COMPUTATIONAL STUDY OF CHEMICAL REACTIONS

By

BOJANA GINOVSKA

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

WASHINGTON STATE UNIVERSITY
School of Electrical Engineering and Computer Science

DECEMBER 2007

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of BOJANA GINOVSKA find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

ACKNOWLEDGMENT

I would like to acknowledge the help and support I was given by my family, the faculty in the Computer Science Department at Washington State University Tri-Cities, and the staff from the Molecular Interactions and Transformations group in the Chemical and Materials Sciences Division at Pacific Northwest National Laboratory (PNNL).

COMPUTATIONAL STUDY OF CHEMICAL REACTIONS

Abstract

by Bojana Ginovska, M.S.
Washington State University
December 2007

Chair: Donald J. Lynch

Computers can be used to obtain detailed pictures of chemical reactions. A brief overview of a number of computational approaches which can be used to this end will be given. Computational advantages and disadvantages of selected algorithms will be discussed, as well as their applicability and their accuracy. Application of some of the methods will be presented in a study of the reaction of hydrogen peroxide and hydroxyl radical (i.e $H_2O_2 + OH \rightarrow H_2O + O_2H$). The reaction was studied in gas as well as in condensed phase. In the gas phase two distinct reaction pathways were identified, and the rate was calculated using variational transition state theory for a temperature range of 250-500 K. The calculations explain how the unusual temperature dependence observed for temperatures above 900 K is due to the reaction occurring on the low-lying excited state surface, rather than on the ground state surface. In solution, the reaction is studied using the QM/MM methodology. The free energy barrier was found to be higher than the barrier in the gas phase, which is in accord with the experimental findings of the rate being slower in solution. This work demonstrates the power of computational studies to explain and predict characteristics of chemical reactions, as well as interpret experimental observations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Chemical Reactions by Computations

The change of chemical substances under specific conditions of pressure, volume, and temperature constitutes a chemical reaction. When a chemical reaction occurs, one is generally concerned with the reaction mechanism, the path that it follows, and the rate at which it proceeds. The determination of the mechanisms and rates of chemical reactions is the subject of chemical kinetics. In some instances, the way a chemical system behaves over time may be of interest, rather than some property at a specific instance of time, or how the system changes in the presence of a solvent. All those aspects of chemistry can be studied simply with computers, using appropriate theoretical models. When studying chemical reactions computationally, a number of issues must be addressed. What is the shape of the potential energy surface (PES) that the molecules follow in the course of their motion? What is the best theoretical model that can describe the PES? How high is the barrier to reaction? What is the rate of the reaction? A number of computational concerns are also relevant. Is a given theory captured in the computer code used? How much computer time will the computations take? How much memory will be required? What is the best computer for the calculation?

Theoretically, answers to the first three questions can be found using quantum theories. In practice, quantum theories can only be used to study systems of limited size, and other approaches must be taken for larger problems. In 1929, the famous British theoretician Dirac stated the following:

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble." [1] Although nearly eighty years later we are still unable to solve the complex equations, theories and computational tools exist that allow us to find quantitative solutions for increasingly complex and realistic models of chemical systems.

The questions about the proper choice of computer program, platform, and computational resources (i.e. CPU time and memory) are just as important as the questions about theory. There are many programs that are available for chemical computation, but choosing the most appropriate one is not an easy task. The program of choice must work on an appropriate platform, and it should also do so efficiently. This is especially important when using large resources, parallel computing models, and architectures. Although we can use petascale data storage and gigahertz speed processors, the challenge to use them properly still remains. The problem of efficient use of resources is crucial when developing both algorithms and theories for computational chemistry. It is also a very important question when carrying out computations. As a result, the advances in computer science and computational chemistry have become inseparable, and the problems of computation must be approached from both aspects. It is clear today, that computational solutions to chemical problems are as important to the advancement of science as theory and experiment are. Chemical systems of various sizes can be simulated computationally in order to give insight into properties that may be impossible or prohibitively expensive to obtain experimentally.

In what follows, a brief overview is given of a number of computational approaches that can be used to study chemical reactions. Some of the advantages and disadvantages of the algorithms, as well as their applicability and accuracy are also discussed. Finally, an application of some of the methods is presented in a study of the reaction of hydrogen peroxide (H_2O_2) and hydroxyl radical (OH).

2 THEORY

There are several approaches to studying chemical reactions. Each study starts by finding the potential energy function, which describes the energetic and structural changes in the molecules that occur during the reaction. Many methods have been developed to estimate this function. Quantum Mechanical (QM) methods are based either on molecular orbital theory or Density Functional Theory (DFT). The high computational costs of these methods limit their usability to very small systems. For the purpose of studying larger systems, classical methods, such as Molecular Mechanics (MM) have been developed. These methods are more approximate and often less accurate, as well as incapable of describing chemical bond making and bond breaking. Alternately, hybrid Quantum Mechanical / Molecular Mechanical (QM/MM) methods can be used. In these methods a subset of the system is identified which is deemed important in the reaction and is treated quantum mechanically. The rest of the system, which is considered to be less involved in the reaction under study, is treated with molecular mechanics.

Once a potential energy surface has been determined, different properties of the system can be extracted. Commonly, a property of interest is the rate of the reaction, which can be calculated using Transition State Theory (TST).

In the following sections, a brief overview is given of the computational chemistry methods mentioned above.

2.1 Molecular Mechanics

Molecular mechanics is the computationally least expensive of the methods. In molecular mechanics, the system is represented classically and is described by a set of bonding and non-bonding interactions. Visually, one may imagine a classical molecule as a collection of spheres representing the atoms with chemical bonds described by springs (Fig. 1).

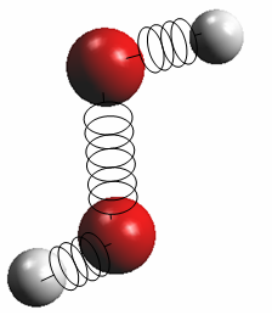

Figure 1. Classical representation of H_2O_2 molecule

The bonding interactions account for the stretching, bending, and torsion within the molecule, while the non-bonding interactions describe the electrostatic and repulsive interactions between the atoms of different molecules.

In general, all MM models are based on force-field parameters (atomic distances and spring force constants) which describe these interactions. These parameters have been determined by fitting to experimental or high-level QM data, and are usually found in libraries of parameters distributed with MM programs. Although quite appealing due to their computational efficiency, the MM methods lack the capability to describe many important features of a reaction. An example is the failure to represent breaking and making of chemical bonds, which are intrinsically quantum mechanical phenomena. In spite of their shortcomings, the MM methods can be very useful. For instance, they are commonly used as a molecular description of solvents in QM/MM or molecular dynamics (MD) simulations.

In molecular mechanics, the potential energy function is very simple and computationally fast. It is a sum of the interacting terms, derived from the harmonic oscillator approximation, and non-bonded interactions:

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{non-bonding} \tag{1}$$

where:

$$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2 \tag{2}$$

$$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \tag{3}$$

$$E_{tosions} = \sum_{torsions} \frac{V_n}{2} \left[1 + \cos(n\phi - \gamma)\right] \tag{4}$$

$$E_{non-bonding} = \sum_{i,j} \left[\frac{q_i q_j}{\varepsilon R_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \tag{5}$$

In the expressions above, the K_r, K_θ and V_n are force constants, controlling the fluctuations of the 'springs' (stretches, bends, and torsions respectively) from their equilibrium positions.

In the non-bonding expression (5), the first element is the attractive electrostatic interaction between the pairs of atoms belonging to different molecules, where the q_i and q_j values are the partial charges of the atoms i and j, R_{ij} is the distance between the atoms, and ε is the dielectric constant for the medium. The second and the third term represent the van-der-Waals interactions through a Lennard-Jones potential function. The values for A and B are determined as constants for each atom type and the r_{ij} are the distances between the atoms. These two terms define a balance between attractive and repulsive forces, allowing for the atoms to repel each other when close enough, and attract when at an appropriate distance. It is the calculation of these non-bonding interactions that is most expensive in an MM calculation, giving the algorithm a quadratic scaling denoted $O(N^2)$, where N is the number of MM atoms in the system. The simplicity of the force field makes the calculation of the energy terms above very fast. However, when attempting to improve performance by parallelizing the calculation, the algorithm faces a network communication bottleneck. This issue will be revisited in the discussion on algorithm performance in Chapter 3.

2.2 Quantum Mechanical Methods

Classical mechanics describes molecular systems as atoms bound together by springs. In quantum mechanics a physical system is described by a function of all the particles in the system, electrons and nuclei, called a wavefunction $\Psi(x, y, z, t, ..)$. The wavefunction reveals the properties of the system when an appropriate operator (i.e. an observable of a property) is applied to it. The Hamiltonian operator (H) is an observable of the energy of the system. When H is applied to the wavefunction, it gives rise to the Schrodinger equation:

$$H\Psi = E\Psi \tag{6}$$

The solutions of this equation are the eigenvalues E, which give the energy of the system. The molecular Hamiltonian describes the interactions between all the particles in the system. In order to simplify the system, the Born-Oppenheimer approximation may be invoked. This approximation states that since the nuclei are heavier and much slower than the electrons, one can separate the motion of the nuclei and the electrons. The motion of the electrons is then determined only by the electronic interactions. In practice this means that the molecular Hamiltonian can be replaced with the electronic Hamiltonian, giving the electronic energy for fixed nuclear positions. Because in QM the electrons have no fixed position, they are represented by an electron density (Fig. 2), which is the probability of finding the electron at a point in the space. This definition of the electronic structure allows for making and breaking of covalent bonds during the evolution of the system.

Figure 2. Electron density surface of H_2O_2

The electronic energies for all possible geometries of the system make up the potential energy surface (PES).

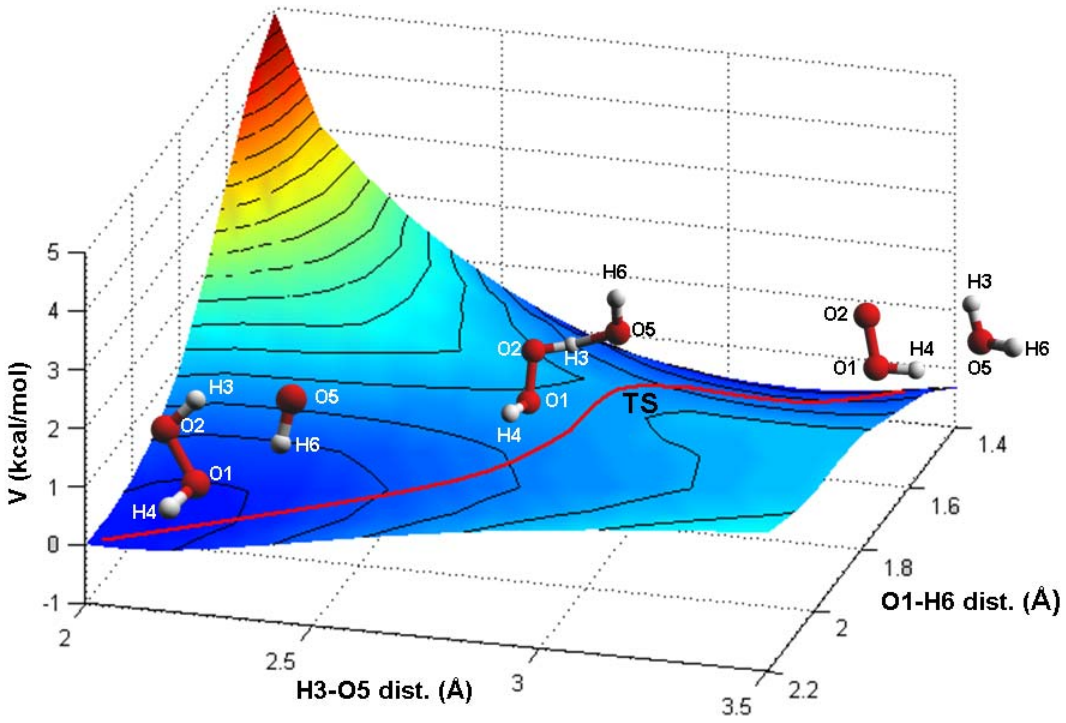

Figure 3. PES of H2O2+OH reaction in gas phase

It is in the light of the PES that we can distinguish structures that are either minima (stable chemical structures), or saddle points (unstable chemical structures), and the reaction paths connecting them. The minima on the PES correspond to reactants or products in the reaction, and the first-order saddle points correspond to intermediate structures called transition states (TS), over which the reactions proceed from reactants to products. The path on the PES that

connects the minima via TS is called the Intrinsic Reaction Coordinate (IRC).[2] Determining a pathway allows us to calculate the rate of the reaction.

Due to its complex form, the evaluation of the QM potential function is computationally very demanding and can be between $O(N^3)$ and $O(N^7)$, depending on the level of theory used. In this case, N does not refer to the number of atoms in the system, but to the number of basis functions used to describe the wavefunction. A more detailed explanation of this dependence will be provided later on (in the section 3.2, *Algorithm Performance in Computational Chemistry* and Appendix A).

2.3 Quantum Mechanics/Molecular Mechanics (QM/MM)

The hybrid QM/MM methods allow for a reaction to be studied with different levels of theory for different regions. This method is very appropriate when reactions are studied in solutions, since the solvent molecules can be represented by MM and the solute by QM (Fig. 4).

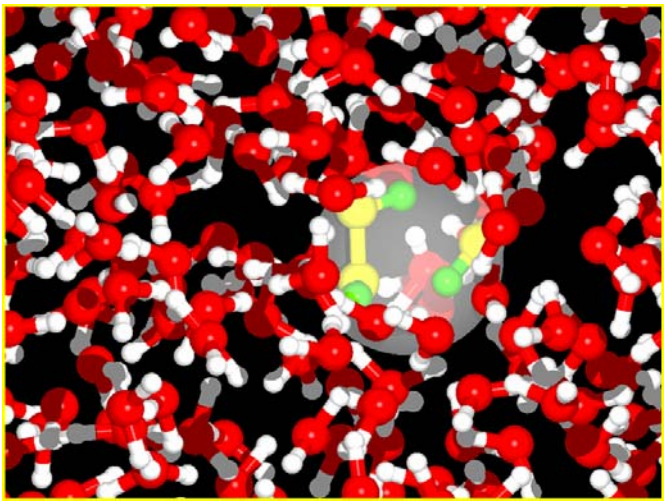

Figure 4. QM (H_2O_2+OH) and MM regions $(H_2O)_n$ in a QM/MM model

It is also commonly used in a study of enzymatic reactions where the species studied are large

biomolecules. In this case, apart from the solvent, most of the biomolecule is represented at the MM level of theory, and only the active site is treated quantum mechanically.

In QM/MM methodology, the potential function is a hybrid function, expressed in terms of an effective Hamiltonian $H_{Eff,}$ which has different terms for the different regions:

$$H_{Eff} = H_{QM} + H_{MM} + H_{QM/MM} \tag{7}$$

In this case the Schrodinger equation becomes:

$$H_{Eff}\Psi = E\Psi \tag{8}$$

Most of the intermolecular interactions in QM/MM belong to one of the two models described above (MM for the solvent-solvent interaction and QM for the solute-solute interaction). There is a need to properly account for the interactions between the QM and the MM atoms, since the quality of the QM/MM calculation depends greatly on the way these interactions are handled.

At finite temperature the solvent adopts many different configurations very close in energy and that capture the fluctuations of the solvent in accord with a Boltzmann distribution. Properties of the system under study must be reported as a mean value resulting from a statistical average over solvent configurations. The averaging can be done by a Monte Carlo method that determines the solvent configurations (from which the energy is averaged) or by molecular dynamics (Newton's equations of motion where the energy is averaged over time). According to the ergodic hypothesis, [3] the two methods of sampling give equivalent ensemble averages.

When studying reactions in solutions, the free energy change along the reaction coordinate is usually the property of interest, as it reflects the effect of the solvent environment on the barrier to reaction and can be used to calculate the rates of reactions in solution. A procedure for calculating the free energy barier in solution is explained in section *4.2.4 Free energy barriers in solution*.

2.4 Rates of Reaction: Transition State Theory (TST)

A reaction is most simply defined as a transition from one state of the system to another. The transition may be conformational, where only the rearrangement of the atoms in the molecule changes, or there may be changes from one chemical species to another. The rate at which the transition occurs is measured as a change in the concentration of the species in the system over time. Mathematically, this is described by a kinetic equation,

$$\frac{d[A]}{dt} = k_{B\to A}[B] - k_{A\to B}[A] \tag{9}$$

where [A] is the concentration of the species A and [B] is the concentration of the species B. The rate is a macroscopic property, so it is meaningful only if we look at a system of a large number of molecules (an ensemble rather then a single molecule).

Conceptually, the simplest way to calculate the macroscopic rate of a reaction is to carry out a dynamics simulation, where trajectories* are started with randomly assigned initial velocity, and to observe the number of trajectories that cross the barrier from reactants to products. This method however is very inefficient, since it requires unfeasible simulation times in order to obtain statistically meaningful results.

As an alternative, the rate can be expressed in terms of TST. Within the framework of this theory, we assume that the reaction arrives at the product state over an activated complex, called a transition state. It is further assumed that the transition state is in equilibrium with the reactants and that none of trajectories return to the reactant region once they reach the transition state (Fig. 5).

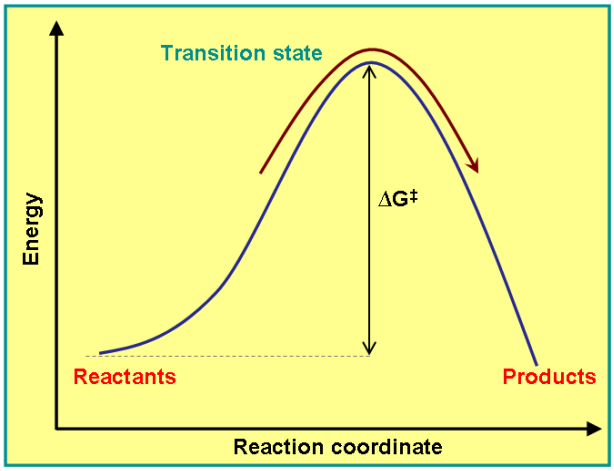

Figure 5. Transition state barrier

TST assumes that the reaction moves from reactant to products over a bottleneck at the PES, and relates the rate of the reaction to the height of the energetic barrier at the bottleneck with the following equation:

* Trajectory is the series of points in space that results from numerically integrating Newton's equation of motion ($F = ma$) where the force F acting on the particles is a result of the potential energy surface.

$$k(T) = \frac{K_b T}{h} \frac{Q^{GT}[T]}{\Phi^R(T)} e^{\left[-\frac{V_{mep}}{K_b T}\right]} \tag{10}$$

In equation (10), Q^{GT} is the partition function* of the transition state, and Φ^R of the reactants. V_{mep} is the potential barrier to reaction, T is temperature, K_b is the Boltzmann's constant and h is Planck's constant. Within the harmonic oscillator approximation, the partition function can be calculated from the structure of a single molecule, providing the bridge between the microscopic and macroscopic (thermodynamic) properties of the system. Because of this connection, the rate can be derived from the kinetic equation, and one obtains the form given in equation (9). The details of the rate expression derivation can be found in reference (4).

In the simplest approximation, TST defines the bottleneck of the reaction as a dividing surface placed at the transition state. Because TST usually overestimates the rate, improvements of this theory, such as the Canonical Variational Transition State Theory (CVT), allow for the dividing surface to be variationally placed along the reaction coordinate in order to minimize the rate:

$$k^{CVT}(T,s) = \min_s k^{TST}(T,s) \tag{11}$$

Additional improvements, such as quantum corrections for nuclear tunneling, can also be included in the rate expression.

Following the formulation (10), to estimate the rate of the reaction, one only needs a minimum energy path and a vibrational analysis of the reactants and the transition state.

* The partition function gives the number of energy states accessible to the system at a given temperature ($Q = \sum_i e^{-\varepsilon_i / K_b T}$) and links the macroscopic properties of a system in terms of averages of microscopic quantum states.

Therefore, TST gives invaluable savings in computing time compared to statistically averaging trajectories that explore a much wider area of the PES than just the IRC.

Once the rate is calculated, it can easily be compared to experiment. For most reactions, the measured experimental rates obey exponential relation as a function of the temperature, known as the Arrhenius equation:

$$k = Ae^{-\frac{\Delta Ea}{RT}} \tag{12}$$

From the similar form of the equations (10) and (12), it is easy to see how the Arrhenius activation energy *Ea* and the pre-exponential factor *A* can be related to the TST terms. Some reactions however, show non-Arrhenius temperature dependence, which is usually attributed to a complex mechanism of the reaction. The reaction of hydrogen peroxide and hydroxyl radical that is described later, is an example of a reaction with non-Arrhenius behavior.

3 COMPUTATIONAL TOOLS

The application of the theories described above would be impossible without the proper computational tools. The scale of the problems that computational chemistry addresses usually demands substantial computational resources, and makes extensive use of high performance computing (HPC). The HPC utilization requires availability of the proper hardware, system tools and libraries, as well as applications designed to work with them.

In this chapter, aspects of high performance computing that are relevant to chemical computations are discussed, followed by an analysis of some of the main algorithmic bottlenecks in the calculations, and a summary of the hardware and software used.

3.1 High Performance Computing (HPC)

HPC really means using parallel programs on multiprocessor machines. Computational chemistry programs utilize both distributed memory architectures with message passing interface (MPI) and shared-memory architectures with direct memory access. Furthermore, tools have been developed for implementing logically shared-memory on architectures with physically distributed memory.

In MPI type parallelization, the problem is divided into tasks, each computed simultaneously by a different processor. The result is then collected and assembled by one of the processors. In this programming model, the inter-process communication is two-sided and employs 'send' / 'receive' messages via a message passing interface. The efficiency of this type of communication is dependent on the network. If the program requires frequent communication between the processors or transfer of large amounts of data, the network can become a bottleneck of the calculation. Therefore, this type of parallelization is most appropriate for programs that are either embarrassingly parallel, that is there is no dependency between the tasks of different processors, or the communication time between the tasks is negligible compared to the total execution time.

To alleviate the network bottleneck in distributed memory hardware, a shared-memory model can be simulated in software. In shared-memory architectures, each process can directly access the entire memory of the system, so there is no communication over the network. Global Arrays (GA)[5] are tools that provide shared-memory interface for both distributed-memory and shared-memory systems. This programming model has been developed at Pacific Northwest

National Laboratory for massively-parallel distributed-memory and scalable shared-memory systems used in HPC.

When global arrays are used on a distributed-memory machine, the array is physically distributed across the memory of all the nodes used by the program. Logically the memory appears shared to each processor, and the call to any location in the array is programmed as a local memory call. If the memory needed is physically located at a different node, it can be copied locally to the calling node, instead of accessing the remote data. In this case, all the subsequent updates are written to the local copy, and when the update is completed, the new data is copied to the appropriate location in the shared object. Fundamentally, the communications are one-sided without any explicit cooperation by the other processes, through 'put' / 'get' instructions.

Apart from the ease of programming it provides for the application programmer, the GA toolkit has been built to be compatible with MPI, offering flexibility to use the different communication models in a single program. Applications built with this parallelization model are designed to run very efficiently on supercomputers and perform best when using hundreds of processors. The GA tools are used in the NWCHEM[6] program.

3.2 Algorithm Performance in Computational Chemistry

Conducting a successful computational study of a chemical reaction requires finding the appropriate tradeoff between the accuracy of the methods applied, and their algorithmic performance. The more accurate QM methods are computationally very expensive and can only be applied to small systems of up to a few hundreds atoms. The less demanding methods, such

as those relying on molecular mechanics, can be applied to systems of millions of atoms. The feasibility of the calculation can also be seen from the aspect of MD simulation times. With the available resources, and the level of theory used, how long can we run a simulation of a system to study its dynamical behavior? This is a very important question, for example, when trying to detect rare events. Whereas the QM methods can only be used to produce trajectories of a few picoseconds, the MM methods can be run on the microsecond time scale.

The significant computing power that is currently available, allows for more rigorous testing of chemical theories, and better prediction of experimental results. This brings the dual challenge of developing more sophisticated theories and more efficient algorithms to implement the theories.

For various theories, different computational bottlenecks can be identified. They can be the result of a poor scaling in terms of CPU time, memory requirements, disk I/O, and in parallel computations, of the communication over the network.

In MM calculations, the bottleneck arises when calculating the non-bonded interactions, i.e. the van der Walls and the electrostatic terms. The calculation scales $O(N^2)$, with N being the number of atoms. Although quadratic scaling may not sound too bad, it can be a problem when the system has millions of atoms. One way of dealing with this bottleneck is by introducing a cut-off distance from the site of interest, beyond which the intermolecular interactions are considered insignificant and are not calculated. This approximation is acceptable, because the energetic contributions are proportional to the distance between the atoms and fall rapidly as a function of that distance. Therefore, the contributions of the interaction between atoms that are beyond 10 - 15 angstroms apart can be considered not worth the computational cost. In parallel implementations, the poor scaling is due to the interprocess communication, which is much

slower and too frequent compared to the time it takes to compute the energy contributions. These two tasks become comparable only when the system is very large, and parallelization improves the performance.

In QM energy calculations, the algorithms are more complicated, and so are the bottlenecks. The size of the calculation arises from the complexity of the functions that describe the wavefunction. These functions are called basis functions, and the set of them used for a specific system is called a basis set. The size of the basis set is determined by the number of the basis functions (N). Because a system of the same number of atoms can be represented by basis sets of different sizes, the performance of QM calculations is expressed in terms of the number of basis functions, not the number of atoms.

The energy in a QM system is evaluated by an iterative algorithm called Self-Consistent Field (SCF) method. Without going into the details of the algorithm, which are available in Appendix A, its performance will be discussed, focusing on two different parts that define the bottleneck in the calculation: the evaluation of the two-electron integrals and the matrix operations.

The wavefunction of a system is expressed in terms of the one-electron and two-electron integrals over the basis functions. The two-electron integrals are computationally expensive, so there are different ways of handling them. In traditional calculations, called conventional SCF, these integrals are calculated once, and are stored on a disk. Every time the values are needed in the calculation, they must be read from the disk, making the disk I/O a bottleneck. One variation to this algorithm stores the integrals in main memory (in-core SCF), thus eliminating the I/O bottleneck. This makes the algorithm very fast, but requires very large amount of memory. Another variation, direct SCF, recalculates the integrals every time they are needed. Although, it

is an improvement over the previous methods in terms of memory usage and disk I/O, direct SCF may have significant computational overhead. Nevertheless, when the system under study is large, this method often gives the best performance. The table below summarizes the performance of these algorithms[7]:

Table 1. Scaling of SCF algorithms

SCF Method	CPU	Memory	Disk I/O
Conventional	N^4	N^2	N^4
In-core	N^4	N^4	----
Direct	N^4	N^2	----

N= number of basis functions.

Another important aspect in the SCF calculation is its reliance on efficient matrix multiplication and diagonalization. At each step of the iterative procedure, the so-called Fock matrix is constructed (by multiplying the two-electron integrals with a density matrix) and diagonalized. The Fock matrix construction requires $O(N^4)$ operations, and the diagonalization is $O(N^3)$. In a single processor calculation, the two-electron integrals and the Fock matrix construction are the performance determining step.

When the SCF algorithm is parallelized, an improvement is gained in the two-electron integral evaluation, since they are independent from each other, and can be easily distributed among the processors. The diagonalization of the Fock matrix, however, can not be efficiently parallelized, and it dominates the calculation in parallel implementations of SCF.

When calculating stationary structures on the PES, we employ algorithms that minimize the energy in order to find first order stationary points (minima), or search the surface to find

higher order stationary points (transition states). The optimization algorithms are iterative procedures, which require not only a substantial number of energy calculations, but also calculations of the first derivatives and possibly the second derivatives of the energy (the gradient and the Hessian). Efficient calculation of gradients greatly influences the performance of the algorithms. Analytical gradients are most efficient, albeit not always available. The second best option is using numerical gradients. The second derivatives are calculated from the gradients by numerical difference methods and are very expensive. Not all search algorithms require Hessian calculation, but the ones that do, are generally more accurate and better at finding stationary points. To reduce the cost of algorithms which use Hessians, those are often approximated, rather than calculated exactly. This is a characteristic of a wide range of minimization algorithms known as quasi-Newton methods. The most commonly used quasi-Newton optimization algorithm is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, and derivations of it, such as LMBFGS and Berny's algorithm.[8] In minimization algorithms, the performance is measured by the number of iterations the algorithm takes to converge to a point with zero gradients within a cut-off value. The BFGS methods have linear convergence. HONDO,[9] NWCHEM and GAUSSIAN 98,[10] the programs that were used for optimization in this work, all use variations of the BFGS method.

3.3 Software and Hardware

The software used for computational chemistry comes in many different varieties. Although many programs have very extensive functionalities, there is no such thing as a comprehensive computational chemistry program. Very often one must combine different

programs, to obtain the desired result. In the work presented here, several programs were used and a few different hardware architectures.

GAUSSIAN 98 was used to optimize the equilibrium structures and find transition states in the system. To find the reaction path the program DRDYGAUSS[11] was used, which is an extension to GAUSSIAN 98 and uses its electronic structure functionality to calculate energies, gradients, and Hessians at points along the reaction coordinate. The output of this program was used as an input to the reaction kinetics program POLYRATE,[12] which calculates rates of reaction. The excited state calculations were done with NWCHEM and HONDO for selected geometries along the reaction path calculated by DRDYGAUSS. Apart from GAUSSIAN 98 and DRDYGAUSS, all the other programs were used in parallel. More details about how these calculations were performed are given in Chapter 4. For visualization of the molecular structures and for producing the images, we used the molecular viewers MOLDEN,[13] ECCE,[14] and Jmol.[15]

Based on the resources required, the calculations were carried out either on a desktop, a multiprocessor machine, or a cluster. For all serial calculations, a 2-CPU Linux desktop was used. The parallel calculations were carried out either on an IBM multiprocessor machine with 64 Power 5 processors on a single node, 128 GB of shared-memory, and AIX operating system, or a Linux cluster of 980, 1.5-GHz, dual processor nodes with Linux operating system (EMSL MPP2 supercomputer).

4 APPLICATION

Reaction Pathways and Rates in $H_2O_2 + OH \rightarrow HO_2 + H_2O$

Hydroxyl radical (OH·) is an important reactive species in a variety of chemical environments. In atmospheric chemistry it plays an important role as a catalyst in the destruction of O_3 in the stratosphere.[16] The reactions leading to loss and production of OH· are considered among the most important ones in atmospheric chemistry. The reactivity of OH radicals is also relevant to several technologies, such as nuclear technology, medical technologies, and electron-driven processes in water.[17] Indeed OH is a very reactive species that results from exposure of water to radiation.[18] When it is subjected to photoexcitation, the hydrogen peroxide molecule H_2O_2 is an important source of OH radical. Finally, in biochemistry recently there has been a considerable interest in the chemistry of H_2O_2 and its derivatives, in part due to their involvement in various biological processes. In fact, H_2O_2 is often used to inactivate cells and microorganisms.[19] There is growing evidence that H_2O_2 plays a role in regulating cellular functions, as well.[20] Therefore it is important to understand the reaction mechanisms and rates for the reaction of OH radical with hydrogen peroxide both in gas phase and solution.

Several computational studies have been published dealing with the reaction of OH radical with hydrogen peroxide.[21,22,23]

$$H_2O_2 + OH \rightarrow HO_2 + H_2O \qquad (13)$$

The reaction is a simple hydrogen abstraction that leads to the formation of a peroxy radical and water. The gas phase reaction kinetics are complex and a hydrogen-bonded complex exists in the

entrance channel the reaction. Experimentally, the rates have been measured in a temperature range from 96 to 1680 K,[24,25] and show inverted temperature dependence below room temperature.

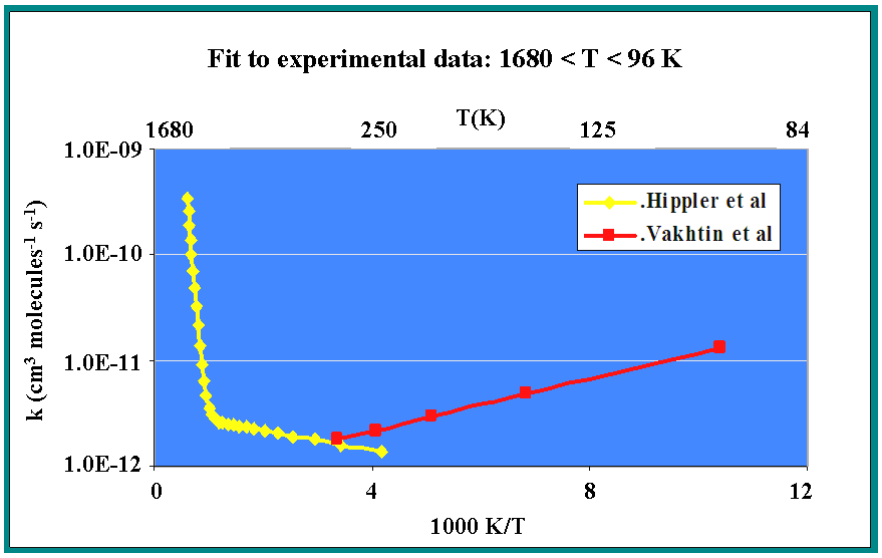

Figure 6. Experimental rates for $H_2O_2 + \bullet OH \rightarrow H_2O + \bullet O_2H$ in gas phase

The Arrhenius plot of the measured rates in gas phase is U-shaped, unlike the standard linear-shape, which is indicative of a complex reaction mechanism.

The calculation of the rate of the reaction in the gas phase for temperatures in the range of 250-500 K is reported. In this range, the experimental rate measurements show Arrhenius temperature behavior with a small activation energy that is consistent with our ground state potential energy surface. Our calculations for the excited state show that the barrier to reaction on the excited state surface is much higher than the barrier on the ground state surface, and is in fact comparable to the activation energy (~29 kcal/mol) for the strong temperature dependence experimentally observed above 900 K.[24] Thus, we suggest that the high temperature reaction involves reaction on the ground and the excited state surfaces.

In aqueous phase the reaction is known to proceed more slowly than in the gas phase,[26] an indication that the activation barrier for the process is higher in water than in the gas phase and that the aqueous solvation is stronger for the reactants than for the transition state.

Using density functional theory and other highly correlated wavefunction theories, we revisited the characterization of the gas phase reaction pathway for reaction (13) on the ground state potential energy surface. The reaction scheme is depicted in Figure 7.

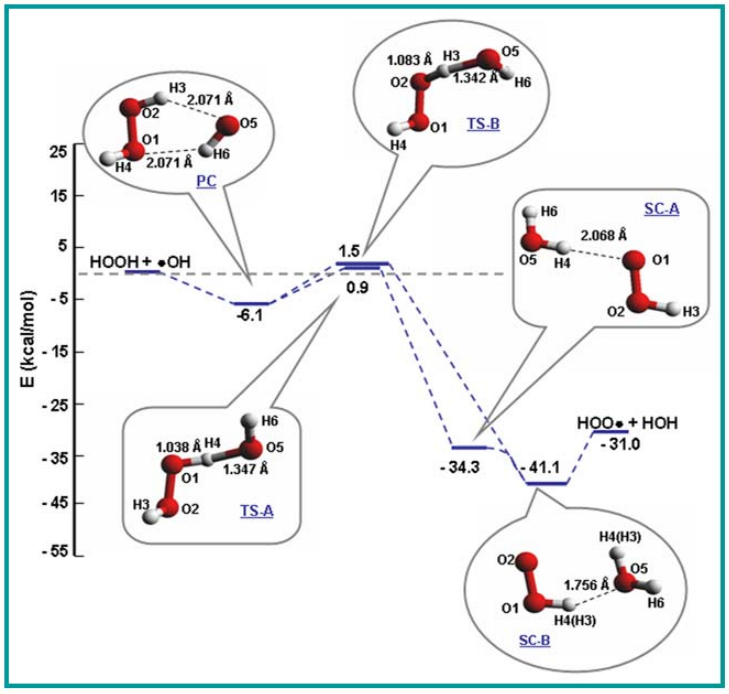

Figure 7. Structures of stationary points and schematic representation of the potential energy surface of the gas phase reaction $HO+H_2O_2 \rightarrow H_2O+HO_2$.

A precursor complex (hereinafter PC), a transition state (hereinafter TST-A), and a successor complex (SC-B), all have been identified in previous work.[21,22,23] In this work, a second transition state was found, (hereinafter TST-B) over which this reaction proceeds and it involves the distal OH bond of H_2O_2. Following the reaction pathway from TST-A, another local

minimum (SC-A) was found, which is higher in energy than the successor complex. A detailed characterization of these two reaction pathways and the potential energy surface of the ground and the first excited state along each of the pathways is presented. Preliminary results for the barrier to reaction in solution, calculated using the QM/MM methodology are also reported.

4.1 Computational methods

The geometries of the minima and the saddle points on the ground state potential energy surface were optimized with GAUSSIAN 98 using the DFT level of theory with the 6-31+G** basis set and the MPW1K hybrid density functional.[27] The minimum energy paths were calculated using the corrected local quadratic approximation Page-McIver integrator, with the program DIRDYGAUSS. The ground and the excited state calculations were carried out for points along the minimum energy paths using the time-dependent DFT (TDDFT).[28] The satisfactory accuracy of the methods was confirmed by comparing the results with higher levels of theory and bigger basis sets.

The reaction rates were calculated using canonical variational transition state theory, with quantum effects included through the method of small curvature tunneling (SCT).[29,30] The calculations were carried out with the program POLYRATE 9.3.1 for the temperature range of 250-500 K. The total rate was estimated as the sum of the rates for each of the reaction pathways. A detailed account of the calculations in gas phase can be found in reference (31).

The reaction in solution was studied with the QM/MM method implemented in NWCHEM. The QM part was treated with DFT/MPW1K and 6-31+G**, and the MM was described using the AMBER99 force field.[32]

4.2 Results and Discussion

4.2.1 The reaction pathways and ground state energetics

This mechanism of the reaction is illustrated in Figures 3 and 7. The reaction of hydrogen abstraction from H_2O_2 proceeds via a precursor complex (PC), a five member ring-like structure, with two hydrogen bonds. Except for one hydrogen atom of the peroxide molecule, all the other atoms essentially reside in a plane. From this precursor complex the reaction follows one of two pathways.

The most chemically intuitive pathway involves the direct attack on the proximal –OH bond of the peroxide molecule by the oxygen atom of the OH radical. The H atom transfers from peroxide to OH and following the transition state along the pathway, the system evolves toward a successor complex where the newly formed water molecule is hydrogen-bonded to the newly formed peroxy radical. This pathway is depicted in figure 8.

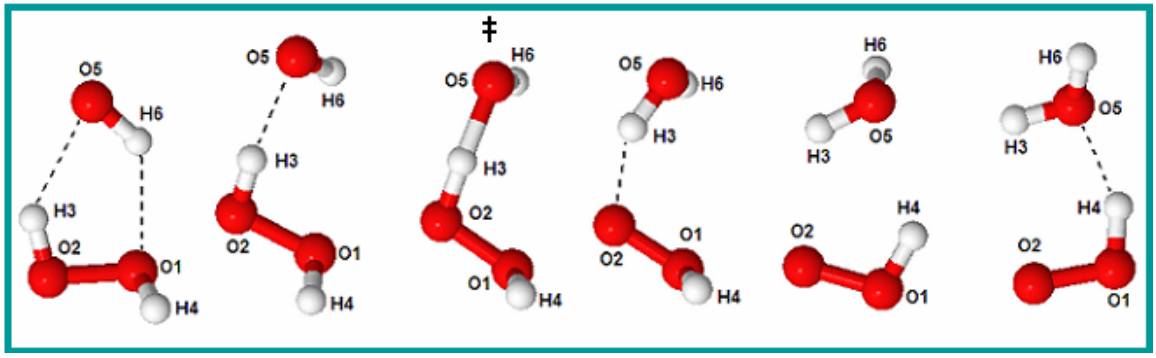

Figure 8. Reaction path B

A less chemically intuitive pathway involves a hydrogen atom abstraction from the distal OH bond of the peroxide molecule. It involves concerted rotations of the OH radical and of the H_2O_2 molecule, specifically a rotation of the peroxide molecule about the O-O bond concurrent

with a rotation of the OH radical and breaking of the peroxide-OH hydrogen bonds. The OH radical abstracts the distal H atom on its way to a successor complex. This pathway is depicted in figure 9.

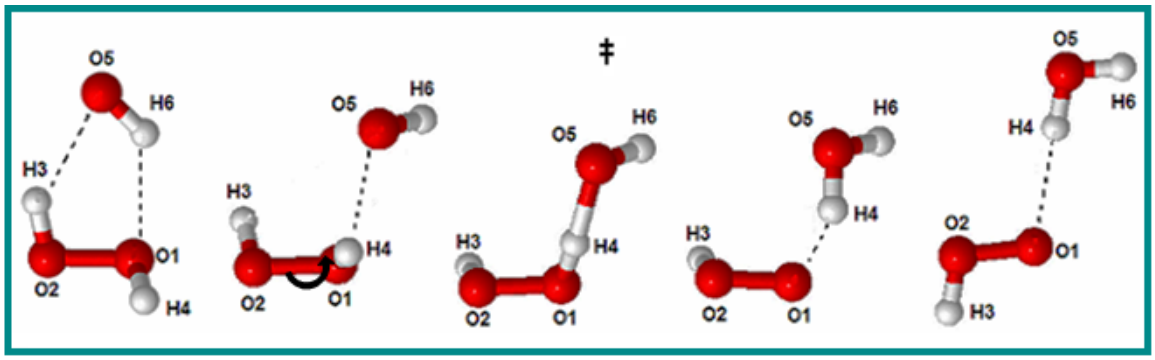

Figure 9. Reaction path A

The two pathways involve two notably different transition state structures labeled TST-A or TST-B (Fig. 7). They differ in the orientation of the hydroxyl-hydrogen and the incipient hydroperoxyl-hydrogen relative to the plane formed by the three oxygen atoms in the system.

The energies of the species and the complex structures of the reaction are shown in Fig. 1. The barrier heights on the ground state pathways are 7.3 (TST-A) and 7.8 kcal/mol (TST-B) above the precursor complex. The precursor complex is 6.2 kcal/mol lower in energy than the separated H_2O_2 and OH species.

4.2.2 Lowest excited state and activation barriers

The vertical excitation energies were calculated with TDDFT for the lowest excited state of the precursor and the successor complexes, the two transition states, and also of a selected number of points along the ground state reaction pathways.

On the reactant side, the first excited state arises from the electronic transition in the OH radical. For an isolated gas phase hydroxyl radical, the ground state is doubly degenerate, which means that there are two molecular orbitals on the oxygen atoms that share the electron occupancy and have the same energy. In other words, the ground and the first excited state in OH are the same. On the product side, the terminal O atom in the HO_2 molecule has the same type of first excited state. When the reactant and the products are complexed making PC and SC-B, the degeneracy is lifted, and the ground state prefers a specific orbital occupancy. This gives the low lying excited states, corresponding to a similar electronic transition in the O atom in both OH and HO_2. In the precursor complex, the excitation energy was calculated to be ~ 0.3 eV and in the successor complex it is ~ 1 eV.

To compare the results obtained in these calculations with experimental values, we calculated the second excited state in gas phase OH radical, and found it to be 4.2 eV, in very good agreement with the experimental value of 4.0 eV.[33] For the isolated HO_2, this excitation was ~ 1.0 eV. In the precursor complex and in the successor complex, the excitation energies are slightly larger than in the isolated species, due to the hydrogen bonding that holds the complexes together. The calculated energies for different theories are displayed in Table 2.

Table 2. Vertical excitation energies (relative to the separated reactant species).

	PC	*TST-A*	*TST-B*	*SC-A*
EOM-CCSD(T)[34]				
Ground state	-5.5 (-0.2)	6.4 (0.3)		-38.4 (-1.7)
First excited state	2.0 (0.1)	22.0 (1.0)		-9.3 (-0.4)
TDDFT				
Ground state	-6.2 (-0.3)	1.6 (0.1)	1.1 (0.0)	-38.9 (-1.7)
First excited state	2.4 (0.1)	14.3 (0.6)	18.2 (0.8)	-10.4 (-0.5)
MCQDPT2[35]				
Ground state	-4.2 (-0.2)	7.8 (0.3)	7.1 (0.3)	-45.2 (-2.0)
First excited state	-2.1 (-0.1)	20.2 (0.9)	16.5 (0.7)	-16.8 (-0.7)

* Energies in kcal/mol (values in parenthesis are in eV).

The three theories yield the same near quantitative pictures of the H_2O_2 + OH system in its ground state and the first excited state.

The excited state barrier was found by calculating the excitation energy along selected points of the reaction pathways. Figure 10 shows the energies of the ground state and the first excited state for points of the reaction pathway near the transition state TST-A. For the pathway proceeding via TST-B, the potential energy surfaces are very similar to those going through TST-A.

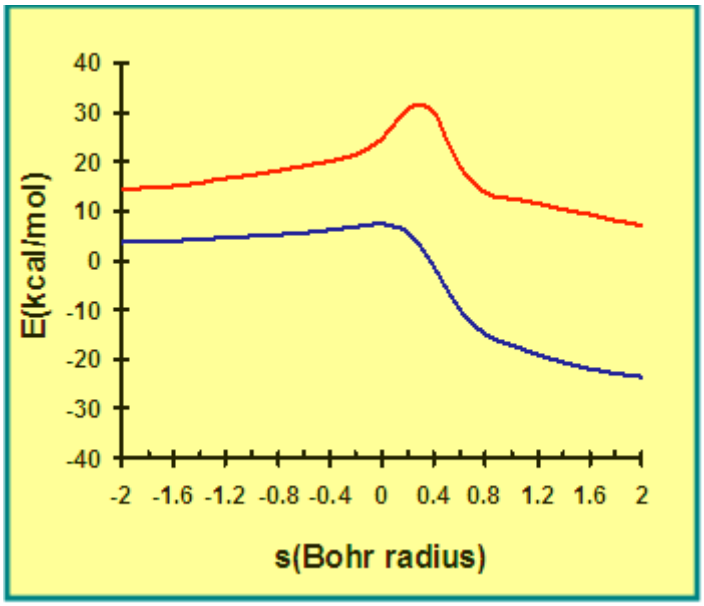

Figure 10. Ground and first excited state along pathway over TST-A

The zero of energy in Fig. 10 is assigned to the precursor complex (PC). The barrier heights (measured from the energy of the separated reactants) on the excited state potential surfaces are 24.6 kcal/mol for the TST-A pathway and 22.0 kcal/mol for the TST-B pathway. The barriers

are slightly displaced towards the product region compared to the ground state barriers and are significantly narrower.

4.2.3 Reaction rate calculations in gas phase

There have been a number of studies, both experimental and theoretical, investigating the rate constant of the reaction (11) in the gas phase. The experimental results show the rate to vary anywhere from 8.4×10^{-13} to 2.0×10^{-12} cm^3 $molecule^{-1}$ s^{-1} for room temperature.[36] Theoretical calculations done by Atadinç *et al.*[23] with transition state theory estimate the rate to be between 2.6×10^{-13} to 4.2×10^{-13} cm^3 $molecule^{-1}$ s^{-1}, depending on the level of theory used.

In this work, canonical variational transition state theory was used. The rate expression is similar to the one for TST (eq. 10), but it also includes correcting factors.

$$k^{CVT/SAG}(T,s) = \kappa^{CVT/SAG}\sigma\frac{K_bT}{h}\frac{Q^{GT}[T,s^{CVT}(T)]}{\Phi^{R}(T)}e^{\left[-\frac{V_{mep}[s^{CVT}(T)]}{K_bT}\right]} \qquad (14)$$

In the rate expression above, σ is the symmetry factor accounting for the two possibilities of hydrogen abstraction reaction from H_2O_2 and s^{CVT} is the optimum position of the dividing surface. To account for quantum effects along the reaction mode, the coefficient $\kappa^{CVT/SAG}$ is included.

Equation (14) is valid for reaction (13) provided PC is in rapid equilibrium with OH and H_2O_2 [37] and the conversion of PC to SC is rate limiting under experimental conditions. Since the

rate constants deviate from normal Arrhenius behavior at low and high temperatures, we use equation (14) to calculate rate constants within the range 250-500 K.

The rate constants were calculated separately for each of the reaction paths, k_A and k_B, and the total rate is the sum of k_A and k_B. When calculating the frequencies in the species, harmonic oscillator approximation is employed to represent the movements of the atoms. This approximation tends to introduce lager errors when the motion described is torsion (hindered rotation). To correct for this error, Truhlar's method of interpolating the partition functions between the free-rotor and harmonic-oscillator limit[38] was included. To account for quantum effects over the barrier, a small curvature tunneling (SCT) correction was included in the calculations. For a temperature of 298 K, the calculated rate is 9.8×10^{-13} cm^3 $molecule^{-1}$ s^{-1}. The rates calculated for temperatures in the range 250 – 500 K, for paths A and B described above (Figs. 7-9), are given in Table 3.

Table 3. Rates, quantum corrections, Arrhenius activation energies and preexponential factors for the two reaction pathways.

T(K)	$\kappa^{CVT/SAG}$		*[a]k ×10^{-13} (cm^3 molecule^{-1} s^{-1})*		*k ×10^{-13} (cm^3 molecule^{-1} s^{-1})*		*Ea (kcal/mol)*		*log(A)*	
	A	B	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
250	2.03	2.34	2.8	2.9	1.7	6.3	0.84	1.41	-12.03	-10.96
298	1.67	1.85	3.2	3.5	2.1	7.7	0.98	1.54	-11.96	-10.98
400	1.35	1.42	4.4	5.1	3.1	11.9	1.33	1.89	-11.79	-10.89
500	1.21	1.26	6.0	7.1	4.3	17.4	1.73	2.29	-11.61	-10.76

[a] Rates with SCT but without hindered rotation correction.

This table also contains the corresponding Arrhenius activation parameters for the two pathways. Path B is the faster of the two reaction paths, despite having the higher barrier. We analyzed the factors that contribute to the rates to find that factors which depend on quantum and hindered rotation corrections are larger for path B than the corresponding corrections for path A.

The total rates are plotted in Figure 11 along with experimental data taken from the NIST[36] database. The experimental data cluster into two groups with the preferred data being the group that exhibits higher rates and lower activation barriers.

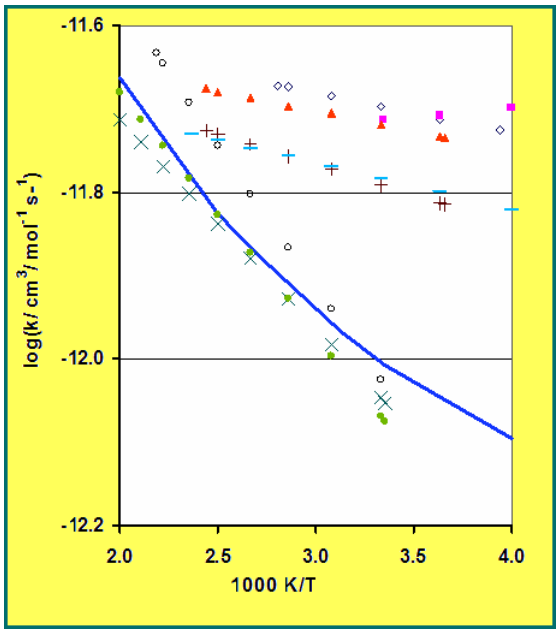

Figure 11. Rate constants for the gas phase reaction $HO+H_2O_2 \rightarrow H_2O+HO_2$. Experiments: ◊ Ref.39; ■ Ref. 40; ▲Ref. 41: + Ref. 42; – Ref.16; × Ref. 43; ● Ref. 44; ○ Ref. 45. Full line: from this work.

Although our calculated rates agree better with the group that exhibits lower rates and higher activation barriers, the differences are not significant considering the accuracy of the MPW1K method and the rigid rotor-harmonic oscillator approximation. An Arrhenius analysis of our total rates leads to activation parameters that are larger than the preferred experimental values by less than a factor of 2 in the preexponential factor and less than 1 kcal/mol in the activation energy. Finally, it must be reiterated that in these rate calculations, it was assumed that the precursor complex is in rapid equilibrium with the reactants. Absent this experimental condition, the measured rates would be limited to a degree by the rate of association of OH and H_2O_2, in which case faster rates and smaller activation barriers may be observed.

Measurements of the rate at lower temperatures show that there is an inversion of the rate dependence below room temperature, which has been attributed to the existence of the precursor complex during the reaction.[24,46] A kinetic analysis that takes into account the pressure-dependent association and dissociation rate constants of the precursor complex[47] and tunneling along the path from PC to products should be applied to capture the non-Arrhenius behavior at low temperatures.

Unusual behavior of reaction (13) has also been observed at high temperatures. Hippler *et al.*[25] found that the rate of reaction increases greatly with T in the range $800 < T \leq 1600$ K. It corresponds to an Arrhenius activation energy of ~30 kcal/mol. To account for this behavior, Hippler[25] and later Vakhtin[24] suggested a change in mechanism from complex formation in the lower T range to direct H transfer above 800 K. Bahri *et al.*[21] have rationalized this behavior by suggesting that at high temperature the reaction occurs from the complex, whereas at lower temperature, collisions that form the TS directly without going into the complex are productive. The latter explanation does not seem consistent with TST. At 300 K, the PC is not favored relative to the separated reactants ($\Delta G° = 2.4$ kcal/mol). Higher temperatures will have the effect of making it increasingly less stable ($\Delta G° = 7.2$ kcal/mol at 500K). Therefore, the change in T dependence must be due to an alternative reaction path with a much higher barrier contributing to the reaction rate. We suggest that reaction via the low-lying first excited state of the reactive complex contributes at high T. From Table 3, the lowest excited transition states are ~22 and ~32 kcal/mol above the separated reactants according to TDDFT and EOM-CCSD(T) levels of theory, respectively. These values are in very good accord with the experimentally predicted activation barrier (~29 kcal/mol) for temperatures over 900 K.

4.2.4 Free energy barriers in solution

To find the free energy barrier in water, we used the gas phase reaction path and applied the QM/MM model of solvation implemented in NWCHEM. The quantum region included the H_2O_2 and OH molecules, for which the DFT/MPW1K level of theory was used in conjunction with the 6-31+G** basis set. The QM system was solvated in a cubic box of 30 angstroms, with 889 SPC/E[48] water molecules, which made up the MM region. The Lenard-Jones parameters for H_2O_2 and OH used in the QM/MM model were modified from those of Freindorf and Gao.[49] The original parameters were found to bind the water too strongly. To correct the parameters, we calculated the binding energy between the reaction species and a single water molecule with QM/MM, and varied the parameters until we found a combination that reproduces the binding energy and structure found with QM.

The free energy is calculated between two successive points (A and B) along the solute reaction pathways by applying Zwanzig's equation,[50]

$$\Delta F^{AB} = -kT \ln\left\langle e^{-\left(E^A - E^B\right)/KT} \right\rangle_A \qquad (15)$$

In eq. (15) the angular brackets denote the statistical average of the energy difference (E_A-E_B) between the solute in the A and B conformations, with the averaging carried over the solvent configurations around the solute in its A conformation. By introducing an effective intermediate classical representation of the QM region, we define a lower level model denoted MM/MM. This model is computationally economical and can be used to calculate the changes in free energy along the reaction pathway by defining the thermodynamic cycle shown below:

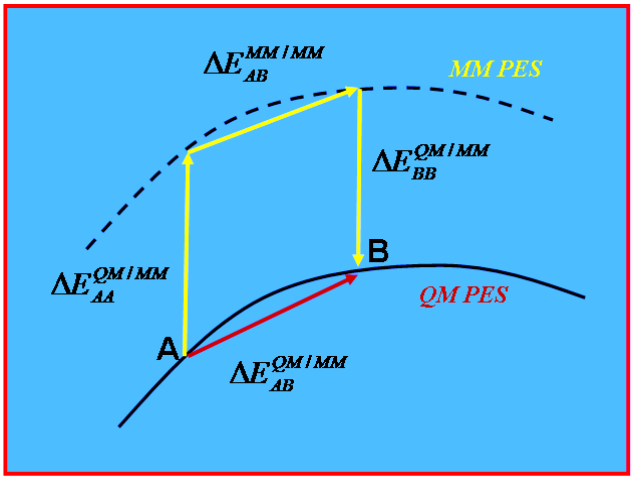

Figure 12. QM/MM thermodynamic cycle

From this formulation, it follows that $\Delta E_{AB}^{QM/QM}$ can be expressed by relation (16),

$$\Delta E_{AB}^{QM/MM} = \Delta E_{AA}^{QM/MM} - \Delta E_{BB}^{QM/MM} + \Delta E_{AB}^{MM/MM} \qquad (16)$$

where the first 2 terms are the difference between the QM and MM representation of the QM atoms and the third term is the energy difference averaged at the MM/MM level.

To create the averaging ensemble of the solvent molecules for a fixed solute structure, one uses the MM representation of the solute subsystem. For each solute configuration, a large number (~10^5-10^6) of solvent configurations are generated at the MM/MM level of theory, at constant temperature and volume (NVT ensemble), giving the change $\Delta E_{AB}^{MM/MM}$. Then, a subset (~10^{2}-10^{3}) is selected randomly from the solvent configurations and the total energies of the system with the solute structures A and B are calculated at the QM/MM level of theory. The subset gives the averaged values $\Delta E_{AA}^{QM/MM}$ and $\Delta E_{BB}^{QM/MM}$. This approach is called double -

perturbation.[51] It can be further simplified if instead of finding the average values of $\Delta E_{AA}^{QM/MM}$ and $\Delta E_{BB}^{QM/MM}$, one takes a representative solvent configuration to calculate $\Delta E_{BB}^{QM/MM}$ and $\Delta E_{BB}^{QM/MM}$, and use those values in equation (16).

Using the above described approach, we calculated the free energy difference (ΔG) relative to the precursor complex geometry. A total of 17 points along the reaction coordinate were used, giving the energetic profile shown in Figure 13.

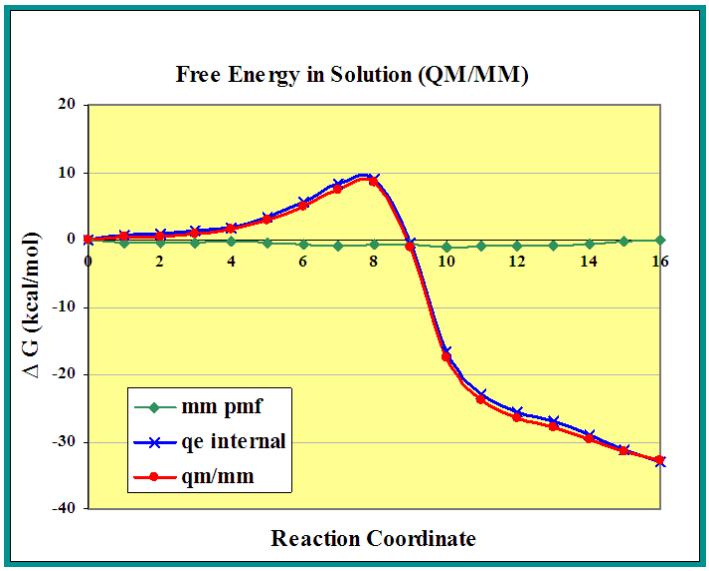

Figure 13. Free energy barriers in QM/MM

The free energy barrier relative to the precursor complex, in solution was calculated to be ΔG_{sol} = 8.5 kcal/mol, which is 3.6 kcal/mol higher than the free energy barrier in gas phase (ΔG_{gas}= 4.9 kcal/mol).

The rate of the reaction in solution has been examined experimentally[26,52] and determined to be ~5×10^{-14} cm^3 $molecule^{-1}$ s^{-1}. The rate is about 2 orders of magnitude slower than in the gas phase,[36] which agrees qualitatively with the calculated difference in the barriers.

The initial results follow the experimental observations, but to be able to properly characterize the reaction in solution, the section of the reaction path from the precursor complex to the separated species must be studied. Also, to obtain quantitative rate constant in solution, rate calculations need to be carried out with the solution state energetics for the reaction path.

4.3 Summary

The reaction was studied in both gas phase and condensed phase. In gas phase two distinct reaction pathways were identified, and the rate was calculated using variational transition state theory. The calculations provided a basis for explaining the unusual observed dependence of the rate at temperature above 900 K. In solution, the reaction was studied using the QM/MM methodology. The free energy barrier was found to be higher than in the gas phase, which is in accord with the experimental findings of the rate being slower in solution.

5 CONCLUSION

Computational solutions to chemical problems are as important to the advancement of science as theory and experiment are. Chemical systems of various sizes can be simulated on computers and one can derive detailed pictures of chemical reactions. An overview of a number of computational approaches was given which can be used to this end, and the computational advantages and disadvantages of selected algorithms, as well as their applicability and accuracy were addressed. An application of some of the methods was presented in a study of the reaction of hydrogen peroxide and hydroxyl radical (i.e. $H_2O_2 + OH \rightarrow H_2O + O_2H$). The computations carried out led to a kinetic model that better explains the experimental findings for the gas phase rate than work previously published. Some preliminary work was also done on the reaction in

solution, indicating that the QM/MM approach gives a good starting point for such study, and should be further pursued. This work demonstrates the power of computational studies in explaining and predicting properties of chemical reactions in different environments.

APPENDIX A: Self-Consistent Field Algorithm

The Schrodinger equation was already introduced, and here follows an expanded explanation of how one obtains the energy of the system computationally. When the Born-Oppenheimer approximation is used, one finds the electronic energy of the system. This simplifies the Schrodinger equation, which now becomes:

$$H_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{A.1}$$

The wavefunction can be approximated so that it only depends on the fixed coordinates of the nuclei and of one electron. This is a simplification to the many-electron wavefunction.

Commonly the eigenfunctions of the equation are represented by molecular orbitals, i.e. the *N*-electron wavefunction Ψ is replaced by a product of single-particle orbitals and can mathematically be represented as a so-called Slater determinant, an *N* x *N* matrix, where *N* is the total number of electrons. The Slater determinant is expanded in terms of the set of basis functions, and a density matrix P which contains the expansion coefficients for the basis functions. Hartree-Fock ground state energy is obtained by minimizing with respect to the variation of the orbitals, through the density matrix, subject to the constraint that the orbitals remain orthonormal. Within the Hartree-Fock (H-F) theory, the energy is dependent both on the Fock matrix and the density matrix. Because the Fock matrix is also dependent on the density matrix, the calculation is converged by self-consistently varying the coefficients in the density matrix, until both the energy and the density matrix coefficients are converged.

The iterative procedure executes the following steps:

1. Initial guess for the density matrix P
2. Construct the Fock Matrix from P, 1-e integrals and 2-e integrals
3. Translate the equation into an eigenvalue problem, and solve it (matrix diagonalization)
    a. Eigenvalues: used to compute the energy
    b. Eigenvectors: used to construct new density matrix
4. Check convergence for the energy and the density matrix
5. Repeat steps 2 to 4 until the convergence is achieved

The Fock matrix is dependent on a density matrix and two-electron integrals, both of which are basis set dependent. The construction of the Fock matrix is $O(N^4)$, but can be efficiently parallelized, because the calculation of the two-electron integrals can be done independently. However, following this is a step of matrix diagonalization (step 3). This step cannot be easily divided into pieces, and it is known as the bottleneck associated with $O(N^3)$ diagonalization of the Fock matrix within each iteration of the SCF procedure. [53]

REFERENCES:

[1] P. A. M. Dirac, Proc. Roy. Soc.(London), 123714 (1929)

[2] K. Fukui, J. Phys. Chem., **74**, 4161 (1970).

[3] T. L. Hill, *An Introduction to Statistical Thermodynamics*, (Addison-Wesley Publishing Company, Inc. London, England,1962), pp. 120.

[4] A. Fernandez-Ramos, B. A. Ellingson, B. C. Garrett, and D. G. Truhlar, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley-VCH, Hoboken, NJ, 2007), Vol. 23, pp. 125-232.

[5] J. Nieplocha, R. J. Harrison, R. J. Littlefield, J. Supercomput., **10**, 197, (1996).

[6] E. J. Bylaska, W. A. de Jong, K. Kowalski, T. P. Straatsma, M. Valiev, D. Wang, E. Aprà, T. L. Windus, S. Hirata, M. T. Hackler, Y. Zhao, P.-D. Fan, R. J. Harrison, M. Dupuis, D. M. A. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, A. A. Auer, M. Nooijen, E. Brown, G. Cisneros, G. I. Fann, H. Früchtl, J. Garza, K. Hirao, R. Kendall, J. A. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, L. Pollack, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z.

Zhang, NWCHEM, A Computational Chemistry Package for Parallel Computers, version 5.0, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA,2006; R.A. Kendall, E. Aprà, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G.I. Fann,R.J. Harrison,J. Ju, J.A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus, A. T. Wong, Computer Phys. Comm. **128**, 260 (2000).

[7] H. B. Schlegel and M. J. Frisch, in *Theoretical and Computational Models for Organic Chemistry*, edited by S.J. Formosinho, I.G. Csizmadia and L.G. Arnaut (Kluwer Academic Publishers , Netherlands, 1991), pp. 5-34.

[8] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, J. Comp. Chem. **17**, 49 (1996).

[9] M. Dupuis, A. Marquez, and E. R. Davidson, HONDO 2000, based on HONDO 95.3, Quantum Chemistry Program Exchange (QCPE), Indiana University, Bloomington, In 47405, 2000.

[10] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong,

J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, GAUSSIAN 98, Revision A.7, Gaussian, Inc., Pittsburgh, PA, 1998.

[11]B. C. Garrett, DIRDYGAUSS, Environmental Molecular Sciences Laboratory, Pacific Northwest Laboratory; Y.-Y. Chuang and D. G. Truhlar, Department of Chemistry and Super Computer Institute, University of Minnesota.

[12] J. C. Corchado, Y.-Y. Chuang, P. L. Fast, W.-P. Hu, Y.-P. Liu, G. C. Lynch, K. A. Nguyen, C. F. Jackels, A. Fernandez-Ramos, B. A. Ellingson, B. J. Lynch, V. S. Melissas, J. Villa, I. Rossi, E. L. Coitino, J. Pu, T. V. Albu, R. Steckler, B. C. Garrett, A. D. Isaacson, and D. G. Truhlar, POLYRATE, version 9.3.1. University of Minnesota, Minneapolis, 2005, http://comp.chem.umn.edu/polyrate

[13] G. Schaftenaar and J. H. Noordik, "MOLDEN: a pre- and post-processing program for molecular and electronic structures", J. Comput.-Aided Mol. Design, **14**, 123 (2000).

[14] G. Black, J. Daily, B. Didier, T. Elsethagen, D. Feller, D. Gracio, M. Hackler, S. Havre, D. Jones, E. Jurrus, T. Keller, C. Lansing, S. Matsumoto, B. Palmer, M. Peterson, K. Schuchardt, E. Stephan, L. Sun, K. Swanson, H. Taylor, G. Thomas, E. Vorpagel, T. Windus, C. Winters, ECCE, A Problem Solving Environment for Computational Chemistry, version 4.0.2, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA, (2006).

[15] Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

[16] P. H. Whine, D. H. Semmes, A. R. Ravishankara, J. Chem. Phys. **75**, 4390 (1981).

[17] B. C. Garrett, D. A. Dixon, D. M. Camaioni, D. M. Chipman, M. A. Johnson, C. D. Jonah, G. A. Kimmel, J. H. Miller, T. N. Rescigno, P. J. Rossky, S. S. Xantheas, S. D. Colson, A. H. Laufer, D. Ray, P. F. Barbara, D. M. Bartels, K. H. Becker, K. H. Bowen, Jr., S. E. Bradforth, I. Carmichael, J. V. Coe, L. R. Corrales, J. P. Cowin, M. Dupuis, K. B. Eisenthal, J. A. Franz, M. S. Gutowski, K. D. Jordan, B. D. Kay, J. A. LaVerne, S. V. Lymar, T. E. Madey, C. W. McCurdy, D. Meisel, S. Mukamel, A. R. Nilsson, T. M. Orlando, N. G. Petrik, S. M. Pimblott, J. R. Rustad, G. K. Schenter, S. J. Singer, A. Tokmakoff, L. S. Wang, C. Wittig, and T. S. Zwier, Chem. Rev. **105**, 355 (2005).

[18] G. V. Buxton, C. L. Greenstock, W. P. Helman, A.B. Ross, J. Phys. Chem. Ref. Data, **17**, 513 (1988).

[19] A. Delabie, S. Creve, B. Coussens, M. T. Nguyen, J. Chem. Soc., Perkin Trans. 2, 977 (2000).

[20] G. Nindl, Cellscience Reviews, **1**, 1 (2004).

[21] M. Bahri, Y. Tarchouna, N. Jaïdane, Z. Ben Lakhdar, J.P. Flament, J. Mol. Struct.: THEOCHEM **664**, 229 (2003).

[22] B. Wang, H. Hou, Y. Gu, Chem. Phys. Lett. **309**, 274 (1999).

[23] F. Atadinç, H. Gunaydin, A. S. Ozen, A. Aviyente, Int. J. Chem. Kin. **37**(8), 502 (2005).

[24] A. B. Vakhtin, D.C. McCabe, A. R. Ravishankara and S. R. Leone, J. Phys. Chem. A **107**, 10642 (2003).

[25] H. Hippler, H. Neunaber, J. Troe. J. Chem. Phys, **103**, 3510 (1995).

[26] H. Christensen, K. Sehested, H. Corfitzen, J. Phys. Chem, **86**, 1588 (1982).

[27] B. J. Lynch, P. L. Fast, M. Harris, D. G. Truhlar, J. Phys. Chem. A **104**, 4811 (2000).

[28] C. Jamorski, M. E. Casida, and D. R. Salahub, J. Chem. Phys. **104,** 5134 (1996); R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256,** 454 (1996); R. Bauernschmitt, M. Häser, O. Treutler, and R. Ahlrichs, Chem. Phys. Lett. **264,** 573 (1997).

[29]D.-h. Lu, T. N. Truong, V. S. Melissas, G. C. Lynch, Y.-P. Liu, B. C. Garrett, R. Steckler, A. D. Isaacson, S. N. Rai, G. Hancock, J. G. Lauderdale, T. Joseph, and D. G. Truhlar, Comp. Phys. Comm. **71**, 235 (1992).

[30] Y.-P. Liu, G. C. Lynch, T. N. Truong, D. Lu, and D. G. Truhlar, J. Am. Chem. Soc. **115**, 2408 (1993).

[31] B. Ginovska, D. M. Camaioni and M. Dupuis, J. Chem. Phys. **127**, 84309 (2007).

[32] J. Wang, P. Cieplak, P. A. Kollman. J. Comp. Chem. **21**, 1049 (2000).

[33] K. Minschwaner, T. Canty, and C. Burnett, J. Atmos. Terr. Phys. **65**, 335 (2003).

[34] K. Kowalski, P. Piecuch, *J. Chem. Phys.* **120**, 1715 (2004).

[35] H. Nakano, *J. Chem. Phys.* **99**, 7983 (1993); H. Nakano, *Chem. Phys. Lett*. **207**, 372 (1993).

[36] NIST Chemical Kinetics Database on the Web, NIST Standard Reference Database 17, Version 7.0 (Web Version), Release 1.4, August 2006, http://kinetics.nist.gov/kinetics

[37] A. Galano, J. R. Alvarez-Idaboy, L. A. Montero, A. Vivier-Bunge, J. Comp. Chem. **22,** 1138 (2001).

[38] Y.-Y Chuang and D. G. Truhlar, J. Chem. Phys. **112**, 1221 (2000); **121**, 7036(E) (2004).

[39] E. Jimenez, T. Gierczak, H. Stark, J. B. Burkholder, and A. R. Ravishankara, J. Phys. Chem. **108**, 1139 (2004).

[40] E. R. Lovejoy, T. P. Murrells, A. R. Ravishankara, and C. J. Howard, J. Phys. Chem. **94**, 2386 (1990).

[41] G. L. Vaghjiani and A. R. Ravishankara, J. Phys. Chem. **93**, 7833 (1999).

[42] L. F. Keyser, J. Phys. Chem. **84**, 1659 (1980).

[43] R. R. Baldwin and R. W. Walker, J. Chem. Soc., Faraday Trans. 1 **75**, 140 (1979).

[44] W. Hack, K. Hoyermann, and H. G. Wagner, Proc. Symp. Chem. Kinet. Data Upper Lower Atmos. **1974**, 329 (1975).

[45] N. R. Greiner, J. Phys. Chem. **72**, 406 (1968).

[46] I.W.M. Smith, A. R. Ravishankara. J. Phys. Chem. A **106**, 4798 (2002).

[47] S.S. Brown, J.B. Burkholder, R.K. Talukdar, A.R. Ravishankara. J. Phys. Chem. A **105**, 1605 (2001).

[48] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[49] M. Freindorf and J. Gao, J. Comp. Chem. **17**, 386 (1996).

[50] R. W. Zwanzig, J. Chem. Phys. **22,** 420 (1954).

[51] M. Dupuis, G. K. Schenter, B. G. Garrett, E. E. Arcia, J. Mol. Struct.: THEOCHEM **632**, 173 (2003).

[52] M. S. Alam, M. Kelm, B. S. M. Rao, and E. Janata, Radiat. Phys. Chem, **71**, 1087 (2004).

[53] R. Shepard, Theoretica Chim. Acta. **84**, 343 (1993).