Ab Initio QM/MM Simulations with a Molecular Orbital-Valence Bond (MOVB) Method: Application to an S_N2 Reaction in Water

YIRONG MO, JIALI GAO

Department of Chemistry and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455
Center for Computational Research, State University of New York, Buffalo, New York 14260

Received 4 April 2000; accepted 24 July 2000

ABSTRACT: A mixed molecular orbital and valence bond (MOVB) method is described in combined *ab initio* QM/MM simulations of the S_N2 reaction of $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in water. The method is based on the construction of individual charge-localized, diabatic states using a block-localized wave function approach, followed by configuration interaction calculations to obtain the adiabatic potential energy surface. To examine the performance of the MOVB method, modern *ab initio* VB calculations were performed. The MOVB gas phase results are found to be in reasonable agreement in the overall potential energy surface in comparison with Hartree–Fock, MP2, and *ab initio* VB calculations. The only exception is that the activation energy is predicted to be about 4 kcal/mol higher in MOVB than in other methods. This is attributed to the configuration interaction procedure, which does not further optimize orbital coefficients in MOVB calculations, and it emphasizes the importance of orbital optimization in these calculations. The adiabatic ground-state potential surface can also be approximate by other quantum chemical models in Monte Carlo QM/MM simulations. At the HF/6-31G(d) level, the calculated activation free energy of 26 kcal/mol in water is in good agreement with experiment and with previous computational results. Importantly, the MOVB method allows for the solvent reaction coordinate to be used to define the reaction path in *ab initio* QM/MM simulations. J Comput Chem 21: 1458–1469, 2000

Keywords: combined QM/MM method; valence bond theory; mixed molecular orbital and valence bond method; solvent effects on S_N2 reaction; Monte Carlo simulations of chemical reaction in solution

Correspondence to: J. Gao; e-mail: gao@chem.umn.edu
Contract/grant sponsors: The National Science Foundation and The National Institutes of Health

Introduction

A great challenge in computational chemistry is to incorporate explicit solvent effects into quantum mechanical treatments of chemical reactions in solution and in enzymes.[1] The difficulty is associated with the large number of degrees of freedom in a condensed phase system and the collective motions of the solvent molecules that influence and mediate the chemical process. Because it is not practical in the near future to carry out *ab initio* molecular dynamics simulations of reasonably large systems with meaningful statistical sampling, approximate methods must be used at the present time. Fortunately, the past decade saw significant progresses in the development of combined quantum mechanical and molecular mechanical (QM/MM) methods.[1–3] This approach has been applied to a wide range of problems of chemical and biological interest.[4–13] Although combined QM/MM methods have been remarkably successful, there are still questions to be further addressed. First, it is of interest to develop an efficient procedure to describe the potential energy surface suitable for chemical reaction in solution. Furthermore, it is desirable to be able to represent the collective solvent motions in the reaction coordinate within the combined QM/MM framework in computer simulations. The primary motivation in this study is to derive a computational method that can be effectively used in QM/MM calculations to incorporate the solvent reaction coordinate for chemical reactions in solution. This method is tested on a model S_N2 reaction in water using *ab initio* QM/MM Monte Carlo simulations.

The combined QM/MM method was first introduced by Warshel and Levitt in a seminal study of the lysozyme reaction,[14] and its recent developments have been described in several review articles.[3, 15] Following that work, Singh and Kollman used *ab initio* molecular orbital methods along with a molecular mechanics to study a number of reactions in solution and in enzymes.[16, 17] These ideas have also been explored by other researchers.[18] However, most applications during this period involved only energy optimization to determine the minimum energy path in the enzyme environment, with the exception of Warshel's work employing the empirical valence bond (EVB) approach in molecular dynamics simulations. It was not until 1987, when Bash, Field, and Karplus reported a molecular dynamics simulation of the S_N2 reaction of Cl^- and CH_3Cl in water using a molecular orbital-based QM/MM potential.[19] In that study, the reactant molecule was represented by the semiempirical Austin model 1 (AM1) method. The succinct approach that was carefully documented[20] has inspired numerous studies in the 1990s,[4–13, 21] including our own efforts in the study of solvent effects on chemical reactions and enzymatic processes with the use of semiempirical,[1] *ab initio* molecular orbital,[22, 23] and the density functional theory.[24]

A majority of combined QM/MM calculations makes use of a characteristic or a combination of geometrical variables to define the reaction coordinate.[25, 26] Although this approach is ideal for studying gas phase processes and can still yield major insights into solvent effects on chemical reactions, the complexity in hydrogen bonding interactions coupled with solvent reorganization brings about nonequilibrium solvation effects.[27–29] This problem has been the subject of a number of recent computational studies.[30–32] In one case, it was shown that the shape and barrier height for a simple proton transfer reaction in water were significantly altered if the reaction coordinate exclude solvent degrees of freedom.[30] On the other hand, other simulations showed minimal difference in activation free energy computed using geometrical or solvent reaction coordinate.[33] Although we do not aim to resolve this problem in this study, it is still useful to device a procedure that can enforce the solvent dipoles to orient in the proper direction as the reaction proceeds from the reactant, through the transition state, to the final product so that the problem can be studied at the *ab initio* level in future applications. In fact, a reasonable approach has been available to characterize the solvent reaction coordinate by the energy gap between the diabatic potential energy surfaces of the reactant and the product state in the solution, making use of the EVB method.[2, 34] However, this procedure has not been used in combined QM/MM simulations because it is difficult to efficiently define specific diabatic states within the charge-delocalized framework of molecular orbital or density functional theory. Again, an exception is the EVB approach, which is a QM/MM method, but does not suffer from the charge delocalization problem as in the MO theory because empirically fixed partial charges are used to describe the diabatic states.

Recently, a block-localized wave function (BLW) method was developed for studying the resonance stabilization, charge delocalization, and hyperconjugation effects of organic compounds.[35–37] The BLW method can be formulated to define valence bond-like resonance structures to specify diabatic

states in a chemical reaction.[33] These diabatic states, which are determined in the framework of the molecular orbital theory, can be used in combined QM/MM simulations. Furthermore, the potential energy surface for a chemical reaction can be constructed following the approach of the modern valence bond (VB) theory.[38–41] This procedure thus synthesizes both features of the molecular orbital and the valence bond theory, and is referred to as the MOVB method. We have recently applied the MOVB method to a model proton transfer reaction between NH_4^+ and NH_3 in water, demonstrating its utility in combined QM/MM Monte Carlo simulations.[33] In this article, the MOVB-QM/MM simulation method is tested on the S_N2 reaction of $Cl^- + CH_3Cl$ in water. This system has been extensively studied previously by a variety of theoretical approaches, which serves as a standard for validation of the present method.[25, 28, 42–48]

We first present a summary of the construction of diabatic states and the MOVB potential energy surface in the Method section, which is followed by a description of computational details in the Computational Details section. In Results and Discussion, we discuss the simulation results and make comparisons with previous findings. The Summary section summarizes the conclusions of the present study.

Method

THE BLOCK-LOCALIZED WAVE FUNCTION METHOD

We begin by briefly describe the block-localized wave function (BLW) method,[35–37] which is used in the present study to construct specific valence bond-like resonance structures. The main assumption in the BLW method is that the total electrons and primitive basis functions can be partitioned into k subgroups, corresponding to a specific form of the Lewis resonance, or VB structure. For simplicity, we consider only closed-shell systems within each subgroup. Thus, the ath subgroup contains n_a electrons and m_a primitive orbitals $\{\chi_\mu,\ \mu = 1, \ldots, m_a\}$. Unlike the standard Hartree–Fock (HF) theory, molecular orbitals in the BLW method are linear combinations of primitive basis orbitals that are restricted in each individual subgroup. Consequently, by construction, the charge density of the entire molecular system is localized according to the electron and orbital partition. This is made possible by taking advantage of the rather localized features of Gaussian basis functions, even when a large basis set with diffuse functions such as aug-cc-pVTZ is used.[37]

In the BLW method, molecular orbitals in subgroup a $\{\phi_i^a,\ i = 1, \ldots\}$ are expanded over atomic orbitals located on atoms within that group. The molecular wave function for resonance state r is:

$$\Psi_r = \hat{A}\left\{\Phi_1^r \Phi_2^r \ldots \Phi_k^r\right\} \tag{1}$$

where $\hat{A}$ is an antisymmetrizing operator, and Φ_a^r is a successive product of the occupied molecular orbitals in the ath subgroup.

$$\Phi_a^r = \varphi_1^a \alpha \varphi_1^a \beta \ldots \varphi_{n_a/2}^a \beta \tag{2}$$

where α and β are electronic spin orbitals, and n_a is the number of electrons in subgroup a. It is important to note that molecular orbitals in eq. (2) satisfy the following orthonormal constraints:

$$\left\langle \varphi_i^a \middle| \varphi_j^b \right\rangle = \begin{cases} \delta_{ij}, & a = b \\ w_{ij}, & a \neq b \end{cases} \tag{3}$$

where w_{ij} is the overlap integral between two molecular orbitals i and j. Clearly, molecular orbitals within each fragment are orthogonal, whereas orbitals in different subgroups are nonorthogonal.

The MOs in eq. (2) can be optimized in two ways: (1) a Jacobi rotation method that sequentially and iteratively optimizes each individual orbital, and (2) a reorthogonalization technique that has been described by Gianinetti et al.[49] The energy of the BLW resonance structure is determined by the expectation value of the Hamiltonian, which is given as follows:

$$E_r = \left\langle \Psi_r \middle| \hat{H} \middle| \Psi_r \right\rangle = \sum_{\mu=1}\sum_{\nu=1} d_{\mu\nu} h_{\mu\nu} + \sum_{\mu=1}\sum_{\nu=1} d_{\mu\nu} F_{\mu\nu} \tag{4}$$

where $h_{\mu\nu}$ and $F_{\mu\nu}$ are elements of the usual one-electron and Fock matrix, and $d_{\mu\nu}$ is an element of the density matrix $\boldsymbol{D}$ [eq. (5)].[35, 50]

$$\boldsymbol{D} = \boldsymbol{C}(\boldsymbol{C}^+\boldsymbol{S}\boldsymbol{C})^{-1}\boldsymbol{C}^+ \tag{5}$$

where $\boldsymbol{C}$ is the molecular orbital coefficient matrix, $\boldsymbol{S}$ is the overlap matrix of the basis functions, $\{\chi_\mu^a;\ a = 1, 2;\ \mu = 1, \ldots, m_a\}$, with m_a being the number of primitive basis orbitals in subgroup a.

THE MOLECULAR ORBITAL-VALENCE BOND (MOVB) METHOD

The MOVB method is perhaps most conveniently illustrated by a specific example involving the S_N2 reaction of $Cl^- + CH_3Cl$, which is studied in this work.

$$Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^- \tag{6}$$

In this system, there are four electrons and three orbitals that directly participate in bond forming

and breaking during the chemical reaction. The valence bond (VB) wave function for this process can thus be represented by a linear combination of six Slater determinants corresponding to the VB configurations resulting from this active space. In practice, however, this is not necessary because three determinants, which have very high energies, do not make significant contributions.[51] Consequently, we only need to use three configurations in the VB calculation.[51] These VB configurations, which are described by the BLW method, are listed below:

$$\begin{aligned}\Psi_1 &= \hat{A}\{\Phi(Cl^-)\Phi(CH_3Cl)\}\\ \Psi_2 &= \hat{A}\{\Phi(ClCH_3)\Phi(Cl^-)\}\\ \Psi_3 &= \hat{A}\{\Phi(Cl^-)\Phi(CH_3^+)\Phi(Cl^-)\}\end{aligned} \quad (7)$$

Here, Ψ_1 and Ψ_2 correspond to the reactant and product state, respectively, and Ψ_3 is a zwitterionic state, having two chloride anions separated by a carbocation. It is important to note that these charge states are determined irrespective of the value of the reaction coordinate in VB calculations. Thus, the product state can have very high energies in the reactant state geometry, whereas the reactant configuration will have high energies in the product region. More specifically, for the reactant state, the BLW function, Ψ_1 is constructed in the following manner: $\Phi(Cl^-)$ is a product of nine occupied MOs for a total of 18 electrons in the nucleophile chloride ion, and $\Phi(CH_3Cl)$ is a product of the 13 doubly occupied MOs in the CH_3Cl subunit. These spin-orbitals are then antisymmetrized to yield the BLW wave function for the resonance structure of the reactant state.

The adiabatic ground-state wave function for the S_N2 reaction as a function of the reaction coordinate, X, is given in eq. (8):

$$\Theta(X) = \boldsymbol{a}_1\Psi_1 + \boldsymbol{a}_2\Psi_2 + \boldsymbol{a}_3\Psi_3 \quad (8)$$

Because we are interested in developing a method for the study of chemical reactions in solution, the condensed phase system is partitioned into a reactant region that is treated quantum-mechanically, and a solvent region that is represented by a molecular mechanics force field. Thus, the effective Hamiltonian of the system is[15, 20]

$$\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}}^0_{qm} + \hat{\boldsymbol{H}}_{qm/mm} + \hat{\boldsymbol{H}}_{mm} \quad (9)$$

where $\hat{\boldsymbol{H}}^0_{qm}$ is the electronic Hamiltonian for the isolated reactant, $\hat{\boldsymbol{H}}_{qm/mm}$ is the interaction term between the reactant and solvent, and $\hat{\boldsymbol{H}}_{mm}$ is tne solvent–solvent interaction energy. QM/MM calculations have been well documented; here, we only note that the QM/MM interaction term contains two empirical parameters per atom type for the van der Waals interaction between QM and MM atoms. Parametrization procedures for obtaining these parameters can be found in ref. 22.

The ground-state potential energy is obtained by solving the secular equation by diagonalizing the Hamiltonian matrix $\boldsymbol{H}$ to yield

$$\boldsymbol{E}_g(X) = \boldsymbol{a}^+\left(\boldsymbol{O}^{-1/2}\boldsymbol{H}\boldsymbol{O}^{1/2}\right)\boldsymbol{a} \quad (10)$$

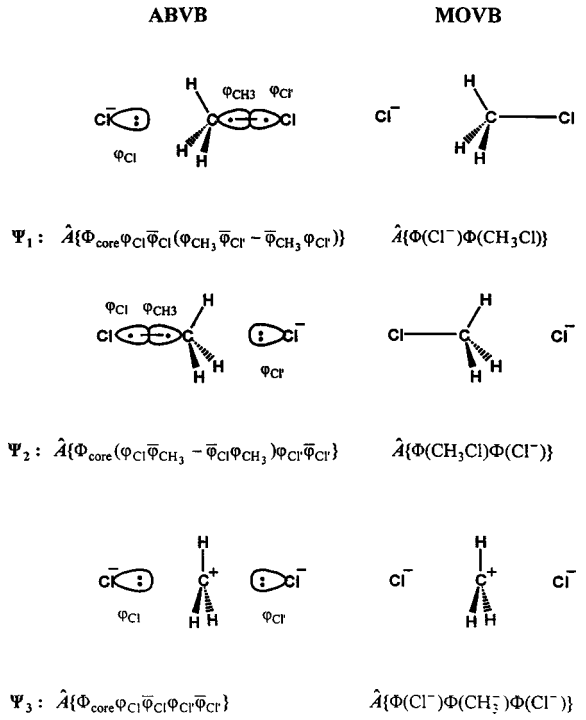
where $\boldsymbol{a}$ is the coefficient matrix, whose elements are defined in eq. (8), $\boldsymbol{O}$ is the overlap matrix of the Slater determinants of eq. (8), which are non-orthogonal, and $\boldsymbol{H}$ is the Hamiltonian matrix, whose elements are defined as $H_{st} = \langle\Psi_s|\hat{\boldsymbol{H}}|\Psi_t\rangle$. Evaluation of these matrix elements is straightforward for a given basis set, and the computational details have been described in ref. 33. It is important to point out that the Hamiltonian $\hat{\boldsymbol{H}}$ is the standard QM/MM effective Hamiltonian. Thus, the effect of solvation is directly incorporated into eq. (10), i.e., the adiabatic potential energy surface for the S_N2 reaction of $Cl^- + CH_3Cl$ corresponds to that in solution.

Computational Details

AB INITIO VALENCE BOND AND MOVB CALCULATIONS

To examine the performance of the MOVB method, we have carried out both modern *ab initio* valence bond self-consistent field (VBSCF) calculations using the 6-31G(d) basis set, and the present MOVB calculations using the 6-31G(d), and cc-pVTZ basis set on the S_N2 reaction of $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in the gas phase. In addition, the corresponding HF and MP2 results were obtained. In this comparative study, we have used the same three resonance structures in VBSCF and MOVB calculations, which include the reactant and product configurations, plus the zwitterionic state, $(Cl^-)(CH_3^+)(Cl^-)$, where the electron pairs are localized on the two chlorine atoms. For clarity, the definitions of the VBSCF and MOVB configurations are schematically illustrated in Scheme 1.

In the VBSCF calculation, each configuration is defined by an active space consisting of three localized fragment (or atomic) orbitals with two electron pairs, and a set of doubly occupied "core" molecular orbitals, Φ_{core}, for the rest of the system (Scheme 1). Electron correlation is thus treated in the VB computation for the S_N2 reaction with two determinations for the reactant and product configuration. Specifically, φ_R in Scheme 1 is a localized



SCHEME 1.

atomic or fragment orbital that is expanded over basis functions of fragment R, with R being Cl, CH_3 or Cl, and the overstrike in $\overline{\varphi}_R$ indicates occupation by an electron with the β spin. In VBSCF, both orbital coefficients and configuration coefficients are simultaneously optimized, while the active VB fragment (atomic) orbitals are kept strictly localized within each fragment. To further increase the flexibility in the VBSCF computation, φ_R and $\overline{\varphi}_R$ are optimized independently in each VB configuration, analogous to the breathing-orbital valence bond (BOVB) appraoch that has been used by Hiberty and coworkers.[52, 53] With these treatments, the electronic configuration is strictly maintained for each Lewis resonance structure in the traditional VB sense.

The difference between MOVB and VBSCF is clearly seen in Scheme 1. In MOVB, each resonance structure is treated by a single Slater determinant as opposed to two Slater determinants in VBSCF for Ψ_1 and Ψ_2. Of course, the ionic structure Ψ_3 requires only one determinant in both methods. Thus, for the C—Cl bond cleavage and formation, the MOVB method is of HF quality for the reactant and product configuration, although some electron correlation effects are introduced by configuration interaction in the VB-like calculation. The MOVB method can be extended to include spin coupling for the treatment of covalent bonds.

Another major difference between VBSCF and the present implementation of MOVB is that each VB configuration in the MOVB calculation is optimized independently, which is followed by a (3×3) CI calculation. The latter only optimizes configurational coefficients. On the other hand, in VBSCF, the VB configurations are optimized within the full 3×3 (or 6×6) self-consistent field calculation. The energies for the same configuration obtained from VBSCF optimization and from a separate, variational optimization, as is done in MOVB, are generally different, although both are of interest in interpreting bonding interactions.[54] To illustrate this difference, we also perform similar CI calculations within the *ab initio* VB framework, for which we use the notation VB-CI to distinguish it from the SCF optimization procedure. It should be pointed out that there is no theoretical restriction in MOVB, which would prevent us from optimizing orbitals in these configurations; its implementation will be described in a future publication. The present study is focused on the use of the diabatic potential surface as a mapping potential in *ab initio* QM/MM simulations to generate the correct solvent configurations.

MONTE CARLO SIMULATIONS

Statistical mechanical Monte Carlo simulations have been performed for $Cl^- + CH_3Cl$ in a box (ca. $24.6 \times 24.6 \times 36.9$ Å^3) containing 740 water molecules. Periodic boundary conditions in the isothermal-isobaric ensemble at 25°C and 1 atm are used along with the Scheraga–Owicki preferential sampling technique to enhance the local sampling. Intermolecular interactions are scaled to zero between 9.5 and 10 Å using a switching function based on heavy atom separations. Thus, if any pair atomic distance is less than 10 Å between two molecules, the entire molecular interaction is included in the energy evaluation. In the present study, the solute molecule is represented by the MOVB and HF theory using the 6-31G(d) basis function. The three-point charge TIP3P model is used for water. Van der Waals parameters for the solute atoms are taken from ref. 22. This parameter set was originally developed for combined QM/MM calculations using the HF/3-21G theory.[22] These parameters are found to be reasonable for the present S_N2 reaction without further modification.

New configurations are generated by randomly selecting a molecule, translating it in all three Cartesian directions, and rotating around a randomly chosen axis. Solute moves are attempted in every 180 configurations, which include translation and

rotation moves. In addition, all bond distances and bond angles are varied in each solute move, except that the three heavy atoms are constrained to be collinear during the simulation. Volume moves are attempted in every 4625 Monte Carlo moves with a maximum change of 370 Å^3. All simulations were maintained with an acceptance rate of about 45% by using rages of ±0.15 Å and 15° for solvent moves, and ±0.10 Å and 5° for solute moves. Internal degrees of freedom are changed in the following ranges: ±0.005 Å for C—H bond distances, ±0.05 Å for C—Cl distances, and 5° for all bond angles. A total of six simulation windows have been executed, each consisting of 2×10^6 configurations of equilibration and 3–4 $\times 10^6$ configurations averaging.

FREE ENERGY CALCULATIONS

To begin, we define the reaction coordinate X^S for the S_N2 reaction in water as the difference between the energies of the reactant and product diabatic state in solution.[34] Thus,

$$X^S = E_1(\Psi_1) - E_2(\Psi_2) \tag{11}$$

where $E_1(\Psi_1)$ and $E_2(\Psi_2)$ are the diabatic potential energies for the reactant and product state, respectively. Because, $E_1(\Psi_1)$ and $E_2(\Psi_2)$ include solute–solvent interaction terms, the change in X^S accompanying the chemical process naturally reflects the collective motions of the solvent configuration. Indeed, eq. (11) has been utilized previously, especially in the work of Warshel and coworkers.[2, 34] Note that X^S is negative when the system is in the reactant state, because the solvent configurations strongly disfavor the product state leading to large positive values in $E_2(\Psi_2)$. X^S is positive when the system is in the product state because $E_1(\Psi_1)$ will be positive and $E_2(\Psi_2)$ will be negative. Therefor, X^S can be conveniently used to monitor the progress of the chemical reaction as the solvent reaction coordinate. Clearly, there is no single reactant structure that defines the transition state, rather, an ensemble of transition states will be obtained from the simulation, which may contain solute geometries closely resembling the reactant or product structure, but having solvent orientations suitable for the transition state.[31, 55, 56]

The potential of mean force as a function of X^S is determined by a coupled free-energy perturbation and umbrella-sampling method.[33, 51] Because the computational details have been described previously,[33] only a brief summary of the method is presented here. The computational procedure in fact includes two steps. The first is to use a reference potential E_{RP} to enforce the orientation polarization of the solvent system such that significant amount of time during the simulation is spent along the entire reaction path, particularly in the transition state region.

$$E_{RP}(\lambda) = (1 - \lambda)E_1(\Psi_1) + \lambda E_2(\Psi_2) \tag{12}$$

Thus, a series of free energy perturbation calculations are executed by moving the variable λ in eq. (12) from 0 to 1 to "drive" the reaction from the reactant state to the product state.[57] However, the free energy change obtained using the reference potential, E_{RP}, does not correspond to the adiabatic ground-state potential surface. The true ground state potential of mean force is derived from the second step of the computation via an umbrella-sampling procedure,[58] which projects the E_{RP} potential on to the adiabatic potential energy surface $E_g(X^S)$:

$$\Delta G(X^S) = \Delta G_{RP}(\lambda) - RT \ln\left\langle e^{-[E_g(X^S) - E_{RP}(\lambda)]/RT}\right\rangle_\lambda - RT \ln \rho\left[X^S(\lambda)\right] \tag{13}$$

where $\Delta G_{RP}(\lambda)$ is the free-energy change obtained in the first step using the reference potential, $E_g(X^S)$ is the adiabatic ground-state potential energy at $X^S(\lambda)$, and ρ $[X^S(\lambda)]$ is the normalized distribution of configuration that has a value of X^S during the simulation carried out using $E_{RP}(\lambda)$.

There are two important points that should be made clear here. First, the umbrella-sampling procedure is performed on the fly during the Monte Carlo free-energy perturbation calculation. It only represents a series of data transformation, with no extra simulations required. Second, the ground-state potential energy $E_g(X^S)$ specified in eq. (13) is not limited to the MOVB energy. Instead, any other quantum mechanical methods can be used to approximate $E_g(X^S)$, including HF, MP2, and DFT. In this regard, the MOVB method can be simply utilized to derive the necessary diabatic-state potential energy surfaces to define the solvent reaction coordinate [eq. (11)]. Consequently, a smaller basis set can be used to save computational costs in configuration sampling in *ab initio* simulations. A higher level of theory, or a larger basis set can be used to obtain the ground state energy $E_g(X^S)$. In the present study, we used the 6-31G(d) basis function to represent the MOVB diabatic state, and the HF/6-31G(d) theory to approximate the adiabatic potential energy function.

SOFTWARE USED

Ab initio VB and MOVB calculations are performed using the Xiamen99, and MOVB program, respectively.[33, 50, 59] The latter program has been implemented into a locally modified version of the GAMESS program,[60] which is also used in QM/MM Monte Carlo simulations using MCQUB.[61] HF and MP2 calculations are done using Gaussian 94.[62] All computations are carried out using Origin 200 workstations in our laboratory, or on an Origin 2000 system at the Center for Computational Research in Buffalo.

Results and Discussion

GAS-PHASE POTENTIAL ENERGY SURFACE

The potential energy reaction profiles from HF, VBSCF, and MOVB calculations for the S_N2 reaction of $Cl^- + CH_3Cl$ in the gas phase are shown in Figure 1. Geometries for the reaction profiles shown in Figure 1 and for energy calculations in Table I are obtained from Hartree–Fock optimizations using the corresponding basis set indicated for both MOVB and VBSCP calculations. Thus, strictly speaking, they do not represent the minimum reaction path of the corresponding level of theory. However, the deviation is not expected to be large. Here, the reaction coordinate X^R is the difference between the two C—Cl distances, i.e., $X^R = R_r(\text{C—Cl}') - R_p(\text{Cl—C})$, where C—Cl′ is the carbon and leaving group distance, and Cl—C is the nucleophile and carbon distance. The classic double well potential for the S_N2 reaction is well characterized by the MOVB method.[63, 64] At the HF/6-31G(d) level, the estimated binding energy for the ion–dipole complex and barrier height relative to the infinitely separated species are, respectively, −10.3 and 3.5 kcal/mol. The corresponding experimental values are −8.6 and 2.5 kcal/mol.[65–67] Glukhovtsev et al. estimated that the binding energy ΔE_b is −10.5 kcal/mol, and the overall barrier $\Delta E^{\neq}$ is 2.7 kcal/mol at the G2(+) level of theory.[47] There is only small variation using much larger basis sets with inclusion of electron correlation effects (Table I). Thus, the 6-31G(d) basis set and the HF theory is adequate for studying the chloride exchange reaction, as has been demonstrated previously.[42, 51] At the VBSCF(3)/6-31G(d) level, where the number in parentheses indicates the number of VB configurations used, the predicted ΔE_b and $\Delta E^{\neq}$ are −10.5 and 3.1 kcal/mol. To validate the use of a three-configuration approach for the S_N2 reaction, we have also performed *ab initio* VB calculations that include all six resonance configurations, i.e., VBSCF(6)/6-31G(d); the corresponding values are −10.0 and 3.9 kcal/mol. The change both in bind-

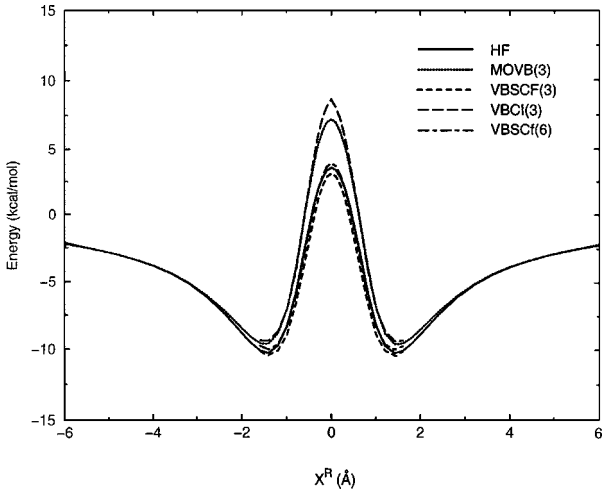


FIGURE 1. Gas-phase potential energy profile for the $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ reaction from calculations using HF/6-31G(d), MOVB(3)/6-31G(d), VBCI(3) 6-31G(d), VBSCF(3)/6-31G(d), and VBSCF(6)/6-31G(d). The reaction coordinate X^R is the minimum energy path computed at the HF/6-31G(d) level, and is given by $X^R = R_r(\text{C—Cl}') - R_p(\text{C—Cl})$, where $R_r(\text{C—Cl}')$ is the distance between C and the leaving Cl′ and $R_p(\text{C—Cl})$ is the distance between C and the nucleophile Cl. The primed Cl′ is to distinguish the leaving and nucleophilic Cl atoms. All geometries are optimized at the Hartree–Fock level with the constraint by X^R.

TABLE I.
Computed Binding Energy for the Ion–Dipole Complex and Activation Energy Relative to the Separate Molecules in the Gas Phase for the Reaction of $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ (kcal/mol).

Method[a]	ΔE_b	$\Delta E^{=}$
MOVB(3)/6-31G(d)	−9.7	7.1
MOVB(3)/cc-pVTZ	−9.6	8.1
VB-CI(3)/6-31G(d)	−9.4	8.6
VBSCF(3)/6-31G(d)	−10.5	3.1
VBSCF(6)/6-31G(d)	−10.0	3.9
HF/6-31G(d)	−10.3	3.5
HF/cc-pVTZ	−10.3	4.7
MP2/cc-pVTZ	−11.7	3.2
Experiment	−8.6[65]	2.5[66]

[a] Geometries are optimized at the Hartree–Fock level using the same basis set as that in energy calculations, i.e., 6-31G(d) or cc-pVTZ.

ing energy and activation energy is minimal with the inclusion of all six VB states. If individually optimized VB structures are used directly in CI calculations (VB-CI), the predicted binding energy (−9.4 kcal/mol) and barrier (8.6 kcal/mol) are noticeably higher than the corresponding VBSCF(3) values (Table I), indicating the importance of SCF orbital optimization in VB calculations. The present MOVB is analogous to the VB-CI in that variationally determined VB configurations are used in configuration interaction calculations without further optimizing the orbital coefficients. Not surprisingly, the quality of the MOVB results is similar to that of VB-CI with estimated ΔE_b and $\Delta E^{\neq}$ of −9.7 and 7.1 kcal/mol. Although the agreement between MOVB and VB-CI is good, orbital optimization is needed for quantitative prediction of the activation energy using MOVB.

The diabatic potential energies determined at the MOVB/6-31G(d) and VBSCF(3)/6-31G(d) level in the region between the reactant and product ion–dipole complex are plotted in Figures 2 and 3, in which the intersect between E_1 and E_2 at X^R = 0 Å corresponds to the location of the transition state.[45, 68, 69] The qualitative features of the potential energy profile are similar between the two computational methods. The energy at the E_1–E_2 crossing point is 18.3 kcal/mol higher than the transition state on the adiabatic potential energy surface, which may be compared to a value of 24.6 kcal/mol by VBSCF(3)/6-31G(d) calculations. The difference in this coupling energy partially reflects a smaller barrier of 13.6 kcal/mol relative to the ion–dipole

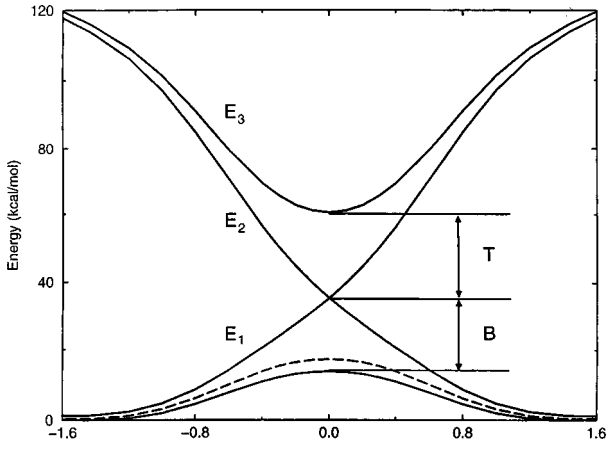


FIGURE 2. Computed gas-phase diabatic potential energy surface for the reactant (E_1), product (E_2), and zwitterionic (E_3) state. The adiabatic MOVB potential energy curve is shown in dashed line along with the HF/6-31G(d) minimum energy path in solid curve.

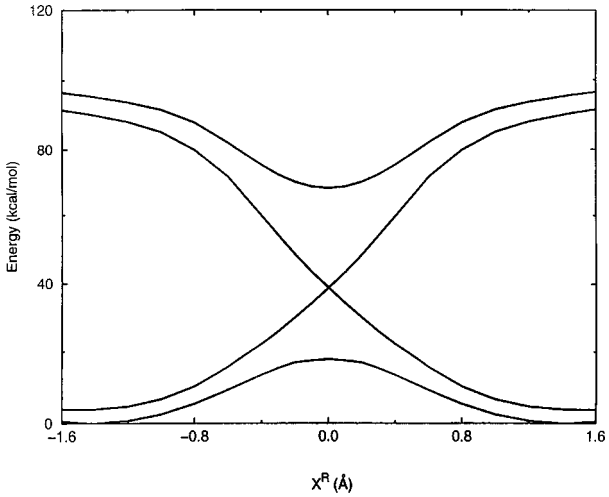


FIGURE 3. Gas-phase diabatic and adiabatic potential energy surface for the $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ reaction from VBSCF(3)/6-31G(d) calculations. See Figure 2 captions for definition.

complex predicted by the latter method than that of 16.8 kcal/mol at the MOVB level. The zwitterionic state Ψ_3 has higher energies than the reactant and product configuration along the entire reaction coordinate, with the asymptotic limit of $E_1 = E_2 = E_3$ in MOVB because of the large intrafragment charge polarization of CH_3Cl in the BLW representation. In VBSCF calculation, the covalent bonding character in the CH_3—Cl bond is well described by two Slater determinants, and thus, E_1 and E_2 do not tend to E_3 at large X^R. At the transition state geometry, E_3 is 25.3 and 29.3 kcal/mol higher in energy than E_1 and E_2, from MOVB and VBSCF computations. Away from the ground state region, the MOVB energies are significantly greater than the corresponding VB-SCF values. Although the difference is much smaller with the use of a larger basis set (cc-pVTZ), these high energy states should have little effects on the simulation results.

POTENTIAL OF MEAN FORCE

An important question in computer simulation of condensed phase reactions is the contribution of the cooperative solvent fluctuation that leads to barrier crossing. A full treatment of the problem requires an adequate sampling of the solute and solvent configurations along the reaction path.[30, 31] In the traditional approach, a solute geometrical parameter, for example, X^R in the present S_N2 reaction, is used to define the reaction coordinate in such a way that the solvent configurations follow the change of the solute reaction coordinate on going from the

reactant state to the transition state.[42] A potential problem in combined QM/MM calculation or in the use of a polarizable potential function, where the solute charge distribution is allowed to polarize, is that solute–solvent interactions tend to localize the solute charge distribution even at the transition state geometry, thereby generating a ⊖–○–○ electronic state in the ⊖–⊕–⊖ state geometry. As a result, the solvent configuration is polarized in favor of the ground state. Although we have not found this to be a serious problem from examination of polarization effects and charge distribution in previous studies of a range of chemical reactions in solution (excluding electron transfer reactions),[1, 33] it is still useful to be able to carry out combined QM/MM simulations that take into account the solvent reaction coordinate. This requires a formulation to specify the necessary charge-localized, diabatic states, which are developed in this work using the MOVB method. To this end, we have used the mapping potential, or reference potential defined in eq. (12), as is done in the EVB approach,[2, 34] to enhance the solvent configurational polarization along the reaction path.

The free energy change alone the mapping potential E_{RP} as a function of the coupling parameter, λ, is determined by the free-energy perturbation theory, and it is shown in Figure 4. The relatively high barrier in this free energy profile is reasonable because it was obtained using the energies of the diabatic state E_1 and E_2, which are 36 kcal/mol greater at the transition state than that in the ground state (compared to 16.8 kcal/mol adiabatic barrier relative to the ion–dipole complex). This is further coupled with solvation effects, favoring the more localized ground state. It is possible to utilize an alternative form of the reference potential, which includes the off-diagonal coupling energy:

$$\boldsymbol{E}_{RP}(\lambda) = (1-\lambda)\boldsymbol{E}_1 + \lambda \boldsymbol{E}_2 - 2\sqrt{\lambda(1-\lambda)}\,\frac{\boldsymbol{H}_{12} - (\boldsymbol{E}_1 + \boldsymbol{E}_2)\boldsymbol{S}_{12}/2}{1+\boldsymbol{S}_{12}} \quad (14)$$

where $\boldsymbol{H}_{12} = \langle \Psi_1 | \hat{\boldsymbol{H}} | \Psi_2 \rangle$, and $\boldsymbol{S}_{12}$ is the overlap matrix between Ψ_1 and Ψ_2. Although results obtained using eq. (14) are not reported here (though it was examined in our study), inclusion of the last term in the reference mapping potential will reduce the energy difference between E_{RP} and E_g,[51] which would accelerate convergence in the subsequent umbrella sampling calculation [eq. (13)].

The potential of mean force for the S_N2 reaction of $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ obtained with the HF/6-31G(d) potential energy, E_g [eq. (13)], is shown in Figure 5. Before we discuss the results, we emphasize that the free energy perturbation simulation described in the previous paragraph was carried out using the reference potential, E_{RP}, whose sole purpose is to force the collective solvent motions to spent a significant amount of time near the transition state region during the calculation. Thus, it is essential to recover the true ground state poten-

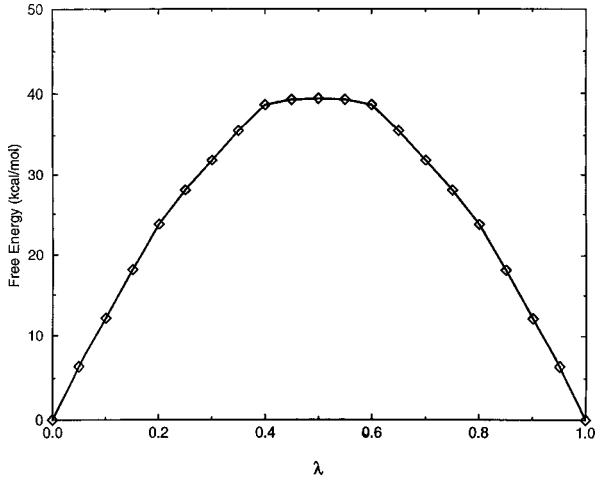


FIGURE 4. Free energy change as a function of the coupling parameter from the reactant ($\lambda = 0$) to the product state ($\lambda = 1$). The free-energy profile is obtained using the free-energy perturbation method on the reference potential energy surface E_{RP}.

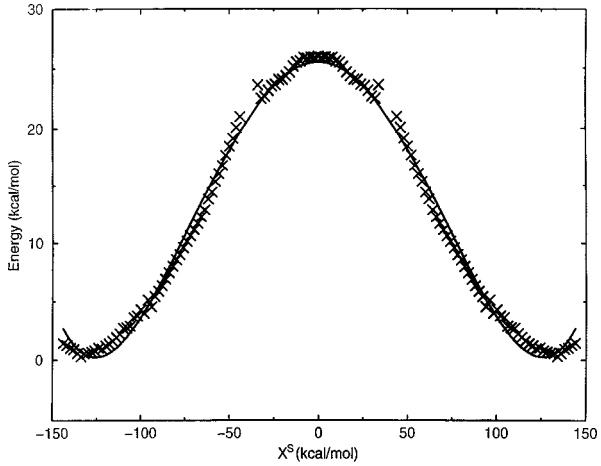


FIGURE 5. Potential of mean force for the nucleophilic substitution reaction of $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in aqueous solution as a function of the reaction coordinate X^S. The ground-state adiabatic potential energy is determined at the HF/6-31G(d) level, and the reaction coordinate is the difference between the potential energies of the reactant and product diabatic state, $X^S = E_1(\Psi_1) - E_2(\Psi_2)$.

tial surface using eq. (13).[33, 51] The computed activation free energy in Figure 5 is 26.0 ± 1.0 kcal/mol, which is in excellent agreement with the experimental value (26.6 kcal/mol) and with previous theoretical results. In particular, it is interesting to point out that the study by Chandrasekhar et al. of this system and later the work by Hwang et al. involved fitting potential energy functions, either in an empirical form or in the EVB formalism, to results obtained at the HF/6-31G(d) level.[42, 51] Thus, the results from these two calculations are directly comparable to the present study. The striking finding of the large solvent effects, which increase the barrier height by more than 20 kcal/mol (or 8 kcal/mol relative to the ion–dipole complex), is reproduced in the present *ab initio* MOVB-QM/MM calculation.[70] The origin of the solvent effects can be readily attributed to differential stabilization between the ground state, which is charge localized and more stabilized, and the transition state, which is more charge-dispersed and poorly solvated.[42, 51]

Although a direct comparison with results from the geometrical mapping procedure at the *ab initio* QM/MM level has not been made due to the high computational costs, the agreement of the present results with that of Chandrasekhar et al.,[42] which was obtained using a geometrical mapping technique, and with that of Hwang et al.,[51] which was derived using a similar energy mapping procedure, indicates that our MOVB method is capable of yielding reasonable results for chemical reactions in solution.[33] Furthermore, the general agreement of these simulation techniques demonstrate that the S_N2 reaction involving Cl^- and CH_3Cl in water can be adequately treated both by the solute reaction coordinate and by a combined solute–solvent reaction coordinate. These studies appear to indicate that for atom transfer reactions such as the present S_N2 reaction, it is not critical to choose a particular representation of the reaction coordinate, either using a simple geometrical mapping procedure or an energy-gap approach that includes collective solvent coordinates in characterizing the reaction path. To unambiguously resolve this question on this reaction, we plan to carry out a simulation that makes use of the traditional geometric mapping procedure as used by Chandrasekhar et al., but with the present *ab initio* QM/MM potential.[42]

Solvent effects on the individual diabatic state are shown in Figure 6, along with the adiabatic potential of mean force from the MOVB and HF ground-state surface. The potential energies of these states increase rapidly as they move away from the corresponding minimum energy state because the diabatic states are more stabilized by solvation in the ion–dipole complex region that in the transition state region. The product state has the wrong solvent orientation in the reactant state, and thus, has a very high energy. The opposite situation exists for the reactant state in the product region. Using variationally optimized VB structures in configuration interaction calculations as is done here (MOVB), the predicted MOVB activation energy is significantly higher than the HP value by about 10 kcal/mol, of which nearly 4 kcal/mol originate from the gas-phase difference (Table I) and 6 kcal/mol from solvent effects. Clearly, the electronic coupling term H_{12} strongly depends on the quality of the individual diabatic state wave functions, and it is important to optimize each VB state in the presence and restrictions of all other states in a full SCF calculation.

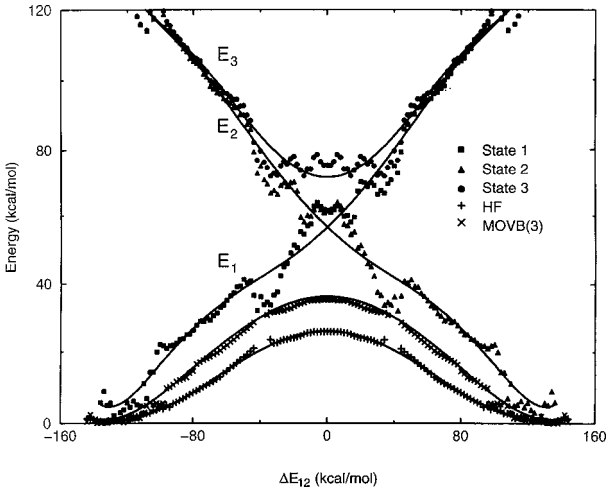


FIGURE 6. Aqueous diabatie-state potential energy surface for the reactant, product, and zwitterionic state, and potential of mean force obtained on the MOVB diabatic potential surface and the Hartree–Fock potential surface. All calculations are done using the 6-31G(d) basis set.

Summary

We have presented a mixed molecular orbital and valence-bond (MOVB) method in combined *ab initio* QM/MM simulations of chemical reactions in solution. The method is based on the construction of individual charge-localized, diabatic states using a block-localized wave function (BLW) approach that was developed for the study of chemical bonding and resonance energies. Then, a configuration interaction Hamiltonian is used to obtain the adiabatic potential energy surface for chemical reaction. The

method is tested on a prototypical S_N2 reaction involving $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in water using the 6-31G(d) basis set. To examine the performance of the MOVB potential surface, modern *ab initio* VB calculations were performed, which include a three-configuration active space, and the full six-configuration space. The MOVB gas phase results are found to be in reasonable agreement with Hartree–Fock, MP2, and *ab initio* VB calculations, although the activation harrier is about 4 kcal/mol higher in MOVB. The discrepancy was traced to originate from the use of variationally optimized individual VB states in configuration calculations. On the other hand, the ground-state adiabatic potential energy surface can be approximated by other quantum chemical models, including Hartree–Fock, MP2, and DFT methods, to determine the potential of mean force for chemical reaction in solution. At the HF/6-31G(d) level, the calculated activation free energy of 26 kcal/mol is in good agreement with experiment and with previous computational results. Importantly, the availability of diabatic potential energy surfaces allows the solvent reaction coordinate to be included in the form of an energy gap between the reactant and product state in *ab initio* QM/MM simulations. Although EVB has been used as a mapping potential in *ab initio* QM/MM free energy calculations,[71] the MOVB approach goes beyond the traditional technique, and represents an important step towards a more systematic approach for molecular simulations at the *ab initio* level.

References

1. Gao, J. Acc Chem Res 1996, 29, 298.
2. Åqvist, J.; Warshel, A. Chem Rev (Washington) 1993, 93, 2523.
3. Amara, P.; Field, M. J. In Computational Molecular Biology; Leszczynski, J., Ed.; Elsevier: Amsterdam, 1999, p. 1.
4. Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. Biochemistry 1991, 30, 5826.
5. Gao, J.; Thompson, M. A., Eds.; Combined Quantum Mechanical and Molecular Mechanical Methods; American Chemical Society: Washington, DC, 1998.
6. Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. J Am Chem Soc 1998, 120, 3858.
7. Hartsough, D. S.; Merz, K. M., Jr. J Phys Chem 1995, 99, 384.
8. Merz, K. M., Jr.; Banci, L. J Phys Chem 1996, 100, 17414.
9. Assfeld, X.; Ferre, N.; Rivail, J.-L. ACS Symp Ser 1998, 712, 234.
10. Liu, H.; Mueller-Plathe, F.; van Gunsteren, W. F. J Mol Biol 1996, 261, 454.
11. Chatfield, D. C.; Brooks, B. R. J Am Chem Soc 1995, 117, 5561.
12. Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G. J Am Chem Soc 1999, 121, 2253.
13. Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. Proc Natl Acad Sci USA 2000, 97, 2017.
14. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227.
15. Gao, J. In Reviews in Computational Chemistry; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH: New York, 1995, p. 119, vol. 7.
16. Singh, U. C.; Kollman, P. A. J Comput Chem 1986, 7, 718.
17. Weiner, S. J.; Singh, U. C.; Kollman, P. A. J Am Chem Soc 1985, 107, 2219.
18. Tapia, O.; Lluch, J. M.; Cardenas, R.; Andres, J. J Am Chem Soc 1989, 111, 829.
19. Bash, P. A.; Field, M. J.; Karplus, M. J Am Chem Soc 1987, 109, 8092.
20. Field, M. J.; Bash, P. A.; Karplus, M. J Comput Chem 1990, 11, 700.
21. Gao, J.; Xia, X. Science 1992, 258, 631.
22. Freindorf, M.; Gao, J. J Comput Chem 1996, 17, 386.
23. Gao, J.; Freindorf, M. J Phys Chem A 1997, 101, 3182.
24. Alhambra, C.; Byun, K.; Gao, J. ACS Symp Ser 1998, 712, 35.
25. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. J Am Chem Soc 1984, 106, 3049.
26. Chandrasekhar, J.; Jorgensen, W. L. J Am Chem Soc 1985, 107, 2974.
27. Van der Zwan, G.; Hynes, J. T. J Chem Phys 1983, 78, 4174.
28. Gertner, B. J.; Bergsma, J. P.; Wilson, K. R.; Lee, S.; Hynes, J. T. J Chem Phys 1987, 86, 1377.
29. Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. J Chem Phys 1987, 86, 1356.
30. Muller, R. P.; Warshel, A. J Phys Chem 1995, 99, 17516.
31. Geissler, P. L.; Dellago, C.; Chandler, D. J Phys Chem B 1999, 103, 3706.
32. Truhlar, D. G.; Garrett, B. C. J Phys Chem B 2000, 104, 1069.
33. Mo, Y.; Gao, J. J Phys Chem A 2000, 104, 3012.
34. Warshel, A. Computer Modeling of Chemical Reactions in Enzymes and Solutions; Wiley: New York, 1991.
35. Mo, Y.; Peyerimhoff, S. D. J Chem Phys 1998, 109, 1687.
36. Mo, Y.; Zhang, Y.; Gao, J. J Am Chem Soc 1999, 121, 5737.
37. Mo, Y.; Gao, J.; Peyerimhoff, S. D. J Chem Phys 2000, 112, 5530.
38. Gerratt, J.; Cooper, D. L.; Karadakov, P. B.; Raimondi, M. Chem Soc Rev 1997, 26, 87.
39. Cooper, D. L.; Gerratt, J.; Raimondi, M. Chem Rev 1991, 91, 929.
40. McWeeny, R. Pure Appl Chem 1989, 61, 2087.
41. Cooper, D. L.; Gerratt, J.; Raimondi, M. Adv Chem Phys 1987, 69, 319.
42. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. J Am Chem Soc 1985, 107, 154.
43. Vande Linde, S. R.; Hase, W. L. J Chem Phys 1990, 93, 7962.
44. Zhao, X. G.; Tucker, S. C.; Truhlar, D. G. J Am Chem Soc 1991, 113, 826.
45. Shaik, S. S.; Schlegel, H. B.; Wolfe, S. Theoretical Aspects of Physical Organic Chemistry: The SN2 Mechanism; Wiley: New York, 1992.
46. Shi, Z.; Boyd, R. J. J Am Chem Soc 1991, 113, 2434.
47. Glukhovtsev, M. N.; Pross, B. A.; Radom, L. J Am Chem Soc 1995, 117, 2024.

48. Glukhovtsev, M. N.; Pross, B. A.; Radom, L. J Am Chem Soc 1995, 117, 9012.
49. Gianinetti, E.; Raimondi, M.; Tornaghi, E. Int J Quantum Chem 1996, 60, 157.
50. Mo, Y.; Gao, J. BLW; Version 0.2; SUNY: Buffalo, NY, 1999.
51. Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. J Am Chem Soc 1989, 110, 5297.
52. Hiberty, P. C.; Flament, J. P.; Noizet, E. Chem Phys Lett 1992, 189, 259.
53. Hiberty, P. C.; Humbel, S.; Archirel, P. J Phys Chem 1994, 98, 11697.
54. Shurki, A.; Hiberty, P. C.; Shaik, S. J Am Chem Soc 1999, 121, 822.
55. Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. J Chem Phys 1998, 108, 1964.
56. Dellago, C.; Bolhuis, P. G.; Chandler, D. J Chem Phys 1999, 110, 6617.
57. Zwanzig, R. J Chem Phys 1961, 34, 1931.
58. Valleau, J. P.; Torrie, G. M. In Modern Theoretical Chemistry; Berne, B. J., Ed.; Plenum: New York, 1977, p. 169, vol. 5.
59. Wu, W.; Mo, Y.; Zhang, Q. Xiamen University: Xiamen, 1998.
60. 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. S. GAMESS; Version 11; Iowa State University: Anes, Iowa, 1993.
61. Gao, J. MCQUB; Version 3.0; Buffalo, NY, 1998.
62. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Oritz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Grodon, M.; Gonzalez, C.; Pople, J. A. GAUSSIAN 94; E2 ed.; Gaussian, Inc.: Pittsburgh, PA, 1994.
63. Olmstead, W. N.; Brauman, J. I. J Am Chem Soc 1977, 99, 4219.
64. Pellerite, M. J.; Brauman, J. I. J Am Chem Soc 1980, 102, 5993.
65. Dougherty, R. C.; Roberts, J. D. Org Mass Spectrom 1974, 8, 81.
66. Wladkowski, B. D.; Brauman, J. I. J Phys Chem 1993, 97, 13158.
67. Barlow, S. E.; Van Doren, J. M.; Bierbaum, V. M. J Am Chem Soc 1988, 110, 7240.
68. Shaik, S. S.; Pross, A. J Am Chem Soc 1982, 104, 2708.
69. Shaik, S. S.; Shurki, A. Angew Chem Int Ed 1999, 38, 587.
70. Albery, W. J.; Kreevoy, M. M. Adv Phys Org Chem 1978, 16, 87.
71. Bentzien, J.; Muller, R. P.; Florian, J.; Warshel, A. J Phys Chem B 1998, 102, 2293.