of oxygen on the silicon surface

Five stationary points were found on the PES for the diffusion of an oxygen atom from one dimer to an adjacent dimer on the Si(100) surface. The structures of these stationary points along with a comparison of the corresponding bond lengths are shown in Figure 6. The predicted singlet PES for the diffusion of an O atom between two adjacent dimers on the Si(100)-2 X 1 surface at several levels of theory using the SIMOMM: UB3LYP/6-31G(d) geometries is shown in Figure 7 (cf., also, Table 4 for the information pertaining to the CC calculations). As noted above, the CC calculations are obtained using only the QM cluster.

As shown in Figures 6 and 7, the starting point for the diffusion pathway is the on-top position a. TS b connects the global minimum (back-bond structure c) on the PES with structure a, shown in Figure 6. The O atom moves towards the second dimer in a perpendicular direction to the dimer rows (cf. Figure 3). The Si6-O bond length decreases from 3.0 Å to 2.6 Å as the oxygen moves from the on-top position a to the TS b. In the back-bond structure c, the oxygen atom has inserted into a bond (Si1-Si6) that connects a surface silicon atom (Si1) to a silicon atom (Si6) in the next layer. A Si6-Si1 bond is broken and a new Si6-O bond is formed (see Figure 6). The SIMOMM:UB3LYP/6-31G(d) energy difference (barrier height) between TS b and on-top a is ~1 kcal/mol (see Figure 7 and Table

4). This barrier disappears when all implementations of CR-CC(2,3)/6-311G(d) are used, including the CIM methods. This indicates that the on-top structure is near a transition state and that the diffusion of the oxygen atom towards the formation of the back-bond structure c is downhill. The SIMOMM:MRMP2(12,11) calculations[22] predict that TS b is 4.2 kcal/mol higher in energy than the on-top structure. This small difference may reflect, in part, the use of B3LYP geometries in the present calculations. In any case it appears that the on-top structure can easily be converted to the back-bond structure c.

TS d (see Figures 6 and 7) connects the back-bond structure c with another back-bond structure e. As shown in Figure 6, TS d has the oxygen atom bonded to three silicon atoms (Si1 and Si3 from the two Si-Si dimers and Si6 from the next layer in the cluster). TS d is formed when the oxygen atom connects the two adjacent silicon dimers via a siloxane bridge structure (sBO).[10] Stationary point e is also a back-bond structure, formed when the oxygen atom moves towards the adjacent silicon dimer (Si3-Si4) from the back-bond structure c. The two structures c and e are similar. The difference in the bond lengths of these two back-bond structures is ~ 0.02 Å and the mixed CIM-CR-CC(2,3)/6-311G(d) energy difference is only 2.6 kcal/mol.

Although structures c and e have similar energies, the conversion of the back-bond structure c to the back-bond structure e via the siloxane bridge (TS d) has a rather large activation energy of more than 70 kcal/mol. As shown in Figure 7, the activation barriers for the conversion of c → e calculated at the SIMOMM:UB3LYP, CCSD(T) and CR-CC(2,3) levels of theory agree reasonably well with each other. The c → e barrier heights calculated at these levels of theory lie in a range of 66-72 kcal/mol. This suggests that using SIMOMM to predict geometries (thereby incorporating bulk effects) and then using the QM cluster

model for high level (i.e., CC) calculations to obtain accurate energies is a reasonable strategy. This also suggests that the major contribution to the activation barrier comes from the QM part of the system at correlated levels of theory and the small QM clusters consisting of 15 silicon atoms reasonably mimic the silicon surface for the diffusion of O on the silicon surface. The overestimated activation barrier (74 and 79.5 kcal/mol) at mixed CIM level of theory relative to the CR-CC(2,3) calculations may be attributed to the use of a loose threshold parameter. The estimated value for the activation barrier at the CIM(0.005)-CR-CC(2,3)/6-311G(d) level of theory, assuming the additivity of improvements in the basis set and level of theory, is 71.3 kcal/mol. This value is in good agreement with the activation barrier calculated using canonical CR-CC(2,3).

Yamasaki et al.[72] predicted that the activation energy corresponding to the c → e conversion decreases when the coverage is more than 3 oxygen atoms and the sBO (d) configuration becomes energetically favorable. Hemeryck et al. studied the diffusion process using plane wave GGA density functional theory.[10] They found two additional metastable structures that connect the sBO structure that in turn connects the two back-bond structures. This is in contrast to the results in the present work where the sBO structure directly connects the two back-bond structures (c and e) and the activation energy is approximately 30 kcal/mol higher than that predicted by Hemeryck et al. In the present work, the pathway along the diffusion process was confirmed using intrinsic reaction coordinate (IRC) calculations that connect the TS (d) with the back-bond structures (c) and (e). No intermediate structure was detected. The low activation energy calculated by Hemeryck et al. has been attributed to the omission of Hartree-Fock exchange in the GGA calculations.[73,74] The results in the present study are likely to be more reliable as compared to the predictions

made by previous DFT studies because in the present study, the relative energies are calculated using correlated high level ab initio calculations. Previous KMC studies employed barrier heights to optimize the agreement with experimental observations. A comparison of the diffusion activation energies corresponding to the conversion of the back-bond c structure to the back-bond e structure with previous calculations and with experiments is given in Table 7. The c → e activation energy calculated in the present study using SIMOMM:UB3LYP and coupled-cluster methods is about 66-74 kcal/mol; the mixed-CIM(0.01)-CR-CC(2,3)/6-311G(d) value is a few kcal/mol higher than this range, very likely due to a combination of a low threshold and a small basis set. The range of barrier heights in Table 7 is a bit higher than that predicted by the previous KMC simulations that included water on the surface.[38] As noted in the Introduction, the range of experimental values[41-45] for this barrier height is quite large and encompasses all of the theoretical values.[10,37,38]

IV. Conclusions

The mechanism of the long range diffusion of atomic oxygen on the Si(100)-2 X 1 surface has been studied using accurate coupled cluster energies that were obtained at geometries determined using unrestricted density functional theory (UB3LYP DFT) in concert with the SIMOMM embedded cluster approach. The diffusion PES reveals that the structure with the back-bond insertion of the oxygen atom into a sub-surface Si-Si bond is the lowest energy structure, in agreement with previous DFT/GGA studies by Hemeryck et al.[10] Another back-bond structure that is formed on the second Si-Si dimer is also a minimum on the diffusion PES. These two back-bond structures are connected via a TS involving a siloxane bridge structure, again in agreement with the DFT calculations of Hemeryck et al.[10] Another stable

intermediate on the PES, the on-top structure, is connected to the back-bond structure via a TS that is similar to the initial oxidation pathway of the etching mechanism observed in previous theoretical studies by Choi et al.[22] The coupled cluster activation energy for the diffusion process is within the rather large experimental range of 41-76 kcal/mol. The theoretical values of the diffusion barrier found previously lie in the range of 42.8 (DFT)-57.7 (KMC simulations) kcal/mol. There have, to our knowledge, been no previous ab initio calculations on this process. The studies presented here employ the most accurate electronic structure theory methods that have been employed to date on this process. While cluster calculations can be limited by "edge effects", the use of large QM/MM clusters as in the present work very likely minimize such effects. The present work does not include an exhaustive of the entire potential energy surface for the Si cluster + O atom. So, it is possible that there are other stationary points that are relevant to the present estimate of the height of the diffusion barrier. However, as noted in earlier work, diffusion perpendicular to dimer rows is not likely due to high energy barriers in that direction.[53]

The SiO desorption barrier for the etching process has also been calculated by taking the energy difference between the final etched product and the back-bond structure, which is the lowest energy structure on the etching PES. The canonical and CIM-CR-CC(2,3) methods have been used to calculate the barrier and assess the accuracy of previous calculations. It has been found that the etching barrier height is 87.3 kcal/mol at the canonical CR-CC(2,3)/6-31G(d) and mixed CIM-CR-CC(2,3)/6-31G(d) levels of theory. This result lies in the higher end of the range of the previous theoretical (73.8-89.8 kcal/mol) and experimental (80-90 kcal/mol) values.

Acknowledgements

This work has been supported by a grant from the U.S. Department of Energy, administered by the Ames Laboratory (M.S.G), and by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (Grant No. DE-FG02-01ER15228; P.P). The authors gratefully acknowledge helpful discussions with Dr. Michael .W. Schmidt.

References

(1) Goddard III, W. A.; Low, J. J.; Olafson, B. D.; Redondo, A.; Zeiri, Y.; Steigerwald, M. L.; Carter, E. A.; Allison, J. N.; Chang, R., in *Proceedings of the Symposium on the Chemistry and Physics of Electrocatalysis*; McIntyre, J. D. E., Weaver, M. J., Yeager, E. B., Eds.; The Electrochemical Society, Inc.: Pennington, NJ, 1984; Vol. 84-12.

(2) Ceiler, M. F.; Kohl, P. A.; Bidstrup, S. A. *Journal of the Electrochemical Society* **1995**, *142*, 2067.

(3) Patrick, W. J.; Schwartz, G. C.; Chapplesokol, J. D.; Carruthers, R.; Olsen, K. *Journal of the Electrochemical Society* **1992**, *139*, 2604.

(4) May; S, G.; Sze; Simon, M. *Fundamentals of Semiconductor Fabrication*; Wiley: New York, 2003.

(5) Peercy, P. S. *Nature* **2000**, *406*, 1023.

(6) Pasquarello, A.; Hybertsen, M. S.; Car, R. *Nature* **1998**, *396*, 58.

(7) Xue, K.; Ho, H. P.; Xu, J. B. *Journal of Physics D-Applied Physics* **2007**, *40*, 2886.

(8) Engstrom, J. R.; Engel, T. *Physical Review B* **1990**, *41*, 1038.

(9) Hemeryck, A.; Richard, N.; Esteve, A.; Rouhani, M. D. *Journal of Non-Crystalline Solids* **2007**, *353*, 594.

(10) Hemeryck, A.; Richard, N.; Esteve, A.; Djafari Rouhani, M. *Surface Science* **2007**, *601*, 2339.

(11) Engstrom, J. R.; Bonser, D. J.; Nelson, M. M.; Engel, T. *Surface Science* **1991**, *256*, 317.

(12) Engstrom, J. R.; Bonser, D. J.; Engel, T. *Surface Science* **1992**, *268*, 238.

(13) Suemitsu, M.; Enta, Y.; Miyanishi, Y.; Miyamoto, N. *Physical Review Letters* **1999**, *82*, 2334.

(14) Engel, T. *Surface Science Reports* **1993**, *18*, 91.

(15) Lewerenz, H. J.; Lubke, M.; Bachmann, K. J.; Menezes, S. *Applied Physics Letters* **1981**, *39*, 798.

(16) Uhlir, A. *Bell System Technical Journal* **1956**, *35*, 333.
(17) Shoemaker, J. R.; Burggraf, L. W.; Gordon, M. S. *Journal of Physical Chemistry A* **1999**, *103*, 3245.
(18) Rintelman, J. M.; Gordon, M. S. *Journal of Physical Chemistry B* **2004**, *108*, 7820.
(19) Jung, Y. S.; Gordon, M. S. *Journal of the American Chemical Society* **2005**, *127*, 3131.
(20) Choi, C. H.; Gordon, M. S. *Journal of the American Chemical Society* **2002**, *124*, 6162.
(21) Tamura, H.; Gordon, M. S. *Journal of Chemical Physics* **2003**, *119*, 10318.
(22) Choi, C. H.; Liu, D. J.; Evans, J. W.; Gordon, M. S. *Journal of the American Chemical Society* **2002**, *124*, 8730.
(23) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chemical Physics Letters* **1989**, *157*, 479.
(24) Piecuch, P.; Włoch, M. *Journal of Chemical Physics* **2005**, *123*, 224105.
(25) Li, S.; Ma, J.; Jiang, Y. *Journal of Computational Chemistry* **2002**, *23*, 237.
(26) Li, S. ; Shen, J.; Li, W.; Jiang, Y. *Journal of Chemical Physics* **2006**, *125*, 074109.
(27) Li, W.; Gour, J. R.; Piecuch, P.; Li, S. *Journal of Chemical Physics* **2009**, *131*, 114109.
(28) Li, W.; Piecuch, P.; Gour, J. R. In *Progress in Theoretical Chemistry and Physics*, Vol. 19, *Advances in the Theory of Atomic and Molecular Systems: Conceptual and Computational Advances in Quantum Chemistry*; Piecuch, P., Maruani, J., Delgado-Barrio, G., Wilson, S., Eds.; Springer: Dordrecht, 2009; pp. 131-195.
(29) Albao, M. A.; Chuang, F. C.; Evans, J. W. *Thin Solid Films* **2009**, *517*, 1949.
(30) Albao, M. A.; Liu, D. J.; Choi, C. H.; Gordon, M. S.; Evans, J. W. *Surface Science* **2004**, *555*, 51.
(31) Albao, M. A.; Liu, D. J.; Gordon, M. S.; Evans, J. W. *Physical Review B* **2005**, *72*, 195420**.**
(32) Fichthorn, K. A.; Weinberg, W. H. *Journal of Chemical Physics* **1991**, *95*, 1090.
(33) Uchiyama, T.; Uda, T.; Terakura, K. *Surface Science* **2001**, *474*, 21.
(34) Perdew, J. P.; Wang, Y. *Physical Review B* **1992**, *45*, 13244.
(35) Laasonen, K.; Pasquarello, A.; Car, R.; Lee, C.; Vanderblit, D. *Physical Review B* **1993**, *47*, 10142.
(36) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. *Reviews of Modern Physics* **1992**, *64*, 1045.
(37) Ebner, C.; Seiple, J. V.; Pelz, J. P. *Physical Review B* **1995**, *52*, 16651.
(38) Esteve, A.; Chabal, Y. J.; Raghavachari, K.; Weldon, M. K.; Queeney, K. T.; Rouhani, M. D. *Journal of Applied Physics* **2001**, *90*, 6000.
(39) Kresse, G.; Furthmuller, J. *Physical Review B* **1996**, *54*, 11169.
(40) Kresse, G.; Joubert, D. *Physical Review B* **1999**, *59*, 1758.
(41) Seiple, J. V.; Pelz, J. P. *Physical Review Letters* **1994**, *73*, 999.
(42) Seiple, J. V.; Ebner, C.; Pelz, J. P. *Physical Review B* **1996**, *53*, 15432

(43) Mikkelsen, J. C. *Applied Physics Letters* **1982**, *40*, 336.
(44) Abe, T.; Yamada-Kaneta, H. *Journal of Applied Physics* **2004**, *96*, 4143.
(45) LEE, S. T.; Nichols, D. *Applied Physics Letters* **1985**, *47*, 1001
(46) Hay, P. J.; Wadt, W. R. *Journal of Chemical Physics* **1985**, *82*, 270.
(47) Hehre, W. J.; Ditchfield, R; Pople, J. A. *Journal of Chemical Physics* **1972**, *56*, 2257.
(48) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *Journal of Chemical Physics* **1982**, *77*, 3654.
(49) Piecuch, P.; Włoch, M.; Gour, J. R.; Kinal, A. *Chemical Physics Letters* **2006**, *418*, 467.
(50) Wloch, M.; Gour, J. R.; Piecuch, P. *Journal of Physical Chemistry A* **2007**, *111*, 11359.
(51) Redondo, A.; Goddard, W. A. *Journal of Vacuum Science & Technology* **1982**, *21*, 344.
(52) Ge, Y. B.; Gordon, M. S.; Piecuch, P. *Journal of Chemical Physics* **2007**, *127*.
(53) Zorn, D. D.; Albao, M. A.; Evans, J. W.; Gordon, M. S. *Journal of Physical Chemistry C* **2009**, *113*, 7277.
(54) Li, W.; Piecuch, P.; Gour, J. R. In *Theory and Applications of Computational Chemistry - 2008*; *AIP Conference Proceedings*, Vol. 1102; Wei, D.-Q., Wang, X.-J., Eds.; American Physical Society: Melville, NY, 2009; pp. 68-113.
(55) Garrett, B. C.; Redmon, M. J.; Steckler, R.; Truhlar, D. G.; Baldridge, K. K.; Bartol, D.; Schmidt, M. W.; Gordon, M. S. *Journal of Physical Chemistry* **1988**, *92*, 1476.
(56) Gonzalez, C.; Schlegel, H. B. *Journal of Physical Chemistry* **1990**, *94*, 5523.Gonzalez, C.; Schlegel, H. B. *Journal of Chemical Physics* **1991**, *95*, 5853.
(57) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds. (Elsevier, Amsterdam, 2005); p. 1167.
(58) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *Journal of the American Chemical Society* **1989**, *111*, 8551.
(59) Lii, J. H.; Allinger, N. L. *Journal of the American Chemical Society* **1989**, *111*, 8566.
(60) Lii, J. H.; Allinger, N. L. *Journal of the American Chemical Society* **1989**, *111*, 8576.
(61) Ponder, J. W.; Richards, F. M. *Journal of Computational Chemistry* **1987**, *8*, 1016.
(62) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. *Journal of Computational Chemistry* **1991**, *12*, 402.
(63) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musiał, M. *Computer Physics Communications* **2002**, *149*, 71.
(64) Kinal, A.; Piecuch, P. *Journal of Physical Chemistry A* **2007**, *111*, 734.

(65) Cramer, C. J.; Włoch, M.; Piecuch, P.; Puzzarini, C.; Gagliardi, L. *Journal of Physical Chemistry A* **2006**, *110*, 1991; **2007**, *111*, 4871 [Addition/Correction].
(66) Cramer, C. J.; Kinal, A.; Włoch, M.; Piecuch, P.; Gagliardi, L. *Journal of Physical Chemistry A* **2006**, *110*, 11557; **2007**, *111*, 4871 [Addition/Correction].
(67) Cramer, C. J.; Gour, J. R.; Kinal, A.; Włoch, M.; Piecuch, P.; Shahi, A. R. M.; Gagliardi, L. *Journal of Physical Chemistry A* **2008**, *112*, 3754.
(68) Piecuch, P.; Włoch, M.; Varandas, A. J. C. *Theoretical Chemistry Accounts* **2008**, *120*, 59.
(69) Song, Y. Z.; Kinal, A.; Caridade, P. J. S. B.; Varandas, A. J. C.; Piecuch, P. *Journal of Molecular Structure: THEOCHEM* **2008**, *859*, 22.
(70) Ge, Y.; Gordon, M. S.; Piecuch, P. *Journal of Chemical Physics* **2007**, *127*, 174106.
(71) Ge, Y.; Gordon, M. S.; Piecuch, P.; Włoch, M.; Gour, J. R. *Journal of Physical Chemistry A* **2008**, *112*, 11873.
(72) Yamasaki, T.; Kato, K.; Uda, T. *Physical Review Letters* **2003**, *91*.
(73) Gritsenko, O. V.; Ensing, B.; Schipper, P. R. T.; Baerends, E. J. *Journal of Physical Chemistry A* **2000**, *104*, 8558.
(74) Umrigar, C. J.; Gonze, X. *Physical Review A* **1994**, *50*, 3827.

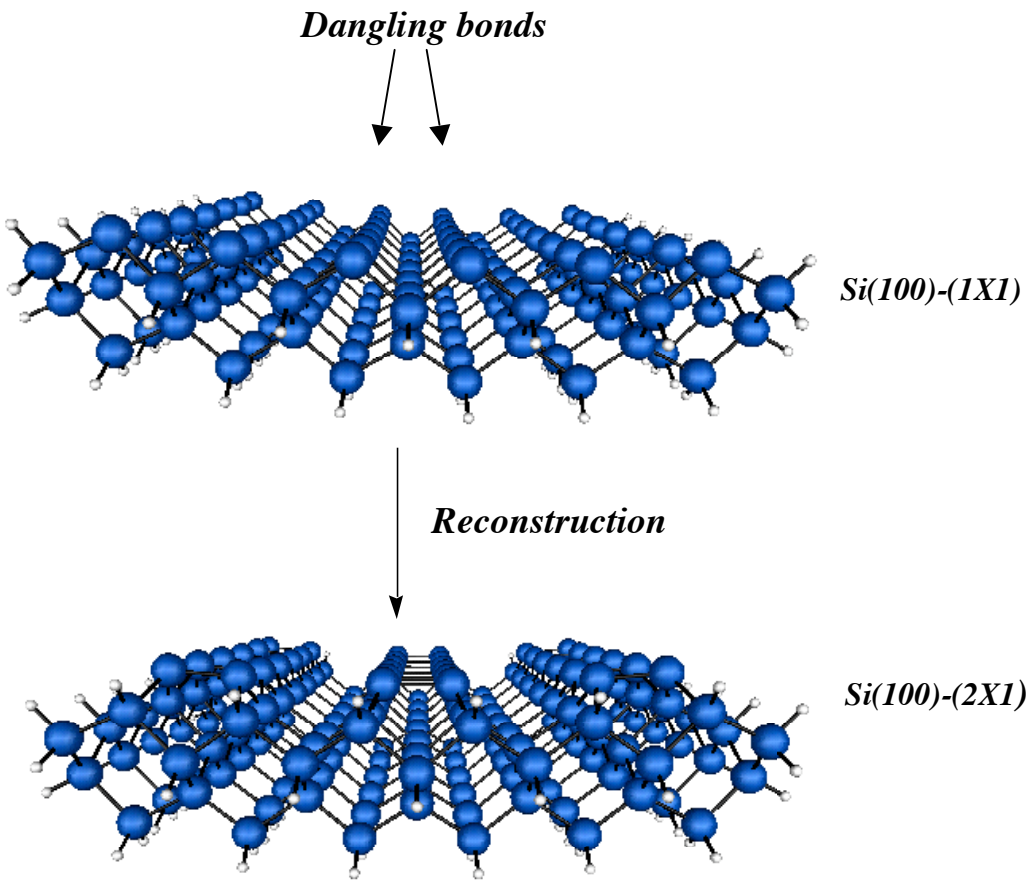


Figure 1: Si(100) surface reconstruction.

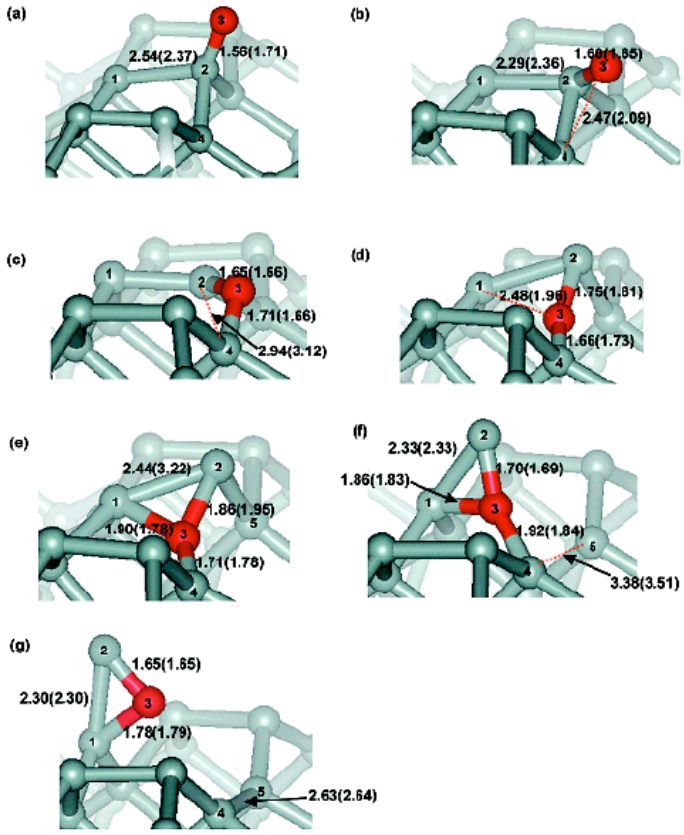


Figure 2: Previously suggested stationary points for the etching mechanism obtained by Choi et al. at CASSCF(8,7)/HW(d) geometries: (a) on-top structure; (b) transition state between a and c; (c) back-bond structure; (d) transition state connecting c and e; (e) minimum with trivalent O atom; (f) transition state connecting e and g; (g) minimum with triangle configuration. See Ref. 22.

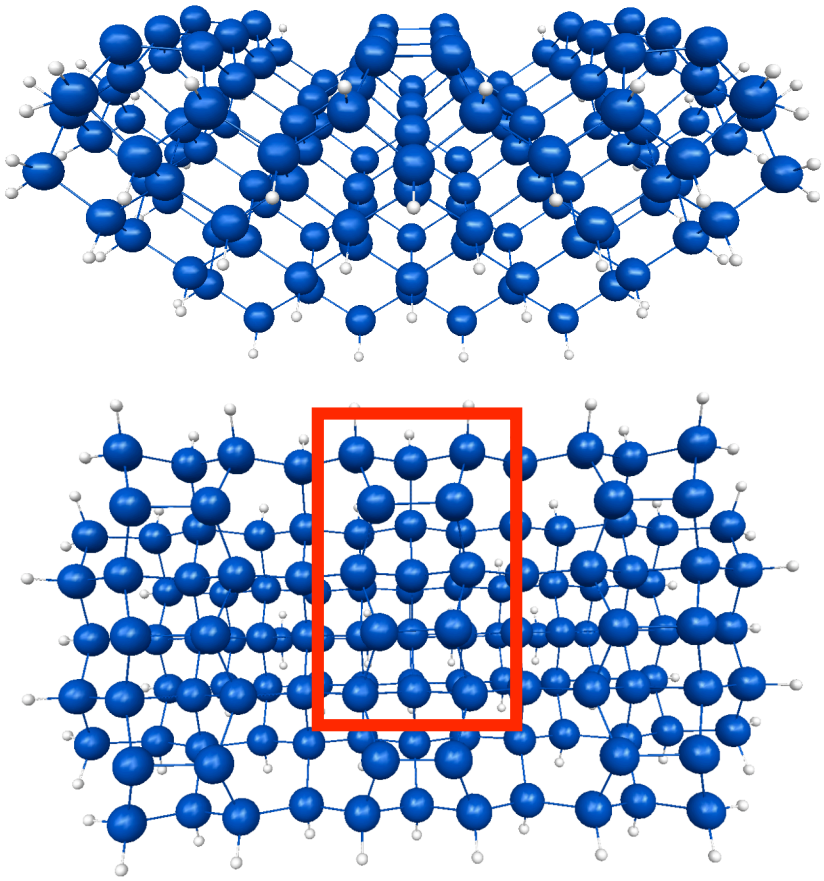


Figure 3: The SIMOMM model of $Si_{15}H_{16}$QM region embedded in an $Si_{136}H_{92}$ MM cluster. The region inside the red box represents the QM cluster.

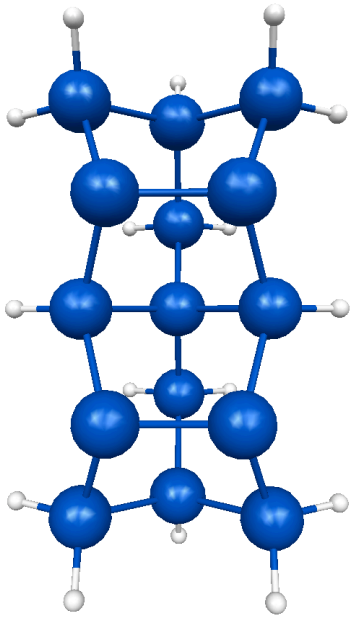


Figure 4: The $Si_{15}H_{16}$ QM cluster

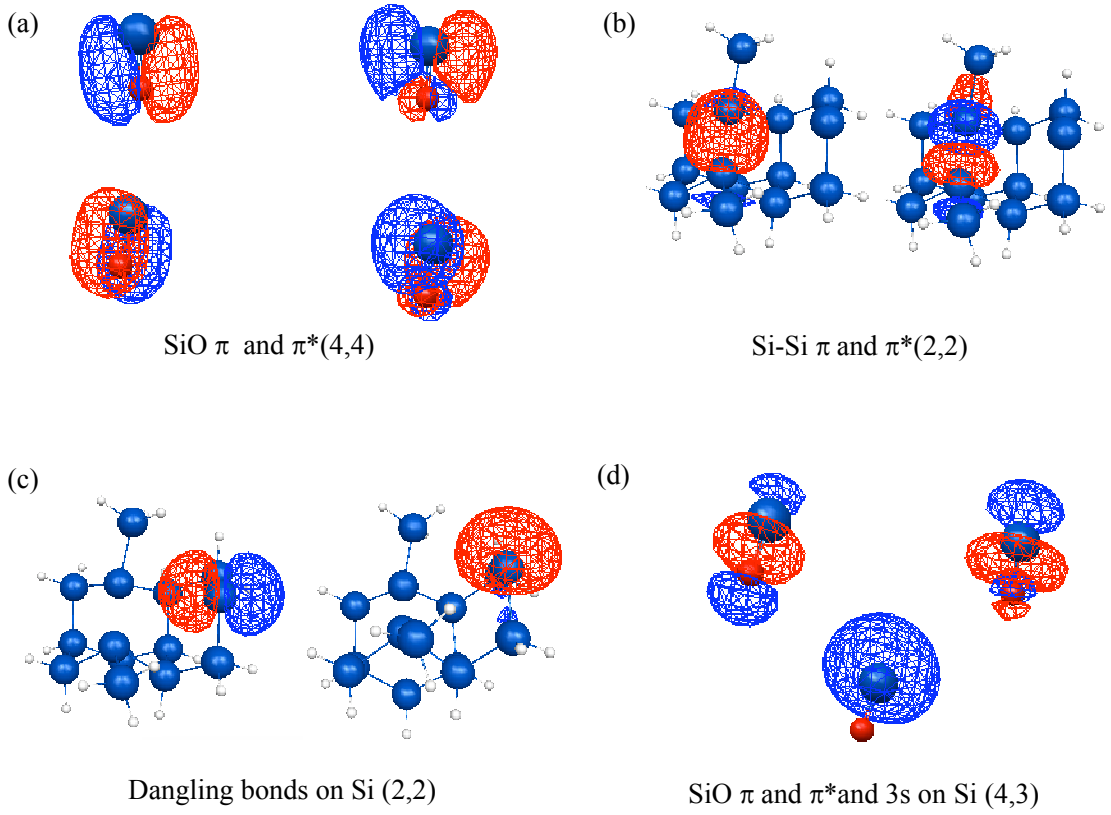


Figure 5: The active space orbitals used in the calculation of SiO desorption barrier

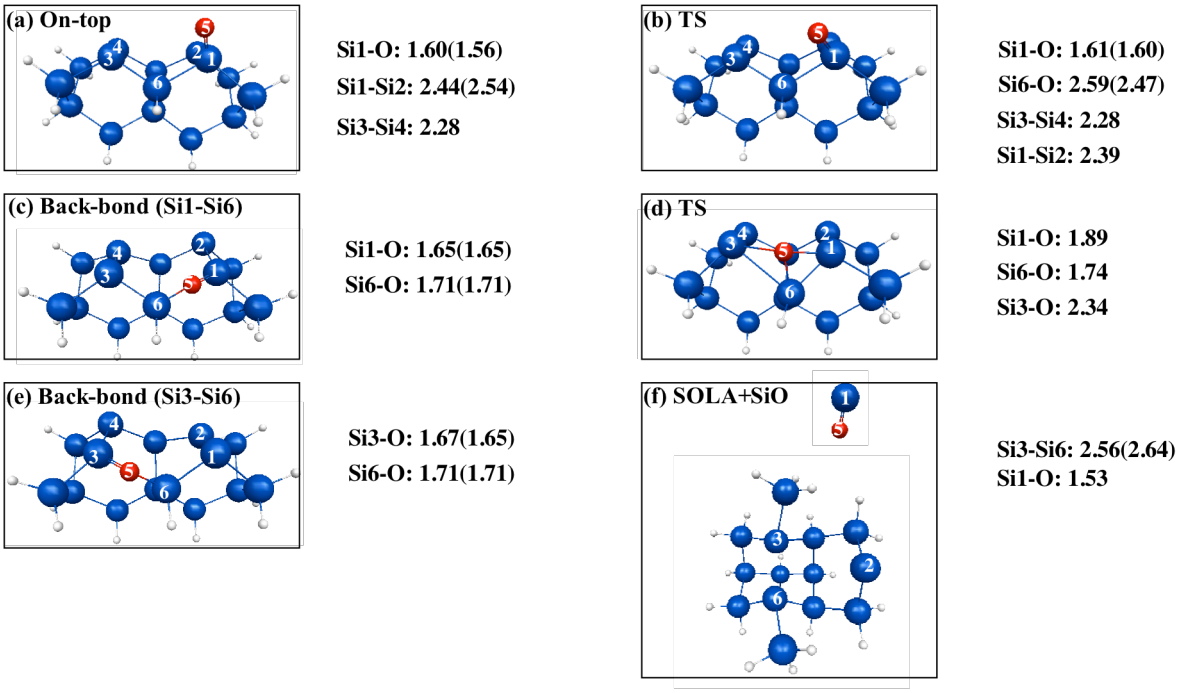


Figure 6: Stationary points (a) to (e) are on the O atom diffusion path along adjacent Si-Si dimers obtained with SIMOMM:UB3LYP/6-31G(d): (a) on-top structure; (b) transition state connecting a and c; (c) back-bond (Si1-Si6) structure; (d) transition state with a trivalent oxygen connecting c and e; (e) back-bond (Si3-Si6) structure. Stationary points (c) and (f) are on the etching PES: (f) SOLA+SiO (etched products). In parentheses are the CASSCF values from ref 22.

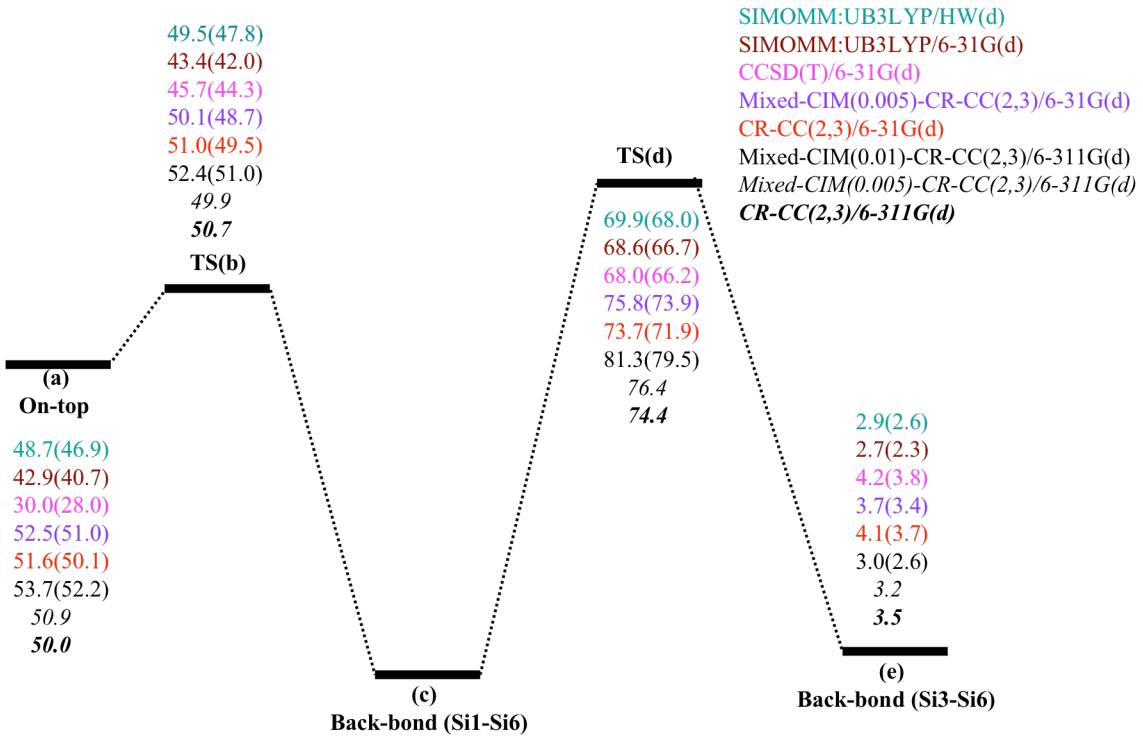


Figure 7: Relative energy diagram for structures (a) to (e) in Figure 6. Energies are obtained using SIMOMM:UB3LYP/6-31G(d) geometries. All values are in kcal/mol. Values in parentheses include SIMOMM:UB3LYP zero point corrections. The values shown in italics correspond to the estimated values based on the assumption of the additivity of basis set and level of theory.

Table 1: Comparison of SIMOMM:CASSCF(12,11)/HW(d) and SIMOMM:UB3LYP/HW(d) bond lengths (Angstroms) and bond angles (degrees) for the structures and atom numbering shown in Figure 6.

Bond lengths (Å)								
	Diffusion stationary points				Etching stationary points			
	on-top (a)		TS (b)		Back-bond (c)		Etched product (f)	
	UB3LYP	MCSCF	UB3LYP	MCSCF	UB3LYP	MCSCF	UB3LYP	MCSCF
Si1-O	1.61	1.56	1.62	1.62	1.65	1.65	1.53	1.54
Si6-O			2.64	2.54	1.71	1.71		
Si3-Si4	2.26	2.25	2.26	2.25				
Si1-Si2	2.40	2.54	2.36	2.35	2.26	2.25		
Si3-Si6							2.57	2.58
Bond angles (degrees)								
Si6-Si1-O			80	76	125	123		
Si2-Si1-O	124	118	127	126				
Si8-Si7-Si9							98	97

Table 2: The canonical and CIM CCSD, and canonical (can), CIM, and mixed CIM CR-CC(2,3) relative energies, obtained with the 6-31G(d) basis set, for the etching stationary points shown in Figure 6.

Structure	CCSD		CR-CC(2,3)		
	Can.[a]	CIM(0.003)[a]	Can.[a]	CIM(0.003)[a]	Mixed CIM(0.003)[a]
Back-bond (**c**)	0	0	0	0	0
Etched products (**f**)	86.8	87.3	87.3	87.7	87.3

[a] The relative energies are reported in kcal/mol.

Table 3: CPU times and RAM requirements characterizing the calculations of the canonical (can) and CIM CCSD energies and the canonical and CIM CR-CC(2,3) triples corrections for the etching stationary points shown in Figure 6. All of the calculations were performed using the 6-31G(d) basis set on an SGI Altix 3700 Bx2 system, equipped with 1.6GHz Itanium2 processors.

Structure	CCSD		Triples correction of CR-CC(2,3)	
	Can.[a]	CIM(0.003)[a,b]	Can.[a]	CIM(0.003)[a,b]
Back-bond (**c**)	34 (6.4)	9 (2.6)	166 (14.4)	129 (8.6)
Etched products (**f**)	42 (6.4)	8 (2.1)	165 (14.4)	81 (6.7)

[a] CPU times are reported in hours. The numbers in parentheses correspond to RAM requirements in GB.

[b] CPU times and RAM requirements characterizing each CIM-CC calculation correspond to the largest CIM orbital subsystem.

Table 4: Relative energies of the stationary points along the diffusion pathway shown in Figure 6 obtained in the canonical and mixed CIM CR-CC(2,3) calculations using the 6-31G(d) basis set and the mixed CIM-CR-CC(2,3) calculations using the 6-311G(d) basis set. All of the reported energy values include zero point energy corrections calculated using the UB3LYP method.

Structure	6-31G(d)				6-311G(d)	
	Can. CR-CC(2,3)[a]	Mixed CIM(0.005)-CR-CC(2,3)[a]	Mixed CIM(0.01)-CR-CC(2,3)[a]	Can. CCSD(T)[a]	Mixed CIM(0.01)-CR-CC(2,3)[a]	Mixed CIM(0.005)-CR-CC(2,3)[a,b]
Back-bond (Si1-Si6) (c)	0.0	0.0	0.0	0.0	0.0	0.0
TS (d)	71.9	73.9	77.0	66.2	79.5	76.4
Back-bond (Si3-Si6) (e)	3.7	3.4	2.8	3.8	2.6	2.4
TS (b)	49.5	48.7	49.8	44.3	51.0	49.9
On-top (a)	50.1	51.0	52.3	28.0	52.2	50.9

[a] The relative energies are reported in kcal/mol; can. = canonical.

[b] The values are estimated assuming the additivity of basis set and level of theory

Table 5: The CPU timings and RAM requirements characterizing the calculations of the triples corrections of the canonical and mixed CIM CR-CC(2,3) approaches using the 6-31G(d) and 6-311G(d) basis sets for the stationary points along the diffusion pathway shown in Figure 6. All of the calculations were performed using the Altix 3700 Bx2 system from SGI, equipped with the 1.6GHz Itanium2 processors.

Structure	6-31G(d)			6-311G(d)	
	Can.[a]	CIM(0.005)[a,b]	CIM(0.01)[a,b]	Can.[a,b,c]	CIM(0.01)[a,b]
Back-bond (Si1-Si6) (c)	119 (12.6)	51 (5.9)	15 (2.5)	881 (50.5)	134 (14.7)
TS (d)	141 (12.6)	88 (8.0)	28 (4.2)	1051 (50.5)	219 (19.5)
Back-bond (Si3-Si6) (e)	145 (12.6)	50 (5.9)	17 (3.2)	1081 (50.5)	145 (15.8)
TS (b)	146 (12.6)	85 (8.3)	29 (4.9)	1082 (50.5)	231 (22.6)
On-top (a)	141 (12.6)	103 (8.0)	13 (2.8)	1046 (50.5)	139 (13.4)

[a] The CPU times are reported in hours. The numbers in parentheses correspond to RAM requirements in GB; can. = canonical.

[b] The CPU timings and RAM requirements characterizing each CIM-CR-CC(2,3) calculation correspond to the largest CIM orbital subsystem.

[c] The CPU timings characterizing the triples parts of the canonical CR-CC(2,3)/6-311G(d) calculations were obtained by extrapolating the analogous CPU timings of the canonical CR-CC(2,3)/6-31G(d) calculations using the theoretical $2n_o^3n_u^4$ scaling of the CPU time with the numbers of occupied (n_o) and unoccupied (n_u) orbitals employed in the post-Hartree-Fock calculations. The RAM requirements of the canonical CR-CC(2,3) calculations were determined using the equation $n_o + n_u + 2n_o^2 + 2n_u^2 + 6n_on_u + 2n_o^2n_u + 2n_on_u^2 + 3n_u^3 + 3n_o^3n_u + 5n_o^2n_u^2 + 3n_on_u^3$ which defines the memory requirements of the computer implementation of CR-CC(2,3) described in Ref. 24 and used in this study.

Table 6: Energies (kcal/mol) relative to the back-bond structure (see Figure 6). The calculations were done with the SIMOMM $OSi_{15}H_{16}$;$OSi_{136}H_{92}$ model shown in Figure 4 and

MRMP2(12,11)	CCSD(T)	Mixed CIM(0.003)-CR-CC(2,3)	Can. CR-CC(2,3)	[a]MRMP2(12,11)/MIXED[22]	GGA[33]	KMC[30]	Expt[8,11,14]
6-31G(d)							
81.4 (78.7)	88.6 (85.9)	87.3 (84.6)	87.3 (84.6)	89.8	83.0-87.6	73.8-80.7	80.0-90.0

the geometries were obtained with CASSCF(12,11)/6-31G(d). All relative energies include both QM and MM contributions. The values in parenthesis include zero point energy corrections calculated using the CASSCF method. Can. = Canonical; Expt. = Experiment

[a]geometry at CASSCF(8,7)/HW(d).

Table 7: Comparison of the barrier heights (back-bond (Si1-Si6) (**c**) to back-bond (Si3-Si6) (**e**)) for the long-range diffusion process (See Figure 6) with previous calculations and experiments. All values are zero point corrected at UB3LYP level of theory. All the values are in kcal/mol; Can. = Canonical; Expt. = Experiment

HW(d)		6-31G(d)			6-311G(d)				
UB3LYP		CCSD(T)	Mixed CIM(0.005)-CR-CC(2,3)	Can. CR-CC(2,3)	Mixed CIM(0.01)-CR-CC(2,3)	GGA[10]	KMC/water covered Si(100)[38]	Pelz[37]	Expt.[41-45]
68.0	66.7	66.2	73.9	71.9	79.5	42.8	57.6	55.3-57.7	41.5-76.1

CHAPTER 5. EFFECT OF AQUEOUS MICROSOLVATION ON THE IONIZATION POTENTIAL OF NA

Pooja Arora, Mark S. Gordon

Abstract

The adiabatic ionization potential (AIP) of sodium in water clusters has been calculated by micro-solvating Na and Na^+. A Monte Carlo/simulated annealing technique has been used to search for the lowest energy structures of $Na(H_2O)_{n=1-12}$ and $Na^+(H_2O)_{n=1-12}$. The solute is treated with ab initio quantum chemistry and the EFP (effective fragment potential) method is used for the water molecules. The global and local minimum energy structures of the $Na(H_2O)_{n=1-12}$ clusters suggest that the Na atom resides on the surface of the water cluster. The global minimum search of $Na^+(H_2O)_{n=1-12}$ suggests that at least 12 water molecules are required to completely solvate the Na^+ cation. The EFP method accurately predicts the geometries and relative energies between the local and global energy structures of $Na(H_2O)_{n=1-12}$ and $Na^+(H_2O)_{n=1-12}$. The evolution of the adiabatic ionization potential of $Na(H_2O)_n$ with the increase in cluster size is studied using a combined Hartree-Fock (HF)/EFP method, as well as fully quantum calculations using both second order perturbation theory and a coupled-cluster approach with singles, doubles, and non-iterative perturbative triples. The Na AIP in water clusters decrease up to n = 7 and then becomes almost constant. The vertical ionization potential (VIP) of $Na(H_2O)_n$ is also calculated to illustrate the importance of the reorganization of the solvent molecules around the sodium ion. The influence of electron correlation and the basis set on the predicted ionization potential of $Na(H_2O)_{n=0-12}$ is also discussed.

I. Introduction

Solvation of alkali metal atoms is a subject of extensive research due to the formation of solvated electrons in polar solvents such as water and ammonia.[1-9] The valence electron in alkali metals is loosely bound, allowing it to be detached by polar solvent molecules. This leads to the formation of solvated electrons. Several photo-ionization and photo-absorption[4,10-14] studies have been performed to measure the ionization potential (IP) of alkali metals in solvent clusters, in pursuit of a better understanding of manner in which solvated electrons are generated. The experiments reveal that the IPs of sodium[4] and cesium[12] atoms in water clusters ($Na(H_2O)_n$ and $Cs(H_2O)_n$) decrease rapidly with an increase in cluster size, up to n=4. The IP becomes approximately constant for n>4. Photo-ionization experiments on Li in water clusters also find that the IP values appear to converge when n>5[15].

Several theoretical studies have been reported for the calculation of ionization energies of alkali metals in water clusters. Barnett et al[16] calculated the adiabatic IP of $Na(H_2O)_n$ ($n \leq 8$) using local spin density functional theory (LSDFT), and found a reduced rate of change in the IP as the cluster size increases. Hashimoto et al[17] also studied the ionization potential of $Na(H_2O)_n$ clusters for up to 6 water molecules using second order perturbation theory (MP2)[18]. They found that the IP is nearly constant over the range of n=4-6 for $Na(H_2O)_n$ clusters in which the sodium atom resides on the surface of the water clusters (surface clusters). They also found that the interior $Na(H_2O)_n$ structures (in which the sodium atom is surrounded by the water molecules) do not exhibit the same trend of ionization potential as the surface structures; instead, the IP continues to decrease even at n=6. The Hashimoto et al. theoretical study of lithium water clusters using ab initio methods also revealed a nearly

constant IP after n>4[19]. Cwiklik et al used ab initio molecular dynamics simulations to study the IP and the dynamics of the delocalization of the valence electron of sodium in $Na(H_2O)_{n=34}$ clusters.[20] They concluded that in finite size clusters (up to n=34), the electron is delocalized between the sodium ion and the water cluster. Recently, an experimental and theoretical study by Dauster et al. on sodium in methanol clusters also revealed the same trend in the IP values as observed for $Na(H_2O)_n$. These authors concluded that there is a large contribution of structural diversity (due to several isomers of the sodium water clusters) to the ionization potential measurements.[13] A study by Gao et al., using density functional theory and the Becke three-parameter Lee-Yang-Parr (B3LYP) hybrid functional (DFT/B3LYP),[21,22] showed that the photo-ionization experiments on $Na(H_2O)_n$ can be correlated with an adiabatic process in which the structural reorganization of the solvent molecules is critical.[23]

The primary goal of the present work is to study the evolution of the ionization potential of the $Na(H_2O)_n$ clusters using correlated levels of electronic structure theory. The importance of the reorganization of the water molecules around the sodium ion is investigated by calculating both the adiabatic ionization potential (AIP) and vertical ionization potential (VIP) for $Na(H_2O)_n$ clusters. This is accomplished by micro solvating both sodium atom and sodium ion with up to 12 water molecules using the Hartree-Fock based effective fragment potential (EFP1/HF)[24,25] method to treat the water molecules. The effective fragment potential (EFP1) is a quantum mechanics (QM) based model potential that provides a computationally cost effective way to treat intermolecular interactions including solvent effects. Indeed, Merrill and Webb[26] have used the EFP method to study the hydration of Na^+, Li^+, K^+, Ca^{2+} and Mg^{2+} by adding up to six water molecules to each cation. These authors

concluded that the EFP method is capable of reproducing the binding energies for the alkali metal-water clusters and demonstrated the agreement between the structures obtained at HF and HF/EFP levels of theory.

In the present work, the energetics of $Na(H_2O)_n$ and $Na^+(H_2O)_n$ are discussed based on the number of water molecules directly H-bonded to the solute, sometimes referred as the coordination number or the first solvent shell. The preference of Na and Na^+ to exist as surface vs. interior ions is also considered. The differential and total binding energies of the solvent water molecules to the solute for both $Na(H_2O)_n$ and $Na^+(H_2O)_n$ clusters are calculated for up to n=12 and compared with the available experimental values.

II. Computational methods

The $Na(H_2O)_n$ and $Na^+(H_2O)_n$ clusters were formed by the addition of water molecules to the solute in a stepwise manner with n ranging from 1 to 12. Minimum energy configurations of $Na(H_2O)_n$ and $Na^+(H_2O)_n$ clusters were determined using a Monte Carlo[27]/simulated annealing (MC/SA) code[28]. The MC/SA calculations were performed with a starting temperature of 600K gradually cooling to 300K over 14 steps. Local geometry optimizations were performed after every 10 steps of the simulation and 14 temperatures were sampled in each simulation. About 1400 structures were analyzed in a given MC/SA run to obtain the global minimum structure that was found for each n. Starting structures for the MC/SA runs were determined by either addition or removal of one water molecule at various systematically chosen positions of the n water global minimum structure to obtain the n-1 or n+1 clusters.

For the MC/SA calculations, the Hartree-Fock (HF) level of theory with the 6-31++G(d,p)[29-32] basis set was used for the solute. The water molecules were treated using the Hartree-Fock based effective fragment potential (EFP1/HF) method. Full ab initio MP2/6-31++G(d,p) geometry optimizations were also performed for each $Na(H_2O)_n$ and $Na^+(H_2O)_n$ cluster starting from HF/EFP1 global minimum geometries. MP2 calculations were performed to account for the contributions of electron correlation to the computed properties. In addition, single point MP2 and coupled-cluster (CC) singles, doubles and non iterative triples (CCSD(T))[33] calculations were performed using the aug-cc-pVDZ[34] basis set at the MP2/6-31++G(d,p) geometries. The CCSD(T) calculations were done explicitly for clusters up to n=5. For n=6-8, the CCSD(T) values were obtained using the extrapolation technique given in Eq.(1):

$$\Delta E_{EST}(CCSD(T))_{aug\text{-}cc\text{-}pVDZ} = \Delta E(MP2)_{aug\text{-}cc\text{-}pVDZ} + |\Delta E(CCSD(T)\text{-}\Delta E(MP2)|_{6\text{-}31+G^*} \qquad (1)$$

In order to verify that all stationary points are minima, the structures were characterized by calculating the Hessians (matrix of energy second derivatives) using the Hartree-Fock method for both the solute and the solvent water molecules. A positive definite Hessian implies a local minimum. HF zero point energy (ZPE) corrections scaled by an empirical factor of 0.89[35] were included in all of the reported energy differences.

The adiabatic ionization potentials (AIP) for Na in water clusters ($Na(H_2O)_{n(0\text{-}12)}$) were calculated by taking the energy differences between the global minimum structures of $Na(H_2O)_{n(0\text{-}12)}$ and $Na^+(H_2O)_{n(0\text{-}12)}$ clusters for each value of n. Another set of AIP values were obtained by computing the energy differences between the global minimum $Na(H_2O)_{n(0\text{-}12)}$ clusters and the local minimum $Na^+(H_2O)_{n(0\text{-}12)}$ clusters for each value of n. The $Na(H_2O)_{n(0\text{-}12)}$ global minimum geometry was used as an initial geometry for the structural

optimization of $Na^+(H_2O)_{n(0-12)}$ clusters for each n, using ab initio MP2/6-31G++(d,p) level of theory. These AIP values are called AIP(L) where 'L' indicates a local optimization (formation of a local minimum structure) of the $Na^+(H_2O)_{n(0-12)}$ clusters. The vertical ionization potentials (VIP) were calculated by taking the energy difference between $Na(H_2O)_{n(0-12)}$ and $Na^+(H_2O)_{n(0-12)}$ clusters at the global minimum geometry of the $Na(H_2O)_{n(0-12)}$ clusters.

The HF/EFP1 and MP2 total and differential binding energies of the solute to the water clusters were calculated at HF/EFP1 and MP2 levels of theory. All the calculations were done using the General Atomic and Molecular Electronic Structure System (GAMESS)[36] program package except for the CCSD(T) calculations which were performed using ACESII[37].

III. Results and Discussions

A. $Na^+(H_2O)_{n(0-12)}$: Figures 1-3 illustrate the $Na^+(H_2O)_{n(1-12)}$ local and global minimum energy structures for each value of n. Each structure in Figures 1-3 is given a name "nI" where n represents the number of water molecules in the cluster. The alphabetical letter 'I' is used for uniqueness of each structure. (HF/EFP1) and [MP2] relative energies (kcal/mol) between the global and local minimum energy structures are given beneath each structure. Three lowest energy structures of $Na^+(H_2O)_n$ are presented for each value of n and the energies are relative to the zero energy structure (global minimum structure), indicated as (0) or [0]. The label (X+Y+Z) represents the solvation shells around the sodium ion with the first (X), second (Y) and third (Z) shell solvent water molecules. The first shell solvent water molecules are those

that participate in direct H-bonding with the solute. The second shell solvent water molecules H-bond with the first shell water molecules. The third shell water molecules H-bond with the second shell and so on. For example, the notation (4+2+0) in structure 6A in Figure 2 means that there are four water molecules in the first shell, two in the second shell and no water in the third shell. The total number of water molecules can be obtained by adding the first, second and third solvation shell water molecules. Both HF/EFP1 and MP2 levels of theory predict the same global minimum structure for a given value of n, except at n=10 where only ~1 kcal/mol of difference in the relative energy is observed (10A and 10B in Figure 3). The relative energy between the local and global minimum structure changes with the level of theory and the quantitative agreement between the two methods is typically within ~ 2 kcal/mol. The global minimum search using MC/SA simulations predicted that at least 12 water molecules are required to completely solvate the sodium ion. The global minimum structures of $Na^+(H_2O)_n$ at n=1, 2 and 3 (1A, 2A and 3A in Figure 1) are surrounded by water molecules but are planar and therefore cannot be considered as interior structures. The local minimum structures at n=4, (4B and 4C, Figure 1) are also planar but the global minimum structure (4A) is non planar and can be vaguely (because the cluster still has some empty space) considered as an interior structure. Therefore the first interior global minimum structure occurs at n=4 (4A) at both HF/EFP1 and MP2 levels of theory. The energy difference between the interior (4A) and planar structures (4B and 4C) is of the order of ~1 kcal/mol at MP2 level. For $Na^+(H_2O)_5$ the local and global minimum structures are found to be the interior structures with 4-5 water molecules directly interacting with the sodium ion. MP2 global minimum structure, $Na^+(H_2O)_5$ (5A), has a coordination number of 4 and is ~0.6-0.7 kcal/mol lower in energy relative to the structure with a coordination number of 5 (5B

and 5C). Although there is a large amount of space that is not filled with water molecules around the $Na^+(H_2O)_n$ clusters for n=4-5, these can still be considered interior clusters because the sodium ion is not lying on the surface of the water cluster (see surface structures of $Na(H_2O)_n$ in Figures 4-6). At n=6, the coordination number of sodium ion in the global minimum structure (6A) remains 4 at both the levels of theory (Figure 2). The local minimum structures 6C and 6A with coordination number 5 are 0.8-1.2 kcal/mol higher in energy than 6A at MP2 level. As the cluster size grows from n=7 to n=10, the sodium ion begins to be surrounded by more water molecules. However, no local minimum structure is observed that is completely solvated until n=12. A coordination number of 6 is preferred in the global minimum structures of $Na^+(H_2O)_n$ for n=8-10 (8A, 9A and 10A in Figure 2 and 3) at the MP2 level of theory. The first completely solvated global minimum structure for $Na^+(H_2O)_n$ is found at n=12 with a coordination number of 6 (See 12A Figure 6) . A local minimum structure (12C) with a coordination number of 6 is about 4 kcal/mol higher in energy relative to the global minimum structure at the MP2 level of theory. The $Na^+(H_2O)_n$ clusters show an increase in the coordination number as the cluster size grows from n=1-12 and the preferred coordination number is 6 at n=12 where the sodium ion is completely solvated.

The average H-bond distances between the solute and the first shell solvent molecules of $Na^+ (H_2O)_{n(1-12)}$ clusters are listed in Table 1 at each n. As n increases, the average H-bond lengths between the sodium ion and the first shell water molecules increase. However, the HF/EFP1 bond lengths are longer than the MP2 bond lengths but the trend agrees qualitatively. The average MP2 Mulliken charges on sodium the ion and water molecules of $Na^+(H_2O)_{n(1-12)}$ clusters are also listed in Table 1. The MP2 Mulliken charge on the sodium

ion becomes less positive as the cluster size grows. Also, the overall average charge on the first shell water molecules becomes more positive with increasing n, suggesting a charge transfer from the solvent water molecules to the solute. This is evident form the gain of 0.02 - 0.21 a.u of Mulliken charge is observed on the first shell water molecules in $Na^+(H_2O)_{n(1-12)}$ clusters, as n increases from 1 to 12. It is also observed that as n increases, the water molecules in $Na^+(H_2O)_n$ clusters become more polarized than the water molecule in the free state. The oxygen atom becomes more negative and hydrogen atom becomes more positive.

B. $Na(H_2O)_{n(0-12)}$: Figures 4-6 present up to three lowest energy structures of $Na(H_2O)_n$ for each value of n. The labeling beneath the structures is the same as that for $Na^+(H_2O)_n$ clusters. In Figures 4-6 it is found that the relative energies between the $Na(H_2O)_n$ global and local minimum structures predicted by the HF/EFP1 method are in good agreement with the corresponding values at MP2, with deviations of ~ 2.5 kcal/mol. The two methods also predict the same global minimum structures of $Na(H_2O)_n$ for a given value of n except for n=4 where the deviation is up to ~2 kcal/mol. The global minimum search for $Na(H_2O)_n$ using MC/SA simulations suggests that the sodium atom stays on the surface of the water cluster for n=1-12. It is also found that all the $Na(H_2O)_n$ local and global minimum structures that are considered here are surface structures. This is in line with the studies by Hashimoto et al where it is predicted that the most stable structures of $Na(H_2O)_n$ are surface structures and not interior for $n \geq 4$.[17] In the present paper, there were no structures found in the global minimum search that were interior structures for n=1-12. An attempt to form an interior structure at n=10 resulted in a very high-energy structure (10I) shown in Figure 6. The

energy difference of 8.9 kcal/mol is observed between the surface global minimum structure (10A) and the interior structure (10I) of $Na(H_2O)_n$ at the MP2 level of theory.

In the global minimum structures of $Na(H_2O)_n$ for n=1-3 (see Figure 4), the first solvation shell has only 1 water molecule. The global minimum structure 3A has only one water molecule in the first solvation shell and is 0.4 kcal/mol lower in energy than the structure 3B with 3 water molecules directly coordinated to the sodium atom at the MP2 level of theory. When n=4, the global minimum structure (4A) at the Hartree-Fock level of theory has 3 water molecules directly H-bonded to the sodium atom whereas MP2 predicts a 3 coordinated global minimum structure (4C) which is 1.5 kcal/mol lower in energy than 4A. Even after adding water molecules beyond n=5 to the sodium atom, the number of water molecules directly interacting with the sodium atom generally remains either 2 or 3 in all the global minimum structures. (See structures 6A, 7A, 8A, 9A, 10A and 12A in Figures 5 and 6). This observation is in contrast to $Na^+(H_2O)_n$ clusters, where the coordination number of sodium ion increases with the cluster size (Section A). The reticence of sodium atom to forming H-bonds with water molecules and the propensity to stay on the surface of the water clusters is due to the weaker Na--H_2O interactions as compared to the water-water interactions. This is evident from the smaller CCSD(T)/aug-cc-pVDZ binding energy for Na--H_2O (-6.5 kcal/mol) than water--water binding energy (-5.48 kcal/mol). A very high binding energy of Na^+--H_2O (-24.1 kcal/mol) in comparison to the Na--H_2O and H_2O-H_2O binding energy illustrates the formation of Na^+ as an interior solute in water clusters.

C: Ionization Potential: The adiabatic ionization potential (AIP) of sodium in water clusters is computed by calculating the energy difference between the global minimum structures of

$Na(H_2O)_{n(0-12)}$ and $Na^+(H_2O)_{n(0-12)}$ clusters at HF/EFP1, MP2 and CCSD(T) levels of theory. Figure 7 shows a plot between the AIP values of $Na(H_2O)_n$ clusters and the number of water molecules (n=0-12) calculated using HF/EFP1, MP2 and CCSD(T) methods. It is observed that the MP2 and HF/EFP1 AIP values reduce at a faster rate when the cluster size is small (n=1-6) and then become almost insensitive to the change in the cluster size. Although, it is observed that the MP2 and HF/EFP1 AIP's appear to be converging very slowly contrary to the experimental findings where the IP values become constant after n=4.[4] The CCSD(T) AIP values also followed a similar trend as HF/EFP1 AIP values.

Table 2 shows the comparison of the calculated AIP values with experimental and previously calculated AIP values. The AIP values at HF/EFP1,MP2 and CCSD(T) methods are very close to each other for smaller (n=1-5) clusters of $Na(H_2O)_n$ (See Table 2) suggesting a negligible influence of electron correlation. The effect of electron correlation increases as the cluster size grows. The importance of electron correlation in the AIP values is accentuated when more sophisticated levels of theory with larger basis sets are used and the AIP values show a better agreement with the experimental values[4,11]. Table 2 shows that the computed AIP values are below the experimental values and previously calculated values by Gao et al[23] using density functional theory (DFT) based molecular dynamics with a plane wave basis set. The AIP values calculated in the present study are lower at all three levels of theory than the predicted values by Gao et al. This is mainly because in the present work, the AIP values were calculated at the global minimum geometry of $Na^+(H_2O)_{n(0-12)}$ and $Na(H_2O)_{n(0-12)}$ clusters, whereas in the work by Gao et al, the $Na^+(H_2O)_n$ clusters were only locally optimized. The AIP(L) values were also calculated here by locally optimizing the $Na^+(H_2O)_n$ clusters starting at the global minimum geometry of $Na(H_2O)_n$ clusters. Table 3 lists the

AIP(L) values calculated at the HF/EFP1, MP2 and CCSD(T) levels of theory and it was found that these values were in closer agreement with the experimental values relative to the AIP values.

In order to explore the importance of relaxation of the solvent water molecules around sodium ion in $Na^+(H_2O)_{n(0-12)}$ structures, the vertical ionization potential (VIP) values were also calculated and the values are listed in Table 4. The VIP values were calculated by taking the energy difference between the $Na^+(H_2O)_{n(0-12)}$ and $Na(H_2O)_{n(0-12)}$ clusters at the global minimum geometry of $Na(H_2O)_{n(0-12)}$ clusters. This calculation of the VIP represents the scenario where the water molecules around the sodium atom are not allowed to relax once the sodium ion is formed.

Figure 8 shows the comparison of AIP, AIP(L) and VIP values calculated using MP2/aug-cc-pVDZ method with the experimental values. Large fluctuations in the VIP values are observed as the cluster size grows. This can be attributed to the lack of relaxation of the water molecules around the $Na^+(H_2O)_{n(0-12)}$ clusters. The VIP values are found to be higher relative to the experimental values, where as both AIP and AIP(L) values lie below the experimental curve (see Figure 8). The comparison of the calculated IP values clearly indicates that the VIP values, that involve no solvent water relaxation around the sodium ion, cannot be correlated with the experimental findings. This suggests that some reorganization of the surrounding solvent molecules around sodium ion, is necessary to correlate the IP values with the experimental measurements. Previous studies[13,23] have also argued that the VIP values calculated theoretically cannot be correlated with the experimental observations and there is a large reorganization (of solvent molecules around sodium ion) involved in the experimental determination of the IP values of $Na(H_2O)_n$ clusters.

D: Binding energies: The total binding energies (TBE) and differential binding energies (DBE) are calculated for $Na(H_2O)_{n(1-12)}$ and $Na^+(H_2O)_{n(1-12)}$ clusters at the HF/EFP1 and MP2 levels of theory using Eqs. (2) and (3) respectively.

DBE: $E[A\ (H_2O)_n] - E[A(H_2O)_{n-1}] - E(H_2O)$ (2)

TBE: $E(A\ (H_2O)_n) - \{E(A\ ) + nE(H_2O)\}$ (3)

where A=Na (Na^+) and n=1-12. $E[A(H_2O)_n]$ and $E[A(H_2O)_{n-1}]$ are the Boltzmann-averaged energies calculated using Eq. (4)

$$E_n = \frac{\sum_i X_i Exp(-\Delta E_i / RT)}{\sum_i Exp(-\Delta E_i / RT)} \quad (4)$$

Where X_i is the energy of the i^{th} structure including zero point energy correction. ΔE_i is the energy difference between the i^{th} and the global minimum structure at a particular value of n (T=298K).

The calculated and the experimental[38] DBE values for $Na^+(H_2O)_{n(1-12)}$ are shown in Table 5. Both the experimental and calculated DBE values using HF/EFP1 and MP2 methods decrease for n=1-6 with some fluctuations observed in the calculated values at n=4 and 6. For n>6, the calculated DBE values at both the levels of theory still show some fluctuations. HF/EFP1 DBE values show larger deviations from the experimental values with percentage errors of 4-18% relative to the deviations of 5-13%) observed for MP2 DBE values. Table 6 illustrates the HF/EFP1 and MP2 calculated TBE values for $Na^+(H_2O)_{n(1-12)}$. The MP2 TBE values are in better agreement with the experimental values[38] as shown by small percentage

errors of 1%-5%. Although the HF/EFP1 TBE values show larger percentage errors but the trend agrees qualitatively with the MP2 values.

The TBE and DBE values for $Na(H_2O)_{n(1-12)}$ clusters at the HF/EFP1 and MP2 levels of theory are given in Table 7 and 8 respectively. The TBE values show an increasing trend as the cluster size grows at both the HF/EFP1 and MP2 levels of theory. Comparison of the calculated DBE values of $Na(H_2O)_{n(1-12)}$ with the experimental values are presented in Table 8. DBE values of $Na(H_2O)_{n(1-12)}$ show an increase as the cluster size increases from n=1-4. However, at larger n, the DBE values show fluctuations at both the HF/EFP1 and MP2 levels of theory. There are significant percentage errors observed in the DBE values calculated at the HF/EFP1 level of theory.

IV. Conclusions

The study herein shows a systematic calculation of the ionization potential (IP) of $Na(H_2O)_n$ clusters using correlated levels of theory. The IP values were calculated by adding up to 12 water molecules to the sodium and sodium ion. For both the $Na(H_2O)_n$ and $Na^+(H_2O)_n$ clusters, the global minimum structures were found using Monte Carlo simulation annealing method at HF/EFP1 level of theory. The AIP and VIP values were calculated by taking the energy difference between the $Na(H_2O)_n$ and $Na^+(H_2O)_n$ clusters. In order to correlate the calculated IP values with the experimental measurements, the influence of electron correlation and the importance of solvation structure (relaxation) around sodium cation in $Na^+(H_2O)_n$ clusters was investigated in this study. The comparison of the IP values with experimental values shows that it is important to account for the relaxation of the water

molecules around the sodium cation and the VIP values cannot be correlated with the experimental measurements. This observation is found to be consistent with the previous theoretical studies by Gao et al[23] and Cwiklik et al[20]. The results also show that the AIP (L) values are in better agreement with the experimental measurements than AIP values. This suggests that $Na^+(H_2O)_n$ clusters do not reorganize to form global minimum structures on ionization, however, some structural reorganization do occur. The IP values calculated at MP2 and CCSD(T) levels of theory at larger basis set are in better agreement with the experimental IP values relative to the HF/EFP1 level of theory, suggesting the importance of electron correlation and basis set. However, it can be concluded that both the structural changes in the ionic clusters and the electron correlation effects are important in theoretical calculation of the IPs to be quantitatively correlated with the experiments.

Acknowledgements

This work has been supported by a grant from the U.S. Department of Energy, administered by the Ames Laboratory (M.S.G), and by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (Grant No. DE-FG02-01ER15228; P.P).

References

(1) Hart, E. J.; Boag, J. W. *Journal of the American Chemical Society* **1962**, *84*, 4090.
(2) Baxendal.Jh; Wardman, P. *Nature* **1971**, *230*, 449.
(3) Kraus, C. A. *Journal of the American Chemical Society* **1908**, *30*, 1323.
(4) Hertel, I. V.; Huglin, C.; Nitsch, C.; Schulz, C. P. *Physical Review Letters* **1991**, *67*, 1767.
(5) Buck, U.; Dauster, I.; Gao, B.; Liu, Z. F. *Journal of Physical Chemistry A* **2007**, *111*, 12355.

(6) Ferro, Y.; Allouche, A.; Kempter, V. *Journal of Chemical Physics* **2004**, *120*, 8683.
(7) Jeevarajan, A. S.; Fessenden, R. W. *Journal of Physical Chemistry* **1989**, *93*, 3511.
(8) Graf, P.; Nitzan, A.; Diercksen, G. H. F. *Journal of Physical Chemistry* **1996**, *100*, 18916.
(9) Belloni, J.; Marignier, J. L. *Radiation Physics and Chemistry* **1989**, *34*, 157.
(10) Schulz, C. P.; Haugstatter, R.; Tittes, H. U.; Hertel, I. V. *Physical Review Letters* **1986**, *57*, 1703.
(11) Schulz, C. P.; Haugstatter, R.; Tittes, H. U.; Hertel, I. V. *Zeitschrift Fur Physik D-Atoms Molecules and Clusters* **1988**, *10*, 279.
(12) Misaizu, F.; Tsukamoto, K.; Sanekata, M.; Fuke, K. *Chemical Physics Letters* **1992**, *188*, 241.
(13) Dauster, I.; Suhm, M. A.; Buckb, U.; Zeuch, T. *Physical Chemistry Chemical Physics* **2008**, *10*, 83.
(14) Steinbach, C.; Buck, U. *Journal of Chemical Physics* **2005**, *122*.
(15) Takasu, R.; Misaizu, F.; Hashimoto, K.; Fuke, K. *Journal of Physical Chemistry A* **1997**, *101*, 3078.
(16) Barnett, R. N.; Landman, U. *Physical Review Letters* **1993**, *70*, 1775.
(17) Hashimoto, K.; Morokuma, K. *Journal of the American Chemical Society* **1994**, *116*, 11436.
(18) Moller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 0618.
(19) Hashimoto, K.; Kamimoto, T. *Journal of the American Chemical Society* **1998**, *120*, 3560.
(20) Cwiklik, L.; Buck, U.; Kulig, W.; Kubisiak, P.; Jungwirth, P. *Journal of Chemical Physics* **2008**, *128*.
(21) Becke, A. D. *Physical Review A* **1988**, *38*, 3098.
(22) LEE, C. T.; Yang, W. T.; Parr, R. G. *Physical Review B* **1988**, *37*, 785.
(23) Gao, B.; Liu, Z. F. *Journal of Chemical Physics* **2007**, *126*.
(24) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *Journal of Chemical Physics* **1996**, *105*, 1968.
(25) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *Journal of Physical Chemistry A* **2001**, *105*, 293.
(26) Merrill, G. N.; Webb, S. P.; Bivin, D. B. *Journal of Physical Chemistry A* **2003**, *107*, 386.
(27) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* **1953**, *21*, 1087.
(28) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(29) Hehre, W. J.; Ditchfie.R; Pople, J. A. *Journal of Chemical Physics* **1972**, *56*, 2257.
(30) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *Journal of Chemical Physics* **1982**, *77*, 3654.
(31) Harihara.Pc; Pople, J. A. *Theoretica Chimica Acta* **1973**, *28*, 213.
(32) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *Journal of Computational Chemistry* **1983**, *4*, 294.
(33) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Headgordon, M. *Chemical Physics Letters* **1989**, *157*, 479.
(34) Dunning, T. H. *Journal of Chemical Physics* **1989**, *90*, 1007.

(35) Scott, A. P.; Radom, L. *Journal of Physical Chemistry* **1996**, *100*, 16502.
(36) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *Journal of Computational Chemistry* **1993**, *14*, 1347.
(37) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *International Journal of Quantum Chemistry* **1992**, 879.
(38) Dzidic, I.; Kebarle, P. *Journal of Physical Chemistry* **1970**, *74*, 1466.

FIGURES CAPTIONS

Figure 1*:* Lowest energy structures of $Na^+(H_2O)_n$ for n=1-5. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 2*:* Lowest energy structures of $Na^+(H_2O)_n$ for n=6-8. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 3: Lowest energy structures of $Na^+(H_2O)_n$ for n=9-12. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 4: Lowest energy structures of $Na(H_2O)_n$ for n=1-5. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 5: Lowest energy structures of $Na(H_2O)_n$ for n=6-8. (X+Y+Z+W+V) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z) and so on. The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 6: Lowest energy structures of $Na(H_2O)_n$ for n=9-12. (X+Y+Z+W+V) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z) and so on. The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Figure 7: Plot of adiabatic ionization potential (AIP) of $Na(H_2O)_n$ with increasing number of water molecules (n). The n=6-8 AIP values at CCSD(T) level of theory are extrapolated using eq. (1) (see text).

Figure 8: Comparison of calculated ionization potential (IP) of $Na(H_2O)_n$ with experimental measurements

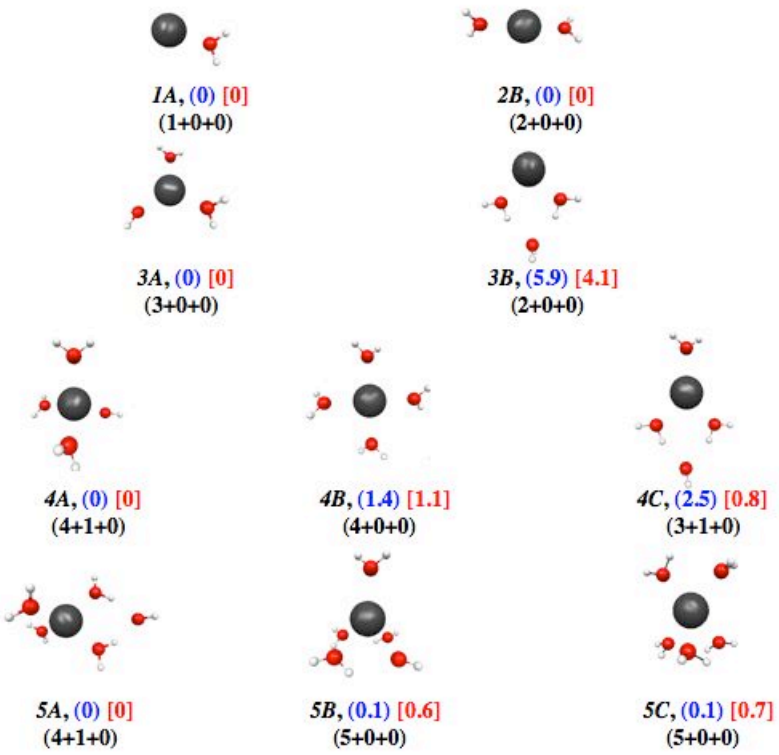


Figure 1: Lowest energy structures of $Na^+(H_2O)_n$ for n=1-5. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

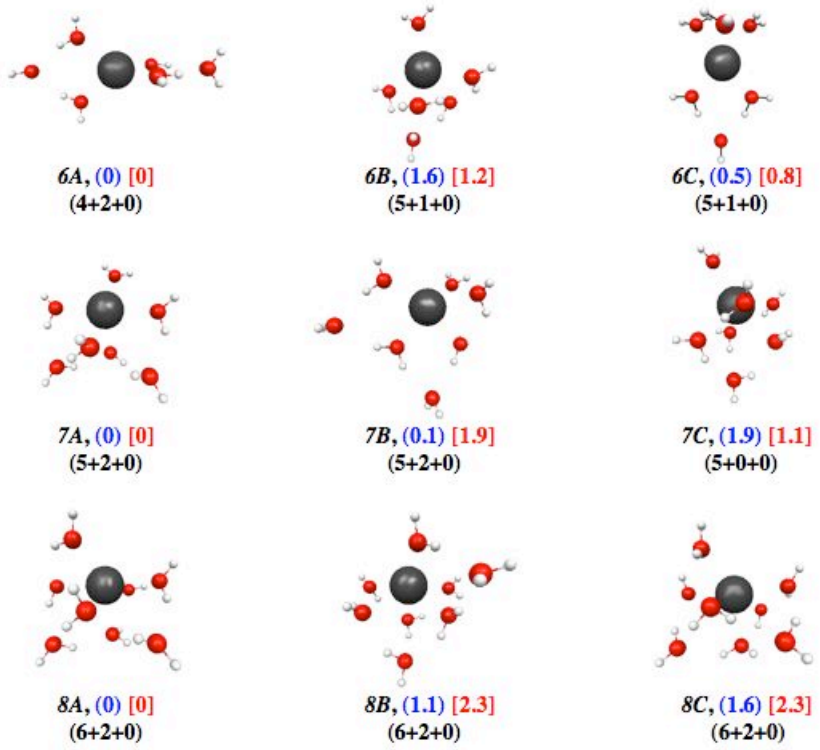


Figure 2: Lowest energy structures of $Na^+(H_2O)_n$ for n=6-8. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

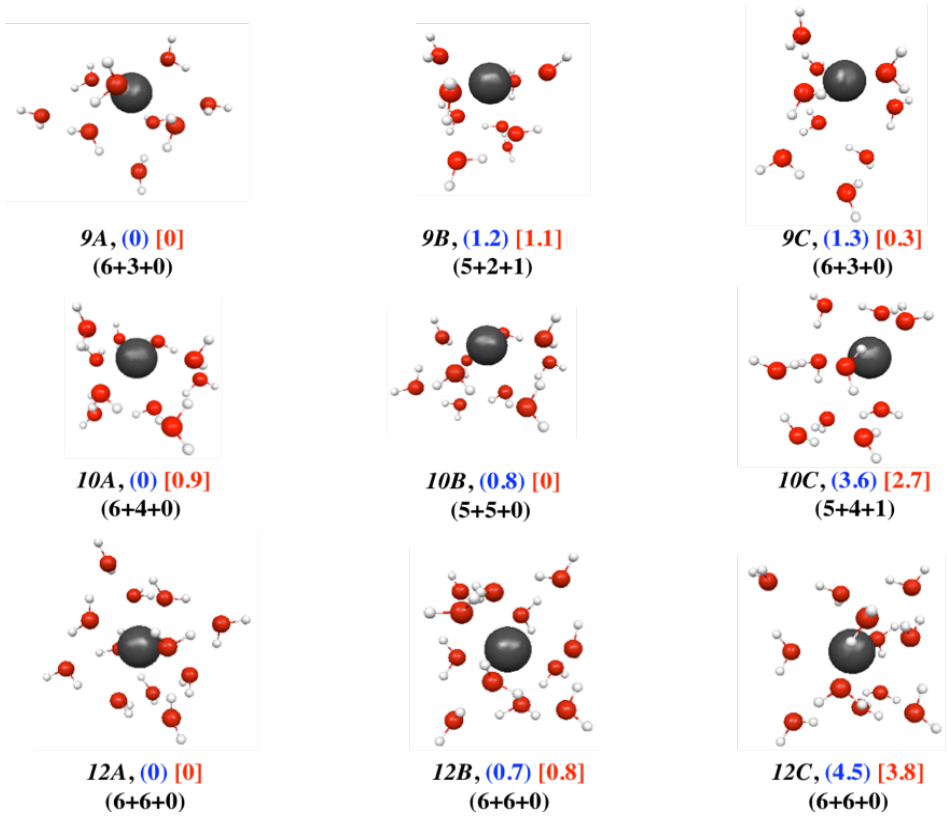


Figure 3: Lowest energy structures of $Na^+(H_2O)_n$ for n=9-12. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

Table 1: The average MP2 Mulliken charges and H-bond distances for the global minimum structures of $Na^+(H_2O)_n$

number of water molecules	Mulliken Charges (a.u)				R(Å)	
	q(Na^+)	(H_2O)			HF/EFP1	MP2
		q(O)	q(H)	q(H_2O)		
0		-0.71	0.35	0.00		
1	0.98	-0.81	0.41	0.02	2.39	2.30
2	0.97	-0.80	0.41	0.02	2.40	2.29
3	0.99	-0.80	0.40	0.00	2.43	2.32
4	0.98	-0.84	0.42	0.01	2.46	2.35
5	1.02	-0.83	0.41	0.00	2.45	2.35
6	1.02	-0.85	0.43	0.00	2.45	2.34
7	0.97	-0.86	0.43	0.00	2.50	2.43
8	0.97	-0.86	0.43	0.00	2.53	2.47
9	1.00	-0.87	0.44	0.00	2.53	2.52
10	0.90	-0.88	0.45	0.02	2.49	2.44
12	0.81	-0.90	0.55	0.21	2.50	2.48

R is the average H-bond distance between the sodium ion and the first shell solvent molecules. *q*(O) and *q*(H) are the average Mulliken charges on oxygen and hydrogen atoms of H_2O molecules respectively. *q*(H_2O) is the average Mulliken charge on a water molecule.

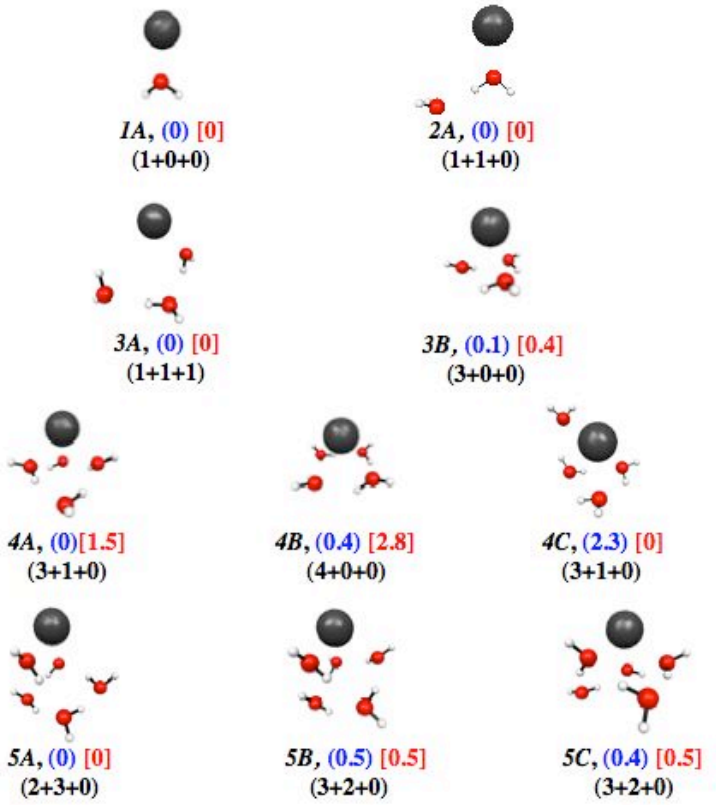


Figure 4: Lowest energy structures of $Na(H_2O)_n$ for n=1-5. (X+Y+Z) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z). The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

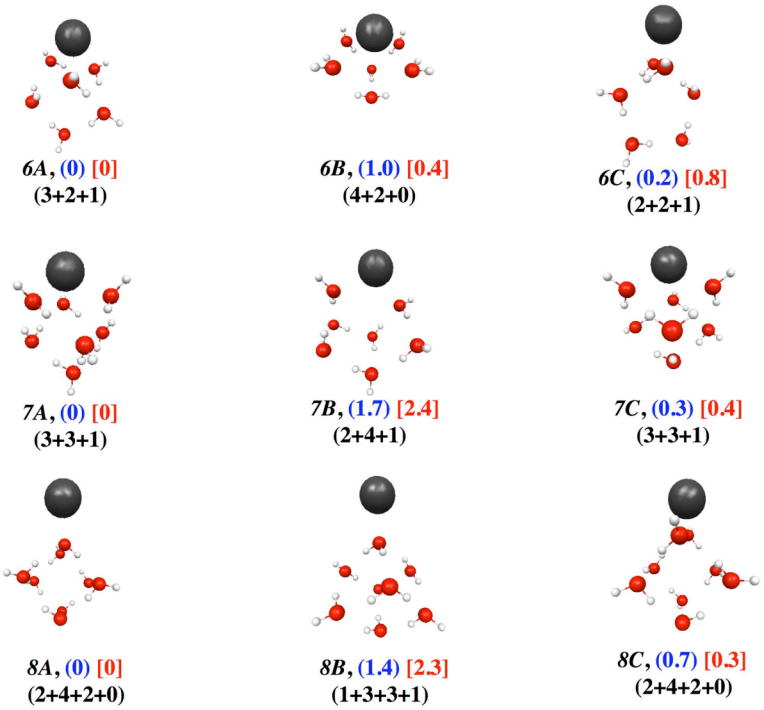


Figure 5: Lowest energy structures of $Na(H_2O)_n$ for n=6-8. (X+Y+Z+W+V) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z) and so on. The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

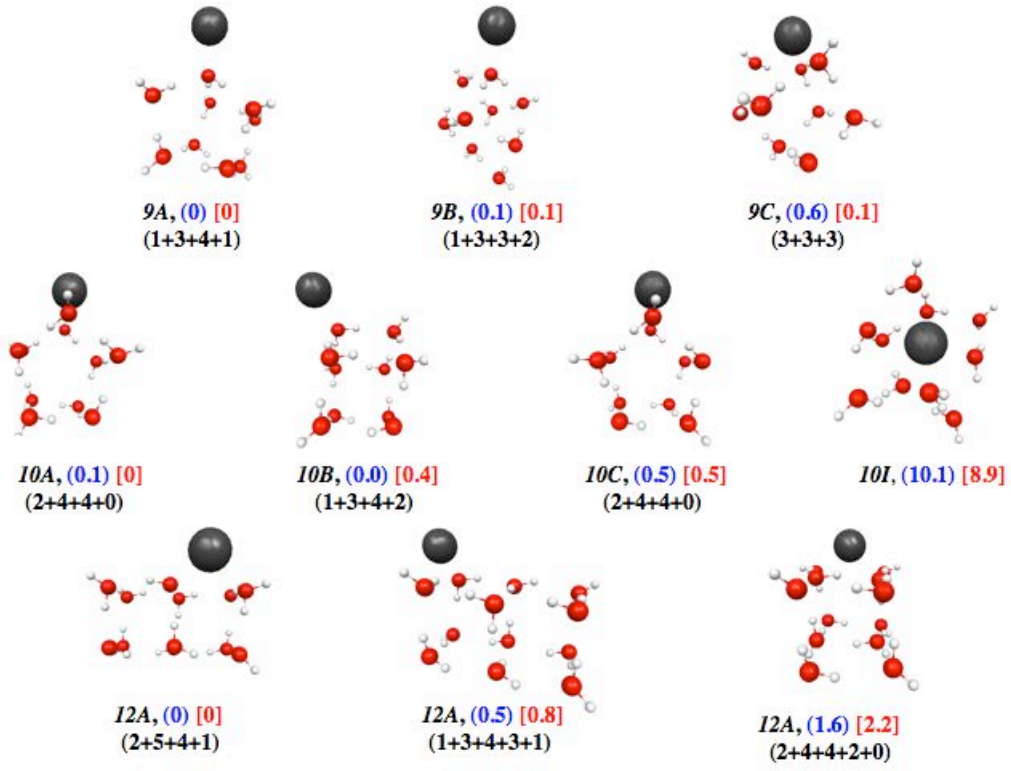


Figure 6: Lowest energy structures of $Na(H_2O)_n$ for n=9-12. (X+Y+Z+W+V) indicate the number of solvent molecules in the first shell (X), second shell (Y), and third shell (Z) and so on. The (HF/EFP1) and [MP2] relative energies are given in kcal/mol. A (0) and [0] relative energy suggests the global minimum structure.

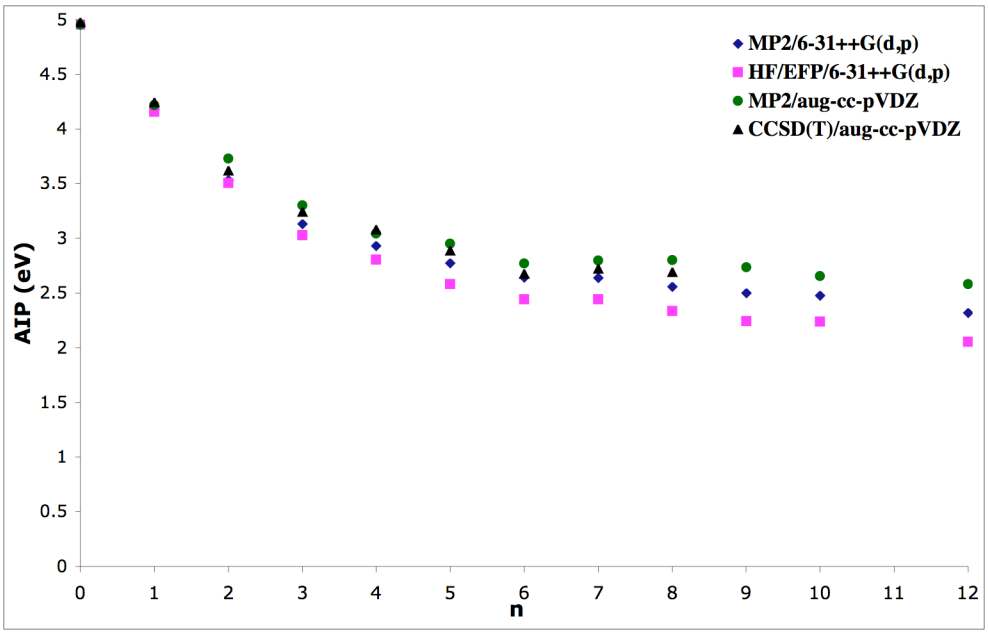


Figure 7: Plot of adiabatic ionization potential (AIP) of $Na(H_2O)_n$ with increasing number of water molecules (n). The n=6-8 AIP values at CCSD(T) level of theory are extrapolated using eq. (1) (see text).

Table 2: Adiabatic ionization potential of $Na(H_2O)_n$ at HF/EFP1, MP2 and CCSD(T) levels of theory. The comparison is done with previous calculations and experiments. All the values are in eV. All the values are ZPE corrected.

number of water molecules	6-31++G(d,p)		aug-cc-pVDZ		Gao et al[23]	Expt[4,11]
	HF/EFP1	MP2	MP2	CCSD(T)		
0	4.96	4.96	4.98	4.98 (5.00)		5.14
1	4.16	4.18	4.22	4.25		4.38±0.03
2	3.51	3.53	3.73	3.62		3.80±0.05
3	3.03	3.13	3.30	3.24		3.48±0.06
4	2.80	2.93	3.05	3.08	3.32-3.37	3.20±0.10
5	2.58	2.77	2.95	2.90	3.13-3.45	3.20±0.10
6	2.44	2.65	2.77	2.68	3.10-3.42	3.20±0.10
7	2.43	2.64	2.79	2.72	3.11-3.29	
8	2.33	2.56	2.79	2.70	3.05-3.18	
9	2.24	2.50	2.73			
10	2.23	2.48	2.66		2.96-3.09	
12	2.05	2.32	2.58			

[a]The value in parenthesis is at aug-cc-pVTZ basis set

Table 3: Adiabatic ionization potential AIP (L) of $Na(H_2O)_n$ at HF/EFP1, MP2 and CCSD(T) levels of theory. The comparison is done with previous calculations and experiments. All the values are in eV. All the values are ZPE uncorrected.

number of water molecules	6-31++G(d,p)		aug-cc-pVDZ		Gao et al[23]	Expt[4,11]
	HF/EFP1	MP2	MP2	CCSD(T)		
0	4.96	4.96	4.98	4.98		5.14
1	4.13	4.15	4.19	4.14		4.38±0.03
2	3.51	3.83	3.64	3.62		3.80±0.05
3	3.05	3.15	3.32	3.26		3.48±0.06
4	2.88	3.10	3.19	3.18	3.32-3.37	3.20±0.10
5	2.78	2.92	3.08	3.01	3.13-3.45	3.20±0.10
6	2.56	2.75	2.90	2.82	3.10-3.42	3.20±0.10
7	2.55	2.79	2.95	2.81	3.11-3.29	
8	2.82	2.93	3.21	2.81	3.05-3.18	
9	2.66	2.79	2.94			
10	2.50	2.69	2.85		2.96-3.09	
12	2.80	2.88	3.01			

Table 4: Vertical ionization potential (VIP) of $Na(H_2O)_n$ at HF/EFP1, MP2 and CCSD(T) levels of theory. The comparison is done with previous calculations and experiments. All the values are in eV.

number of water molecules	6-31++G(d,p)		aug-cc-pVDZ		Gao et al[23]	Expt[4,11]
	HF/EFP1	MP2	MP2	CCSD(T)		
0	4.96	4.96	4.98	4.98		5.14
1	4.15	4.17	4.21	4.40		4.38±0.03
2	3.99	3.97	4.00	4.05		3.80±0.05
3	4.15	4.24	4.24	4.30		3.48±0.06
4	3.50	3.51	3.54	3.63	3.77-3.98	3.20±0.10
5	3.67	3.70	3.72	3.80	3.76-4.10	3.20±0.10
6	3.37	3.40	3.43		3.58-3.91	3.20±0.10
7	3.27	3.32	3.35		3.49-3.92	
8	4.15	4.11	4.11		3.59-3.80	
9	4.08	4.04	4.05			
10	3.98	3.93	3.65		3.22-3.97	
12	3.64	3.67	3.68			

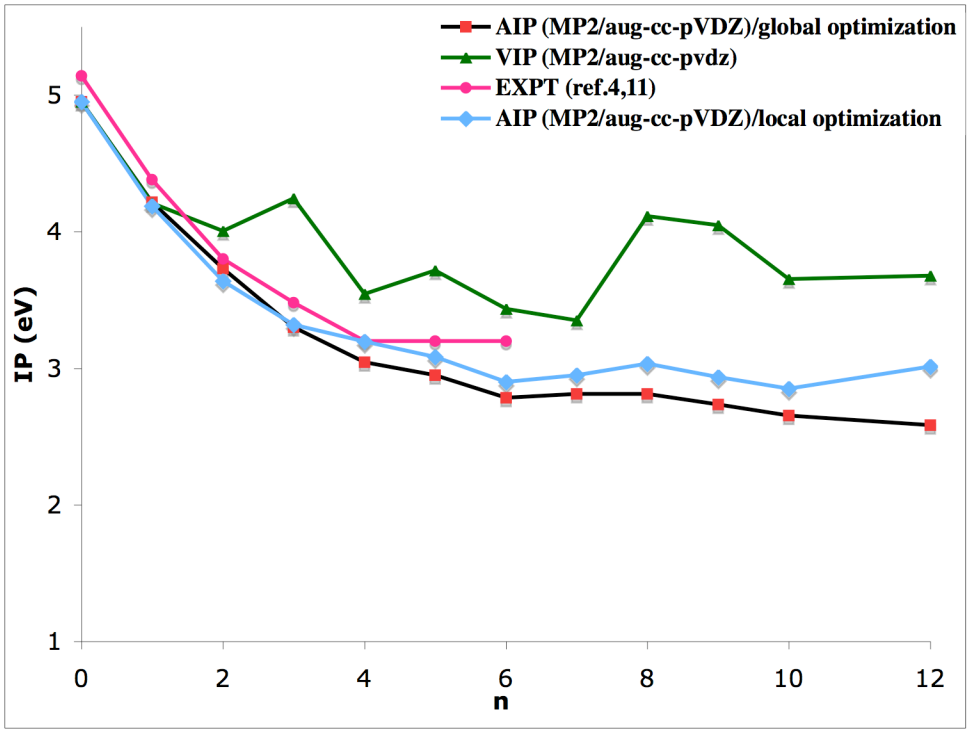


Figure 8: Comparison of calculated ionization potential (IP) of $Na(H_2O)_n$ with experimental measurements.

Table 5: HF/EFP1 and MP2 differential binding energies for $Na^+(H_2O)_n$

number of water molecules	Experiment[40]	HF/EFP1		MP2	
	binding energies (kcal/mol)	binding energies (kcal/mol)	%error[b]	binding energies (kcal/mol)	%error[b]
1	-24.0	-22.3	-7.1	-22.8	-4.9
2	-19.8	-20.3	2.8	-20.8	5.1
3	-15.8	-13.9	-11.7	-14.9	-5.7
4	-13.8	-15.8	14.5	-15.6	13.1
5	-12.2	-10.1	-17.9	-13.1	6.4
6	-10.6	-11.1	3.9	-11.5	8.0
7		-8.3		-10.4	
8		-8.3		-11.0	
9		-8.1		-11.4	
10		-7.7		-11.3	
12		-10.3		-7.7	

[a]Data taken from ref 40. [b]The percentage errors are calculated by taking the difference between the calculated and the experimental value at each value of n and dividing with the experimental value and multiplying by 100.

Table 6: HF/EFP1 and MP2 total binding energies for $Na^+(H_2O)_n$

number of water molecules	Experiment[40]	HF/EFP1		MP2	
	binding energy (kcal/mol)	binding energy (kcal/mol)	%error[b]	binding energy (kcal/mol)	%error[b]
1	-24.0	-22.3	-7.1	-22.8	-5.0
2	-43.8	-42.7	-2.6	-43.6	-0.4
3	-59.6	-56.6	-5.0	-58.5	-1.8
4	-73.4	-72.4	-1.4	-74.1	1.0
5	-85.7	-82.5	-3.7	-87.2	1.8
6	-96.4	-93.6	-2.9	-98.8	2.5
7		-101.9		-109.2	
8		-110.2		-120.2	
9		-118.3		-131.6	
10		-125.9		-143.0	
12		-146.7		-158.2	

[a]Data taken from ref 40. [b]The percentage errors are calculated by taking the difference of the calculated value and the experimental value at each value of n and dividing with the experimental value and multiplying by 100.

Table 7: HF/EFP1 and MP2 total binding energy for $Na(H_2O)_n$

number of water molecules	HF/EFP1	MP2
1	- 3.7	- 4.9
2	-9.1	-10.8
3	-14.5	-18.4
4	-21.2	-26.6
5	-27.7	-36.5
6	-34.5	-45.7
7	-41.6	-55.7
8	-49.7	-65.6
9	-56.6	-75.4
10	-64.2	-85.9
12	-78.8	-104.5

Table 8: HF/EFP1 and MP2 differential binding energies for $Na(H_2O)_n$

number of water molecules	Expt[11,16]	HF/EFP1		MP2	
	binding energies (kcal/mol)	binding energies (kcal/mol)	%error[b]	binding energies (kcal/mol)	%error[b]
1	-6.4±0.92	-3.7	-42.1	- 4.9	- 23.4
2	-6.4±0.92	-5.4	-16.3	-5.8	-9.2
3	-8.5±3.45	-5.4	-36.7	-7.6	-10.4
4	-7.4±4.61	-6.7	-9.5	-8.1	10.5
5	-12.2±5.53	-6.6	-46.1	-9.9	-18.1
6	-10.6±5.53	-6.8	-13.8	-9.1	-13.8
7		-7.1		-10.0	
8		-8.1		-9.8	
9		-6.9		-9.8	
10		-7.6		-10.5	
12		-7.3		-9.3	

[b]The percentage errors are calculated by taking the difference of the calculated value and the experimental value at each value of n and dividing with the experimental values and multiplying by 100.

CHAPTER 6. GENERAL CONCLUSIONS

This dissertation discussed the methodology and application of QM/MM (quantum mechanics/molecular mechanics) methods and illustrated the importance of QM/MM methods in solving problems that involved large ensembles, with accuracy and a reasonable computational cost.

The second chapter of the thesis illustrates the development and implementation of approaches to interface the configuration interaction singles (CIS) method with EFP method. Test examples indicate that method 1, which indirectly includes the polarization effect due to the solvent molecules only in the ground state Hartree-Fock (HF) orbitals, makes the overwhelming majority of solvent effects. Method 2, which adds a one-time correction term to account for the solvent response for the excited state density, contributes a miniscule improvement over method 1 in predicting the solvent induced shifts. It was also concluded that the fully self consistent method that involves a coupled iterative procedure to solve both the solute wavefunction (represented by CIS) and the solvent induced dipoles (represented by EFP1/HF), to obtain an excited state, is not expected to be necessary. The calculated solvent induced shifts using method 1 and method 2 were found to be consistent with the full ab initio results. Finally, the application of CIS/EFP method to acetone and coumarin151 in the condensed phase resulted in a qualitative agreement with previous experimental measurements.

The third chapter presented the micro-solvation effects on the adiabatic electron affinity (AEA) of hydroxyl radical using a QM/effective fragment potential (EFP) method. The global minimum search for $OH^-(H_2O)_n$ clusters showed that a minimum of 15 water

molecules are required to completely solvate the hydroxide ion. In contrast, the global minima search for $OH\cdot(H_2O)_n$ clusters predicted that the OH radical has a propensity to stay at the surface even up to 15 waters. The structural analysis of $OH^-(H_2O)_n$, $OH\cdot(H_2O)_n$ and $(H_2O)_n$ clusters showed that an explicit solvent model, EFP, is able to capture the correct structural features at a reasonable computational cost. The AEA value in the presence of a continuum solvent illustrated that the value of the aqueous electron affinity has not converged at 15 water molecules.

The fourth chapter focused on the application of a hybrid QM/MM embedded cluster model called SIMOMM in predicting the mechanism and energy barriers for the O atom diffusion and the etching of the Si(100) surface. The results also highlighted the utility of correlated single reference "black-box" methods called CR-CC and CIM-CC in predicting the energy barriers of the oxidation reaction on Si(100) surface.

Finally, chapter 5 consists of a study of the evolution of the ionization potential (IP) value for the microsolvated sodium atom. It was predicted that the IP values of $Na(H_2O)_n$ exhibit a cluster size dependence and was found to be qualitatively consistent with the experimental predictions. The results also indicate that the IP values are influenced by electron correlation effects and vary significantly with the solvent structure around the sodium ion.

FUTURE PROSPECTS

The methodologies that were derived to interface the CIS method with the EFP1 method are general and can be applied to more sophisticated excited state approaches, such as EOM-CC, MRCI and CISD to capture the quantitatively correct solvent effects on excited states. In addition, the CIS/EFP1 method can be extended to interface CIS with a general EFP method (EFP2) that can account for the solvent effects due to solvents other than water. The development described here can be further extended to the gradient derivation and implementation, which will be very useful in studying solvent dynamics in excited states.

The global minimum search for $OH^-(H_2O)_n$ clusters can be used to investigate the transport mechanism of the hydroxide ion in $OH^-(H_2O)_n$ clusters. To study the mechanism the first shell water molecules can be treated by an ab initio method and the remaining solvent molecules can be represented using the EFP method. The study of the electron affinity of hydroxyl radical in aqueous phase may be extended to investigate the interaction between the hydroxyl radical and a C=C double bond in phospholipids, which are important for the mobility of the cell membrane components.

The study of a long range O atom diffusion on the Si(100) surface can be extended to study the diffusion of molecular oxygen on the Si(100) surface. It will also be interesting to investigate the diffusion mechanism for the hopping of O along the Si dimer columns instead of dimer rows. The SIMOMM method can be utilized to treat the noble metal surfaces like Au(111) to study the self-assembly of alkyl thiols on the surfaces.