NEW COMPUTATIONAL ALGORITHMS AND MOLECULAR

STRUCTURE STUDIES

FAN YANPING

NATIONAL UNIVERSITY OF SINGAPORE

2007

NEW COMPUTATIONAL ALGORITHMS AND MOLECULAR STRUCTURE STUDIES

FAN YANPING

(B. Sc., Shandong University, China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgements

I would like to express my immense gratitude to my supervisor, Dr. Ryan P.A. Bettens for his invaluable guidance of this work. He introduced me into the wonderful field of quantum chemistry. And his encouragement, support and friendly personalities were helpful and precious to the success of this research work. I will remember his kindness all whole my life.

I deeply appreciate the kind assistance from Dr. Adrian Michael Lee for his stimulating discussion and useful suggestions.

I also profoundly give my sincere thanks to my colleagues and friends who helped and supported me through the whole Ph.D studies, Enyi Ye, Jing Shi, Jiong Ran, Yifan, Xinming, Weiqiang for their advices and friendship.

Last but not least, my acknowledgement goes to National University of Singapore for awarding me the research scholarship and for providing the facilities to carry out the research work reported herein.

Table of Contents

Summary

New computational algorithms for predicting molecular energies and evaluating vibrational frequencies for large molecular systems are developed. Predicting energy and other related molecular properties accurately within in a short time period is a rigorous task. A fragmentation approach has been applied to transition metal complexes successfully.

The use of symmetry coupled with fragmentation allows the calculation of essentially infinitely large systems within a CPU budget. An extensive study of the harmonic frequencies of a large set of small polyatomic closed-shell molecules computed at both full ab initio and composite approximations using various combinations of basis sets and composite methods are capable of predicting full ab initio CCSD(T) level harmonic frequencies to within 5 cm^{-1} on average, which suggests a computationally affordable means of obtaining highly accurate vibrational frequencies compared to the CCSD(T) level. These new methods obtain high accuracy results in a very efficient way.

Interesting aspects of a few important molecules are well studied, such as hydroxyacetaldehyde, cyclodextrine based rotaxane. The conformers of hydroxyacetaldehyde are studied both with ab initio method and quantum diffusion Monte Carlo method. The potential energy surface (PES) of hydroxyacetaldehyde has been mapped and all the critical points identified. The rotational constants predicted from the simulations were found to be in excellent agreement with experiment for the only yet observed conformer in the gas phase. Cyclodextrine based rotaxane is a good candidate

for the design of molecular machines. This 184 atom rotaxane is investigated on its conformation, energy, geometries and movement by employing QM semi empirical AM1 method. The results well explain the experiment findings. And the movement mechanism of the molecular machine is provided and well explained. Based on this study it can also be inferred that the AM1 semi-empirical method is a good tool for analyzing mechanisms of large molecular systems especially motions of nano system.

Chapter 1 General Introduction

1.1 Computational Chemistry

In the studies of chemical phenomena, we want to comprehend various mechanisms in a systematic way. The purpose of theoretical studies is to find such methods and to formulate them into a mathematical and/or conceptual form. Experiment has its own significance. However, knowing the common concept underlying the principles behind it is valuable. Thus theoretical chemistry has become very important to make an active effort to correlate the real world to theory. The term "computational chemistry" is generally used when a mathematical method is sufficiently well developed that it can be automated for the implementation on a computer. Computational chemistry is an independent research area and has become a powerful tool in other research fields, especially chemistry, biology and interdisciplinary subjects, which focus on simple and complicated molecular-system studies. Computational chemistry can generally be classified into two categories: molecular mechanics and quantum mechanics.

1.2 Molecular Mechanics (MM)

Molecular mechanics is formulated based on the laws of classical (Newtonian) physics. In many cases, large molecular systems can be modeled successfully while avoiding quantum mechanics calculations entirely. Molecular mechanics simulations, for example, use a single classical expression for the energy of a compound, such as the harmonic oscillator. The CPU time for calculations scales as the square of the number of

atoms. As it is quick and simple, MM is being widely used in biology system studies, such as in proteins and enzymes. However the accuracy of MM is heavily dependent on parameterization. The accuracy of the database of compounds used for paramount to the success of molecular mechanics calculations. Without having fully documented molecules, MM often gives unacceptable results.

1.3 Quantum Mechanics

Molecules consist of nuclei and electrons, and thus the methods of quantum mechanics can be applied. Quantum mechanics states that the energy and other related properties of a molecule may be obtained by solving the Schrödinger equation or the Dirac equation in relativistic quantum chemistry using the electronic molecular Hamiltonian. This is to determine the electronic structure of the molecule.

1.3.1 Molecular Structure and Energy

Studies on molecular structure are very basic, yet very important in computational chemistry. Molecules' physicochemical properties are dependent on their structures and environments. One of the most explicit, and significant properties of a molecule is its energy. Energy is one of the most useful concepts in science. Analysis of energetic can predict what molecular processes are likely to occur, or are able to occur. All computational chemistry techniques defined energy as the function of the position of a specific structure.

The potential energy surface (PES) describes the energy in term of its structure. In other words, PES is the assembly of all molecular conformers with different structures. PES, which governs the interatomic motions, is the sum of the electronic energy and the

nuclear repulsion. Once a PES has been computed, it can be analyzed to determine a lot of information about the chemical system. PES is the most complete description of all the conformers, isomers, and energetically accessible motions of a system. Minima on this surface correspond to the optimized geometries. The lowest-energy minimum is called the global minimum. There can be many local minima, such as higher-energy conformers or isomers. The transition state structure between the reactants and products of a reaction is a saddle point on this surface. A PES can be used to locate both saddle points and reaction coordinates.

For non-linear polyatomic molecules having N atoms, the energy surfaces depend on 3N-6 internal coordinates and thus can be very difficult to visualize. A slice through such a surface (i.e., a plot of the energy as a function of two out of 3N-6 coordinates) is shown below and various features of such a surface are detailed. Figure 1.1 illustrates these topological features.

Molecular potential energy surfaces (PES) are required for the computation of reaction dynamics. In the Born-Oppenheimer approximation, the molecular potential energy is the total electronic energy, which can be evaluated using the methods of ab initio quantum chemistry. The potential energy surfaces of macroscopic systems are fundamental to the understanding of their structural, thermodynamic, and dynamic properties. An energy surface resembles a mountain range, complete with peaks (energy barriers), valleys (energy minima), and passes (saddle points). Many reactions' mechanisms can be successfully explained by many researchers upon getting the potential energy surface. Photodissociation reactions are the most commonly occurring reactions in the atmosphere, and attract much interest from the world as people now pay more attention to environment pollutions issues and effects on human health. However due to

the interaction between many surfaces, reaction mechanisms often show very complicated features. Dr.Liu[1-4] focuses on reaction mechanisms of the atmosphere and on combustion chemistry using accurate quantum chemical methods. Some radicals or ions (CHX(X=H, F, Cl), CH, CCN^+ CNC^+) and molecules (NO, N_2O, NO_2, CH_3OH, H_2O) play an important role in the atmosphere and in combustion processes. By constructing the profile of potential energy surfaces to determine the reaction mechanisms and the reaction rate constants.

Evidently, accuracy and efficiency in calculating energy and the related properties of molecule are the key objective in computational chemistry.

There are several Quantum Mechanics methods to resolve this issue.

1.3.2 VB Method

The commonly used method is the empirical valance bond (EVB) method. The bond functions and ionic terms have a simple and clear physical meaning. As a result, it is conceptually easier to define different states along a chemical reaction path in terms of VB configurations. A chemical reaction is described using a valence bond approach, i.e., the system wave function is represented by a linear combination of the most important ionic and covalent resonance forms and the potential energy is found by solving the related equation. The electronic interaction Hamiltonian is built using parameter terms extracted from empirical values and ab initio surfaces.

All the results obtained from EVB are quantitatively good; however the drastic limitation of EVB is the size of molecule. Results from EVB have to solve the Schrödinger equation. The Schrödinger equation provides the mathematical apparatus of the quantum physical description of the wave functions. It is, however, a more tricky

business to try to solve the Schrodinger equation by expanding the wave function in terms of the usual VB functions. If the system has to deal with an increase in number of particles, the calculation would need to get more and more resource-consuming. As a result, traditional EVB is only suitable for very simple systems.

1.3.3 Chemical Dynamics

A further step can consist of solving the Schrödinger equation with the total molecular Hamiltonian in order to study the motion of molecules. Direct solution of the Schrödinger equation, the classical mechanics framework is called quantum molecular dynamics, semiclassical molecular dynamics and molecular dynamics (MD) respectively. Statistical approaches, for example Monte Carlo methods, are also possible.

Molecular dynamics solves Newton's laws of motion for atoms on a potential energy surface; it can locate the minimum energy conformations but it is temperature dependent. While using MD, some problems are usually encountered, time and system size limitation, quantum effects when the temperature is sufficiently low.

Monte Carlo simulations require less computer time to execute each iteration than a molecular dynamics simulation using the same system. However, Monte Carlo simulations are more limited in that they cannot yield time-dependent information, such as diffusion coefficients and viscosity. The accuracy of the results is also very dependent on the sampling or biased sampling.

1.3.4 Ab initio Methods

The term "Ab initio" is Latin for "from the beginning". This implies that no or few assumptions are made, and that the method is 'pure' from a theoretical standpoint.

The name is given to computations that are derived directly from theoretical principles with no inclusion of experimental data. It computes solutions to the Schrödinger equation using a series of rigorous mathematical approximations.

Ab initio calculations give very good qualitative results and can yield increasingly accurate quantitative results as the molecules in question get smaller. The advantage of ab initio methods is that they eventually converge to the exact solution once all the approximations made are sufficiently small in magnitude.

However, ab initio methods are much more computationally expensive. These methods often take enormous amount of computer CPU time, memory, and disk space. The HF method scales as N^3, where N is the number of basis functions. This means that a calculation takes 16 (2^4) times more to complete. Correlated calculations often scale much worse than this. In practice, extremely accurate solutions are only obtainable when the molecule contains a dozen electrons or less. However, results with an accuracy rivaling that of many experimental techniques can be obtained for only moderate-size organic molecules.

1.3.5 Modified Methods

Currently, no single method can solve all problems in chemistry perfectly; Ab initio methods can do prediction very accurate on the cost of too much CPU time, so some researchers started to explore new methods, while others tried to combine different current methods' advantages, to develop hybrid methods. Today popular modified methods are linear scaling approach, compound method and QM/MM method.

a)**The linear scaling approach** Various linear-scaling methods have been developed since 1991. These methods are based on the locality of the electron correlation. Recently,

several linear scaling approaches have been introduced which replace the time dominating diagonalization step in semiempirical methods, enabling practical calculations to be performed on both small and big systems. It changes the way quantum calculations are being done. Numerous examples of calculations on systems with more than 1000 atoms are carried out.[5,6] However, while this new methodology seems promising, the CPU time involved in today's calculations is rather expensive and only allows for single energy point calculation. Some improvements, both in linear scaling algorithms and computing power, are still needed to make an addressful advance for linear scaling method.

b)Compound methods The G1, G2, and G3 methods modeled by Pople and co-workers calculate energies in cells of their matrix, then project more accurately.

Petersson's CBS methods (CBS-Q, CBS-QB3, etc.) are compound methods that give impressively accurate results by extrapolating basis-set effects to infinite limits.

c) QM/MM method To study macromolecular processes such as enzymatic reactions, it is necessary to use a chemical model that is capable of describing the forming and breaking of chemical bonds and is also suitable for capturing the complexity of the system. A fully quantum mechanical treatment of the entire enzyme system, in principle, satisfies these criteria, and quantum mechanical algorithms designed to scale linearly with system size have been developed and applied to protein systems in energy calculations.[7-11] hough this approach has many attractive features, it is very expensive, and thus limits its application in biological problems. Luckily, in most enzymatic reactions it is not necessary to treat the electronic structure of the entire enzyme-solvent system quantum mechanically. QM/MM comprises of two methods. QM is used to resolve the most essential parts of the system, and MM is used to non-reactive parts. That's why this method is a good choice to investigate huge systems. Nowadays, this methodology is the

most commonly used to handle the reactivity of biochemical systems. Warshel[12]as used this method to study chemical reactions in enzymes and solutions. Other biology researchers like Monard, G and co workers[13]vitt, M[14]se QM/MM to study enzymes, proteins and solutions.

The main advantage of QM/MM method is its easy implementation in computational codes while giving good chemical results. Its main disadvantage, especially in enzymatic systems, is to go beyond qualitative results and, thus, obtain quantitative numbers out of QM/MM computations.

Spatial extrapolation such as embedded-atom models of catalysts and Morokuma's ONIOM method: They connect or extrapolate domains of different-level calculations, but kind of hard to determine the area for different level.

If consider plotting a potential energy surface for one system, Collins developed a new scheme, interpolation scheme to make the whole calculation faster than pure ab initio approach.

1.4 General Introduction of the Collins' Interpolation Scheme

Many important chemical reactions occur for molecules in a single electronic state; that is where the wavefunction for the electrons is given by a single eigenfunction of the time-independent Schrödinger equation with the position of the nuclei fixed in space. The total electronic energy is the eigenvalue. This energy depends on the position of the nuclei. The value of the PES at any molecular configuration can now be evaluated with adequate accuracy for many small to medium-sized molecules using the method of ab initio quantum chemistry. However, the cost of such calculations can be quite high and the number of configurations at which the energy must be evaluated is very large for a

molecule undergoing chemical reaction. Pessimistically, this number is of the order of d(3N-6), where d is the number of configurations needed for each configurational degree of freedom.

Recently, a substantial progress has been made in evaluating PES by interpolation of ab initio data. For triatomic molecules, Ho, Rabitz and coworkers have constructed very accurate PES by using a "reproducing kernel Hilbert space" method.[15-18] Collins group has pursued a modified form of Shepard interpolation and applied this successfully to reactions involving several atoms.[19-24] In order to construct a PES in this way, an accurate interpolation method and an efficient method for deciding where in configuration space the (inevitably) limited number of ab initio calculations are performed.

Collins's interpolation method employs classical trajectory calculations of the reaction dynamics in the PES construction process, and the resultant surface is consequently particularly appropriate for such trajectory studies of the dynamics. The PES obtained could be used in quantum dynamical studies or in statistical reaction rate theories.

The algorithm presented here does not assume a functional form for the global PES,[25] nor is it a numerical "surface fitting method.[26] The PES is given by a moving interpolant[27,28] which represents the PES exactly at all the configurations where data was evaluated. At all other configurations, the PES is constructed from local Taylor series expansions which take advantage of the energy derivatives. As we shall see, the algorithm can be expected to converge more rapidly when higher order of derivatives are available.

The Collins PES interpolation scheme expresses the PES as:

$$V(\mathbf{z}) = \sum_{i=1}^{N} w_i(\mathbf{z}) T_i(\mathbf{z})$$

where z is the coordinates describing a molecular structure. $T_i(z)$, is a Taylor series expansion about point, i, in configuration space. It is truncated after second order, so the energy, first and second derivatives of the potential are needed at each point i. $w_i(z)$ is the weight given to the energy estimate made by Taylor series T_i for the geometry z.

1.5 Objective of the Thesis

Predicting energy and other related properties of molecule accurately within in a short time period is a rigorous task in computational chemistry. The present approaches require calculation of this energy at each node in a very large grid of molecular configurations. The numerical implementation of classical reaction dynamics requires the gradient of the energy with respect to the nuclear positions for a very large number of molecular configurations. The direct determination of these energies and/or energy gradients by ab initio calculations is an extremely expensive task, and has only been applied to small molecules or with relatively low level ab initio methods. During the past two decades, there have been dramatic improvements in both the accuracy and efficiency of high-level electronic structure calculations.[29-32] These advances, along with the increasing speed of modern computers, have made possible very high-quality ab initio calculations for small polyatomic systems.[33,34] For three- and four-atom systems, calculations with errors less than 4kJ/mol are feasible. Gradients and Hessians are also becoming widely available. In my research work, I would like to focus on improving the present methods, developing new algorithms to deal with different problems such as structures, energies, frequencies and other related molecular properties, within different chemical systems. According to Collins and co-workers' study results,[35-38] interpolation between ab initio data points, provides a method for obtaining a representation of the

global PES with high accuracy. Based on the PES by Collins' interpolation method over a number of ab initio calculations which is in the orders of magnitude smaller than that required for the direct approach, quantum observables—vibrational averaged internal coordinates, fully anharmonic zero-point energies and nuclear radial distribution functions can be calculated by combining the quantum Diffusion Monte Carlo (DMC).The advantages of this PES are twofold. Firstly, the exact ground state energy and properties can be calculated by performing DMC calculations directly on the analytic surface. Secondly, the calculations of the energies, first and second derivatives required for the interpolation are trivial. The computational effort require for DMC calculations on the interpolated potential energy surfaces is therefore determined only by the DMC convergence properties, allowing different possible regimes in the interpolation to be examined with relatively little computational expense. These PES-building regimes with high quality ab initio calculations are able to construct very accurate PES for either loosely bound complexes or condensed systems. A fragmentation method will be developed to predict energy accurately and efficiently for organometallic molecules.

We want to develop new methods by optimizing and combining current methods and applying them in my studies. We evaluated the current methods by examining the applications of the semi empirical method, diffusion Monte Carlo (DMC) (a method belonging linear scaling approach) and ab initio method on different cases. In the fragmentation scheme, accurate energy of a molecule can be computed by fragmenting the molecule and taking a linear combination of the resulting fragment subunits. This also allows for different levels of fragmentation, where higher levels fragmentation involving larger fragments lead to more accurate total energy calculations. It is the first time that charged compounds have been studied using fragmentation techniques. This would be a

major revolution in the field of quantum chemistry. Successively, an attempt will be made

to calculate the total electronic potential energy of a very large organometallic molecule.

An ab initio calculation of such a molecule is not possible using standard techniques. This

would be a major breakthrough in the areas of drug chemistry and chemical biology.

Composite G3X(MP2) [39-41] methods were originally developed to estimate energies of

molecules at very high levels of ab initio theory such as CCSD(T) and QCISD(T) by

performing a series of lower level calculations This optimized composite CCSD(T)

approach is further tested and an estimation of the potential CPU time-savings may be

obtained. It is envisaged that the results from this study should provide a clear general

indication of the applicability of composite method for calculating fundamental

frequencies and would contribute towards an alternative procedure for predicting highly

accurate energies of larger molecular system with significant reductions in computational

cost.

1.6 Scope

In the following chapter, Chapter 2, is the summary of the theoretical theory and

methods which are being used through out my whole work mentioned in the thesis.

In Chapter 3, an eight atom molecule CH_2OHCHO is thoroughly studied.

Using ab initio Density Function theory at B3LYP/6-31g** method, the potential

energy surface of hydroxyacetaldehyde has been mapped and all the critical points

identified.   A total of four unique minima, and the transition states connecting them,

were located.   In order to identify the observable conformers' two dimensions (2D)

and 18 dimensions (18D) quantum diffusion Monte Carlo (DMC) simulations were conducted.

The rotational constants for 2D were predicted from the simulations and were found to be in excellent agreement with experimental results for the observed conformer in the gas phase. It was predicted that among the remaining three minima only one conformer is effectively observable.   The predicted rotational constants for this conformer are provided.

In Chapter 4, a stilbene-cyclodextrin rotaxane system is studied by the QM semi empirical AM1 method. This huge complex system is a nano molecular machine with more than 150 atoms. The main force dominates in this system is intermolecular hydrogen bonding. We studied its conformation, energy, geometries, and movement. Finally it's possible that the mechanism is provided and explained.

Chapter 5 describes the composite methods. An extensive study of the harmonic frequencies of a large set of small polyatomic closed-shell molecules computed at both full ab initio and composite approximations of coupled cluster CCSD(T) method combined with augmented Dunning's basis sets is presented here. Using various combinations of basis sets, composite methods are capable of predicting full ab initio CCSD(T) level harmonic frequencies to within 5 cm^{-1} on average, which suggests a computationally affordable means of obtaining highly accurate vibrational frequencies compared to the CCSD(T) level. A general approach for calculating the composite level equilibrium geometries and harmonic frequencies (including the bends and torsions motions) for 19 tri- and 18 tetra non linear molecules that uses the Collin's method of interpolating potential energy surfaces is also described here. This approach is further

tested on tetrafluoromethane, and an estimation of the potential CPU time-savings that may be obtained is also presented. It is envisaged that the findings here will enable theoretical studies of fundamental frequencies and energetics of significantly larger molecular systems.

In Chapter 6, the fragmentation code is programmed; we apply this new algorithm to first row organometallic compounds. Such compounds necessarily require a charged metal centre. Octahedral compounds and tetrahedral compounds with closed and non closed shell are studied respectively. For the same transition metal (centre atom in all molecules), different oxidation states are also being studied. Once the studying samples are decided, we fragmented them in a hierarchied way at L1, L2, L3. For L1, there're two subunits, one without point charge and another with point charge in the fragments. Finally we compared these results at different levels with full ab initio results.

1.7 Reference:

1   Liu J-J, Ding Y-H., Feng, J., Sun, C-C., J. Phys.Chem. A. 2002, 106(11), 2695-2706.

2   Liu J-J, Ding Y-H., Feng, J-k., Tao, Y-G., Sun, C-C., J. Phys.Chem. A. 2002, 106(9), 1746-1764.

3   Liu J-J , Ding Y-H., Tao Y-G.,Feng J-K., Sun C-C., J Comput Chem 2002, 23(11), 1031-44.

4   Liu J-J., Ding Y-H., Tao Y-G., Feng J-K., Sun C-C., J Comput Chem 2002 , 23(6), 625-49.

5   Goedecker, S., Rev Mod Phys.1999,71,1085

6   Van der Vaart, A., Gogonea, V.,Dixon, S. L., Kenneth, M.,Merz, J. J Comput Chem. 2000, 21, 1494.

7   Van der Vaart A., Merz KM Jr., J. Am. Chem. Soc. 1999. 121:9182–90

8   York DM., Lee T-S., Yang WT., Phys. Rev. Lett. 1998. 80:5011–14

9   Van der Vaart A., Merz KM Jr., J. Phys. Chem. A. 1999.103:3321–29

10   Dixon SL, Merz KM Jr., J. Chem. Phys. 1996. 104:6643–49

11   Nadig G, Van Zant L.C., Dixon SL., Merz KM Jr., 1999. ACS Symp. Ser. 721,439–47

12   Warshel, A., In: Computer Modeling of Chemical Reactions in Enzymes and Solutions; Wiley & Sons, New York, 1992.

13   Monard, G., Merz, K. M., Acc Chem Res 1999, 32, 904.

14   Levitt, M., J Mol Biol 1976, 103, 227

15   Hollebeek T., Ho,T.-S., Rabitz, H., Harding, L. B., J. Chem. Phys. 2001,114, 3945

16   Hollebeek T., Ho,T.-S., Rabitz, H., J. Chem. Phys.1997, 106, 7223 .

17   Schatz ,G. C., Papaioannou ,A., Pederson,L. A., Harding, L. B., Hollebeek, T., Ho,T.-S., Rabitz, H., J. Chem. Phys. 1997,107, 2340 .

18   Ho,T. , Rabitz, H. , J. Chem. Phys. 1996,104, 2584 .

19   Collins, M. A., Bettens, R. P. A., Physical Chemistry Chemical Physics 1999, 1, (6), 939-945.

20   Collins, M. A., Petrie, S., Chalk, A. J., Radom, L., J. Chem. Phys. 2000, 112, (15), 6625-6634.

21   Bettens, R. P. A., Collins, M. A., Jordan, M. J. T., Zhang, D. H., J. Chem. Phys. 2000, 112, (23), 10162-10172.

22   Collins, M. A., Radom, L., J. Chem. Phys. 2003, 118, 6222-6229.

23   Crespos, C., Collins, M. A., Pijper, E., Kroes, G. J., J. Chem. Phys. 2004, 120, 2392-2404.

24   Moyano, G. E., Collins, M. A., Theo. Chem. Acc.s 2005, 113, (4), 225-232.

25   Murrell,J N, Carter ,S., Farantos S. C., Huxley, E, and Varandas ,A.. C., Molecular Potential Energy Functions (Wily, Chichester, 1984).

26   P G Jasien and R. Shepard, Int. J. Quantum Chem. 1988, 22, 183.

27   R. Farwig, in Algorithms For Approximation, edited by J. C. &son and . G. Cox (Clarendon, Oxford, 1987), p. 194.

28   Lancaster,P., Salkauskas,K.,Curve,Surje Fitting, An Introduction(Academic, London, 1986), Chap. 10.

29 Hehre,WJ., Radom L, Schleyer P, Pople J.A., 1996. Ab Initio Molecular Orbital Theory. New York: Wiley

30 Head-Gordon M., J. Phys. Chem. 1996. 100:13213–25

31 Yamaguchi Y, Osamura Y, Goddard JD, Schaefer HF III. 1994. A New Dimension to Quantum Chemistry. New York-Oxford: Oxford Univ.

32 Kohn W, Becke AD, Parr R.G., J. Phys. Chem. 1996. 100,12974–80

33 Jeziorski B, Moszynski R, Szalewicz K., Chem. Rev. 1994.   94,1887–930

34 Dunning TH Jr., ed. 1990. Advances in Molecular Electronic Structure Theory, Vol. 1. London

35 Jordan M. J. T, Thompson,K. C., Collins, M. A., J. Chem. Phys. 1995,102, 5647 .

36 Bettens,R. P. A., Collins ,M. A., J. Chem. Phys.1999, 111, 816.

37 Collins, M. A., Theor. Chem. Acc.2002, 108, 313.

38 Thompson,K. C., Jordan M. J. T., Collins ,M. A., J. Chem. Phys. 1998,108, 8302 .

39 Curtiss, L.A, Raghavachari,K, Redfern,P.C., Rassolov,V, Pople, J.A., J.Chem.Phys. 1998, 7746,109

40 Curtiss, L.A,Redfern,P.C,Raghavachari, K.Rassolov, V,Pople, J. A, J.Chem.Phys. 1999, 4703, 109

41 Curtiss, L.A,Redfern,P.C,Raghavachari, K, Pople, J.A.,J.Chem.Phys.2001,114,108

1.8 Appendix
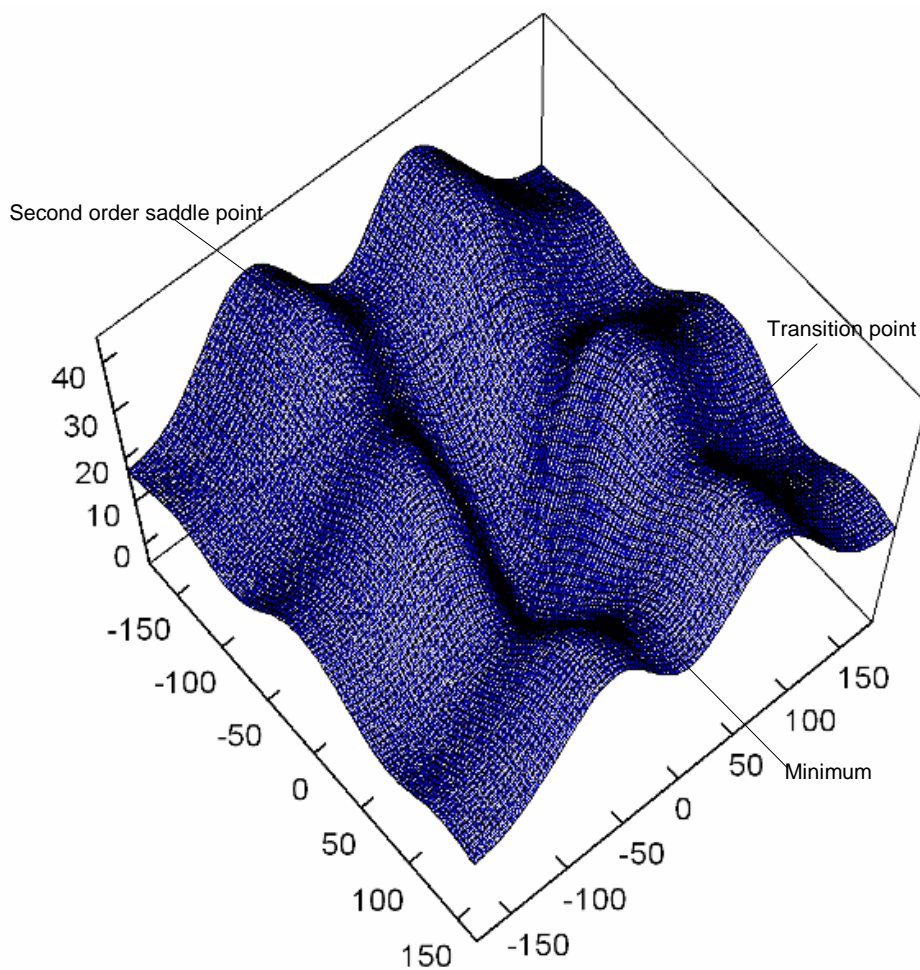
Figure 1.1 Points on a potential energy surface.

Chapter 2 Theoretical Methodology

2.1 Introduction

Quantum chemistry is a branch of theoretical chemistry, which applies quantum mechanics and quantum field theory to address issues and problems in chemistry. The description of the electronic behavior of atoms and molecules as pertaining to their reactivity is one of the applications of quantum chemistry. Quantum chemistry lies on the border between chemistry and physics, and significant contributions have been made by scientists from both fields.

The first step in solving a quantum chemical problem is usually solving the Schrödinger equation.

2.2 Schrödinger Equation

In 1925, Erwin Schrödinger and Werner Heisenberg independently developed the new quantum theory. Schrödinger's method involves partial differential equations, whereas Heisenberg's method employs matrices; however, a year later the two methods were shown to be mathematically equivalent. The Heisenberg wrote the Schrödinger equation as such

$$H|\Psi\rangle = E|\Psi\rangle \qquad (2.1)$$

where here H is the Hamiltonian operator[1] for a system consisting of nuclei and electrons, Ψ is the wavefunction known as the eigenfunction and E is the energy of the system known as the eigenvalue. The Hamiltonian operator is a sum of the kinetic (T) and potential (V) energy operators of the system.

$$H = V + T \tag{2.2}$$

For a molecule, it is reasonable to split the kinetic energy into two summations: one over electrons, and one over nuclei. Similarly, we can split the potential energy into terms representing interactions between nuclei, between electrons, or between electrons and nuclei. Using i and j to index electrons, and A and B to index nuclei, we have (in atomic units)

$$H = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2.3}$$

where $r_{ij} = |r_i - r_j|$, $R_{Ai} = |r_A - r_i|$, and $R_{AB} = |r_A - r_B|$.

The Laplacian operators ∇_i^2 and ∇_A^2 involve differentiation with respect to the coordinates of the ith electron and the Ath nucleus.

$$\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2} \tag{2.4}$$

The first term in Eq. 2.3 is the operator for the kinetic energy of the electrons; the second term is the operator for the kinetic energy of the nuclei; the third term represents the Coulomb attraction between electrons and nuclei; the fourth and fifth terms represent the repulsion between electrons and between nuclei, respectively.

2.3 Approximations Used to Solve the Schrödinger Equation

In fact it is impossible to obtain an exact solution to the Schrödinger equation for any system except for hydrogen atom or H_2^+ fixed nuclear coordinates, yet it is possible to obtain a fairly good approximate solution for a variety of systems. Therefore a number of approximations are incorporated to solve the Schrödinger equation. They are as follows:

1. The Neglect of Relativistic Effects

2. The Born-Oppenheimer Approximation

3. The One-Electron Approximation

4. The Linear Combination of Atomic Orbital (LCAO) Approximation

5. The Time Independence Approximation

2.3.1 The Neglect of Relativistic Effects[2]

The first and foremost approximation which is introduced while solving the Schrödinger equation is the neglect of relativistic effects. Though there is no direct method to measure relativity yet, there is an easily measurable property called the *spin-orbit coupling* which can account at least partially for the effects of relativity on the energy of a system under consideration. One should bear in mind that spin-orbit coupling is just one term in the relativistic energy expression. It arises due to the interaction between the spin moment of the electron and the magnetic moment of the orbital due to its motion.

Now let us consider the consequences of neglecting the relativistic effects. Solving the Dirac equation for the H-atom gives the following results. The magnitude and the z-component of the angular momentum can be expressed as

$$L = \hbar\sqrt{j(j+1)} \tag{2.5}$$

$$L_Z = \hbar m_j, \tag{2.6}$$

where the values of j are different from those of the nonrelativistic l:

$$j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, ...$$

while the relationship between j and m_j is the same as that between l and m:

$$m_j = j, j-1, ..., -j .$$

Therefore, the energy depends on both n and j on using Dirac's theory. If, for example, $n = 1$, j has only one possible value, $j = \dfrac{1}{2}$, which is doubly degenerate according to $m_j = \pm(1/2)$. If the electron is excited to the $n = 2$ level of the hydrogen atom, this would give rise to two states, $^2P_{1/2}$ and $^2P_{3/2}$ for $j = 1/2$ and $3/2$, respectively. The first state is doubly degenerate and the second one is quadruply degenerate and so on. Due to this tiny splitting of the spectral lines in the atomic spectrum, the spin-orbit coupling is introduced in order to account for the interaction between the two different moments.

In the relativistic case it is not necessary to introduce the spin-orbit coupling because it is explicitly introduced in the theory. In the non-relativistic case the spin-orbit interaction can be expressed by a term in the Hamiltonian containing the product of the two operators, $\hat{L} \bullet \hat{S}$. The omission of this term makes it possible to separate the Hamiltonian from the spin operator, which means that the energy of the system will be independent of spin. Another consequence of this approximation is the appearance of l and s quantum numbers instead of j.

The separation of the two states, $^2P_{1/2}$ and $^2P_{3/2}$, in the case of the H-atom can be measured by optical spectroscopy and about 0.4 cm^{-1} is neglected when the spin-orbit term is omitted from the Hamiltonian. The value of 0.4 cm^{-1} can be compared, for example, with the $1s \rightarrow 2p$ transition energy in the atomic hydrogen spectrum of 82260 cm^{-1}; it is negligibly small for most purposes. However, the importance of relativistic effect increases for heavier elements. The spin-orbit splitting from the photoelectron spectra of argon, krypton and xenon can be measured as about 1436 cm^{-1}, 5364 cm^{-1} and 10534 cm^{-1}, respectively. These values are considerably

larger. Therefore, relativistic effects can be safely ignored for lighter elements (in the atoms H-Ne) for most purposes but have to be taken into account for heavier elements.

2.3.2 The Born-Oppenheimer Approximation[2]

The Born-Oppenheimer (BO) approximation plays a most important role in quantum chemistry. According to this approximation, one can consider the electrons in a molecule to be moving in a field of fixed nuclei since the nuclei are much heavier than the electrons. Therefore, Ψ can be approximated as a product of electronic and nuclear wavefunctions.

$$\Psi = \Psi_{elec}\Psi_{nucl} \qquad (2.7)$$

The electronic wavefunction, Ψ_{elec} can be obtained by assuming the electrons to be moving in a field of fixed nuclei and the nuclear wavefunction, Ψ_{nucl} can be obtained by assuming the nuclei to be moving in an average electronic field.

Upon applying the Born-Oppenheimer approximation to Eq. 2.3 the second term representing the kinetic energy of the nuclei can be removed from consideration of the electronic energy and the fifth term representing the repulsion between the nuclei becomes a constant. Any constant added to an operator adds only to the operator eigenvalues but has no effect on the operator eigenfunctions. Therefore Eq. 2.3 becomes

$$H_{elec} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}, \qquad (2.8)$$

where H_{elec} is known as the electronic Hamiltonian, i.e. Hamiltonian describing the motion of N electrons in a field of M point charges. Solution of the electronic Schrödinger equation,

$$H_{elec}\Psi_{elec} = E_{elec}\Psi_{elec}, \qquad (2.9)$$

gives the electronic wavefunction, Ψ_{elec} and the electronic energy, E_{elec}. The electronic wavefunction,

$$\Psi_{elec} = \Psi_{elec}\left(\{r_i\};\{R_A\}\right), \tag{2.10}$$

describes the motion of the electrons or represents the molecular orbitals and the electronic energy,

$$E_{elec} = E_{elec}\left(\{R_A\}\right), \tag{2.11}$$

represents the energies of the molecular orbitals. The electronic wavefunction and electronic energy obtained by solving the electronic Schrödinger equation depends explicitly on the electronic coordinates and depends parametrically on the nuclear coordinates. Parametric dependence means that, for different arrangements of the nuclei, Ψ_{elec} is a different function of the electronic coordinates. The total energy of a system with fixed nuclei is given by

$$E_{tot} = E_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}. \tag{2.12}$$

Eqs. 2.8 to 2.12 constitute the electronic problem. If one has solved the electronic problem, it is possible to solve for the motion of nuclei as well by using the same assumption as that used to solve the electronic problem. Since the electrons move much faster than the nuclei, it is a reasonable approximation to replace the electronic coordinates in Eq. 2.3 by their average values, averaged over the electronic wavefunction. This then generates a nuclear Hamiltonian (H_{nucl}) for the motion of the nuclei in an average electronic field.

$$H_{nucl} = -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 + \left\langle -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \right\rangle + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}$$

$$= -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 + E_{elec}\left(\{R_A\}\right) + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}$$

$$= -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 + E_{tot}\left(\{R_A\}\right) \tag{2.13}$$

The total energy $E_{tot}\left(\{R_A\}\right)$ provides a potential for the nuclear motion. Therefore, the nuclei in the Born-Oppenheimer approximation moves on a potential energy surface obtained by solving the electronic problem. Solving the nuclear Schrödinger equation,

$$H_{nucl}\Psi_{nucl} = E\Psi_{nucl}, \tag{2.14}$$

gives the nuclear wavefunction Ψ_{nucl} which describes the rotation, vibration and translation of a molecule and the energy E which is a sum of the rotational, vibrational and translational energy of a molecule.

2.3.3 The One-Electron Approximation

Applying the Born-Oppenheimer approximation to the Schrödinger equation helps to split this complex Schrödinger equation into two parts, namely the electronic (Eq. 2.9) and nuclear (Eq. 2.14) Schrödinger equations. Now let us consider solving the electronic Schrödinger equation. The electronic wavefunction, Ψ_{elec}, is a function of the spatial coordinates of all the n electrons and it would be easier to solve the electronic Schrödinger equation if we can approximate Ψ_{elec} as a product of n one-electron wavefunctions:

$$\Psi_{elec}(1,2,...,n) = \Psi_1(1)\Psi_2(2)...\Psi_n(n), \tag{2.15}$$

where $\Psi_i(i)$ is a function of only the three coordinates of the ith electron. In order to achieve this, we have to express the Hamiltonian operator as a sum of one-electron operators. The Hamiltonian can be written as a function of zero, one and two electron terms.[2]

$$H_0 = e^2 \sum_{\alpha < \beta} \sum_{\alpha < \beta} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \tag{2.16}$$

$$H_1 = \sum_{i=1}^{n} \left(\frac{\hbar^2}{2m} \nabla_i^2 - \sum_\alpha \frac{e^2 Z_\alpha}{r_{\alpha i}} \right) = \sum_{i=1}^{n} h_i(i) \tag{2.17}$$

$$H_2 = \sum_{i<j} \sum_{i<j} \frac{e^2}{r_{ij}} \tag{2.18}$$

H_0, H_1 and H_2 correspond to the Hamiltonians which are function of zero, one and two electrons, respectively. H_0 is a constant since the nuclei are considered to be stationary and H_1 presents no obstacle to the separation of variables since it is a function of one-electron terms. It is the H_2 operator which causes a problem in separating the Hamiltonian into a sum of one-electron operators. We can simplify our problem by simply ignoring the H_2 operator. For example, let us consider a three electron system and construct its Schrödinger equation using a product wavefunction.

$$H\Psi(1,2,3) = E\Psi(1,2,3) \tag{2.19}$$

$$[h_1(1) + h_1(2) + h_1(3)]\phi_1(1)\phi_2(2)\phi_3(3) = (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)\phi_1(1)\phi_2(2)\phi_3(3)$$

On dividing the above equation by $\phi_1(1)\phi_2(2)\phi_3(3)$ we get

$$\frac{1}{\phi_1(1)} h_1(1)\phi_1(1) + \frac{1}{\phi_2(2)} h_1(2)\phi_2(2) + \frac{1}{\phi_3(3)} h_1(3)\phi_3(3) = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \tag{2.20}$$

We find that the original equation splits into three independent one-electron Schrödinger equations:

$$h_1\phi_1 = \varepsilon_1\phi_1 \tag{2.21}$$

$$h_1\phi_2 = \varepsilon_2\phi_2 \tag{2.22}$$

$$h_1\phi_3 = \varepsilon_3\phi_3 \tag{2.23}$$

Since h_1 is the same in all the three equations, we just have to solve only one equation. Therefore we find that it is quite simple to solve the Schrödinger equation by

neglecting the two electron terms. However, the two electron terms are so important in the molecular energy expression that their omission would lead to unreliable results. Therefore we should try to separate the Hamiltonian by taking the two electron terms into consideration. While considering the two electron terms, it should be borne in mind that the resulting total wavefunction satisfies the Pauli principle of antisymmetry with respect to the exchange of electrons. If the total wavefunction is a product of n wavefunctions, we find that it does not satisfy the Pauli principle. To exemplify the problem let us consider a two electron system and the product wavefunction of which would be

$$\Psi_1(1)\Psi_2(2)$$

The above product wavefunction is surely not antisymmetric. However, an antisymmetric linear combination of the above wavefunction

$$N\left[\Psi_1(1)\Psi_2(2) - \Psi_1(2)\Psi_2(1)\right]$$

is antisymmetric (N is a normalization constant) with respect to the exchange of two electrons. This wavefunction includes only the spatial coordinates of the electrons. It is necessary to include the spin coordinates as well. Therefore the one-electron wavefunction can be written as a product of one-electron orbital Ψ and one-electron spin η functions:

$$\phi_i = \Psi_i \eta_i \tag{2.24}$$

Rewriting the above antisymmetric wavefunction by including the spin-orbit functions:

$$N\left[\phi_1(1)\phi_2(2) - \phi_1(2)\phi_2(1)\right]$$

Guessing an antisymmetric wavefunction for a many-electron system is not as simple as that for a two electron system discussed above. However, it is very

straightforward to obtain an antisymmetric wavefunction for any system by writing the complete spin-orbital wavefunction in the form of a determinant. For two electrons:

$$N \begin{vmatrix} \phi_1(1) & \phi_1(2) \\ \phi_2(1) & \phi_2(2) \end{vmatrix} = N \left[\phi_1(1)\phi_2(2) - \phi_1(2)\phi_2(1) \right] \tag{2.25}$$

and for *n* electrons (including the normalization constant):

$$\Phi \equiv \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(1) & \cdots & \phi_1(n) \\ \vdots & & \vdots \\ \phi_n(1) & \cdots & \phi_n(n) \end{vmatrix} \tag{2.26}$$

This determinant is known as a *Slater determinant*. Interchanging two electrons leads to the exchange of two rows in the determinant, which changes sign as a consequence. If two one-electron functions of a given system happen to be the same then two rows in the Slater determinant will be identical and hence the determinant value will become equal to zero. This is in fact the mathematical consequence of Pauli's exclusion principle according to which no two electrons can have the same set of quantum numbers. Now if we pair up the electrons having the same orbital wavefunction but differ only in the spin wavefunction, then the number of functions in the Slater determinant (Eq. 2.18) reduces from *n* to *n/2*.

$$\Phi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \Psi_1(1)\alpha(1) & \Psi_1(2)\alpha(2) & \cdots & \Psi_1(n)\alpha(n) \\ \Psi_1(1)\beta(1) & \Psi_1(2)\beta(2) & \cdots & \Psi_1(n)\beta(n) \\ \vdots & & & \vdots \\ \Psi_{n/2}(1)\beta(1) & \Psi_{n/2}(2)\beta(2) & \cdots & \Psi_{n/2}(n)\beta(n) \end{vmatrix} \tag{2.27}$$

So far we have seen that the neglect of two-electron term simplifies the Schrödinger equation and at the same time it also leads to unreliable results. Therefore inclusion of the two-electron term in the Hamiltonian is essential in order to obtain moderately good results. Now that we have taken the two-electron term into consideration, let us assume that each electron moves in a field of all other electrons, i.e., each electron experiences an average field of all other electrons. This would mean

that each electron is formally independent of all the other electrons. This is known as the *independent particle model*. This model behaves computationally as a one-electron model, even though in practice the effective field depends on all the electrons; i.e., for the calculation of the effective potential, we should know the states of all the electrons. Due to this interdependence, the equation has to be solved by an iterative procedure. The quality of the model depends on how well the effective one-electron potential approximates the real two-electron potential:

$$\sum_{i<j}^{n}\sum_{i<j}^{n}\frac{e^2}{r_{ij}} \rightarrow \sum_{i}^{n}V_1^{eff}(i),$$

where $V_1^{eff}(i)$ depends on all the electrons except the ith electron. Now the Hamiltonian can be expressed as

$$H \cong \sum_{i=1}^{n}\left[h_1(i)+V_1^{eff}(i)\right] = \sum_{i=1}^{n}F(i) \qquad (2.28)$$

and

$$F(i)\phi_i = \varepsilon_i\phi_i \qquad (2.29)$$

The one-electron wavefunction ϕ_i can be used to construct the many-electron determinant wavefunction Φ and the energy ε_i can be used to determine the energy E, of the system.

2.3.4 The Linear Combination of Atomic Orbital (LCAO) Approximation

In the LCAO approximation, a molecular orbital can be constructed by a linear combination of one-electron basis functions, usually called the atomic orbitals (AOs), which are normally centered on each nucleus,

$$\Psi = \sum_{\mu}^{AO}C_\mu^i\chi_\mu . \qquad (2.30)$$

C_μ^i is the coefficient of the μ th atomic orbital χ_μ in the ith MO.[3] The LCAO approximation provides us an efficient approach to obtain a trial linear variational function to describe the MOs in a molecule. The orbital coefficients are the variational parameters of the quantum mechanical calculations and their best values will give the optimum calculated energy.

2.3.5 The Time Independence Approximation

According to the time independence approximation, the Hamiltonian is considered to be independent of time. It does not depend explicitly on time and hence the corresponding wavefunction is a function of only the spatial coordinates and it corresponds to a stationary state of the system under consideration.

2.4 Approximate Methods Used to Solve the Schrödinger Equation

For most chemical problems, the Schrödinger equation is not strictly separable and the differential equation cannot be easily solved by analytic means. The techniques that are best used to find wavefunctions for complicated problems often turn out to be indirect, or at least they appear so. The *variation method* and *perturbation method* represent two alternative approaches to the problem of calculating approximate wavefunctions and energies of systems for which direct solution of the Schrödinger equation is difficult or impossible.[4]

2.4.1 The Variation Method

The variational principle is the basis for the variational determination of a wavefunction. The variational principle states that the expectation value or average value of the energy for an approximate wave function always lies above or equal to the exact solution of the Schrödinger equation for the same Hamiltonian operator[5]. This

means that if we have a wave function that contains adjustable parameters and we adjust them to minimize the expectation value of the energy, then we are approaching the exact result.

Assume the ground state energy[6] of the system be E_0 with the corresponding wavefunction Ψ_0, and let Ψ be an arbitrary function, the expectation value of energy with the trial wavefunction Ψ can be expressed as

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{2.31}$$

where the denominator is required for normalization. As stated above, according to the variation theorem for any Ψ it is true that $E_0 \leq E$ and equality holds only if $\Psi = k\Psi_0$ (and $|k| = 1$). In order to find the ground-state energy of the system, we have to minimize Eq. 2.31. To achieve this, we have to select a set of known basis functions $\{\phi_n\}$, express the trial wavefunction, Ψ, as a linear combination of these, and substitute the result into Eq. 2.31:

$$E = \frac{\left\langle \left(\sum_i c_i \phi_i \right) \middle| H \middle| \left(\sum_j c_j \phi_j \right) \right\rangle}{\left\langle \sum_i \sum_j c_i \phi_i \middle| c_j \phi_j \right\rangle} \tag{2.32}$$

On expanding the above equation, we get

$$E = \frac{\sum_i \sum_j c_i {}^* c_j \langle \phi_i | H | \phi_j \rangle}{\sum_i \sum_j c_i {}^* c_j \langle \phi_i | \phi_j \rangle} \tag{2.33}$$

where

$$\langle \phi_i | H | \phi_j \rangle = H_{ij} \quad \text{(Coulomb Integral)} \tag{2.34}$$

and

$$\langle \phi_i | \phi_j \rangle = S_{ij} \quad \text{(Exchange Integral)}. \tag{2.35}$$

By substituting Eqs. 2.34 and 2.35 into Eq. 2.33, we get

$$E = \frac{\sum_i \sum_j c_i * c_j H_{ij}}{\sum_i \sum_j c_i * c_j S_{ij}} \tag{2.36}$$

In order to determine the minimum value of E, we have to calculate the $\dfrac{\partial E}{\partial c_i}$ partial derivative for each c_i and set it equal to zero. This results in the following set of linear equations:

$$c_1 (H_{11} - S_{11} E) + c_2 (H_{12} - S_{12} E) + \cdots = 0$$

$$c_1 (H_{21} - S_{21} E) + c_2 (H_{22} - S_{22} E) + \cdots = 0 \tag{2.37}$$

or more briefly:

$$\sum_j c_j (H_{ij} - E S_{ij}) = 0 \quad (i = 1, 2, \ldots) \tag{2.38}$$

This equation can also be written in the matrix form:

$$[\mathbf{H} - E\mathbf{S}]\mathbf{c} = 0, \tag{2.39}$$

which has a trivial solution ($\mathbf{c} = 0$) and also a non-trivial solution $|\mathbf{H} - E\mathbf{S}| = 0$. This determinant is called the *secular determinant*. The solutions of the resulting

polynomial, E_0, E_1, E_2,…, E_k… are the energy eigenvalues of the system. Once the value of c is known, the trial wavefunction can be written as:

$$\Psi = \sum_{i=1}^{k} c_i \phi_i .$$
(2.40)

2.4.2 The Perturbation Method

Perturbation theory is the second most widely used approximation method in quantum chemistry. It allows one to estimate the splittings and shifts in energy levels and changes in wavefunctions that occur when an external field (e.g., an electric or magnetic field or a field that is due to a surrounding set of 'ligands'- a crystal field) or a field arising when a previously-ignored term in the Hamiltonian is applied to a species whose 'unperturbed' states are known. In perturbation theory, the Hamiltonian for any problem is partitioned into two or more parts. The first part is one for which the eigenfunctions and eigenvalues are known, while everything else represents the perturbation. This first part and the associated eigenfunctions and eigenvalues are distinguished in notation by a zero superscript. Assume the Schrödinger equation for the model system is written as

$$H^{(0)}\Psi^{(0)} = E^{(0)}\Psi^{(0)} ,$$
(2.41)

then the Schrödinger equation of the true system can be expressed as

$$H\Psi = E\Psi \ , \text{ where } \ H = H^{(0)} + \lambda V$$
(2.42)

Here V is the potential representing the difference between the two systems and λ is a dimensionless parameter, the *perturbation parameter*. If we can describe the true

system as a small perturbation of the model, Ψ and E will not be very different from $\Psi^{(0)}$ and $E^{(0)}$, and both can be expressed using powers of λ:

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \tag{2.43}$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \tag{2.44}$$

To simplify the mathematics, we choose the perturbed wavefunctions to be orthogonal to $\Psi^{(0)}$. By inserting the preceeding two equations into Eq. 2.42, we obtain

$$\left(H^{(0)} + \lambda V\right)\left(\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots\right) \\ = \left(E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots\right)\left(\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots\right) \tag{2.45}$$

Collecting the powers of λ, leads to:

$$H^{(0)}\Psi^{(0)} + \lambda\left[V\Psi^{(0)} + H^{(0)}\Psi^{(1)}\right] + \lambda^2\left[V\Psi^{(1)} + H^{(0)}\Psi^{(2)}\right] + \dots \\ = E^{(0)}\Psi^{(0)} + \lambda\left[E^{(1)}\Psi^{(0)} + E^{(0)}\Psi^{(1)}\right] + \lambda^2\left[E^2\Psi^{(0)} + E^{(1)}\Psi^{(1)} + E^{(0)}\Psi^{(2)}\right] + \dots \tag{2.46}$$

This equation can only be satisfied for an arbitrary value of λ if the coefficients for different powers of λ are the same:

$$H^{(0)}\Psi^{(0)} = E^{(0)}\Psi^{(0)}$$
$$V\Psi^{(0)} + H^{(0)}\Psi^{(1)} = E^{(1)}\Psi^{(0)} + E^{(0)}\Psi^{(1)}$$
$$V\Psi^{(1)} + H^{(0)}\Psi^{(2)} = E^{(2)}\Psi^{(0)} + E^{(1)}\Psi^{(1)} + E^{(0)}\Psi^{(2)} \tag{2.47}$$
$$\vdots$$
$$V\Psi^{(k-1)} + H^{(0)}\Psi^{(k)} = \sum_{i=0}^{k} E^{(k-i)}\Psi^{(i)}$$

Depending on the power of λ we truncate this expression. We refer to them as the first, second, ..., etc. order of perturbation. First-order perturbation theory delivers $\Psi^{(1)}$ and $E^{(1)}$, while we obtain $\Psi^{(2)}$ and $E^{(2)}$ at the second order.

Firstly the first-order equations was examined. By multiplying from the left with $\Psi^{(0)}$ and integrating over all space, we get

$$\left\langle \Psi^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle + \left\langle \Psi^{(0)} \middle| H^{(0)} \middle| \Psi^{(1)} \right\rangle = E^{(1)} \left\langle \Psi^{(0)} \middle| \Psi^{(0)} \right\rangle + E^{(0)} \left\langle \Psi^{(0)} \middle| \Psi^{(1)} \right\rangle . \qquad (2.48)$$

The second term of this equation can be written as

$$\left\langle \Psi^{(0)} \middle| H^{(0)} \middle| \Psi^{(1)} \right\rangle = \left\langle H^{(0)} \Psi^{(0)} \middle| \Psi^{(1)} \right\rangle = E^{(0)} \left\langle \Psi^{(0)} \middle| \Psi^{(1)} \right\rangle = 0 . \qquad (2.49)$$

The first term of the right-hand side of equation 2.48 is $E^{(1)}$, while the second term is zero due to the orthogonality condition. Therefore, the first-order energy correction is

$$E^{(1)} = \left\langle \Psi^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle . \qquad (2.50)$$

Since the perturbation operator V and $\Psi^{(0)}$ are known, $E^{(1)}$ can be calculated without having to determine the perturbed wavefunction. We of course, need this wavefunction to determine the higher-order energy corrections.

$$\begin{aligned} E^{(2)} &= \left\langle \Psi^{(0)} \middle| V \middle| \Psi^{(1)} \right\rangle \\ &\vdots \\ E^{(k)} &= \left\langle \Psi^{(0)} \middle| V \middle| \Psi^{(k-1)} \right\rangle \end{aligned} \qquad (2.51)$$

and the true energy of the system is

$$E = E^{(0)} + \sum_{i=1}^{\infty} \lambda^i \left\langle \Psi^{(0)} \middle| V \middle| \Psi^{(i-1)} \right\rangle . \qquad (2.52)$$

After a short mathematical skirmish (refer the appendix for derivation), $\Psi^{(1)}$ can be expressed as follows:

$$\Psi^{(1)} = \sum_i \frac{\left\langle \Psi_i^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle}{E^{(0)} - E_i^{(0)}} \Psi_i^{(0)} \ . \tag{2.53}$$

The conclusion from all of this is that if we only require the perturbed energy to the first-order, it is sufficient to know the unperturbed wavefunction of the given state. In contrast, if we are after the perturbed wavefunction, we need to know all the eigenfunctions of the unperturbed system. By inserting, Eq. 2.53, into the second-order energy expression (Eq. 2.51), we can calculate $E^{(2)}$. In a similar way we can go on to $\Psi^{(2)}$, and so on.

Inspecting the denominator of expression (Eq. 2.53), it appears that only states energetically close to $E^{(0)}$ contribute appreciably to the true energy of the system. On the other hand, we might argue that the sheer numbers of higher-lying states might affect the results. However, it can be proved that many of these higher states vanish from the numerator. To estimate the zero or nonzero value of an integral, e.g., $\left\langle \Psi_i^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle$, we recall our knowledge of group theory. The result of an integration of a function from $-\infty$ to $+\infty$ is zero if the function is antisymmetric and nonzero if it is symmetric. Therefore, if the product function $\Psi_i^{(0)} V \Psi^{(0)}$ is totally symmetric, the integration gives a nonzero result. In this case, the direct product $\Gamma^{(i)} \otimes \Gamma^{(pert)} \otimes \Gamma^{(0)}$ must contain the totally symmetric species. In other words, if the distortion picked up by the system has the same symmetry as the disturbing perturbation, the close-in energy state indeed plays a dominant role.

An important question that arises is whether the perturbation series (Eq. 2.43) converges. In most practical cases, it does converge well (although it cannot be taken for granted). An important shortcoming of the perturbation theory is that it is not

variational; i.e., perturbation theory does not provide an upper bound to the energy of the system. It is reasonable to expect, however, that by involving higher-energy terms, we improve the quality of the results.

2.5 The Hartree-Fock Method

Let us now return back to the one-electron approximation problem discussed in section 2.3.3. Now we have to find out a means of determining the best one-electron wavefunction, ϕ_i, in order to calculate the energy and other related properties of the wavefunction. For achieving this we have to start with writing the expectation value of the energy, based on Eqs. (2.16) - (2.18):

$$E = \left\langle \Phi | H | \Phi \right\rangle = \left\langle \Phi | H_0 | \Phi \right\rangle + \left\langle \Phi \left| \sum_{i=1}^{n} h_1(i) \right| \Phi \right\rangle + \left\langle \Phi \left| \sum_{i<j}^{n} \sum_{i<j}^{n} h_2(ij) \right| \Phi \right\rangle \quad (2.54)$$

where Φ is the Slater determinant of the system under study, $h_1(i)$ is the one-electron operator and $h_2(ij) = \left(e^2 / r_{ij} \right)$, the two-electron operator. We know that the determinant is simply the linear combination of product wavefunctions. Since the $h_1(i)$ operators affect only one function (the ith) from such a product and $h_2(ij)$ affects only two (the ith and j^{th}), a number of integrals will vanish from Eq. 2.54. After a long and strenuous derivation, we finally reach the following complex-looking expression of the total energy:

$$E = E_0 + \sum_{i=1}^{n} \left\langle \phi_i(i) | h_1(i) | \phi_i(i) \right\rangle$$

$$+ \sum_{i<j}^{n} \sum_{i<j}^{n} \left[\left\langle \phi_i(i)\phi_j(j) | h_2(ij) | \phi_i(i)\phi_j(j) \right\rangle - \left\langle \phi_i(i)\phi_j(j) | h_2(ij) | \phi_i(j)\phi_j(i) \right\rangle \right]$$

$$= E_0 + \sum_{i=1}^{n} \left\langle \phi_i(i) | h(i) | \phi_i(i) \right\rangle + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \phi_i(i)\phi_j(j) | h_2(ij)(1 - P_{ij}) | \phi_i(i)\phi_j(j) \right\rangle \quad (2.55)$$

We have replaced $\sum\limits_{i<j}^{n}\sum\limits_{i<j}^{n}$ term by $\dfrac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}$ in Eq. 2.55. P refers to a permutation of electron labels. This is because the first term of the double sum (the *Coulomb interaction*) and the second term (the *exchange interaction*) are equal and cancel out if $i=j$. This is why the $V_1^{eff}(i)$ potential indeed describes the interaction of an electron with the field of n-1 other electrons. For finding the minimum of the energy expression (refer the appendix for derivation) described by Eq. 2.55, we have to follow the conditions $\langle \phi_i | \phi_j \rangle = \partial_{ij}$ with the parameters ε_{ij}, and add them to the above equation. Following this procedure, we obtain the following much simpler equations:

$$F(i)\phi_i = \sum_j \varepsilon_{ij}\phi_j \qquad i=1,2,...,n \qquad (2.56)$$

where n is the number of electrons and the *Fock operator*, $F(i)$, has the following form:

$$F(i) = h_1(i) + \sum_{j=1}^{n}\left\langle \phi_j(j)\left|h_2(ij)\left(1-P_{ij}\right)\right|\phi_j(j)\right\rangle. \qquad (2.57)$$

The equations in (2.56) are called the *Hartree-Fock* (abbreviated as HF) *equations* (there are n equations). We can also write them in the matrix form:

$$F\phi = \phi\varepsilon = F\left(\phi_1\phi_2...\phi_n\right) = \left(\phi_1\phi_2...\phi_n\right)\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1n} \\ \varepsilon_{21} & \cdots & & \vdots \\ \vdots & & & \\ \varepsilon_{n1} & \cdots & & \varepsilon_{nn} \end{pmatrix} \qquad (2.58)$$

where ϕ is a row vector and ε is a square matrix. Since the latter is symmetric, it can be made diagonal using similarity transformation. Using the appropriate transformation matrix, **Q**, the HF equation can be written as:

$$F \underbrace{\phi \times \mathbf{Q}}_{\phi'} = \underbrace{\phi \mathbf{Q}}_{\phi'} \underbrace{\mathbf{Q}^{-1} \varepsilon \mathbf{Q}}_{\varepsilon'} = F\phi' = \phi'\varepsilon' \, , \qquad (2.59)$$

which has the component

$$F(i)\phi_i = \varepsilon_i \phi_i \qquad i = 1,2,...,n \qquad (2.60)$$

eigenvalue equations which are called *canonical Hartree-Fock equations*. We see that the Fock operator, F, remains unchanged, whereas, the individual functions, ϕ_j, are altered after the similarity transformation. It can be easily proved (refer to the appendix for the proof) that similarity transformation does not change the value of any determinant. Since the individual functions, ϕ_j, have no physical meaning, unlike the determinant Φ, there is no harm in performing a similarity transformation in order to transform the original HF equations into the canonical HF equations, which are much easier to handle.

It is obvious from Eq. 2.57 that the Fock operator, F, itself contains the ϕ_i functions that are to be evaluated. Therefore, the HF equation has to be solved *iteratively*. First we choose a series of one-electron functions ϕ_1^0, ϕ_2^0, ϕ_n^0, construct an initial Fock operator $\left(F_0\right)$, and by solving the HF equations we obtain the new series ϕ_1^1, ϕ_2^1, ϕ_n^1. We then construct a new F^1 from the ϕ_1^1, ϕ_2^1, ϕ_n^1 one-electron functions. This procedure is repeated until convergence has been reached. In other words, a *self-consistent field* (SCF) is reached and the algorithm is therefore, called the SCF procedure.

Despite the fact that the one-electron function, ϕ, and the energy ε have no physical meaning, they can be associated with a descriptive model called the

molecular orbital model (MO model). According to this model, the one-electron functions, ϕ_i, are called *orbitals* and the energies, ε_i, associated with it are called *orbital energies*. The list of filled one-electron orbitals is called the *electron configuration*. A number of orbitals which are left vacant, $(\phi_{n+1},...)$, are called the *virtual orbitals*.

2.5.1 Restricted Hartree-Fock Method

A restricted Hartree-Fock calculation is commonly used for closed-shell systems. It forces each electron pair in a molecule to occupy a single molecular orbital. RHF treatment of a closed-shell system will therefore result in all doubly occupied MOs, whereas, that of an open-shell system will result in both singly and doubly occupied MOs (Figure 2.1). In RHF treatment, the ground state energy[7] for a closed-shell system can be written as:

$$E = 2\sum_{i=1}^{N/2}\varepsilon_i - \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \qquad (2.61)$$

where N is the number of MOs, J_{ij} and K_{ij} are the Coulomb and exchange integrals, respectively. The corresponding *n* electron wavefunction is an eigenfunction of the total spin operator S^2.
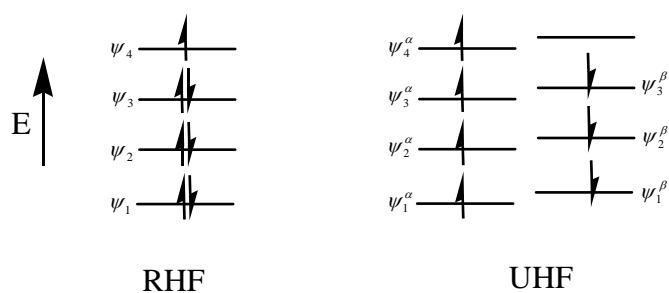
Figure 2.1 MOs for RHF and UHF theories

2.5.2 Unrestricted Hartree-Fock Method

Unrestricted Hartree-Fock calculations are commonly used for open-shell systems such as radicals. In this approach, different spatial orbitals are assigned to α and β electrons (Figure 2.1). Therefore, there are two distinct sets of molecular orbitals, which are defined by two sets of MO expansion coefficients:

$$\psi_i^\alpha = \sum_{\mu=1}^{N} c_{\mu i}^\alpha \phi_\mu; \qquad \psi_i^\beta = \sum_{\mu=1}^{N} c_{\mu i}^\beta \phi_\mu \qquad (2.62)$$

The coefficients $c_{\mu i}$ are varied independently, leading to the UHF generalization of the Roothaan-Hall equation, the Pople-Nesbet equations

$$\sum_{\upsilon=1}^{N} (F_{\mu\upsilon}^\alpha - \varepsilon_i^\alpha S_{\mu\upsilon}) c_{\upsilon i}^\alpha = 0; \qquad (2.63)$$

$$\sum_{\upsilon=1}^{N} (F_{\mu\upsilon}^\beta - \varepsilon_i^\beta S_{\mu\upsilon}) c_{\upsilon i}^\beta = 0, \qquad \mu = 1, 2, ..., N \qquad (2.64)$$

If we assume N_α and N_β are the total number of spatial orbitals for α and β spins, then the HF ground state energy[8] using unrestricted spin orbitals can be written as:

$$E = \sum_{i=1}^{N_\alpha} \varepsilon_i^\alpha + \sum_{i=1}^{N_\beta} \varepsilon_i^\beta + \frac{1}{2} \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\alpha} (J_{ij}^{\alpha\alpha} - K_{ij}^{\alpha\alpha}) + \frac{1}{2} \sum_{i=1}^{N_\beta} \sum_{j=1}^{N_\beta} (J_{ij}^{\beta\beta} - K_{ij}^{\beta\beta}) + \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} J_{ij}^{\alpha\beta} \qquad (2.65)$$

The UHF method is capable of providing a qualitatively correct description of bond dissociation, and it is mathematically more straightforward than RHF for open-shell systems. One of the major drawbacks of the UHF method is that the resulting wavefunction is not always an exact eigenfunction of the total spin operator $<S^2>$, and it may be contaminated by states of higher spin multiplicity. The true eigenvalue of $<S^2>$ is $S(S+1)$, where S is the total electronic spin of the wavefunction, and the

degree of contamination is reflected in the amount by which $<S^2>$ (UHF) exceeds $S(S+1)$.

2.6 Electron Correlation

The correlation between the motions of electrons is referred to as "electron correlation". The HF method provides an approximate solution to Schrödinger equation by simplifying the wavefunction ψ to a single Slater determinant. Due to this simplification, it provides an inadequate treatment of the correlation between the motions of electrons within a molecular system, especially that arising between electrons of opposite spins. It however, accounts for the correlation between the motions of electrons of same spin. This correlation is termed as exchange correlation and is automatically taken into account by the antisymmetry requirement of the wavefunction.

The difference between HF energy and the exact (for a simplified non relativistic Hamiltonian) energy of a system is referred to as "correlation energy". Correlation energy is small compared to the total energy but it is of the same order of magnitude as the quantities of chemical interest. Any method which goes beyond SCF in attempting to treat electron correlation properly is known as an electron correlation method (despite the fact that HF theory does include some correlation effects) or a post-SCF method.

Most ab initio methods dealing with electron correlation are based on the HF reference wavefunction. Almost all the post-SCF methods, such as, configuration interaction (CI), coupled cluster (CC), multi-reference configuration interaction (MRCI), multi-configuration self-consistent field (MCSCF) and complete active space

self-consistent field (CASSCF), use one of the following approaches to improve upon the HF wavefunction:

1) Optimizing only the coefficients of the Slater determinants.

2) Optimizing both the coefficients of the Slater determinants and the coefficients of the one-electron wavefunctions forming the Slater determinant.

Approach 1) is the basic idea behind all the CI methods. The Hartree-Fock wavefunction, ψ_{HF}, is only one of the $N!/(N-n)!n!$ possible arrangements (or configurations) of n electrons in the N spin orbitals. In this approach, all other determinantal wavefunctions are derived from the reference HF wavefunction by substitution of occupied spinorbitals by virtual spinorbitals. This substitution is actually the excitation of electrons from the occupied orbitals to the unoccupied or virtual orbitals. Depending on the number of electrons excited, we have singly-excited configurations, doubly-excited configurations and so on. These excited configurations are often abbreviated as singles, doubles, and so forth. The full CI wavefunction is a linear combination of all single, double and multiple substitutions:

$$\Psi = \alpha_{HF}\Psi_{HF} + \sum \alpha_S \Psi_S \qquad (2.66)$$

The above expression represents a mixing of all possible electronic configurations of the molecules, all of which have some probability of being attained according to the laws of quantum mechanics. A full CI method is the most complete non-relativistic treatment possible within the limitations imposed by the basis set. As the basis set becomes more and more flexible, the results of a full CI treatment approaches the exact solution of the non-relativistic, Born-Oppenheimer approximated,

Schrödinger equation. The full CI method has the advantage of being well defined, size-consistent (energy of well separated molecules is equal to the sum of energies of individual molecules) and variational (provides an upper bound to energy). Because of the enormous amount of computation time required to perform a full CI calculation, it is practically impossible to treat molecules with more than a few heavy atoms.

Therefore, limited CI methods, such as, CIS, CID, CISD are used in which the CI series is truncated at a given level of substitution. For example, in the CISD method, the wavefunction is composed of only single and double excitation terms along with the reference HF determinant. Similarly, a CID wavefunction is composed of determinants resulting from only double excitations and the reference HF determinant. The greatest disadvantage of these limited CI methods is that they are not size-consistent. To overcome this deficiency, the quadratic configuration interaction (QCI) method was developed. Corresponding to CID and CISD methods, are the QCID and QCISD methods.[9] QCISD(T) is the QCI method obtained by adding triplet substitutions to QCISD in an iterative manner. Coupled cluster (CC) methods were also developed to correct the size-consistency problem of limited CI methods. The CCD, CCSD and CCSD(T) methods[10,11] include the double, single and double, single, double and triple excitations, respectively. CC methods are size-consistent but they are not variational and they are computationally less efficient than the limited CI methods.

Approach 2 is the basic idea behind multi-configuration self-consistent field (MCSCF) method and the complete active space self-consistent field (CASSCF) method. The MCSCF wavefunction is a truncated CI expansion

$$\left| \psi_{MCSCF} \right\rangle = \sum_I C_I \left| \psi_I \right\rangle \tag{2.67}$$

in which both the expansion coefficients (C_I) and the orthonormal orbitals contained in $\left|\psi_I\right\rangle$ are optimized. For a closed-shell system, if only one determinant is included in the expansion (Eq. 2.67), the MCSCF and HF methods become identical. A major problem with the MCSCF calculations is the selection of appropriate configurations to be used in the wavefunction. Selection of inappropriate configurations would lead to biased results.

The most commonly used MCSCF method is the CASSCF method in which the entire orbital space is divided into three orthogonal subspaces, namely, inactive, active and virtual orbitals. The inactive orbitals represent the core orbitals which are kept doubly occupied in all configurations of the CASSCF wavefunction. The virtual orbitals are those which are kept empty during the CASSCF calculations. The remaining electrons occupy the active orbitals. It is a reasonable choice to take the active orbitals as those MOs that arise from valence orbitals of the atoms that form the molecule. The CASSCF wavefunction is a linear combination of all configurations that can be attained by distributing the active electrons among the active orbitals in all possible ways and that have the same spin and symmetry eigenvalues as the state to be treated.[12,13] The CASSCF method is far less arbitrary than the general MCSCF method. The major drawback of CASSCF method is that the number of configurations increases dramatically with the size of molecule and hence its application is limited to small molecules.

Another electron correlation method which is developed to improve upon the HF wavefunction is the Møller-Plesset perturbation theory. A brief discussion of the perturbation theory can be found in section 2.4.2. In this approach, electron interaction is treated as a perturbation to the sum of one-electron Hamiltonians. If the perturbation

correction to energy is truncated at the second, third, fourth or fifth order, the method is known as MP2[14], MP3[15,16], MP4[17] and MP5[18] respectively. The commonly used MPn methods are based on the HF reference wavefunction. CASPT2[19] and CASPT3[20] methods are MP2 and MP3 methods using CASSCF reference wavefunction.

2.8 Basis Set

Simply a basis set is a set of functions used to describe the orbital of a system (which in turn combine to approximate the total electronic wavefunction). The molecular orbitals Ψ_i in a Hartree-Fock treatment are expressed as a linear combination of a pre-defined set of one-electron functions or N nuclear-centered functions known as basis functions ϕ_μ $(\mu = 1,2,...,N)$,

$$\Psi_i = \sum_{\mu=1}^{N} c_{\mu i} \phi_\mu \qquad (2.68)$$

Basis sets assign a group of basis functions to each atom within a molecule to approximate its orbitals. The following two types of basis functions are most widely used: (1) Slater-type functions and (2) Gaussian-type functions.

The Slater-type orbitals (STOs) are characterized by the exponential factor $\exp(-\xi r)$ and are represented by the following expression:

$$\phi_i = N_j(x,y,z)\exp(-\xi r) \qquad (2.69)$$

STOs provide a very good representation of atomic orbitals because they possess a cusp at the nucleus. It is however, very difficult to evaluate the two-electron integrals using STOs.

Gaussian-type functions (GTFs) are characterized by the exponential factor $\exp(-\alpha r^2)$ and are represented by the following expression:

$$\phi_i = N_j(x, y, z)\exp(-\alpha r^2) \tag{2.70}$$

GTFs lack the proper cusp behavior of the STOs as the distance between the nucleus and electrons approaches zero and they die off quickly at large distances. Even though the GTFs do not represent atomic orbitals as well as STOs, they are widely used because the evaluation of two-electron integrals is much easier using the GTFs. Due to the ease of calculating two-electron integrals with GTFs, most ab initio electronic structure programs use GTFs rather than STFs as basis functions. In order to provide an improved description, the individual basis functions in a Gaussian basis set are often taken as a linear combination of GTFs

$$\phi_\mu = \sum_k d_{\mu k} g_k \tag{2.71}$$

where the coefficients $d_{\mu k}$ are fixed and the individual functions g_k are all of the same type. Such basis functions ϕ_μ are known as "contracted Gaussians" and the individual functions g_k are known as "primitives". A basis function consisting of a single Gaussian function is referred to as "uncontracted". A brief description of various types of basis set is given below.

2.8.1 Minimal Basis Sets

A minimal basis set is the one which contains the minimum number of basis functions needed for each atom, while maintaining the overall spherical symmetry. Minimal basis set use fixed atomic-type orbitals. For instance, the "STO-KG" basis set is a minimal basis set which approximates the expansion of Slater-type atomic orbitals

(STOs) by taking a linear combination of K gaussian functions. The commonly used STO-KG minimal basis set is STO-3G[21,22] which uses three gaussian primitives (3G) per basis function. "STO" stands for Slater-type orbitals and the STO-3G basis set approximates the Slater-type atomic orbitals using three gaussian primitives. An example of the atomic orbitals required by a minimal basis set for any theoretical calculation with carbon and hydrogen atoms is shown below.

H: 1s

C: 1s, 2s, $2p_x$, $2p_y$, $2p_z$

2.8.2 Split Valence Basis Sets

A basis set has to be very flexible in order to be able to provide a realistic description of atomic orbitals. To increase the flexibility of a basis set one has to increase its size and the simplest way of doing this is to increase the number of basis functions on each atom. A basis set obtained by doubling all the functions of a minimal basis set is referred to as a "double-zeta" basis set. An example of a double-zeta basis set is the Dunning-Huzinaga basis set (D95), in which all the molecular orbitals are formed by a linear combination of two different functions for each atomic orbital.

A much simpler way of extending a basis set is to just double the valence functions of a minimal basis set. Such a basis set is known as a "split valence" basis set in general and a "double split valence" basis set in particular. The commonly used double split valence basis sets are 3-21G[23-25] and 6-31G[26,27] basis set. A 3-21G basis set is formed by taking a linear combination of three GTO's to form the core orbitals and the valence orbitals are split into two parts, formed by taking a linear combination

of two and one GTO's for the inner and outer parts respectively. For example, hydrogen and carbon atoms using a double split valence basis set are represented as follows:

H: 1s, 1s´

C: 1s, 2s, 2s´, $2p_x$, $2p_x´$, $2p_y$, $2p_y´$, $2p_z$, $2p_z´$

A triple split valence basis set, such as 6-311G, is formed by splitting the valence orbitals into three parts.

2.8.3 Polarized Basis Set

Split valence basis sets allow orbitals to change size, but do not allow them to change shape. Polarized basis sets remove this limitation by adding orbitals with angular momentum beyond what is required for the ground state to the description of each atom. For example, polarized basis sets add *p* functions to hydrogen atoms, *d* functions to carbon atoms and *f* functions to transition metals. The most commonly used polarized basis set, 6-31G(*d*) (also represented as 6-31G*) is formed by adding *d* functions to all the heavy atoms. The other commonly used basis set, 6-31G(*d,p*) (also represented as 6-31G**) is formed from the 6-31G(*d*) basis set by adding *p* functions to hydrogen atoms. In a similar manner, the 6-31G(*mdf, npd*) basis set is formed from the 6-31G basis set by adding *m* sets of *d* functions and one set of *f* function to heavy atoms and *n* sets of *p* functions and one set of *d* function to hydrogen and helium. For cases where the description of hydrogen atoms is important, a set of *p* functions is usually added to hydrogen atoms.

2.8.4 Diffuse Basis Sets

Diffuse functions are larger-size versions of s-and p-type functions. They allow orbitals to occupy a larger region of space. Basis sets with diffuse functions are important for systems where electrons are relatively far from the nucleus: molecules with lone pairs, anions and other systems with significant negative charge, systems in their excited states, systems with low ionization potentials, description of absolute acidities, and so on. The 6-31+G(d) basis set is an example of a diffuse function basis set. It is formed from the 6-31G(d) basis set by incorporating a set of *s* and *p* diffuse functions to the heavy atoms.

In selecting a basis set, the number of expansion and the nature of the functions ϕ_i need to be considered. A limiting HF treatment would involve an infinite set of basis functions ϕ_i. This is clearly impractical in terms of the basis set expansion required to describe various properties satisfactorily. In general, a larger basis set, more accurately approximates the orbitals by imposing fewer restrictions on the location of electrons in space.

2.8.5 High Angular Momentum Basis Sets [28]

High angular momentum basis sets consists of split valence basis-set plus polarization and diffuse functions. Larger basis sets add multiple polarization function per atom to the triple zeta basis set. Multiple polarizations are now practical for many systems and although not generally required for a HF calculation, multiple polarization functions are useful for describing the interactions between electrons in electron correlation methods. The examples of high angular momentum basis set are as follows:

* 6-31G (2d) – In this basis set two d functions are added to heavy atoms.

* 6-311G (2df, pd) – Besides the (311) valence functions two d functions and one f functions are added to heavy atoms, p and d function are added to the hydrogen atom.

* 6-311G (3df, 2df, p) – three d functions and one f function are added to atoms with Z >11, two d, functions and one f function to first-row atoms (Li to Ne) and one p function to hydrogen. High angular momentum basis sets augmented with diffuse functions represent the most sophisticated basis sets available in the Gaussian program. Most widely used high accurate ab initio calculation would be produced by reasonably sophisticated polarized split-valence basis sets augmented with high angular momentum and diffuse atomic orbitals.

2.9 G3(MP2) Theory

The G3(MP2) theory[29] developed by Pople et al., is a general procedure based on ab initio molecular orbital theory for the accurate calculation of energies. In the G3(MP2) approach, all the structures are optimized at the MP2(full)/6-31G(d) level using both the core and the valence electrons. Followed by the geometry optimization, a series of single point energy calculations are carried out at higher levels of theory. All the subsequent single point energy calculations include only the valence electrons in the treatment of electron correlation, i.e., frozen core (fc) approximation. The first high level calculation is performed at the quadratic configuration interaction level of theory[10] with the 6-31G(d) basis set, i.e., QCISD(T)/6-31G(d). This energy is then modified by a series of corrections to obtain the total energy, $E_0[G3(MP2)]$

$$E_0[G3(MP2)] = QCISD(T)/6\text{-}31G(d) + \Delta E_{MP2} + \Delta E(SO) + E(HLC) + E(ZPE), \quad (2.72)$$

where ΔE_{MP2} is the correction at the second order Møller-plesset level[15] (MP2) given by

$$\Delta E_{MP2} = [E(MP2/G3MP2large)] - [E(MP2/6\text{-}31G(d))] \qquad (2.73)$$

The G3MP2 large basis set is the same as the G3 large basis set used in the G3 theory, except that the core polarization functions are not included. $\Delta E(SO)$ is the spin-orbit correction and it is included only for the atomic species. The zero-point energy correction, $E(ZPE)$ is obtained from scaled HF/6-31G(d) frequencies. The frequencies are scaled by a factor of 0.8929. $E(HLC)$ is the "high-level correction" term which is added to take into account the remaining deficiencies in the energy calculation. $E(HLC)$ is given by

$$E(HLC) = -An_{\beta} - B(n_{\alpha} - n_{\beta}) \qquad (2.74)$$

for molecules and by

$$E(HLC) = -Cn_{\beta} - D(n_{\alpha} - n_{\beta}) \qquad (2.75)$$

for atoms and atomic ions, where n_{α} and n_{β} are the number of α and β valence electrons respectively, with $n_{\alpha} \geq n_{\beta}$. A, B, C and D are constants. For G3(MP2) theory, A=9.279 mhartrees, B=4.471 mhartrees, C=9.345 mhartrees and D=2.021 mhartrees. The average absolute deviation from experiment of G3(MP2) theory is 1.30 kcal mol^{-1} for energies and 4.72 kcal mol^{-1} for enthalpies for small and median sized systems.

2.10 Density Functional Theory (DFT) [13]

Density functional theory is a quantum mechanical method used in physics and chemistry to investigate the electronic structure of many-body systems, in particular molecules and the condensed phases. In recent years Density functional theory (DFT) has become very popular. Based on the pragmatic observation, DFT is less

computationally intensive than other methods with similar accuracy. The main objective of density functional theory is to replace the many-body electronic wavefunction with the electronic density as the basic quantity.

In 1964, Hohenberg and Kohn proved that the ground-state molecular energy, wavefunction, and all other molecular electronic properties are uniquely determined by the electron probability density $\rho(x, y, z)$, a function of only three variables.[30] Therefore, the ground-state energy E_0 is a functional of electron probability density ρ and can be written as $E_0 = E_0[\rho]$.

In the traditional quantum-chemical approach, one determines the wavefunction ψ first and then the electron probability density ρ by integrating ψ. The Hohenberg-Kohn theorem on the other hand tells us that if we know the ground-state electron density $\rho(x, y, z)$, then all the ground-state molecular properties can be calculated from it. It however, does not tell us how to calculate E_0 from ρ or how to find ρ without first finding ψ. Later in 1965, Kohn and Sham[31] showed that the exact ground-state purely electronic energy E_0 of an n-electron system with ground-state electron probability density ρ can be written as

$$E_0 = -\frac{1}{2} \sum_{i=1}^{n} \left\langle \psi_i(1) \middle| \nabla_1^2 \middle| \psi_i(1) \right\rangle - \sum_{\alpha} \int \frac{Z_\alpha \rho(1)}{r_{1\alpha}} dv_1 + \frac{1}{2} \iint \frac{\rho(1)\rho(2)}{r_{12}} dv_1 dv_2 + E_{XC}[\rho], \quad (2.76)$$

where $\psi_i(1)$ are the Kohn and Sham orbitals and $E_{XC}[\rho]$ is the exchange-correlation energy. Kohn-Sham also showed that the exact ground-state ρ can be found from ψ_i's, according to

$$\rho = \sum_{i=1}^{n} |\psi_i|^2 \tag{2.77}$$

The Kohn-Sham orbitals are obtained by solving the one-electron equations

$$\widehat{F}_{KS}(1)\psi_i(1) = \varepsilon_{i,KS}\psi_i(1), \tag{2.78}$$

where the Kohn-Sham operator $\widehat{F}_{KS}$ is given by

$$\widehat{F}_{KS} \equiv -\frac{1}{2}\nabla_1^2 - \sum_\alpha \frac{Z_\alpha}{r_{1\alpha}} + \sum_{j=1}^{n} \widehat{J}_j(1) + V_{XC}(1) \tag{2.79}$$

where $\widehat{J}_j(1)$ is the Coulomb operator defined by

$$\widehat{J}_j(1) = \int |\phi_j(2)|^2 \frac{1}{r_{12}} dv_2 \tag{2.80}$$

and V_{XC} is the exchange-correlation potential defined by

$$V_{XC} = \delta E_{XC}[\rho]/\delta\rho \tag{2.81}$$

The Hartree-Fock operator $\widehat{F}$ for electron m in a molecule of n-electrons is given by

$$\widehat{F} \equiv -\frac{1}{2}\nabla_m^2 - \sum_\alpha \frac{Z_\alpha}{r_{m\alpha}} + \sum_{j=1}^{n} \left[\widehat{J}_j(m) - \widehat{K}_j(m)\right] \tag{2.82}$$

This equation is surprisingly similar to Eq. 2.67 except that the exchange operators $-\sum_{j=1}^{n} \widehat{K}_j(m)$ are replaced by V_{XC}, which handles the effects of both exchange (antisymmetry) and electron correlation. The density functional theory (DFT) methods are self-consistent methods like the Hartree-Fock method.

The only problem with density functional methods is that the correct functional form of $E_{XC}[\rho]$ is not known and it is too complicated to be evaluated analytically. Therefore, numerical quadrature must be used which may lead to significant loss of precision. In practice E_{XC} is divided into two parts, exchange E_X and correlation E_C,

$$E_{XC} = E_X + E_C \tag{2.83}$$

2.10.1 Exchange Functionals

Listed below are some of the commonly used exchange functionals.

Exchange functional proposed by Slater[32]

$$E_X^{LDA}[\rho] = -\int \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \rho(r)^{\frac{4}{3}} dr \tag{2.84}$$

The corresponding potential is

$$\varepsilon_X^{LDA}[\rho] = -\left(\frac{3}{\pi}\rho(r)\right)^{\frac{1}{3}} \tag{2.85}$$

Becke's 1988 exchange functional (**B88**)[33]

$$\varepsilon_X^{B88}[\rho] = -\varepsilon_X^{LDA}[\rho]\left[1 - \frac{\beta}{2^{\frac{1}{3}}A_X}\frac{z^2}{1+6\beta z \sinh^{-1}(z)}\right] \tag{2.86}$$

where

$$z = 2^{\frac{1}{3}}\frac{|\nabla\rho|}{\rho^{\frac{4}{3}}}, \qquad A = \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \qquad \text{and} \qquad \beta = 0.0042 .$$

Perdew-Wang (PW91)[34,35] exchange functional

$$\varepsilon_X^{PW91} = \varepsilon_X^{LDA}(\rho)\left[\frac{1 + sa_1\sinh^{-1}(sa_2) + \left(a_3 + a_4 e^{-100s^2}\right)s^2}{1 + sa_1\sinh^{-1}(sa_2) + a_5 s^4}\right] \tag{2.87}$$

where

$$s = \frac{|\nabla\rho|}{\left(24\pi^2\right)^{\frac{1}{3}}\rho^{\frac{4}{3}}}, \quad a_1 = 0.19645, \; a_2 = 7.7956, \; a_3 = 0.2743, \; a_4 = -0.1508 \text{ and } a_5 = 0.004.$$

2.10.2 Correlation Functionals

Listed below are some of the commonly used correlation functionals.

Vosko, Wilk and Nusair correlation functional (**VWN**)[36]

$$\varepsilon_C^{VWN}[\rho] = \frac{A}{2}\left[\ln\frac{x}{X(x)} + \frac{2b}{Q}\tan^{-1}\frac{Q}{2x-b} - \frac{bx_0}{X(x_0)}\left(\ln\frac{(x-x_0)^2}{X(x)} + \frac{2(b+2x_0)}{Q}\tan^{-1}\frac{Q}{2x+b}\right)\right] \tag{2.88}$$

where the functions x, X and Q are given by

$$x = r_s^{\frac{1}{2}}, \quad r_s = \left(\frac{4\pi\rho}{3}\right)^{-\frac{1}{3}}, \quad X(x) = x^2 + bx + x, \quad Q = (4c - b^2)^{1/2}$$

and the constants are $A = 0.0621814$, $x_0 = -0.409286$, $b = 13.0720$ and $c = 42.7198$.

Lee, Yang and Parr correlation functional (**LYP**)[37]

$$\varepsilon_C^{LYP}[\rho] = -a\frac{1}{1 + d\rho^{-\frac{1}{3}}}\left\{\rho + b\rho^{-\frac{2}{3}}\left[C_F\rho^{\frac{5}{3}} - 2t_w + \frac{1}{9}\left(t_w + \frac{1}{2}\nabla^2\rho\right)\right]e^{-c\rho^{\frac{1}{3}}}\right\} \tag{2.89}$$

where

$$t_w = \frac{1}{8}\left(\frac{|\nabla\rho|^2}{\rho} - \nabla^2\rho\right), \quad C_F = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}}, \quad a=0.04918, \, b=0.132, \, c=0.2533 \text{ and } d=0.349.$$

Perdew-Wang (PW91)[35,36] correlation functional

$$V_C^{PW91}[\rho] = \varepsilon_C^{LDA}[\rho] + \rho H[\rho, s, t] \tag{2.90}$$

where

$$H = \frac{\beta^2}{2\alpha}\ln\left[1 + \frac{2\alpha}{\beta}\frac{t^2 + At^4}{1 + At^2 + A^2t^4}\right] + C_{c0}\left[C_c(\rho) - C_{c1}\right]t^2 e^{-100s^2} \tag{2.91}$$

with

$$A = \frac{2\alpha}{\beta}\left[e^{\frac{2\alpha\varepsilon_C^{LDA}(\rho)}{\rho\beta^3}} - 1\right]^{-1} \tag{2.92}$$

s is the same as in PW91 exchange functional, and

$$t = \frac{\left(\frac{\pi}{3}\right)^{\frac{1}{6}}}{4}\frac{|\nabla\rho|}{\rho^{\frac{7}{6}}}, \quad \alpha = 0.09, \beta = 0.0667263212, C_{c0} = 15.7559, C_{c1} = 0.0035521, \text{ and}$$

$$C_c(\rho) = C_1 + \frac{C_2 + C_3 r_s + C_4 r_s^2}{1 + C_5 r_s + C_6 r_s^2 + C_7 r_s^3} \tag{2.93}$$

with $C_1 = 0.001667$, $C_2 = 0.002568$, $C_3 = 0.023266$, $C_4 = 7.389 \times 10^{-6}$ $C_5 = 8.723$, $C_6 = 0.472$ and $C_7 = 0.07389$.

There are three different categories of DFT methods and they all differ in the way they handle $E_{XC}[\rho]$.

1) The **Local DFT methods** based on the local density approximation (LDA), approximates $E_{xc}[\rho]$ as

$$E_{XC}^{LDA}[\rho] = \int \varepsilon_{XC}^{LDA}[\rho(r)]\rho(r)dr \qquad (2.94)$$

For example, the S-VWN method which is a combination of Slater-type exchange functional and the VWN correlation functional parameterized on the homogenous electron gas.

2) The **Non-Local or Gradient-Corrected DFT methods** which contain gradient-corrected exchange and correlation functionals (e.g., B-LYP and B-P86 methods).

$$E_{XC}^{NL}[\rho] = \int \varepsilon_{XC}[\rho(r)|\nabla\rho(r)|\nabla^2\rho(r)]dr \qquad (2.95)$$

3) The **Hybrid DFT methods** contain a mixture of Hartree-Fock exchange energy and DFT exchange correlation energy. For example, B3LYP, B3P86 and B3PW91 methods. The three-parameter mixing scheme proposed by Becke in 1993 is[38]

$$E_{XC} = E_{XC}^{LDA} + a_0\left(E_X^{HF} - E_X^{LDA}\right) + a_X\Delta E_X^{B88} + a_C\Delta E_C^{non-local} \qquad (2.96)$$

The correlation functional Becke used in his original paper is PW91. The B3LYP functional incorporated in Gaussian 94[39] and Gaussian 98[40] suit of programs is

$$A * E_X^{Slater} + (1-A)E_X^{HF} + B * E_X^{Beck88} + E_C^{VWN} + C * \Delta E_C^{LYP} \qquad (2.97)$$

with A = 0.80, B = 0.72 and C = 0.81 obtained by fitting to G2 test set.

2.11 Natural Bond Orbital (NBO) Analysis[41]

Natural bond orbital analysis (NBO) is the name of a whole set of analysis techniques. One of these is the natural population analysis (NPA) for obtaining

occupancies (how many electrons are assigned to each atom) and charges. Some researchers use the acronyms NBO and NPA interchangeably.

Rather than using the molecular orbitals directly, NBO uses the natural orbitals.Natural orbitals (NOs) are the intrinsic orbitals (θ_k) that arise as eigenfunctions of the first-order reduced density operator $\hat{\Gamma}$,

$$\hat{\Gamma}\theta_k = q_k\theta_k \tag{2.98}$$

which is formed by 'reducing' the wavefunction probability distribution to the single-particle level,

$$\hat{\Gamma} = N\int \psi(1,2,...,N)\psi^*(1',2,...,N)d\tau_2...d\tau_N \tag{2.99}$$

and whose eigenorbitals are hence 'natural' to ψ itself. As shown by Löwdin[42] and others, rigorous quantum-mechanical questions involving subsystems of an N-particle system are best formulated in terms of reduced density operators. In particular, the squared probability amplitude $\left|\langle\psi(1,2,...N)|\phi(1)\rangle\right|^2$ that an electron of $\psi(1,2,...N)$ is 'in' orbital ϕ (i.e., the population of $\phi(1)$ in the wavefunction) is rigorously expressed, for any possible orbital ϕ, as

$$q_\phi = \left\langle\phi\left|\hat{\Gamma}\right|\phi\right\rangle \tag{2.100}$$

The occupancies q_ϕ are intrinsically non-negative and limited by the Pauli exclusion principle, e.g., for spatial orbital $\phi(r)$,

$$0 \le q_\phi \le 2 \tag{2.101}$$

(The analogous restriction $q_\phi \leq 1$ applies to spin orbitals). The sum of occupancies q_k over any complete orthonormal set $\{\phi_k\}$ accounts for all N electrons,

$$\sum_k \left\langle \phi_k \left| \hat{\Gamma} \right| \phi_k \right\rangle = \sum_k q_k = Tr\{\hat{\Gamma}\} = N \qquad (2.102)$$

The Mulliken populations generally fail to satisfy the physical constraints (Eq. 2.101 and Eq. 2.102)

The chemist's idealized Lewis structure picture describes the $N/2$ electron pairs as localized in one-centre (lone pair) or two-centre (bond) regions of the molecule. The natural bond orbital (NBO) algorithm[43,44] leads to an optimal set of one- and two-centre orbitals b_i that are in close correspondence with this picture. In effect, the algorithm searches the density matrix for the set of $N/2$ localized Lewis-type lone pair and bond orbitals of near double occupancy that best describe the given wavefunction, with the residual weakly occupied non-Lewis-type antibonding and *Rydberg orbitals* representing small corrections to delocalization.

2.12 Outline of Diffusion Monte Carlo Method[45]

The diffusion Monte Carlo (DMC) is a numerical method used to solve the Schrodinger equation by using Monte Carlo sampling. It is based on rewriting the Schrödinger equation in imaginary time[46] $\tau = it$. The imaginary time Schrödinger equation is:

$$\frac{\partial |\Psi\rangle}{\partial \tau} = H |\Psi\rangle \qquad (2.103)$$

DMC method was first used in the electronic structure studies by Anderson before being applied to nuclear motions by Coker and Watts.[47] This quantum treatment is widely used and gains insights in the understanding of Van der Walls complexes because it gives the exact energy of the vibrational ground state. The high anharmonicity of the interactions govern these complexes and the zero point energy must be included. The DMC algorithm, is easy to implement, but has not the popularity because of its difficulties to get the excited states. However, a good knowledge of the dynamics of the ground vibrational state will do because it can probe important physical properties of the system, such as dissociation energy, vibrationally averaged structures, rotational constants and wavefunctions. The accuracy of the results depends only on the quality of the potential used.

Assuming the multidimensional potential energy surface V of an electronic state is known, DMC solves the time dependent Schrödinger equation for a system of N atoms of mass m_j. By transforming in this equation the time frame into an imaginary one, we get the expression of a classic time dependent diffusion process with a sink or source term.

The random walk technique developed by Metropolis and Ulam,[48] is a generally successful method to reproduce a diffusional behavior.[49] As a result, the Schrödinger equation can be simulated by both a random walk for the kinetic energy part and a continuous weight assessment for the potential energy term. This is achieved by generating a population of certain number of replicas and by adjusting their weights according to their energy as they move randomly on the potential energy surface. Each of these "wavefunction particles" describes one possible geometry of the system and also represents (depending on its weight) a part of the wavefunction.

2.13 References

1   Szabo, A., Ostlund, N. S. Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory; Dover Pub. Inc. Mineola, New York, 1996.

2   Veszprémi, T., Fehér, M. Quantum Chemistry: Fundamentals to Applications; New York, 1999.

3   Yamaguchi, Y., Osamura, Y.,Goddard, J. D., Schaefer, H. F. A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory; Oxford University Press, New York, 1994.

4   Phillips, L. F. Basic Quantum Chemistry; John Wily & Sons, Inc. New York, 1965.

5   Dykstra, C. E. Quantum Chemistry and Molecular Spectroscopy; Prentice Hall, New Jersey, 1992.

6   D. R. Hartree, Proc. Cambridge Phil. Soc. 1928, 24, 111.

7   (a) Roothaan, C. C. J. Rev. Mod. Phys. 1960, 32, 179. (b) Binkley, J. S., Pople, J. A., Dobosh, P. A., Mol. Phys. 1974, 28, 1423.

8   Pople, J. A., Nesbet, R. K., J. Chem. Phys. 1954, 22, 571.

9   Pople, J. A., Head-Gordon, M., Raghavachari, K., J. Chem. Phys. 1987, 87, 5968.

10   Purvis, G. D., Bartlett, R. J., J. Chem. Phys. 1982, 76, 1910.

11   Raghavachari, K., Trucks, G. W., Pople, J. A., Head-Gordon, M., Chem. Phys. Lett. 1989, 157, 479.

12   (a) Levine, I. N. Quantum Chemistry; Prentice Hall, New Jersy, 2000. (b) Koch, W., Holthausen, M. C. A Chemist's Guide to Density Functional Theory; Wiley-VCH, Weinheim, 2000.

13   Hehre, W. J., Radom, L., Schleyer, P. v. R., Pople, J. A. Ab initio Molecular Orbital Theory; John Wiley and Sons, New York, 1996.

14   Almloff, J., Korsell, K., Faegri, Jr., K., J. Comp. Chem. 1982, 3, 385.

15   Pople,J. A.,Binkley, J. S., Seeger, R., Int. J. Quantum Chem. Symp.1976,10, 1.

16   Pople, J. A., Seeger, R., Krishnan, R., Int. J. Quantum Chem. Symp. 1977, 11, 149.

17   Krishnan, R., Pople, J. A., Int., J. Quantum Chem. 1978, 14, 91.

18   Raghavachari, K., Pople, J. A., Replogle, E. S., Head-Gordon, M., J. Phys. Chem. 1990, 94, 5579.

19   McDouall, J. J.,Peasley, K., Robb, M. A., Chem. Phys. Lett. 1988, 148, 183.

20   Werner, H., J. Mol. Phys. 1996, 89, 645.

21   Hehre, W. J., Stewart, R. F., Pople, J. A., J. Chem. Phys. 1969, 51, 2657.

22   Collins, J. B., Schleyer, P. v. R., Binkley, J. S., Pople, J. A., J. Chem. Phys. 1976, 64, 5142.

23   Binkley, J. S., Pople, J. A., Hehre, W. J., J. Am. Chem. Soc. 1980, 102, 939.

24   Gordon, M. S., Binkley, J. S., Pople. J. A., Pietro, W. J., Hehre, W. J., J. Am. Chem. Soc. 1982, 104, 2797.

25   Pietro, W. J., Francl, M. M., Hehre, W. J., Defrees, D. J., Pople, J. A., Binkley, J. S., J. Am. Chem. Soc. 1982, 104, 5039.

26   Ditchfield, R., Hehre, W. J., Pople, J. A., J. Chem. Phys. 1971, 54, 724.

27   Hehre, W. J., Ditchfield, R., Pople, J., A., J. Chem. Phys. 1972, 56, 2257.

28   Galabov B,Yamaguchi Y, Remington R B., Schaefer III H F., Phys. Chem. A, 2002 106,819 -832.

29 Curtiss, L. A., Redfern, P. C., Raghavachari, K., Rassolov, V., Pople, J. A. J. Chem. Phys. 1999, 110, 4703.

30 Hohenberg, P., Kohn, W., Phys. Rev. B 1964, 136, 864.

31 Kohn, W., Sham, L. J., Phys. Rev. A 1965, 140, 1133.

32 Slater, J., The Self-Consistent Field of Molecules and Solids: Quantum Theory of Molecules and Solids, Vol. 4, McGraw-Hill, New York, 1974.

33 Becke, A. D., Phys. Rev. A 1988, 33, 3098.

34 Perdew, J. P., Wang, Y., Phys. Rev. B 1992, 45, 13244.

35 Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R.; Singh, D. J., Fiolhais, C., Phys. Rev. B 1992, 46, 6671.

36 Vosko, S. H., Wilk, L., Nusair, M. Can., J. Phys. 1980, 58, 1200.

37 Lee, C., Yang, W., Parr, R. G., Phys. Rev. B 1988, 37, 785.

38 Becke, A. D., J. Chem. Phys. 1993, 98, 5648.

39 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.; Ortiz, 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-Gordon, M., Gonzalez, C.,Pople, J. A. Gaussian 94 (Revision E.2), Gaussian, Inc., Pittsburgh PA, 1995.

40 Frisch, M. J., Trucks, G. W., Schlegel, H.B., Scuseria, G. E., Robb, M. A, Cheeseman, J. R., Zakrzewski, V. G., Montgomery, Jr. J. A., Startmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B, Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., and Pople, J. A. Gaussian 98 (Revision A.2), Gaussian, Inc., Pittsburgh PA, 1998.

41 Weinhold, F. In Encyclopedia of Computational Chemistry; Schleyer, P. V. R., Allinger, N. L., Kollman, P. A., Clark, T., Schaefer, H. F., Gasteiger, J., Scheiner, P. R., Eds., John Wiely & Sons, 1999, Vol 3, p1792.

42 Löwdin, P. O., Phys. Rev. 1955, 97, 1474.

43 Foster, J. P., Weinhold, F., J. Am. Chem. Soc. 1980, 102, 7211.

44 Reed, A. E., Weinhold, F., J. Chem. Phys. 1983, 78, 4066.

45 Ioan K., Byron F., Klaus S. American Journal of Physics, 64,633-644, 1996.

46 Anderson, J. B., J. Chem. Phys. 1975, 63.1499.

47 Coker D.F., Watts R.O., Mol. Phys. 1986 58,1113 .

48 N. Metropolis and S. Ulam, J.Am. Stat. Assoc. 1949 44, 335 .

49 Einstein A., Phys.,1905, 17,549.

2.14 Appendix

1. Derivation of Eq. 2.52

Let us express the perturbed wavefunction $\Psi^{(1)}$ as a function of the exact eigenfunctions, $\Psi_i^{(0)}$, of the model operator $H^{(0)}$:

$$\Psi^{(1)} = \sum_i c_i \Psi_i^{(0)}$$

and insert it into the second equation of Eq. 2.39. After rearranging, we get

$$\sum_i c_i \left(H^{(0)} - E^{(0)} \right) \Psi_i^{(0)} = \left(E^{(1)} - V \right) \Psi^{(0)}$$

Multiplying the expression by $\Psi_k^{(0)}$ from the left and forming the scalar products, we obtain:

$$\sum_i c_i \left\langle \Psi_k^{(0)} \middle| \left(H^{(0)} - E^{(0)} \right) \middle| \Psi_i^{(0)} \right\rangle = \left\langle \Psi_k^{(0)} \middle| \left(E^{(1)} - V \right) \middle| \Psi^{(0)} \right\rangle$$

and further

$$\sum_i c_i \left(E_i^{(0)} \left\langle \Psi_k^{(0)} \middle| \Psi_i^{(0)} \right\rangle - E^{(0)} \left\langle \Psi_k^{(0)} \middle| \Psi_i^{(0)} \right\rangle \right) = E^{(1)} \left\langle \Psi_k^{(0)} \middle| \Psi^{(0)} \right\rangle - \left\langle \Psi_k^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle$$

If we take it into account that the $\Psi_i^{(0)}$ eigenfunctions are orthonormal (that is, orthogonal and normal), this further simplifies to

$$c_k \left(E^{(0)} - E_k^{(0)} \right) = \left\langle \Psi_k^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle$$

We can easily express c_k, and with it the first-order perturbed wavefunction, $\Psi^{(1)}$:

$$\Psi^{(1)} = \sum_i \frac{\left\langle \Psi_i^{(0)} \middle| V \middle| \Psi^{(0)} \right\rangle}{E^{(0)} - E_i^{(0)}} \Psi_i^{(0)}$$

2. Derivation of Eq. 2.55

We wish to find the minimum of the energy expression

$$E = E - \sum \sum \varepsilon_{ij} \left\langle \phi_i \middle| \phi_j \right\rangle$$

In general, if $A = \langle f | \hat{A} g \rangle$, then $\partial A = \langle \partial f | \hat{A} g \rangle + \langle f | \hat{A} \partial g \rangle$; hence the first variation of

E is:

$$\partial E = \sum_{i=1}^{n} \int \partial \phi_i^*(i) \left[\left[h_1(i) + \sum_{j=1}^{n} \langle \phi_j(j) | h_2(ij)(1 - P_{ij}) | \phi_j(j) \rangle \right] \phi_i(i) - \sum_{j=1}^{n} \varepsilon_{ij} \phi_j(i) \right] d\tau_i$$

$$+ \sum_{i=1}^{n} \int \partial \phi_i(i) \left[\left[h_1(i) + \sum_{j=1}^{n} \langle \phi_j(j) | h_2(ij)(1 - P_{ij}) | \phi_j(j) \rangle \right]^* \phi_i^*(i) - \sum_{j=1}^{n} \varepsilon_{ij} \phi_j^*(i) \right] d\tau_i$$

In the minimum $\partial E = 0$. This can arise if the coefficients of $\partial \phi_i(i)$ or $\partial \phi_i^*(i)$ (i.e., the expressions in square brackets) are zero:

$$\left[h_i(i) + \sum_{j=1}^{n} \langle \phi_i(i) | h_2(ij)(1 - P_{ij}) | \phi_j(j) \rangle \right] \phi_i(i) = \sum_{j=1}^{n} \varepsilon_{ij} \phi_j(j) \qquad i = 1, 2, ..., n$$

The term in the square brackets is the *Fock operator*, $F(i)$. A similar equation can be derived from the second square bracket, which is equivalent to the first one, because the second part of the expression is the complex conjugate of the first part.

3. **Proof to show that a similarity transformation does not change the value of any determinant**

Let **A** be a quadratic matrix and apply a similarity transformation using the **Q** and $\mathbf{Q^{-1}}$ matrices. We need to find the determinant of the transformed matrix $\mathbf{Q^{-1}AQ}$:

$$\left| \mathbf{Q^{-1}AQ} \right| = \left| \mathbf{Q^{-1}} \right| \left| \mathbf{A} \right| \left| \mathbf{Q} \right| = \left| \mathbf{Q^{-1}} \right| \left| \mathbf{Q} \right| \left| \mathbf{A} \right| = \left| \mathbf{Q^{-1}Q} \right| \left| \mathbf{A} \right| = \left| \mathbf{A} \right|$$

Here we used the following two characteristics of determinants:

1. The determinant of a product matrix is equal to the product of the determinants of the original matrices.

2. The determinant of a product matrix is independent of the order of matrices.

Chapter 3 The Conformers of Hydroxyacetaldehyde

3.1 Introduction

3.1.1 Hydroxyacetaldehyde

Hydroxyacetaldehyde, (hydroxyethanal, glycolaldehyde, CH_2OHCHO) which is an isomer of methyl formate and acetic acid and the simplest monosaccharide, have been recently detected toward the Galactic center cloud Sgr B2(N) (Hollis, Lovas and Jewell 2000).[1] It represents a well-known molecule atmospherically relevant[2] and a potential intermediary in the prebiotic syntheses of sugars. The impact of the detection has been significant although its abundance in the ISM is very low. It accomplishes the set of $C_2O_4H_2$ isomers discovered, which has increased the interest in the study of isomerism in interstellar chemistry[1]. Characteristic millimeter-wave rotational transitions have been observed in emission toward Sagittarius B2 (N) (SgrB2(N)) with the National Radio Astronomy Observatory (NRAO). Later investigations have confirmed its insignificant abundance in other sources.[3] Recently, Hollis et al.[4] have spatially mapped the galactic center SgrB2(N) for evaluating the abundance ratio of the detected isomers (glycolaldehyde/acetic acid/methyl formate) 0.5:1:26. The molecule is of considerable biochemical interest and, upon condensation with amines, various natural products are formed and also for its significance in tracing the origin of life on earth and in universe.

Organic species such as glycolaldehyde bring up new questions to the astronomers related to the chemical evolution of the ISM and the interpretation of the observed spectra. Generally, it is believed that saturated molecules in hot cores are synthesized on interstellar dust grains in a low temperature era.[2-4] However the

synthesis of glycolaldehyde is currently unknown towards its source. Ab initio calculations can help the search for solutions of many astrophysical problems like the mechanism of glycolaldehyde formation.

Several research groups have paid attention to glycolaldehyde, its origin in interstellar clouds.[5] Huntress and Mitchell proposed the radioactive association reactions in gas phase for methyl formate[2] which seem not to be able to be extended to all the observed large molecules. Because they become visible on surfaces, it is generally accepted that the surface of the interstellar dust grains plays an important role, it may be assumed that molecular groups forming these species were synthesized on the grains and afterward the molecules were deposited in the gas phase. And other researchers studied the oxidation[6], and its reaction with the OH radical[7,8] both experimentally and theoretically.

Marstokk and Møllendal[9-11] first systematically studied the structure of glycolaldehyde in the gas phase, but only observed one isomer. Their measurements included the dipole moment, the microwave spectra of the parent molecule and deuterated species as well as three other isotopic species. They also pointed out that the *cis* form, denoted GM in this chapter, is the most stable conformer based on low level theoretical calculations of three possible conformers, denoted here as GM, L1 and L3. Later, it was found that there was a fourth conformer of glycolaldehyde (cf. Figure 3.1), denoted in this work as L2, in the theoretical work of Antero *et al.*[12] Recently, Senent[13] studied the torsional spectrum and interconversion process between the four conformers at the MP2/cc-pVQZ level using a two-dimensional variational approach. In addition, Senent computed the rotation parameters corresponding to respective conformers.

3.1.2 DMC

Diffusion Monte Carlo (DMC) method is well studied in recent years, basically it is a very simple but superior method to obtain the ground quantum state of a many body system. Highly accurate DMC calculations of ground-state energies have already been demonstrated.[14-18]

Loosely bound molecular complexes are common in chemistry. They can occur as clusters, as transition states and as reactive intermediates and are involved in many biological processes. In many cases a number of stable local minima are found on the molecular potential energy surface (PES) describing the complex. The past decade has brought tremendous advance in the study of weakly bound states of atoms and molecules. One of the driving forces behind this research has been the need to understand how the properties of atomic or molecular clusters vary as a function of size, particularly the stepwise development of condensed phase attributes, thus allowing us to bridge the gap between isolated molecules and condensed matter. The experimental and theoretical methods which are now being applied to these systems are often capable of revealing quantum-state-specific information and are therefore quite different from those applied in the bulk, resulting in the promise that these finite systems will provide us with fundamentally new insight into condensed-phase behavior.

While the multi-atom bound state problem for a single potential energy surface (PES) is essentially solved (scarcity of reliable global PESs being the only, albeit serious, remaining problem), the situation is much less satisfactory for larger floppy clusters. The only methods available thus far for an accurate treatment of the quantum systems with very many degrees of freedom are based on interpolation approaches and are collectively referred to as quantum Monte Carlo, Their main advantages are the absence of size-dependent properties and the favorable scaling of the computational

effort with the number of degrees of freedom. Successive studies were devoted to the analysis of $SF_6(He)_n$[19] and $HF(He)_n$[20] clusters using the diffusion Monte Carlo (DMC) method to calculate ground-state energetics, structural properties, and the solvent-induced spectral shift of the vibrations of the dopant molecule.

Because hydroxyacetaldehyde is an eight-atom system the total number of nuclear degrees of freedom is eighteen. The only feasible technique available to study systems with such high dimensionality is quantum DMC, which is used throughout this work.

Quantum DMC is now routinely used to solve for the ground-state nuclear wavefunction, this wave function may be calculated from the molecular PES using variational methods or Monte Carlo methods such as DMC.[21-24] Quantum diffusion Monte Carlo can be used to calculate the ground state wave function as it has the ability to determine the exact solution to the nuclear Schrödinger equation. Furthermore, Monte Carlo methods tend to have better scaling properties with increasing dimension than grid-based methods. It has been applied to a wide variety of systems including the thirty-dimensional intermolecular modes of the water hexamer[25] and 142 and 452 torsions of a bimolecular system.[26] Systems where no degrees of freedom have been frozen have also been studied. Examples include nine dimensions for of FCH_3[27] and twelve dimensions for CH_5^+,[28] and most recently the water dimer.[29,30]

3.1.3 Objectives

In this work our main focus is on the implications of introducing the full dimensionality of the Potential Energy Surface (PES) of hydroxyacetaldehyde in the identification and assignment of different conformers. Different structures of the same molecule have distinct physicochemical properties. One of the most explicit posteriori properties of the structure of a molecule is its energy. The energy of a molecule is dependent on parameters such as its environment and conformations. As we change the variable bond length, bond angles and torsional angles in a molecule, the energy of

the molecule will be different. In evaluation of the conformational space of structurally flexible molecules, such a glycolaldehyde, only torsional angles are subject of significant variations. If the calculation is done with two or more parameters, we obtain potential energy surface (PES) or a hypersurface (PEHS).

The potential energy surface of hydroxyacetaldehyde will be mapped and all the critical points identified. All minima, and the transition states connecting them, would be located. From the graph of PES (see Figure 3.2), it is very easy to approximately locate minima point and the transition point (saddle point) as well as the lowest energy pathway connecting. Clearly to see that totally there're 4 conformers (GM,L1,L2 and L3) of glycolaldehyde. In order to identify the observable conformers two dimensions (2D) and 18 dimensions (18D) quantum diffusion Monte Carlo (DMC) simulations were conducted. Both two and eighteen dimensional potential energy surfaces were generated and used in the DMC runs. The rotational constants for 2D were predicted from the simulations and an approximate equilibrium structure can be obtained by combining our theoretical results with the experimentally observed rotational constants. It was predicted that of the remaining three minima effectively only one conformer is observable. The predicted rotational constants for this conformer are provided.

3.2 Computational Methods

All *ab initio* calculations reported in this work were computed by using the Gaussian 98 suite of programs[31] with B3lyp/6-31G(d,p) level of theory. Then a modified G3 theory --G3XMP2 was performed during the calculation of the energies of the stable points. This method adds little computer expense(10%-15%)[32] compared to G3, but gives significantly better agreement with experiment and also have advantage that can be used to study potential energy surfaces.

The PES was mapped in the two torsional angles by performing B3LYP/6-31G** constrained optimizations from ϕ_1 = -180 - 180 in steps of 5^0 and ϕ_2= 0 - 180 in steps of 10^0. That is, a total of 1387 constrained optimizations were performed. These 1387 points are actually a data pot. From the pot a contour plot of this two-dimensional PES is given in Figure 3.3. Indicated on this figure are all the minima and saddle points with the corresponding energies given in Scheme 3.1. Table 3.1 includes the energies of the minima at the B3LYP/6-31G(d,p), B3LYP/cc-pVTZ and G3XMP2 levels.

In diffusion Monte Carlo small displacements are made to the Cartesian coordinates of the atoms. The size of the displacements depends on the mass of the nucleus and the imaginary time step size. As imaginary time passes the structure can change dramatically depending on the nature of the potential energy surface. In this work we ran the diffusion Monte Carlo simulations on the B3LYP/6-31G(d,p) surface given in Figure 3.3. In the work reported here we considered explicitly only two torsional angles. In doing so we have assumed that as these angles change the molecule is able to readjust its structure to the most stable form for the given values of ϕ_1 and ϕ_2. Hence after each time step we computed the two values of the torsional angles then reset the remaining internal coordinates (and hence the structure) to an interpolation of the minimum energy structure that corresponded to these two angles. A simple bilinear interpolation was used to obtain the above internal coordinates based on the four sets of optimized internal coordinates, extracted from the grid described previously, that corresponded to the four bracketing pairs of ϕ_1 and ϕ_2. Similarly, a bilinear interpolation was also used to obtain the potential energy of the molecule for the given values of the torsional angles.

Here we specifically implemented discrete sampling diffusion Monte Carlo with 1000 initial replicas. The population was first pre-equilibrated using a step size

of 5 au for 7000 steps. After the pre-equilibration, data sampling then occurred every 50 steps over a period of 10 000 steps using a step size of 1 au. The rotational constants were computed also during this period using the method of descendant weights. Descendants were followed for 1000 steps with a new set of descendants initiated and followed every 100 time steps. The reported results for energies and rotational constants are from 20 separate runs. The reported errors are two standard deviations of the respective means.

In order to compute the rotational constants it is necessary to ensure that the Eckart conditions are enforced. The Eckart conditions minimize the coupling between the vibrational and rotational motions within the molecule. A speedy algorithm was developed to ensure this. We also utilized the molecular symmetry of the system to effectively double the population size in computing the inverted moment of inertia tensor. Note that while we are always in the Eckart axis system, the inverted moment of inertia tensor is not exactly diagonal, except for the reference configuration. However, the absolute value of the off-diagonal elements for all isotopomers of the GM was never more than 31 MHz for the inverted product of inertia about the *a-b* axes, and never more than 6 MHz for the *a-c* and *b-c* axes. Interestingly, isotope $CH_2OH-CDO$ showed a small inverted product of inertia element about the *a-b* axes of 3.5 MHz.

3.3 Results and Discussion

3.3.1 The Ab initio Molecular Structure and Energies of Glycolaldehyde.

Glycolaldehyde exhibits two large amplitude vibrations responsible for the nonrigidity of the molecule: the central C-C bond torsion and the alcoholic hydroxyl torsion. The corresponding independent variables ϕ_1 and ϕ_2 can be identified with the $O_1C_1C_2O_2$ and $H_1O_1C_1C_2$ dihedral angles (see Figure 3.4). The OH torsion cancbe interpreted as the internal rotation of a top (OH) with respect to a fixed frame (COH-CH_2). In contrast, the C-C torsion represents the internal rotation of two quasi-equivalent tops, COH and CH_2OH, with respect to each other.

When going from one conformer to another, the angles CCO and OCH are seen to vary in accordance with the trans-angle rule If the structure is fully optimized at the B3LYP/cc-pVTZ level from different starting points, four conformers, GM (0,0), L1 (180,180), L2 ((197.9,75.8), or (162.1,-75.8)), and L3 (0,180), are found (see Scheme 3.1). Three of them are planar with *Cs* symmetry, whereas the other one is a double minimum. In this case, the molecular plane is lost for minimizing the nonbonding steric repulsions between hydrogens H1 and H3 (see Figure 3.4). Hence, the potential energy surface shows a total number of five holes, two cis minima (L1 and L3) and three trans minima (L1, L2,L2'). The most stable cis geometry (cis-GM) is stabilized by the formation of a 2.0661 Å hydrogen bond between the hydroxyl hydrogen and the carbonyl oxygen. It breaks down during the torsions to produce the secondary minima of which the relative energies are higher than 1200 cm^{-1}. Table 3.1 compares the four conformers energies by different approach, the order is consistent by using B3LYP at 6-31G(d,p) and cc-pVTZ, however by G3XMP2 L1 is lower than L2 in energy as was also noted by Senent. And at G3XMP2, GM, L1, L2, L3 are ordered following an increasing energy criterion. Conformers L1 and L2 lie at 1161.13 and 1223.11 cm^{-1} over the I-cis conformer, whereas the relative energy of the cis-L3 is 1786.72. The planar geometry (ϕ_1, ϕ_2) =(180,180) represents a maximum.
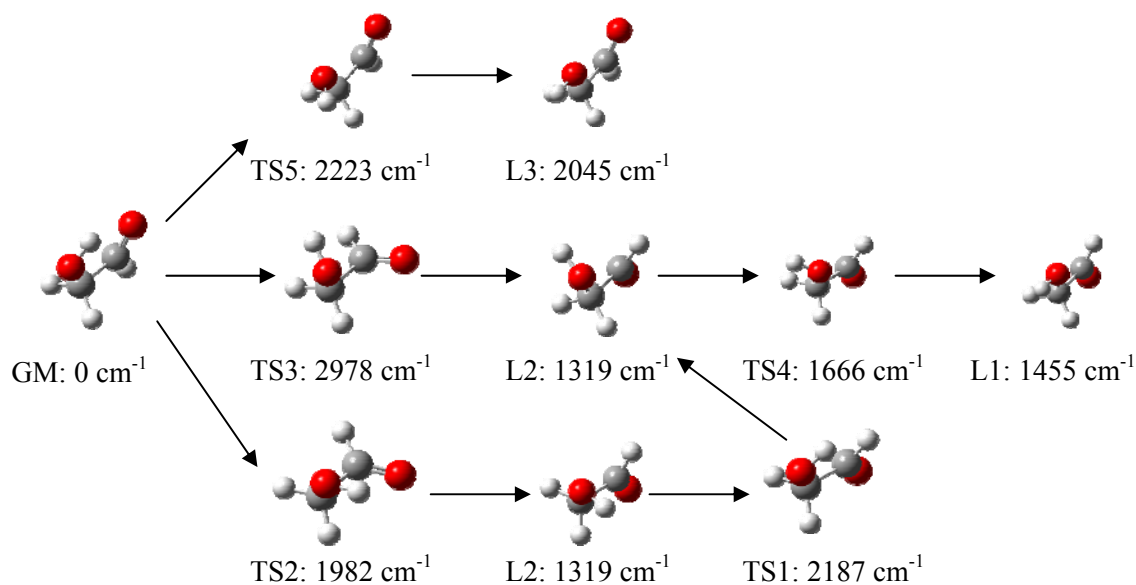
Scheme 3.1 Local minima and transition states for hydroxyacetaldehyde at the B3LPY/6-31G** level of theory.

The OH torsion interconverts trans forms (L1 and L2) of similar stabilities combining Table 3.1 and Scheme 3.1. Their low energy difference (ΔE(B3LYP/cc-pVTZ)) =27 cm^{-1}) impedes definitive assertion of which one of the two geometries is the most stable. With B3LYP/63-1G(d,p) and B3LYP/cc-pVTZ, the most stable form is L2, but if calculations are performed wit G3XMP2, L1 appears slightly more stable (ΔE =62 cm^{-1}) with the zero point correction. This issue happened to other theoretical studies(ref.13). Because the barrier between is very low between the two, we guess that the two cis conformer L1 or L2 can not be isolated in the real or may only one can exist, depending on their energies, further study for ordering the energy by using more advanced method need to be done.

3.3.2 Intramolecular Interaction

It is well known[9,33-35] that there is an intramolecular hydrogen-bond type interaction in the species GM of GA. In this lowest-energy species, the position of the OH group is eclipsed instead of the usual staggered; its ν_{OH} fundamental band is at low frequencies compared, e.g., with methanol and ethanol,[34] the non-bonded OH...O distance is only 206 pm, whereas the sum of the van der Waals radii of H and 0 atoms is 260 pm[9], and the overlap population of this "bond" is considerable.[35] The GM's geometry and the experiment results are listed in table 3.2.

For the energy of this intramolecular H bond, values of 13-26 kJ mol^{-1} [9] or 27 kJ mol^{-1} have been suggested. Since there are attractive and repulsive interactions and there is no suitable reference structure, it is not easy to estimate this energy. Newton and Jeffrey[35] had the structure L3 as the reference structure (φ_O rotation), but there is obviously some repulsion between the lone-pair electrons of two oxygens in this structure; however, the O...O, overlap population is then small (ref.35) and table 3.2). Our 6-31G** energy difference of species GM and L3 is 22.1 kJ mol^{-1} and 19.6 kJ mol^{-1} by B3LYP/6-31G (d,p) and G3XMP2 method seperately. Thus the intramolecular interaction energy might be of the order of 15-21 kJ mol^{-1}.

The shifts caused by the intramolecular interaction on the ν_{OH} band are discussed below, in connection with the spectra. The calculations predict that the intramolecular interaction results in a small lengthening of the OH bond as well as in a decrease of the COH angle (cf. also ref. 35).

3.3.3 Calculated Harmonic Frequencies and Rotation Constants Spectra

Table 3.3 shows the harmonic frequencies of the four conformers of GA calculated with B3LYP/6-31G(d, p) and compared with experimental ones. The ν_{OH} value found for species GM is bit higher than the experimental result, this is maybe due to the intramolecular interaction in this species. The calculated L1- L2 difference is 22 cm^{-1}.

The behaviour of the O=C-C bending band is also interesting: the calculated, reduced value for species GM is 767 cm^{-1}, that for species L1 being 544 cm^{-1} and 527 cm^{-1} for species L2. This shift is exceptionally large, and the calculated values are in excellent agreement with the experimental ones.

For the alcoholic C-C-O bending fundamental band of species GM in vapour, values of 260 ± 40 cm^{-1} been reported.[9] Our calculated value (298 cm^{-1}) is in agreement with the value. Estimates of 195 ± 30 cm^{-1} have been reported[9] for the C-C torsional fundamental of GA, whereas our calculated, value is 192.1 cm^{-1}. Also for the C-C stretch, the predicted value (870 cm^{-1}) is in good agreement with the experimental one (860 cm^{-1}).

Considering symmetry and energy criteria, it may be expected that the aldehydic hydrogen wagging will be the single mode that can interact significantly with the torsions. Table 3.4 shows some characteristic structural and spectroscopic properties. It may be remarked that the cis and trans rotational constants are quite different, whereas the OH torsion is basically responsible for the dipole moment variation. The dipole moment is one the few properties of glycolaldehyde available in the literature.[9,11] For the absolute cis minimum, the B3LYP/6-31G(d,p) calculations lead to a dipole moment value of 2.31 D, which is in great agreement with the experimental one 2.33 D of Marstokk and Møllendal and all are lower than Senet's at MP2, which could mean the geometries we obtained is very close to the equilibrium

structures. There is a reasonable agreement between experiments and calculations. The differences can arise from the failure of the MP2 approximation to describe the electronic distribution between the aldehydic and alcoholic groups. The trans form shows a relatively low dipole moment (1.79 D).

The experimental rotational constants (errors less than the last significant digit given) are compared with the DMC rotational constants (error arises from the random Monte Carlo component of the simulation) for various isotopomers of the GM in Table 3.5. We have also included in this table the values expected for the rotational constants using perturbation theory at the B3LYP/6-31G(d,p) level as implemented in Gaussian 03[36].

The agreement between experiment and the DMC 2D predictions is remarkable (and most likely fortuitous) considering the level of theory used and the two-dimensional approximation. A closer examination of Table 3.5 reveals that the B_0 rotational constant is consistently predicted too low by about 25 MHz, which may indicate the equilibrium structure is marginally too "tight" about this axis at the B3LYP/6-31G(d,p) level. However, the 18 DMC's results differ a bit from the experiment, we relaxed 18 parameters of hydroxyacetaldehyde from its Cs symmetry. It becomes too floppy and the vib rotation movements deviate far from equilibrium geometry. So we did a fit to the equilibrium geometry(cf. Table 3.6), for reducing the effect of strong vib rotation behaviors of C-H and the results improved a lot. The average of errors deviated from experiments lows down from 0.52% to only 0.19%.

A contour plot of the ground state wavefunction is given in Figure 3.5. It is evident from this figure that the hydroxy hydrogen undergoes substantial excursions away from the equilibrium position. Where the probability amplitude falls to half its maximum value ϕ_2 is about $\pm 25^0$, while ϕ_1 travels over $\pm 15^0$. Figure 3.5 also illustrates

bears out a correlated motion between ϕ_1 and ϕ_2 – the intramolecular H-bonded hydrogen tends to follow the electronegative carbonyl oxygen.

3.4 Conclusions

In this study we presented ab initio potential energy surfaces for CH_2OHCHO . These surfaces are based on B3LYP calculations with the6-31G(d,p) basis using Gaussian software package. The GA PES is constructed from 1387 single point calculations at BL3LY/6-31G(d,p) level. Conformational spaces contain critical points, as minima saddle points and maxima. Minima have significance in forming stable conformations, while saddle points show the conformation for transitions form a minimum to another one and also give the activation energy. From Figure 3.6, the positions of a transition form could be predicted and the pathways from one conformer to another through the saddle points are decided. It has been suggested in several studies (cf. ref.37) that intramolecular hydrogen bonds play an important role in the mechanisms of conformer interconversion processes. Conformer GM of GA has a stronger intramolecular hydrogen bond (= 15-21 kJ mol^{-1}) than any other compound we have studied so far.

The visualization of PESs was considered as efficient tools of understanding conformational behaviors of flexible molecules. As the advance of computational technologies made presentation of such surfaces possible, numerous nicely shaped landscapes were reported offering visual adventures. However, as the dimensionality of these surfaces grow with the number of flexible torsional angles the graphical representation the PEHS becomes hardly possible. Furthermore, exploration of such surfaces showed that they cannot reveal more potential energy minima.

DMC is a simulation technique that is valuable and feasible of treating large numbers of coupled degrees of freedom. It has been used to solve lots of electronic

structure problems[38,39] and gains very success in applications to molecular vibrations[40] and intermolecular modes in molecular clusters.[41-43] In its simplest form, this theory readily gives the ground state energy of a quantum mechanical system.

Here both two and eighteen dimensional quantum diffusion Monte Carlo (DMC) calculations were used to study the isomers of hydroxyacetaldehyde. The DMC method that has been provided in which a first principal calculation is shown to be possible for the zero-point energy and vibrationally averaged rotational constants for a eight atom system CH_2OHCHO. The method can, in principle, be applied to much larger systems. The method (a) is completely automated, (b) is fully ab initio, (c) includes all nuclear degrees of freedom, (d) requires no assumptions regarding the functional form of the PES, (e) possesses the full symmetry of the system, (f) is generally applicable to any system amenable to quantum chemical calculations and Collins' interpolation method.

Results were compared with available experimental references. Excellent agreement has been found. A total of four unique minima, and the transition states connecting them, were located. Both two and eighteen dimensional potential energy surfaces were generated and used in the DMC runs. The rotational constants for the global minimum were predicted for all experimentally identified isotopomers and an approximate equilibrium structure obtained by combining our theoretical results with the experimentally observed rotational constants. The results obtained for the remaining isomers indicate that not all of them can be isolated.

3.5 References

1   Ehrenfreund, P., Schutte, W. A., in Astrochemistry: From Molecular Clouds to Planetary Systems, ed. Y.C.Minh & E. F. Van Dishoeck (Chelsea:Sheridan Books), 2000, p 135.

2   Huntress, W. T., Jr.,Mitchell, G. F., J.Astrophys. 1979, 231, 456

3   Kaiser, R. I., Chem. ReV. 2002, 102 (5), 1309.

4   Sorrell, W. H., J.Astrophys. 2001, 129,555,.

5   Wilfred H.S., Ap. J. 2002 ,129,555.

6   Wieland B., Lancaster J.P, Hoaglund C S., Holota P.,and Tornquist WJ., Langmuir, 1996,12, 2594.

7   Galano A, Alvarez-Idaboy JR, Ruiz-Santoy ME, Vivier-Bunge A., J. Phys. Chem. A, 2005, 109 , 169 - 180.

8   Ochando-Pardo M, Nebot-Gil I, Gonzalez-Lafont A, Lluch J.M., J. Phys. Chem. A, 2004, 108 , 5117.

9   Marstokk KM and Møllendal H., J. Mol. Struct., 1973, 16, 259.

10   Marstokk KM and Møllendal H., J. Mol. Struct. 1971, 7, 101.

11   Marstokk KM and Møllendal H., J. Mol. Struct., 1970, 5 , 205.

12   Asplala A, Murto J, Sten P., J. Chem. Phys. 1986,106, 339-412 .

13   Senent, M. L, J. Phys. Chem. A 2004,108 (30) ,6286-6293

14   Ceperley, D. M., Alder, B., J. Phys Rev Lett 1980, 45, 566.

15   Hammond, B. L.,Lester, W. A., Reynolds, P., J.Monte Carlo Methods in Ab Initio Quantum Chemistry;World Scientific: Singapore, 1994.

16   Foulkes, W. M. C., Mitas, L., Needs, R. J., Rajagopal, G., Rev Mod Phys 2001, 73, 33

17   Filippi, C., Umrigar, C. J., J Chem Phys 1996, 105, 213.

18   Kent, P. R. C., Hood, R. Q., Williamson, A. J., Needs, R. J., Foulkes,W. M. C.; Rajagopal, G., Phys Rev B 1999, 59, 1017.

19   Barnett R. N., Whaley K. B., J. Chem. Phys. 1993, 99, 9730.

20   Blume D., Lewerenz M., Huisken F.,Kaloudis M., J. Chem. Phys.1996,   105, 8666 .

21   Anderson B., J. Chem. Phys. 1975, 63, 1499.

22   Anderson B., J. Chem. Phys. 1976, 65, 4121.

23   Anderson B., J. Chem. Phys. 1980, 73, 3897.

24   Anderson B., Freihaut B. H., J. Comput. Phys. 1979, 31, 425.

25   Liu, K., Brown, M. G., Carter, C., Saykally, R. J., Gregory, J. K., Clary, D. C. Nature 1996, 381, 501.

26   Clary, D. C., J. Chem. Phys. 2001, 114, 9725.

27   Bettens, R.P.A., J. Am. Chem. Soc. 2003, 125 , 584.

28   McCoy A.B., Braams B.J., Brown A., Huang X.C., Jin Z., Bowman J.M., J. Phys. Chem. A 2004, 108, 4991.

29   Huang X.C., Braams B.J., Bowman J.M., J. Phys. Chem. A 2006, 110, 445.

30   Huang X.C., Braams B.J., Bowman J.M., J. Chem. Phys.2005, 122, 44308

31   Gaussian 98, Revision A.6 & A.11.2, 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, 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, Inc., Pittsburgh PA, 1998.

32   Curtiss L.A, Redfern C., Raghavachari K., Pople J.A., J.Chem.Phys. 2001,114, Number1

33   Michelsen H., Klaboe P., J. Mol. Struct. 1969, 4 ,293.

34   Jensen H.A., Mellendal H., J. Mol. Struct. 1976, 30 ,145.

35   Newton M.D., Jeffrey G.A., J. Am. Chem. Sot.1977, 99,2413.

36   Gaussian 03, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.

37   Aspiala, T., Lotta, J., Murto,Rsten M., Gupta V.P., Chem. Phys. Letters 1984,109 ,179.

38   Anderson J. B., J. Chem. Phys. 1975, 63, 1499.

49   Ceperley D. M., Adv. Chem. Phys. 1996, 93 .

40   Suhm M. A., Watts R. O., Phys. Rep. 1991, 204, 293.

41   Dykstra C. E., VanVoorhis T. A., J. Comput. Chem. 1997, 18, 702.

42   Buch V., J. Chem. Phys. 1992, 97, 726.

43   Gregory J. K., Clary D. C., J. Phys. Chem. 1996, 100, 18014.

3.6 Appendix

Table 3.1 Relative energies (cm^{-1}) of minima at the various levels of theory.

Minima	B3LYP/63-1G(d,p)	B3LYP/cc-pVTZ	G3XMP2
GM	0	0	0
L1	1455	1301.87	1161.13
L2	1319	1274.62	1223.11
L3	2045	1875.59	1786.72

Table 3.2 Structure comparison between experiment and theoretical calculation.

	Expt.	Calculated
C=O	1.2094±0.0003Å	1.210 Å
C-O	1.4366±0.0007Å	1.3977 Å
C-C	1.4987±0.0004Å	1.6102 Å
C-H	1.1021±0.0003Å	1.1045 Å
O…H	2.0069±0.0004Å	2.060 Å
O…O	2.6974±0.0004Å	2.670 Å
∠C-C=O	122º44'±2'	121.405º
∠C-C-O	111º28'±2'	111.930º
∠C-C-H	109º13'±1'	107.620º
∠H-C-H	107º34'±2'	105.840º
∠H-C-O	109º39'±1'	111.740º
∠O-H…O	120º33'±2'	118.490º
∠H…O-C	83º41'±1'	83.090º
Dipole μ	2.34±0.01D	2.3143D

(Note: the experiment data were obtained from substitution structure, and the calculation was done at B3LYP/6-31G(d, p) level of theory)

Table 3.3 Harmonic frequencies of the glycolaldehyde conformers (in cm-1) calculated with B3LYP/6-31G(d, p)

assignment	GM	expt (a)	L1	L2	L3
	A′		A′	A′	A′
C-C-O bend	298.5	260 ± 40 [9]	320.9	337.1	263.9
O=C-C bend	767.4	751.6	544.2	527.8	726.4
C-C stretch	870.9	860.4	995.9	1032.2	862.7
C-O stretch	1141.7	1114.6	1107.2	1100.3	1149.3
H-C-O bend	1308.5	1275.3	1226.9	1208.8	1226.7
O-C-H bend	1402.77		1382.4	1343.5	1426.2
H_2C twist	1451.6		1438.1	1404.5	1447.4
H-C-H bend	1497.2		1500.3	1474.9	1502.7
C=O stretch	1815.4	1754.2	1842.7	1836.9	1855.5
C-H stretch	2949.4	2708.2	2924.9	2893.8	2873.9
CH_2 stretch	2976.8	2832	2986.5	3023.7	2946.9
O-H stretch	3697.0	3548.8	3842.6	3819.6	3833.9
	A″		A″	A″	A″
C-C tor	192.1	195 ± 30 [9]	67.1	86.1	172.2
O-H tor	406.3		226.6	318.1	255.2
H wag	728.2		735.0	717.0	734.3
CH_2 twist	1112.5		1108.6	1064.8	1112.8
CH_2-C bend	1246.3		1240.2	1394.0	1255.4
CH_2 stretch	2987.5	2881.1	3018.9	3076.1	2965.5

a From ref (10); some experimental values have been reassigned

Table 3.4 Under total 10m Torr pressure, each conformer's pressure and total dipole moment vs temperature.

	298.15K	323.15K	373.15K	Total Dipole
Conformer	P (mTorr)	P (mTorr)	P (mTorr)	(Debye)
GM	9.59	9.27	8.41	2.31
L1	2.96E-01	5.14E-01	1.13	2.97
L2	1.05E-01	2.03E-01	4.35E-01	1.79
L3	5.19E-03	9.87E-03	2.77E-02	4.05

Table 3.5 Experimental rotational constants for parent and isotopomers of the global minimum isomer compared with the perturbation theory and DMC 2D and 18D

constants. The former two were computed at the B3LYP/6-31G (d,p) level, all values are in MHz.

	Experiment [a]	Exp. – Perturb	Exp.–DMC(2D)	Exp.-fitted(18D)
GM				
A	18446.4	106.5	-0.5 ± 1.0	-2.8 ± 50.2
B	6526.0	-52.8	25.1 ± 0.5	14.0 ± 38.2
C	4969.3	-22.3	3.7 ± 0.2	3.4 ± 19.4
$CH_2OD\text{-}CHO$				
A	17490.8	68.8	-3.4 ± 0.7	-8.0 ± 44.6
B	6499.8	-44.8	29.9 ± 0.4	-6.3 ± 35.5
C	4883.0	-18.9	4.7 ± 0.1	-9.8 ± 16.5
$CH_2OH\text{-}CDO$				
A	17151.3	103.0	16.0 ± 0.7	171.3 ± 51
B	6363.0	-47.7	26.5 ± 0.4	7.2 ± 29.7
C	4779.0	-18.3	4.1 ± 0.1	7.1 ± 19.6
CHDOHCHO				
A	16988.0	104.1	14.5 ± 1.2	143.4 ± 5.3
B	6385.5	-51.0	20.5 ± 0.4	-6.9 ± 38.7
C	4843.8	-20.3	6.5 ± 0.2	0 ± 26.1
$^{13}CH_2OH\text{-}CHO$				
A	18126.9	88.5	-13.9 ± 0.7	122 ± 47.3
B	6487.5	-49.9	26.5 ± 0.5	-2.3 ± 30.5
C	4923.0	-22.4	2.8 ± 0.2	-1.6 ± 19.9
$CH_2OH\text{-}^{13}CHO$				
A	18259.5	109.4	4.5 ± 0.7	-0.4 ± 55.4
B	6472.3	-51.9	24.1 ± 0.4	9.1 ± 31.1
C	4924.6	-21.6	3.6 ± 0.2	1.2 ± 14.7
$CH_2OH\text{-}CH^{18}O$				
A	18087.0	101.5	-0.8 ± 0.8	-16.7 ± 41.5
B	6242.8	-49.0	24.8 ± 0.4	5.5 ± 23.8
C	4778.5	-21.3	4.4 ± 0.1	2.3 ± 11.9

(a) From ref.6

Table 3.6 Comparison parameters of the fitted structure and reference structure Ref 6.

	Expt. r_s structure	Fitted r_e structure
C=O	1.2094±0.0003Å	1.2106 Å
C-O	1.4366±0.0007Å	1.3937 Å
C-C	1.4987±0.0004Å	1.5079 Å
O-H	1.0510±0.0004Å	0.9712 Å
H_{ald}-C	1.1021±0.0003Å	1.1058 Å
H_{alc}-C	1.0930±0.0003Å	1.0897 Å
∠C-C=O	122º44'±2'	121 º26'
∠C-C- H_{ald}	115º16'±2'	116 º52'
∠C-C-O	111º28'±2'	111 º59'
∠C-O-H	101.34º±2'	105 º19'
∠C-C- H_{alc}	109.13º±1'	107 º50'
∠H-C-H	107º34'±2'	104 º56'
∠H-C-O	109º39'±1'	111 º56'

GM(Global Minimum)

L1(Local minimum 1)

L2 (Local minimum 2)

L3 (Local minimum 3)

Figure 3.1 The structures of four conformer of glycolaldehyde
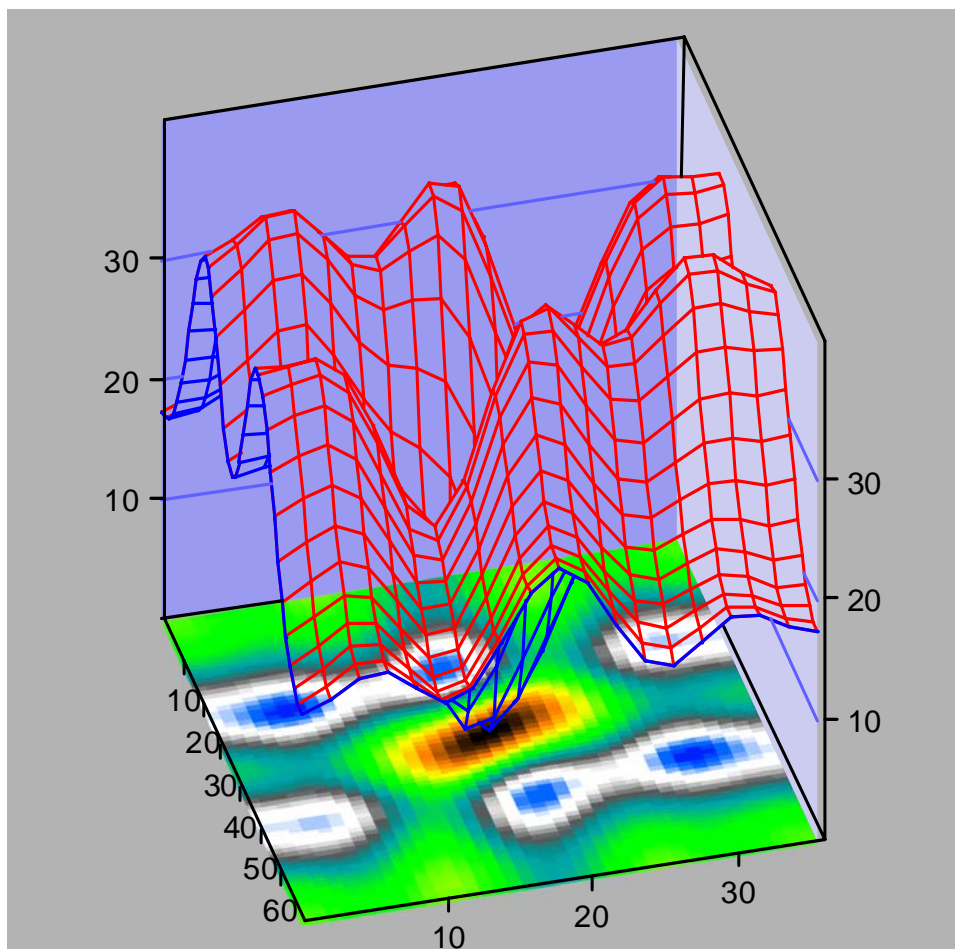
**Figure 3.2** The 3D potential energy surface of Glycolaldehyde(in KJ/mol).
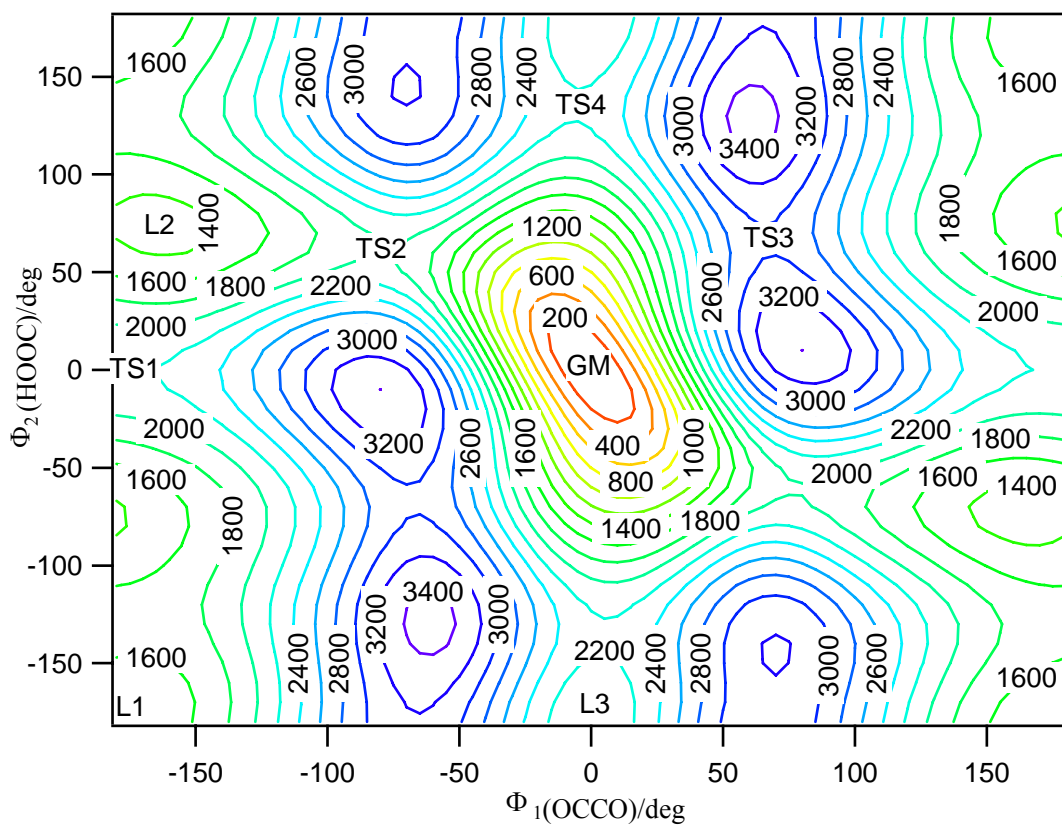
Figure 3.3 Contour plot of hydroxyacetaldehyde (energies in cm^{-1}) as a function of the two torsional angles
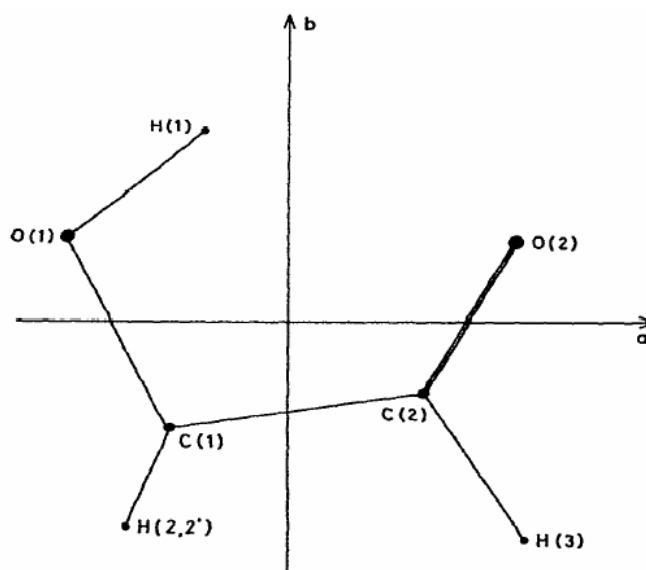
Figure 3.4 Projection of stable conformer of glycolaldehyde in the a-b principal axes plane.
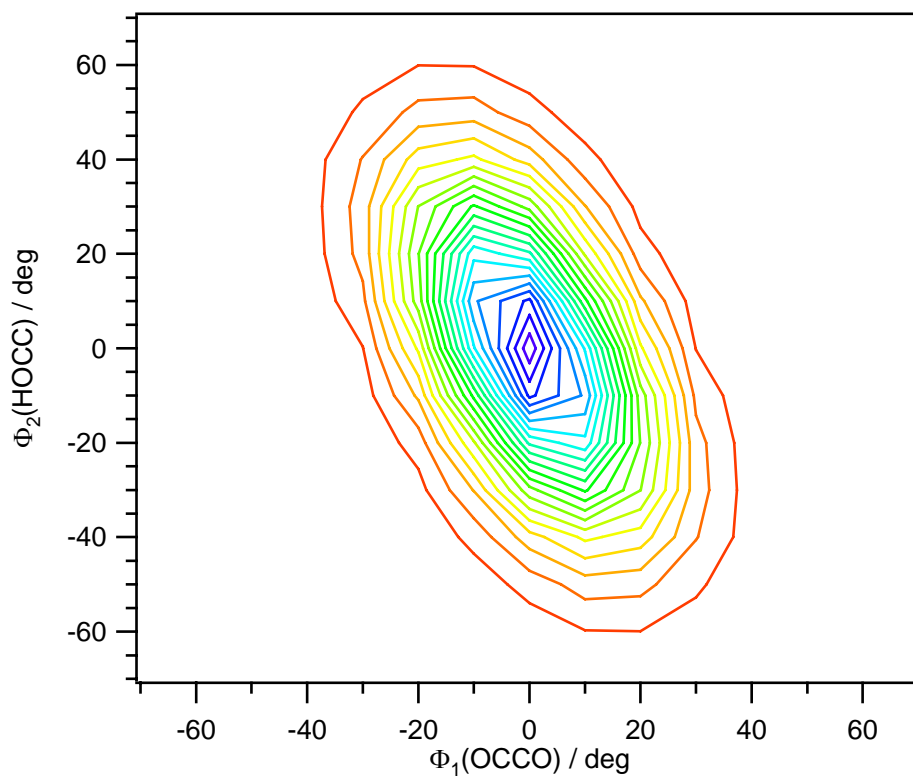
Figure 3.5 Contour plot the ground state wavefunction for the global minimum. Each contour represents a fall of about 5% in probability amplitude.
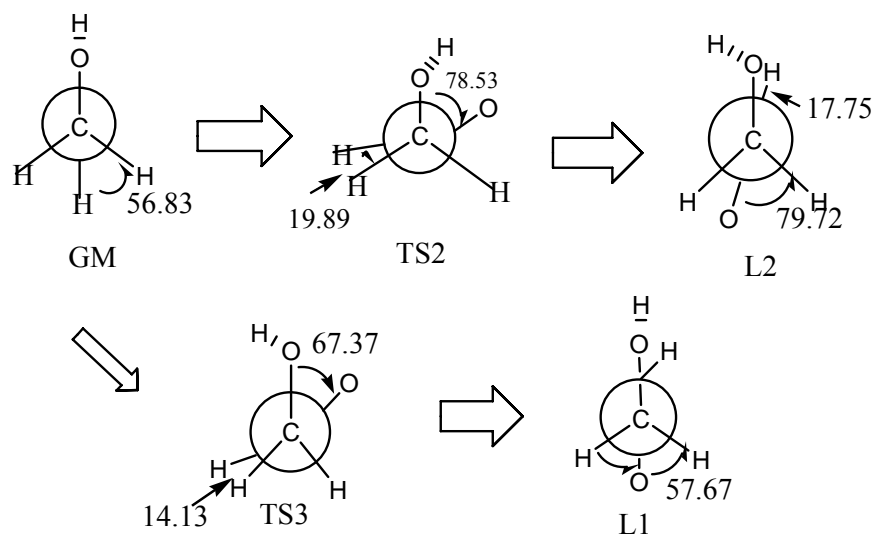
Figure 3.6 Example of Transformation of Glycolaldehyde

Chapter 4 A study of the Shuttling of a Rotaxane-Based Molecular Machine Device

4.1 Introduction

4.1.1 Definition of Molecular Machine

A machine is defined as "an assembly of parts that transmit and modify forces, motion, and energy one to another in a predetermined manner".[1] When the word "parts" is replaced by "molecules", a machine turns into a molecular or supramolecular machine. Therefore, a molecular-level machine can be defined as an assembly of a discrete number of molecular components designed to perform machine-like movements in response to an appropriate external stimulus. In addition, a molecular machine has features characteristic of the molecules. So it is not surprising that artificial molecular machines and motors appear as one of the emerging fields of chemistry in the past decade.[2-5] In biology, many "molecular motors" or "machines" play an essential role. Many examples are known of proteins which undergo important shape modifications (folding-defolding) after a signal has been sent to the protein[6] and it is fascinating to see how nature has found elegant solutions to control movement at the molecular level by the conversion of chemical energy into mechanical energy.[7]

4.1.2 Application of Molecular Machine

The extension of the concept of a machine to the molecular level is important not only for basic research but also for the growth of nanoscience and nanotechnology. The miniaturization of components for the construction of useful devices, which is an essential feature of modern technology, is currently pursued by a large downward (top-down) approach. This approach, however, which leads physicists and engineers to

manipulate progressively smaller pieces of matter, has its intrinsic limitations. An alternative and promising strategy is offered by the small-upward (bottom-up) approach. Chemists are already at the bottom, since they are able to manipulate molecules (i.e., the smallest entities with distinct shapes and properties) and are therefore in the ideal position to develop bottom-up strategies for the construction of nanoscale machines. The idea of constructing artificial molecular-level machines is quite recent. This topic was briefly discussed for the first time by Richard P. Feynman, Nobel Laureate in Physics, in his famous address, "There's Plenty of Room at the Bottom", to the American Physical Society in 1959, and only in the past few years have systematic studies been performed in this field. [8,9]

4.1.3 Type of Energy Supply and Requirement for Constructing a Molecular Machine

To make a machine work, energy inputs have to be supplied, a chemical system is through an exergonic chemical reaction.There are usually three kinds of energy resources: chemical energy (from chemical reaction); electricity energy (battery, voltage from outside); light energy (laser or other resource). If a molecular-level machine works by inputs of chemical energy, it will need addition of fresh reactants at any step of its working cycle, with the concomitant formation of waste products.[3] Accumulation of waste products, however, will compromise the operation of the machine unless they are removed from the system. The need to remove waste products introduces noticeable limitations in the design and construction of artificial molecular-level machines based on "chemical fuel" inputs. Photochemical energy inputs offer other advantages compared to chemical energy inputs. For example, they can be switched on and off easily and rapidly. A system is exclusively governed by light excitation, a "light-fueled" motor, but without generation of any waste product. A

further advantage offered by the use of photochemical techniques is that photons besides supplying the energy needed to make a machine work can also be useful to "read" the state of the system and to control and monitor the operation of the machine.

Light-powered molecular machines are conjectured to be essential constituents of future nanoscale devices.[10] Nanomechanical devices or molecular machines will, for a broad range of applications, most likely be powered by light or other kinds of electromagnetic radiation. The dominant reasons are ease of addressability. The light-driven molecular machine can undergo repetitive movement. The effort in my project led ultimately to the light-driven molecular machine, which can undergo repetitive movement in a unidirectional manner.

The construction of stimuli-responsive molecular devices and molecular machines is a great challenge. The construction of molecular shuttles in which a ring moves back and forth like a shuttle between two or more "stations" in response to external stimuli is one target.[11-16]

4.1.4 Pseudorotaxanes, Rotaxanes, and Catenanes

Since the molecular-level machines are mostly based on pseudorotaxanes,[17-19] rotaxanes,[20] and catenanes.[21] Rotaxanes are compounds in which a ring is threaded by a linear chain bearing bulky end units (cf. Figure 4.1).Such compounds have been made long ago.[22,23] Recently, rotaxanes underwent a real revival due to the newly developed efficient procedures that make them relatively easy to prepare[24-26] and also because of their electro- and photochemical properties[27,28] and their aptitude to undergo controlled molecular motions.[29-31] The rotaxane described here contains two different recognition sites in the dumbbell component, it is possible to switch the position of the ring between the two "stations" by an external stimulus.

Mechanically interlocked molecules, such as rotaxanes [20] and catenanes,[21] are suitable candidates for the construction of molecular machines, and they are now studied as a type of molecular motors and machines. Many kinds of molecular components can be used to construct molecular-level machines. Crown ethers,[28] cyclophanes,[32] and calixarenes,[33,34] have been extensively used as cyclic components in the construction of the supramolecular structures.

4.1.5 Cyclodextrin

Rotaxanes was often considered to be a typical prototype of molecular machines, because they have a rotor and an axle in the molecule.[35] Choosing a proper candidate molecular to be used as cyclic components in the construction of various supramolecular architectures, we consider these cyclodextrin molecules as promising molecular components for the construction of molecular-based machines because they have a rigid, well-defined ring structure and an ability to bind various low-molecular-weight compounds, and to be threaded onto a long axle and to slide along a chain or to rotate around an axle.

Cyclodextrins (CDs) are cyclic oligomers of α-D(+)-glucose units formed by the action of certain enzymes on starch. The glucose units of cyclodextrins are connected through glycosidic α-(1,4) linkages. In general, a cyclodextrin molecule can be briefly described as a torus, but is somewhat more realistically pictured as a shallow truncated cone possessing multiple stereogenic centers with a partially blocked base, a hydrophobic interior cavity and hydrophilic edges due to the presence of hydroxyl groups. All the glucose units in this toroidal structure are in their chair-conformation.[36] The structure of cyclodextrins is depicted in Figure 4.2; three CDs are commonly available: α-, β-, and γ-CD (Figure 4.2a), with six, seven, and eight glucose units,

respectively. The glucose units are connected through glycosidic α -1,4 bonds, as illustrated in Figure 4.3 .

The CD molecule has secondary 2- and 3- hydroxyl groups lining at the mouth of the cavity and primary 6-hydroxyl groups at the rear end of the molecule (Figure 4.2b).

The interior cavity of CD is lined with glucosidic oxygen linkages, while the nonbonding electron pairs of the glycosidic oxygen bridges are directed toward the inside of the cavity, producing a high electron density and lending it some Lewis base character. As a result of this special arrangement of the functional groups in the CD molecules, the cavity itself is relatively hydrophobic and permits inclusion of hydrophobic portions of solute molecules. Meanwhile, the C-H forms two rings at the outer rims of the cyclodextrin, which are rather hydrophilic.[36,37] In cyclodextrin molecules, a ring of hydrogen bonds is also formed intramolecularly between the 2- and 3-hydroxyl groups of adjacent glucose units, which gives the CD a remarkably rigid structure.[38-42] Table 4.1 lists some interesting properties of the native CDs. The cavity dimensions given in the table are approximate, other properties have been measured, such as rate of acid hydrolysis,[43] molar volumes,[44] activity coefficients,[45] diffusion coefficients,[46] and solubility in dimethylformamide.[47]

As shown in Table 4.1, the physical properties of the three CDs are quite different[52-56]. The cavity diameter narrows on proceeding from the secondary hydroxyl rim to the primary hydroxyl rim, and of course within the cavity the van der Waals radii of the oxygens and hydrogens contribute further variability.[57] The table also tells that β-CD being (apparently) an anomaly among the CDs. The thermodynamics of solution show that the relatively low solubility of β-CD is associated with both a less favorable $\Delta H°$ and a less favorable $\Delta S°$. These solubility results are one piece of information

suggesting that the CDs may not universally behave as a monotonically graded series. Low solubility of β-CD is due to the interruption by aggregated β-CD, with its 7-fold symmetry, of the hydrogen-bond structure of water, the even symmetries of α-CD and γ-CD not behaving in this way.

CDs are able to be regarded as "hosts" for "guest" molecules capable of entering (in whole or in part) the cavity and forming noncovalent host-guest inclusion complexes. Almost all applications of CDs involve complexation. A combination of weak forces enables the complexations such as hydrophobic interaction, electrostatic interaction, van der Waals interaction, hydrogen bonding, and dipole-dipole interaction, etc. For example if a ligand for inclusion into CD is dissolved in water and CD is added, complex formation occurs if the ligand L as a whole or a part can ben enclosed in the cavity.

$$CD + L = CD·L$$
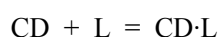
The complexation can be followed by some suitable signal. If the ligand had an absorption spectrum in the visible region of light , there's generally a shift in the spectrum which can be monitored by spectrophotometers.If the ligand is achiral, inclusion by the chiral CD gives rise to an anisotrophy which can be measured by circular dichroism.Other spectroscopic methods to follow the equilibrium are NMR-spectroscopy,calorimetry and competition experiments where one ligand that cannot be measured directly substitutes a ligand in the cavity which can be easily detected.[58]

Therefore the efforts in this study we choose α-cyclodextrin (α-CD) (See Figure4.4) was chosen as a cyclic component that led ultimately to the light-driven molecular machine, which can undergo repetitive movement in a unidirectional manner. This light driven stilbene-cyclodextrin rotaxane system was computationally investigated by semi-empirical AM1 calculations. The results well explain the

experiment findings.[59] The configuration that is energetically preferred is the one in which the α-cyclodextrin is bound to the middle of trans- form of stilbene. After being subjected to a laser irradiation, the trans- isomer isomerizes to its cis form, the α-cyclodextrin (α-CD) moves in one direction along the stilbene shaft.

4.2 Computational Methods

This rotaxane CD-stilbene system here is containing 184 atoms:108 heavy atoms and 76 protons. The shaft has 58 atoms (42 heavy atoms).While there are 126 atoms in the α-CD shuttle (66 heavy atoms). Such a huge system makes a daunting computational challenge for ab initio methods. Because ab initio techniques are more preferable to study small molecules or simple systems.To handle the formidable computational problem, semiempirical methods AM1[60] is a reasonable alternative choice. Additional support is from experience, quite a lot of successes were obtained in treating this kind of inclusion phenomena with semiempirical methods.[61-63] Here AM1 was chosen by its computational efficiency, well studied for complexes only containing H,C,N,O,Na atoms and shows advantages to other semiempirical methods like MNDO and PM3.AM1 reproduces better results for the steric interactions and gas-phase transition-state energies than MNDO.[64-66] In addition , in contrast to the MNDO and PM3 methods, AM1 has the ability to reproduce the intromolecular and intermolecular hydrogen bonds,[67] which are abundantly presents in the CD and is the dominate interaction between shaft and shuttle.[68]

Firstly, the initial CD, stilbene and CD- stilbene complex structures were generated with the Hyperchem software package.[69] To make give reasonable starting structures and save computer time for the optimization with AM1 method, preliminary optimizations were done on all structures with the MM+ molecular mechanics

method.[70] After the first step, full geometry optimizations, were carried out with AM1 method using Gaussian 03[71] set of program. The UV predictions of this rotaxane were performed at AM1 by VAMP model in the Materials Studio Modelling software.[72]

4.3 Results and Discussion

Molecular-level machines operate via electronic and nuclear rearrangements, i.e., through photoisomerization reactions. Basic requirements rational for designing a macular machine should be met. (i) input energy supplied to make them work, (ii) the way in which their operation can be controlled and monitored, (iii) the possibility to repeat the operation at will, (v) the time scale needed to complete a cycle of operation, and (vi) the function performed.

To control and monitor the machine operation [point (ii)], the motions of the component parts should cause readable changes in some properties of the system; any kind of chemical or physical techniques can be useful, particularly the various types of spectroscopies. Since a machine has to work by repeating cycles [point (iii)], an important requirement is that any chemical reaction taking place in the device has to be reversible.

Our stilbene-cyclodextrin rotaxane system had two different recognition sites in the dumbbell component; middle site of shaft is an excellent binding site in the ground state. Meanwhile on one end of shaft is a poor holding site except when stilbene turns from trans-form to cis-form via excitation by an external stimulus. So it is possible to switch the position of CD between the two "stations" UV light irradiation occurred during the interconversions between trans- and cis- .UV spectroscopy and NMR could be used to monitor whole process. The calculated optical UV spectra agreed well with experiment results, also proved the shaft undergoes an intersystem cross to the triplet

state. On the basis of these results, mechanism for this photoinduced shuttling of stilbene-cyclodextrin rotaxane is likely to occur first by motion of the α-CD to one end of the stilbene, then the photoisomerization of the stilbene to localize the α-CD at this end.

4.3.1 Geometry

A cyclodextrin (CD) is a cyclic oligomer of α-Dglucose.According to Connors's view,[73] cyclodextrins are fairly flesxible. It is really not easy to construct the structure. And to optimize α-CD often took more or less one week to complete because of high mass weight of α-CD.So the preliminary step is necessary.

The trans conformational isomer of stilbene was generated with GUI editor.Cis-isomer was obtained by rotating the C-C=C-C torsional angle in to apporopriate orientation.Then optimizing them with MM+ molecular mechanics method. With pre-optimized structures,we firstly optimize the two monomer components in free space with AM1.The structures of α-CD ,trans and cis stilbene conformers are shown in Figure 4.5.

The CD molecule is often described as a torus, but is somewhat more realistically pictured as a shallow truncated cone, the primary hydroxyl rim of the cavity opening having a somewhat reduced diameter compared with the secondary hydroxy rim. This cavity makes the CDs attractive subjects for studying inclusion complexes. Complex formation occurs if the molecular as a whole or in part can be enclosed in the cavity as show in Figure 4.6.

To make the formation of an inclusion complex between α-CD and stilbene, optimized shaft stilbene was inserted to the shuttle. Optimization is to look for the lowest energy of the rotaxane system, by inserting shaft through the centre of α-CD

conformer, the widest cavity produced a stable trans conformer of the rotaxane. Similarly, a stable cis rotaxane conformer was produced by putting the shuttle with secondary rim nearer to the end of shaft. The optimized rotaxane structures are show in Figure 4.7.Both of the two isomers lost any symmetry because of the presence of α -cyclodextrin, this phenomenon was also detected by experiment.[74]

4.3.2 Energy and Photoisomerization Barrier

For the trans conformer of stilbene, the most stable configuration is when the shuttle binds in the middle of the shaft. In the ground state, any other part of the shaft is poor site for binding with α-CD. While when the trans converted to cis conformer, the global energy minimum move to the site –one end of stilbene which oriented to the secondary rim of CD. This process is repeatable and can be controlled by a remote laser. The energy differences between the two isomerization was found to be 85.6 kJ mol^{-1}.Very interesting from Table it can also be found that the sign of Bonding Energy(BD) is opposite, BD for trans configuration is around -75.2 kJ mol^{-1} mol ,BD for cis configuration is 73.3 kJ mol^{-1}, the difference is near 150 kJ mol^{-1}. This indicated the inconversion of system is not that easy, to make it happen a strong energy resource is needed.

To obtain the potential energy curve for the conformational isomerization of the rotaxane from trans to cis, a series of constrained partial optimization were carried out. Starting with trans conformer, the C–C=C–C dihedral angle (θ) was stepped in 10° increments and all other internal coordinates optimized at each step. Unfortunately, the HF-SCF procedure failed to converge for θ < 110°; at the same time, the same procedure was followed from cis conformer. The HF-SCF procedure failed to converge for θ > 80°. I have tried several methods to force the SCF convergence, but unsuccessfully they all didn't work. The resulting potential energy curve, though incomplete, it gives some

insight into the conformational isomerization process, and is shown in Figure 4.8. From the figure it is clear that the barrier to isomerization is at least 40 kcal/mol, and possibly considerably higher. This large energy barrier, coupled with the significant structural transformation make the isomerization process considerable hard.

Another possible energy potential curve is plotted by the rout of moving the shuttle to the end first and increasing up C–C=C–C dihedral angle (θ) of cis conformer with 10° for each step. The potential curve (Figure 4.8 A) has a low barrier which is at least 8 kJ mol^{-1} lower compared to Fig 4.8 B. Seemly, the second route is least to happen.

4.3.3 UV Spectra

Governed by light energy, molecular shuttle in which α-CD moves back and forth between the mid and end site of shaft accompanying stilbene rotating around the C=C double bond. Upon Carol A. Stanier .co workers' results, irradiation with 340-nm light of a aqueous solution containing α-CD and stilbene causes photoisomerization of trans to cis, causing a general decrease in absorbance, and an increase in the absorption at 265. My calculated UV data are to be at 300nm and 261nm respectively for the trans, cis configurations. No counter anions or solvation effects were included.

Table 4.3 lists the UV spectra from experimental measurement and theoretical predicted. Compared to the experiment, the predicted data 261 nm is rather good while the other one for trans configuration is not that close to the experimental findings, with a 40 nm differences. That is because optical spectrum depends strongly on the structure of the system as well as the solvent effects. My system is totally in the gas phase. Accounting on this point, the prediction is satisfied.

4.3.4 Mechanism

Cyclodextrin (CD) is known to form inclusion complexes. This stilbene cyclodextrin rotaxane system revealed to have functioned as a light-driven molecular shuttle and are stabilized more greatly by trans-stilbene base rotaxane than by cis one. Light irradiation induces the isomerization.

Carol A. Stanier et al suggested two mechanisms for photoinduced shuttling in a photoisomerizable rotaxane, which are show in Figure 4.9. Simulated process for this molecular machine model comprised of two parts, isomerization of stilbene (trans to cis) and the movement of the shuttle. The main difference between the two routes is which part happens first, in a word, the sequence. The UV spectra, and comparison of potential curves between two isomerization routs, both of them support that route 2 is much more preferable. According to steric effect, the size of the cavity and the height of α-CD, it can't reach to time when photoisomerization completed and start to move. It rather moves off the center of shaft, and then being blocked there by isomerized shaft. The machine was localized there and waiting to set out until the light signal appeared, the cis conformations turns back to trans conformation. This mechanism clearly describes the way how a molecular machine operates and may apply to other similar photoisomerizable rotaxane system.

4.4 Conclusion

The field of molecules in motion, for which movements and shape changes are triggered and controlled from the outside, has indisputably been one of the most rapidly developing areas of the past decade. CD based rotaxane photonic process represents an interesting class of such compounds, due to theirs unique physic co chemical properties. In this account, a modeling was set up to simulate and describe a rotaxane system. In our

computational study of stilbene-cyclodextrin rotaxane with photoinduced isomerization which was previously demonstrated experimentally, we investigated the characteristic of configurations, simulated the process of the movement of shuttling and explain the possible mechanism. All the AM1 semi-empirical electronic structure calculation results demonstrate that the preferable formation of rotaxane structure with α-CD and stilbene. Stilbene which holds $C2$ symmetry axis loses the symmetry after the formation of rotaxane systems completed. External light irradiation onto the complex caused the photoisomerization from the trans to the cis configuration of the stilbene unit in the meantime α-CD moved to one end of the shaft. In another words, the rotaxane stilbene-cyclodextrin rotaxane functions as a light-driven molecular shuttle in which α-CD machine moves back and forth from the middle station to another station by alternating photoirradiation of light or behaves like a switch changing between ON and OFF by laser or other light resources. This finding is also supported by the experimental work.[59]

On the basis of these results, the mechanism is provided: the photoinduced shuttling in this system is likely to occur first by motion of the α-CD to one end of the stilbene, then the photoisomerization of the stilbene to localize the α-CD at this end. The mechanism is different from other macroscopic analogues [11,12,15,74-78] by random walks, it is unidirectional movement.

The present results also suggested that semi-empirical techniques especially AM1methods may be of considerable value in the analysis of molecular machines or molecular device, where large molecular systems and large amplitude mechanical motions make ab initio methods computationally unfeasible. These arguments are also consistent with other researcher's.

4.5 References

1   Balzani, V., Credi, A., Raymo, F. M., Stoddart, J. F., Angew. Chem., Int. Ed. 2000, 39, 3348-3391;Yamaguchi, H., Kamachi, M. H,, Angew. Chem., Int. Ed. 2000, 39, 3829-3831.

2   Fabbrizzi, L., Licchelli, M., Pallavicini P., Acc. Chem. Res 1999,32, 846- 853.

3   Kelly, T. R., De Silva, H., Silva, R. A., Nature 1999.401,150-152.

4   Koumura, N., Zijlstra, W. J., Van Delden, R. A.,Harada, N., Feringa, B. L, Nature 1999,401,152-155.

5   Sauvage, J.-P., Acc. Chem. Res Res 1998, 31: 611-619.

6   SchenckH.L.,Dada,G.P.,Gellman,S.H.,J.Am.Chem.Soc.1996,118:12487-12494. Dado, G. P., Gellman S. H., J. Am. Chem. Soc. 1993, 115, 12609-12610.;Bixler, J.,Bakker,G.,McLendon,G.J. Am. Chem. Soc. 1992, 114,6938-6939.Thirumalai, D.,Woodson, S. A. Acc. Chem. Res. 1996, 29,433-439. Pascher, T.,Chesick, J. P. Winkler, J. R., Gray H. B., Science 1996, 271, 1558-1560; Jones C. M.,Henry, E. R., Hu, Y., Chan, C.-K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., Eaton W. A. Proc. Natl. Acad. Sci. U.S.A. 1993, 90,11860-11864.

7   Special IssuesMovement: Molecular to Robotic. Science 2000, 288, 79-106.

8   Balzani,V.,Gomez-Lo pez, M., Stoddart, J. F., Acc. Chem. Res. 1998, 31, 405-414

9   Balzani,V.,Credi, A., Venturi, M., Molecular-Level Devices. In Supramolecular Science; Ungaro, R., Dalcanale, E., Eds., Kluwer: Dordrecht, 1999, 1-22.

10   Ballardini,R.,Balzani,V.,Credi,A.,Gandolfi,M.,T.,Venturi,M.,Acc.Chem.Res.2001,3 4,445-455.

11   Amabilino, D. B., Stoddart, J. F., Chem.ReV. 1995, 95: 2725.

12   Anelli, P. L., Spencer, N. ,Stoddart, J. F., J. Am. Chem. Soc 1991, 113,5131

13   Ashton, P. R., Johnston, M. R., Stoddart, J. F., Tolley, M. S., Wheeler, J. W., J. Chem. Soc., Chem. Commun. 1992, 1128-1131.

14   Benniston, A. C., Harriman, A., Angew. Chem., Int. Ed. Engl 1993, 32,1459.

15   Bissell, R. A., Co´rdova, E., Kaifer, A.E. ,Stoddart, J. F. , Nature 1994, 369,133.

16   Philp, D., Stoddart, J. F., Angew. Chem., Int. Ed.Engl.1996,   35: 1154-1196.

17   For reviews of pseudorotaxanes and rotaxanes see: Carlucci, L., Ciani, G., Proserpio, D. M. Coordin. Chem. Rev. 2003, 246, 247-289.

18   Grabuleda X., Ivanov P., Jaime C., J.Org. Chem., 2003,68 , 1539 .

19   Sarah J. Vella, Jorge Tiburcio, James W. Gauld, and Stephen J. Loeb, Org. Lett., 2006, 8,3421 -3424.

20   Harada, A. Adv. Polym. Sci. 1997, 133, 141-191. Nepogodiev, S. A., Stoddart, J. F., Chem. Rev. 1998, 98, 1959-1976.; Wenz, G., Angew. Chem., Int. Ed. Engl. 1994, 33, 803-822.

21   Sauvage, J.-P., Dierich-Buchecker, C., Molecular Catenanes, Rotaxane and Knots; Wiley-VCH: Weinheim, 1999.

22   Harisson, I. T., Harrison, S., J. Am. Chem. Soc. 1967, 89, 5723

23   Schill, G., Zollenkopf, H., Rotaxan-Verbindungen, I. Liebigs Ann. Chem. 1969, 721, 53-74.

24   Belohradsky, M., Raymo, F. M., Stoddart, J. F. Template-Directed Syntheses of Rotaxanes. Collect. Czech. Chem. Commun. 1996, 61, 1-49.

25   Gibson, H., Bheda, M. C., Engen, P. T. Rotaxanes, Catenanes, Polyrotaxanes, Polycatenanes and Related Materials. Prog. Polym. Sci. 1994, 19, 843-945.

26   Ogino, H., J. Am. Chem. Soc. 1981, 103, 1303-1304.

27   Ballardini, R., Balzani, V., Gandolfi, M. T., Prodi, L., Venturi, M.; Philp, D., Ricketts, H. G., Stoddart, J. F., Angew. Chem., Int. Ed. Engl. 1993, 32, 1301-1303.

28   Gokel, G. W., Crown Ethers and Cryptands. In Large Ring Molecules; Semlyen, J. A., Ed.; Wiley: Chichester, 1996.

29   Collin, J.-P., Gavina, P., Sauvage, J.-P., Chem. Commun. 1996, 2005-2006.

30   Collin, J.-P., Gavina, P., Sauvage, J.-P. ,New J. Chem. 1997, 21, 525-528.

31   Bissell, R. A., Cordova, E., Kaifer, A. E., Stoddart, J. F. A., Nature 1994, 369, 133-137.

32   Diederich, F. N. Cyclophanes, Royal Society of Chemistry: Cambridge, 1991.

33   W. Sliwa, Polish J. Chem., 2001 75, 921–940

34   Notestein, J.M., Katz, A., Iglesia, E., Langmuir 2006, 22, 4004.

35   Ogino, H., Ohta, K., Inorg. Chem. 1984 23:3312-3316.

36   M. L. Bebder, M. Komiyama, Cyclodextrin Chemistry, Springer-Vaerlag: Berlin, 1978

37   T. J. Ward and D. W. Armstrong, J. Liq. Chromatogr., 1986. 9, 407

38   J. Szejtli, in Comprehensive Supramolecular Chemistry, T.,Osa (Editor), Pergamon Elsevier: Oxford, 1996

39   Casu B., Reggiani M., Gallo G. G., Vigevani A., Tetrahedron, 1966, 22, 3061

40   Casu B., Reggiani M., Gallo G. G., Vigevani A., Tetrahedron, 1968, 24, 803

41   Saenger W. R., Jacob J., Gessler K., Steiner T., Hoffmann D., Sanbe H., Koizumi K., Smith S. M., Takaha T., Chem. Rev., 1998, 98, 1787

42   Schneider H. J., Hacket F., Rudiger V., Ikeda H., Chem. Rev., 1998, 98(5), 1755

43   Szejtli, J., Budai, Z., Acta Chim. Acad. Sci. Hung. 1976, 91, 73.

44   Miyajima, K., Sawada, M., Nakagaki, M., Bull. Chem. Soc. Jpn. 1983, 56, 3556.

45   Craig, L. C., Pulley, A. O., Biochemistry 1962, 1, 89.

46   Danil de Namor, A. F., Traboulssi, R., Lewis, D. F. V., J. Am. Chem. Soc. 1990, 112, 8442.

47   Coleman, A. W., Nicolis, I.; Keller, N., Dalbiez, J. P., J. Inclusion Phenom. Mol. Recogn. Chem. 1992, 13, 139.

48   Enssentials of Carbohydrate Chemistry John F.Robyt QD321.Roy]

49   Gelb, R. I., Schwartz, L. M., Laufer, D. A. Bioorg. Chem. 1982,11, 274.

50   Gelb, R. I., Schwartz, L. M., Bradshaw, J. J., Laufer, D. A., Bioorg. Chem. 1980, 9, 299.

51   Jozwiakowski, M. J., Connors, K. A., Carbohydr. Res. 1985, 143,51.

52   Szejtli J., Cyclodextrins and Their Inclusion Complexes; Academiai Kiado: Budapest, 1982

53   Jozwiakowski M. J., Connors K. A., Carbohydr. Res., 1985, 143, 51

54   Thoma J. A., Stewart L., Starch: Chemistry and Technology, Whistler, R. L.; Paschall, E. F. (Eds.); Academic: New York, Vol. I, 1965

55   Briggner L. E., Wadso I., J. Chem. Thermodyn., 1990, 22, 1067

56   Cannors K. A., Chem. Rev., 1997, 97, 1325

57   Lichtenthaler F. W., Immel S., Liebig's Ann., 1996, 22, 1067

58   Bebder M. L., Komiyama M., Cyclodextrin Chemistry, Springer-Vaerlag: Berlin, 1978

59   Carol A.Stanier, Aderman S. J., Claridge T. D.W., Anderson H. L., Angew.Chem.Int.Ed. 2002, 41,No 10.

60   Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P., J.Am. Chem. Soc.1985, 107, 3902.

61   Sohlberg, K., Sumpter, B. G., Noid, D. W. THEOCHEM 1999, 491:281-286.

62   Sohlberg, K., Tarbet, B. J., J. Inclusion Phenom. Mol. Recognit.Chem. 1995, 23: 203-212.

63   Liu, L., Li, X. S., Song, K. S., Guo, Q. X., THEOCHEM 2000, 531,127.

64   Lluch, J. M., Bertran, J., Dannenberg, J. J. Tetrahedron 1988, 44:7621.

65   Saba, G.,Sc ano, P., Sedda, P. A, Thomson, C., THEOCHEM 1990, 204:379

66   Nakayama, A., Richards, W. D. Quant. Struct.-Act. Relat. 1987, 6,153.

67   Khalil, M., Woods, R. J., Weaver, D. F., Smith, V. H., Jr., J. Compur. Chem. 1991, 12,584.

68   Buemi, G., Zuccarello, F., Raudino, A., THEOCHEM 1988, 164,379-389.

69   Hyperchem version 6.0 windows.

70   Allinger N.L., J. Am. Chem. Soc. 1977, 99,8127.

71   Gaussian 03 version 6.0 Copyright 1995-03 Gaussian, Inc.

72   Materials Studio Modelling 3.0 Copyright 2001-2005 Accelrys Software Inc.

73   Connors K.A., Chem. Rev. 1997, 97 ,1325.

74   Murakami H., Kawabuchi A., Kotoo K., Kunitake M., Nakashima N., J. Am. Chem. Soc. 1997, 119 ,7605.

75   Ashton, P. R., Bissell, R. A., Spencer, N., Stoddart, J. F., Tolley,M. S. , Synlett 1992, 914-918.

76   Glink, P. T., Oliva, A. I., Stoddart, J. F., White, A. J. P., Williams,D. J. , Angew. Chem., Int. Ed. Engl 2001, 40,1870-1875.

77   Kanamathareddy, S., Gutsche, C. D. , J. Am. Chem. Soc 1993, 115,6572-6579.

78   Pease, A. R., Stoddart, J. F., Mol. Mach. Mot. 2001, 99,189-236.

3.6 Appendix

Table 4.1 Characteristic properties of cyclodextrin

	α	β	γ	Ref
No. of glc.	6	7	8	48
Mol.wt.	972	1134	1296	48
Depth.	7.8Å	7.8 Å	7.8 Å	48
Diam.top	13.7 Å	15.3 Å	16.9 Å	48
Diam.cavity	5.7 Å	7.8 Å	9.5 Å	48
I_2/I^- color	Blue	Brown	Yellow	48
H_2O^a solubility(S.)	10.1	1.6	23.2	48
Benzene S.	1.24	0.075	1.45	48
Fluorobenzene S.	1.14	0.000	1.6	48
p-Cymene S.	2.92	0.04	0.15	48
Cyclohexanol S.	1.85	0.99	4.77	48
Chlorobenzene S.	1.32	0.000	0.36	48
Bromobenzene S.	1.62	0.01	1.64	48
pKa (25°)	12.33	12.20	12.08	49,50
$\Delta H°$ (ionization), kcal mol-1	8.36	9.98	11.22	49,50
$\Delta S°$ (ionization), cal mol-1 K-1	-28.3	-22.4	-17.6	49,50
$\Delta H°$ (water, 25°), kcal mol-1	7.67	8.31	7.73	51
$\Delta S°$ (water, 25°), cal mol-1 K-	13.8[b]	11.7[b]	14.7[b]	51

a g/100mL

b Mole fragtion standard state

**Table 4.2** Energies of all molecules from structure optimized at AM1 method.

	Energy with zero-point energies	Energy without zero point energy
trans stilbene	-0.59342	-0.99838
cis stilbene	-0.61782	-1.02257
trans rotaxane	-1.81171	-3.28081
cis rotaxane	-1.77755	-3.24140
α-CD	-1.18830	-2.24953

Note: All energies are in unit of Hartree. E_0(trans rotaxane)- E_0(α-CD)- E_0(trans stilben)≈-0.03 H.

Table 4.3 Comparison of UV absorption between experiment and calculation.

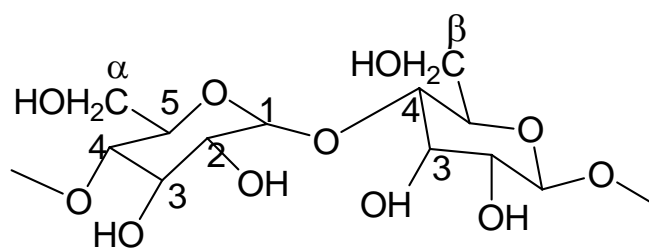|	UV spectra (nm)	
	trans rotaxane	cis rotaxane
Experimetal[59]	340	265
Calculated	300	260
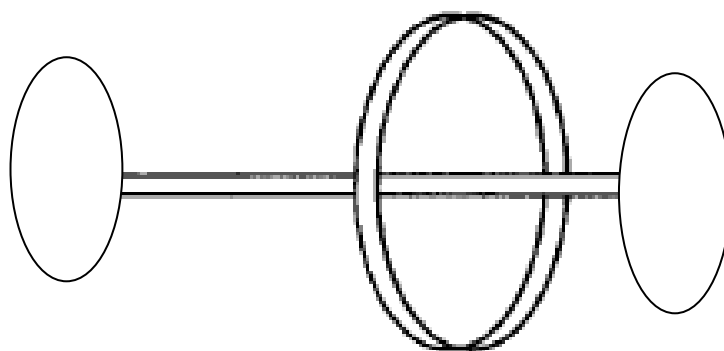
Figure 4.1 Stick and ring model of rotaxane
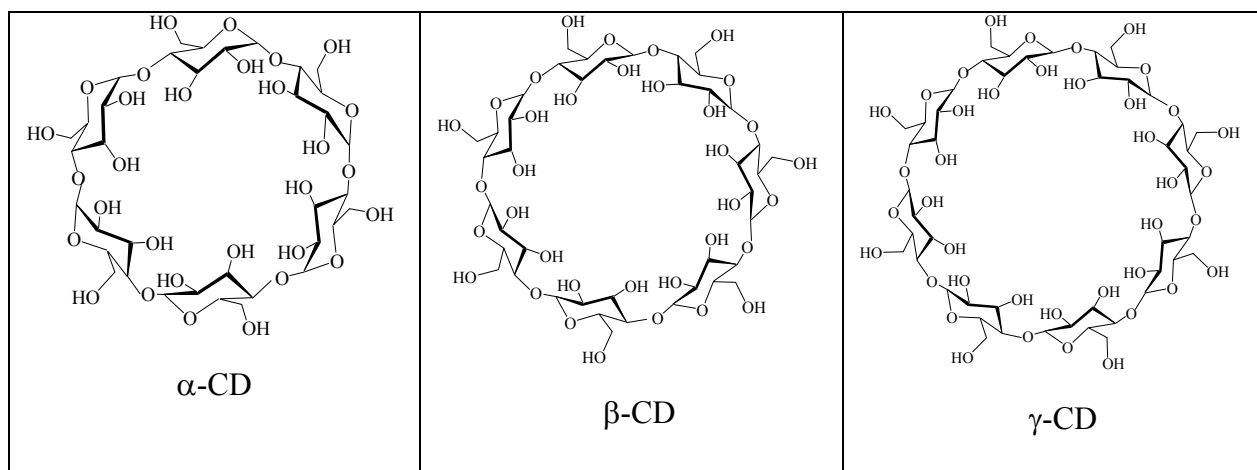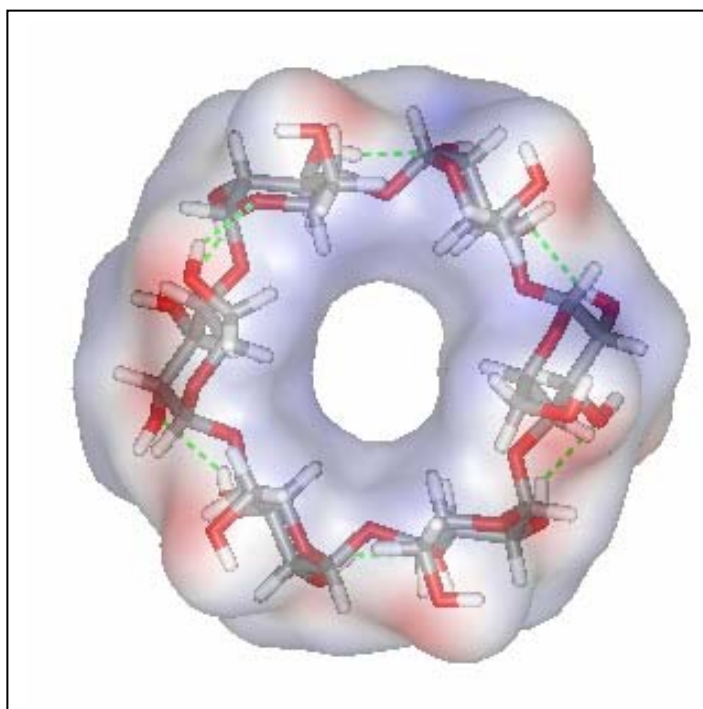
a) Molecular structure of cyclodextrins
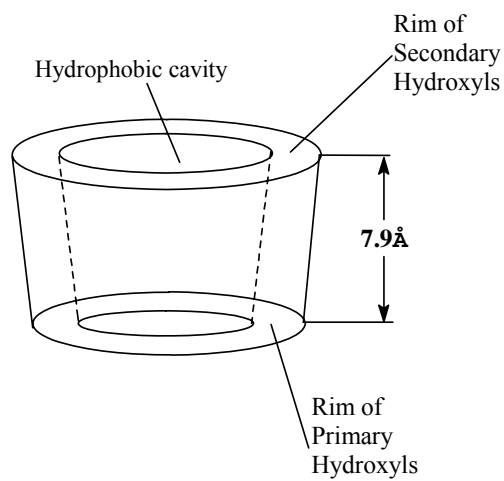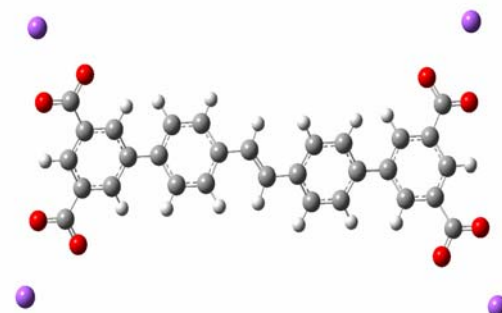

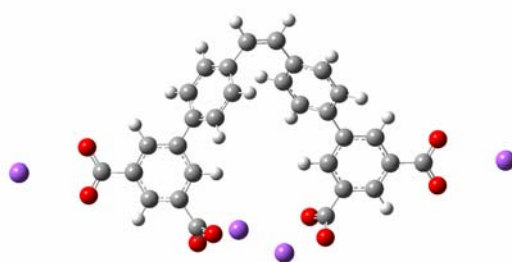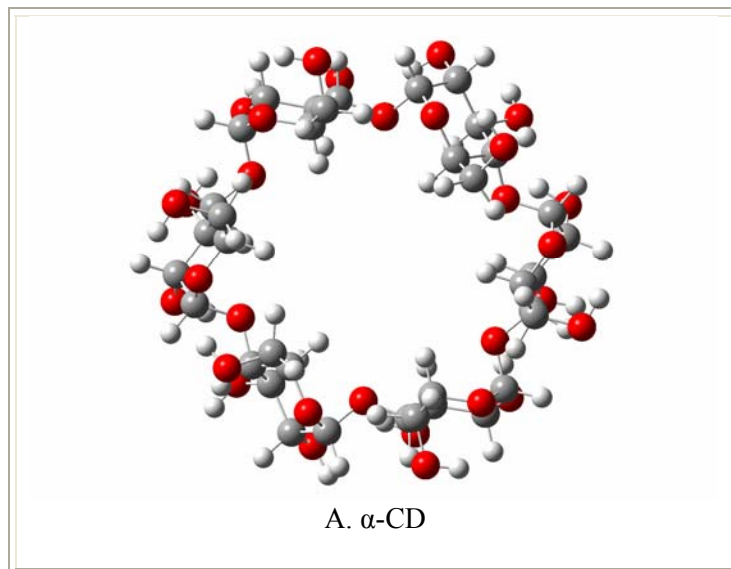b) Schematic representation of the cyclodextrin torus
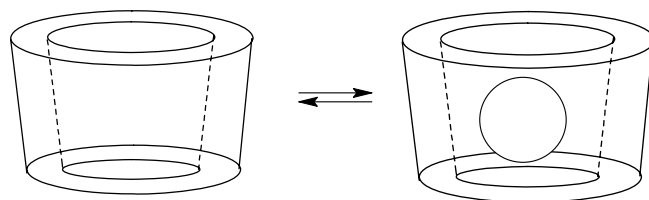
Figure 4.2 Structure of cyclodextrins

Figure 4.3 Structure of glycosidic

Figure 4.4 Structure model of α-cyclodextrin. The green dash line shows the intramolecule hydrogen bonding, which dominates and stabilize the cyclic form.

A. α-CD

B. Trans-form of stilbene

C. Cis-form of stilebene.
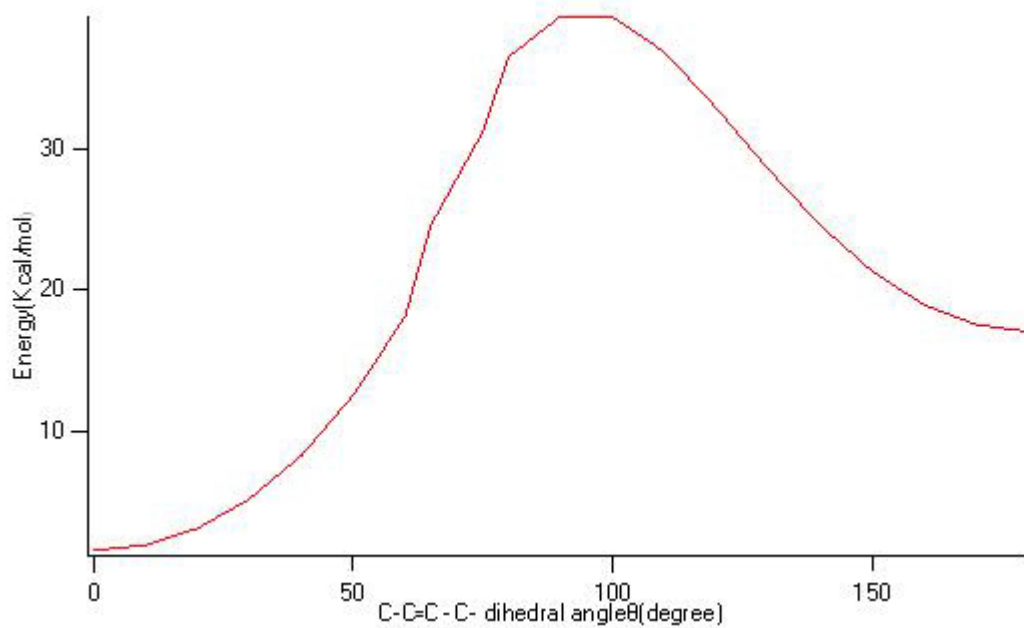
Figure 4.5 Optimized structures of α-CD ,trans and cis stilbene conformers
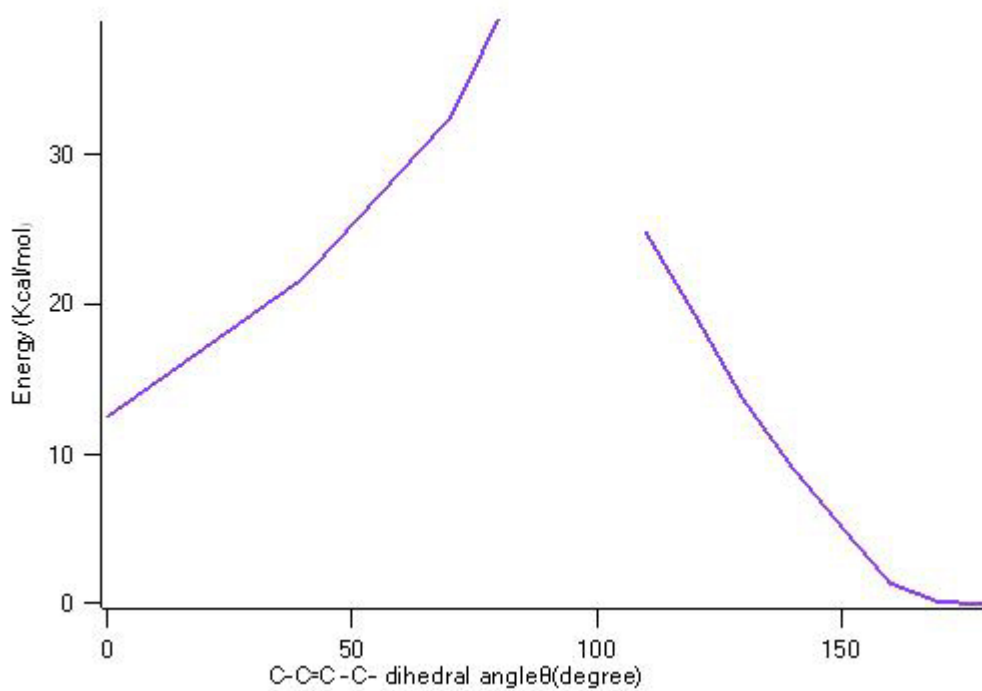
Figure 4.6 Example of cyclodextrin combinds a molecule.
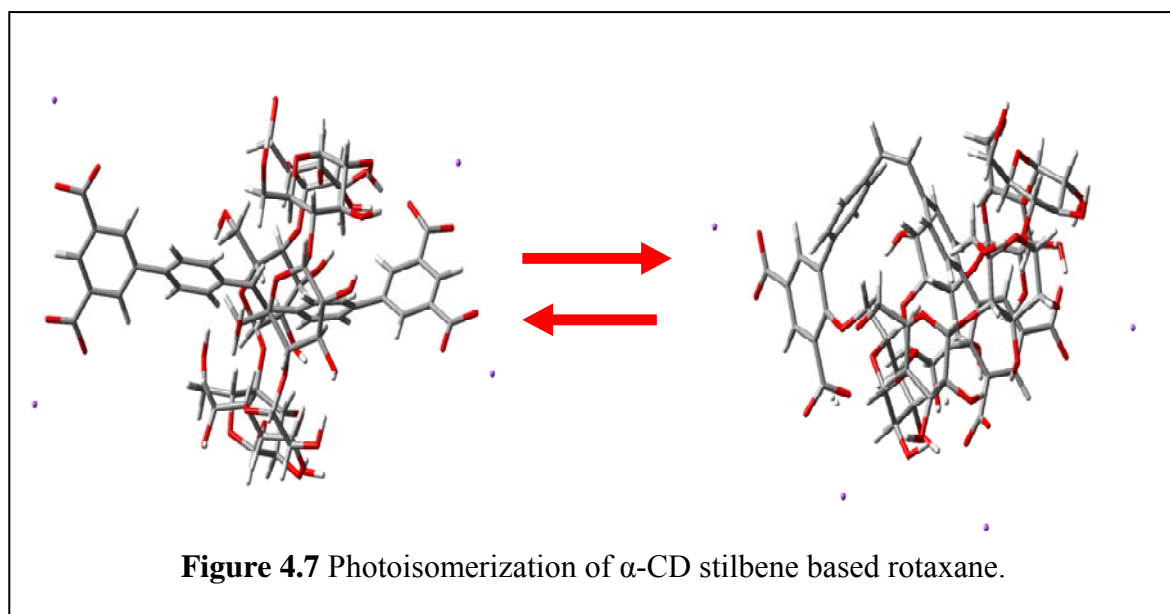
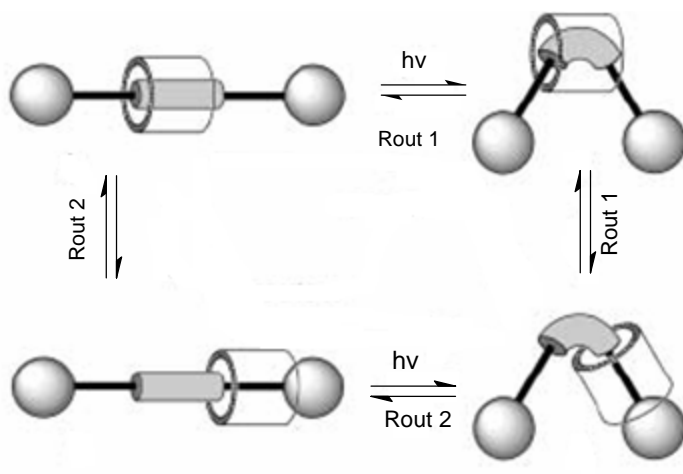**Figure 4.7** Photoisomerization of α-CD stilbene based rotaxane.

A

B

Figure 4.8 The potential curve of isomerization from the two routes.

Figure 4.9 The mechanism of the molecular machine. And route 2 is preferable

Chapter 5 Predicting Harmonic Frequencies with Composite Methods Based on the Collins' Interpolation Scheme

5.1 Introduction

The construction of molecular force fields from experiments is mainly based on Infrared spectroscopy. The experimental data are however insufficient for a unique determination of force fields in most polyatomic molecules. Thus *ab initio* calculations have contributed considerably to the determination of harmonic force fields as well as the assignment of vibrational spectra.

To determine vibrational frequencies by ab initio computational method is becoming more and more important in many areas of chemistry. Theoretically predicted accurate frequencies can serve as fingerprints.

Calculations of harmonic frequencies at the coupled cluster CCSD(T) method accompanying of accuracy are now being carried out on averager to within 8 cm^{-1} of experimental values .

The coupled cluster method that includes all single and double excitations plus a perturbative estimate for triple excitations, CCSD(T)[1] provides unrivaled accuracy for a wide range of molecules see the review by Lee and Scuseria,[2] as well as the more papers of Dunning and co-workers,[3] Martin and co-workers,[4] Helgaker and co-workers[5] and Feller and co-workers.[6]

The coupled cluster CCSD(T) method provides a theoretically sound, accurate description of the electronic structure of a wide range of molecules. To obtain accurate results, however, very large basis sets must be used or a series of calculations must be extraplated to the complete basis set limit. Since the computational cost of CCSD(T)

calculations formally increases with the seventh power of the number of basis functions (N^7),[7] where N is the number of basis functions. Thus using a basis set with double the number of functions in the set or doubling the number of atoms in the molecule with the same basis set increases the cost of the calculation by two orders of magnitude. Although there are some researchers are working on the reduction of the dependence of CCSD(T) on N through a series of controlled approximations,[8] even in the best of circumstances it is unlikely that the dependence will be reduced below N^{4-5} for many of the molecules of interest in chemistry. This steep dependence of the CCSD(T) method on the number of basis functions greatly restricts the range of applicability of this otherwise promising theoretical approach. The CCSD(T) method can only be applied to a restricted range of molecules. Calculations with CCSD(T) method are still confined to small molecules and the evaluations of the first and particularly the second order energy derivatives by such methods are still rather demanding.

In this work we are going to show that the composite methods were originally developed to estimate energies of molecules at very high levels of ab initio theory such as CCSD(T) and QCISD(T) by performing a series of lower level calculations.[9-11] It has been demonstrated that such methods (e.g. G3X(MP2)) can predict molecular energies to within 1.19 kcal/mol accuracy.[11] Furthermore, the CPU time of CCSD(T) calculations scale as $O(N^7)$, whereas lower level methods such as MP2 scale as $O(N^5)$, where N is the number of basis functions.[10] If the success of composite methods for computing energies can be carried over into the calculation of frequencies, then significantly larger molecular systems can be studied with high accuracy.As such, composite methods provide a reliable yet computationally economical means for calculating molecular energies.

Since the introduction of G1 theory by Pople and co-workers in 1989[12], a sizable literature has appeared that utilizes composite methods or, more generally, methods that use various lower levels of ab initio or DFT theory to approximate significantly higher levels of theory. The advantage in doing so lies in the very significant saving in computational expense resulting from the lower level computations. G1 theory and its descendants, G2,[10,13,14] G3,[15] G3S,[16] G3X[9] were originally developed to achieve "chemical accuracy" (energies to within 4 kJ mol^{-1} when compared with experiment) in the computation of thermochemical properties (enthalpies, ionization energies, electron affinities, etc.) of gases. Indeed, this level of accuracy has been achieved for many molecules.

The Gn theories of Pople and co-workers are by no means the only methods that aim to, and achieve, chemical accuracy by approximating expensive higher level methods using several lower level results and empirical parameters. Some of the more popular include the complete basis set (CBS) methods from Petersson and co-workers,[17,18] the Weizmann-n theories and their variants of Martin and co-workers[19,20] and the multicoefficient correlation methods (MCCMs)[21,22] of Truhlar's group.

Significantly fewer studies have appeared in the literature that utilize composite methods for predicting potential energy surfaces (PES). Collins and co-workers have successfully utilized a G3X(MP2) type method in the construction of PES for reactive systems and the calculation of various kinetic parameters[23]. Such methods have also been used in a nine-dimensional bound-state problem for the determination of zero-point energies and ground-state rotational constants[24]. Császár and co-workers utilized a CBS approach to generate a base PES for water and then added in a core-correlation surface, a relativistic correction surface, a quantum electrodynamics correction surface and an adiabatic correction surface.[25,26]

Other groups have considered up to quartic expansions of the potential about an equilibrium configuration. In these studies it is the fundamental frequencies of vibration that are of interest, as well as other spectroscopic constants. Bose and Martin[27] published a detailed study on the azabenzene series, which included considering the possibility of combining DFT anharmonic force fields with coupled cluster geometries and harmonic frequencies. Pouchan and co-workers have also combined harmonic ab initio force constants with DFT anharmonicity constants in a number of studies.[28]

Although high accuracy can be obtained using the above approach for computing fundamentals, high-level ab initio calculations are still required of the harmonic frequencies. Furthermore, such approaches to obtaining a PES, although perfectly suited for the determination of spectroscopic observables of tightly bound systems, are not applicable over the entire PES but presumably can only be applied to turning points. An alternative approach is to define a potential energy that can be computed for any single configuration that is composed of contributions from various levels of theory in a manner similar to Gn theory. In this way, not only can composite force constants and anharmonic force constants to be computed, but composite energies, gradients and second and higher derivatives can also be evaluated for *any* configuration.

Though the high accuracy of composite methods has been demonstrated by numerous studies for total energies, at least at and around minima on the PES, almost no work has been done on examining the general accuracy of the approach for first and higher order derivatives. One way of measuring the accuracy of the curvature of the PES is by comparing composite harmonic frequencies to those obtained using a high single level of theory.

The computation of vibrational frequencies has seen much interest in recent years, with frequencies determined on average to within 8 cm^{-1} of experimental values using CCSD(T) and large basis sets.[29] However, the CPU time associated with this method scales as the seventh power of the number of basis functions, which makes the calculation for even medium-sized molecules prohibitive. Of course, one must include the effects of correlating the core-electrons to achieve such high levels of accuracy. Dunning and Peterson have examined the use of composite methods for making reliable estimates of the electronic energy, spectroscopic properties (D_e, r_e, ω_e, $\omega_e x_e$), ionization energy, and electron affinities compared with the single level CCSD(T)/aug-cc-pV5Z for a test set of diatomic molecules.[30] Specifically, the authors calculated an energy at the CCSD(T) level using a smaller basis aug-cc-pVXZ, X = D, T, and Q, and then added to this energy a correction, Δ_{basis}, to account for the inadequate basis set. Their study revealed that for the test set of molecules, the composite approach is capable of predicting single level CCSD(T) harmonic frequencies to within 2 cm^{-1} on average when X = T. This approach has also been successfully applied for the calculation of harmonic and fundamental frequencies of 12 for first-row closed shell diatomic molecules. Thus, if the success of composite methods for computing energies could be carried over into the calculation of vibrational frequencies, then significantly larger molecular systems can be studied with high accuracy.

However, for this to be possible, it is first necessary to establish the general applicability of composite methods for the calculation of other vibrational modes viz. molecular bends and torsions. To the best of the my knowledge, all previous studies utilizing a Gn-type approach have been restricted to simple diatomic systems, where only uncomplicated stretching modes are assessed. In this work, the harmonic

frequencies at both single level ab initio and G*n*-type composite approximations of CCSD(T) theory, are reported for 19 tri- and 18 tetratomic nonlinear molecules where the bends and torsions are examined as well. Additionally, a general scheme for calculating the composite level equilibrium geometries and harmonic frequencies for polyatomic systems that utilizes the Collins' method of interpolating potential energy surfaces is also described. The accuracy of the composite-level harmonic frequencies is evaluated through comparison with the corresponding single level CCSD(T) calculations.

It is envisaged that the results of this study should provide a clearer indication of the general applicability of composite methods for the calculation of vibrational frequencies of more complicated molecules. Furthermore, this would also contribute toward an alternative procedure for calculating highly accurate ab initio frequencies of larger molecules with significant reductions in computational cost.

5.2 Computational Methods

5.2.1 Collin's Method of Interpolating Potential Energy Surfaces (PES)

The nature of the PES which is a continuous function or algorithm which evaluates the molecular electronic energy at all configurations through which the molecule may pass during the reaction, can be probed either experimentally or by using *ab initio* quantum chemical calculations.

Usually, *ab initio* calculations can be used to obtain, at a reasonable computational cost, many parameters about the shape of PES in the mathematical function for instance, the first derivatives and second derivatives or Hessian matrix of the potential energy with respect to the coordinates of the at discrete points on the PES.

We can characterize the point configuration, zero point, energy, vibrational frequencies, rotational constants etc. In recent years, several scientists have investigated more direct methods of interfacing electronic structure calculations for constructing PES. Collins and co-workers developed a method which combined it with classical trajectory simulations to provide an iterative scheme for successively improving the surface. The interpolation procedure used by Ischtwan and Collins is a modified Shepard interpolation[31,32] that expresses a molecular PES as an interpolation over data which is scattered throughout the relevant region of configuration space. At each data point configuration, the molecular energy, the energy gradient, and matrix of second derivatives are determined from *ab initio* calculations. The PES at any configuration is then given by a weighted sum of Taylor series expansions truncated after second order, about each data point in inverse interatomic distance space. The interpolation procedure forms part of an iterative process in which classical simulations of the chemical reaction provide a sample of molecular configurations, one of which is chosen as a new data point, thus creating a new expanded data set and PES, so that, in the next iteration, the classical simulations explore the configuration space in subtly new ways. This method has a number of desirable attributes: it does not assume a functional form for the PES and so avoids any requirement of chemical intuition as well as the possibility of unphysical features being introduced, the iterative procedure automatically places new data points in regions of the surface that are important dynamically, and most importantly the procedure itself is inherently simple and may be completely automated. There are two distinct features to this scheme: the nature of the Shepard interpolation and the iterative process for updating the PES through classical trajectory simulation.

5.2.2 Composite Methods

The single level ab initio calculations were carried out at CCSD(T) level of thoery with aug-cc-pVXZ basis set, where X = D, T and Q. The calculations were performed using the MOLPRO 2002.1[33] and Gaussian 98[34] suite of programs. The composite energies were based on the ad hoc expression

$$E_{L/S} = E[CCSD(T)/S] + \Delta_{basis} \tag{5.1}$$

Where $\Delta_{basis} = E[MP_n/L] - E[MP_n/S]$, MP$n$ refers to nth-order Moller Plesset perturbation theory, and S and L denote small and large basis sets, respectively. $E_{L/S}$ is an approximation to the energy at the CCSD(T)/L level of theory. This was achieved through three independent, lower level, calculations .Note that if the MPn treatment in the basis set correction term Δ_{basis} was substituted with the CCSD(T) treatment, then this would yield exactly the CCSD(T)/L energies. This expression is similar to the electronic energy given in G3X(MPn) theory in refs 23 and 24 and is the same as that used in refs 7 and 30. It can thus be seen from eq 5.1 that one significant source of error in this approximation is the difference in treatment of electron correlation between the MPn and CCSD(T) levels.

In the subsequent sections, short-hand notations to describe the above calculations are D, T and Q for CCSD(T)/aug-cc-pVDZ, CCSD(T)/aug-cc-pVTZ and CCSD(T)/aug-cc-pVQZ, respectively. Similarly, composite methods are denoted by L/S, where L and S are shorthand notations of the above basis sets. It should be noted from eq 5.1 that the energy is defined for *any* molecular configuration, not just locally at and around minima, and provides a means to generate a composite potential energy surface (PES), as discussed earlier. Because each term in eq 5.1 is differentiable with respect to Cartesian displacements of the atoms so too is the composite energy. Thus we are able to obtain a composite equilibrium structure and harmonic frequencies. All

molecular structures in this work have been optimized using both composite and single level ab initio methods specifying tight convergence. A threshold for the convergence of the energy in the SCF procedure of 10^{-10} Hartree has also been chosen in all calculations.

To calculate the L/S harmonic frequencies of a molecule, a PES is first required to locate its L/S optimized geometry. The PES was constructed using Collins' method of interpolation and has been described in detail previously part.[35-37] Once the PES minimum is located, the second derivative matrix is calculated numerically at this geometry and the harmonic frequencies obtained.

5.2.3 The Algorithm for Obtaining the L/S Harmonic Frequencies of an N-atom Nonlinear Polyatomic Molecule

a) Obtain an approximate set of normal coordinates $(Z_1, Z_2, ..., Z_{3N-6})$ at a lower level ab initio method such as MP2/6-31G(d), where analytic calculation of the Hessian matrix is possible. The optimized geometry, Z_o, at this level of theory serves as an initial guess to the composite method equilibrium structure.

b) The L/S gradient and hessian matrices are evaluated via numerical differentiation using equation (5.2) and (5.3) at $\mathbf{Z}_o$. The step size used in the central difference formulae in Eq. (5.2) and (5.3) is denoted h. This generates the initial L/S PES, which corresponds to a second-order Taylor polynomial about $\mathbf{Z}_o$.

$$\frac{\partial V}{\partial Z_i}\bigg|_{\mathbf{Z}} = \frac{V(i+h) - V(i-h)}{2h} + O(h^2) \tag{5.2}$$

$$\frac{\partial^2 V}{\partial Z_i \partial Z_j}\bigg|_{\mathbf{Z}} = -\frac{1}{2h^2}[V(i+h, j) + V(i-h, j) + V(i, j+h) + V(i, j-h) - 2V(i, j) \\ -V(i+h, j+h) - V(i-h, j-h)] + O(h^3) \tag{5.3}$$

c) The minimum point, Z_1, of this PES is located using the Newton-Raphson method. The process repeats from step two, generating the next data point. After more than one data point is generated, the PES is expressed as an interpolation over the total number of data points or configurations, N_{data}, based on eq 5.4, where $w_n(Z)$ and $T_n(Z)$ refer to the normalized distance-based weight function and second-order Taylor approximation of the nth data point at Z. Each Taylor series requires the energy, gradient, and second derivative of the potential, which are determined from quantum chemical calculations.

$$V(\mathbf{Z}) = \sum_{n=1}^{N_{data}} w_n(\mathbf{Z})T_n(\mathbf{Z}) \tag{5.4}$$

where

$$w_n(\mathbf{Z}) = \frac{v_n(\mathbf{Z})}{\sum_{i=1}^{N_{data}} v_i(\mathbf{Z})}$$

and

$$v_n(\mathbf{Z}) = \left\| \mathbf{Z} - \mathbf{Z}(n) \right\|^{-2p},$$

$$2p > 3N - 3 \tag{5.5}$$

d) The optimization is deemed converged if all the calculated gradient elements (∇V_i, $i = 1, ..., 3N$ - 6) of the newest data point are less than or equal to an ad hoc value, ε_{tol}; otherwise the algorithm repeats from step b. The final data point is the L/S optimized geometry, Z_{eq} expressed in terms of the MP2/6-31G(d) normal coordinates.

e) Z_{eq} is expressed in terms of the $3N$ - 6 standard Z-matrix internal coordinates, where the Hessian with respect to these coordinates is calculated numerically (using Eq 5.3). The L/S harmonic frequencies are obtained in the usual manner from the Hessian and atomic masses.

The Table 5.1 shows parameter specifications during calculations, all numerical derivatives were evaluated using a step size of 5×10^{-4} au and ε_{tol} was specified as 5×10^{-5} au, which corresponds to the tight convergence criteria in the Gaussian software package. In all the molecules examined, the geometry optimization converged within three cycles.

5.3 Results and Discussion

Table 5.2 shows the full list of molecules that were examined in this study. Unless otherwise stated, all composite frequencies were evaluated using MP2 theory in eq 5.1. We first examined the results for the triatomic molecules system, followed by the tetratomic and larger systems. Finally, the efficiency of the composite method is evaluated.

5.3.1 Triatomics

The single level D, T and composite T/D harmonic frequencies for 19 triatomic molecules have been calculated, providing a sample of 57 bending and stretching frequencies for comparison. The data for the T/D and D harmonic frequencies are compared to the T frequencies and are summarized in Table 5.3. As mentioned earlier, the CCSD(T) theory has an intrinsic error of about 8 cm^{-1} in terms of the calculation of experimental vibrational frequencies, and including core-correlation (not included in this work). Thus, it is desirable that the composite harmonic frequencies lie within the same range of their single level CCSD(T) counterparts. It is clear by examining the data in Table 5.3 that a substantial improvement in the accuracy of the harmonic frequencies is achieved using composite methods compared with the D frequencies. For example, the T/D mean absolute deviation (MAD) and root-mean-square (RMS)

values are 4.46 cm^{-1} and 6.88 cm^{-1}, which are about 8 times smaller compared to the D frequencies with MAD and RMS values of 37.0 and 45.7 cm^{-1} respectively.

The distribution of the absolute deviation (AD) values for the 57 T/D and D frequencies is illustrated in Figure 5.1. From the distribution curves, it was observed that the absolute deviations in the D frequencies are fairly evenly distributed, with errors as large as 108 cm^{-1}. On the other hand, about 70% of the T/D frequencies are within 5 cm^{-1} of the T frequencies, and at least 95% within 15 cm^{-1}. However, it was noted that two (originating from HCO$^-$ and HON) of the 57 T/D frequencies had absolute deviations in excess of 20 cm^{-1}, where the maximum was 27.1 cm^{-1}. Likewise, the absolute deviations in the corresponding D frequencies were found to be in excess of 40 cm^{-1}. Further inspection revealed that these frequencies arose from the highest frequency stretching modes of these two molecules. Curiously, the remaining T/D vibrational frequencies of the two molecules are relatively accurate and fall within 12 cm^{-1} of the corresponding T frequencies.

Generally speaking, the errors in the composite expression in eq 5.1 are likely to propagate and impact most on the high frequency vibrational modes. The fact that the two outlying frequencies correspond to the highest stretching frequencies of two molecules attest to this. There are two main sources of error in the composite frequencies: First, the gradient vectors and Hessians were evaluated numerically via central difference and must therefore incur some errors in the harmonic frequencies. More significantly, the use of MPn in the basis set correction term Δ_{basis} must be taken into consideration. Presumably, the anomalously large deviations in the composite frequencies for the two systems are due to the inadequate treatment of electron correlation by the MP2 procedure. As pointed out earlier, this error can be improved

by systematically increasing the level of electron correlation in the basis set correction term.

As such, the harmonic frequencies for the two molecules were reevaluated by substitution of MP3(see Table 5.4) for MP2 in eq 5.1. This led to a marked improvement in the two outlying frequencies where the deviations were reduced to less than 7 cm^{-1}. There was also further improvement in the other frequencies of these molecules where the AD with the T frequencies was reduced to less than 3 cm^{-1}. Similarly, upon substitution with the corresponding MP3 T/D frequencies for the two molecules, the MAD and RMS values were further reduced from 4.46 and 6.88 cm^{-1} to 3.31 and 4.95 cm^{-1}, respectively. These observations suggest that the high-frequency vibrations tend to be more sensitive to the inexactness of the composite expression.

Additionally, the single level Q and composite levels Q/T and Q/D were also computed for a subset of the six lightest triatomic molecules shown in Table 5.6 of the Appendix, and summarized in Table 5.5. Also provided in Table 5.5 and Table 5.6 are the results for T/D, T and D harmonic frequencies versus the Q frequencies. The Q/T frequencies were of comparable accuracy to the Q frequencies, with a MAD of only 1.3 cm^{-1}, compared to a MAD of 9.1 cm^{-1} in the T frequencies. It was also noted that the performance of the Q/D frequencies was slightly worse compared to the Q/T frequencies, with an MAD of 4.1 cm^{-1}, although this is within the acceptable error range. Not surprisingly, the T/D frequencies do not predict the Q frequencies as accurately as the former two but compares well with the T frequencies as illustrated by the good agreement between their MAD and $|\Delta\omega|_{median}$ values.

The above observations imply that the optimal combination of basis sets (L and S) for predicting single level L harmonic frequencies is when they differ by no more than one in the valence designation. It is possible that the widening difference in the

valence designation of the two basis sets (L and S) would deteriorate the quality of the basis set correction term Δ_{basis}, thereby leading to poor agreement with the CCSD(T)/L frequencies.

5.3.2 Tetratomics

The single level T, D and composite T/D harmonic frequencies are also calculated for a set of 18 tetratomic molecules, providing a sample of 108 stretching, bending and torsional modes for comparison. These molecules have geometries ranging from tetrahedral, trigonal pyramidal to planar structures. Table 5.7 summarizes our results from the CCSD(T)/aug-cc-pVXZ (X = D and T) and composite, T/D, harmonic frequencies (cm^{-1}) for the tetratomic molecules.

The performance of the composite frequencies in the tetratomic systems is consistent with the triatomic systems. Here, the MAD value of the T/D frequencies from the T calculations is merely 4.2 cm^{-1}, which is about a 5-fold reduction compared to that of the D frequencies at 20 cm^{-1}. The distribution of the AD of the 108 T/D and vibrational frequencies is plotted in Figure 5.2.

The distribution curves in Figure 5.2 illustrates a trend similar to that in Figure 5.1 where about 95% of the T/D frequencies lie within 10 cm^{-1} of the T frequencies, although it was observed that a small number had absolute deviations greater than 12 cm^{-1} with $\left|\Delta\varpi\right|_{\text{max}}$ of 20 cm^{-1}. Further examination revealed that these frequencies arose from high frequency stretching modes of several tetratomic molecules. On the contrary, the remaining vibrations of these molecules generally showed good agreement with deviations of 10 cm^{-1} or less. To assess the errors due to the composite approximation, the composite frequencies were reevaluated using the MP3 rather than

MP2 in eq 1 for the two of molecules, H_2CN^- and H_2NN, with the largest deviations (19.3 and 20.0 cm^{-1}).

Consequently, both deviations were substantially reduced to 0.36 and 10.7 cm^{-1}, respectively (see Table 5.4). Likewise, the deviations for the remaining frequencies were further reduced to less than 4 cm^{-1}. Substitution of these frequencies for the two molecules with the MP3 composite frequencies led to improved MAD and RMS values of 3.75 and 4.58 cm^{-1}, respectively.

Thus far, the results have been supportive of the capacity of the composite procedure to make reliable predictions of the harmonic frequencies corresponding to bending and torsional modes. However, it has also been noted that the high frequency vibrational modes, specifically stretches, tend to be more sensitive to the errors incurred in the composite approximation. These errors are primarily due to the inaccuracy of the basis set correction term in eq 5.1. Our preliminary assessment shows that the correction term may be systematically refined by using higher-order perturbation methods such as the MP3 procedure. This observation was also reported in the study by Dunning and Peterson on diatomic molecules, where the MP3 composite procedure out-performed its MP2 counterpart.[15]

Despite the higher accuracy and consistency in the MP3 approximation, there is also the added computational cost as the CPU time associated with this method scales as the sixth power of the number of basis functions. On the other hand, the MP2 composite procedure is generally very accurate with errors less than 5 cm^{-1} on average. Hence, for a given CPU time budget, the MP2 approximation should be useful for many molecular studies.

5.3.3 CPU Time Savings

The major advantage with the composite approach is the ability to predict single level CCSD(T) harmonic frequencies accurately, while only requiring a significantly shorter CPU time. Based on the MP2 procedure, the composite approach is approximately a factor of n times faster:

$$n = \frac{t\{CCSD(T)/L\}}{t\{MP2/L\} + t\{CCSD(T)/S\}} \qquad (5.6)$$

where $t\{CCSD(T)/L\}$ refers to the CPU time incurred for the CCSD(T) and large basis set calculation, and so forth.

To estimate the CPU time-savings that may be obtained, the composite procedure was applied to tetrafluoromethane, which is composed of five heavy atoms. Based on a single point calculation at the T/D equilibrium geometry, the CPU times required by the T and T/D procedures are tabulated in Table 5.8.

Accordingly, it is estimated that the CCSD(T)/aug-cc-pVDZ calculations are approximately 14.5 times faster than CCSD(T)/aug-cc-pVTZ. Quite remarkably, the CPU times associated with the composite approximations are exceedingly close, where n has been estimated to be 13.5 and 10.7 for the MP2 and MP3 procedures, respectively.

Additionally, the T/D frequencies for CF_4 have also been computed and compared with the corresponding T harmonic frequencies from earlier work of Wang et al.[38]. The frequencies are tabulated in Table 5.9.

As shown in Table 5.9, all the T/D frequencies are in excellent agreement with the T frequencies, with errors of 3.0 cm^{-1} or less. This result is most noteworthy considering the mere additional cost of performing a MP2 energy calculation. It also appears that for a medium-sized system molecule such as tetrafluoromethane, the difference in the CPU times required for MP3 and MP2 is somewhat small when

compared with the single level CCSD(T) calculations. Accordingly, the MP3 approximation may be more advantageous in terms of reliability for small to medium-sized molecules.

5.4 Conclusion

In this chapter, the harmonic frequencies for a test set of closed shell triatomic and tetratomic molecules have been calculated at both single level and composite approximations of the CCSD(T) method. The results of this study demonstrate the ability of the composite approximation to make very accurate predictions of the harmonic frequencies that are within 5 cm^{-1} of the corresponding single level CCSD(T) calculation. All previous studies have focused exclusively on simple diatomic molecules, where only stretching modes were examined. Through the work presented here it is established that the composite procedure is equally capable of making accurate predictions of other vibrational frequencies corresponding to bending and torsional modes for more complicated polyatomic systems.

The poorer estimation of the stretching frequencies for polyatomic molecules has been attributed to the fact that stretching modes are invariably the high-frequency vibrations and are therefore more sensitive to the errors in the energy expression in eq 5.1. Nevertheless, it has been demonstrated in problematic systems that the large deviations in the T/D harmonic frequencies are readily remedied through the use of MP3 procedure. The tradeoff, however, is the increased computational cost associated with this method, which scales as the sixth power of the number of basis functions.

To summarize, it is conceivable that the combination of efficient Hessian update schemes combined with the theoretical procedure presented here should enable the study of significantly larger molecular systems.

5.5 References

1   Purvis G. D., Bartlett R. J., J. Chem. Phys. 1982 76, 1910 ; Raghavachari K., Trucks G. W., Pople J. A, Head-Gordon M., Chem. Phys. Lett. 1989 157, 479. See also J. F. Stanton, ibid.1997 281, 130.

2   Lee T.J.,Scuseria G.,in Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy, edited by S. Langhoff Kluwer, Dordrecht, 1995.

3   Peterson K. A., Dunning T. H., Jr., J. Chem. Phys.1995, 102, 2032 ; Peterson K. A., Dunning T. H., Jr., J. Phys. Chem. 1995, 99, 3898; Woon D. E., Dunning T. H., Jr., Peterson K. A., J. Chem. Phys. 1996 104, 5883; Xantheas S. S., Dunning T. H., Jr., A Mavridis., ibid. 1997 106, 3280; Peterson K. A., Dunning T. H., Jr., J. Mol. Struct. THEOCHEM 1997 400, 93; van Mourik T. , Dunning T. H., J. Chem. Phys. 1997 107, 2451; Peterson K. A., Wilson A. K., Woon D. E.,, Dunning T.H., Jr., Theor. Chem. Acc. 1997 97, 251; Woon D. E., Peterson K. A., Dunning T. H., Jr., J. Chem. Phys.1998 109, 2233 ; vanMourik T., Wilson A. K., Dunning T. H., Jr., Mol. Phys.1999 96, 529 ; van Mourik T. , Dunning T. H., Jr., J. Chem. Phys.1999 111, 9248.

4   Martin J. M. L., Taylor P. R., Chem. Phys. Lett. 1996 248, 336 ; Martin J. M.L. and Lee T. J., ibid. 1996 258, 129 ; ibid. 1996, 258, 136 ; Martin J. M.L., ibid. 259,679,1996; Martin J. M. L, in Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics, edited by K. K. Irikura and D. J. Frurip American Chemical Society, Washington, D.C., 1996, p. 212; Martin J. M. L., Taylor P. R., J. Chem. Phys. 106, 8620 1997; Martin J. M. L., Theor. Chem. Acc. 97, 227 1997; Martin J. M. L., Chem. Phys. Lett. 273, 98 1997, 292, 411 1998; Martin J. M. L., Lee T. J., Taylor P. R., J. Chem. Phys. 1998 108, 676.

5   Halkier, O. C, Sundholm D., Pyykkoo P., Chem. Phys.Lett. 271, 273 1997; Halkier A., Jorgensen P, Gauss J., Helgaker T., ibid. 1997, 274, 235 ; Helgaker T., Gauss J., Jørgensen P., Olsen J., J. Chem. Phys. 1997, 106, 6430; Helgaker T., Klopper W., Koch H., Noga J., ibid. 1997,106, 9639 ; Halkier A., Koch H., Jorgensen P., Christiansen O., Beck Nielsen I. M., Helgaker,T. Theor. Chem. Acc. 1997, 97,150 ; Halkier A., Taylor P. R., Chem. Phys. Lett. 1998, 285, 133 ; Halkier A., Helgaker T., Jorgensen P., Klopper W., Koch H., Olsen J., Wilson A. K., ibid. 286, 243   1998; Halkier A., Coriani S.,Jorgensen P., ibid.1998, 294, 292 ; Halkier A., Larsen H., Olsen J., Jorgensen P., J. Chem. Phys. 1991, 110, 734; Halkier A., Coriani S., Chem. Phys. Lett. 1999, 303, 408; Halkier A., Klopper W., Helgaker T., Jorgensen P., Taylor P. R., J. Chem. Phys.1999 111, 9157.

6   Feller D., Peterson K. A., J. Chem. Phys. 1998 108, 154 .

7   Bettens, R. P. A., J. Phys. Chem. A 2004, 108, 1826-1829.

8   Lauderdale W. J., Stanton J. F., Gauss J., Watts J. D., Bartlett R. J., Chem. Phys. Lett. 1991 187, 21.

9   Curtiss, L. A., Redfern, P. C., Raghavachari, K., Pople, J. A., J. Chem. Phys. 2001, 114, (1), 108-117.

10   Curtiss, L. A.,Redfern, P. C., Raghavachari, K., Pople, J. A., J. Chem. Phys. 1998, 109, (1), 42-55.

11   Curtiss, L. A., Raghavachari, K., Redfern, P. C., Pople, J. A., J. Chem. Phys. 1999, 109, 4703.

12   Pople, J. A., Headgordon, M., Fox, D. J., Raghavachari, K., Curtiss, L. A., J. Chem. Phys. 1989, 90, 5622-5629.

13   Curtiss, L. A., Raghavachari, K., Trucks, G. W., Pople, J. A., J. Chem.Phys. 1991, 94, 7221.

14   Curtiss, L. A., Raghavachari, K., Redfern, P. C., Pople, J. A., J. Chem. Phys. 1997, 106, 1063.

15   Curtiss, L. A., Raghavachari, K., Redfern, P. C., Rassolov, V.,Pople, J. A. J. Chem. Phys. 1998, 109, 7764; Curtiss, L. A., Redfern, P. C., Raghavachari, K., Rassolov, V., Pople, J. A., J. Chem. Phys. 1999, 110, 4703.; Curtiss, L. A., Redfern, P. C., Raghavachari, K., Pople, J. A., Chem. Phys. Lett. 1999, 313, 600.; Kedziora, G. S., Pople, J. A., Rassolov, V. A., Ratner, M., Redfern, P. C,Curtiss, L. A. J. Chem. Phys. 1999, 110, 7123. Curtiss, L. A., Redfern, P. C., Rassolov, V., Kedziora, G., Pople, J. A., J. Chem. Phys. 2001, 114, 9287.

16   Curtiss, L. A., Raghavachari, K., Redfern, P. C., Pople, J. A., J. Chem. Phys. 2000, 112, 1125-1132.

17   Montgomery, J. A., Frisch, M. J., Ochterski, J. W., Petersson, G. A., J. Chem. Phys. 2000, 112, (15), 6532-6542.

18   Ochterski, J. W., Petersson, G. A., Montgomery, J. A., J. Chem. Phys. 1996, 104, (7), 2598-2619.

19   Martin, J. M. L., de Oliveira, G., J. Chem. Phys. 1999, 111, (5), 1843-1856.

20   Atasoylu, O., Martin, J. M. L., Ka'llay, M., Gauss, J., J. Chem. Phys. 2004, 120, 4129.

21   Fast, P. L., Corchado, J., Sanchez, M. L.; Truhlar, D. G., J. Phys. Chem. A 1999, 103, 3139-3143

22   Zhao, Y., Lynch, B. J., Truhlar, D. G., Physical Chemistry Chemical Physics 2005, 7, (1), 43-52.

23   Bettens, R. P. A., Collins, M. A., Jordan, M. J. T., Zhang, D. H., J. Chem. Phys. 2000, (23), 10162-10172.

24   Bettens, R. P. A., J. Am. Chem. Soc. 2003, 125, 584-587.

25   Csaszar, A. G., Czako, G., Furtenbacher, T., Tennyson, J., Szalay, V., Shirin, S. V., Zobov, N. F., Polyansky, O. L., J. Chem. Phys. 2005, 122, (21).

26   Polyansky, O. L.,Csaszar,A.G., Shirin, S. V., Zobov, N. F., Barletta, P.; Tennyson, J., Schwenke, D.W., Knowles, P. J., Science 2003, 299, 539-542.

27   Boese, A. D., Martin, J. M. L., J. Phys. Chem. A 2004, 108, (15), 3085-3096.

28   Begue, D., Carbonniere, P., Barone, V., Pouchan, C. Chem. Phys. Lett. 2005, 415, 25. Gohaud, N., Begue, D., Pouchan, C. ,Int. J. Quantum Chem. 2005, 104, 773.; Begue, D., Carbonniere, P., Pouchan, C., J. Phys. Chem. A 2005, 109, 4611.

29   Preface in Spectrochim. Acta, Part A 1997, 53, vii.

30   Dunning, T. H., Jr., Peterson, K. A., J. Chem. Phys. 2000, 7799, 113.

31   R. Farwig, in Algorithms for Approximation, edited by J. C. Mason and M.G. Cox Clarendon, Oxford, 1987.

32   Lancaster P., Salkauskas K., Curve and Surface Fitting, An Introduction Academic, London, 1986.

33   Werner,H.-J.,Knowles,P. J.version 2002.1,Amos,R.D.; Bernhardsson, A. Berning, A., Celani, P., Cooper, D. L., Deegan, M. J.O., Dobbyn, A. J., Eckert, F., Hampel, C., Hetzer, G., Knowles, P. J., Korona, T., Lindh, R., Lloyd, A. W.,McNicholas, S. J., Manby, F. R., Meyer, W., Mura, M. E., Nicklass, A., Palmieri,P.,PitzerR.,RauhutG.,Schu¨tz,M.,Schumann,U.,Stoll,H.,Stone,A.J.,Tar roni,R., Thorsteinsson, T.,Werner, H.-J. MOLPRO 2002.6, a package of ab initio programs.

34   Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E.; Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C.,Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C.,Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B. G., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle, E. S., Pople, J. A. Gaussian 98, revision A.6 and A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

35   Bettens, R. P. A., Collins, M. A., J. Chem. Phys. 1999, 111, 816-826.

36   Jordan, M. J. T., Thompson, K. C., Collins, M. A., J. Chem. Phys. 1995, 102, (14), 5647-5657.

37   Thompson, K. C., Jordan, M. J. T., Collins, M. A., J. Chem. Phys. 1998, 108, (20), 8302-8316.

38   Wang, X. G., Sibert, E. L., Martin, J. M. L., J. Chem. Phys. 2000, 112, (3), 1353-1366.

5.6 Appendix

Table 5.1 Parameters specifications for calculation of L/S harmonic frequencies

Molecule	h /Bohr	p	ε_{tol} /hartree
Three-atom	5×10^{-4}	4	5×10^{-5}
Four-atom	5×10^{-4}	5	5×10^{-5}
Six-atom	5×10^{-4}	8	5×10^{-5}

Table 5.2 Test Set of Molecules Used

triatomic	H_3^+ H_2F^+ CH_2 CHF H_2O HNO HON NH_2^+ NH_2^- HO_2^+ HO_2^-
	OCF^- HF_2^+ HOF HNF^- HCO^- CF_2 C_2O F_2O
tetratomic	CFH_2^- NFH_2 H_2CO H_2O_2 OFH_2^+ *cis*-N_2H_2 *trans*-N_2H_2
	trans-HCOH *cis*-HCOH *cis*-$HCNH^-$ *trans*-$HCNH^-$,
	CH_3^- NH_3 OH_3^+ H_2NO^+ H_2NO^- H_2NN H_2CN^-

Table 5.3 Comparison of D and T/D frequencies with T harmonic frequencies for three-atom systems

Method	MAD	RMS	$\left\|\Delta\omega\right\|_{median}$	$\left\|\Delta\varpi\right\|_{max}$
D	38.8	44.2	25.5	108.1
T/D	5.2	7.5	4.0	27.1

Table 5.4 Two tetratomics H_2CN^- and H_2NN, and two triatomics, HCO^- and HON, CCSD(T)/aug-cc-pVTZ and composite T/D harmonic frequencies (cm^{-1}) using MP3 and MP2 in the Δ_{basis} correction.

Molecule	aug-cc-pVTZ	T/D(MP2)	T/D(MP3)
HCO^-	1284.06	1292.97	1284.42
	1380.94	1369.60	1383.52
	1800.83	1775.35	1805.55
HON	1266.05	1260.99	1265.74
	1457.34	1454.92	1457.69
	3294.83	3267.78	3288.28
H_2CN^-	973.92	969.54	971.34
	1131.30	1128.36	11131.62
	1475.76	1474.14	1477.23
	1622.10	1623.36	1625.70
	2254.63	2234.61	2243.96
	2529.56	2520.82	2529.11
H_2NN	987.48	982.97	983.92
	1315.56	1313.40	1316.08
	1573.15	1571.87	1573.11
	1725.84	1724.82	1729.36
	3123.55	3111.12	3126.43
	3132.44	3113.17	3132.07

Table 5.5 Comparison of Q/T, Q/D, T and D frequencies with Q harmonic frequencies for three-atom systems

Method	MAD	RMS	$\left|\Delta\omega\right|_{median}$	$\left|\Delta\varpi\right|_{max}$
T	9.1	11.0	10.2	20.7
D	44.7	54.0	34.9	118.0
Q/T	1.3	1.61	1.2	2.9
Q/D	4.1	5.0	3.9	9.8
T/D	12.2	14.3	12.9	24.7

Table 5.6 The CCSD(T)/aug-cc-pVXZ (X = D, T and Q) and the composite Q/T, Q/D and T/D harmonic frequencies (cm^{-1}) for the six lightest triatomics.

Molecule	aug-cc-pVQZ	aug-cc-pVTZ	aug-cc-pVDZ	Q/T	Q/D	T/D
CH$_2$	1398.69	1395.4	1387.515	1398.474	1399.299	1396.312
	2920.43	2912.163	2886.512	2919.326	2913.642	2906.464
	2994.87	2982.7	2961.553	2994.049	2987.98	2976.743
H$_2$O	1649.799	1645.602	1637.912	1649.864	1648.298	1644.337
	3831.173	3810.932	3786.987	3829.005	3824.482	3806.395
	3940.791	3920.044	3904.953	3938.239	3934.226	3916.019
NH$_2^-$	1488.17	1486.242	1465.605	1488.434	1487.402	1485.265
	3290.5	3276.33	3233.759	3290.846	3286.565	3271.632
	3375.61	3359.186	3323.316	3375.785	3372.368	3355.45
H$_2$F$^+$	1458.04	1454.466	1450.483	1459.381	1458.99	1454.328
	3536.46	3535.267	3458.182	3535.202	3530.718	3530.02
	3540.31	3538.485	3463.572	3539.779	3536.307	3534.285
H$_3^+$	2771.45	2760.014	2653.387	2769.23	2767.746	2758.763
	2772.189	2760.766	2654.113	2769.541	2768.062	2759.081
	3436.457	3426.335	3466.174	3433.676	3436.074	3428.452
NH$_2^+$	1406.68	1407.812	1399.923	1406.232	1406.396	1408.03
	3188.9	3178.625	3150.254	3186.602	3180.423	3172.036
	3278.95	3267.688	3250.653	3276.795	3269.185	3259.817

**Table 5.7** Comparison of D and T/D frequencies with T harmonic frequencies for four-atom systems

Method	MAD	RMS	$|\Delta\omega|_{median}$	$|\Delta\varpi|_{max}$
D	20.3	23.7	18.9	63.9
T/D	4.2	5.4	3.5	20.0

Table 5.8 CPU Times Associated with the MP2 and CCSD(T) Calculation at the Equilibrium Geometry of CF_4

basis set	no. of basis functions	MP2	MP3	CCSD(T)
aug-cc-pVDZ	115	-	-	1702.90
aug-cc-pVTZ	230	112.68	593.96	24687.76

Table 5.9 Computed CCSD(T) Harmonic Frequencies (cm^{-1})

Vibrational mode	T/D	T	|T - D|
1	434.1	435.2	*1.1*
2	628.4	630.4	*2.0*
3	912.4	915.2	*2.8*
4	1298.5	1301.3	*2.8*

Absolute Deviation (AD) Distribution Curve Triatomic Systems
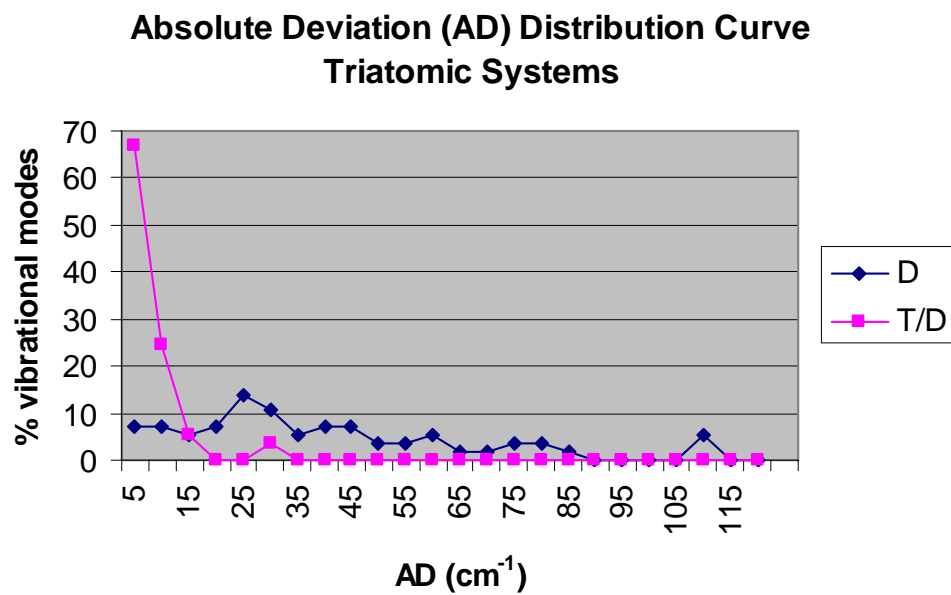
Figure 5.1 Plot of the percentage of vibrational modes against the absolute deviation from the T frequencies for the triatomic systems.

**Absolute Deviation (AD) Distribution Curve
Tetra-atomic Systems**
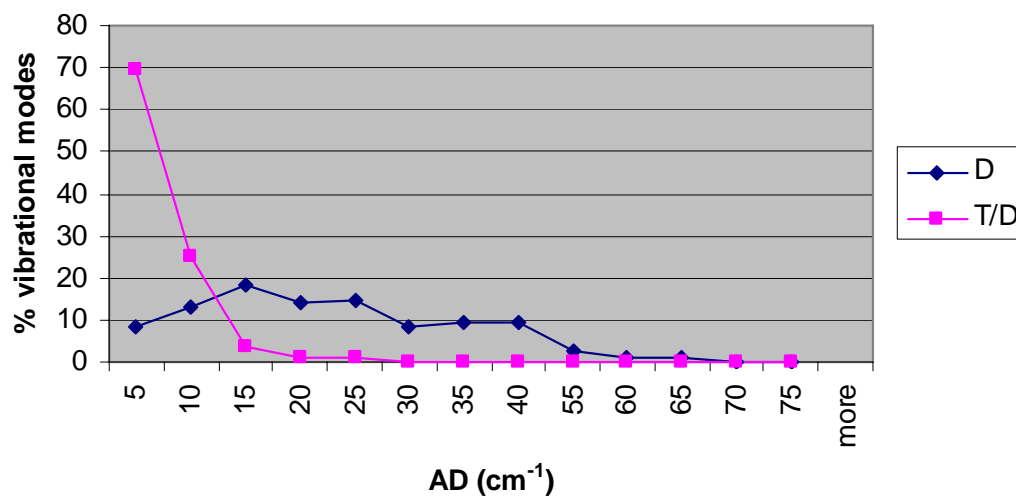
Figure 5.2 Plot of the percentage of vibrational modes against the absolute deviation from the T frequencies for the test set of tetratomic systems.

Chapter 6 The Accurate Prediction of Energies via Fragmentation

6.1 Introduction

A challenge in the field of biology and computational chemistry has been the *ab initio* energy calculations of large biomolecules, such as proteins. As is known, the main task of computational chemistry is to predict molecular system properties and interactions by solving Schrödinger–Dirac equations. As these biomolecules often have large numbers of atoms, and the present ab inito methods suffer from a steep scaling problem that generally prohibits the accurate estimation of properties for large systems, like proteins computationally prohibitive.

The computational time for predicting molecular properties scales drastically with the number of atoms (or electrons) in the molecule, and also the level of theory and basis set used. The dependence is even bigger if post Hartree−Fock methods which involve electron correlation are employed. For instance, the CPU time for couple-cluster approximation scales N^7,[1] where N is the number of basis functions used. It is desirable if the *ab initio* calculations can achieve linear scaling with respect to the size of the system.

Lately, a lot of works have been done to modify traditional *ab initio* methods to near linearity.Yang[2,3] has generated a new category of approaches known as fragment-based methods. The methods treat the system as a whole, and it is divided into a set of subsystems (or fragments). All the fragments are calculated by traditional quantum-mechanical approach, the results can be combined to determine various properties of the whole system. After Yang's methods were born, many fragment-based methods are developed, such as the adjustable density matrix assembler

approach[4-6] by Exner and Mezey; the fragment molecular orbital approach (FMO)[7-10] by Kitaura; the elongation method[11,12] developed by Imamura etc.; the molecular tailoring approach[13-15] by Gadre; the *ab initio* fragment orbital-based theory[16] of Das *et al.*; and the molecular fractionation with conjugate caps (MFCC) scheme[17-21] proposed by Zhang and co-workers. All these methods compute molecular properties that scale linearly with molecular size and produce very accurate properties for different types of large molecules.

A very recent breakthrough in computational chemistry on the subject of fragmentation methods are two studies by Li *et al* [22] and Deev and Collins.[23] It has shown that the total ground-state energy of a molecule can be directly computed by fragmenting the molecule and taking a linear combination of the resultant fragment total energies. These results were achieved at both the Hartree–Fock and post-Hartree–Fock levels of theory. In Li *et al*'s study, each individual system under consideration needs to decide which fragments should be considered. While, Deev and Collins showed more generality that accurate gradients and second derivatives of the energy can also be obtained through the use of their fragment-based method. In their fragmentation scheme, accurate energies of a molecule can be computed by breaking up the molecule into small fragments and taking a linear combination of the resulting fragment subunits. The number of fragments is linearly proportional to the number of atoms in the molecule, thus computational time scales linearly with the number of atoms and the basis functions used. There are three levels of fragmentation, whereby higher levels involving larger fragments lead to more accurate total energy calculations due to the better estimation of the local bonding environment up the hierarchy. The absolute error in energy is linearly proportional to the number of fragments in the molecule.

According to Deev and Collins's findings, the fragmentation approximations are weakly dependent on the level of theory and basis functions used. This is because of the deficiencies in basis set and level of theory which is nearly equal for the true fragmented calculations. This would mean that the accuracy of the fragmentation method can be tested at a lower level of theory and basis set and then relied on at higher levels where calculation of true energy would have been impossible. Higher level calculations can now be carried out than what was deemed feasible for full energy calculations.

The work in this chapter is derived from the approach by Deev and Collins, but with improvements done to it by Bettens and Lee.[24] In the original study done by Deev and Collins, only small closed-shell organic molecules with no explicit charge separation were considered. The aim of this study is to prove that the Collin's method of fragmentation can be used to accurately calculate the total electronic energies of charged organometallic compounds. This will be the first time that charged compounds have been studied using fragmentation techniques. The energies of the fragments and of the full molecule will be computed using the GAUSSIAN 03 computational package.[25]

6.2 Methodology

6.2.1 Computational Procedure

The organometallic molecule' structure was optimized at HF/6-31g(d) 5d or B3LYP/6-31g(d)5d level of theory to find its ground state energy configuration. The molecule was then fragmented using the scheme developed by Bettens and Lee with the central metal charge manually added to all fragments that do not contain group G_1* as described. The total electronic energies of the fragments and the full molecule were

then calculated .The energy calculated for all three levels were then compared with the full energy calculations. This was repeated for all the molecules investigated.

6.2.2 Testing Samples

To avoid ambiguity in the electronic spin structure, our work was first restricted to closed-shell octahedral and tetrahedral d^0, d^6 and d^{10} compounds, at the Hartree−Fock (HF) level of theory, using the 6-31G basis set. Next, the study was extended to open-shell octahedrals, using B3LYP/6-31g. The ligands used for the close-shell compounds tested were ethyl amine, $NH_2−CH_2−CH_3$; propyl amine, $NH_2−CH_2−CH_2−CH_3$ and butyl amine, $NH_2−CH_2−CH_2−CH_2−CH_3$. This is to test for the accuracy of the energy calculations with varying ligand length. For the open-shell compounds, ethylenediamine, $NH_2−CH_2−CH_2−NH_2$, was used as the ligand.

To further explain fragmentation method, we will take $[Sc(ethylamine)_6]^{3+}$ for instance. Firstly, the molecule $[Sc(ethylamine)_6]^{3+}$ is made up of a set of groups, G_1, G_2 etc and then fragments are generated in a hierarchy of three levels using the following method.

During fragmentation, positive and negative fragments are generated for each level seen in Table 6.1. Each fragment is in a form of "A Frag B = C, D…" whereby A refers to the multiplicity of that fragment, B refers to the numbering of that fragment and C, D… refers to the groups that are in that fragment.

Note that in the attempt to preserve the crystal field splitting energy of the metal centre, the group involving the charged metal centre (G_1*) also includes the amine group of the surrounding ligands as shown in the diagram above.

As shown, larger fragments are generated up the hierarchy, thus, we can expect the accuracy of energy approximations to improve up the hierarchy, as the distance of the two simultaneous breaks in the molecule increases.

In theory, we are able to fragment any molecule type by following the rules for each level as shown above. The number of fragments is expected to increase with the molecule size. Note that even higher levels e.g. 4 or 5 level of fragmentation is also possible but not feasible for smaller molecules as that may mean including the whole molecule in the fragments which would be meaningless.

Level 1: We consider every possible pair of groups in the molecule. Subtract all double counts (i.e. overlapping subfragments).

$G_1*G_1...G_{12}$ → $G_1*G_1 + G_1*G_3 + G_1*G_5 + G_1*G_7 + G_1*G_9 + G_1*G_{11} + G_1G_2 + G_3G_4 + G_5G_6 + G_7G_8 + G_9G_{10} + G_{11}G_{12} - (5 G_1* + G_1 + G_3 + G_5 + G_7 + G_9 + G_{11})$

Level 2: We consider all non-end groups and connect all the groups that are directly connected to that group. Subtract all double counts.

$G_1*G_1...G_{12}$ → $G_1*G_1G_2 + G_1*G_3G_4 + G_1*G_5G_6 + G_1*G_7G_8 + G_1*G_9G_{10} + G_1*G_{11}G_{12} + G_1*G_1 G_3 G_5 G_7 G_9 G_{11} - (G_1*G_1 + G_1*G_3 + G_1*G_5 + G_1*G_7 + G_1*G_9 + G_1*G_{11})$

Level 3: We consider all non-end groups connecting pairs (group pairs) and connect all the groups directly to each group pair. Subtract all double counts.

$G_1*G_1...G_{12}$ → $G_1*G_1G_2 G_3 G_5 G_7 G_9 G_{11} + G_1*G_3G_4 G_5 G_7 G_9 G_{11} G_1 + G_1*G_5G_6 G_7 G_9 G_{11}G_1 G_3 + G_1*G_7G_8 G_9 G_{11} G_1 G_3 G_5 + G_1*G_9G_{10} G_{11} G_1 G_3 G_5G_7 + G_1*G_{11}G_{12} G_1 G_3 G_5 G_7 G_9 - (5 G_1*G_1G_3G_5G_7G_9G_{11})$

6.2.3 Alternative Fragmentation

Here gives a simple example where the fragmentation method differs from that of Deev and Collin can be seen by the level-3 fragmentation of a molecule with the morphology shown in Figure 6.2. Table 6.2 provides the fragmentation using the Deev and Collins method and Bettens and Lee's method. Clearly the two methods, for this morphology, give different fragments. At level-1 and level-2 identical fragments and coefficients are obtained, but at level-3, the fragments given by the method described

151

in this work are smaller. Here both methods give unique fragments.

The Molecule is illustrated in Scheme1. The arbitrary group numbering is also indicated in Figure 6.2. For brevity we shall consider only the level-3 decomposition. Included in Scheme 1 are all the bonding positive and negative fragments that can be formed from ML_2 complex, with the positive fragments noted by big case of letter "P". The negative fragments from which the positive fragments are derived are indicated by the letter "N".
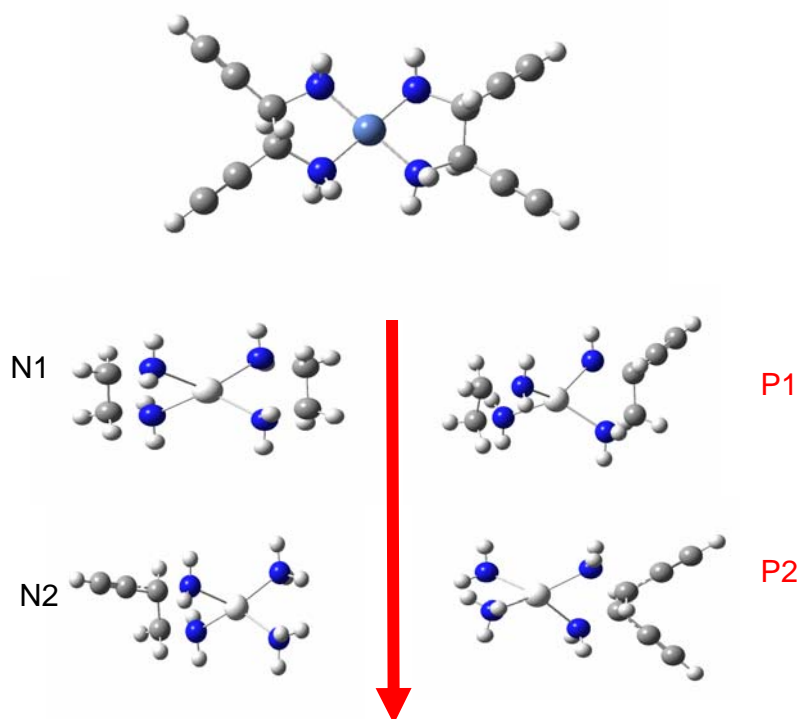
Scheme 1 ML_2 and its level-3 decomposition (see text).M is the central transition metal, L is the notation for ligand hexa-1,5-diyne-3,4-diamine. Fragments labeled Pn are positive fragments, while those labeled Nn are negative fragments.

6.3 Results and Discussion

6.3.1 Effect of Addition of Metal Charge to Fragments

Capping hydrogens are added during the fragmentation to replace broken bonds. The use of hydrogen atoms assumes that the connections between the groups

are single bonds and thus ensures that the number of single bonds in the reaction is conserved.

The central metal charge was also added at the exact position of the metal centre, to all the fragments that do not contain G_1*. This is to account for non-bonding interactions of the central metal charge and the ligands surrounding it.

In each level of fragmentation, there are fragments which are added and some that are subtracted, these are the positive and negative fragments respectively. The negative fragments are there to subtract off the subfragments that have been counted two or more times in the positive fragments. The energies of each of these fragments were added (or subtracted) and compared with the full molecule energy calculation.

We begin by considering the need of the addition of the central metal charge to all the fragments that do not contain group G_1*. The addition of charge was placed exactly where the metal was and takes into account the non-bonding interaction between the charge metal centre and the ligands attached to it. This is important as long-range coulombic interaction cannot be simulated within individual fragments but encompasses the entire molecule.

Table 6.3 compares the error in the total electronic energy, with and without the addition of charge, for charged closed-shell octahedral complexes with Sc^{3+}, Fe^{2+} and Zn^{2+} as metal centres. The error is calculated by the calculated value by fragmentation minus the full molecule energy calculation. Due to the way the molecules are fragmented, for metals with ethylamine ligands, fragments in level 2 and 3 all have group G_1* in them thus, all its fragments do not need the addition of charge and are not included in this table. For this reason, level 3 errors for metals with propylamine ligands are not included in this table as well.

Comparing the errors of the addition of charge with those without adding of charges for each molecule, it is clear that this is useful for more accurate energy calculations. The addition of charge is most needful for the level 1 calculations seen by the significant improvement of errors. For example: for $[Fe(butylamine)_6]^{2+}$, the error in level 1 decreased the most from −3.51 to 0.016 H when charge was added whereas the error in level 2 decreased from −0.87 H to −0.015 H and for level 3, the error decreased from −0.236 H to −0.0069 H. This is due to the decrease of the number of fragments which does not include G_1* up the hierarchy, i.e. less number of fragments would need the addition of charge at level 3 as compared to level 2 and level 1, thus the significance of the addition of charge would decrease up the hierarchy.

However, not all the molecules tested reflected this norm. For $[Sc(propylamine)_6]^{3+}$, it is observed that the addition of charge to level 2 fragments seemed to have increased the error, from 0.0116 H to −0.0230 H. This presumed increase in error is also seen for level 2 and 3 for $[Sc(butylamine)_6]^{3+}$. This could be due to the cancellation of errors between the positive and negative fragments.

Overall, it is observed for all the fragment energies tested, that the addition of charge lowers the energy of the fragment. This shows an energy stabilization of the fragment when charge is placed. Let's take Frag = 2, 3 and Frag = 2 as examples to illustrate this point. Without addition of charge, Frag = 2, 3 is essentially an ethane fragment and Frag = 2 is a methane fragment after hydrogen capping. With the addition of charge, Frag = 2, 3 and Frag = 2 are still an ethane and methane fragment respectively but with a metal charge added at the central metal position. Table 6.4 shows the fragment energy of Frag = 2, 3 and Frag = 2 for $[Sc(ethylamine)_6]^{3+}$, $[Fe(ethylamine)_6]^{2+}$ and $[Zn(ethylamine)_6]^{2+}$ complexes, with and without the addition of charge. It is shown that for $[Sc(ethylamine)_6]^{3+}$, the energy of both fragments

decreased by about 30−40 mH with the addition of charge. This energy change of 30−40 mH is seen for all of the Sc^{3+} complexes tested. For $[Fe(ethylamine)_6]^{2+}$ and $[Zn(ethylamine)_6]^{2+}$, however, there was a drastic decrease in energy with the addition of charge for both Frag = 2, 3 and Frag = 2 by about 1450 mH. This energy change of about 1450 mH is also seen for all of the Fe^{2+} and Zn^{2+} complexes tested. It is noted that a 1450 mH difference in energies is far too large, however, this huge difference in energy cannot be accounted for till further tests is done.

6.3.2 Octahedral Complexes

Let us now consider the fragmentation of closed-shell octahedral organometallic complexes. Table 6.5 presents the error in the total electronic energy for octahedral complexes at levels 1, 2 and 3. A variety of first row transition metals were used to test the accuracy of the fragmentation method. All the octahedral complexes have positive charge ranging from +1 to +4 so as to test for the effect of charge on the errors in the total electronic energy. Figure 6.3, 6.4 and 6.5 compare the total electronic energy errors[2] on a log scale based on the d state of the metal centres. The error values were squared to allow comparison of both the positive and negative errors. As we take errors that are below 0.01 H to be in the acceptable range, thus from the figures, errors that are below the 0.0001 H mark are considered to be in good range. From the data collected, a few trends can be determined.

Firstly, it is observed that the longer the ligand is, the bigger the error incurred. Taking Cu^{1+} compounds at level 3; for $[Cu(ethylamine)_6]^{1+}$, the error was −0.069 mH; for $[Cu(propylamine)_6]^{1+}$, the error was −0.21 mH; and for $[Cu(butylamine)_6]^{1+}$, the error was −1.98 mH. This increase in error with the ligand length is consistent with the level 3 errors of all the octahedral complexes in Table 6.5 and is due to the increase

of the number of fragments with the size of the compound, which increases the error involved.

Secondly, the larger the charge on the metal centre is, the bigger the error at level 3. This is also consistent with all the octahedrals in Table 6.5. This trend can be explained by the fact that the higher the charge on the metal, the greater the electron withdrawing capacity it has. Thus in the full molecule, electrons can be drawn into the space in and around the metal since basis functions (i.e. orbitals) are present on the metal. By replacing the metal by a simple charge without the associated basis functions (also called ghost functions in this case) can lead to error. This technique of charge correction would have to be refined to yield better results. However, this trend did not seem to apply to level 1. For the d^0 compounds in Figure 6.3, it is observed that for the Ti^{4+} compounds, the errors^2 are at the range of $5-7$ mH^2 which were much larger compared to the errors for the Sc^{3+} compounds, which were at the range of 0.05 to 0.3 mH^2 at level 1. However for the d^{10} compounds in Figure 6.4, it is observed that the Cu^{1+} compounds has errors^2 which were larger than the errors^2 for the Zn^{2+} compounds at level 1 although in this case, Zn has a higher charge than Cu .

Combining these 2 trends, it is observed that at level 3, the errors which were way out of the acceptable few millihartrees range were compounds that had both a highly charged metal centre and longer ligands attached to it. For example: $[Ni(butylamine)_6]^{4+}$ had an error of -67 mH; $[Ti(butylamine)_6]^{4+}$ have an error or -39mH; and $[Co(butylamine)_6]^{3+}$ have an error of -20 mH at level 3.

Thirdly, the higher the level of fragmentation, the smaller the error involved. This can be clearly seen from Figures 6.3, 6.4 and 6.5. This trend is clearly observed for d^0 and d^{10} compounds. To explain this, one has to consider the electronic environment of any particular group within a molecule. In order to accurately describe

the electronic structure about this group in a fragment, one has to include as much of the parent molecule as possible that surrounds this group or region. Thus one may assume that the bigger the fragments, the better the energy approximation.

To further elaborate, we refer to the method of grouping of $[\text{Sc(ethylamine)}_6]^{3+}$ in Table 6.1 , using group G_1 to illustrate this point. Referring to Table 6.1, for level 1 fragmentation, we first identify all the fragments that contain G_1. They are found to be positive fragments Frag 1 = G_1^*,G_1 and Frag 7 = G_1,G_2, and with negative fragment Frag 14 =G_1, which accounts for the double counting of G_1. In Figure 6.1, the G_1 unit in the full molecule is a $-CH_2$. However, in positive fragments, Frag 1 and 7, the G_1 unit is actually a $-CH_3$ after hydrogen capping, whereas for negative fragment, Frag 14, the G_1 unit is a CH_4. This would mean that different "types" of CH ligand bonds are cancelled as though they were equivalent at level 1. Thus when you add up the positive and negative fragments up, you would get:

CH_3 (Frag 1) + CH_3 (Frag7) $-$ CH_4 (Frag 14)

= CH_2 (full molecule)
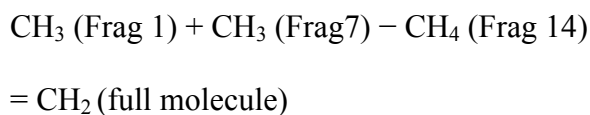
This would be one of the major sources of error in level 1.

At level 2, however, the same type of C$-$H bonds is cancelled off. Taking the G_1 unit again, for level 2 fragmentation, we have positive fragments Frag 1 = G_1^*, G_1, G_3, G_5, G_7, G_9, G_{11} and Frag 2 = G_1^*, G_1, G_2 with negative fragment Frag 8 = G_1^*, G_1. The G_1 unit in Frag 1 is a $-CH_3$ after hydrogen capping, whereas the G_1 unit in Frag 2 is a $-CH_2$, which is the same as the $-CH_2$ G_1 unit of the full molecule. For the negative fragment, Frag 8, the G_1 unit is a $-CH_3$ after hydrogen capping. Thus when you add the positive and negative fragments up, you would get:

CH_3 (Frag 1) + CH_2 (Frag 2) $-$ CH_3 (Frag 8)

= CH_2 (full molecule)
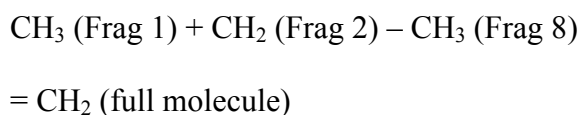
Thus, it is shown that the same type of C–H bonds is cancelled out at level 2. This would give better approximations of the energy calculations in level 2 compared to level 1.

At level 3, due to the larger fragments involved, local bonding CH−CH interactions are accounted for correctly as more of the parent molecule is included in a fragment, thus more accurate energy calculations are obtained.

However, if the absolute errors of all 3 levels are compared for d^6 compounds seen from Table 6.5, it can be observed that this trend is not strictly followed. For example: for $[Fe(ethylamine)_6]^{2+}$, the level 2 error at −3.3 mH was larger than the level 1 error at 2.3 mH. This is observed for $[Fe(propylamine)_6]^{2+}$ and $[Co(butylamine)_6]^{3+}$ as well. This does not imply that the fragmentation at level 1 is better than level 2 and could be due to competing errors canceling each other out at level 1.

Overall, it was observed that the fragmentation method works better for compounds with low charge and shorter ligands. Generally, the d^0, d^6 and d^{10} systems all were able to produce errors close to a few millihartrees. The most significant errors were observed for $[Ni(butylamine)_6]^{4+}$ and $[Ti(butylamine)_6]^{4+}$ systems. This is likely due to the combination of its high charge and of longer ligands used.

The general trend is that the energy error gets smaller up the hierarchy but only on an average. It is observed that the higher the level of fragmentation, the more reliable the estimates of energy calculations becomes.

6.3.3 Tetrahedral Complexes

Table 6.6 lists the full information from tetrahedral compounds calculated in this study. The entire molecule possessed planar structure. Different from octahedral compounds, we relaxed the constriction of the characters of sample molecules. We

investigated molecules in an open shell, not only in a closed shell whose multiplicity is 1. And the same molecular centre but different d orbital electrons. For this set transition metals are chosen in a wide range varying from d^0 to d^{10}. Under these conditions, it would meaningful for analyzing the effect of different charges, the relationship between different multiplicity and the accuracy.

6.3.4 Accuracy Analysis

All the results from fragmentation and full theory results are shown in Table 6.5 and table 6.7, it can be seen that this new approach is able to reproduce the conventional HF and density function theory (B3LYP) energies within a few millihartree in most cases, average error is -5.2 mH on L3, 8.7 mH, which are improved a lot than previous researches.[4-6,17-21] For HF the results is slicely worse than B3LYP, this is due to the inter shortcomings of HF theory, density function theory deals electron correlation better than HF on accounting using the same basis sets.

From Table 6.7, it can be indicated that both L3 and L2 underestimated the total energies, while L1 overestimated. It is obviously that Level 3 results show the most accurate results compared with using full ab intio theory, the second better is L2, then L1.This conclusion is also supported the assumptions of Deev and Colins. V^{3+} however showed contradicting results, with errors of 46 mH and 17mH respectively at level 3. The approximations also worsen up the hierarchy. This is surprising as better approximations are expected up the hierarchy. $V(L)_2$ is an exception, the error value for L3 is -46 mH,-24 mH for L2 and 43 mH for L1.Absolute error value for L1 is half reduced and more or less the same as L3's.Checking the population analysis, the place where we cleaved the bond between carbon-carbon is actually not exactly a single bond ,it is partial double bond, in other words, the electron density is not that localized. Considering the development of fragmentation theory, this problem about

delocalized electrons, generally double or triple bonds, couldn't be broken within fragmentation scheme.

The fifth column of table lists the results putting point charge in all Ll fragments. The point charge is a ghost charged point without mass. Therefore it didn't increase total calculation time. It can be seen that the results are improved greatly compared to the ones without point charge in column 4. The addition of charge is most needful for the level 1 calculations which could be seen by the significant improvement of errors. The findings agreed with our assumption. This is due to the fact that level 1 does not take into account the local non-bonding interactions of the fragments thus the addition of the charge would improve the error to an acceptable range of millihartrees. Another possible reason may be that if we put the point charge in the all L1 fragments, the electrostatic environment is very close to the real electron environment in the parent molecule without fragmenting.

Preliminary tests by Hatree-Fock method was done in Sc^{3+}, Fe^{2+} and Zn^{2+} tetrahedral complexes to test the accuracy of this fragmentation method on tetrahedral structures. From Table 6.8, it can be observed that d^0 and d^{10} complexes works well, with the level 3 errors for the Sc^{3+} and Zn^{2+} complexes in an acceptable range of a few millihartrees. The trends that applied to the octahedral complexes were also true for the d^0 and d^{10} tetrahedral complexes. Indeed it can be seen that as the size of the complex increase, the error increased as well. For example, for $[Zn(ethylamine)_4]^{2+}$, the error was -0.469 mH at level 3; for $[Zn(propylamine)_4]^{2+}$, the error was -1.77 mH; and for $[Zn(butylamine)_4]^{2+}$, the error was -4.79 mH. From Figure 6.5, it is observed that the higher the level of fragmentation, the smaller the errors involved. This trend is clearly observed for the d^0 and d^{10} compounds.

For $[Fe(ethylamine)_4]^{2+}$, level 3 error was 640 mH. This value is way out from the acceptable error range. The spin for the Fe tetrahedral complexes were fixed at $(2S+1) = 1$. Combining this with the t_{2g} configuration, this forces the two electrons at the t_2 level to be paired. The huge error value could be due to the 3 orbitals in t_2 all slightly differing in energy. Thus, the energy of the positive fragments (refer to Table 6.1) corresponds to an excited state and not the lowest energy state. This is shown from the energy calculated for the positive fragments in level 3, all of which were all +160 mH higher than the energy values of same fragments in $[Fe(propylamine)_4]^{2+}$. Coincidently, with all 4 positive fragments of $[Fe(ethylamine)_4]^{2+}$ showing the same error, this would give an error of +640 mH, which also corresponds to the error calculated in level 3, seen in Table 6.8 This is also shown for $[Fe(butylamine)_4]^{2+}$ at level 3.

Referring to Table 6.9, the energy values for fragments 1, 2, 3 are all about +165 mH higher compared to fragment 4. The energy value for fragment 5 is also about +165 mH higher compared to fragment 6, 7 and 8. This differing energy values for similar fragments could account for the huge errors for Fe tetrahedral complexes. To further confirm this, it is observed that the Sc and Zn tetrahedral complexes, which gave acceptable error values of a few millihartree range at level 3, did not have this differing energy values for similar fragments.

6.4 Conclusion

In this chapter, we presented an efficient systematic fragment-based approach for predicting the total energies of macromolecules at the HF and B3LYP levels. The approach done by Bettens[24] and Lee is based on fragmentation of a molecule and a combination of fragments' energies, which is similar to the approach adopted by Zhang and Deev to calculate interaction energies between two molecules. The new

algorithms fragment a molecule or simple radical in a hierarchy of larger fragments was presented. In this way, the energy and energy derivatives could be obtained in a series of approximations of increasing reliability. Tests of this new technique were carried out for the first three members of this hierarchy, levels 1, 2, and 3.

6.4.1 Purpose and Achievements

Initial studies on charged organometallic compounds using this fragmentation method was done and the total electronic energies were found to be accurate to a few millihartrees for the d^0 to d^{10} for both the octahedrals and tetrahedrals. This method provides accurate molecular total energies involving large organometallic compounds through full quantum mechanical electron structure calculations. By breaking the complexes into individual fragments that are properly capped on different levels, the total energy of a molecule of interest at a given structure can be obtained by proper combination of the fragments' energies. The extra interactions between the molecule and the introduced caps are canceled out by subtraction of the artificial molecules formed by hydrogen caps.

The method is shown to converge from level 1 to level 3 for 21 octahedral and 10 tetrahedral complexes with a first row transition metal, by considering errors in the total energy. By performing a systematic study of the scheme for levels of theory, it is demonstrated that the errors in these quantities do not depend on the level of *ab initio* theory or basis set used. Our results indicate that higher levels in the hierarchy produce molecular fragments of larger size and approximate the total electronic energy more reliably. This is maybe due to the non-bonding interactions between different groups. As an example, at Level 1, we broke the adjacent bonds, that is , the bonds on either side of a single group, and we always get the smallest fragments on Level 1.However

on Level 2, there must be a common bond between the breaking point, and this produces bigger fragments than L2, and Level 3, Level4 etc. L1's the fragments are so close to each other, if we summed up the adjacent intra-atomic interactions it would be a huge number, however, when we calculated the individual fragments' energies, we don't include the effect of fragments non bonding interaction. For higher hierarchy, the effect reduced error due to the relative long distance between different groups.

The inclusion of ghost point charge was examined. Comparing the errors of the addition of charge with those without adding of charges for each molecule, it is clear that the addition of charge is needful for more accurate energy calculations. The addition of charge is most needful for the level 1 calculation seen by the significant improvement of errors. For example: The error with point charge at L1 was found to be under 0.16 kcal mol^{-1} for at B3LYP/6-3G(d,p)5d for 10 tetrahedral compounds, and 0.07 kcal mol^{-1} lower than L1. This is due to the fact that level 1 does not take into account the local non-bonding interactions of the fragments thus the addition of the charge would improve the error to an acceptable range of millihartrees. Although this few tested molecular number of approximation appears to be at a useful level of accuracy for moderate complicated cases, further development of this aspect of the model may prove to be necessary for completely general applications.

We find that the fragmentation scheme can be used to predict the total energy of large molecules only from energy calculations on a series of small subsystems. The approach computationally achieves linear scaling with the molecular size. Our test calculations on a broad range of octahedral and tetrahedral organometallic molecules demonstrate that this approach is able to reproduce the conventional HF and Density Function theory (B3LYP) energies within a few millihartrees in most cases. The fragmentation method works better for compounds with low charge and shorter

ligands. In addition, the approach could be applied to estimate the heats of formation for a series of organic compounds, NMR spectroscopic. This approach provides an appealing alternative approach to the traditional additivity rules based on either bond or group contributions for the estimation of thermochemical properties**.**

6.4.2 Advantages

Compared to previous fragment-based methods, the present approach has several advantages. First, the approach may be used in a wide applications, HF, MP2, or CCSD, and DFT methods. Second, the approach is computationally the simplest one among existing fragment-based methods For example, the approach does not really construct the total density matrix of the whole macromolecule from the density matrixes of the capped fragments, like the ADMA[4-6] approach and other methods. In addition, it is important to note that the accuracy of these fragmentation approximations is only very weakly dependent on the level of *ab initio* theory and basis set comparing table 6.4 and 6.5. This independence of basis set and level of theory means that the accuracy of the fragmentation approximations can be tested at the HF/6-31G(d,p) 5d level and then relied on at higher levels, where calculation of the real energy is impractical, but calculation of the fragment energies is relatively much easier. Thus, much higher level *ab initio* calculations can be carried out for the whole molecule.

Finally, the systematic fragment-based approach enables one to calculate the total energy of molecule without performing the calculation of whole molecule. The method is, therefore, expected to reduce the computational time of large molecule drastically. Since the computational cost of the present method scales linearly with the molecular size in large molecules, the total energy of systems with hundreds or

thousands of electrons can now be approximately computed at the ab initio HF and B3LYP levels with existing quantum chemistry programs Gaussian 03 package.

6.4.3 Problems and Limitations

Some limitations of the current approach were demonstrated by the calculation of the energies of partial double bond molecules like the one with centre metal Vanadium, with errors up to 0.05 hartrees in relative energies for the case of $V(L)_2$. It still needs more improvement to the up present research. Finally, as an illustration, the scheme was tested on a larger series of same type of structure with same ligand. Convergence in the energy error was observed by going from level 1 fragmentation to level 3.

Generally, a point to take note of when doing fragmentation of organometallics is the crystal field splitting. Discrepancies in errors are often to do with fragments that include the metal centre. Since the extent of crystal field splitting is dependant on the nature of the ligands, there will have to be a difference in crystal field splitting when you break away the ligands from the central metal. Referring to Table 6.1, at level 1 for $[Sc(ethylamine)_6]^{3+}$, the structures of Frag 1−5 are essentially the same, all of which are Sc^{3+} metal binding to five amine ligands and one methylamine ligand. This would have a different crystal field splitting compared to the original $[Sc(ethylamine)_6]^{3+}$ molecule which in turn would cause differences in the energy calculations. However, looking at the level 3 fragments for $[Sc(ethylamine)_6]^{3+}$, the structures of Frag 1−6 are also essentially the same, all of which are Sc^{3+} metal binding to five methylamine ligands and one ethylamine ligand. Thus, better approximation of crystal field splitting is expected up the hierarchy since the fragments better resembles the original parent compound at level 3. This problem of

crystal field splitting of the metal centre has to be tackled in future work in order to prove that this fragmentation method can work for organometallics.

Fragmentation of charged organometallics proved to be more complex and difficult to do compare to neutral organic molecules due to its charge and crystal field splitting. For organometallics, the charge of the metal centre proved to cause significant deviation of energy calculations from its true value. An effective and systematic way of accounting for the charge influence of the metal centre to the ligands attached has to be considered carefully and employed for charged molecules. In this paper, the approach taken to account for the metal charge was the introduction of the metal charge to all the fragments. Another approach that would be feasible as well is to include the charged metal centre itself to all the fragments. The challenge would be to find the best possible approach to correct for the long range electrostatic interactions for charge molecules such that the energies calculations would be as accurate as possible. This is crucial as many biomolecules contain a charged metal centre, such as hemoglobin, cytochromes and other bioenzymes.

6.4.4 Future Work and Applications

As this method of fragmentation is fairly new, there is a wide range of work that has to be done to determine its accuracy of *ab initio* calculations. As with what Deev and Collins have pointed out, accurate results may not be seen for certain systems of fragmentation, however, there exists a wide variety of systems in which this method of fragmentation has the potential to work well. Thus accuracy of this fragmentation method has to be tested on other organometallic complexes in order to conclude that it is indeed useful for transition metal complexes and biomolecules. A variety of ligands can be tested on, such as complexes with phosphine ligands or

porphyrin rings systems. The effect of different geometries such as trigonal planar and trigonal bipyramidal should also be investigated. Fragmentation of organometallics that involve ligands with double bonds or aromatic ring structures can also be studied. The inclusion of non-bonding interactions in the energy calculations could be introduced as well. Other practical methods to make corrections to the errors of this approach will be desirable.

The accuracy of the fragmentation approximations for organic compounds was found to be only weakly dependent on the level of *ab initio* theory and basis set. However, this can be extended to organometallic compounds by using higher level of theories such as B3LYP, MP2 and QCISD and various basis sets.

The energy approximation improves up the hierarchy but only on an average. Future work will be necessary to handle special cases with multiple chemical bonds and non bonding interaction between fragments. In addition, it is also desirable to introduce practical methods to make corrections to the errors of the systematic fragment-based approach.

To summarize, preliminary studies on the fragmentation approach proposed by Deev and Collins is demonstrated to be effective in approximating the total electronic energies of charged organometallic compounds. Further developments have to be done to make this approach a useful tool for performing *ab initio* calculations on very large molecules.

6.5 References

1   Bettens, R. P. A., J.Phys. Chem. A 2004, 108, 1826-1829.
2   Yang, W., Phys. Rev. Lett. 1991, 66, 1438.
3   Yang, W., J. Chem. Phys. 1991, 94, 1208.
4   Exner, T. E., Mezey, P. G., J. Phys. Chem. 2002, 106, 11791.
5   Exner, T. E., Mezey, P. G., J. Comput. Chem. 2003, 24, 1980.
6   Exner, T. E., Mezey, P. G., J. Phys. Chem. 2004, 108, 3599.
7   Kitaura, K., Ikeo, E., Asada, T., Nakano, T., Uebayashi, M., Chem. Phys. Lett. 1999, 313, 701.
8   Kitaura, K., Sugiki, S.-I., Nakano, T., Komeiji, Y., Uebayashi, M., Chem. Phys. Lett. 2001, 336, 163.
9   Fedorov, D. G., Kitaura, K., J. Chem. Phys. 2004, 120, 6832.
10   Fedorov, D. G., Kitaura, K., Chem. Phys. Lett. 2004, 389, 129.
11   Imamura, A., Aoki, Y., Maekawa, K., J. Chem. Phys. 1991, 95, 5419.
12   Korchowiec, J., Gu, F. L., Imamura ,A., Kirtman, B., Aoki, Y., Int. J. Quantum Chem. 2005, 102, 785.
13   Gadre, S. R., Shirsat, R. N., Limaye, A. C., J. Phys. Chem. 1994, 98, 9165.
14   Babu, K., Gadre, S. R., J. Comput. Chem. 2003, 484.
15   Babu, K., Ganesh, V., Gadre, S. R., Ghermani, N. E., Theor. Chem. Acc. 2004, 111, 255.
16   Das, G. P., Yeates, A. T., Dudis, D. S., Int. J. Quantum Chem. 2004, 120, 6832.
17   Zhang, D. W., Zhang, J. Z. H., J. Chem. Phys. 2003, 119, 3599.
18   Zhang, D. W., Xiang, Y., Zhang, J. Z. H., J. Phys. Chem. B 2003, 107, 12039.
19   Gao, A. M., Zhang, D. W., Zhang, J. Z. H., Zhang, Y., Chem. Phys. Lett. 2004, 394, 293.
20   Mei, Y., Zhang, D. W., Zhang, J. Z. H., J. Phys. Chem. A 2005, 109, 2.
21   He, X., Zhang, J. Z. H., J. Chem. Phys. 2005, 122, 12039.
22   Li, X., Millam, J. M., Scuseria, G. E., Frisch, M. J., Schlegel, H. B., J. Chem. Phys. 2003, 119, 7651.
23   Deev, V., Collins, M. A., J. Chem. Phys. 2005, 122, 154102
24   Bettens, R. P. A., Lee, A. M., J. Phys. Chem. A 2006, 110, 28, 8777-8785.
25   M. J. Frisch, et al., GAUSSIAN 03, Gaussian. Inc., Pittsburgh, PA, 2003.

6.6 Appendix

Table 6.1 The positive and negative fragments generated during fragmentation for [Sc(ethylamine)$_6$]$^{3+}$ in levels 1, 2 and 3.

Level 1		Level 2		Level 3	
Positive Fragments	Negative Fragments	Positive Fragments	Negative Fragments	Positive Fragments	Negative Fragments
1 Frag 1 = G_1^*,G_1	−5 Frag 13 = G_1^*	1 Frag 1 = $G_1^*, G_1, G_3, G_5, G_7, G_9, G_{11}$	−1 Frag 8 = G_1^*, G_1	1 Frag 1 = $G_1^*, G_1, G_2, G_3, G_5, G_7, G_9, G_{11}$	−5 Frag 7 = $G_1^*, G_1, G_3, G_5, G_7, G_9, G_{11}$
1 Frag 2 = G_1^*,G_3	−1 Frag 14 = G_1	1 Frag 2 = G_1^*, G_1, G_2	−1 Frag 9 = G_1^*, G_3	1 Frag 2 = $G_1^*, G_1, G_3, G_4 , G_5, G_7, G_9, G_{11}$	
1 Frag 3 = G_1^*,G_5	−1 Frag 15 = G_3	1 Frag 3 = G_1^*, G_3, G_4	−1 Frag10 = G_1^*, G_5	1 Frag 3 = $G_1^*, G_1, G_3, G_5, G_{6,} G_7, G_9, G_{11}$	
1 Frag 4 = G_1^*,G_7	−1 Frag 16 = G_5	1 Frag 4 = G_1^*, G_5, G_6	−1 Frag11 = G_1^*, G_7	1 Frag 4 = $G_1^*, G_1, G_3, G_5, G_7, G_{8,} G_9, G_{11}$	
1 Frag 5 = G_1^*,G_9	−1 Frag 17 = G_7	1 Frag 5 = G_1^*, G_7, G_8	−1 Frag 12 = G_1^*, G_9	1 Frag 5 = $G_1^*, G_1, G_3, G_5, G_7, G_9, G_{10,} G_{11}$	
1 Frag 6= G_1^*,G_{11}	−1 Frag 18 = G_9	1 Frag 6 = G_1^*, G_9, G_{10}	−1. Frag 13 = G_1^*,G_{11}	1 Frag 6 = $G_1^*, G_1, G_3, G_5, G_7, G_9, G_{11}, G_{12}$	
1 Frag 7 = G_1,G_2	−1 Frag 19 = G_{11}	1 Frag 7 = G_1^*, G_{11},G_{12}			
1 Frag 8 = G_4,G_3					
1 Frag 9 = G_5,G_6					
1 Frag 10 = G_7,G_8					
1 Frag 11 = G_9,G_{10}					
1 Frag 12 =G_{11},G_{12}					

Table 6.2 Deev and Collins (D&C) Fragmentation vs the Fragmentation given in this work for the tetrahedral organometallics molecules. (cf.Figure 6.2)

	D&C Level-3		Level-3	
	coef	frag	coef	frag
1	1	1234567	2/3	12367
2	1	1234589	2/3	12389
3	-1	12345	2/3	123456
4			2/3	123457
5			2/3	123458
6			2/3	123459
7			-1/3	1238
8			-1/3	1239
9			-1/3	1456
10			-1/3	1457
11			-5/3	12345

Table 6.3 The comparision of the errors in the total electronic energy (in H) of octahedral closed shell octahedral complexes with and without the addition of the central metal charge to all the fragments that do not include group G1* .Calculated at HF/6-31G (d) 5d.

		Error (E) = (F) − (T), whereby (E) is the absolute error in Hartrees, (F) is the calculated value by fragmentation and (T) is the full molecule energy calculation					
		Level 1		Level 2		Level 3	
Organometallic compound	config.	No charge added	Charge added	No charge added	Charge added	No charge added	Charge added
$[Sc(ethylamine)_6]^{3+}$	d^0	0.0519	−0.0188				
$[Sc(propylamine)_6]^{3+}$		0.0798	−0.0161	0.0116	−0.0230		
$[Sc(butylamine)_6]^{3+}$		0.0988	−0.0076	0.0200	−0.0277	0.0035	−0.0144
$[Fe(ethylamine)_6]^{2+}$	d^6	−0.8484	0.0013				
$[Fe(propylamine)_6]^{2+}$		−2.3445	0.0083	−0.4114	−0.0107		
$[Fe(butylamine)_6]^{2+}$		−3.5145	0.0161	−0.8767	−0.0151	−0.2361	−0.0069
$[Zn(ethylamine)_6]^{2+}$	d^{10}	−1.1500	0.0045				
$[Zn(propylamine)_6]^{2+}$		−2.3433	0.0116	−0.4114	−0.0096		
$[Zn(butylamine)_6]^{2+}$		−3.5236	0.0201	−0.8788	−0.0130	−0.3197	−0.0064

Table 6.4 Energy (in H) of selected fragments, Frag = 2, 3 and Frag = 2 for [Sc(ethylamine)$_6$]$^{3+}$, [Fe(ethylamine)$_6$]$^{2+}$ and [Zn(ethylamine)$_6$]$^{2+}$ complexes, with and without the addition of charge. Calculated at HF/6-31G (d) 5d.

	Frag = 2, 3		Frag = 2	
	Without adding of charge	With adding of charge	Without adding of charge	With adding of charge
[Sc(ethylamine)$_6$]$^{3+}$	−79.222	−79.266	−40.183	−40.214
[Fe(ethylamine)$_6$]$^{2+}$	−77.937	−79.244	−38.710	−40.199
[Zn(ethylamine)$_6$]$^{2+}$	−77.960	−79.242	−38.720	−40.197

Table 6.5 Absolute error in the total electronic energy (in H) of close shell octahedral complexes at fragmentation levels 1, 2, 3 using HF/6-31g (d) 5d.

First row transition metal complex	Config.	Error=calculated value – true value (H)		
		Level 1	Level 2	Level 3
$[Sc(ethylamine)_6]^{3+}$	d^0	-0.0188	-0.0079	-0.0017
$[Sc(propylamine)_6]^{3+}$		-0.0161	-0.0230	-0.0070
$[Sc(butylamine)_6]^{3+}$		-0.0076	-0.0277	-0.0144
$[Ti(ethylamine)_6]^{4+}$		-0.0853	-0.0245	-0.0047
$[Ti(propylamine)_6]^{4+}$		-0.0811	-0.0630	-0.0201
$[Ti(butylamine)_6]^{4+}$		-0.0726	-0.0687	-0.0395
$[Fe(ethylamine)_6]^{2+}$	d^6	0.0013	-0.0033	-0.0007
$[Fe(propylamine)_6]^{2+}$		0.0083	-0.0107	-0.0027
$[Fe(butylamine)_6]^{2+}$		0.0161	-0.0151	-0.0069
$[Co(ethylamine)_6]^{3+}$		-0.0424	-0.0114	-0.0023
$[Co(propylamine)_6]^{3+}$		-0.0378	-0.0338	-0.0091
$[Co(butylamine)_6]^{3+}$		-0.0301	-0.0390	-0.0202
$[Ni(ethylamine)_6]^{4+}$		-0.1383	-0.0397	-0.0068
$[Ni(butylamine)_6]^{4+}$		-0.1186	-0.1043	-0.0665
$[Cu(ethylamine)_6]^{+}$	d^{10}	0.0228	-0.0002	-0.0001
$[Cu(propylamine)_6]^{+}$		0.0322	-0.0022	-0.0002
$[Cu(butylamine)_6]^{+}$		0.0397	-0.0058	-0.0020
$[Zn(ethylamine)_6]^{2+}$		0.0045	-0.0029	-0.0007
$[Zn(propylamine)_6]^{2+}$		0.0116	-0.0096	-0.0026
$[Zn(butylamine)_6]^{2+}$		0.0201	-0.0130	-0.0064

Table 6.6 Studied tetrahedral complexes. L is the notation for ligand hexa-1,5-diyne-3,4-diamine

orbital occupancies	charges	Multiplicity (2S+1)	Studied sample molecules
d^0	3+	1	ScL2
d^2	2+	3	TiL2
	3+	3	ViL2
d^4	3+	1	MnL2
d^5	2+	6	MnL2
d^7	2+	4	CoL2
d^8	2+	1	NiL2
d^9	2+	2	CuL2
d^{10}	2+	1	ZnL2
	1+	1	CuL2

Table 6.7 The comparison of the errors in mH between full theory B3LYP and fragmentation method at L1, L2, L3

[M(L)$_2$]	L3-Full	L2-Full	L1-Full	L1 with point charge
Sc	-5.2	-12.7	36.6	19.0
V	-46.0	-24.6	43.3	10.7
Ti	-1.27	-3.48	7.69	-83.7
Mn(II)	-10.5	-25.9	49.3	-4.11
Mn(III)	-1.23	-3.51	8.45	27.8
Co	16.8	-3.45	3.14	19.4
Ni	-1.09	-2.81	12.8	24.5
Cu(II)	-1.31	-3.12	22.1	26.5
Cu(I)	-1.40	-3.93	10.6	34.9
Zn	-1.24	-3.63	9.64	25.3

Table 6.8 Absolute error in the total electronic energy (H) of closed-shell tetrahedral complexes using HF/6-31g (d) 5d, with spin = 1.

Complex	Config.	Error=calculated value – true value (H)		
		Level 1	Level 2	Level 3
$[Sc(ethylamine)_4]^{3+}$	d^0	-0.0180	-0.0060	-0.0014
$[Sc(propylamine)_4]^{3+}$		-0.0278	-0.0196	-0.0058
$[Sc(butylamine)_4]^{3+}$		-0.0275	-0.0250	-0.0119
$[Fe(ethylamine)_4]^{2+}$	d^6	0.0175	0.4845	0.6402
$[Fe(propylamine)_4]^{2+}$		0.0045	0.0291	-0.0379
$[Fe(butylamine)_4]^{2+}$		0.0039	0.4538	0.1626
$[Zn(ethylamine)_4]^{2+}$	d^{10}	0.0017	-0.0020	-0.0005
$[Zn(propylamine)_4]^{2+}$		0.0047	-0.0071	-0.0018
$[Zn(butylamine)_4]^{2+}$		0.0103	-0.0091	-0.0048

Table 6.9 Energy values of positive fragments at level 3 containing metal centre for $[Fe(butylamine)_4]^{2+}$, using HF/6-31g (d) 5d.

Level 3 positive fragments containing metal centre	Energy value (H)
Frag 1 = 2, 6,10,14, 1, 3,	−1681.501410
Frag 2 = 2, 6,10,14, 1, 7,	−1681.504539
Frag 3 = 2, 6,10,14, 1,11,	−1681.504619
Frag 4 = 2, 6,10,14, 1,15,	−1681.668689
Frag 5 = 1, 3, 2, 4,	−1603.466851
Frag 6 = 1, 7, 6, 8,	−1603.635018
Frag 7 = 1,11,10,12,	−1603.618039
Frag 8 = 1,15,14,16,	−1603.634907
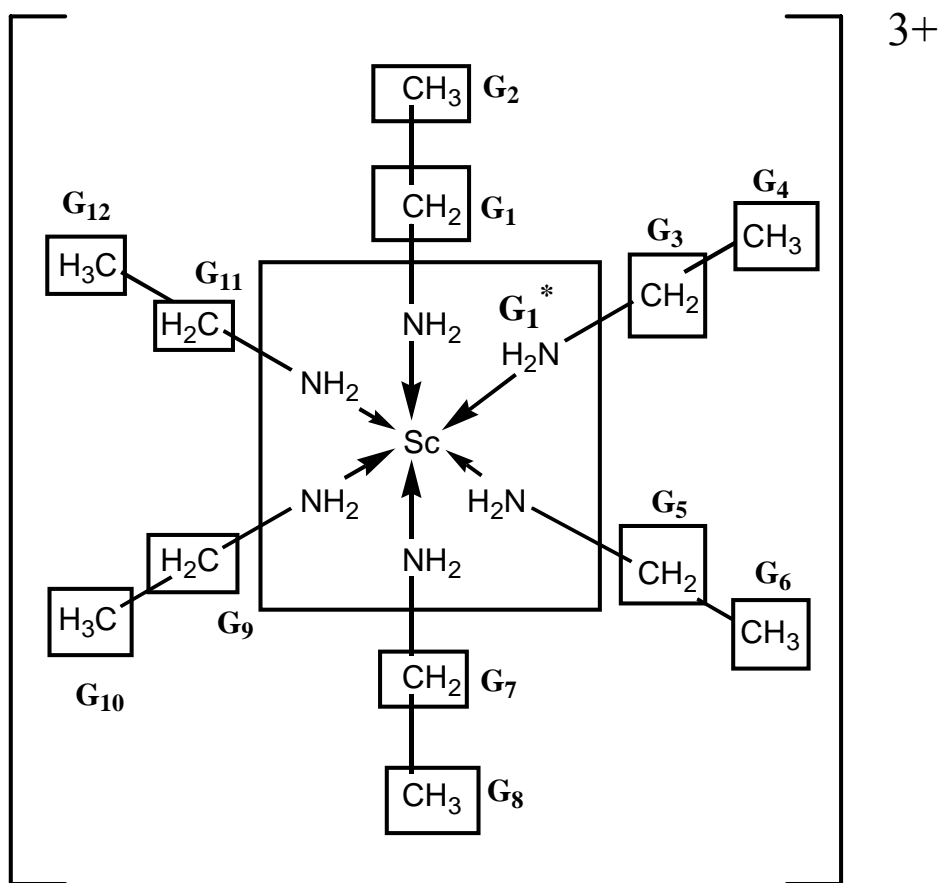
Figure 6.1 Method of grouping of [Sc(ethylamine)$_6$]$^{3+}$.
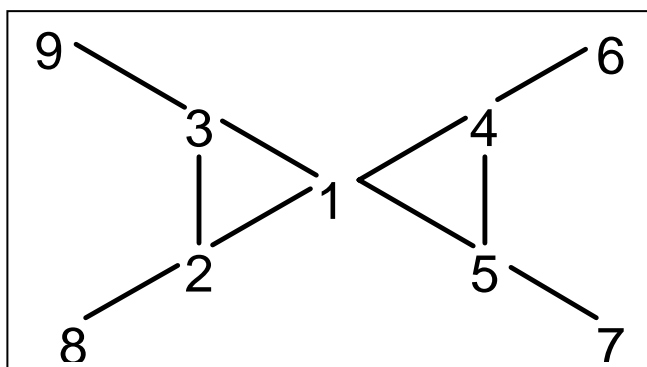
Figure 6.2 Example of the tetrahedral morphology to be fragmented under the

Deev and Collins scheme and the scheme presented in this work. Groups
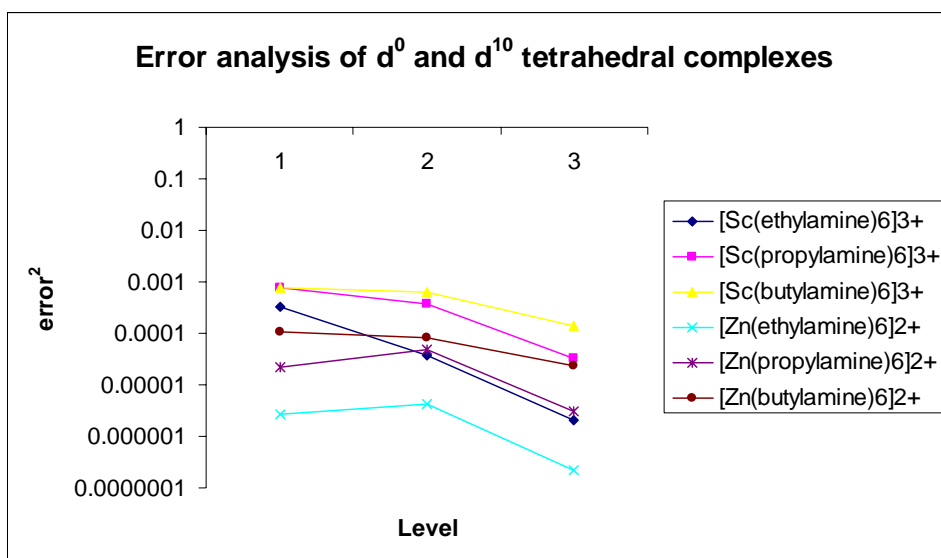
are labeled with numbers.
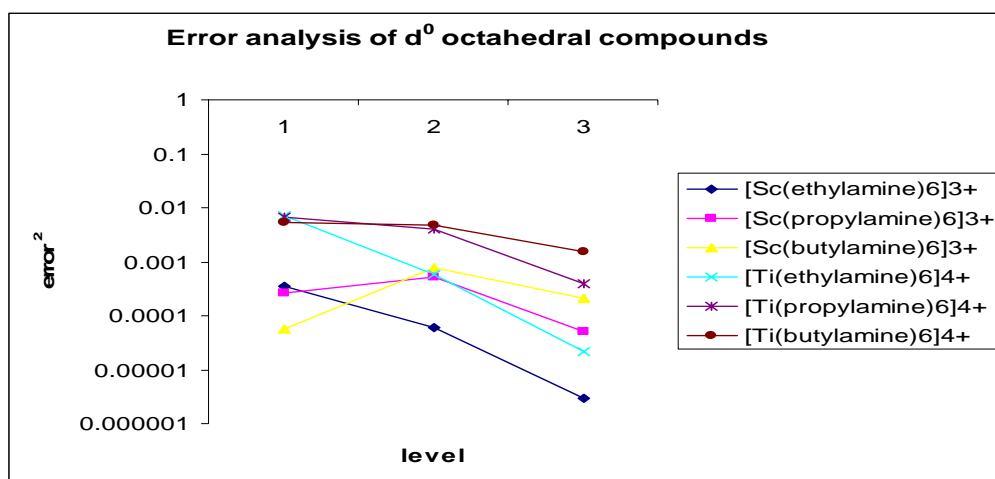
Figure 6.3 Error analysis of the total electronic energy (H) of close shell d^0 octahedral

complexes using HF/6-31g.
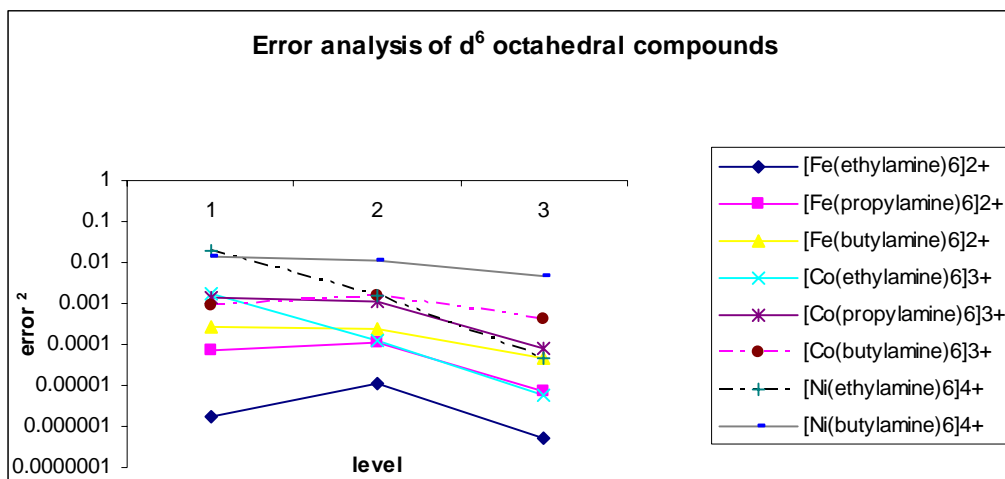
Figure 6.4 Error analysis of the total electronic energy of close shell d^6 octahedral complexes using HF/6-31g.
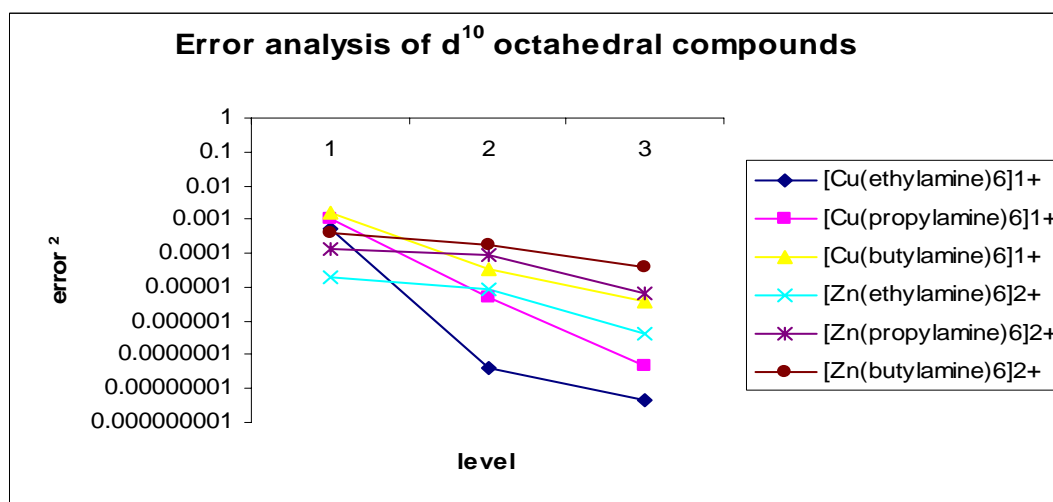
Figure 6.5 Error analysis of the total electronic energy of close shell d^{10} octahedral complexes using HF/6-31g.

Figure 6.6 Error analysis of the total electronic energy of close shell d^0 and d^{10} tetrahedral complexes using HF/6-31g.

SUPPORTING PUBLICATIONS

Approximating Coupled Cluster Level Vibrational Frequencies with Composite Methods

Yanping Fan, Junming Ho, and Ryan P. A. Bettens*

Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543

Received: October 28, 2005; In Final Form: December 23, 2005

An extensive study of the harmonic frequencies of a large set of small polyatomic closed-shell molecules computed at both single level ab initio and composite approximations is presented here. Using various combinations of basis sets, composite methods are capable of predicting single level ab initio CCSD(T) harmonic frequencies to within 5 cm^{-1} on average, which suggests a computationally affordable means of obtaining highly accurate vibrational frequencies compared to the CCSD(T) level. A general approach for calculating the composite level equilibrium geometries and harmonic frequencies for polyatomic systems that uses the Collin's method of interpolating potential energy surfaces is also described here. This approach is further tested on tetrafluoromethane, and an estimation of the potential CPU time savings that may be obtained is also presented. It is envisaged that the findings here will enable theoretical studies of fundamental frequencies and energetics of significantly larger molecular systems.

1. Introduction

Since the introduction of G1 theory by Pople and co-workers in 1989[1] a sizable literature has appeared that utilizes composite methods or, more generally, methods that use various lower levels of ab initio or DFT theory to approximate significantly higher levels of theory. The advantage in doing so lies in the very significant saving in computational expense resulting from the lower level computations. G1 theory and its descendants, G2[2], G3,[3] G3S,[4] G3X[5] were originally developed to achieve "chemical accuracy" (energies to within 4 kJ mol^{-1} when compared with experiment) in the computation of thermochemical properties (enthalpies, ionization energies, electron affinities, etc.) of gases. Indeed, this level of accuracy has been achieved for many molecules.

The Gn theories of Pople and co-workers are by no means the only methods that aim to, and achieve, chemical accuracy by approximating expensive higher level methods using several lower level results and empirical parameters. Some of the more popular include the complete basis set (CBS) methods from Petersson and co-workers,[6] the Weizmann-n theories and their variants of Martin and co-workers[7] and the multicoefficient correlation methods (MCCMs)[8] of Truhlar's group.

Significantly fewer studies have appeared in the literature that utilize composite methods for predicting potential energy surfaces (PES). Collins and co-workers have successfully utilized a G3X(MP2) type method in the construction of PES for reactive systems and the calculation of various kinetic parameters.[9] Such methods have also been used in a nine-dimensional bound-state problem for the determination of zero-point energies and ground-state rotational constants.[10] Császár and co-workers utilized a CBS approach to generate a base PES for water and then added in a core-correlation surface, a relativistic correction surface, a quantum electrodynamics correction surface and an adiabatic correction surface.[11]

Other groups have considered up to quartic expansions of the potential about an equilibrium configuration. In these studies

it is the fundamental frequencies of vibration that are of interest, as well as other spectroscopic constants. Bose and Martin[12] published a detailed study on the azabenzene series, which included considering the possibility of combining DFT anharmonic force fields with coupled cluster geometries and harmonic frequencies. Pouchan and co-workers have also combined harmonic ab initio force constants with DFT anharmonicity constants in a number of studies.[13]

Although high accuracy can be obtained using the above approach for computing fundamentals, high-level ab initio calculations are still required of the harmonic frequencies. Furthermore, such approaches to obtaining a PES, although perfectly suited for the determination of spectroscopic observables of tightly bound systems, are not applicable over the entire PES but presumably can only be applied to turning points. An alternative approach is to define a potential energy that can be computed for any single configuration that is composed of contributions from various levels of theory in a manner similar to Gn theory. In this way, not only can composite force constants and anharmonic force constants be computed, but composite energies, gradients and second and higher derivatives can also be evaluated for *any* configuration.

Though the high accuracy of composite methods has been demonstrated by numerous studies for total energies, at least at and around minima on the PES, almost no work has been done on examining the general accuracy of the approach for first and higher order derivatives. One way of measuring the accuracy of the curvature of the PES is by comparing composite harmonic frequencies to those obtained using a high single level of theory.

The computation of vibrational frequencies has seen much interest in recent years, with frequencies determined on average to within 8 cm^{-1} of experimental values using CCSD(T) and large basis sets.[14] However, the CPU time associated with this method scales as the seventh power of the number of basis functions, which makes the calculation for even medium-sized molecules prohibitive. Of course, one must include the effects of correlating the core−electrons to achieve such high levels of accuracy. Dunning and Peterson have examined the use of composite methods for making reliable estimates of the elec-

tronic energy, spectroscopic properties (D_e, r_e, ω_e, $\omega_e x_e$), ionization energy, and electron affinities compared with the single level CCSD(T)/aug-cc-pV5Z for a test set of diatomic molecules.[15] Specifically, the authors calculated an energy at the CCSD(T) level using a smaller basis aug-cc-pVXZ, X = D, T, and Q, and then added to this energy a correction, Δ_{basis}, to account for the inadequate basis set. Their study revealed that for the test set of molecules, the composite approach is capable of predicting single level CCSD(T) harmonic frequencies to within 2 cm^{-1} on average when X = T. This approach has also been successfully applied for the calculation of harmonic and fundamental frequencies for first-row closed shell diatomic molecules.[16] Thus, if the success of composite methods for computing energies could be carried over into the calculation of vibrational frequencies, then significantly larger molecular systems can be studied with high accuracy.

However, for this to be possible, it is first necessary to establish the general applicability of composite methods for the calculation of other vibrational modes viz. molecular bends and torsions. To the best of the authors' knowledge, all previous studies utilizing a G*n*-type approach have been restricted to simple diatomic systems, where only uncomplicated stretching modes are assessed. In this work, the harmonic frequencies at both single level ab initio and G*n*-type composite approximations of CCSD(T) theory, are reported for 19 tri- and 18 tetratomic nonlinear molecules where the bends and torsions are examined as well. Additionally, a general scheme for calculating the composite level equilibrium geometries and harmonic frequencies for polyatomic systems that utilizes the Collins' method of interpolating potential energy surfaces is also described. The accuracy of the composite-level harmonic frequencies are evaluated through comparison with the corresponding single level CCSD(T) calculations.

It is envisaged that the results of this study should provide a clearer indication of the general applicability of composite methods for the calculation of vibrational frequencies of more complicated molecules. Furthermore, this would also contribute toward an alternative procedure for calculating highly accurate ab initio frequencies of larger molecules with significant reductions in computational cost.

2. Computational Details

The single level ab initio calculations were carried out at CCSD(T)/aug-cc-pVXZ, where X = D, T and Q. The calculations were performed using the MOLPRO 2002.1[17] and Gaussian 98[18] suite of programs. The composite energies were based on the ad hoc expression

$$E_{L/S} = E[CCSD(T)/S] + \Delta_{basis} \qquad (1)$$

where $\Delta_{basis} = E[MPn/L] - E[MPn/S]$, MP*n* refers to *n*th-order Moller Plesset perturbation theory, and S and L denote small and large basis sets, respectively. $E_{L/S}$ is an approximation to the energy at the CCSD(T)/L level of theory. Note that if the MP*n* treatment in the basis set correction term Δ_{basis} was substituted with the CCSD(T) treatment, then this would yield exactly the CCSD(T)/L energies. This expression is similar to the electronic energy given in G3X(MP*n*) theory in refs 9 and 10 and is the same as that used in refs 15 and 16. It can thus be seen from eq 1 that one significant source of error in this approximation is the difference in treatment of electron correlation between the MP*n* and CCSD(T) levels.

In the subsequent sections, short-hand notations to describe the above calculations are D, T and Q for CCSD(T)/aug-cc-

pVDZ, CCSD(T)/aug-cc-pVTZ and CCSD(T)/aug-cc-pVQZ, respectively. Similarly, composite methods are denoted by L/S, where L and S are shorthand notations of the above basis sets. It should be noted from eq 1 that the energy is defined for *any* molecular configuration, not just locally at and around minima, and provides a means to generate a composite potential energy surface (PES), as discussed earlier. Because each term in eq 1 is differentiable with respect to Cartesian displacements of the atoms so too is the composite energy. Thus we are able to obtain a composite equilibrium structure and harmonic frequencies. All molecular structures in this work have been optimized using both composite and single level ab initio methods specifying tight convergence. A threshold for the convergence of the energy in the SCF procedure of 10^{-10} Hatree has also been chosen in all calculations.

To calculate the L/S harmonic frequencies of a molecule, a PES is first required to locate its L/S optimized geometry. The PES was constructed using Collins' method of interpolation and has been described in detail elsewhere.[19] Once the PES minimum is located, the second derivative matrix is calculated numerically at this geometry and the harmonic frequencies obtained. The algorithm for obtaining the L/S harmonic frequencies of an *N*-atom nonlinear polyatomic molecule is described below:

1. Obtain an approximate set of normal coordinates (Z_1, Z_2, ..., Z_{3N-6}) at a lower level ab initio method such as MP2/6-31G(d), where analytic calculation of the Hessian matrix is possible. The optimized geometry, $\mathbf{Z_o}$, at this level of theory serves as an initial guess to the composite method equilibrium structure.

2. The L/S gradient and Hessian matrices are evaluated numerically by central difference formulas at $\mathbf{Z_o}$. This generates the initial L/S PES, which corresponds to a second-order Taylor polynomial about $\mathbf{Z_o}$.

3. The minimum point, $\mathbf{Z_1}$, of this PES is located using the Newton−Raphson method. The process repeats from step two, generating the next data point. After more than one data point is generated, the PES is expressed as an interpolation over the total number of data points, N_{data}, based on eq 2, where $w_n(\mathbf{Z})$ and $T_n(\mathbf{Z})$ refer to the normalized distance-based weight function and second-order Taylor approximation of the *n*th data point at $\mathbf{Z}$.

$$V(\mathbf{Z}) = \sum_{n=1}^{N_{data}} w_n(\mathbf{Z}) \, T_n(\mathbf{Z}) \qquad (2)$$

where

$$w_n(\mathbf{Z}) = \frac{v_n(\mathbf{Z})}{\sum_{i=1}^{N_{data}} v_i(\mathbf{Z})} \qquad v_n(\mathbf{Z}) = ||\mathbf{Z} - \mathbf{Z}(n)||^{-2p}$$

$$2p > 3N - 3 \quad (3)$$

4. The optimization is deemed converged if all the calculated gradient elements (∇V_i, $i = 1, ..., 3N - 6$) of the newest data point are less than or equal to an ad hoc value, ϵ_{tol}; otherwise the algorithm repeats from step 2. The final data point is the L/S optimized geometry, $\mathbf{Z_{eq}}$ expressed in terms of the MP2/6-31G(d) normal coordinates.

5. $\mathbf{Z_{eq}}$ is expressed in terms of the $3N - 6$ standard **Z**-matrix internal coordinates, where the Hessian with respect to these coordinates is calculated numerically. The L/S harmonic

TABLE 1: Test Set of Molecules Used

triatomic	H_3^+ H_2F^+ CH_2 CHF H_2O HNO HON NH_2^+ NH_2^- HO_2^+ HO_2^- OCF^- HF_2^+ HOF HNF^- HCO^- CF_2 C_2O F_2O
tetratomic	CFH_2^- NFH_2 H_2CO H_2O_2 OFH_2^+ *cis*-N_2H_2 *trans*-N_2H_2 *trans*-HCOH *cis*-HCOH *cis*-$HCNH^-$ *trans*-$HCNH^-$, CH_3^- NH_3 OH_3^+ H_2NO^+ H_2NO^- H_2NN H_2CN^-

TABLE 2: Comparison of D and T/D Frequencies with T Harmonic Frequencies for Triatomic Systems

method	MAD	RMS	$|\Delta\omega|_{median}$	$|\Delta\omega|_{max}$
D	37.0	45.7	28.4	108.1
T/D	4.46	6.88	3.12	27.1

frequencies are obtained in the usual manner from the Hessian and atomic masses.

All numerical derivatives were evaluated using a step size of 5×10^{-4} au and ϵ_{tol} was specified as 5×10^{-5} au, which corresponds to the tight convergence criteria in the Gaussian software package. In all the molecules examined, the geometry optimization converged within three cycles.

3. Results and Discussion

Table 1 shows the full list of molecules that were examined in this study. Unless otherwise stated, all composite frequencies were evaluated using MP2 theory in eq 1. We shall first examine the results for the triatomic molecules, followed by the tetratomic and larger systems.

3.1. Triatomics. The single level D, T and composite T/D harmonic frequencies for 19 triatomic molecules have been calculated, providing a sample of 57 bending and stretching frequencies for comparison (see Table S1 in the Supporting Information). The data for the T/D and D harmonic frequencies are compared to the T frequencies and are summarized in Table 2. As mentioned earlier, the CCSD(T) theory has an intrinsic error of about 8 cm^{-1} in terms of the calculation of experimental vibrational frequencies, and including core-correlation (not included in this work). Thus, it is desirable that the composite harmonic frequencies lie within the same range of their single level CCSD(T) counterparts. It is clear by examining the data in Table 2 that a substantial improvement in the accuracy of the harmonic frequencies is achieved using composite methods compared with the D frequencies. For example, the T/D mean absolute deviation (MAD) and root-mean-square (RMS) values are 4.46 cm^{-1} and 6.88 cm^{-1}, which are about 8 times smaller compared to the D frequencies with MAD and RMS values of 37.0 and 45.7 cm^{-1} respectively.

The distribution of the absolute deviation (AD) values for the 57 T/D and D frequencies is illustrated in Figure 1. From the distribution curves, it was observed that the absolute deviations in the D frequencies are fairly evenly distributed, with errors as large as 108 cm^{-1}. On the other hand, about 70% of the T/D frequencies are within 5 cm^{-1} of the T frequencies, and at least 95% within 15 cm^{-1}. However, it was noted that two (originating from HCO$^-$ and HON) of the 57 T/D frequencies had absolute deviations in excess of 20 cm^{-1}, where the maximum was 27.1 cm^{-1}. Likewise, the absolute deviations in the corresponding D frequencies were found to be in excess of 40 cm^{-1}. Further inspection revealed that these frequencies arose from the highest frequency stretching modes of these two molecules. Curiously, the remaining T/D vibrational frequencies of the two molecules are relatively accurate and fall within 12 cm^{-1} of the corresponding T frequencies.

Generally speaking, the errors in the composite expression in eq 1 are likely to propagate and impact most on the high frequency vibrational modes. The fact that the two outlying

Absolute Deviation (AD) Distribution Curve
Triatomic Systems
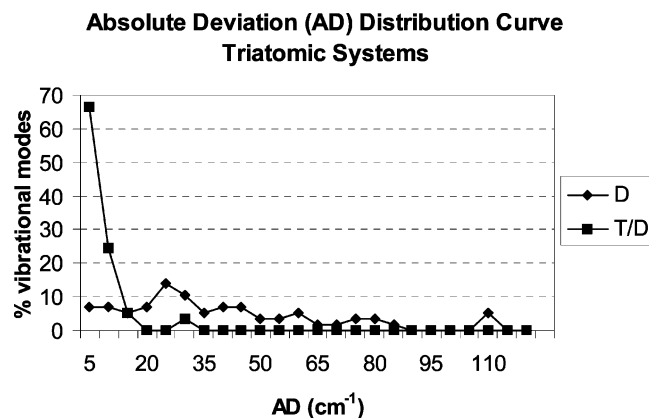
Figure 1. Plot of the percentage of vibrational modes against the absolute deviation from the T frequencies for the triatomic systems.

TABLE 3: Comparison of Q/T, Q/D, T and D Frequencies with Q Harmonic Frequencies for Triatomic Systems

| method	MAD	RMS	$|\Delta\omega|_{median}$	$|\Delta\omega|_{max}$
T	9.1	11.0	10.2	20.7
D	44.7	54.0	34.9	118.0
Q/T	1.3	1.61	1.2	2.9
Q/D	4.1	5.0	3.9	9.8
T/D	12.2	14.3	12.9	24.7

frequencies correspond to the highest stretching frequencies of two molecules attest to this. There are two main sources of error in the composite frequencies: First, the gradient vectors and Hessians were evaluated numerically via central difference and must therefore incur some errors in the harmonic frequencies. More significantly, the use of MPn in the basis set correction term Δ_{basis} must be taken into consideration. Presumably, the anomalously large deviations in the composite frequencies for the two systems are due to the inadequate treatment of electron correlation by the MP2 procedure. As pointed out earlier, this error can be improved by systematically increasing the level of electron correlation in the basis set correction term.

As such, the harmonic frequencies for the two molecules were reevaluated by substitution of MP3 (see Table S4 in the Supporting Information) for MP2 in eq 1. This led to a marked improvement in the two outlying frequencies where the deviations were reduced to less than 7 cm^{-1}. There was also further improvement in the other frequencies of these molecules where the AD with the T frequencies was reduced to less than 3 cm^{-1}. Similarly, upon substitution with the corresponding MP3 T/D frequencies for the two molecules, the MAD and RMS values were further reduced from 4.46 and 6.88 cm^{-1} to 3.31 and 4.95 cm^{-1}, respectively. These observations suggest that the high-frequency vibrations tend to be more sensitive to the inexactness of the composite expression.

Additionally, the single level Q and composite levels Q/T and Q/D were also computed for a subset of the six lightest triatomic molecules shown in Table S2 of the Supporting Information, and summarized in Table 3. Also provided in Tables S2 and 3 are the results for T/D, T and D harmonic frequencies versus the Q frequencies. The Q/T frequencies were of comparable accuracy to the Q frequencies, with a MAD of only 1.3 cm^{-1}, compared to a MAD of 9.1 cm^{-1} in the T frequencies. It was also noted that the performance of the Q/D frequencies was slightly worse compared to the Q/T frequencies, with an MAD of 4.1 cm^{-1}, although this is within the acceptable error range. Not surprisingly, the T/D frequencies do not predict the Q frequencies as accurately as the former two but compares

Coupled Cluster Level Vibrational Frequencies

J. Phys. Chem. A, Vol. 110, No. 8, 2006 **2799**

TABLE 4: Comparison of D and T/D Frequencies with T Harmonic Frequencies for Tetratomic Systems

method	MAD	RMS	$|\Delta\omega|_{median}$	$|\Delta\omega|_{max}$
D	20.3	23.7	18.9	63.9
T/D	4.2	5.4	3.5	20.0

Absolute Deviation (AD) Distribution Curve Tetratomic Systems
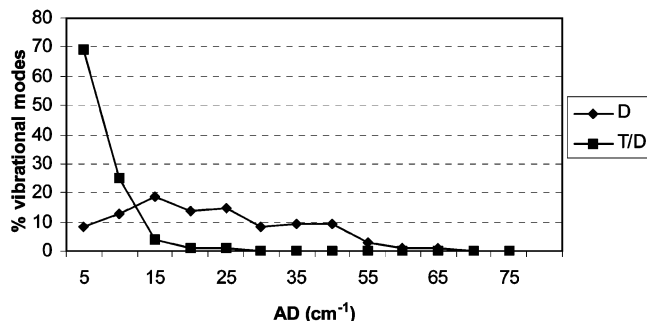
Figure 2. Plot of the percentage of vibrational modes against the absolute deviation from the T frequencies for the test set of tetratomic systems.

well with the T frequencies as illustrated by the good agreement between their MAD and $|\Delta\omega|_{median}$ values.

The above observations imply that the optimal combination of basis sets (L and S) for predicting single level L harmonic frequencies is when they differ by no more than one in the valence designation. It is possible that the widening difference in the valence designation of the two basis sets (L and S) would deteriorate the quality of the basis set correction term Δ_{basis}, thereby leading to poor agreement with the CCSD(T)/L frequencies.

3.2. Tetratomics Systems. The single level T, D and composite T/D harmonic frequencies are also calculated for a set of 18 tetratomic molecules, providing a sample of 108 stretching, bending and torsional modes for comparison. These molecules have geometries ranging from tetrahedral, trigonal pyramidal to planar structures. Table 4 summarizes our results, whereas Table S3 in the Supporting Information provides all the frequencies.

The performance of the composite frequencies in the tetratomic systems is consistent with the triatomic systems. Here, the MAD value of the T/D frequencies from the T calculations is merely 4.2 cm^{-1}, which is about a 5-fold reduction compared to that of the D frequencies at 20 cm^{-1}. The distribution of the AD of the 108 T/D and vibrational frequencies is plotted in Figure 2.

The distribution curves in Figure 2 illustrates a trend similar to that in Figure 1 where about 95% of the T/D frequencies lie within 10 cm^{-1} of the T frequencies, although it was observed that a small number had absolute deviations greater than 12 cm^{-1} with $|\Delta\omega|_{max}$ of 20 cm^{-1}. Further examination revealed that these frequencies arose from high frequency stretching modes of several tetratomic molecules. On the contrary, the remaining vibrations of these molecules generally showed good agreement with deviations of 10 cm^{-1} or less. To assess the errors due to the composite approximation, the composite frequencies were reevaluated using the MP3 rather than MP2 in eq 1 for the two of molecules, H_2CN^- and H_2NN, with the largest deviations (19.3 and 20.0 cm^{-1}).

Consequently, both deviations were substantially reduced to 0.36 and 10.7 cm^{-1}, respectively (see Table S4 in the Supporting Information). Likewise, the deviations for the remaining fre-

TABLE 5: CPU Times Associated with the MP2 and CCSD(T) Calculation at the Equilibrium Geometry of CF_4

basis set	no. of basis functions	MP2	MP3	CCSD(T)
aug-cc-pVDZ	115			1702.90
aug-cc-pVTZ	230	112.68	593.96	24687.76

quencies were further reduced to less than 4 cm^{-1}. Substitution of these frequencies for the two molecules with the MP3 composite frequencies led to improved MAD and RMS values of 3.75 and 4.58 cm^{-1}, respectively.

Thus far, the results have been supportive of the capacity of the composite procedure to make reliable predictions of the harmonic frequencies corresponding to bending and torsional modes. However, it has also been noted that the high frequency vibrational modes, specifically stretches, tend to be more sensitive to the errors incurred in the composite approximation. These errors are primarily due to the inaccuracy of the basis set correction term in eq 1. Our preliminary assessment shows that the correction term may be systematically refined by using higher-order perturbation methods such as the MP3 procedure. This observation was also reported in the study by Dunning and Peterson on diatomic molecules, where the MP3 composite procedure out-performed its MP2 counterpart.[15]

Despite the higher accuracy and consistency in the MP3 approximation, there is also the added computational cost as the CPU time associated with this method scales as the sixth power of the number of basis functions. On the other hand, the MP2 composite procedure is generally very accurate with errors less than 5 cm^{-1} on average. Hence, for a given CPU time budget, the MP2 approximation should be useful for many molecular studies.

3.3. CPU Time Savings. The major advantage with the composite approach is the ability to predict single level CCSD-(T) harmonic frequencies accurately, while only requiring a significantly shorter CPU time. Based on the MP2 procedure, the composite approach is approximately a factor of n times faster:

$$n = \frac{t\{CCSD(T)/L\}}{t\{MP2/L\} + t\{CCSD(T)/S\}} \qquad (4)$$

where $t\{CCSD(T)/L\}$ refers to the CPU time incurred for the CCSD(T) and large basis set calculation, and so forth.

To estimate the CPU time-savings that may be obtained, the composite procedure was applied to tetrafluoromethane, which is composed of five heavy atoms. Based on a single point calculation at the T/D equilibrium geometry, the CPU times required by the T and T/D procedures are tabulated in Table 5.

Accordingly, it is estimated that the CCSD(T)/aug-cc-pVDZ calculations are approximately 14.5 times faster than CCSD-(T)/aug-cc-pVTZ. Quite remarkably, the CPU times associated with the composite approximations are exceedingly close, where n has been estimated to be 13.5 and 10.7 for the MP2 and MP3 procedures, respectively.

Additionally, the T/D frequencies for CF_4 have also been computed and compared with the corresponding T harmonic frequencies from earlier work of Wang et al.[20] The frequencies are tabulated in Table 6.

As shown in Table 6, all the T/D frequencies are in excellent agreement with the T frequencies, with errors of 3.0 cm^{-1} or less. This result is most noteworthy considering the mere additional cost of performing a MP2 energy calculation. It also appears that for a medium-sized system molecule such as tetrafluoromethane, the difference in the CPU times required

TABLE 6: Computed CCSD(T) Harmonic Frequencies (cm^{-1})

vibrational mode	T	T/D	T-T/D
1	435.2	434.1	1.1
2	630.4	628.4	2.0
3	915.2	912.4	2.8
4	1301.3	1298.5	2.8

for MP3 and MP2 is somewhat small when compared with the single level CCSD(T) calculations. Accordingly, the MP3 approximation may be more advantageous in terms of reliability for small to medium-sized molecules.

4. Concluding Remarks

In this paper, the harmonic frequencies for a test set of closed shell triatomic and tetratomic molecules have been calculated at both single level and composite approximations of the CCSD-(T) method. The results of this study demonstrate the ability of the composite approximation to make very accurate predictions of the harmonic frequencies that are within 5 cm^{-1} of the corresponding single level CCSD(T) calculation. All previous studies have focused exclusively on simple diatomic molecules, where only stretching modes were examined. Through the work presented here it is established that the composite procedure is equally capable of making accurate predictions of other vibrational frequencies corresponding to bending and torsional modes for more complicated polyatomic systems.

The poorer estimation of the stretching frequencies for polyatomic molecules has been attributed to the fact that stretching modes are invariably the high-frequency vibrations and are therefore more sensitive to the errors in the energy expression in eq 1. Nevertheless, it has been demonstrated in problematic systems that the large deviations in the T/D harmonic frequencies are readily remedied through the use of MP3 procedure. The tradeoff, however, is the increased computational cost associated with this method, which scales as the sixth power of the number of basis functions.

To summarize, it is conceivable that the combination of efficient Hessian update schemes combined with the theoretical procedure presented here should enable the study of significantly larger molecular systems.

Supporting Information Available: Table S1 contains the CCSD(T)/aug-cc-pVXZ (X = D and T) and the composite, T/D, harmonic frequencies (cm^{-1}) for the triatomic molecules. Table S2 contains the CCSD(T)/aug-cc-pVXZ (X = D, T and Q) and the composite Q/T, Q/D and T/D harmonic frequencies (cm^{-1}) for the six lightest triatomics. Table S3 contains the CCSD(T)/aug-cc-pVXZ (X = D and T) and composite, T/D, harmonic frequencies (cm^{-1}) for the tetratomic molecules. Table S4 contains the two tetratomics H_2CN^- and H_2NN, and two triatomics, HCO^- and HON, CCSD(T)/aug-cc-pVTZ and composite T/D harmonic frequencies (cm^{-1}) using MP3 and MP2 in the Δ_{basis} correction. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

(2) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem Phys.* **1998**, *109*, 42.

(3) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *Chem. Phys. Lett.* **1999**, *313*, 600. Kedziora, G. S.; Pople, J. A.; Rassolov, V. A.; Ratner, M.; Redfern, P. C.; Curtiss, L. A. *J. Chem. Phys.* **1999**, *110*, 7123. Curtiss, L. A.; Redfern, P. C.; Rassolov, V.; Kedziora, G.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 9287.

(4) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 1125.

(5) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108.

(6) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys.* **1996**, *104*, 2598. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(7) Martin, J. M. L.; Oliveira, G. d. *J. Chem. Phys.* **1999**, *111*, 1843. Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Ka'llay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(8) Fast, P. L.; Corchado, J.; Sanchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 3139. Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.

(9) Bettens, R. P. A.; Collins, M. A.; Jordan, M. J. T.; Zhang, D. H. *J. Chem. Phys.* **2000**, *112*, 10162.

(10) Bettens, R. P. A. *J. Am. Chem. Soc.* **2003**, *125*, 584.

(11) Polyansky, O. L.; Császár, A. G.; Shirin, S. V.; Zobov, N. F.; Barletta, P.; Tennyson, J.; Schwenke, D. W.; Knowles, P. J. *Science* **2003**, *299*, 539. Császár, A. G.; Czakó, G.; Furtenbacher, T.; Tennyson, J.; Szalay, V.; Shirin, S. V.; Zobov, N. F.; Polyansky, O. L. *J. Chem. Phys.* **2005**, *122*, 214305.

(12) Boese, A. D.; Martin, J. M. L. *J. Phys. Chem. A* **2004**, *108*, 3085.

(13) Begue, D.; Carbonniere, P.; Barone, V.; Pouchan, C. *Chem. Phys. Lett.* **2005**, *415*, 25. Gohaud, N.; Begue, D.; Pouchan, C. *Int. J. Quantum Chem.* **2005**, *104*, 773. Begue, D.; Carbonniere, P.; Pouchan, C. *J. Phys. Chem. A* **2005**, *109*, 4611.

(14) Preface in *Spectrochim. Acta*, *Part A* **1997**, *53*, vii.

(15) Dunning, T. H., Jr.; Peterson, K. A. *J. Chem. Phys.* **2000**, *7799*, 113.

(16) Bettens, R. P. A. *J. Phys. Chem.* **2004**, *108*, 1826.

(17) Werner, H.-J.; Knowles, P. J. version 2002.1, Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Knowles, P. J.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Werner, H.-J. MOLPRO 2002.6, a package of ab initio programs.

(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6 and A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(19) Jordan, M. J. T.; Thompson, K. C.; Collins, M. A. *J. Chem. Phys* **1995**, *102*, 5647. Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302. Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 816.

(20) Wang, X. G.; Silbert III, E. L.; Martin, J. M. L. *J. Chem. Phys.* **2000**, *112*, 1353.

The Conformers of Hydroxyacetaldehyde

Yanping Fan, Lai Peng Leong, and Ryan P. A. Bettens*

Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543

*Received: July 12, 2006; In Final Form: April 11, 2007*

Both two and eighteen dimensional quantum diffusion Monte Carlo (DMC) calculations were used to study the isomers of hydroxyacetaldehyde. A total of four unique minima, and the transition states connecting them, were located. Both two and eighteen dimensional potential energy surfaces were generated and used in the DMC runs. The rotational constants for the global minimum were predicted for all experimentally identified isotopomers and an approximate equilibrium structure obtained by combining our theoretical results with the experimentally observed rotational constants. The results obtained for the remaining isomers indicate that not all of them can be isolated in the gas phase.

Introduction

Hydroxyacetaldehyde, (hydroxyethanal, glycolaldehyde, CH_2-OHCHO), an isomer of methyl formate and acetic acid, has been recently seen toward the Galactic center cloud Sgr B2(N) (Hollis, Lovas, and Jewell 2000).[1] Generally, it is believed that saturated molecules in hot cores are synthesized on interstellar dust grains in a low-temperature era. However the synthesis of glycolaldehyde is currently unknown toward its source. Several research groups have paid attention to glycolaldehyde, its origin in interstellar clouds,[2] oxidation,[3] and its reaction with the OH radical[4,5] both experimentally and theoretically.

Marstokk and Møllendal[6–8] first systematically studied the structure of glycolaldehyde in the gas phase but only observed one isomer. Their microwave measurements of the parent molecule and deuterated species as well as three other isotopic species also included the dipole moment. They pointed out that the cis form, denoted GM in this paper, is the most stable conformer based on low level theoretical calculations of three possible conformers, denoted here as GM, L1, and L3. Later, it was found that there was a fourth conformer of glycolaldehyde (cf. Figure 1), denoted in this work as L2, in the theoretical work of Antero et al.[9] Recently, Senent[10] studied the torsional spectrum and interconversion process between the four conformers at the MP2/cc-pVQZ level using a two-dimensional variational approach. In addition, Senent[10] computed the rotation parameters corresponding to respective conformers.

In this paper our main focus is on the implications of introducing the full dimensionality of the potential energy surface (PES) of hydroxyacetaldehyde in the identification and assignment of different conformers. Because hydroxyacetaldehyde is an 8-atom system the total number of nuclear degrees of freedom is 18. The only feasible technique available to study systems with such high dimensionality is Diffusion Monte Carlo (DMC), which is used throughout this work. Quantum DMC is now routinely used to solve for the ground-state nuclear wavefunction and has been applied to a wide variety of systems including the 30-dimensional intermolecular modes of the water hexamer[11] and 142 and 452 torsions of a bimolecular system.[12] Systems where no degrees of freedom have been frozen have

* To whom correspondence should be addressed. Email: chmbrpa@nus.edu.sg. Fax: +65 6779 1691.
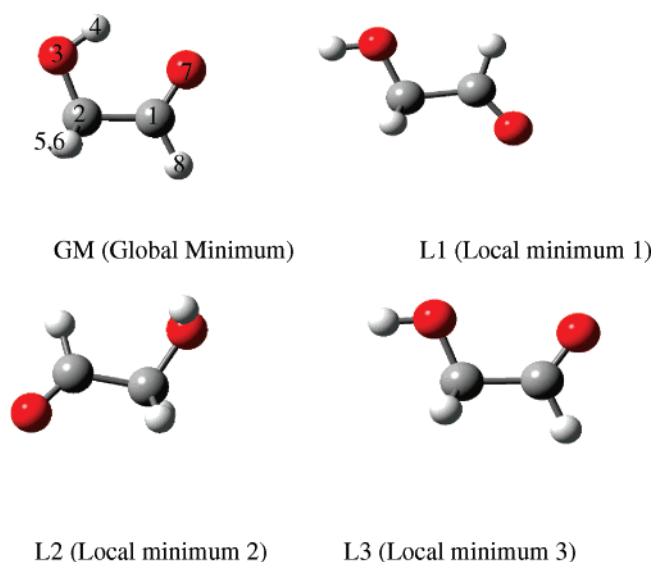
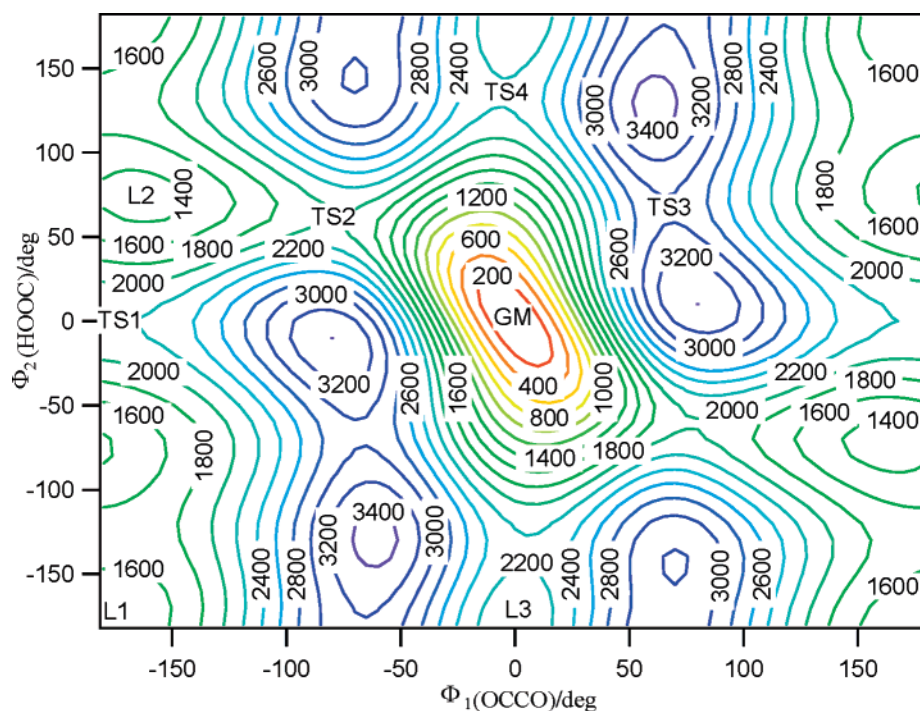
Figure 1. The four conformer structures of glycolaldehyde.

also been studied. Examples include 9 dimensions for of FCH_3[13], and 12 dimensions for $CH_5{}^+$[14], and the water dimer.[15,16]

Methods

All ab initio calculations reported in this work were computed by using the Gaussian 98 suite of programs.[17] The PES was mapped in the two torsional angles, $\phi_1 = -$OCCO and $\phi_2 = -$HOCC by performing B3LYP/6-31G** constrained optimizations from $\phi_1 = -180-180$ in steps of 5° and $\phi_2 = 0-180$ in steps of 10°. That is, a total of 1387 constrained optimizations were performed. A contour plot of this two-dimensional PES is given in Figure 2. Indicated on this figure are all the minima and saddle points with the corresponding energies given in Scheme 1. Table 1 includes the energies of the minima at the B3LYP/6-31G(d,p), B3LYP/cc-pVTZ, CBS^{MP2}, and G3XMP2 levels.

Table 1 compares the four conformers energies at different levels of theory, the order is consistent for the two B3LYP calculations; however, the CBS^{MP2} and G3XMP2 methods both predict L1 to be slightly lower in energy than L2 as was also found by Senate.[10] The CBS calculations were performed in an

Figure 2. Contour plot of hydroxyacetaldehyde (energies in cm^{-1}) as a function of the two torsional angles.

SCHEME 1: Local Minima and Transition States for Hydroxyacetaldehyde at the B3LPY/6-31G** Level of Theory
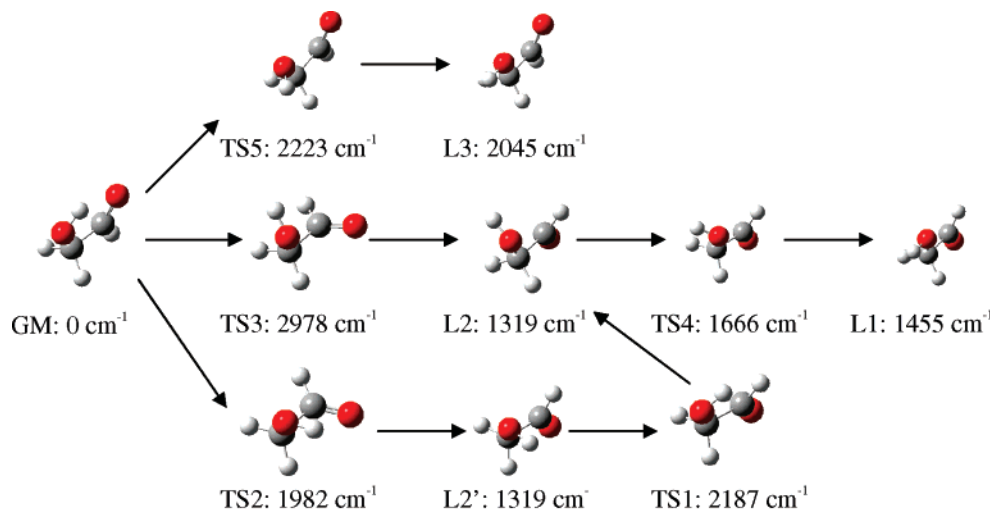
TABLE 1: Relative Energies (cm^{-1}) of Minima at the Various Levels of Theory

minima	B3LYP/6-31G(d,p)	Senent[a]	B3LYP/cc-pVTZ	CBS$_{MP2}$[b]	G3XMP2
GM	0	0	0	0	0
L1	1455	1278	1300	1260	1161
L2	1319	1297	1274	1340	1223
L3	2045	1865	1874	1860	1786
L3 → GM	178	175	158		
L2 → L2′	868	713	716		
L2 → L1	347	191	222		

[a] MP4(SDTQ)/cc-pVQZ//MP2/cc-pVQZ from ref 10. [b] MP2/aug-cc-pVXZ//MP2/aug-cc-pVTZ, X = T, Q and 5 energies were fit to $1/N^{1.5}$, where N was the number of basis functions. Relative energies from extrapolation accurate to ±20 cm^{-1}.

attempt to eliminate any basis set superpositioin error (BSSE) and to correctly describe the H-bond interaction. It is of note that the much higher level calculations agree reasonably well with the lower level B3LYP results. By combination of Table 1 and Scheme 1, it can be seen that L2 and L1 are very similar in energy and the barrier between them is small. Furthermore, the lowest barrier from conformer L3 to GM was calculated to be only 178 cm^{-1} at the B3LYP/6-31G(d,p) level

and 175 cm^{-1} at the much higher level of Senent[10] (MP4-(SDTQ)/cc-pVQZ//MP2/cc-pVQZ). Our results and that of Senate both show that there exists a significant barrier between conformers L1 and the GM as well as L2 and the GM.

In DMC, small displacements are made to the Cartesian coordinates of the atoms. The size of the displacements depends on the mass of the nucleus and the imaginary time step size. As imaginary time passes the structure can change dramatically

Conformers of Hydroxyacetaldehyde

J. Phys. Chem. A, Vol. 111, No. 23, 2007 **5083**

depending on the nature of the potential energy surface. In this work we ran DMC simulations on the B3LYP/6-31G(d,p) surface given in Figure 2 and an 18-dimensional surface derived from a subset of the grid points used to generate the two-dimensional surface, as described later.

In the two-dimensional calculations we have assumed that as the two torsional angles change the molecule is able to readjust its structure to the most stable form for the given values of ϕ_1 and ϕ_2. Hence after each time step we computed the two values of the torsional angles and then reset the remaining internal coordinates (and hence the structure) to an interpolation of the minimum energy structure that corresponded to these two angles. A simple bilinear interpolation was used to obtain the above internal coordinates based on the four sets of optimized internal coordinates, extracted from the grid described previously, that corresponded to the four bracketing pairs of ϕ_1 and ϕ_2. Similarly, a bilinear interpolation was also used to obtain the potential energy of the molecule for the given values of the torsional angles.

Here we specifically implemented discrete sampling DMC with 1000 initial replicas. The population was first pre-equilibrated using a step size of 5 au for 7000 steps in the 2-dimensional surface and 20000 steps on the 18-dimensional surface. After the pre-equilibration the zero-point energy was noted to have converged, and data sampling then occurred every 50 steps over a period of 10 000 steps using a step size of 1 au. The rotational constants were computed also during this period using the method of descendant weights. Descendants were followed for 1000 steps with a new set of descendants initiated and followed every 100 time steps. The reported results for the rotational constants are from 20 and 320 separate runs for the 2- and 18-dimensional surfaces, respectively. The reported errors are two standard deviations of the respective means.

To compute the rotational constants it is necessary to ensure that the Eckart conditions are enforced. A speedy algorithm was developed to ensure this and is essentially the same as that described by Kohn et al.[18] We also utilized the molecular symmetry of the system to effectively double the population size in computing the inverted moment of inertia tensor. Note that while we are always in the Eckart axis system, the inverted moment of inertia tensor is not exactly diagonal, except for the reference configuration. However, the absolute value of the off-diagonal elements for all isotopomers of the GM was never more than 21 MHz for the inverted product of inertia about the $a-b$ axes and never more than 2 MHz for the $a-c$ and $b-c$ axes.

We used two approaches for modeling the PES in 18 dimensions. One utilized the B3LYP/6-31G(d,p) level of theory, and the other the energies given by Senent[10] at the MP4(SDTQ)/cc-pVQZ level on the provided grid, but the first and second derivatives of the potential at these points at the B3LYP/6-31G-(d,p) level. At each of the grid points we performed constrained optimizations then obtained the energies, gradients, and second derivatives of the energy with respect to the coordinate system (described below). This data is then used in the Collins interpolation method[19] for evaluating the energy for any given configuration of the system in all 18 dimensions. We did apply one modification to standard Collins scheme, however, and that was to use $3n-6$ internal coordinates rather than the $n(n-1)/2$ interatomic distances. The following set of 18 internal coordinates were chosen (the atom labels can be found in Figure 1 for the GM), $\{r_1(C_1,C_2),\ r_2(O_3,C_1),\ r_3(H_4,C_1),\ r_4(H_5,C_2),\ r_5$-$(H_6,C_2),\ r_6(O_7,C_2),\ r_7(H_8,O_7),\ a_1(O_3,C_1,C_2),\ a_2(H_4,C_1,C_2),\ a_3$-$(H_5,C_2,C_1),\ a_4(H_6,C_2,C_1),\ a_5(O_7,C_2,C_1),\ a_6(H_8,O_7,C_2),\ d_1$-$(H_4,C_1,C_2,O_3),\ d_2(H_5,C_2,C_1,H_4),\ d_3(H_6,C_2,C_1,H_4),\ d_4(O_7,C_2,C_1,O_3),$

TABLE 2: Experimental Rotational Constants for Parent and Isotopomers of the Global Minimum Isomer Compared with the Perturbation Theory and DMC 2D and 18D Constants (Former Two Were Computed at the B3LYP/6-31G(d,p) Level, All Values Are in MHz)

|	experiment[a]	exptl−perturb	exptl−DMC(2D)	exptl−fitted(18D)
			GM	
A	18446.4	106.5	−0.5 ± 1.0	−77.9 ± 17.6
B	6526.0	−52.8	25.1 ± 0.5	−3.6 ± 11.9
C	4969.3	−22.3	3.7 ± 0.2	7.4 ± 6.0
			$CH_2OD-CHO$	
A	17490.8	68.8	−3.4 ± 0.7	−47.7 ± 15.1
B	6499.8	−44.8	29.9 ± 0.4	−13.7 ± 11.7
C	4883.0	−18.9	4.7 ± 0.1	1.4 ± 5.8
			$CH_2OH-CDO$	
A	17151.3	103.0	16.0 ± 0.7	157 ± 18.1
B	6363.0	−47.7	26.5 ± 0.4	1.1 ± 9.3
C	4779.0	−18.3	4.1 ± 0.1	19.8 ± 6.1
			CHDOHCHO	
A	16988.0	104.1	14.5 ± 1.2	69.5 ± 17.4
B	6385.5	−51.0	20.5 ± 0.4	−19.0 ± 10.4
C	4843.8	−20.3	6.5 ± 0.2	13.0 ± 6.7
			$^{13}CH_2OH-CHO$	
A	18126.9	88.5	−13.9 ± 0.7	154.0 ± 19.5
B	6487.5	−49.9	26.5 ± 0.5	6.6 ± 11.5
C	4923.0	−22.4	2.8 ± 0.2	22.9 ± 7.4
			$CH_2OH-^{13}CHO$	
A	18259.5	109.4	4.5 ± 0.7	−49.5 ± 19.1
B	6472.3	−51.9	24.1 ± 0.4	−8.6 ± 11.5
C	4924.6	−21.6	3.6 ± 0.2	5.8 ± 5.6
			$CH_2OH-CH^{18}O$	
A	18087.0	101.5	−0.8 ± 0.8	−66.7 ± 19.0
B	6242.8	−49.0	24.8 ± 0.4	−7.8 ± 11.6
C	4778.5	−21.3	4.4 ± 0.1	4.8 ± 5.9

[a] From ref 6.

TABLE 3: Comparison Parameters of the Fitted Structure and Reference Structure Ref 6

	exptl r_s structure	fitted r_e structure
C=O	1.2094 ± 0.0003 Å	1.2106 Å
C−O	1.4366 ± 0.0007 Å	1.3937 Å
C−C	1.4987 ± 0.0004 Å	1.5079 Å
O−H	1.0510 ± 0.0004 Å	0.9712 Å
H_{ald}−C	1.1021 ± 0.0003 Å	1.1058 Å
H_{alc}−C	1.0930 ± 0.0003 Å	1.0897 Å
−C−C=O	122°44′ ± 2′	121°26′
−C−C−H_{ald}	115°6′ ± 2′	116°52′
−C−C−O	111°28′ ± 2′	111°59′
−C−O−H	101.34° ± 2′	105°19′
−C−C−H_{alc}	109.13° ± 1′	107°50′
−H−C−H	107°34′ ± 2′	104°56′
−H−C−O	109°39′ ± 1′	111°56′

$d_5(H_8,C_2,C_1,O_3)\}$, where r is an interatomic distance, a is bond angle, and d is dihedral angle. Note that $d_4 = \phi_1$ and $d_5 = \phi_2$. The Taylor series about each point of the surface was expanded in inverse r but directly in a and d. The above coordinate set transforms to the following set under the permutation-inversion operation of the molecular symmetry group of hydroxyacetaldehyde, i.e., the operation $(H_5,H_6)^*$, $\{r_1, r_2, r_3, r_5, r_4, r_6, r_7, a_1, a_2, a_4, a_3, a_5, a_6, -d_1, -d_3, -d_2, -d_5, -d_6\}$. Thus this choice of coordinates ensures that the potential possess the correct symmetry properties.

Note that the grid step size of 30° for the 18 dimensional surface is coarse and may result is some non-smooth behavior of the potential when interpolating between ab initio data points. However, DMC, being a statistical method, is well suited to dealing with such potentials. The reader should also note that, while ϕ_1 and ϕ_2 were varied over their entire range of values,
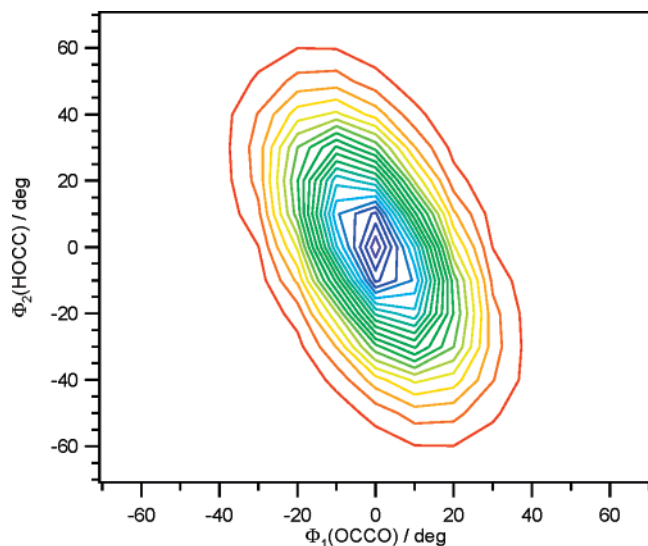
Figure 3. Contour plot the ground state wavefunction for the global minimum. Each contour represents a fall of about 5% in probability amplitude.

Figure 4. Contour plot the ground state wavefunction for the L2 minimum. Each contour represents a fall of about 5% in probability amplitude.

many of the remaining 16 dimensions changed little over the grid of points (e.g., bond lengths and angles). Thus interpolated energies for significantly different values of these coordinates cannot be expected to be accurate. Our 18 dimensional surface is constructed to reasonably accurately describe $V(\phi_1, \phi_2)$ but only qualitatively describe how the potential varies for the remaining 16 degrees of freedom.

Results and Discussion

The experimental rotational constants (errors less than the last significant digit given) are compared with the DMC rotational constants (error arises from the random Monte Carlo component of the simulation) for various isotopomers of the GM in Table 2. We have also included in this table the values expected for the rotational constants using perturbation theory at the B3LYP/6-31G(d,p) level as implemented in Gaussian 03.[20]

The agreement between experiment and the DMC 2-dimensional predictions is remarkable (and most likely fortuitous) considering the level of theory used and the 2-dimensional approximation. A closer examination of Table 2 reveals that the B_0 rotational constant is consistently predicted too low by about 25 MHz, which may indicate the equilibrium structure is marginally too "tight" about this axis at the B3LYP/6-31G(d,p) level. However, the 18-dimensional DMC results differed more significantly from experiment. In this case we varied the reference structural parameters to obtain the best agreement with the experimental rotational constants and obtained the geometry given in Table 3. The corresponding rotational constants using the fitted structure are also given in Table 2. The percentage average absolute error between experimental and theoretical rotational constants was reduced from 0.52 to 0.29% using the fitted structure.

A contour plot of the two-dimensional projection of the ground state wavefunction is given in Figure 3. It is evident from this figure that the hydroxy hydrogen undergoes substantial excursions away from the equilibrium position. Where the probably amplitude falls to half its maximum value Φ_2 is about $\pm 25°$, while Φ_1 travels over $\pm 15°$. Figure 3 also bears out a correlated motion between Φ_1 and Φ_2; the intramolecular H-bonded hydrogen tends to follow the electronegative carbonyl oxygen.
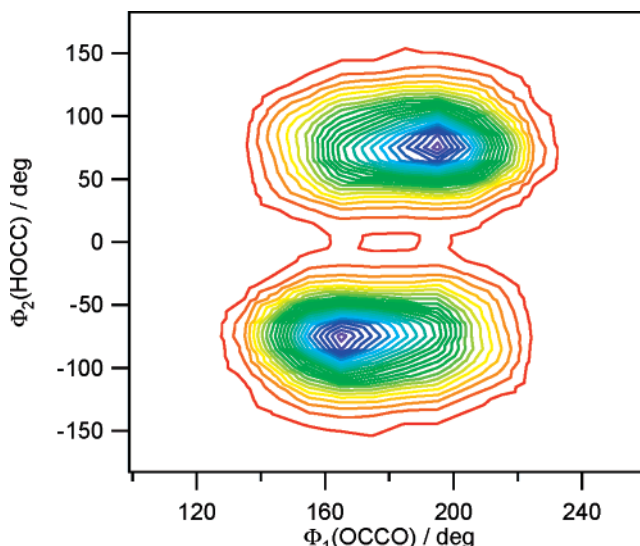
All attempts to localize the wavefunction about the minimum L3 failed. This was clearly due to the small barrier associated with the interconversion of L3 to GM. While a DMC calculation is not a dynamics simulation, we also recognize a relationship between the imaginary time taken for a population to migrate from one minimum down to another and the ability to isolate a system as an independent conformer, rather than detect it spectroscopically as a transient excited vibrational state. Based on the results of the two-dimensional and 18-dimensional calculations we propose that isomer L3 cannot be isolated in the gas phase.

Because of the slight energy differences between conformers L1 and L2 and the relatively low barrier between them it is difficult to predict which conformer may be isolated. By use of the 2- and 18-dimensional surfaces at the B3LYP/6-31G(d,p) level the wavefunction localizes around minima L2 and L2′ as indicated in Figure 4. However, using the energies of Senent,[10] in which isomer L1 is lower in energy, the wavefunction tends to localize about both L2 and L1 to some extent, although the projected wavefunction is somewhat difficult to interpret due to the crudeness of using a coarse grid for the PES and energies with derivatives of the energies obtained from different levels of theory. On the basis of our results, we conclude that only one other isomer of hydroxyacetaldehyde should be observable in the gas phase, and that isomer is most likely to be L2.

References and Notes

(1) Ehrenfreund, P.; Schutte, W. A. In *Astrochemistry: From Molecular Clouds to Planetary Systems*; Minh, Y. C., Van Dishoeck, E. F., Eds.; Sheridan Books: Chelsea, 2000; p 135.
(2) Wilfred, H. S. *Ap. J.* **2002**, *555*, L129.
(3) Bonnie, W.; Patrick, J. P.; Cherokees.L.; Hoaglund, P. H.; Wade, J. T. *Langmuir* **1996**, *12*, 2594.
(4) Galano, A.; Alvarez-Idaboy, J. R.; Ruiz-Santoy, M. E.; Vivier-Bunge, A. *J. Phys. Chem. A* **2005**, *109*, 169−180.
(5) Ochando-Pardo, M.; Nebot-Gil, I.; Gonzalez-Lafont, A.; Lluch, J. M. *J. Phys. Chem. A* **2004**, *108*, 5117.
(6) Marstokk K.-M.; Møllendal, H *J. Mol. Struct.* **1973**, *16*, 259.
(7) Marstokk K.-M.; Møllendal, H *J. Mol. Struct.* **1971**, *7*, 101.
(8) Marstokk, K.-M.; Møllendal, H. *J. Mol. Struct.* **1970**, *5*, 205.
(9) Asplala, A.; Murto, J.; Sten, P. *J. Chem. Phys.* **1986**, *106*, 339.
(10) Senent, M. L. *J. Phys. Chem. A* **2004**, *108*, 6286−6293.

(11) Liu, K.; Brown, M. G.; Carter, C.; Saykally, R. J.; Gregory, J. K.; Clary, D. C. *Nature* **1996**, *381*, 501.

(12) Clary, D. C. *J. Chem. Phys.* **2001**, *114*, 9725.

(13) Bettens, R. P. A. *J. Am. Chem. Soc.* **2003**, *125*, 584.

(14) McCoy, A. B.; Braams, B. J.; Brown, A.; Huang, X. C.; Jin, Z.; Bowman, J. M. *J. Phys. Chem. A* **2004**, *108*, 4991.

(15) Huang, X. C.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 445.

(16) Huang, X. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2005**, *122*, 44308.

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revisions A.6 and A.11.2; Gaussian, Inc.: Pittsburgh, PA, 1998.

(18) Kohn, D. W.; Robles, E. S. J.; Logan, C. F.; Chen, P. *J. Phys. Chem.* 1993, *97*, 4936.

(19) Jordan, M. J. T.; Thompson, K. C.; Collins, M. A. *J. Chem. Phys.* **1995**, *102*, 5647. Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302. Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 816.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.01; Gaussian, Inc.: Pittsburgh PA, 2003.