Drama in dynamics: Boom, splash, and speed

by

Heather Marie Netzloff

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Physical Chemistry

Program of Study Committee:
Mark S. Gordon, Major Professor
Gordon Miller
William S. Jenks
Xueyu Song
Ricky K. Kendall

Iowa State University

Ames, Iowa

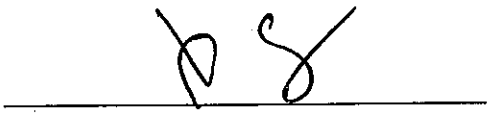
2004

Graduate College
Iowa State University

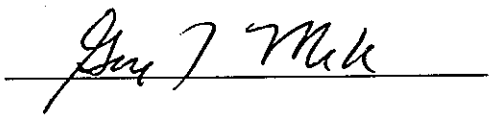
This is to certify that the doctoral dissertation of

Heather Marie Netzloff

has met the dissertation requirements of Iowa State University



Committee Member

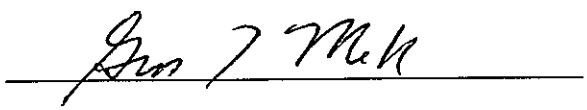


Committee Member

Committee Member

Committee Member

Major Professor



For the Major Program

But those who wait on the Lord shall renew their strength; they shall mount up with wings like eagles, they shall run and not be weary, they shall walk and not faint.
Isaiah 40:31

"For I know the plans I have for you," says the LORD. "They are plans for good and not for disaster, to give you a future and a hope."
Jeremiah 29:11

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

I. General Overview

The full nature of chemistry and physics cannot be captured by static calculations alone. Dynamics calculations allow the simulation of time-dependent phenomena. This facilitates both comparisons with experimental data and the prediction and interpretation of details not easily obtainable from experiments. Simulations thus provide a direct link between theory and experiment, between microscopic details of a system and macroscopic observed properties. Many types of dynamics calculations exist. The most important distinction between the methods and the decision of which method to use can be described in terms of the size and type of molecule/reaction under consideration and the type and level of accuracy required in the final properties of interest. These considerations must be balanced with available computational codes and resources as simulations to mimic "real-life" may require many time steps.

As indicated in the title, the theme of this thesis is dynamics. The goal is to utilize the best type of dynamics for the system under study while trying to perform dynamics in the most accurate way possible. As a quantum chemist, this involves some level of first principles calculations by default. Very accurate calculations of small molecules and molecular systems are now possible with relatively high-level *ab initio* quantum chemistry. For example, a quantum chemical potential energy surface (PES) can be developed "on-the-fly" with dynamic reaction path (DRP) methods. In this way a classical trajectory is developed without prior knowledge of the PES.

In order to treat solvation processes and the condensed phase, large numbers of molecules are required, especially in predicting bulk behavior. The Effective Fragment Potential (EFP) method for solvation decreases the cost of a fully quantum mechanical calculation by dividing a chemical system into an *ab initio* region that contains the solute and an "effective fragment" region that contains the remaining solvent molecules. But, despite the reduced cost relative to fully QM calculations, the EFP method, due to its complex, QM-based potential, does require more computation time than simple interaction potentials, especially when the method is used for large scale molecular dynamics simulations. Thus,

the EFP method was parallelized to facilitate these calculations within the quantum chemistry program GAMESS.

The EFP method provides relative energies and structures that are in excellent agreement with the analogous fully quantum results for small water clusters. The ability of the method to predict bulk water properties with a comparable accuracy is assessed by performing EFP molecular dynamics simulations. Molecular dynamics simulations can provide properties that are directly comparable with experimental results, for example radial distribution functions.

The molecular PES is a fundamental starting point for chemical reaction dynamics. Many methods can be used to obtain a PES; for example, assuming a global functional form for the PES or, as mentioned above, performing "on-the-fly" dynamics with AI or semi-empirical calculations at every molecular configuration. But as the size of the system grows, using electronic structure theory to build a PES and, therefore, study reaction dynamics becomes virtually impossible. The program Grow builds a PES as an interpolation of AI data; the goal is to attempt to produce an *accurate* PES with the *smallest* number of AI calculations. The Grow-GAMESS interface was developed to obtain the AI data from GAMESS. Classical or quantum dynamics can be performed on the resulting surface. The interface includes the novel capability to build multi-reference PESs; these types of calculations are applicable to problems ranging from atmospheric chemistry to photochemical reaction mechanisms in organic and inorganic chemistry to fundamental biological phenomena such as photosynthesis.

II. Dissertation Organization

The present work contains six chapters: chapters 2 through 5 are papers accepted, submitted to, in press with, or in preparation for submission to appropriate journals. The present author is the primary author in all cases.

Chapter 2 describes the investigation of the existence of FN_5, a possible high-energy polynitrogen species. Both experimental (done by co-authors Christie, Wilson, V. Vij, A. Vij, and Boatz) and theoretical (done by the present author) aspects of FN_5 were studied. The theoretical work focused on using high-level *ab initio* electronic structure calculations

with intrinsic reaction coordinate and dynamic reaction path calculations were used to study reaction paths and molecules, as well as to predict life times for FN_5.

Chapter 3 describes the parallelization of the EFP code within the quantum chemistry program GAMESS. As larger systems and dynamics on these systems are studied, the EFP calculation time needs to decrease. An atom decomposition scheme is used to distribute work, as well new schemes which make use of non-blocking communication.

Chapter 4 introduces methods for coupling molecular dynamics (MD) with the Effective Fragment Potential (EFP) method for solvation. The EFP method has been shown to provide excellent results for small water clusters; its ability to predict bulk water properties is tested by performing EFP molecular dynamics.

Chapter 5 details the use of interface between the Grow and GAMESS programs to facilitate the growth of multiple multi-reference potential energy surfaces. Grow builds a molecular potential energy surface as an interpolation of *ab initio* data; this data is supplied by GAMESS. Small $X + H_2$ systems (X = C, N, and O) are studied with the method.

III. Theoretical Methods

This section presents a brief overview of the theoretical methods that will be considered throughout later chapters. A distinction must be made between the two broad types of approaches used and described in this thesis: quantum mechanics and molecular mechanics. Quantum mechanics (QM) describes a molecule as a number of electrons surrounding a set of positively charged nuclei. Molecular mechanics (MM), on the other hand, views a molecule as a collection of atoms held together by bonds. Thus, when considering a chemical problem, it is important to place it in the correct context of *which* mechanics method should be used to describe the system. If treating large systems, molecular mechanics is most appropriate due to its low-level computational cost as compared with quantum mechanics. On the other hand, chemical reactions, involving the movement of electrons from one species to another, are most accurately described by quantum mechanics. Hybrid methods of splitting the system into QM and MM portions (QM to treat the part of the system most affected by a chemical reaction and MM to treat the remainder), for example the Effective Fragment Potential (EFP) method,[1,2] are currently in use. A brief description of each approach will be given below.

A. Quantum Mechanics

Since electrons display both wave and particle characteristics, they must be described in terms of a wavefunction, Ψ. The time-dependent Schrödinger[3] equation postulates that the state of the Ψ will change with time as

$$\mathrm{H(r,t)}\Psi\mathrm{(r,t)} = i\hbar\frac{\partial\Psi\mathrm{(r,t)}}{\partial t} \qquad (1)$$

where H(r,t) is the Hamiltonian operator, i is $(-1)^{1/2}$, and $\hbar = \frac{h}{2\pi}$.

If it is assumed that the Hamiltonian is time independent, this equation can be simplified to give the time-independent Schrödinger equation

$$\mathrm{H}(r)\Psi(r) = E\Psi(r) \qquad (2)$$

where E is the total system energy. The Hamiltonian is composed of both kinetic (T) and potential (V) energy for all particles

$$\mathrm{H} = T_N + T_e + V_{NN} + V_{Ne} + V_{ee} \qquad (3)$$

where T_N is the operator for the nuclear kinetic energy (KE), T_e is the operator for the electronic KE, V_{NN} represents the nuclear-nuclear potential energy (PE), V_{Ne} is the nuclear-electronic PE, and V_{ee} is the electron-electron PE.

Since nuclei are ~1800 times heavier than electrons, a good approximation is that electronic motion will instantaneously follow the nuclear motion. This allows an electronic-nuclear separation of Hamiltonian. Under the Born-Oppenheimer approximation[4], the nuclei remain fixed during the cycle of electronic motion (thus, $T_N \rightarrow 0$ and V_{NN} is a constant), and the focus can shift to the electronic Schrödinger equation

$$\mathrm{H}_{el}\psi_{el} = E_{el}\psi_{el} \qquad (4)$$

where the H_{el} is the electronic Hamiltonian

$$H_{el} = T_e + V_{Ne} + V_{ee} + V_{NN} \tag{5}$$

and ψ_{el} and E_{el} are the electronic wavefunction and energy, respectively, which depend *parametrically* on the nuclear configuration.

Even with the Born-Oppenheimer approximation, it is only possible to solve the electronic Schrödinger equation exactly for one-electron systems, such as H_2^+. In general, more approximations must be applied. The variational principle states that the energy must be minimized as a function of the parameters introduced into the approximate wavefunction. This principle, as well as antisymmetry (change in sign if any two electron coordinates are interchanged) of the total electronic wavefunction, is applied in the approximations. A Slater determinant[5] can be built which fulfills the antisymmetry requirement

$$\Psi(x_1, x_2, \ldots x_N) = \hat{A}\Psi^{HP} \tag{6}$$

where x defines electronic coordinates (N-electron system), A is the antisymmetrization operator, and Ψ^{HP} is the Hartree product wavefunction.[6] Another approximation is used to further simplify the Schrödinger equation in this description. The orbital approximation assumes that each electron moves in its own orbital (independent particles). Since electrons must be described both by position and spin, the one-electron functions are described as products of spatial orbitals and spin (α or β) functions (the product is defined as a spin orbital). The Hartree product thus re-writes a many-electron wavefunction as a product of spin orbitals. The antisymmetrization operator completes the picture in that it interchanges the coordinates of two electrons i and j.

If the trial wavefunction consists of a single Slater determinant and if the variational principle is used to minimize the energy with respect to optimization of the spin orbitals (with the constraint of orthonormality), the Hartree Fock equations[7] are defined as

$$\hat{F}_i\chi(x_i) = \varepsilon_i\chi(x_i) \tag{7}$$

where F is the Fock operator (the effective Hartree-Fock Hamiltonian), χ is the i^{th} spin orbital, and ε_i is the orbital energy of the i^{th} spin orbital. The Fock operator is given by

$$\hat{F}_i = \hat{h}_i + \sum_j^N (\hat{J}_j - \hat{K}_j) \tag{8}$$

where $\hat{h}_i$ is a one-electron operator, describing the motion of electron i in the field of all the nuclei, and $\hat{J}_j$ and $\hat{K}_j$ are two-electron operators giving the electron-electron repulsion. More specifically, $\hat{J}_j$ is the Coulomb operator, representing the average local potential at the coordinates of an electron in orbital i arising from an electron in the j^{th} orbital. $\hat{K}_j$ is the exchange operator and arises due to the antisymmetry requirement; it is a nonlocal operator. The Fock operator is an effective one-electron operator, describing the kinetic energy of an electron, attraction to all nuclei, and the repulsion due to all other electrons (through $\hat{J}$ and $\hat{K}$).

The next approximation used to solve the Fock equations is the representation of each Hartree Fock orbital (molecular orbital (MO) for molecules) as linear combinations of a complete set of K known functions, called basis functions, $(\phi_\mu(r)|\mu = 1,2,\ldots,K)$. The basis functions are conventionally called atomic orbitals (the MOs are then linear combinations of atomic orbitals (LCAO)). The wavefunction then becomes

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu \qquad i = 1,2,\ldots K \tag{9}$$

Determination of the expansion coefficients, $C_{\mu i}$, is accomplished through a self-consistent field (SCF) iterative procedure. This allows us to write the Hartree-Fock equation in matrix notation[8,9]

$$FC = SC\varepsilon \tag{10}$$

where F contains the Fock matrix elements, C contains the expansion coefficients, S contains the overlap elements between basis functions, and ε is a diagonal matrix of orbital energies.

Note that the Fock operators can only be determined if all occupied orbitals are known, therefore the SCF method must be employed.

The above discussion described the MOs used to build the trial wavefunction in general terms, but there are different types of spatial orbitals that can be used for the approximation. If no constraints are placed on the spatial functions, the trial function is called an unrestricted Hartree Fock (UHF) wavefunction. This means that different spatial functions many have different spins. If the system of interest has an even number of electrons and all electrons are paired (closed shell system), the trial function is termed a restricted Hartree Fock (RHF) wavefunction. In the restricted open-shell Hartree Fock (ROHF) method, the paired electrons are given the same spatial function (as with RHF), but no constraints are placed on the other electrons (like the open-shell, UHF method).

Since a single Slater determinant was used as the trial wavefunction, an error has been introduced in the HF equations. The resulting HF energy only accounts for electron-electron repulsion in an average fashion; the true interaction is instantaneous. The difference in energy between the Hartree Fock energy, E_{HF}, and the exact energy, E_{exact}, is the electron correlation energy, E_{corr}

$$E_{corr} = E_{HF} - E_{exact} \tag{11}$$

There are several methods that exist to correct for electron correlation. The first of these is configuration interaction (CI). In this method, wavefunctions are formed that differ from the ground state by various excitations of electrons from occupied orbitals to vacant orbitals. The MOs used for building the excited Slater determinants are taken from a HF calculation and held fixed

$$\Psi_{CI} = a_0\Phi_{HF} + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \ldots \tag{12}$$

where Ψ_{CI} is the CI wavefunction, Φ_{HF} is the HF wavefunction, and the Φ_S, Φ_D, etc indicate wavefunctions that are singly, doubly, etc. excited relative to the HF configuration. The energy is minimized under the constraint that Ψ_{CI} is normalized; this is done by optimizing the CI expansion coefficients (a_O, a_S, etc.).

The multi-configurational self-consistent field (MCSCF) method uses a wavefunction that is a truncated CI expansion. Again, the energy is minimized, but the MCSCF calculation does this by optimizing *both* the CI expansion coefficients, as well as the orbitals contained in the truncated CI expansion (HF orbitals). The excitations are also confined to take place in a well-defined active space that is composed of a given number of electrons in chemically important orbitals.

Perturbation theory is an alternative method used to capture electron correlation. Since the correlation energy can be assumed to be a small part of the HF energy, the full Hamiltonian H can be expressed as

$$H = H_{HF} + \lambda H' \tag{13}$$

where H_{HF} is the reference Hartree Fock Hamiltonian, λ is a perturbation parameter that is either 0 or 1 (turning the perturbation on or off), and H' is the perturbation. If H (Eq. 13) is used as the Hamiltonian of the system, a "perturbed" Schrödinger equation can be expressed as

$$H\Psi = W\Psi \tag{14}$$

where Ψ and W are the perturbed wavefunction and energy, respectively. Since both the perturbed wavefunction and energy must change as λ changes, they can be written as Taylor expansions in λ, and zeroth order, first order, second order, etc. corrections to the wavefunction and energy can be obtained. This scheme is known as many body perturbation theory (MBPT).[10] When the zeroth order Hamiltonian is chosen to be the sum over Fock operators, the method is known as Moller-Plesset[11] (MP) perturbation theory. The most popular choice of this theory is MP2, where corrections up to second order are included.

A third method that can be employed to account for electron correlation is coupled cluster (CC) theory. Perturbation theory adds *all types* of corrections (single, double, triple, etc.) to the reference wavefunction to a *given order*.[12] CC methods attempt to include *all corrections* of a given type to an *infinite order*. The coupled cluster wavefunction Ψ_{CC} can be written as

$$\Psi_{cc} = e^{T}\Psi_{HF} \tag{15}$$

The coupled cluster operator T is given by

$$T = T_1 + T_2 + T_3 + \ldots + T_N \tag{16}$$

The T_i operator acts on the HF wavefunction Ψ_{HF} to give all of the i-particle intereactions. Thus, if Eq. 16 is truncated at the doubles term (including singles and doubles), the method is called CCSD. IF Eq. 16 is truncated at the triples term, the method is CCSDT, etc. Several hybrid methods between CC and perturbation theory exist. One of the most popular is CCSD(T) in which the triples contribution is evaluated by perturbation theory and then added to the CCSD result.

Electron correlation can also be approximately calculated using Density Functional Theory (DFT). This method attempts to calculate the ground state energy and other properties from the electron density. The energy is expressed as a functional of the density, $E_0 = E_0[\rho_0]$. The exact functional is not known, thus many approximate (semi-empirical) functionals are currently in use.

B. Molecular Mechanics

Since electrons are not considered explicitly in molecular mechanics (molecules are considered to be a collection of balls (atoms) held together by springs (bonds)), the solution of the electronic Schrödinger equation (Eq. 4) can be bypassed, as well as the treatment of the quantum aspects of the nuclear motion. The dynamics of atoms is thus treated with classical mechanics (Newton's second law of motion). The energy is expressed as a sum of distortion energies, such as stretching, bending, torsions, non-bonded atom (van der Waals and electrostatic), and cross terms. The parameters (force constants, etc.) included in each of these terms are fit to experimental or other higher-level computational data.

Since much of this thesis describes solvation processes, non-bonded interactions are very important. Non-bonded interactions are composed of van der Waals and electrostatic interactions. The van der Waals energy describes the repulsion or attraction between atoms

that are not directly bonded. A common way to describe the van der Waals energy is with the Lennard-Jones (LJ) potential[13]

$$V^{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{17}$$

where σ and ε are LJ parameters and r describes point coordinates. This potential gives the strong, repulsive short-ranged interaction, as well as long-range van der Waals attraction.

The electrostatic energy is due to the internal distribution of electrons, creating positive and negative parts of the molecule. The basic approach to modeling this effect is by assigning charges to each atom. The Coulomb potential gives the interaction between point charges

$$V^{el}(r_{ab}) = \frac{q_a q_b}{r_{ab}} \tag{18}$$

where q_a and q_b are the charges on points a and b, respectively, and r_{ab} is the distance between those points.

Many common solvation methods (especially for water) describe the interaction potential as a sum of Coulomb interactions between all charges and a single Lennard-Jones interaction between oxygen atoms

$$V = \sum_{ab}\frac{q_a q_b}{r_{ab}} + 4\varepsilon_O\left[\left(\frac{\sigma_O}{r_{OO}}\right)^{12} - \left(\frac{\sigma_O}{r_{OO}}\right)^{6}\right] \tag{19}$$

Where q_a and q_b are again the charges on each atom, a and b, in the system, r_{ab} is the distance between points a and b, r_{OO} is the distance between two oxygen atoms, and ε_O and σ_O are parameterized for each model.

C. The Effective Fragment Potential Method

The EFP method[1,2] has been described in detail elsewhere, so it is only briefly summarized here. The method is based on first principles quantum chemistry but takes into account aspects of both quantum mechanics and molecular mechanics as described above. Thus it is a QM/MM method. The description below is for the version of EFP that is currently in the quantum chemistry program GAMESS[14] and is designed specifically for water (EFP1). A general version of the method (for any type of solvent), called EFP2,[15,16] has been developed.

The EFP method was originally designed and implemented to describe discrete solvent effects[1] with a focus on the effect of solvents on chemical reactions and, subsequently, on the study of small clusters of water molecules. Instead of performing a complete quantum calculation on the system of interest, the chemically important part (solute + some number of solvent molecules, if desired) is treated with some level of quantum mechanics, and the remaining solvent molecules are treated as interacting "fragments". Therefore, the system Hamiltonian (H_{system}) is a sum of the active region Hamiltonian (H_{AR}) and the potential due to the fragment molecules (V):

$$H_{system} = H_{AR} + V \tag{20}$$

The fragment potential V contains only one-electron integrals; thus, an EFP calculation is much less computationally demanding than a full QM calculation. There are three terms in the EFP1 model that has been designed specifically for water:

$$V_{el}(\mu,s) = \sum_{k=1}^{K} V_k^{Elec}(\mu,s) + \sum_{l=1}^{L} V_l^{Pol}(\mu,s) + \sum_{m=1}^{M} V_m^{Rep}(\mu,s) \tag{21}$$

where m represents a fragment molecule, μ the *ab initio* electronic coordinates, and k, l, and m label expansion points for the various terms described below. The first term represents Coulomb interactions between solvent molecules or between a solvent molecule and the QM solute. The second term models the solvent-solvent and solvent-solute induction/polarization interactions, and the third term represents the remaining interactions contained in the Hartree

Fock potential. The potential includes ab initio-fragment, ab initio (nuclei)-fragment, and fragment-fragment interactions.

The Coulomb term is represented by a distributed multi-polar expansion[17,18,19] through octopoles. Since this is a point charge model, screening is included to account for overlapping electron density as molecules (fragment-fragment or fragment-solute) approach each other. Each nuclear center and bond midpoint are chosen as expansion points. Thus, for water, K = 5 in Eq. 21.

The polarization term is based on a dipole-induced dipole potential. Polarizability tensors are located on the localized molecular orbital centroids. The polarization energy is iterated until self-consistency is reached, since a new induced dipole on a fragment affects the total electric field surrounding the QM part. This, in turn, modifies the induced dipole... For water, localized orbital polarizability tensors are placed at the centroids of the two O-H bonds, two oxygen lone pairs, and the inner shell oxygen orbital (L = 5 in Eq. 21).

The third, remainder (repulsion) term in Eq. (2) is represented by simple Gaussian functions for QM-fragment interactions and simple exponentials for fragment-fragment interactions. The exponents in these functions are optimized by a fitting procedure. This is accomplished by performing QM calculations on 192 water dimer structures.[2] For each point, the sum of the electrostatic and polarization energy components is subtracted from the total interaction potential, and the functional form (Gaussian or exponential) is then obtained by fitting to this remainder. Expansion points are the nuclear centers and the center of mass (for water, M = 4 in Eq. 21).

Multipoles and polarizabilities are determined from QM calculations on a single solvent molecule (these can be calculated once and stored or calculated "on-the-fly"). Because of this, the method can be systematically improved, as computational methodologies and hardware improve. Only one-electron integrals are required, making EFP much less computationally demanding than a full QM calculation.

The EFP method was originally based on the HF method, using the DH(d,p) basis set,[20] EFP1/HF.[1,2] In this case, the fitted remainder contains exchange repulsion + charge transfer interactions. More recently, the method has been extended to DFT employing the B3LYP[21,22] functional, EFP1/DFT.[23] For the DFT implementation, the remainder term also

includes some electron correlation effects. An MP2-level implementation, EFP1/MP2, including dispersion, is under development.[24]

References

1 P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, Journal of Chemical Physics **105** (5), 1968 (1996).

2 W. Chen and M. S. Gordon, Journal of Chemical Physics **105** (24), 11081 (1996).

3 E. Schrodinger, Ann. Physik. **79**, 361 (1926).

4 M. Born and R. Oppenheimer, Annalen der Physik (Berlin, Germany) **84**, 457 (1927).

5 J. C. Slater, Phys. Rev. **35**, 509 (1930).

6 D. R. Hartree, Proc. Cambridge Phil. Soc. **24** (Pt. 3), 426 (1928).

7 A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*. (Dover Publications, 1989).

8 C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

9 G. G. Hall, Proc. R. Soc. (London) **A205**, 541 (1951).

10 R. J. Bartlett and D. M. Silver, Int. J. Quant. Chem. **9**, 183 (1975).

11 C. Moller and M. S. Plesset, Phys. Rev. **46**, 618 (134).

12 F. Jensen, *Introduction to Quantum Chemistry*. (Wiley, 1999).

13 J. E. Lennard-Jones, Proc. R. Soc. London. **106** (A), 463 (1924).

14 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Journal of Computational Chemistry **14**, 1347 (1993).

15 J. H. Jensen and M. S. Gordon, Molecular Physics **89** (5), 1313 (1996).

16 J. H. Jensen and M. S. Gordon, Journal of Chemical Physics **108** (12), 4772 (1998).

17 A. D. Buckingham, Quarterly Rev. (London) **8**, 183 (1959).

18 A. J. Stone and M. Alderton, Molecular Physics **56** (5), 1047 (1985).

19 A. J. Stone, *The Theory of Intermolecular Forces*. (Oxford University Press, Oxford, 1996).

20 T. H. Dunning and P. J. Hay, in *Methods of Electronic Structure Theory*, edited by H. F. Shaefer (Plenum, New York, 1977), pp. 1.

[21] A. D. Becke, Physical Review A: Atomic, Molecular, and Optical Physics **38** (6), 3098 (1988).

[22] C. Lee, W. Yang, and R. G. Parr, Physical Review B: Condensed Matter and Materials Physics **37** (2), 785 (1988).

[23] I. Adamovic, M. A. Freitag, and M. S. Gordon, Journal of Chemical Physics **118** (15), 6725 (2003).

[24] J. Song and M. S. Gordon, in preparation.

CHAPTER 2: ON THE EXISTENCE OF FN_5, A THEORETICAL AND EXPERIMENTAL STUDY

Taken from a paper that has been published in the Journal of Physical Chemistry A.
Reprinted with permission from the *Journal of Physical Chemistry A* **2003**, *107*, 6638.

Heather M. Netzloff, Mark S. Gordon, Karl Christe, William W. Wilson, Ashwani Vij, Vandana Vij, and Jerry A. Boatz

Abstract

The possible existence of FN_5 was studied by *ab initio* electronic structure theory. Calculations were carried out at the MP2/6-31+G(d) and CCSD(T)/aug-cc-pVDZ levels of theory for the $N_5^+AsF_6^-$ ion pair and its decomposition to FN_5 and AsF_5. Six different vibrationally stable isomers of FN_5 were identified. Intrinsic reaction coordinate (IRC) and dynamic reaction path (DRP) calculations were used to study the isomerization of FN_5 and its decomposition to FN_3 and N_2. A Rice-Ramsperger-Kassel-Marcus (RRKM) analysis was performed, indicating upper limits to the lifetimes of the FN_5 isomers in the nanosecond range. These theoretical predictions were confirmed by an experimental study of the thermolyses of N_5AsF_6 and $[N_5]_2SnF_6$ and the displacement of FN_5 from N_5SbF_6 with CsF, using FT-IR spectroscopy. In accord with the theoretical predictions, the primary reaction product FN_5 could not be observed, but its decomposition products FN_3, F_2N_2, and NF_3 were identified.

I. Introduction

Polynitrogen compounds are of great interest as high energy density materials (HEDM).[1-5] Although theoretical studies have predicted numerous kinetically stable polynitrogen compounds,[4] almost all attempts to synthesize them have failed due to their very high endothermicities, low energy barriers towards decomposition, and a lack of suitable synthetic methods. The high energy content of polynitrogen compounds arises from an unusual property of nitrogen that sets it apart from most other chemical elements. Its

single and double bond energies are considerably less than one-third and two-thirds, respectively, of its triple bond energy. Therefore, the decomposition of polynitrogen species to N_2 is accompanied by a large release of energy.[1]

Recently, the N_5^+ cation has been synthesized and characterized.[1,2] It represents only the second known homonuclear polynitrogen species after N_3^-,[6] that is stable and can be prepared on a macroscopic scale. Its bent structure of C_{2V} symmetry (Figure 1) was established by a crystal structure determination of $N_5^+Sb_2F_{11}^-$ and vibrational and NMR spectroscopy, and is in accord with *ab initio* and density functional theory (DFT) calculations.[1,2,7] The bent structure avoids the unfavorable neighboring positive charges that would result from a linear structure.[1]

Most salts consisting of an X^+ cation and a complex fluoro anion, MF_6^-, are prepared by the transfer of an F^- anion from the parent FX molecule to the strong Lewis acid MF_5 (eq. 1).

$$FX + MF_5 \Leftrightarrow X^+MF_6^- \qquad (1)$$

Usually this reaction is reversible and the FX molecule can be regenerated by either thermolysis of $X^+MF_6^-$ or a displacement reaction between $X^+MF_6^-$ and a stronger Lewis base, such as CsF (eq. 2).

$$X^+MF_6^- + CsF \rightarrow FX + CsMF_6 \qquad (2)$$

Only a few $X^+MF_6^-$ salts are known for which FX cannot be generated in this manner. Typical examples are the NF_4^+,[8] ClF_6^+,[9] and BrF_6^+[10] salts where the corresponding FX parent compounds cannot exist because the maximum coordination number of the central atom would be exceeded.[11] The reverse case, where an amphoteric FX molecule exists but the corresponding X^+ cation does not, is also known but rare. A typical example is FN_3 that does not form a stable N_3^+ salt with strong Lewis acids.[12]

The availability of several marginally stable N_5^+ salts, such as $N_5^+AsF_6^-$,[1] $N_5^+SbF_6^-$,[2] and $[N_5]^+_2[SnF_6]^{2-}$ [13] that can be readily subjected to thermolysis or displacement reactions, offered an ideal opportunity to probe the possible existence of the unknown FN_5 molecule.

While carrying out the theoretical study, we have also investigated the potential energy surfaces of the $N_5^+AsF_6^-$ ion pair and its FN_5 decomposition product in order to better understand the likely structure and stability of this new polynitrogen species.

II. Experimental Section

Caution! *Reactions of N_5^+ salts can be violent and can result in explosions,[1,2] particularly when highly shock sensitive FN_3[12,14] is formed as a decomposition product. Therefore, these materials should be handled only on a small scale with appropriate safety precautions (face shield, leather gloves, and protective clothing).*

Materials and Apparatus. All reactions were carried out in a demountable Teflon-PFA condensing side arm of a 5 cm path length, Teflon-FEP infrared cell equipped with AgCl windows. Nonvolatile solids were loaded in the dry nitrogen atmosphere of a glove box into the side arm of the IR cell. The cell was then evacuated and placed into the FT-IR spectrometer. The decomposition or displacement reactions were initiated by gentle warming, and the volatile decomposition products were continuously monitored by infrared spectroscopy using a Mattson Galaxy FT-IR spectrometer. Volatile materials were handled on a stainless steel/Teflon-FEP vacuum line.[15]

The $N_5^+AsF_6^-$,[1] $N_5^+SbF_6^-$,[2] and $[N_5]^+_2[SnF_6]^{2-}$ [13] starting materials were prepared by literature methods. The CsF (KBI) was fused in a platinum crucible, transferred while hot into the dry box, and finely powdered.

III. Computational Methods

Initial optimizations of all structures were performed using second order perturbation theory (MP2)[16] and the 6-31+G(d) basis set.[17] Hessians (energy second derivatives) were calculated for the final equilibrium structures to determine if they are minima (positive definite hessian) or transition states (one negative eigenvalue). At the final MP2/6-31+G(d) geometries, improved relative energies were obtained using singles and doubles coupled cluster theory with triples included perturbatively (CCSD(T))[18] and the aug-cc-pVDZ basis set.[19] The MP2 calculations were performed using the electronic structure code GAMESS,[20] while the CCSD(T) calculations were carried out using ACES II.[21]

Intrinsic reaction coordinate (IRC) pathways[22] were employed in the study of the FN_5

species in order to connect isomer minima, transition states, and decomposition products. The IRC method is the minimum energy path (MEP) in mass weighted Cartesian coordinates. IRC calculations were performed with GAMESS using the second-order method developed by Gonzalez and Schlegel[23] with a step size of 0.1 $(amu)^{1/2}$-bohr.

A simple Rice-Ramsperger-Kassel-Marcus (RRKM) analysis was performed on the isomers included in the potential energy surfaces obtained with the IRC calculations. The RRKM theory of reaction dynamics can be used to give an upper limit to the lifetime of the minima.[24] It assumes a microcanonical equilibrium and a locally separable reaction coordinate. The microscopic rate constant is proportional to the sum of states of the i^{th} reaction channel ($W_N^i (E - E_0^i)$) divided by the reaction density of states (ρ_N):

$$k \propto W_N^i (E - E_0^i) / \rho_N (E) \tag{3}$$

$$k = \frac{\prod_{i=1}^{3N-6} \nu_i}{\prod_{i=1}^{3N-7} \nu_i^{TS}} \left(\frac{E - E_0^{barrier}}{E} \right)^{3N-7} \tag{4}$$

where k = rate constant

$\nu_i = i^{th}$ frequency of the minima

$\nu_i^{TS} = i^{th}$ frequency of the transition state

E = applied energy

$E_0^{barrier}$ = barrier energy

Note that the lifetime $\tau = 1 / k$.

Finally, in order to further study the decomposition and stability of several FN_5 isomers, the dynamic reaction path (DRP) method was used to add photons (kinetic energy (KE)) to one or more FN_5 vibrational normal modes.[25] The DRP method, a classical trajectory approach, is based on a quantum chemical potential energy surface (PES) that need not be known ahead of time. Unlike the IRC, energy is strictly conserved along the dynamic reaction path. Thus, larger step sizes can be used. In this way a classical trajectory is

developed "on-the-fly" without prior knowledge of the PES. GAMESS can use normal modes as the initial dynamic reaction coordinates. An initial KE and velocity direction is supplied to one or more modes (in units of quanta).[26] The strategy is to provide energy to those modes that appear to lead to desired reaction products. Because there is often significant mode-mode mixing, the applied energy is usually in excess of the reaction/decomposition barrier. This is a direct dynamics method in that the *ab initio* (in the present case, MP2) gradients/forces are calculated at each step and are then used to solve Newton's equations of motion and propagate the system. Step sizes ranged from 0.1 to 0.2 fs, depending on how well energy conservation criteria were satisfied.

IV. Results and Discussion

The $N_5^+AsF_6^-$ Ion Pair. The starting points for the calculations were the AsF_6^- and N_5^+ ions separated by a distance of ~10.0 Å, in C_1 symmetry, followed by a complete geometry optimization. The resulting ion pair is shown in Figure 2a. At the MP2/6-31+G(d) level of theory, the ion pair is 98.8 kcal/mol lower in energy than the separated ions (Table 1).

MP2 Mulliken charges on the N_5^+ and AsF_6^- units within the ion pair show that the charge separation is ±0.825; thus, there is relatively little charge transfer between the ions. In comparison with the separated ions (Figures 2b and 3b), the N_5^+ unit in the ion pair has a less negative charge on the central N by approximately 0.3, while the terminal N atoms are about 0.2 less positive than in the isolated N_5^+ ion. In the case of AsF_6^- in the ion pair, all As-F bonds except for the bond opposite to the N_5^+ unit have been elongated by about 0.085 Å (Table 2). This effect can be explained by the partial removal of an electron from bonding orbitals. It is interesting to note that the bond distances in N_5^+ do not exhibit a similar effect and are essentially unchanged.

In isolated N_5^+, the central N has a large negative charge (Figure 3b). This charge is reduced by a factor of two in the ion pair. All other N atoms are positive, with the terminal N atoms being most positive. This type of charge distribution can be rationalized by the valence bond structures[1,2,27] given in Figure 1. It must be kept in mind, however, that the magnitude of the charges varies strongly with the calculation method, although their signs and relative order remain the same. Thus, at the NBO (B3LYP/aug-cc-pVDZ) level of theory, the charges

on N1, N2, and N3 are significantly smaller and amount to 0.33, 0.22, and –0.11, respectively.[28]

As can be seen from Figure 2a, the closest N-F distances in the ion pair are N2-F8 = 2.26 Å and N4-F8 = 2.34 Å. They are much longer than a typical N-F bond (~1.3 Å)[29], but significantly shorter than the sum of their Van der Waals radii (3.0 Å)[30] and the shortest N···F contacts observed in the crystal structure of $N_5^+Sb_2F_{11}^-$ (2.72 and 2.78 Å).[2]

Since there is still a large charge separation between the components of the ion pair, it is of interest to consider how much energy would be required to transfer F^- from AsF_6^- to N_5^+ to make gaseous AsF_5 and FN_5. At the MP2/6-31+G(d) level of theory, the ion pair is 46.5 kcal/mol lower in energy than the separated AsF_5 and *bifurcated(bif)*-FN_5 molecules, indicating that it takes at least this much energy to transfer F^- from AsF_6^- to N_5^+ to make *bif*-FN_5 and AsF_5 (Figure 4). Thus, the ion pair is reasonably stable to dissociation to gaseous AsF_5 and *bif*-FN_5. Because an ion pair is only a poor approximation to a crystalline solid, the 25.5 kcal/mol difference between the lattice energy of $N_5^+AsF_6^-$ (124 ± 4 kcal/mol)[31] and the ion pair energy (98.8 kcal/mol) must be added to the above minimum decomposition energy barrier of 46.5 kcal/mol when considering the thermal stability of solid N_5AsF_6. The resulting minimum decomposition energy barrier of 72.0 kcal/mol for crystalline N_5AsF_6 is in accord with the experimental observation that this salt is marginally stable at room temperature.[1] The experimentally observed irreversible decomposition of N_5AsF_6 at higher temperatures[1] is due to the subsequent rapid, highly exothermic, and irreversible decompositions of FN_5 and FN_3 (see below).

In addition to the $[AsF_6]^-[N_5]^+$ structure discussed above, a geometry search revealed a second, lower energy isomer. This C_{2v} structure is obtained from the structure show in Figure 2a by rotating the $[AsF_6]^-$ anion so that the As_6 ,F_7 ,and F_8 atoms are coplanar with $[N_5]^+$. The C_2 rotation axis then passes through atoms N_3 and As_6. The second symmetry plane contains N_3, As_6, and F_9-F_{12} with F_9 and F_{11} pointing towards the $[N_5]^+$ cation and F_{10} and F_{12} pointed away from it. Some of the distinguishing features of this C_{2v} geometry are the As-F distances (As-$F_{7,8}$ = 1.75 Å; As-$F_{9,11}$ = 1.81 Å; As-$F_{10,12}$ = 1.72 Å); the N_1-F_8 (N_5-F_7) distance is 2.80 Å. The F_9-N_2, F_9-N_4, F_{11}-N_2, and F_{11}-N_4 distances are each 2.47 Å. The geometry of the $[N_5]^+$ cation is virtually identical to that in Figure 2a. This second bifurcated isomer is ~10 kcal/mol lower in energy than the structure shown in Figure 2a, most likely

because of the electrostatic interactions between four F atoms with N. However, the higher energy structure is more likely to yield the neutral species $AsF_5 + FN_5$. In either case, the ion complex is clearly much lower in energy than the separated neutrals.

FN_5 Isomers. Because the F^- ion can attach itself in different ways to the three different nitrogen atoms of N_5^+, it is necessary to explore the different possible isomers of FN_5 and their relative energies and stabilities. An exhaustive study was performed and six stable isomers were found. Their relevant energies and geometries are listed in Tables 3 and 4, respectively. Optimized structures were obtained for each isomer and hessians confirmed that these structures are minima on the potential energy surface. The relevant geometrical features are summarized in Figure 5 and Table 4. The N-F and N-N bond lengths are similar for all isomers except for the bifurcated isomer (1) and the cyclic structure (6). All isomers are planar molecules. Both MP2 and CCSD(T) calculations predict the *cis-* and *cyclic-* isomers to be among the lowest energy species. Indeed, except for the bifurcated isomer, the two levels of theory predict similar relative energies (Table 3). Note that the MP2 results are not very basis set dependent. The *cis-*, *trans-*, *wag-*, and *harp-* isomers are all within a few kcal/mol of each other and can be easily interconverted by simple rotations around the N3-N4 and N4-N5 bonds. The more accurate and reliable CCSD(T) method places the bifurcated isomer 3-11 kcal/mol above the other isomers. Since the two N-F distances in this "bifurcated" isomer differ by 0.342 Å, this isomer might reasonably be described as an N_β-bonded isomer with a weak secondary interaction to the other N_β atom, while the other four non-cyclic isomers are all N_α-bonded. The formation of the cyclic isomer (6) is less likely to form in the experiments because it would probably involve a fluoride attacking N_5^+ from the backside at its central N_γ atom. Such an approach is not likely since both reacting atoms apparently carry partial negative charges that should repel each other. Furthermore, *cyclic-* FN_5 has been predicted to have a low barrier to decomposition of 6.7 kcal/mol at the CCSD(T) level of theory.[32]

Because N_2 is so stable, the dissociation of FN_5 to FN_3 and N_2 is 45.5 kcal/mol exothermic at the CCSD(T) level of theory (Table 3) and is irreversible. Therefore, it is

important to explore the possible decomposition pathways for the preferred FN_5 isomers and to determine their energy barriers towards decomposition.

In order to explore both the isomerization and decomposition pathways, IRC studies were performed at the MP2 level of theory to connect minima with the corresponding transition states (TS). Transition states are illustrated in Figure 6, and FN_3 is depicted in Figure 7; detailed geometries are given in Table 5. The first IRC was started from the "bifurcated" isomer, which most closely resembles the original ion pair. Structurally, the isomer that most naturally connects with the *bifurcated*-isomer is the *harp*-isomer, since this simply requires F to move to a neighboring nitrogen atom of the N_5 moiety via TS14. Starting from *harp*, two PESs have been identified (Figure 8) that the *harp*-isomer can take to reach decomposition products: a "direct" decomposition route (through TS4p) or an isomerization via the *cis*-isomer (through TS42). The *cis*-isomer decomposes directly to FN_3 $+N_2$ (through TS2p). As shown in Figure 8 and Table 6, the bottleneck for both reaction paths is the initial barrier towards isomerization from *bifurcated* to *harp* (through TS 14), predicted to be 8.9 kcal/mol. The subsequent activation energies are much smaller, so once the *harp*-isomer is reached, the decomposition to FN_3 $+N_2$ should proceed even more easily.

The two PESs are qualitatively similar at the MP2/6-31+G(d) and CCSD(T)/aug-cc-pVDZ levels of theory. The MP2 activation energies are larger than those from CCSD(T), and all transition states and minima, excluding the initial bifurcated species, are higher in energy than the reference *cis*-isomer. CCSD(T) predicts that all minima and TSs are higher in energy than the reference *cis*-isomer (Table 3).

The results of a *qualitative* RRKM analysis (using the vibrational frequencies in Table 7) for the two isomerization reactions and two decomposition reactions are shown in Tables 8 and 9. Several values were chosen for E, the applied energy, in equation (4) for each minimum. Since the CCSD(T) and MP2 activation energies are different (Table 6), a given value of E corresponds to different excess energies above the barrier. So, in Tables 8 and 9, a given amount of energy above a barrier corresponds to different values of E. Since the CCSD(T) energies were obtained as single energy calculations at the MP2 geometries, the frequencies used in both sets of RRKM calculations are those obtained from MP2. Thus, the CCSD(T) results are only qualitative, but these frequencies are not expected to be very different. Since the CCSD(T) barriers are much lower in energy than those based on MP2,

the corresponding lifetimes will be smaller. Of course, the shorter the predicted lifetime, the more likely it is that the isomerization or dissociation will occur. At the CCSD(T) level of theory, the predicted lifetimes for the two isomerization reactions, *bifurcated* (1) → *harp* (4) and *harp* (4) → *cis* (2), are on the order of a few nanoseconds even in the case where an energy that is only 5 kcal/mol above the barrier is added. When this excess energy rises to 40 kcal/mol, these lifetimes decrease to 1-2 picoseconds. The lifetimes for the decomposition reactions, from *harp* and *cis* to FN_3 and N_2, are even shorter.

DRP Calculations. In order to further study the decomposition and stability of the *harp*- and *cis*-FN_5 isomers, the dynamic reaction path (DRP) method[25,26] was used to provide kinetic energy to one or more vibrational normal modes.

For the *cis*-isomer, mode 10 (Figure 9a) is a good candidate for breaking the N3-N4 bond since ν_{10} is the symmetric N3-N4-N5 stretching mode. At the MP2/6-31+G(d) level, the barrier for the decomposition, *cis*-FN_5 → FN_3 + N_2, is approximately 22 kcal/mol. With the addition of approximately 35 kcal/mol to this mode, the N3-N4 bond undergoes only oscillations, and at least 51 kcal/mol of KE must be applied to this mode in order to break the N3-N4 bond (the bond breaks after 5.6 fs into such a trajectory run). Figures 10 and 11 show various structures along the trajectories with the addition of 35 and 51 kcal/mol of KE, respectively. Further illustration of this phenomenon is shown in Figure 12 where the coordinate changes in terms of the normal modes of the equilibrium structure for both E = 35 and E = 51 kcal/mol are plotted. With the addition of 35 kcal/mol, the KE is quickly redistributed to other modes. This dissipates the applied energy, so bond oscillations due to the activation of other modes result. For E = 51 kcal/mol, the energy remains localized in mode 10; the modes become well separated and do not undergo wild oscillations. There is enough energy available to "push" the N_2 fragment away so it is not recaptured during oscillations

As noted earlier, the *harp*-isomer can decompose in a direct or stepwise manner. The MP2/6-31+G(d) barrier to direct decomposition is 7.5 kcal/mol, while the barrier for the isomerization, *harp* → *cis*, is 5.4 kcal/mol. In the analysis of this structure, ν_{12} (Figure 9b) appears to be a good candidate for N3-N4 bond breaking since ν_{12} is the N4-N5 stretching mode. After the addition of 44 kcal/mol, the N3-N4 bond breaks, but it then reforms and oscillates. As in the case for the *cis*-isomer, this can be seen in the graphs of normal mode

coordinate change versus time. Figure 13 shows that the energy applied to ν_{12} dissipates to other modes even more rapidly than in the case of the *cis*-isomer. Of course, these DRP calculations are only suggestive about possible dynamic processes. One would need to explore many more trajectories to obtain quantitatively meaningful kinetics data.

The MP2 and CCSD(T) RRKM analyses for the addition of energy, corresponding to this DRP for both the *cis*- and *harp*-isomers, are given in Table 9. The examples discussed above are in boldface in Table 9. Consider the decomposition of the *cis*-isomer, using ν_{10}. At least seven (four) vibrational quanta are required to surmount the barrier in this mode for MP2 (CCSD(T)), corresponding to lifetimes of 0.2 and 855 milliseconds, respectively. For an E value of 35 kcal/mol, corresponding to 10 vibrational quanta, the lifetimes are 216 and 0.9 ps, respectively. The trajectory with E = 35 kcal/mol was only run for ~ 0.15 ps, and only bond oscillations were observed. For E = 52 kcal/mol (15 quanta), the predicted lifetimes 116 fs and 1.7 ps for CCSD(T) and MP2, respectively. The E = 52 kcal/mol trajectory shows complete bond dissociation after ~ 6 fs (0.006 ps). For the *harp*-isomer, the addition of E = 44 kcal/mol to mode 12 (7 quanta) shows no dissociation after ~ 600 fs (0.6 ps). Upper limit lifetimes for this isomer, depending on whether one is considering the isomerization or decomposition channel, are ~ 3 ps and 0.3 ps, respectively, at the MP2 level and 1.9 ps and 50.4 fs, respectively, at the CCSD(T) level. It is important to reiterate here the following point: the reason that the amount of energy required to cause dissociation in the DRP trajectories exceeds the calculated barrier heights is that the energy provided to a specific vibrational mode does not localize in the bond that breaks in the reaction.

Experimental Results. The vacuum thermolyses of $N_5^+AsF_6^-$ (eq. 5) and $[N_5]^+_2[SnF_6]^{2-}$ (eq. 6), and the displacement reaction between CsF and $N_5^+SbF_6^-$ (eq. 7) were studied experimentally by fast, *in situ* FT-IR spectroscopy of the gaseous reaction products.

$$N_5^+AsF_6^- \rightarrow [FN_5] + AsF_5 \qquad (5)$$

$$(N_5)^+_2SnF_6^{2-} \rightarrow [FN_5] + N_5^+SnF_5^- \qquad (6)$$

$$N_5^+SbF_6^- + CsF \rightarrow [FN_5] + CsSbF_6 \qquad (7)$$

In excellent agreement with the above theoretical lifetime predictions for FN_5, this molecule could not be observed directly; however its expected decomposition products (eq. 8-11) were observed.

$$[FN_5] \rightarrow FN_3 + N_2 \quad (8)$$

$$FN_3 \rightarrow [FN] + N_2 \quad (9)$$

$$2\,[FN] \rightarrow F_2N_2 \quad (10)$$

$$3\,F_2N_2 \rightarrow 2\,NF_3 + 2\,N_2 \quad (11)$$

Because reactions 8-11 are all strongly exothermic, the thermodynamically most stable products, NF_3 and N_2,[33] were obtained as the major final products. However, the formation of smaller amounts of FN_3 and F_2N_2[34] was also observed, thus confirming the above sequence of reactions.

V. Conclusions

In contrast to FN_3 that can be isolated at room temperature but does not form stable N_3^+ salts with strong Lewis acids,[12] N_5^+ salts are stable but their FN_5 parent molecule has very limited kinetic stability. Because of the general difficulty of correctly modeling infinite crystal lattices,[35] the ion pair $N_5^+AsF_6^-$ and its dissociation products were studied. It is shown that at the MP2/6-31+G(d) level the ion pair and crystalline $N_5^+AsF_6^-$ are more stable than the free FN_5 and AsF_5 molecules by 46.5 and 72.0 kcal/mol, respectively. These values represent minimum dissociation energy barriers; the latter value is in accord with the observed marginal stability of crystalline $N_5^+AsF_6^-$ at room temperature. The FN_5 molecule can exist as at least six different, vibrationally stable isomers (one cyclic,[32] one bifurcated, and four chain-like structures) that differ in energy by only 6.4, 9.0, and 11.2 kcal/mol at the MP2/6-31+G(d), MP2/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ levels, respectively. Decomposition studies of FN_5 reveal two distinct pathways to $FN_3 + N_2$. The energy change from the initial bifurcated FN_5 isomer to $FN_3 + N_2$ is predicted to be -56.0 kcal/mol at the CCSD(T) level of theory.

IRC, DRP, and RRKM calculations suggest that the energy barriers for these decomposition pathways are very low resulting in predicted life times for FN_5 at room

temperature on the order of nanoseconds. These predictions of an extremely short lived FN_5 are confirmed by our experimental FT-IR studies on the thermolyses of $N_5^+AsF_6^-$ and $[N_5]^+_2[SnF_6]^{2-}$, and the displacement reaction between CsF and $N_5^+SbF_6^-$ which showed only the expected decomposition products FN_3, F_2N_2, and NF_3, but no evidence for the intermediate FN_5.

Acknowledgements

This work was supported by a grant (to MSG) from the Air Force Office of Scientific Research (AFOSR) and by a Grand Challenge grant of computer time from the High Performance Computing Modernization Program (to MSG and JAB) at the High Energy Maui High Performance Computation Center. The authors also thank Dr. Galina Chaban and Prof. William Hase for insightful discussions about the DRP calculations. HMN was supported by a Department of Energy Computational Science Graduate Fellowship and a Miller Fellowship from the Iowa State University Department of Chemistry. The work at the Air Force Research Laboratory was supported by the Defense Advanced Research Projects Agency (DARPA) and AFOSR, while that at USC was supported by DARPA, AFOSR, and the National Science Foundation. The authors thank Prof. Jenkins for the lattice energy estimate for N_5AsF_6 and Drs. R. Corley, M. Berman, and D. Woodbury for their steady support. The calculations described here were performed on both local Iowa State University/DOE Ames Laboratory workstations, as well as computers at the Maui High Performance Computation Center.

References

(1) Christe, K.O.; Wilson, W.W.; Sheehy, J.A.; Boatz, J.A. *Angew. Chem. Int. Ed.* **1999**, *38*, 2004.

(2) Vij, A.; Wilson, W.W.; Vij, V.; Tham, F.S.; Sheehy, J.A.; Christe, K.O. *J. Am. Chem. Soc.* **2001**, *123*, 6308.

(3) Lauderdale, W.J.; Stanton, J.F.; Bartlett, R.J. *J.Phys. Chem.* **1992**, *96*, 1173.

(4) Glukhovtsev, M.N.; Jiao, H.; Schleyer, P.v.R. *Inorg. Chem.* **1996**, *35*, 7124.

(5) Bartlett, R. J. *Chem. Ind.-London* **2000**, *4*, 140.

(6) Curtius, T. *Ber. Dtsch. Chem. Ges.* **1890**, *23*, 3023.

(7) Pyyko, P.; Runeberg, N. *J. Mol. Struct. (Theochem.)* **1991**, *234*, 279.

(8) (a) Christe, K. O.; Guertin, J. P.; Pavlath, A. E. *Inorg. Nucl. Chem. Lett.* **1966**, *2*, 83; (b) Guertin, J. P.; Christe, K. O.; Pavlath, A. E. *Inorg. Chem.* **1966**, *5*, 1921; (c) Tolberg, W. E.; Rewick, R. T.; Stringham, R. S. *Inorg. Chem.* **1967**, *6*, 1156.

(9) (a) Christe, K. O. *Inorg. Nucl. Chem. Lett.* **1972**, *8*, 741; (b) Robero, F. Q. *Inorg. Nucl. Chem. Lett.* **1972**, *8*, 737; (c) Christe, K. O. *Inorg. Chem.* **1973**, *12*, 1580.

(10) (a) Gillespie, R. J.; Schrobilgen, G. J. *J. Chem. Soc., Chem. Commun.* **1974**, 90; (b) Gillespie, R. J.; Schrobilgen, G. J. *Inorg. Chem.* **1974**, 13, 1230; Christe, K. O. ; Wilson, R. D. *Inorg. Chem.* **1975**, *14*, 694.

(11) (a) Christe, K. O.; Wilson, W. W.; Schrobilgen, G. J.; Chiracal, R. V. *Inorg. Chem.* **1988**, *27*, 789; (b) Christe, K. O.; Wilson, W. W. *J. Am. Chem. Soc.* **1992**, *114*, 9934.

(12) Schatte, G.; Willner, H. *Z. Naturforsch.* **1991**, *46b*, 483.

(13) Wilson, W.W.; Vij, A.; Vij, V.; Bernhardt, E.; Christe, K.O. *Chem. Eur. J.* **2003**, *9*, 2840.

(14) Gholivand, K.; Schatte, G.; Willner, H. *Z. Inorg. Chem.* **1987**, *26*, 2137.

(15) Christe, K. O.; Wilson, W. W.; Schack, C. J.; Wilson, R. D. *Inorg. Synth.* **1986**, *24*, 39.

(16) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934,** *46*, 618.

(17) Hehre, W. J.;Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972,** *56*, 2237.

(18) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**,*157*, 479.

(19) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(20) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L. *J.Comput. Chem.* **1993**, *14*, 1347.

(21) ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Kucharski, S. A.; Gwaltney, S. R.; Beck, S.; Balková, A.; Bernholdt, D. E.; Baeck, K. K.; Rozyczko, P.; Sekino, H.; Hober, C.; Bartlett, R. J. Integral packages included are VMOL (Almlöf, J.; Taylor, P. R.); VPROPS (Taylor, P. R.) ABACUS; (Helgaker, T.; Jensen, H. J. Aa.;

Jørgensen, P.; Olsen, J.; Taylor, P. R.).

(22) Garrett, B.C.; Redmon, M.; Steckler, R.; Truhlar, D.G.; Baldridge, K. K.; Bartol, D.; Schmidt, M.W.; Gordon, M.S. *J. Phys. Chem.* **1988,** *92*, 1476.

(23) Gonzales, C.; Schlegel, H. B. *J.Chem.Phys.* **1989,** *90,* 2154.

(24) (a) Robinson, P. J.; Holbrook, K. A. Unimolecular Reactions; Wiley-Interscience: New York, 1972; (b) Forst, W. Theory of Unimolecular Reactions; Academic Press: New York, 1973; (c) Schranz, H. W.; Nordholm, S.; Freasier, B. C. *Chem. Phys.* **1986,** *108,* 69.

(25) (a) Stewart, J. J. P.; Davis, L. P.; Burggraf, L. W. *J.Comput.Chem.* **1987**, *8,* 1117. (b) Maluendes, S. A.; Dupuis, M. *J. Chem. Phys.* **1990**, *93*, 5902.

(26) (a) Taketsugu, T.; Gordon, M. S. *J. Phys. Chem.* **1995,** *99,* 8462. (b) Gordon, M. S.; Chaban, G.; Taketsugu, T. *J.Phys.Chem.* **1996,** *100,* 11512.

(27) Harcourt, R. D.; Klapoetke, T. M. *Z. Naturforsch.* **2002,** *57b,* 983.

(28) Fau, S.; Bartlett, R. J. *J. Phys. Chem. A* **2001,** *105*, 4096.

(29) (a) Christe, K. O.; Lind, M. D.; Thorup, N.; Russell, D. R.; Fawcett. J. *Inorg. Chem.* **1988**, *27*, 2450; (b) Vij, A.; Zhang, X.; Christe, K. O. *Inorg. Chem.* **2001**, *40*, 416, and references cited therein.

(30) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(31) Jenkins, H. D. B. private communication.

(32) (a) Hammerl, A. ; Klapoetke, T. M. ; Schwerdtfeger, P. private communication; (b) Inagaki, S.; Goto, N. *J. Am. Chem. Soc.* **1987**, *109*, 3234.

(33) Gmelin Handbook, F Suppl. Vol. 4, pg.385, 1986.

(34) (a) Haller, J., F. *Ph.D. Thesis*, Cornell University, Ithaca, N.Y., Sept. 1942; (b) Bohn, R.K., Bauer, S.H. *Inorg. Chem.* **1967**, *6*, 309; (c) Kuczkowski, R.; Wilson, E.B. *J. Chem. Phys*. **1963**. *39*, 1030; (d) King, S-T.; Overend, J. *Spectrochim. Acta* **1966**, *22*, 689.

(35) Christe, K. O.; Zhang, X.; Sheehy, J. A.; Bau, R. *J. Am. Chem. Soc.* **2001**, *123*, 6338.

Table 1. Energies (hartrees), zero point energy (ZPE) corrections, and relative energies (kcal/mol) for ion pair, separated ions, and separated neutrals (reference = ion pair).

Molecule	E(MP2/6-31+G(d))	ZPE(kcal/mol)	Relative Energy[a]
$N_5^+AsF_6^-$	-3102.913474	23.3	0
N_5^+	-272.602102	12.4	--
AsF_6^-	-2830.155490	9.9	--
$N_5^+ + AsF_6^-$	-3102.757592	22.3	98.8
FN_5 (bifurcated)	-372.481754	14.0	--
AsF_5	-2730.358788	8.6	--
$FN_5 + AsF_5$	-3102.840542	22.6	46.5

[a]Includes ZPE correction.

Table 2. Bond distances and angles for ion complex and separated ions (refer to Figures 2a and 3a).

Ion Complex

	Distances (Å)		Angles (deg)
N1-N2	1.144	N2-N3-N4	113.56
N2-N3	1.305	F10-As6-F8	86.88
N3-N4	1.312		
N4-N5	1.144		
As6-F7	1.722		
As6-F8	1.843		
As6-F9	1.843		
As6-F10	1.843		
As6-F11	1.843		
As6-F12	1.843		
N1-F8	2.720		
N2-F8	2.258		
N3-F8	2.737		
N4-F8	2.338		
N5-F8	2.879		

N_5^+

	Distances (Å)		Angles (deg)
N1-N2	1.143	N2-N3-N4	110.17
N2-N3	1.313	N1-N2-N3	167.15
N3-N4	1.313		
N4-N5	1.143		

AsF_6^-

	Distances (Å)		Angles (deg)
As-F	1.758	F-As-F	90.0

Table 4. Distances, angles, and dihedral angles for FN_5 isomers-MP2/6-31+G(d) (refer to Figure 5).

Distances (Å)

Isomer	N1-N2	N2-N3	N3-N4	N4-N5	F-N1	F-N2	F-N3	F-N4
bifurcated	1.166	1.339	1.308	1.147	--	1.837	2.459	2.179
cis	1.239	1.396	1.279	1.152	1.452			
trans	1.251	1.404	1.275	1.152	1.416			
harp	1.233	1.362	1.293	1.157	1.495			
wag	1.253	1.393	1.285	1.156	1.412			
cyclic	1.318	1.337	1.351	1.337	1.345			

Angles (deg)

Isomer	F-N1-N2	N1-N2-N3	N2-N3-N4	N3-N4-N5
bifurcated	--	137.78	109.38	160.37
cis	112.16	117.42	107.54	169.77
trans	107.39	107.33	170.27	104.98
harp	129.23	119.27	162.54	112.42
wag	106.25	112.88	114.85	169.22
cyclic	121.29	101.59	109.70	109.70

Dihedral angles (deg)

Isomer	F-N1-N2-N3	N1-N2-N3-N4	N2-N3-N4-N5
bifurcated	180	180	180
cis	0	180	180
trans	180	180	180
harp	0	0	180
wag	180	0	180
cyclic	180	0	0

Table 5. Distances, angles, and dihedral angles for FN_5 transition states-MP2/6-31+G(d) (refer to Figure 6).

	Distances(Å)							
Molecule	N1-N2	N2-N3	N3-N4	N4-N5	F-N1	F-N2	F-N3	F-N4
TS14	1.144	1.315	1.356	1.16	2.022	2.351	2.716	2.049
TS2p	1.248	1.27	1.675	1.138	1.489			
TS42	1.230	1.448	1.264	1.155	1.495			
TS4p	1.242	1.264	1.563	1.142	1.509			

	Angles (deg)			
Molecule	F-N1-N2	N1-N2-N3	N2-N3-N4	N3-N4-N5
TS14	91.62	150.36	107.04	145.03
TS2p	107.44	125.92	106.07	150.15
TS42	113.49	118.76	109.55	171.99
TS4p	108.68	138.9	113.77	152.47

	Dihedral angles (deg)		
Molecule	F-N1-N2-N3	N1-N2-N3-N4	N2-N3-N4-N5
TS14	0	0	180
TS2p	0	180	180
TS42	7.33	-92.57	-172.06
TS4p	0	0	180

Table 6. Activation energies (kcal/mol) for isomerization and decomposition reactions.

Barriers	ΔE(MP2/G)[a,d]	ΔE(CCSD(T))[b,d]	ΔE(MP2/aug)[c,d]
TS14	18.8	8.9	18.1
TS42	5.4	3.7	5.5
TS2p	21.8	13.2	21.6
TS4p	7.5	1.0	6.9

[a]Basis set = 6-31+G(d). [b]Basis set = aug-cc-pVDZ. [c]Basis set = aug-cc-pVDZ.

[d]Includes ZPE correction.

Table 7. Vibrational frequencies and corresponding IR intensities for FN_5 isomers and transition states in the isomerization/decomposition PES surfaces (MP2/6-31+G(d)).

Molecule	Frequencies, cm^{-1} (IR intensities, $debye^2/amu\text{-}ang^2$)
bifurcated (1)	130(0.032), 232(0.415), 316(0.479), 418(0.048), 460(0.084), 541(1.693), 565(0.308), 669(0.984), 1007(2.234), 1212(2.616), 1965(5.003), 2269(2.992)
cis (2)	113(0), 178(0.044), 300(0.190), 485(0.064), 554(0.001), 638(0.122), 691(3.290), 822(2.306), 999(1.784), 1196(3.807), 1532(1.153), 2321(9.337)
trans (3)	122(0.001), 164(0.044), 340(0.075), 399(0.226), 460(0.047), 521(0.185), 716(0.321), 947(4.070), 1050(0.308), 1220(4.531), 1481(0.069), 2356(8.660)
harp (4)	49(0.016), 221(0.006), 315(0.533), 452(0.091), 518(1.043), 546(0.001), 721(1.974), 860(0.785), 916(0.790), 1139(2.967), 1546(2.382), 2175(6.153)
wag (5)	146(0.005), 148(0.010), 361(0.023), 395(0.396), 472(0.092), 517(0.012), 766(2.388), 988(0.448), 1024(2.399), 1123(4.000), 1443(0.102), 2252(5.988)
cyclic (6)	272(0.052), 435(0.045), 614(0.002), 701(0.143), 727(0), 1022(0.018), 1034(0.198), 1090(0.088), 1132(0.024), 1211(0.175), 1378(0.134), 1387(1.832)
TS14	493 i, 115, 307, 391, 423, 456, 516, 701, 898, 1120, 1901, 2311
TS4p	823 i, 115, 159, 328, 335, 442, 591, 699, 807, 1206, 1550, 1993
TS42	109 i, 204, 310, 446, 500, 611, 724, 806, 864, 1230, 1538, 2409
TS2p	728 i, 86, 125, 242, 265, 444, 581, 706, 823, 1111, 1505, 2046

Table 8. RRKM analysis for isomerization and decomposition reactions as a function of excess energy.

Bifurcated—TS14—harp

Excess Energy (kcal/mol)	E[CCSD(T)] [b,c] (kcal/mol)	E[MP2/G] [a,c] (kcal/mol)	CCSD(T) lifetime	MP2/G lifetime
5	13.9	23.8	5.5 ns	1.9 ms
10	18.9	28.8	78.3 ps	7.6 ns
20	28.9	38.8	4.0 ps	100.9 ps
30	38.9	48.8	1.2 ps	14.6 ps
40	48.9	58.8	0.6 ps	4.8 ps

Harp—TS42—cis

Excess Energy (kcal/mol)	E[CCSD(T)] [b,c] (kcal/mol)	E[MP2/G] [a,c] (kcal/mol)	CCSD(T) lifetime	MP2/G lifetime
5	8.7	10.4	0.3 ns	2.1 ns
10	13.7	15.4	23.1 ps	78.7 ps
20	23.7	25.4	4.6 ps	9.5 ps
30	33.7	35.4	2.5 ps	4.3 ps
40	43.7	45.4	1.9 ps	2.8 ps

Harp—TS4p—FN_3+N_2

Excess Energy (kcal/mol)	E[CCSD(T)] [b,c] (kcal/mol)	E[MP2/G] [a,c] (kcal/mol)	CCSD(T) lifetime	MP2/G lifetime
5	6.0	12.5	0.3 ps	0.9 ns
10	11.0	17.5	0.1 ps	18.3 ps
20	21.0	27.5	67.0 fs	1.3 ps
30	31.0	37.5	56.0 fs	451.4 fs
40	41.0	47.5	51.2 fs	256.8 fs

Cis—TS2p—FN_3+N_2

Excess Energy (kcal/mol)	E[CCSD(T)] [b,c] (kcal/mol)	E[MP2/G] [a,c] (kcal/mol)	CCSD(T) lifetime	MP2/G lifetime
5	18.2	26.8	7.2 ns	480.2 ns
10	23.2	31.8	50.1 ps	1.5 ns
20	33.2	41.8	1.3 ps	15.3 ps
30	43.2	51.8	0.3 ps	1.9 ps
40	53.2	61.8	0.1 ps	0.6 ps

[a]Basis set = 6-31+G(d). [b]Basis set = aug-cc-pVDZ. [c]Includes ZPE correction.

Table 9. RRKM analysis for isomerization and decomposition reactions as a function of vibrational quanta.

Harp—TS42—cis
Mode 12 (harp)

Applied energy (quanta)	E (kcal/mol)	CCSD(T)[b] lifetime	MP2/G[a] lifetime
1	6.21	18.2 ns	2.5 ms
2	12.42	36.2 ps	353.0 ps
3	18.63	8.2 ps	29.3 ps
4	24.84	4.2 ps	10.1 ps
5	31.05	2.9 ps	5.6 ps
6	37.26	2.2 ps	3.9 ps
7	**43.47**	**1.9 ps**	**3.0 ps**
8	49.68	1.6 ps	2.4 ps
9	55.89	1.5 ps	2.1 ps
10	62.1	1.4 ps	1.9 ps
11	68.31	1.3 ps	1.7 ps
12	74.52	1.2 ps	1.6 ps
13	80.73	1.2 ps	1.5 ps
14	86.94	1.1 ps	1.4 ps
15	93.15	1.1 ps	1.3 ps

Harp—TS4p—FN_3+N_2
Mode 12 (harp)

Applied energy (quanta)	E (kcal/mol)	CCSD(T)[b] lifetime	MP2/G[a] lifetime
1	6.21	0.3 ps	--
2	12.42	99.5 fs	1.0 ns
3	18.63	72.0 fs	11.2 ps
4	24.84	61.5 fs	2.0 ps
5	31.05	56.0 fs	811.5 fs
6	37.26	52.6 fs	459.4 fs
7	**43.47**	**50.4 fs**	**311.4 fs**
8	49.68	48.7 fs	234.6 fs
9	55.89	47.5 fs	189.2 fs
10	62.1	46.5 fs	159.8 fs
11	68.31	45.8 fs	139.4 fs
12	74.52	45.1 fs	124.6 fs
13	80.73	44.6 fs	113.3 fs
14	86.94	44.2 fs	104.6 fs
15	93.15	43.8 fs	97.6 fs

Table 9. (continued)

Cis—TS2p—FN_3+N_2
Mode 10 (cis)

Applied energy (quanta)	E (kcal/mol)	CCSD(T)[b] lifetime	MP2/G[a] lifetime
1	3.49	---	---
2	6.98	---	---
3	10.47	---	---
4	13.96	855.1 ms	---
5	17.45	30.0 ns	---
6	20.94	0.3 ns	---
7	24.43	25.4 ps	0.2 ms
8	27.92	5.5 ps	77.1 ns
9	31.41	1.9 ps	2.0 ns
10	**34.9**	**891.3 fs**	**215.8 ps**
11	38.39	491.2 fs	46.2 ps
12	41.88	306.0 fs	14.8 ps
13	45.37	208.2 fs	6.1 ps
14	48.86	151.3 fs	3.0 ps
15	**52.35**	**115.5 fs**	**1.7 ps**

[a]Basis set = 6-31+G(d). [b]Basis set = aug-cc-pVDZ.

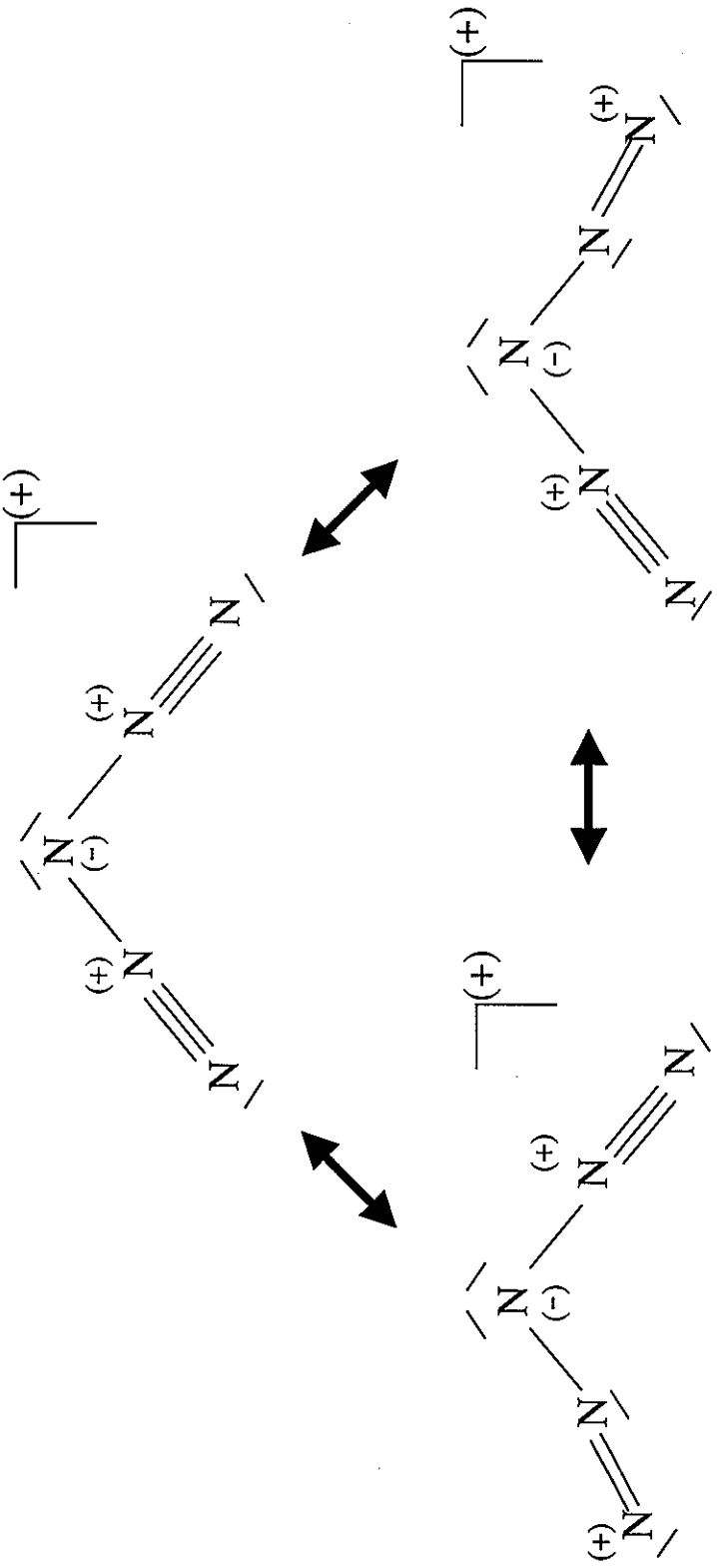
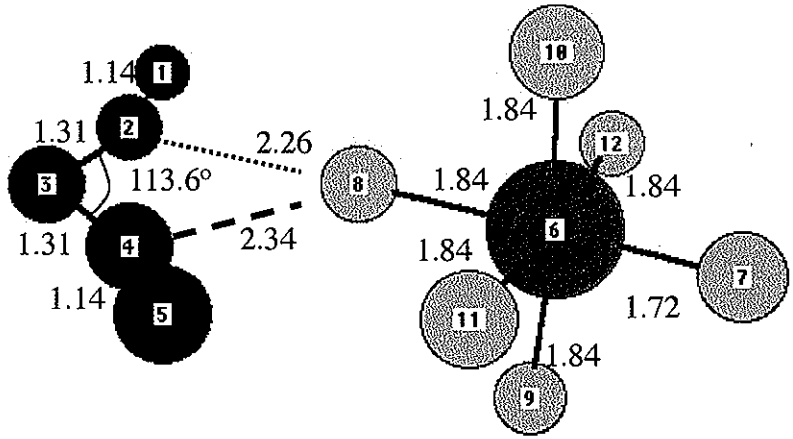
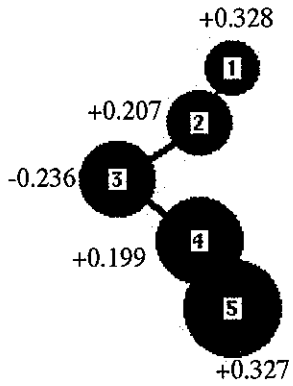
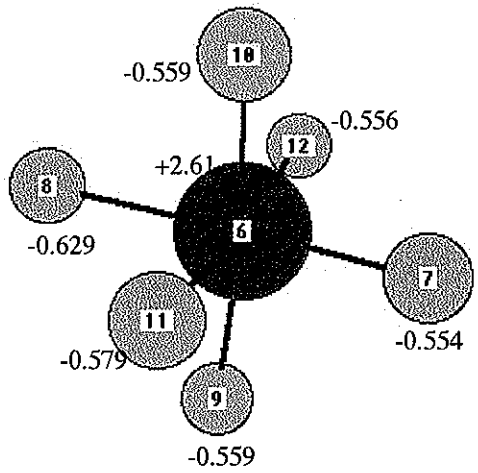


Figure 1. N_5^+ resonance structures.



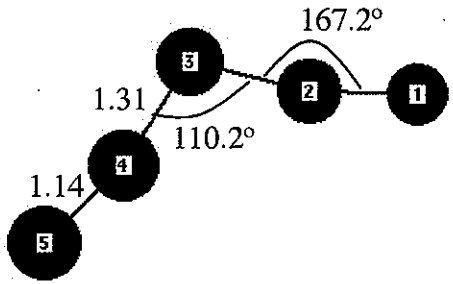
2a.

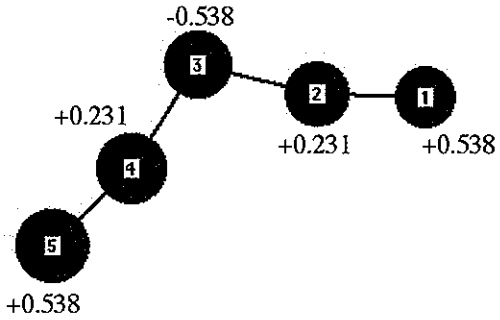


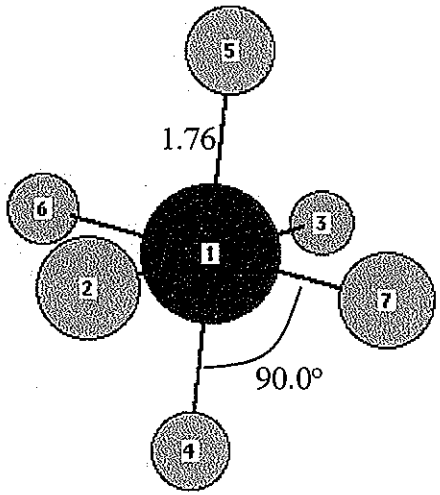


2b.

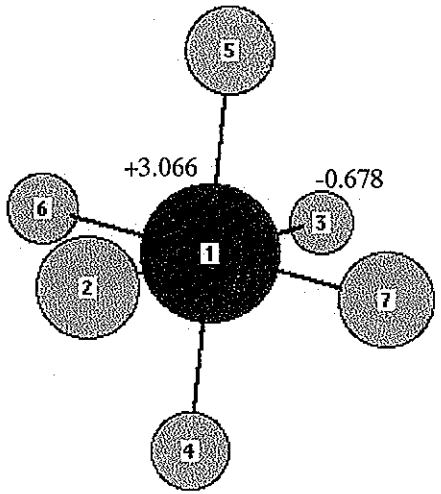
Figure 2. Optimized ion pair, $[N_5]^+[AsF_6]^-$: (a) Bond distances (Å) and angles; (b) Mulliken charges.







3a.



3b.

Figure 3. Separated Ions, N_5^+ and AsF_6^-: (a) Bond distances (Å) and angles; (b) Mulliken charges.

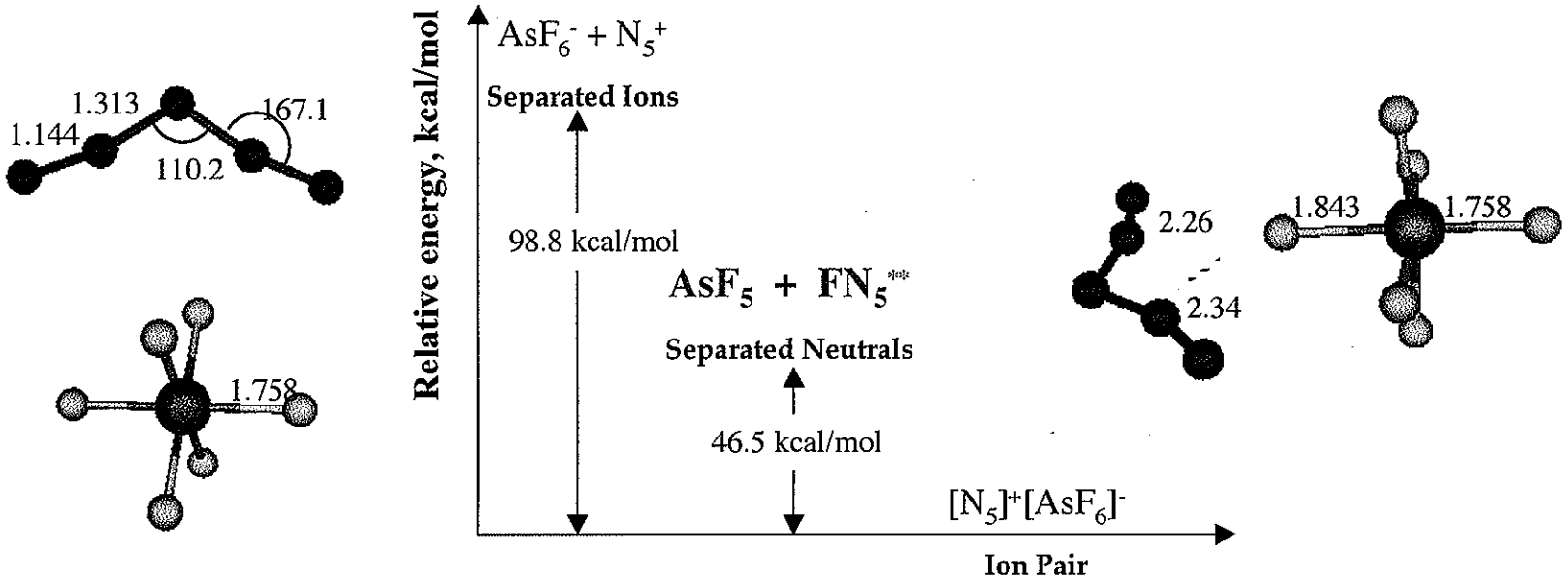


Figure 4. Relative energy diagram (kcal/mol): ion pair, separated ions, and separated neutrals (reference = ion pair). Energies: MP2/6-31+G(d). FN_5 = bifurcated FN_5 isomer.

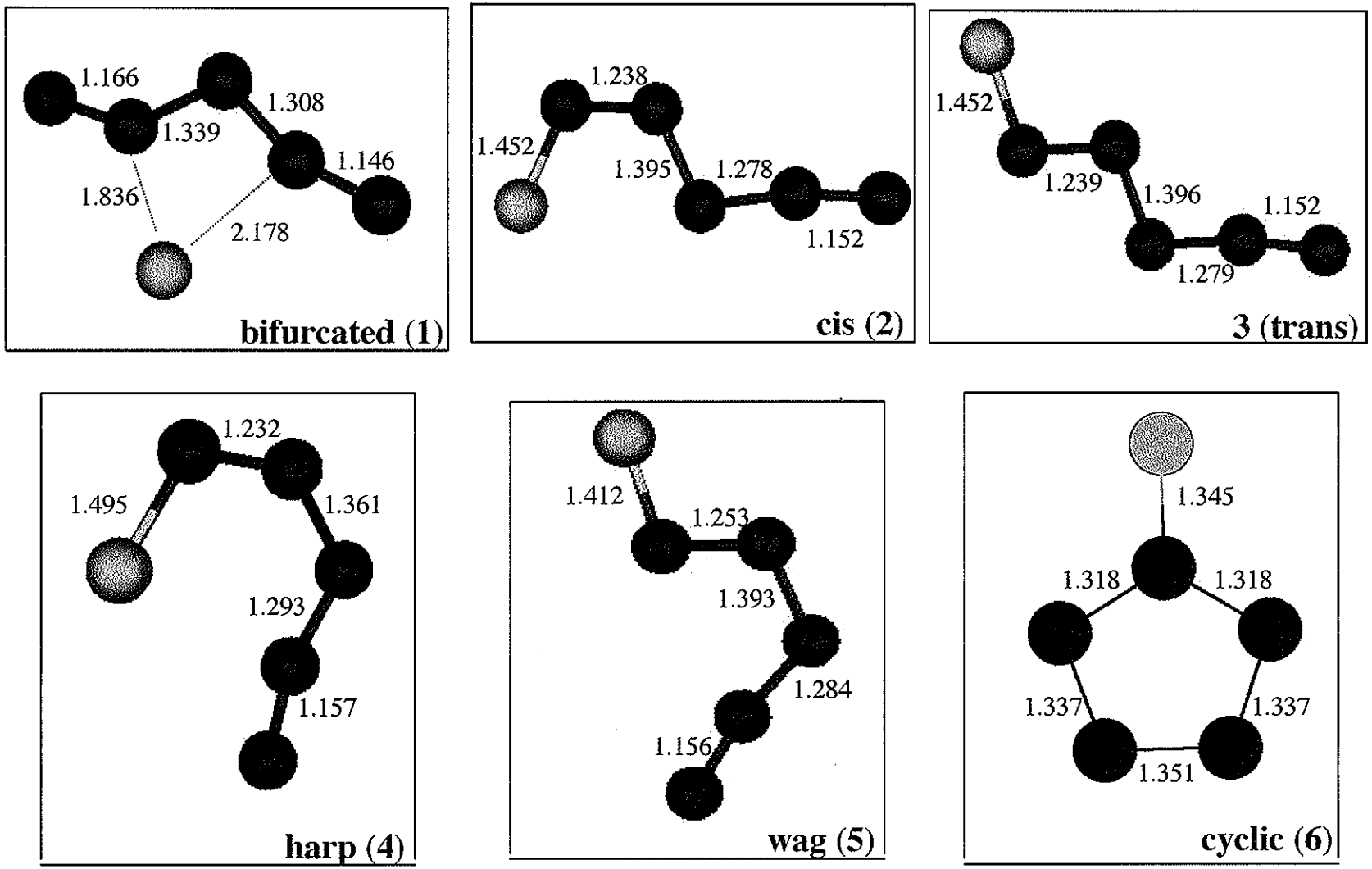


Figure 5. FN_5 isomers (MP2/6-31+G(d) geometries). Bond distances in Å.

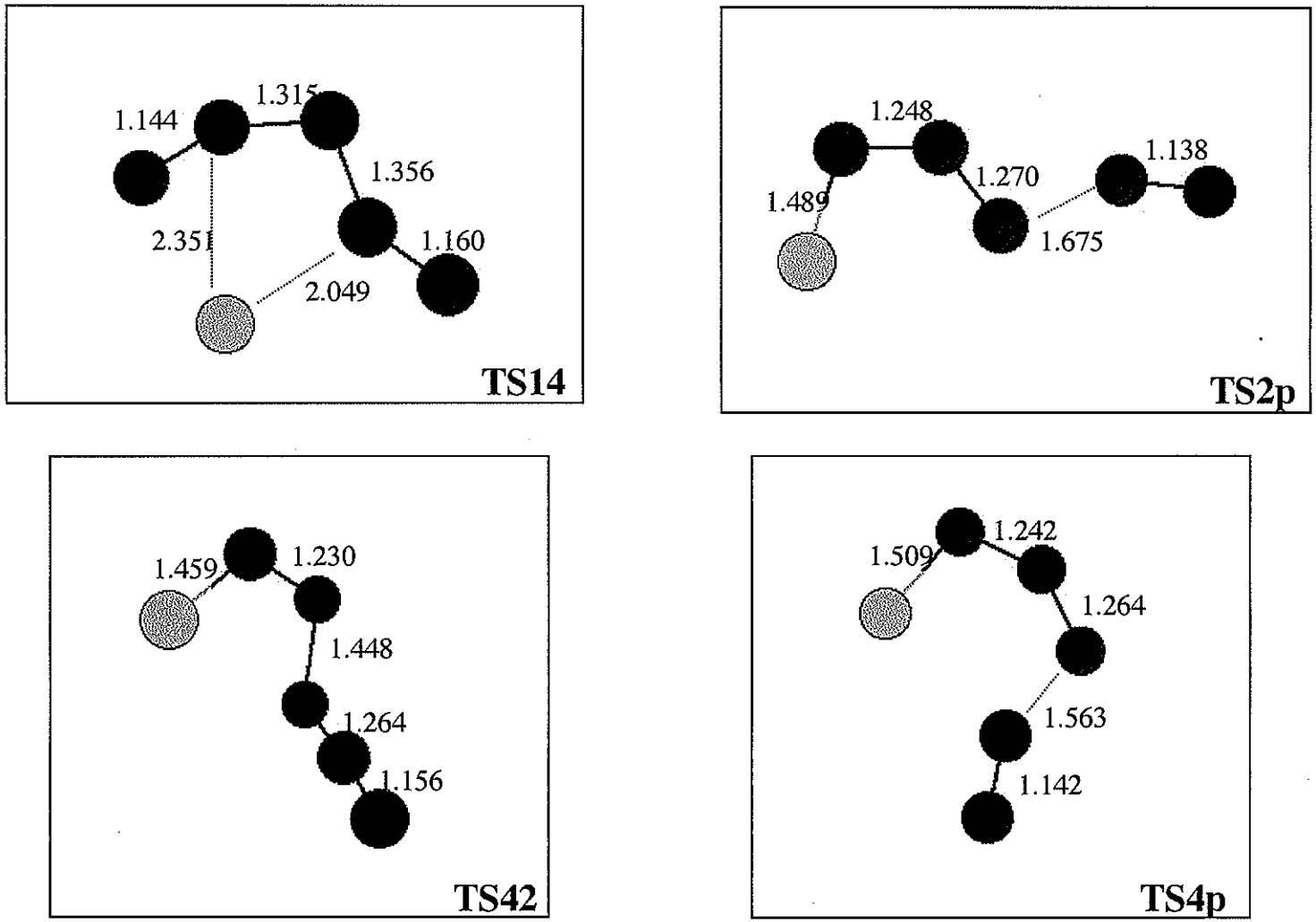


Figure 6. FN_5 transition states (MP2/6-31+G(d) geometries). Bond distances in Å.

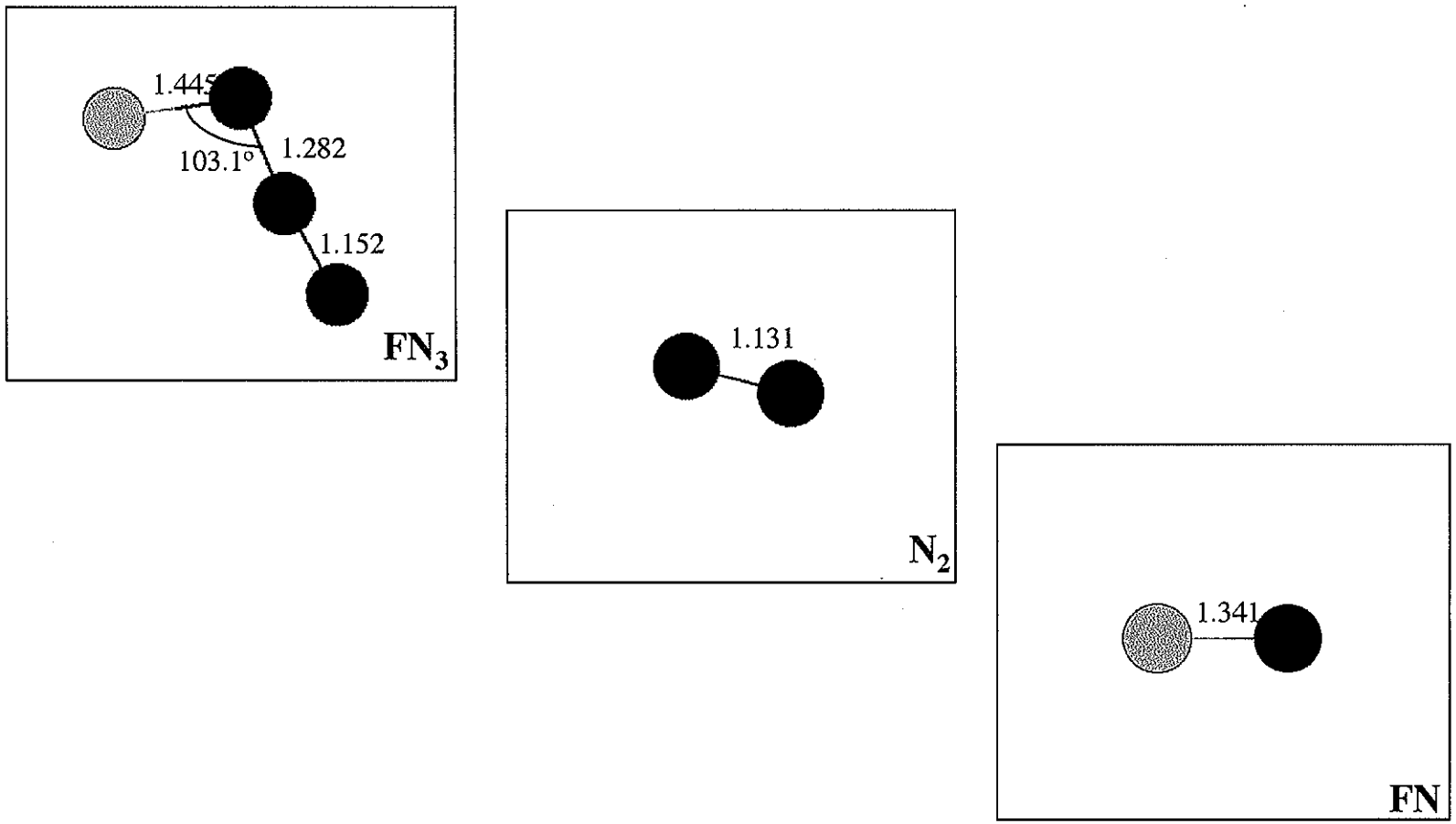


Figure 7. Optimized FN_3, N_2, and FN (MP2/6-31+G(d) geometries). Bond distances in Å.

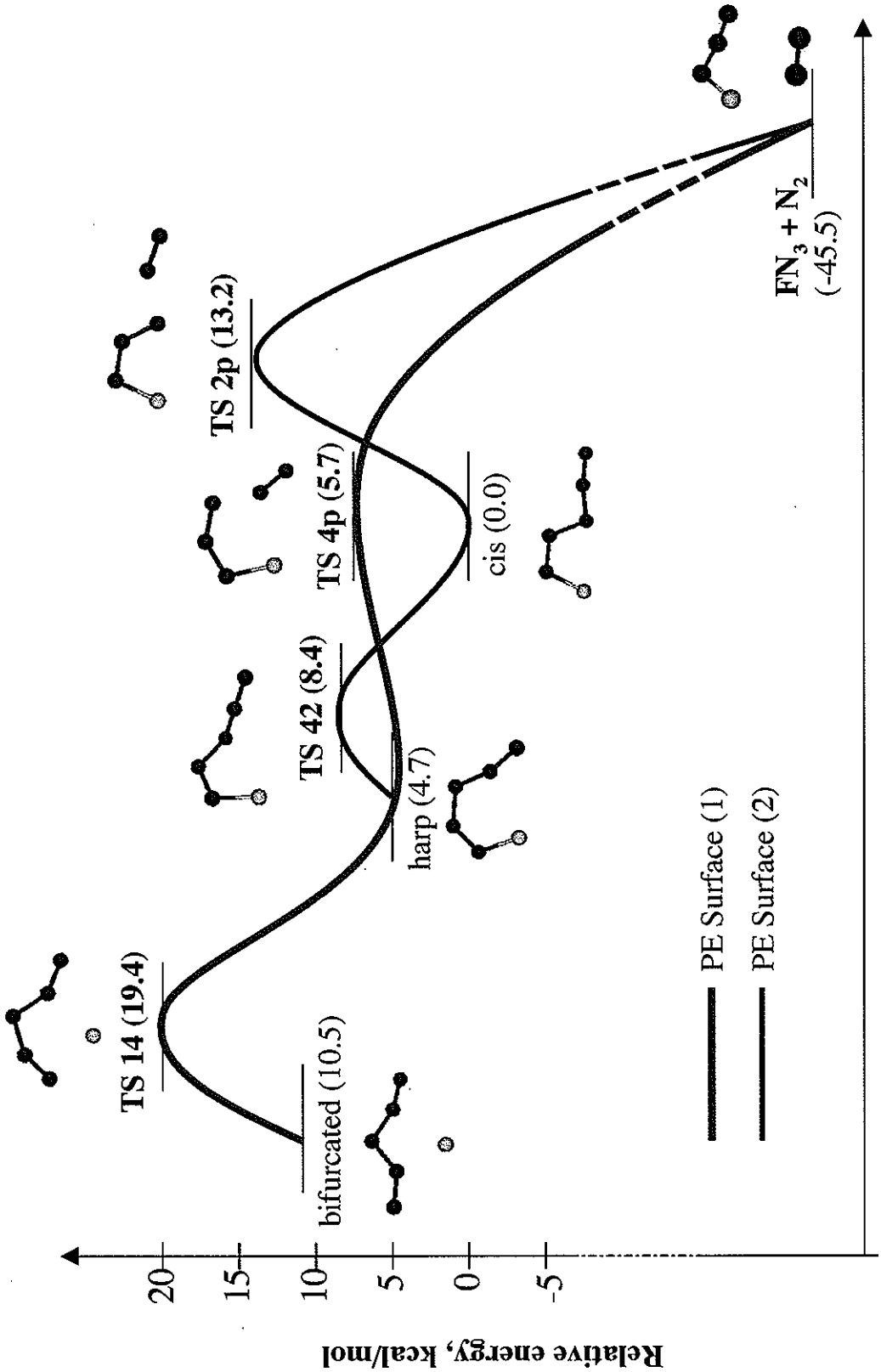
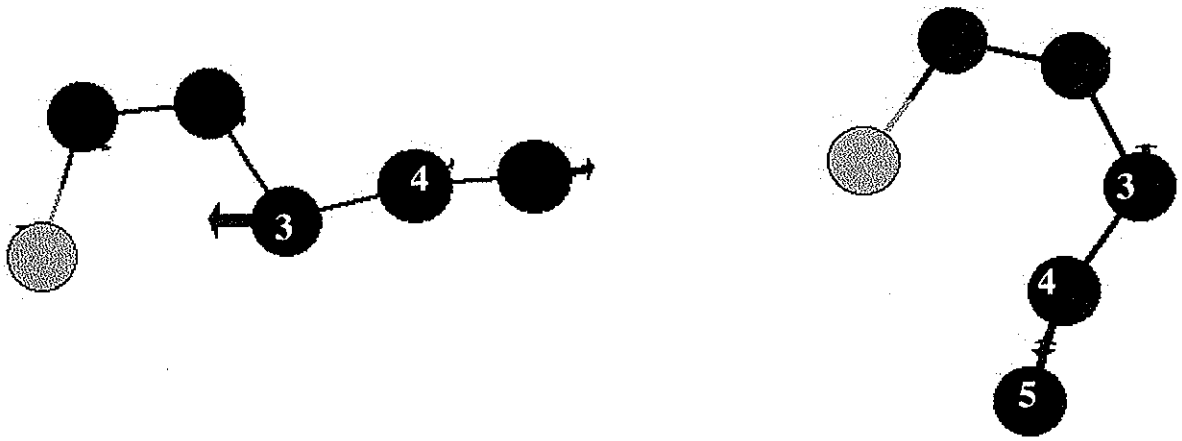


Figure 8. Relative isomerization/decomposition PESs with FN_3 isomers (reference = cis isomer). Energies: CCSD(T)/aug-cc-pVDZ with ZPE.



9a. 9b.

Figure 9. FN_5 vibrational modes selected for DRP analysis: (a) cis: mode 10; (b) harp : mode 12. The molecular structure was distorted along the negative of the vibrational displacements; these displacements are shown in the picture.

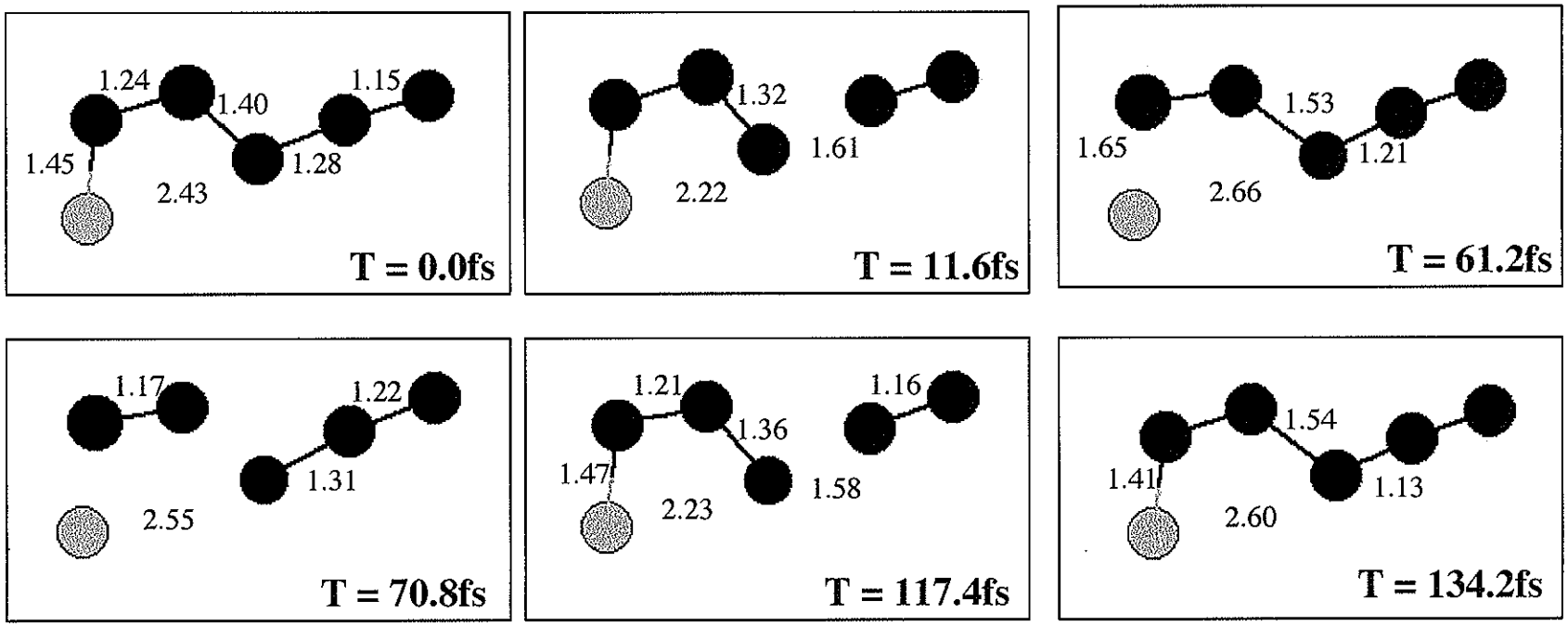


Figure 10. Stages along the trajectory for adding an initial KE of 34.9 kcal/mol to mode 10, cis isomer. Distances in Å(initial bond lengths shown at T = 0.0fs; if bond lengths have changed by > 0.40 Å, new values are reported in next pane).

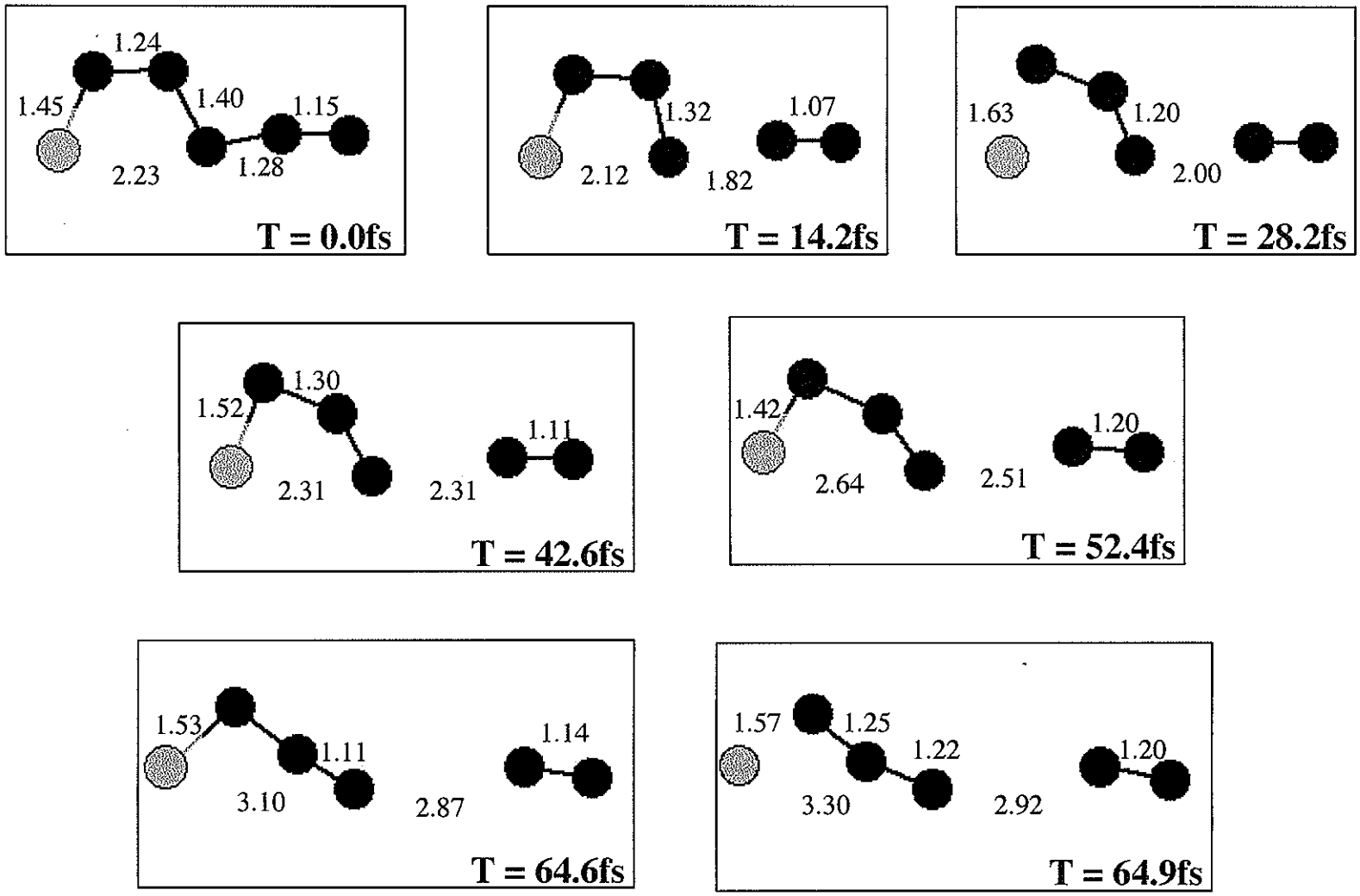


Figure 11. Stages along the trajectory for adding an initial KE of 51.3 kcal/mol to mode 10, cis isomer.

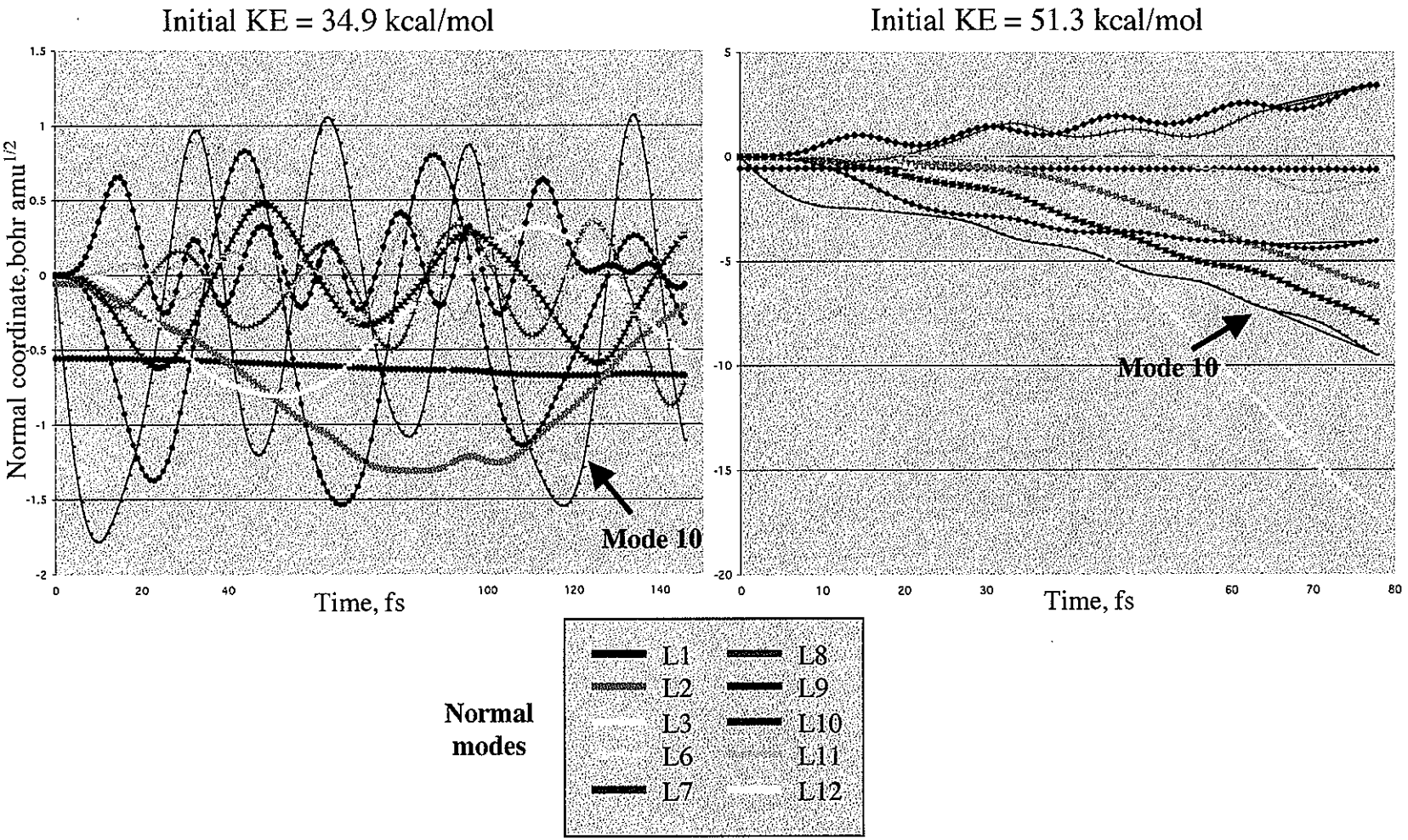


Figure 12. Normal coordinate change for cis isomer: Addition of KE to mode 10.

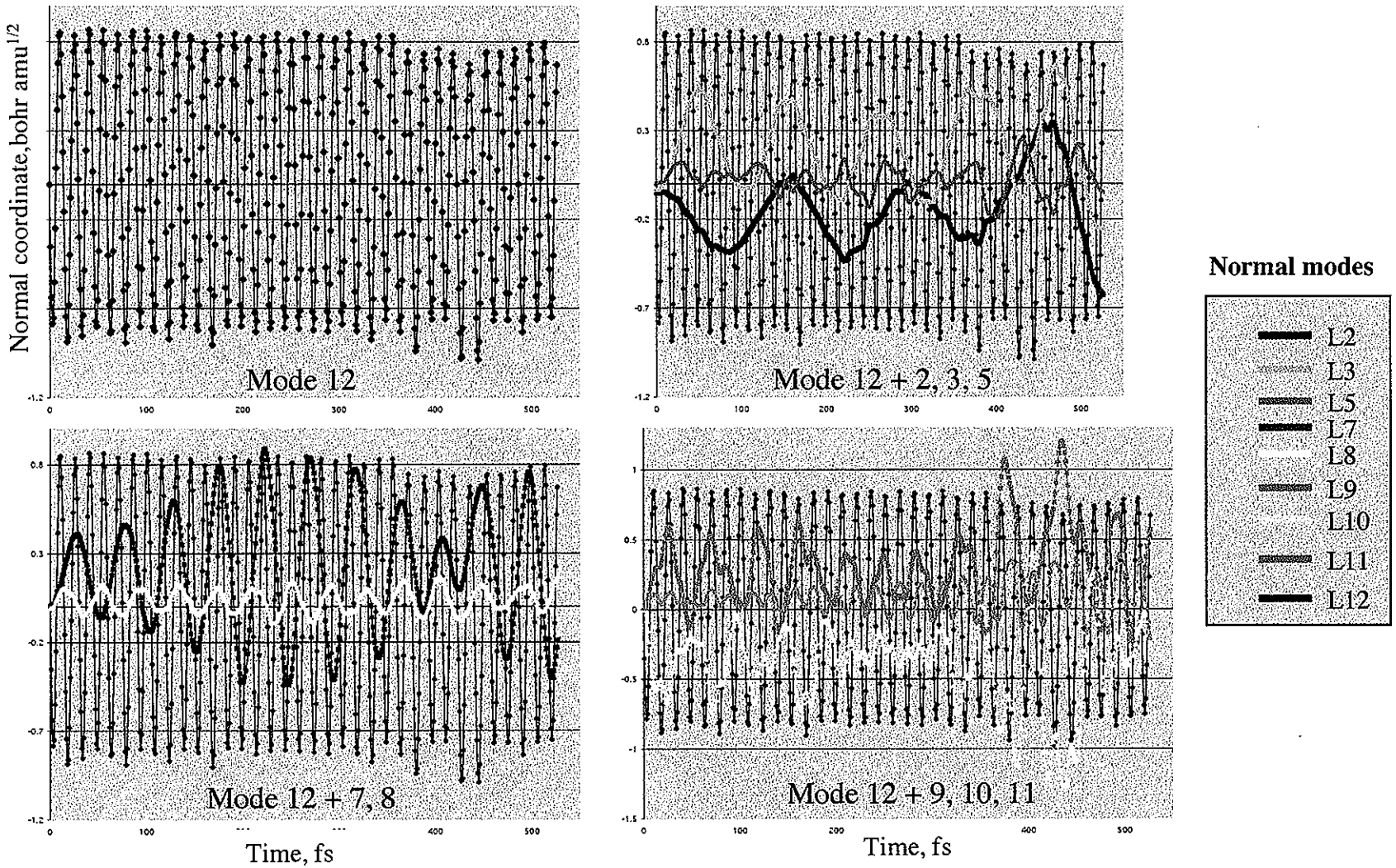


Figure 13. Normal coordinate change for harp isomer: Addition of KE to mode 12.

CHAPTER 3: FAST FRAGMENTS: THE DEVELOPMENT OF A PARALLEL EFFECTIVE FRAGMENT POTENTIAL METHOD

Taken from a paper submitted to the *Journal of Computational Chemistry.*

Heather M. Netzloff and Mark S. Gordon

Abstract

The Effective Fragment Potential (EFP) method for solvation decreases the cost of a fully quantum mechanical calculation by dividing a chemical system into an *ab initio* region that contains the solute plus some number of solvent molecules, if desired, and an "effective fragment" region that contains the remaining solvent molecules. Interactions introduced with this fragment region (for example, Coulomb and polarization interactions) are added as one-electron terms to the total system Hamiltonian. As larger systems and dynamics are just starting to be studied with the EFP method, more needs to be done to decrease the computation time of the method. This paper considers parallelization of both the EFP fragment-fragment and mixed quantum mechanics (QM)-EFP interaction energy and gradient calculations within the GAMESS suite of programs. The iteratively self-consistent polarization term is treated with a new algorithm that makes use of non-blocking communication to obtain better scalability. Results show that reasonable speedup is achieved with a variety of sizes of water clusters and number of processors.

I. Introduction

Quantum chemists realize that, as the size of a system increases, a full *ab initio* quantum chemistry calculation quickly increases in computational cost. This is especially true when trying to model environmental effects in the condensed phase due to the need to treat many more than one or two molecules. The Effective Fragment Potential (EFP) method[1,2] for solvation has been developed, in part, to address these concerns. The major premise of this method is the division of the system into a quantum mechanics (QM) region that contains the solute plus some number of solvent molecules, if desired, and an "effective fragment" region that contains the remaining solvent molecules. Interaction between solvent

molecules, represented by effective fragments, and the quantum mechanics (QM) part of the system is treated via one-electron terms; fragment-fragment interactions are treated in a similar manner. The EFP method has been shown to exhibit excellent results (for example, geometries, relative energies, spectral shifts, and reaction energetics)[3,4,5,6] for chemical reactions and small clusters in comparison with both full QM results and experimental trends. But, despite the reduced cost relative to fully QM calculations, the EFP method, due to its complex, QM-based potential, does require more computation time than simple interaction potentials, especially when the method is used for large scale molecular dynamics simulations.

This paper describes a distributed data approach to a parallel EFP method within the GAMESS program.[7] The EFP method is briefly introduced, followed by a description of the serial EFP code. Parallelization of the EFP code is then discussed. Results and analysis are presented with applications to water clusters (all EFP fragment molecules), as well as a QM glycine molecule solvated by a system of EFP waters. The accuracy of the potential used to describe a system must generally be balanced against computational efficiency and the available computational resources. The development of highly scalable and efficient parallel codes and access to high performance computers facilitates a focus on accuracy.

II. Theoretical/Computational Approach

A. EFP overview

The EFP method and various applications have been described in detail elsewhere,[1,2] so it is only briefly summarized here. The method is based on first principles quantum chemistry; thus, it can be systematically improved as computational methodologies and hardware improve.

The EFP method was originally designed and implemented to describe discrete solvent effects[1] with a focus on the effect of solvents on chemical reactions and, subsequently, on the study of small clusters of water molecules. Instead of performing a complete quantum calculation on the system of interest, the chemically important part (solute + some number of solvent molecules, if desired) is treated with some level of quantum mechanics, and the remaining solvent molecules are treated as interacting "fragments".

Therefore, the system Hamiltonian (H_{system}) is a sum of the active region Hamiltonian (H_{AR}) and the potential due to the fragment molecules (V):

$$H_{system} = H_{AR} + V \tag{1}$$

The fragment potential V contains only one-electron integrals; thus, an EFP calculation is much less computationally demanding than a full QM calculation. There are three terms in the EFP1 model that has been designed specifically for water:

$$V_{el}(\mu,s) = \sum_{k=1}^{K} V_k^{Elec}(\mu,s) + \sum_{l=1}^{L} V_l^{Pol}(\mu,s) + \sum_{m=1}^{M} V_m^{Rem}(\mu,s) \tag{2}$$

where m represents a fragment molecule, s the *ab initio* electronic coordinates, and k, l, and m label expansion points for the various terms described below. The first term represents Coulomb interactions between solvent molecules or between a solvent molecule and the QM solute. The second term models the solvent-solvent and solvent-solute induction/polarization interactions, and the third term represents the remaining interactions contained in the Hartree Fock potential.

The Coulomb term is represented by a distributed multipolar expansion[8,9,10] through octopoles. Since this is a point charge model, screening is included to account for overlapping electron density as molecules (fragment-fragment or fragment-solute) approach each other. Each nuclear center and bond midpoint are chosen as expansion points (Figure 1a).

The polarization term is based on a dipole-induced dipole potential. Polarizability tensors are located on the localized molecular orbital centroids. The polarization energy is iterated until self-consistency is reached, since a new induced dipole on a fragment affects the total electric field surrounding the QM part. This, in turn, modifies the induced dipole, and so on. For water, localized orbital polarizability tensors are placed at the centroids of the two O-H bonds, two oxygen lone pairs, and the inner shell oxygen orbital (Figure 1b).

The third, remainder interaction term in Eq. (2) is represented by simple Gaussian functions for QM-fragment interactions and simple exponentials for fragment-fragment

interactions. The exponents in these functions are optimized by a fitting procedure. This is accomplished by performing QM calculations on 192 water dimer structures.[2] For each point, the sum of the electrostatic and polarization energy components is subtracted from the total interaction potential, and the functional form (Gaussian or exponential) is then obtained by fitting to this remainder. Expansion points are the nuclear centers and the center of mass (Figure 1c).

The EFP method was originally based on the Hartree Fock (HF) method, using the DH(d,p) basis set,[11] EFP1/HF.[1,2] In this case, the fitted remainder term contains exchange repulsion + charge transfer interactions. More recently, the method has been extended to density functional theory (DFT) employing the B3LYP[12,13] functional, EFP1/DFT.[14] For the DFT implementation, the remainder term also includes some electron correlation effects. An MP2-level implementation, EFP1/MP2, that includes dispersion interactions, is under development.[15] A general version of the method, EFP2,[16,17] is derived from first principles and has no fitted parameters. At present, EFP2 includes the Coulomb and polarization terms discussed above, plus an exchange repulsion term derived from an intermolecular overlap expansion. The EFP method has been applied to a variety of problems, including small water clusters,[18,3,19] the mechanism of the Menshutkin reaction,[4] solvent effects on excitation energies,[20] ion solvation,[6] and protein pK_as.[21]

Since EFP does an excellent job of describing small clusters and chemical reactions, testing the method for larger systems and probing the ability of the method to describe bulk properties are logical next steps. It is therefore relevant to consider computation times. Table I compares wall clock times for the EFP method, fully QM calculations, and the simple SPC/E[22] water potential for an energy plus gradient calculation for several water cluster sizes. It is clear that EFP requires more time than SPC/E but still orders of magnitude less time than a full QM calculation. Of course, the more sophisticated EFP captures more of the fundamental physics than SPC/E. The origin of the difference in times for EFP1 and EFP2 is the more complex form used for the exchange potential in the latter.

B. Consideration of the EFP Parallel Code

GAMESS[7] uses the distributed data interface (DDI) for most of its parallel algorithms. This has been described in detail elsewhere;[23] only highlights are given here.

DDI is a data server model in the sense that two processes are assigned to each central processing unit (CPU). One process performs the "traditional" computational work, while the other process stores and services requests for data associated with distributed arrays. DDI is based on virtual shared memory (Figure 2). This means that part of the memory used by parallel processes in each CPU is reserved for storage of distributed data. The distributed memory is either local or remote. Local distributed memory is the memory used to store the distributed data associated with that particular CPU; remote distributed memory is memory reserved by other CPUs for their portions of the distributed data.

Each processor is allowed to access/modify any element in the distributed memory region. There are both replicated and distributed memory operations. Distributed data operations include one-sided gets, puts, and accumulates. Implementation of these functions is either native or non-native, depending on the system architecture. A native implementation implies that the underlying parallel library fully supports all necessary DDI communication operations, i.e. DDI is just a wrapper to this type of library. A non-native implementation means that the underlying parallel libraries are deficient in some area and more explicit programming is required to fully utilize DDI.

The EFP code currently in GAMESS is serial in nature. Each component of the EFP serial code is calculated independently. This means, for example, that the calculation of the charge-charge interaction energy is separate from the charge-octopole (Coulomb) interaction, the charge-induced dipole (polarization) interaction, etc. This is also true for portions of the gradient calculation. A breakdown of the components of the code is given in Figure 3. This format makes it much easier to apply parallelization in a consistent, yet interaction-specific manner.

The development of the parallel EFP code has followed a basic atom decomposition scheme. This means that during an energy or energy + gradient calculation, all processors store the data for all fragments, but that each processor is designated as the "owner" of a subset of fragments, i.e. each processor is responsible for the work necessary for each subset of fragments assigned to it. The size of this subset is based on the ratio n_{EFP}/n_p, where n_{EFP} is the total number of fragments in the system and n_p is the number of processors. For example, during a calculation with 16 fragments using 4 processors, processor 0 would "own" fragments 1-4, processor 1 would "own" fragments 5-8, etc. If the number of fragments is

not evenly divisible by the number of processors, the extra fragments are distributed among the processors as equally as possible.

Parallelization of the *Coulomb* and *exchange repulsion + charge transfer* subroutines is straightforward. In summing all interactions among all fragments, the serial code contains four fragment-fragment interaction loops. The two outermost loops are for fragment i and its expansion points k, while the two innermost loops are for fragment j and its expansion points m. This results in the piecemeal calculation of the total interaction energy of the k^{th} expansion point of fragment i with the m^{th} expansion point of fragment j. Thus, the total number of calculations n_{calc} is

$$n_{calc} = n_{EFP}(n_{EFP} - 1) * n_{expan} * n_{expan} \qquad (3)$$

where n_{EFP} is the total number of fragment molecules and n_{expan} is the number of expansion points used to describe a fragment (assuming that n_{expan} is the same for both fragments in the interaction energy calculation). The parallel code reduces this calculation by having the outer loop run over only those fragments that the particular processor "owns" (Figure 4). The total number of calculations is reduced to

$$n_{calc} = \frac{n_{EFP}}{n_p}(n_{EFP} - 1) * n_{expan} * n_{expan} \qquad (4)$$

where n_p is the number of processors.

In the above example (16 fragments - 4 processors), the outer loop for processor 1 would run over fragments 5-8 and the inner loop would continue to run over all 16 fragments. Thus, according to Eq. (4), in a single point energy calculation for the 16 fragment-4 processor system, each processor would only perform 3840 charge-dipole calculations (where n_{expan} = 8) rather than 15,360 calculations if the run was performed on a single processor. In the serial EFP subroutines such as charge-charge, dipole-dipole, etc., only the fragment i - fragment j interaction must be computed since the fragment j - fragment i interaction energy is the same. The total number of calculations in the serial code is therefore

$$n_{calc} = \frac{n_{EFP}}{2}(n_{EFP} - 1) * n_{\text{expan}} * n_{\text{expan}} \qquad (5)$$

To optimize scalability, the parallel code calculates both *i-j* and *j-i* interactions (n_{calc} is still given by equal Eq. 4) and then divides the resulting energy by two. This is done to assure load balancing among processors. The only communication associated with the Coulomb and exchange repulsion + charge transfer is *one* global sum at the end of the calculation.

The *polarization* term provides a greater challenge because the induced dipoles are calculated iteratively until self-consistency is reached. The initial electric fields and induced dipoles are calculated before entering the iterative subroutine. The induced dipoles are updated within each iteration step. For example, the interactions between fragment 1 and all other fragments are calculated to find the new induced dipole for fragment 1. During the calculation of the interactions of fragment 2 (immediately following), the *new* induced dipole of fragment 1 is used.

In the parallelization of the polarization contribution, only the "local" initial electric fields and induced dipoles are calculated, where "local" means those fragments that are "owned" by a processor. In order to avoid excessive communication of the induced dipoles, the dipoles are *not* updated within each iterative step. Figure 5 compares the serial and parallel implementations. Rather than replacing the induced dipoles with new values within the fragment loops, the new values are saved and then applied after convergence is tested. This approach increases the number of iterations (this is chemical system dependent), but gives the same resultant energy.

The parallel polarization code uses non-blocking communication for the induced dipole computation within the iterative fragment loops. The purpose of this approach is to overlap communication and computation. Figure 6 gives a schematic for the algorithm. This type of communication works in conjunction with saving (vs. replacing) induced dipoles. A cyclic, right-shift of induced dipole data is used. The 'right shift' terminology assumes that the processors are physically arranged in a ring (i.e. processor 0 is connected to the last processor). Each processor sends or receives data from the processor immediately adjacent to it. A 'right shift' means that the data moves in a clockwise direction (for example,

processor 1 would right shift its data to processor 2); a 'left-shift' means that the data movement is counterclockwise.

The parallel algorithm begins with a copy of the "local" induced dipoles to a buffer. A loop running over the total number of processors is initiated. A non-blocking send is posted to transmit the buffer to the right processor neighbor, while a non-blocking receive is posted to retrieve the buffer contents from the left processor neighbor. The fragment-fragment interaction loops are entered with the outer loop again running over the "local" fragments and the inner loop running over those fragments in the buffer. Thus, since no sends or receives are complete for the first set of calculations, the saved buffer will contain "local" fragments, and only "local" interactions are calculated. Waits are posted to complete sends and receives after all calculations between the "local" fragments and the buffer fragments are complete. The received buffer (containing new data from the left neighbor) is then copied to the send buffer space and this process of initiating a send and receive, calculating interactions among "local" and buffer fragments, and collecting new buffer information is continued until all interactions among the "local" fragments and all other fragments is complete.

The current DDI implementation within GAMESS does not contain non-blocking communication. Therefore, the code was rewritten so that DDI becomes a wrapper to non-blocking MPI functions for native implementations. Special code was written for TCP/IP sockets (non-native implementations) based on the use of threads to carry out a normal TCP/IP send/receive; thread termination is the signal that communication is complete and is detected by a wait.

These parallelization techniques were also applied to the mixed QM/EFP code. Most QM calculations are already parallel in GAMESS, so the existing code was utilized. Minor changes to the QM/EFP interfaces were required to complete the QM/EFP parallelization.

II. Results and Discussion

The code was tested on a cluster of IBM workstations composed of 22 dual 200 MHz Power3 nodes and 34 quad 375 MHz Power3-II nodes. The high speed interconnect is Gigabit Ethernet, and the operating system is IBM AIX. Results are presented for several sizes of water clusters (EFP-EFP interactions only), as well as one QM glycine molecule

surrounded by EFP waters. The glycine molecule is treated with HF/6-31++G(d,p) (125 basis functions). All timings are based on wall clock times.

Figure 7 shows the energy + gradient calculation speedup curves for water clusters ranging in size from 64-1024 on 2-16 processors. Speedup is defined as the calculation time on one processor divided by the time on p processors. The serial time for 1024 waters is 648.0 sec, and the parallel time on 16 processors is only 66.0 sec. This speedup will have a significant impact if one is interested in, for example, molecular dynamics simulations. Note that the jump from one to two processors is slightly skewed since the algorithm for the polarization routine changes from the original serial program to the new parallel method described here.

Table II shows timings for the individual energy components + gradient for the 1024 EFP water cluster. The calculation of the polarization energy is geometry dependent due to the iteration procedure, but Coulomb and polarization energy calculation times are comparable. Energy components and the total gradient exhibit good speedup as the number of processors is increased. Calculation times are very small with fewer fragments, but, as the size of the system increases, speedup improves and scalability is apparent. Table III shows the "raw" CPU and wall clock times for the energy + gradient calculation with 1 and 16 processors for each system shown in Figure 7. Differences in CPU and wall times on 1 processor are most likely due to disk I/O. The CPU and wall times are comparable, indicating that not much time is lost in communication.

Table IV compares the wall clock times for QM glycine with an increasing number of EFP waters on one processor. Times for the one-electron integrals and gradients are given, as well as the times for the computation of the total energy + gradient. The introduction of fragments does not affect the two-electron timings for glycine as EFP interactions are added as one-electron terms only. As the size of the QM component grows and/or the level of theory or basis set becomes more complicated, the QM calculation will quickly overshadow the EFP times. But, even for smaller QM systems, as well as very large numbers of EFP molecules, parallelization does have an impact. Figure 8 shows the total energy + gradient speedup curves for glycine plus 16-256 EFP waters. Figures 9 and 10 decompose this in terms of the fragment integral and total one-electron integral times (Figure 9) and fragment gradient and total one-electron gradient times (Figure 10). All graphs show that as the

number of EFP fragments is increased, the speedup curves also improve, indicating good scalability of the present code. The time to compute the entire one-electron gradient consumes approximately three-fourths of the total run time for an energy + gradient calculation; thus, it is important that this part of the parallel code performs well. Super-linear speedup can most likely be attributed to system cache.

III. Conclusions

Parallelization of the EFP method within GAMESS has been accomplished using the distributed data interface. Good speedup is achieved for both EFP-EFP and QM/EFP interactions. As expected, the best use of the parallel code is with large systems on an appropriate number of processors. The polarization energy calculation can be the most time consuming part of the EFP calculation due to the system and/or geometry dependence of the iteration scheme, but the use of the new induced dipole algorithm and non-blocking communication has fit into this calculation in a beneficial manner. Molecular dynamics (MD) code is being developed with the EFP method in GAMESS in order to test the ability of the EFP method to predict bulk properties. Preliminary results for radial distribution functions (RDF)[24] suggest that as the underlying EFP quantum method is improved from HF to DFT to MP2, the agreement with the experimental RDF also improves. The MP2-based EFP method yields a RDF that is in excellent agreement with experiment. Simulations require many thousands or millions of time steps in order to accurately represent phase space, therefore, the parallel EFP development will allow more efficient calculations of larger systems, as well as the exploration of many different areas of configuration space.

Acknowledgements

The calculations in this work were performed on an IBM workstation cluster made possible by grants from IBM in the form of a Shared University Research grant and the United States Department of Energy. The authors would also like to thank Ryan Olson and Dr. Michael Schmidt for writing the special code for TCP/IP sockets for the non-blocking communication routines in GAMESS, as well as many helpful discussions. HMN was supported by a Department of Energy Computational Science Graduate Fellowship and a

Miller Fellowship from the Iowa State University Department of Chemistry. Development of the parallel EFP code was supported by a CHSSI grant from the Department of Defense.

References

1. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. Journal of Chemical Physics 1996, 105(5), 1968-1986.

2. Chen, W.; Gordon, M. S. Journal of Chemical Physics 1996, 105(24), 11081-11090.

3. Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. Journal of Chemical Physics 2000, 112(5), 2063-2073.

4. Webb, S. P.; Gordon, M. S. Journal of Physical Chemistry A 1999, 103(9), 1265-1273.

5. Krauss, M. Computers & Chemistry (Oxford) 1995, 19(1), 33-38.

6. Petersen, C. P.; Gordon, M. S. Journal of Physical Chemistry A 1999, 103(21), 4162-4166.

7. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; J. A. Montgomery. J Comput Chem 1993, 14, 1347.

8. Buckingham, A. D. Quarterly Rev (London) 1959, 8, 183-214.

9. Stone, A. J.; Alderton, M. Molecular Physics 1985, 56(5), 1047-1064.

10. Stone, A. J. The Theory of Intermolecular Forces; Oxford University Press: Oxford, 1996.

11. Dunning, T. H.; Hay, P. J. In Methods of Electronic Structure Theory; Shaefer, H. F., Ed.; Plenum: New York, 1977, p 1-27.

12. Becke, A. D. Physical Review A: Atomic, Molecular, and Optical Physics 1988, 38(6), 3098-3100.

13. Lee, C.; Yang, W.; Parr, R. G. Physical Review B: Condensed Matter and Materials Physics 1988, 37(2), 785-789.

14. Adamovic, I.; Freitag, M. A.; Gordon, M. S. Journal of Chemical Physics 2003, 118(15), 6725-6732.

15. Song, J.; Gordon, M. S. in preparation.

16. Jensen, J. H.; Gordon, M. S. Molecular Physics 1996, 89(5), 1313-1325.

17. Jensen, J. H.; Gordon, M. S. Journal of Chemical Physics 1998, 108(12), 4772-4782.

18. Merrill, G. N.; Gordon, M. S. Journal of Physical Chemistry A 1998, 102(16), 2650-2657.

19. Merrill, G. N.; Webb, S. P. Journal of Physical Chemistry A 2003, 107(39), 7852-7860.

20. Krauss, M.; Webb, S. P. Journal of Chemical Physics 1997, 107(15), 5771-5775.

21. Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. Journal of Physical Chemistry B 2002, 106(13), 3486-3494.

22. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. Journal of Physical Chemistry 1987, 91(24), 6269-6271.

23. Fletcher, G. D.; Schmidt, M. W.; Bode, B. M.; Gordon, M. S. Computer Physics Communications 2000, 128(1-2), 190-200.

24. Netzloff, H. M.; Gordon, M. S. Journal of Chemical Physics 2004, in press.

Table I. EFP performance, energy + gradient calculation (single processor).[1]

Method	20 water molecules	62 water molecules	122 water molecules	512 water molecules
Ab initio[2]	3.19 hrs	---	---	~6 yrs[3]
EFP2	3.3 sec	26.1 sec	95.3 sec	26.8 min
EFP1	0.2 sec	2.6 sec	5.1 sec	97.8 sec
SPC/E[4]	0.02 sec	0.02 sec	0.1 sec	0.7 sec

[1]Run on 1200 MHz Athlon/Linux machine. [2]*Ab initio*: DZP basis set. [3]Assuming N^3 scaling, [4]SPC/E = Simple Extended Point Charge model.

Table II. Energy component wall clock times (sec) for a 1024 EFP water calculation.

n_p	Coulomb	Polarization	Remainder	TOTAL gradient
1	163	118	5	375
2	124	80	5	278
4	63	41	2	139
8	31	21	1	69
16	16	11	0.6	38

*n_p=total number of processors

Table III. CPU and wall times (sec) for EFP-EFP energy + gradient calculation with 1 and 16 processors.

1 processor

n_{EFP}	CPU time	Wall time
64	2.76	3.0
192	22.87	24.0
256	40.79	42.0
512	162.21	163.0
1024	648.86	664.0

16 processors

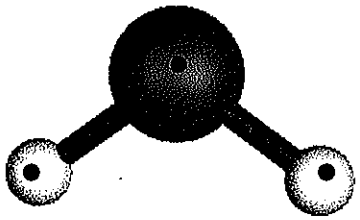
n_{EFP}	CPU time	Wall time
64	0.86	3.0
192	3.0	5.0
256	4.57	7.0
512	15.72	19.0
1024	60.84	65.0

* n_{EFP} = total number of fragments in the system

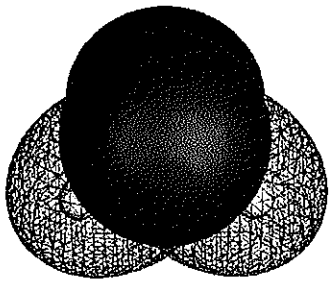
Table IV. Wall clock times (sec) for an energy + gradient calculation for QM glycine + EFP waters (1 processor).

System	One-electron integrals TOTAL time	One-electron gradient TOTAL time	TOTAL time Energy + Gradient
Glycine[1,2]	0.14	0.91	200
Glycine + 16 EFPs	25	252	479
Glycine + 64 EFPs	116	1147	1502
Glycine + 256 EFPs	424	4035	4898

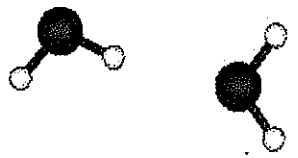
[1]Full QM calculation. [2]Glycine basis set = HF/6-311++G(d,p) = 125 basis functions



a. Electrostatics



b. Polarization



c. Exchange repulsion-
charge transfer

Figure 1. EFP method expansion points for the water molecule.

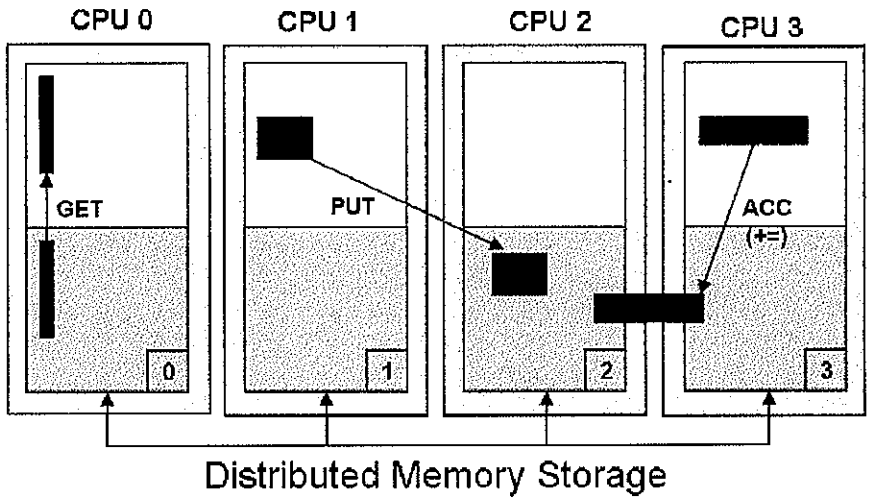


Figure 2. The virtual shared-memory model. Each large box (grey) represents the memory available to a given CPU. The inner boxes represent the memory used by the parallel processes (rank in lower right). The gold region depicts the memory reserved for the storage of distributed data. The arrows indicate memory access (through any means) for the distributed operations: get, put and accumulate.

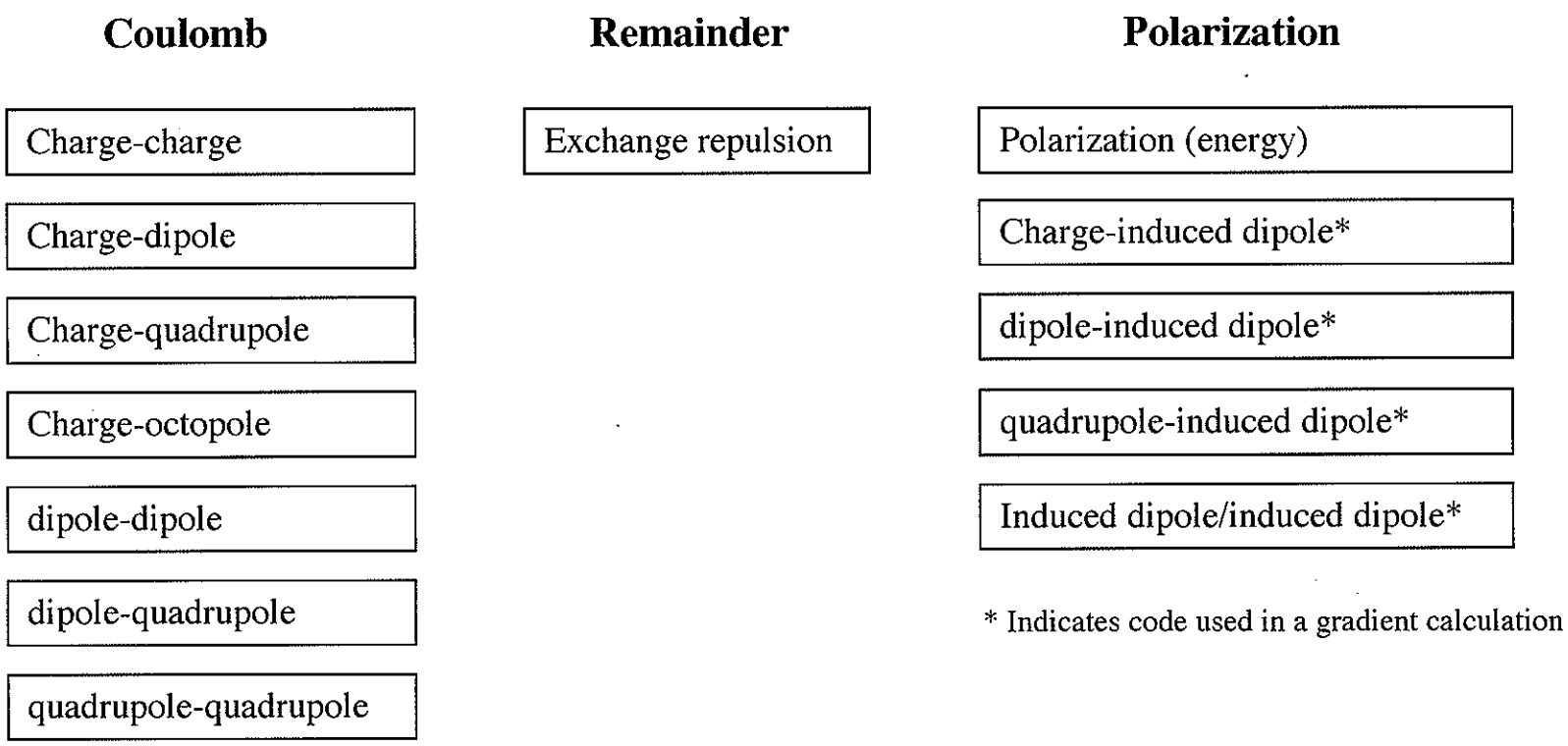
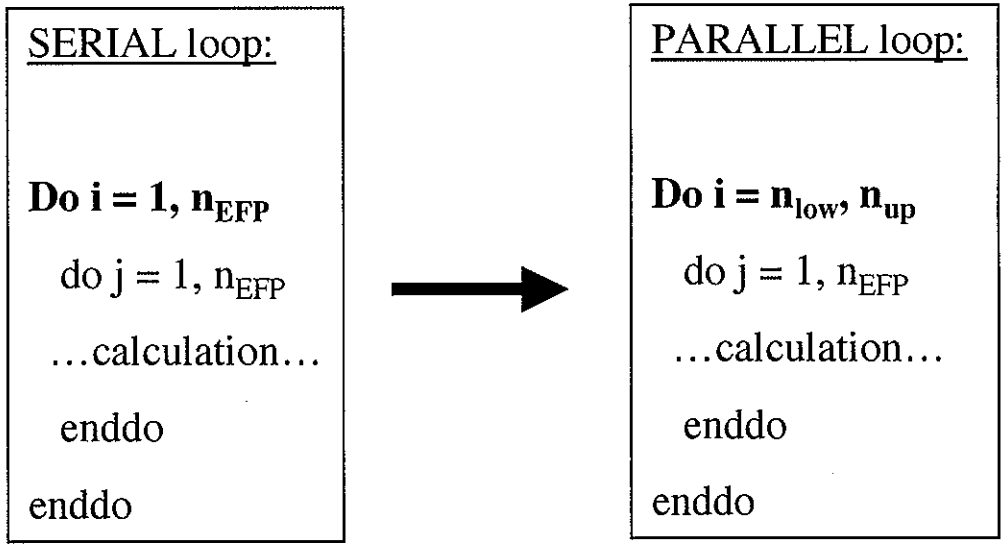


Figure 3. Serial EFP code: independent subroutines for interaction potential calculation.



* n_{EFP} = total number of fragments in the system

* n_{low} and n_{up} = lower limit and upper limit of fragments "owned" by a processor

Figure 4. Serial and parallel algorithms for electrostatics and exchange repulsion + charge transfer subroutines.

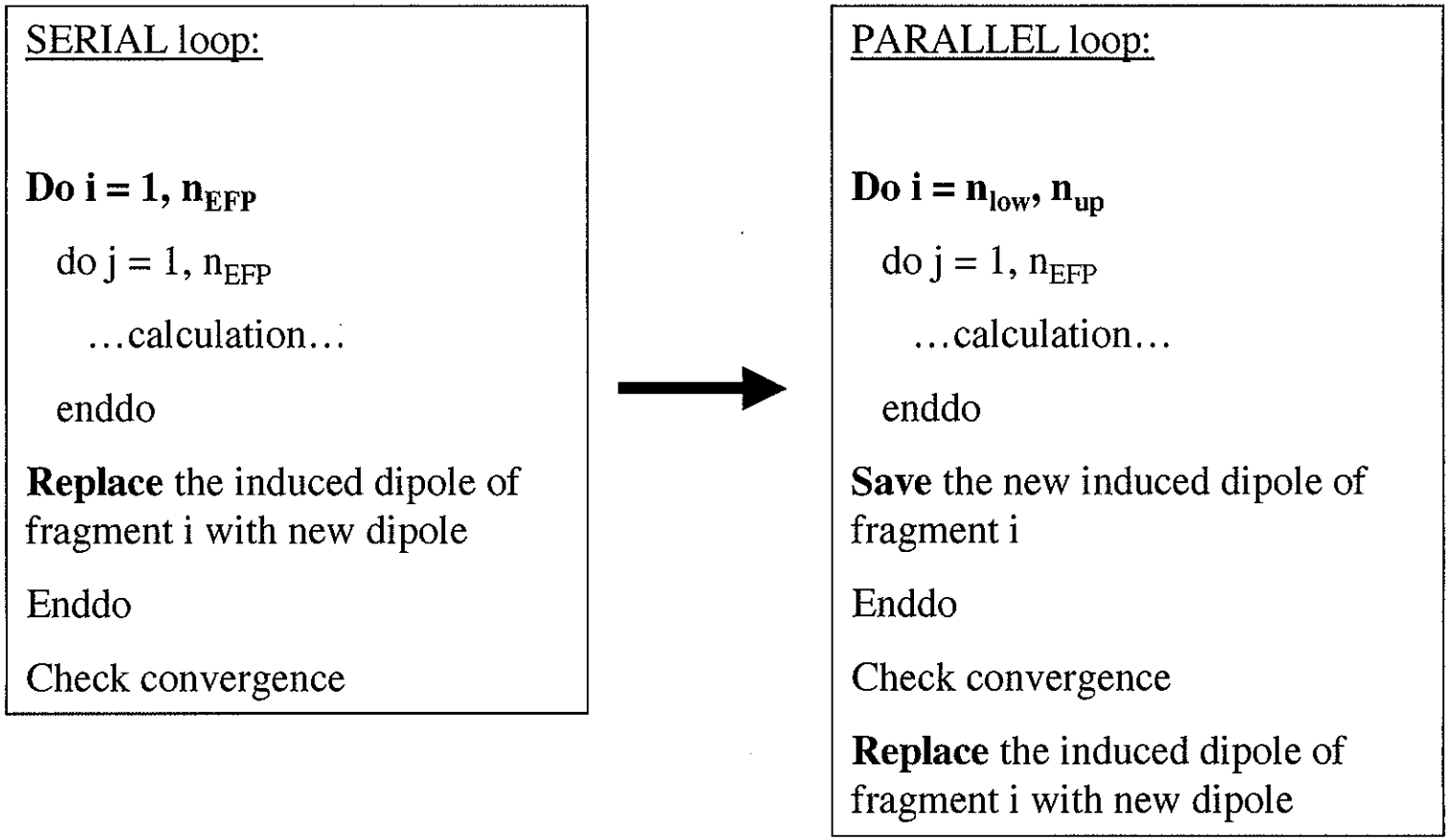


Figure 5. Serial and parallel algorithms for polarization subroutines.

```
Copy LOCAL induced dipoles to send buffer
do p = 1, n_p
  Start isend of send-buffer to right neighbor
  Post irecv from left neighbor
  do i = n_low, n_up
    do j = n_low(buffer), n_up(buffer)
      ...calculation (first iteration will be interaction among
         LOCAL fragments)...
    enddo
  Save the new induced dipole of fragment i
  enddo
  Post waits--complete sends and receives
  Copy received-buffer to send-buffer space
enddo (Continue send/recv until all interactions calculated)
```

*n_p=total number of processors

*isend, irecv indicate non-blocking sends or receives

Figure 6. Non-blocking communication algorithm in parallel polarization subroutine.

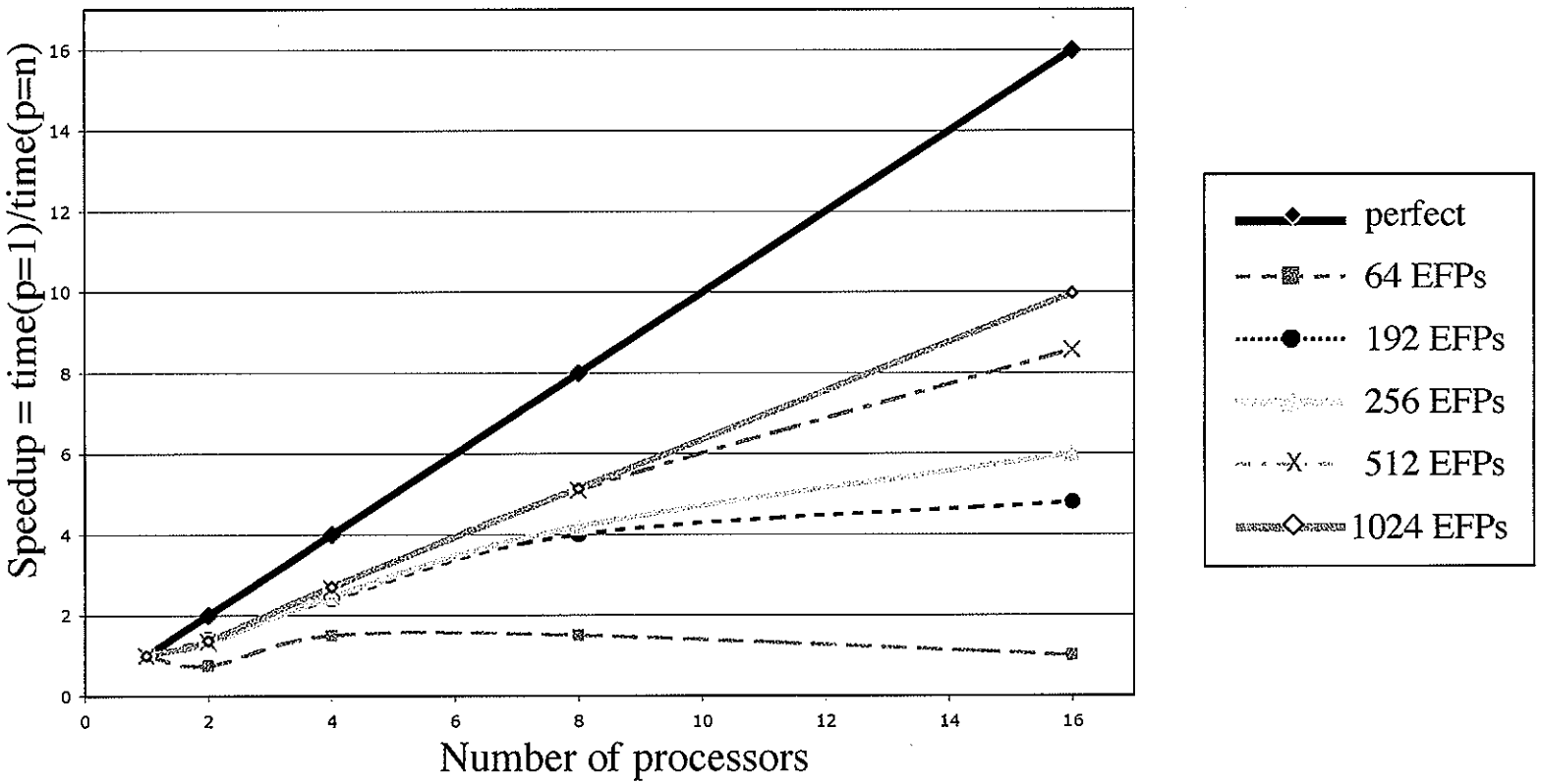


Figure 7. Speedup curve for energy + gradient calculation with EFP-EFP waters.

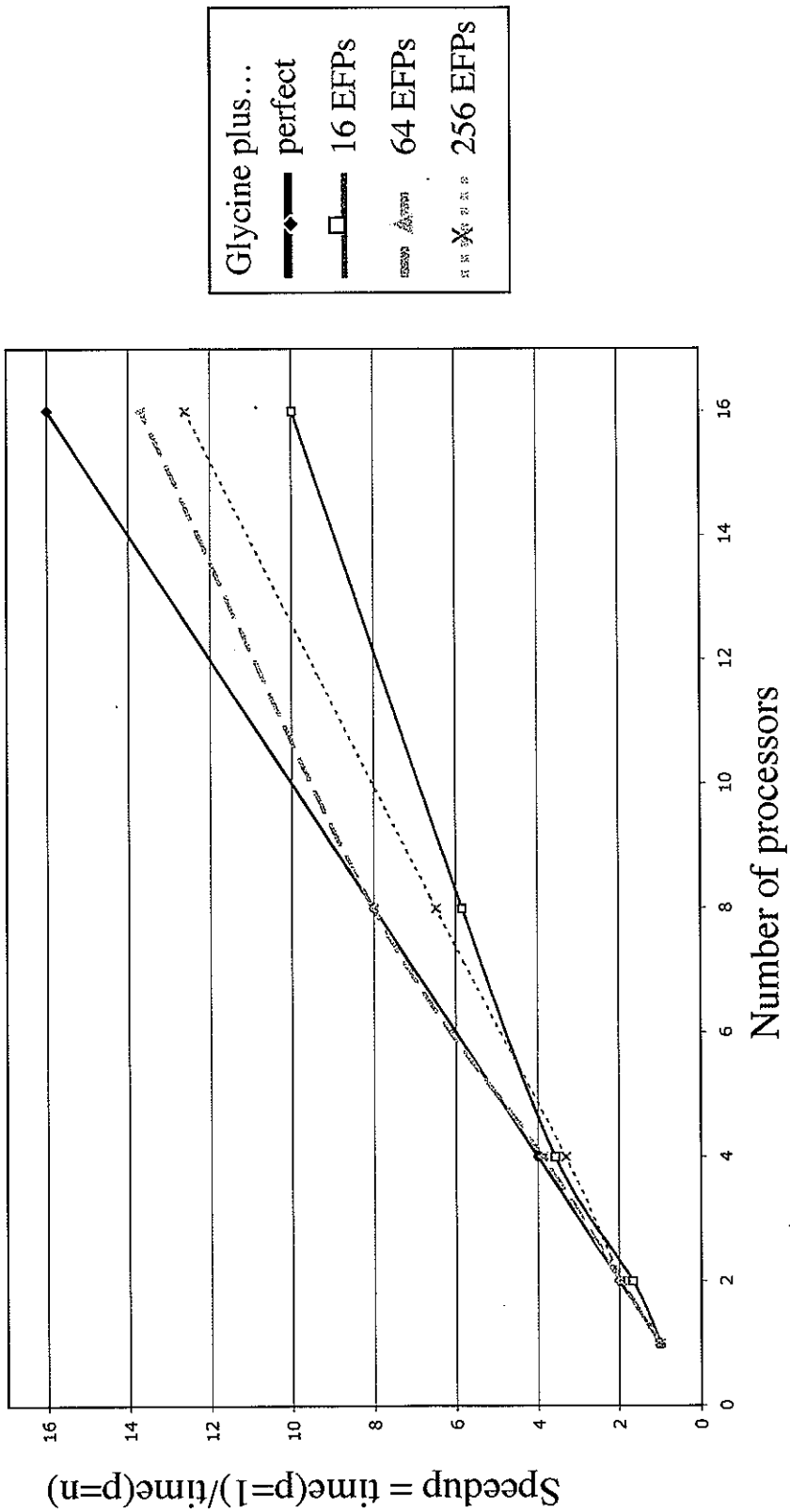


Figure 8. Speedup curve for energy + gradient calculation of glycine + EFP waters.

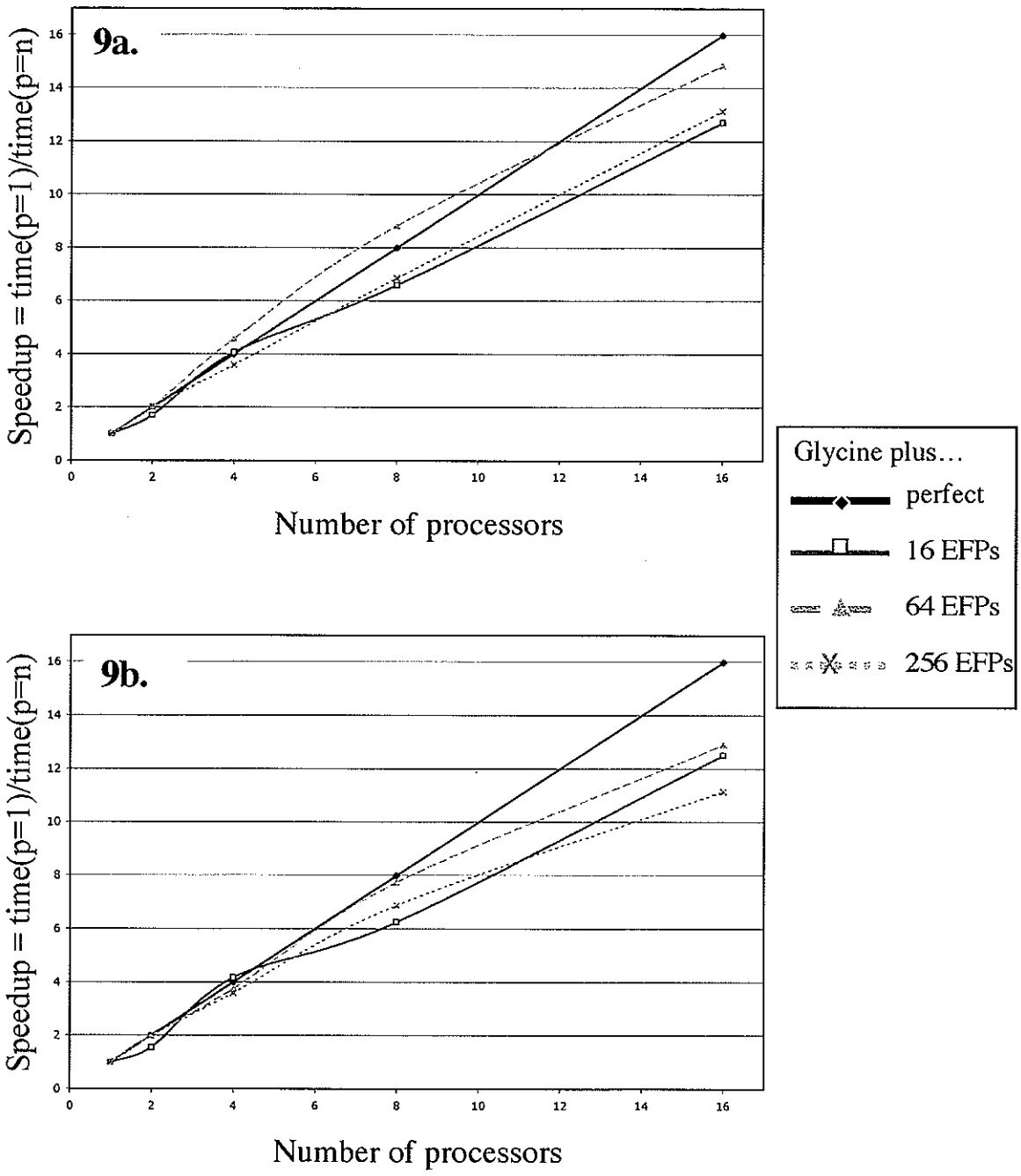


Figure 9. Speedup curve for energy + gradient calculation of glycine + EFP waters: (a) fragment integral speedup, (b) TOTAL one electron integral speedup.

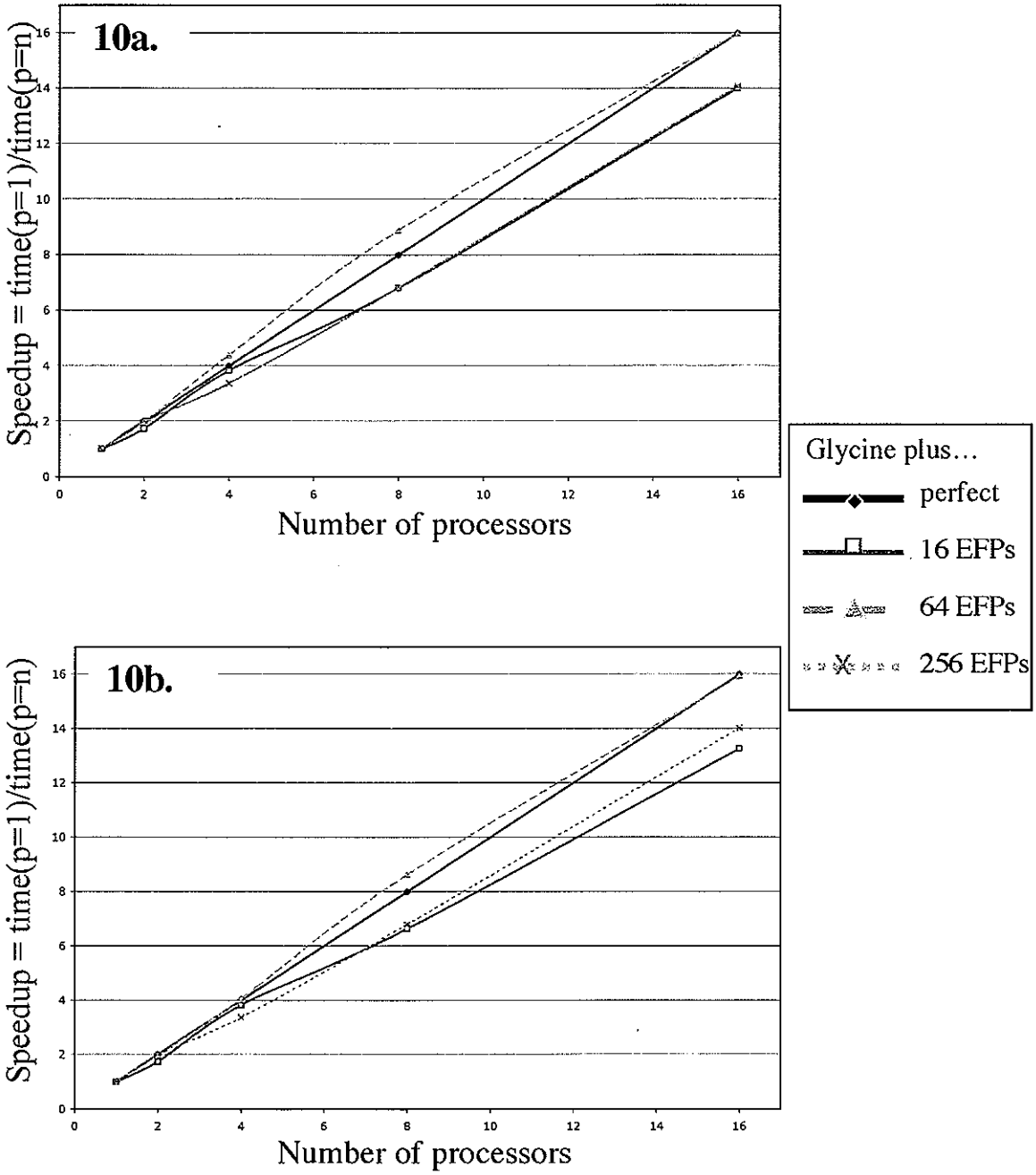


Figure 10. Speedup curve for energy + gradient calculation of glycine + EFP waters: (a) fragment gradient speedup, (b) TOTAL one electron gradient speedup.

CHAPTER 4: THE EFFECTIVE FRAGMENT POTENTIAL: SMALL CLUSTERS AND RADIAL DISTRIBUTION FUNCTIONS

Taken from a paper that is in press with the *Journal of Chemical Physics.*

Heather M. Netzloff and Mark S. Gordon

Abstract

The Effective Fragment Potential (EFP) method for treating solvent effects provides relative energies and structures that are in excellent agreement with the analogous fully quantum (i.e., Hartree-Fock (HF), density functional theory (DFT), and second order perturbation theory (MP2)) results for small water clusters. The ability of the method to predict bulk water properties with a comparable accuracy is assessed by performing EFP molecular dynamics simulations. The resulting radial distribution functions (RDF) suggest that as the underlying quantum method is improved from HF to DFT to MP2, the agreement with the experimental RDF also improves. The MP2-based EFP method yields a RDF that is in excellent agreement with experiment.

Paper

Computational quantum chemistry has been expanding its focus beyond primarily treating individual molecular species with sophisticated methods to developing accurate methods to model environmental effects. Of particular interest is the condensed phase, where most chemical and biological processes occur. Of paramount importance is the development of a method that is accurate and reliable across the full range from molecular to cluster to bulk solvation. Most discrete solvent methods are either most appropriate to the study of small clusters (due to high computational demands) or specifically parameterized to reproduce certain bulk behaviors.[1,2] An important exception is the water potential developed by Xantheas, et al., based on second order perturbation theory.[3]

The Effective Fragment Potential (EFP)[4,5] method was originally designed and implemented to describe discrete solvent effects. The initial focus was on accurate prediction of the effect of solvents on chemical reactions and on the study of small clusters of water molecules. Since the EFP method has been shown to accurately reproduce cluster properties,[6] the focus of the present work is to present preliminary results that illustrate the ability of the method to predict bulk properties. The EFP method and various applications have been described in detail elsewhere,[4,5] so only brief highlights are given here.

The EFP method expresses the energy as a sum of three terms:

$$E_{EFP} = E_{Coul} + E_{Pol} + E_{Rem} \qquad (1)$$

The first term represents Coulomb interactions and is described by a distributed multipolar expansion of fragment charge densities up to octopoles.[7] Nuclear atom centers and bond midpoints are chosen as expansion points. The second term models dipole-induced dipole interactions that are iterated to self-consistency; dipole polarizabilities are centered on bond and lone pair localized molecular orbitals. The final, remainder term E_{Rem} contains all interactions not accounted for by the Coulomb and polarization terms. Internal EFP geometries are fixed.

The EFP method is being developed based on several quantum mechanics (QM) methods. All of these methods treat Coulomb and polarization terms as described above but differ in the remainder term. These 'EFP1' methods (specific to water) each fit E_{Rem} to a functional form. In EFP1/Hartree Fock (HF),[4,5] E_{Rem} contains exchange repulsion + charge transfer. EFP1/Density Functional Theory (DFT),[8] based on the B3LYP functional,[9] includes some electron correlation effects in E_{Rem}. A second order many-body perturbation theory (MP2)-level implementation, EFP1/MP2,[10] contains separate terms for exchange repulsion + charge transfer and dispersion. A general version of the method (applicable to any species), 'EFP2',[11] is derived from first principles and has no fitted parameters. The EFP method has been applied to small water clusters,[6] the mechanism of the Menshutkin reaction,[12] solvent effects on excitation energies,[13] ion solvation,[14] environmental effects on biomolecular systems,[15] protein pKas,[16] and the prediction of shielding constants in solution.[17]

To illustrate the performance of both quantum methods HF, DFT(B3LYP), MP2, and the corresponding EFP methods on small clusters, consider the water hexamer structures given in Table 1. The binding energies for the five lowest energy isomers predicted by the three QM methods and their EFP analogs are compared in the table. Coupled cluster (CCSD(T)) energies are also included.

All EFP based methods track their QM counterparts very well. The maximum energy difference between HF and EFP1/HF is 2.0 kcal/mol. Similarly, the maximum energy differences for the DFT and MP2 methods are 1.3 and 0.5 kcal/mol, respectively. All methods except HF and EFP1/DFT predict the same isomer order. Although HF predicts the cyclic structure to be the most stable, the correct order is predicted for the other four isomers; EFP1/DFT reverses the order of the first three isomers, but the energy difference between these structures is only about 0.7 kcal/mol (within the expected deviation of the method). Assuming CCSD(T) to be the closest to the correct binding energies, HF greatly underbinds the water hexamers, while DFT provides significant overbinding. MP2 binding energies are slightly too high and in the best agreement with CCSD(T).

Since the EFP methods reproduce the corresponding QM results very well for these small clusters, it is not unreasonable to expect that, when applied to the prediction of bulk properties, the EFP methods are likely to perform with an accuracy that is comparable to the corresponding fully quantum methods.

In order to test the ability of the EFP method to predict bulk properties, an EFP molecular dynamics (MD) code has been implemented within the GAMESS (General Atomic and Molecular Electronic Structure System) program.[18] The EFP MD implementation in GAMESS and the results described here currently employ the leapfrog finite-difference algorithm to treat translational motion; quaternions are used to treat rotational motion with a modified leapfrog scheme. Periodic boundary conditions and the minimum image convention are used. Results are presented for a cluster of 62 water molecules. All calculations were done with GAMESS using the EFP1/HF, EFP1/DFT, and EFP1/MP2 methods. For comparison, the SPC/E method[2] was also used with the same MD code. SPC/E has point charges on the oxygen and hydrogen sites. Water-water interactions are described with a simple Lennard Jones 6-12 potential between oxygen atoms and a simple Coulomb interaction between all other atoms. The method includes a polarization correction

and was parameterized to reproduce bulk properties of water (e.g., experimental density and heats of vaporization).

A useful quantity with which to assess the performance of the EFP methods for predicting bulk properties is the radial distribution function (RDF). For effective fragment potentials, the RDF is a site-site distribution function that gives the probability of finding a pair of sites on different molecules a distance r apart, relative to the probability expected from a completely random (ideal gas) distribution at the same density. For water, there are three site-site RDFs: gOO(r), gOH(r), and gHH(r). These distribution functions are directly related to the structure factor of molecular fluids and hence to experimentally observable properties (e.g. those obtained from neutron diffraction or X-ray scattering). In the following paragraphs, HF, DFT, and MP2 refer to the corresponding EFP methods.

The procedures used for obtaining the RDFs involve heating, equilibration, and finally a production run. The simulations use a box length that gives the experimental liquid water density, 0.997 g/L at 300K, for the given number of molecules, 12.30 Å for 62 waters. The simulations used a 1.0 fs time step. gOO(r) may be compared with experimental X-ray results,[19] while gOH(r) and gHH(r) are compared with neutron diffraction experiments.[20] All EFP1 structures were equilibrated for at least 26 ps before the 10 ps NVE production simulation that was used to obtain the RDF results. For the gOO(r) RDF (Figure 1a), HF misses the second structural peak, and the first peak is shifted to a larger value of r. In contrast, the DFT and MP2 RDFs have first and second peak positions that are nearly identical to the experimental results. The DFT intensities are too high, but the MP2 intensities reproduce both peaks, as well as the remainder of the experimental gOO(r) curve. Also, the ratio of the intensities of the first and second peaks predicted by the DFT and MP2 models are in excellent agreement with the experimental ratio. The DFT and MP2 ratios differ from experiment by only 0.02 and 0.01, respectively. Similar agreement is observed for the gOH(r) and gHH(r) RDFs (Figures 1b and 1c, respectively): HF peaks are shifted to lower values of r; DFT, while having a larger intensity than experiment, has peaks and valleys at the same positions. MP2 again reproduces the experimental curve very well.

As the EFP model is improved from HF to DFT to MP2, the predicted radial distribution functions become increasingly accurate. Since the cluster calculations suggest that each EFP level of theory represents the corresponding QM level of theory very well, it is

likely that the MD results reflect the abilities of the corresponding levels of theory to reproduce one of the simplest bulk properties, the RDF. As other important bulk properties are coded and calculated, for example, thermodynamic and other structural data and self-diffusion coefficients, the EFP method will very likely provide insight into how the corresponding QM level of theory would behave, without the cost of performing the full QM calculations.

With regard to CPU requirements, a 20-water EFP1 energy + gradient run on a 1.2 GHz Athlon computer requires 0.12 sec, compared with 0.01 sec for SPC/E and 3.19 hr for HF/DZP. For 62 waters, the EFP method requires 0.71 sec and SPC/E requires 0.02 sec. Assuming N^3 scaling, a HF/DZP energy + gradient calculation for 62 waters would take ~4 days. Thus, while EFP requires more time than SPC/E, it is orders of magnitude faster than HF/DZP. Since EFP is a more sophisticated potential, more of the fundamental physics is captured.

This paper has presented the first test of the ability of the EFP method to model bulk behavior using molecular dynamics with periodic boundary conditions. The radial distribution functions, especially the complex gOO(r), suggest that as the EFP method is systematically improved from HF to DFT to MP2 (adding electron correlation), the agreement with the experimental curves improves. To the extent that the most significant difference between the DFT- and MP2-based methods is the incorporation of dispersion effects in the latter method, this suggests that dispersion is important in accurately reproducing the correct behavior. The success of predicting bulk behavior is particularly exciting since the EFP method has previously been shown to be an accurate approach for predicting explicit solvent effects on small clusters. Future work will focus on developing and testing additional bulk properties of interest.

Acknowledgements

The authors are grateful to Drs. William Swope, Raymond Mountain, and Greg Voth for helpful discussions. This work was supported in part by a grant from the Air Force Office of Scientific Research. HMN was supported by a Department of Energy Computational Science Graduate Fellowship.

References

1. W.L. Jorgensen and J.D. Madura, Mol. Phys. **56**, 1381 (1985).
2. H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
3. S.S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**, 1493 (2002).
4. P.N. Day, et. al. J. Chem. Phys. **105**, 1968 (1996).
5. W. Chen and M.S. Gordon, J. Chem. Phys. **105**, 11081 (1996).
6. (a) G.N. Merrill and M.S. Gordon, J. Phys. Chem. A **102**, 2650 (1998). (b) P.N. Day, et. al. J. Chem. Phys. **112**, 2063 (2000). (c) G.N. Merrill and S.P. Webb, J. Phys. Chem. A **107**, 7852 (2003).
7. (a) A.J. Stone and M. Alderton Mol. Phys. **56**, 1047 (1985). (b) A.J. Stone The Theory of Intermolecular Forces; Oxford University Press: Oxford, UK (1996).
8. I. Adamovic, M.A. Freitag, and M.S. Gordon, J. Chem. Phys. **118**, 6725 (2003).
9. (a) A.D. Becke, Phys. Rev. A **38**, 3098 (1988). (b) C. Lee, W. Yang, and R.G. Parr, Phys. Rev. B **37**, 785 (1988). (c) For the specific flavor of the B3LYP used, consult the GAMESS manual (date accessed 06/10/04): http://www.msg.ameslab.gov/GAMESS/doc.menu.html
10. J. Song and M.S. Gordon, in preparation.
11. (a) J.H. Jensen, J. Chem. Phys. **104**, 7795 (1996). (b) J.H. Jensen and M.S. Gordon, Mol. Phys. **89**, 1313 (1996). (c) J.H. Jensen and M.S. Gordon, J. Chem. Phys. **108**, 4772 (1998).
12. S.P. Webb and M.S. Gordon, J. Phys. Chem. A **103**, 1265 (1999).
13. M. Krauss and S.P. Webb, J. Chem. Phys. **107**, 5771 (1997).
14. M.S. Gordon and C. Peterson, J. Phys. Chem. A **103**, 4162 (1999).
15. (a) M. Krauss and B.D. Wladowski, Int. J. Quantum Chem. **69**, 11 (1998).; (b) R.M. Minikis, V. Kairys, and J.H. Jensen, J. Phys. Chem. A **105**, 3829 (2001).
16. H. Li, et. al. J. Phys. Chem. B **106**, 3486 (2002).
17. M.A. Freitag, et. al. J. Chem. Phys. **20**, 1197 (2004).
18. M.W. Schmidt, et. al. J. Comput. Chem. **14**, 1347 (1993).
19. J.M. Sorenson, et. al. J. Chem. Phys. **113**, 9149, (2000).
20. A.K. Soper, F. Bruni, and M.A. Ricci, J. Chem. Phys. **106**, 247 (1997).

Table 1. Binding energies for the five lowest energy isomers of the water hexamer.[1]

Binding Energy	HF[2]	EFP1/ HF[2]	DFT B3LYP[2]	EFP1/ DFT[2]	MP2[3]	EFP1/ MP2[3]	**CCSD (T)**
prism	42.86	42.42	62.37	61.08	58.25	58.26	**55.10**
cage	42.49	41.90	61.84	61.53	57.52	57.64	**54.30**
book	42.44	41.45	61.34	61.79	56.49	56.67	**53.10**
cyclic	43.10	41.14	60.57	60.65	55.75	55.24	**52.20**
boat	42.12	40.09	59.13	59.37	54.29	53.92	**50.80**

[1]Basis set = DH(d,p); energies in kcal/mol. [2]From ref. 8. [3]From ref. 10.

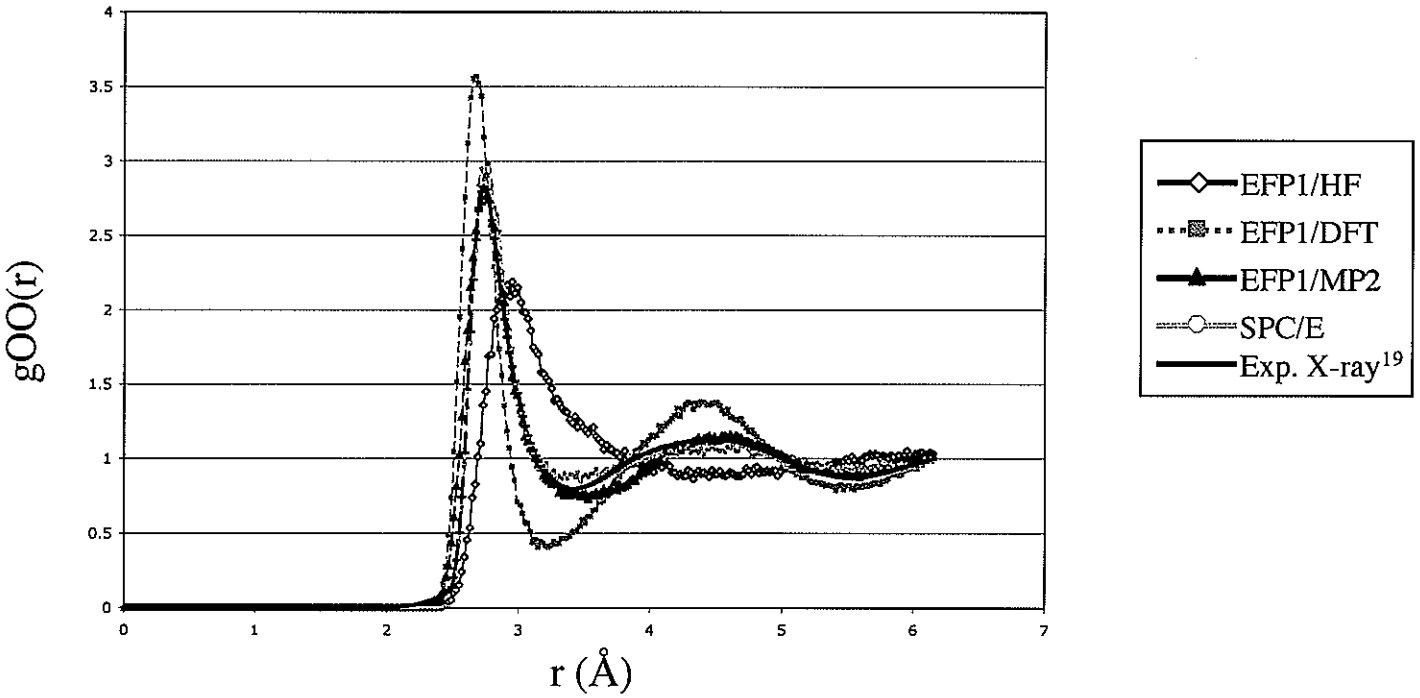


Figure 1. RDFs for 62 water molecules: (a) gOO(r).

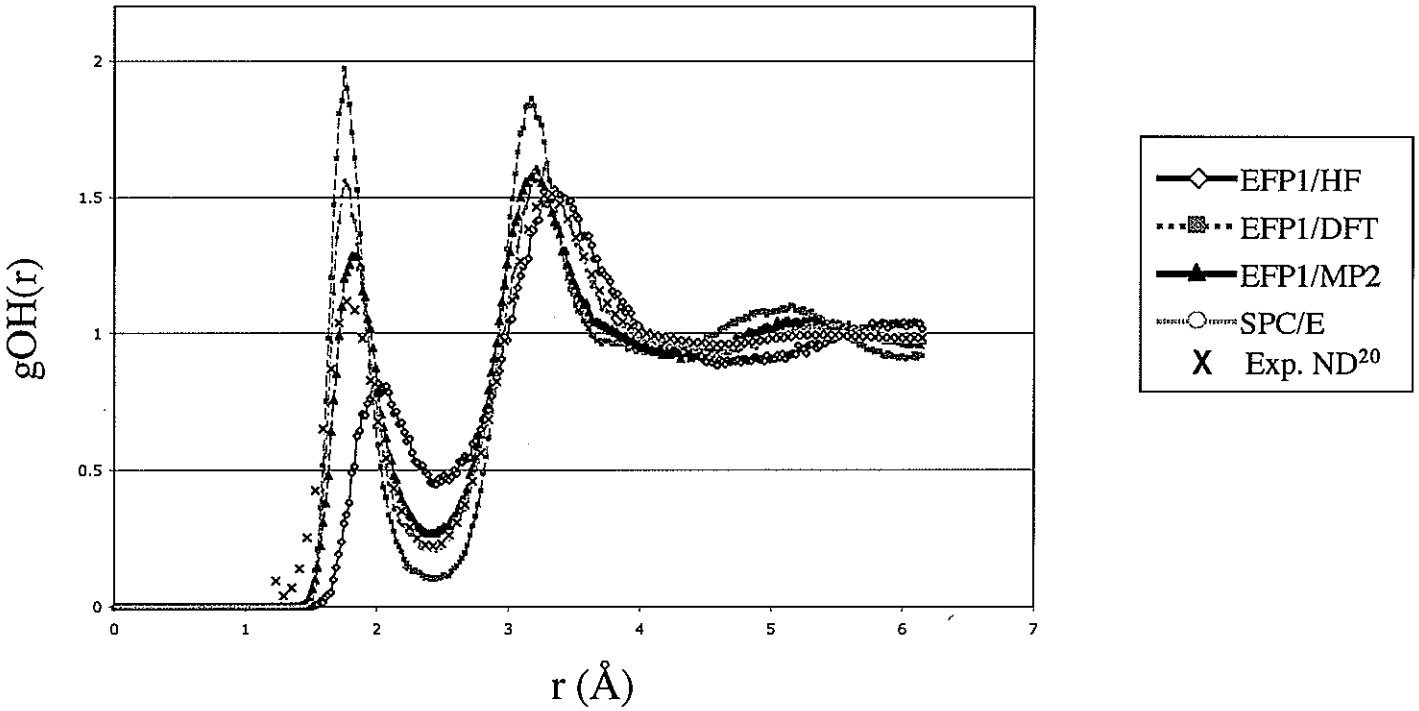


Figure 1. RDFs for 62 water molecules: (b) gOH(r).

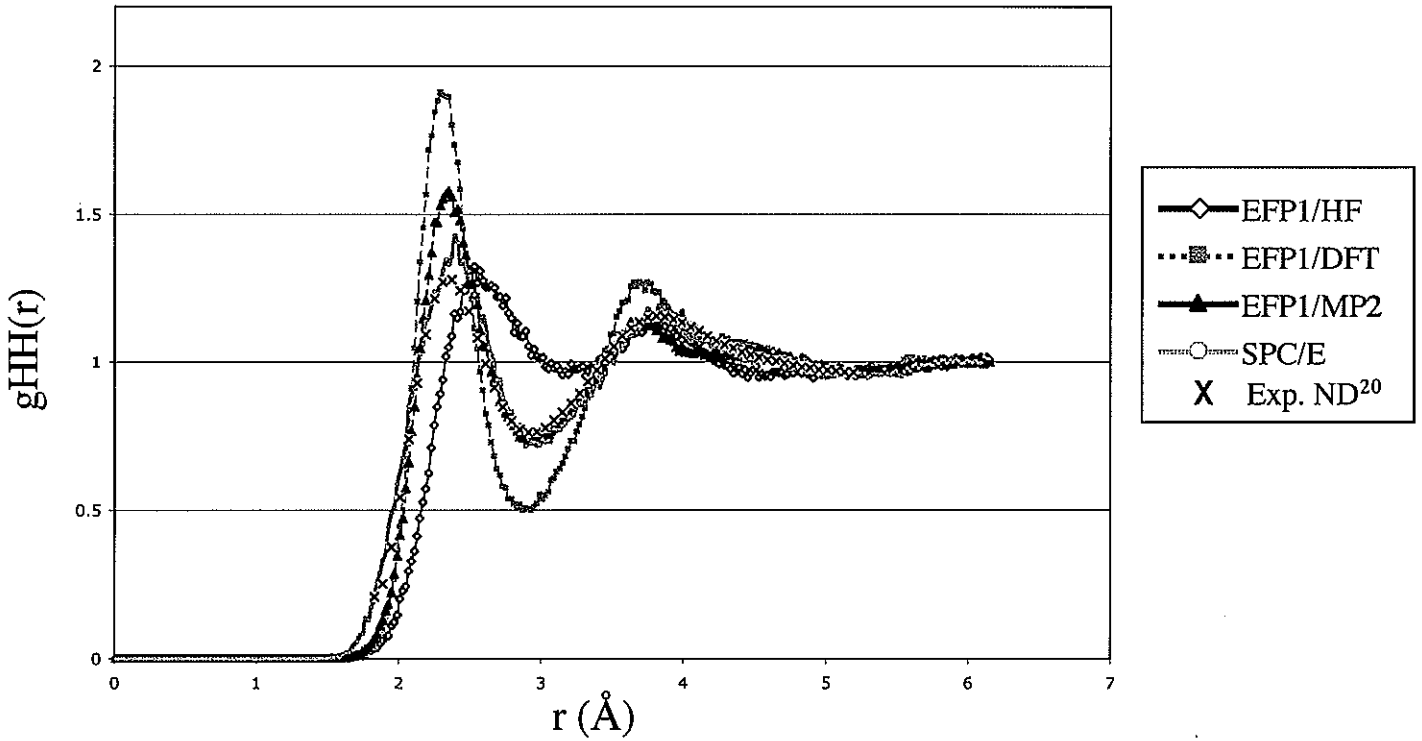


Figure 1. RDFs for 62 water molecules: (c) gHH(r).

CHAPTER 5: GROWING MULTI-CONFIGURATIONAL POTENTIAL ENERGY SURFACES WITH APPLICATIONS TO X + H_2 (X = C, N, O) REACTIONS

Taken from a paper to be submitted to the *Journal of Chemical Physics*.

Heather M. Netzloff, Michael A. Collins, and Mark S. Gordon

Abstract

Very accurate calculations on small molecules and molecular systems are now possible with *ab initio* quantum chemistry. In order to describe the dynamics of a reaction, information about the molecular potential energy surface (PES) is necessary. A previously developed method, based on a Shepard interpolation procedure to automatically construct a quantum mechanical potential PES, is extended to the construction of multiple potential energy surfaces using multi-configurational wavefunctions. These calculations are accomplished with the interface of the PES-building program, Grow, and the GAMESS suite of electronic structure programs. The efficient computation of multi-configuarational self-consistent field surfaces is illustrated with the C + H_2, N + H_2, and O + H_2 abstraction and exchange reactions.

I. Introduction

The investigation of the chemical dynamics of a molecular system allows time-dependent phenomena to be calculated. This facilitates both comparisons with experimental data and the prediction and interpretation of details not easily obtainable from experiments. In order to perform dynamics calculations over a large range of configurations, knowledge of the corresponding molecular potential energy surface (PES) must cover a broad range of coordinate space. The least-biased and most accurate method to obtain data about the surface is through first principles or quantum mechanics (QM) electronic structure calculations. With the advance of high performance computing and scalable codes, QM methods can provide very accurate calculations of small molecules and molecular systems. But, in most cases, QM calculations can only provide a static picture of the PES due to the large computational cost of such methods. The calculation of QM energies and first and second

derivatives (gradients and hessians) at selected discrete points only gives the overall *shape* of the PES. The difficulty arises in that a finite number of unrelated energy and derivative values *alone* do not constitute a molecular PES. The PES may be more formally defined as a function or algorithm that is capable of accurately describing the molecular energy while some interatomic distances change by several Angstroms and others by no more than 0.1 Å.

Many methods have been used to obtain a PES: fitting an assumed global functional form to the available QM data,[1,2] using some expansion of the potential about a reaction path[3], or interpolating in some way between known data points.[4-9] Direct dynamics can also be carried out "on-the-fly", with *ab initio* or semi-empirical calculations at every molecular configuration generated by the dynamics calculation.[10-12] The bottleneck in using electronic structure theory to build a PES and, therefore, the study of reaction dynamics, is the dimensionality of the system and the number of points needed in each dimension to accurately describe the PES. For a diatomic molecule, the PES is just a one-dimensional curve. For a tri-atomic system, a global, three-dimensional surface must be constructed. Four atoms require the determination of a six-dimensional surface. Beyond four atoms, the determination of a global PES in this manner is almost impossible. Of course, most interesting chemical systems contain more than four atoms. The number of points needed in each dimension depends on the molecular system and the desired properties. It is therefore desirable to design a method for constructing a global PES that is accurate, computationally cost-effective, and generic.

Collins and coworkers[4-7,15-17] have derived and developed a method based on a modified Shepard interpolation procedure to construct a molecular PES. The goal is to attempt to produce an accurate PES with the smallest number of QM calculations by carefully selecting the location of data points. The interpolation method used to construct the PES is combined with classical trajectory simulations to provide an iterative scheme for successively improving and updating the PES. The simulations allow configuration space to be broadly sampled. This methodology, implemented in the program Grow, has been developed to minimize the interpolation error for a given number of data points, thereby ensuring that the PES can be reliably constructed for a small set of data points without the need to fit to a functional form. Thus, reliance on chemical intuition and the inadvertent introduction of unphysical features can be avoided. Each iteration automatically places new

data points only in dynamically important parts of the configuration space; QM calculations are then performed at those points. For example, the $OH + H_2 \rightarrow H_2O + H$ reaction only required ~200 data points to accurately describe the PES;[5,6,7] ~1500 data points were needed for a reasonably accurate interpolation of the $H + CH_4 \rightarrow H_2 + CH_3$ reaction.[14]

The original implementation of Grow was confined to the construction of potential energy surfaces for ground electronic states. This paper describes the extension of Grow to enable the study of reactions on multiple surfaces and problems that require a multi-reference description. Since Grow requires QM energies and derivatives at a select number of data points in order to "grow" the PES, the interface of Grow and the quantum chemistry program GAMESS[18] is implemented. The paper begins with a brief introduction to the essential features of Grow, followed by a description of the ability to generate multiple MCSCF surfaces with the Grow-GAMESS interface. Finally, applications to X + H_2 reactions (X = C, N, O) will be discussed.

II. Methods Description and Development

A. Grow

The features of Grow have been described elsewhere[4-7,15-17] so only brief highlights will be given here. Grow constructs a PES as an interpolation of Taylor expansions centered at data points scattered throughout the configuration space of the system.[4] In general, a surface is constructed by expanding the inverse distance matrix **Z**
(**Z** = 1 / **R**, where **R** is the set of interatomic distances) as a Taylor series T_i. More physically reasonable behavior is obtained when the Taylor expansion is written in terms of inverse distances **Z** rather than interatomic distances:[4]

$$T_i(Z) = V[Z(i)] + \sum_{k=1}^{3N-6} [Z_k - Z_k(i)] \left.\frac{\partial V}{\partial Z_k}\right|_{Z=Z(i)} + \frac{1}{2!} \sum_{k=1}^{3N-6} \sum_{j=1}^{3N-6} [Z_k - Z_k(i)] \times [Z_j - Z_j(i)] \left.\frac{\partial^2 V}{\partial Z_k \partial Z_j}\right|_{Z=Z(i)} + \ldots \quad (1)$$

where V[**Z**(i)] is the potential energy at **Z**(i) and derivatives are taken with respect to inverse distances at the i^{th} data point **Z**(i). The expansion is truncated to second order.

For a system with more than four atoms, redundant internal coordinates can occur since the number of internuclear distances is greater than the number of independent internal coordinates, 3N-6. In this case, a set of 3N-6 local internal coordinates ζ are constructed as linear combinations of the N(N-1)/2 reciprocal distances for each **Z**(i):[14]

$$\zeta^{(X_0)}(X) = \tilde{U}^T Z \qquad \text{for Z close to } Z_0 = Z(X_0)$$

where X_0 refers to the initial set of Cartesian coordinates.

The matrix **U** is defined in terms of molecular fragments, both intra- and inter-molecular distances. The QM Cartesian derivatives of the PES at **X**(i) must then be transformed into these internal coordinates.

Once the required QM energy and derivatives have been evaluated at some number (N_d) of molecular configurations, a modified Shepard interpolation[19,20] defines the total potential energy (PE) V(**Z**) at any configuration **Z** as a weighted average of Taylor series about all N_d initial data points and their symmetry equivalents[7]

$$V(Z) = \sum_{g \in G} \sum_{i=1}^{N_d} w_{g \circ i}(Z) T_{g \circ i}(Z) \qquad (2)$$

G indicates the molecular symmetry group (usually the complete nuclear permutation (CNP) group) and g◦i indicates that the i^{th} data point **Z**(i) is replaced by a data point transformed by group element g. The sum over G means that all permutationally equivalent data points are included in the energy expression; no new QM calculations need to be performed for these symmetrically equivalent points. This procedure results in a symmetrized data set, and a PES that exhibits the full molecular symmetry.[21]

The normalized weight function w_i weights the contribution of Taylor expansions at **Z**(i) to the total PE at configuration Z:

$$w_i(Z) = \frac{v_i(Z)}{\sum_{g \in G} \sum_{k=1}^{N_d} v_{g \circ k}(Z)} \qquad (3)$$

Data points that are spatially close to **Z** will have a larger weight. The biasing of data is achieved using the un-normalized weight function $v_i(\mathbf{Z})$:

$$v_i(Z) = \left\{ \left[\frac{\|Z - Z(i)\|}{rad(i)} \right]^{2q} + \left[\frac{\|Z - Z(i)\|}{rad(i)} \right]^{2p} \right\}^{-1} \quad \text{for } q \ll p \qquad (4)$$

Here, q=2 and 2p >> 3N-3,[14] and the term rad(i) effectively defines a type of confidence radius about **Z**(i). If **Z** is within the confidence radius, the weight assigned to **Z**(i) will fall slowly as the distance increases. If **Z** is beyond the confidence radius, the weight assigned to **Z**(i) falls quickly with an increase in distance. As described elsewhere, the confidence lengths are determined through a Bayesian analysis.[15] This two-part weight function allows the relative weights of two or more data points near **Z** to vary only slowly with varying **Z**; this reduces the error in the PES gradient and is especially important for larger systems.[14]

The basic iterative procedure is as follows (see Figure 1).[4,5,7] After a small initial data set is constructed, usually on a relevant reaction path (the Shepard interpolation is thus well-defined on this reaction path), Grow runs a small number of trajectories with initial conditions appropriate to the reaction under study. Molecular configurations are periodically recorded to obtain a set of several thousand geometries. One or more of these geometries is chosen to be added to the data set, and the required QM energies and derivatives are calculated for this point. Finally, the new data point is incorporated into the data set to form a new version of the PES. This process continues with the calculation of more trajectories. As the data set grows, the interpolated PES becomes more accurate. The PES is considered to be converged to chemical accuracy when the average values of observables of interest (obtained from large sets of classical trajectories) are unchanged by further iterations and increase in the size of the data set.

The methods for choosing a new data point have been described in detail elsewhere.[7] The variance sampling method attempts to place data points at configurations where the uncertainty in the interpolated PE is the highest. The "h-weight" method[4,5] attempts to place data in regions that the trajectory often visits, but where few data points are already present. Both methods are used alternately to give a more "well-rounded" description of the PES.

B. MCSCF Surfaces

The initial implementation of the interpolation procedure for growing potential energy surfaces was limited to ground state surfaces and single configuration wavefunctions. For problems that involve significant di-radical character (frequently bond-breaking processes) or multiple electronic states, multi-determinant methods must be employed. The introduction of MCSCF wavefunctions requires the addition of new features to facilitate efficient construction of potential energy surfaces. For example, physically meaningful starting orbitals are especially important when performing MCSCF calculations. Also, when computing surfaces with multi-reference wavefunctions, it is critical to regularly update the orbitals. In a chemical reaction, bonds are breaking and forming, and electrons are constantly moving from orbital to orbital. The energies and occupation numbers of the orbitals must therefore be monitored, as well as the active space of the MCSCF calculation.

The interpolation procedure in Grow is initiated using data points from a reaction path calculation with GAMESS (usually an intrinsic reaction coordinate (IRC)[22] calculation). These geometries are used to generate an initial data file for the interpolated PES. The energy, gradient, and/or hessian are calculated at each of these initial points and saved in the PES data file. The MCSCF interface also saves each geometry and wavefunction in libraries to be used while growing the PES. The iterative growing procedure eventually chooses a new point to add to the data set. The geometry library is searched to find the closest possible matching geometry to this new point. The corresponding wavefunction is then used as the initial wavefunction for the MCSCF calculation on the new data point. A better guess for the initial wavefunction allows for a much more efficient calculation.

III. Application to X+H_2

The X + H_2 system on the ground state PES of the separated reactants is chosen to illustrate the new MCSCF potential energy surface growing process. There are three basic reactions to consider

$$X + H_2 \rightarrow XH_2 \quad \text{(insertion)} \quad (5)$$

$$X + H_2 \rightarrow XH + H \quad \text{(abstraction)} \quad (6)$$

$$H + XH \rightarrow HX + H \quad \text{(exchange)} \quad (7)$$

Following a brief discussion of the X + H_2 reactions (X = C, N, O), the specific potential energy surfaces will be discussed.

A. Background of N, C, and O reactions

The N + H_2 reaction network has been the subject of several previous experimental and theoretical studies [23,24,25] The forward abstraction reaction (Eq. 6) is of interest in the study of rocket fuel propellants[26], and the reverse abstraction reaction is relevant to the pyrolysis of ammonia.[27] Most experimental studies of the forward and reverse abstraction reactions have been focused on reaction rate constants.[23,27] Electronic structure calculations were also done to predict and confirm stationary point properties and rate constants for these reactions.[24,25] Xu and coworkers[24] found that the energy of reaction strongly depends on electron correlation and basis set size. For example, energies calculated at the unrestricted Hartree-Fock (UHF) level were overestimated in comparison with experimental values; the addition of polarization functions reduced the theoretical-experimental difference by ~10 kcal/mol. After the addition of electron correlation to UHF/6-311G(d,p) energies, barrier heights also dropped at the UMP2 and UMP4 levels of theory.

No experimental studies of the ground state N(^{4}S) abstraction reaction have been done to determine differential cross sections and characterize rovibrational and angular distributions of the products, due to the difficulty in preparation of N atoms and NH radicals. Since ground state N(^{4}S) is also not very reactive, most theoretical (and experimental) dynamics studies[25] focused on the dynamics of the lowest excited state N(^{2}D), as it is more reactive. Recently, Pascual and Schatz[28] used quasi-classical trajectory (QCT) and variational

transition state theory to study the dynamics of the forward and reverse abstraction and exchange reactions for ground state N(^{4}S) using an accurate PES derived from high level electronic structure calculations. Their motivation was to determine rate constants, as well as to provide information about reaction cross sections and product state distributions that would be useful to experimentalists. Since the level of theory and computation time must be balanced in a dynamics study, Pascual and Schatz, after assessing several alternative levels of theory and basis set sizes, settled on multi-configurational, quasi-degenerate second order perturbation theory (MCQDPT2) calculations with fully optimized reaction space (FORS) multi-configuration, self-consistent field (MCSCF) (FORS-MCSCF(7,6)/6-311++G(d,p)) molecular orbitals. GAMESS was used for the electronic structure calculations, and Grow was used to generate the PES for the QCT studies. The two transition states and the PES were accurately described based on 156 points, and the predicted rate constants are in reasonable agreement with available experimental and theoretical results.

C and O are adjacent to N, each with a ^{3}P ground state. Thus, both C+H_2 and O+H_2 are reported here for the triplet surfaces. Several theoretical dynamics studies of the CH_2 reaction on the triplet surface have been performed.[29,30] The study done by Harding, Guadagnini, and Schatz[30] fit multi-reference configuration interaction calculations to a functional form to characterize the PES. They found no barrier for the CH + H addition reaction or C + H_2 insertion reaction (both giving CH_2), and determined that reaction kinetics are dominated by activationless processes versus, for example, the abstraction reaction. Theoretical studies of the singlet surface of C + H_2 have also been performed[31]. To our knowledge, no studies of the triplet O + H_2 global PES are in the literature; many such studies have been done on the ground state singlet surface.[32]

B. General Methods

Electronic Structure Methods. The first step in growing a global PES is to choose an appropriate level of theory and basis set. Since a PES constructed using the procedure outlined above is an accurate interpolation of the exact PES *for the level of theory employed*, an inadequate level of theory can result in a PES that does not reflect experimental observables. MCSCF calculations can provide qualitatively correct chemical insights, but dynamic correlation needs to be added to obtain quantitatively correct energies. The PESs

reported here will be used to illustrate the ability to build MCSCF global surfaces and provide fundamental insights about the chemical processes of interest here.

For N + H_2 , the Pascual-Schatz[28] basic method is a FORS-MCSCF wavefunction with a full valence active space of 7 active electrons in 6 active orbitals (FORS-MCSCF(7,6)/6-311++G(d,p)). The active orbitals are the H-H sigma bonding and anti-bonding molecular orbitals (MO) and the 2s and 2p atomic orbitals on N. The N 1s orbital is kept doubly occupied. Once this FORS-MCSCF PES was generated, single point MCQDPT energies were obtained at selected geometries for improved relative energies. Following the lead of Pascual and Schatz, the PES developed in this work, for X = C, N, and O, was generated with FORS-MCSCF wavefunctions: FORS- MCSCF(6,6), FORS- MCSCF(7,6), and FORS- MCSCF(8,6) with the 6-311++G(d,p) basis set for C, N, and O, respectively. Since both C and O are ground state triplets, for consistency with the ground state N(^{4}S) analog, both C+H_2 and O+H_2 PES generation and classical trajectory studies were done on the triplet surface. Of course, for C, the triplet and singlet surfaces are very close in energy and both are chemically interesting. Since a major motivation is the new ability to build multiple potential energy surfaces, subsequent work will report on the lowest singlet surface of CH_2.

As with the Pascual-Schatz calculations, the abstraction and exchange reactions were considered for all systems. Transition states were located, and hessians were calculated for the structures (reactants, products, and TSs) to determine if they are minima (positive definite hessian) or transition states (one negative eigenvalue). Intrinsic reaction coordinate (IRC) pathways[22] were employed in order to connect reactants, transition states, and products.

Many convergence problems were encountered with the C + H_2 MCSCF calculations on the triplet surface so restricted open-shell Hartree Fock theory (ROHF/6-311++G(d,p)) was employed to define the reactants and products, locate the TSs, and calculate IRC pathways. In studying the CH_2 exchange reaction, no TS was found as the system most likely falls into the deep CH_2 well (see below). As with the C + H_2 system, a deep OH_2 well is expected for the O + H_2 reaction so a TS for the exchange reaction was not pursued.

Only select points from the IRC pathways (approximately every third IRC point) are used to define an initial surface for the Shepard interpolation procedure. These points were used to provide an energy profile for each IRC pathway (energy profiles will be given in the

following sections). For N + H_2, this set of initial data consisted of 30 points from the abstraction IRC and 30 points from the exchange IRC. The C + H_2 and O + H_2 systems consisted of 50 and 80 points, respectively, from the abstraction IRC. Finally, all electronic structure calculations were performed using the GAMESS suite of programs.[18]

Classical Trajectory Methods. There are several controls for the classical simulation portion of the PES growth process; explicit details about these controls have been described elsewhere.[14,4] Initial reference geometries for the classical trajectories are required, but different parts of a PES can be grown from different reference conditions and may give a more "well-rounded" characterization of the final surface. The PESs in this study were constructed starting from the forward and reverse IRCs; reference geometries were reactants and products, respectively. Trajectories for building a PES were also initiated at each abstraction transition state (TS).

The initial internal energies of the reference fragments and the relative translational energy of the system must also be supplied in the growth process. Values equal to approximately the zero point energy were given to each fragment in this study. These values will be presented for each reaction in the descriptions given below. The tables should be interpreted to mean that if the reactants are used as the reference, the values for the vibrational energy of H_2 will be supplied to Grow (since the atom X has no vibrational energy). If products are used as the reference, the vibrational energy of XH will be supplied, and if the TS is used as the reference geometry, the vibrational energy of the abstraction TS, H--H--X (all atoms in one fragment), will be used. The initial fragment separation was set at 5.0 Å for all studies. All trajectories start with the fragments separated by this distance, and any trajectory is discontinued if fragments move out of this range. The interpolation scheme employed q=2 and p= 6 (for N + H_2) or p=9 (for C + H_2 and O + H_2) (see Eq. 4); these values have been tested and used for other small systems.[5,3] All permutations of the molecular systems were allowed so that the molecular and PES symmetry group is the CNP group.

A velocity Verlet integrator was used to calculate classical trajectories with a time step of 0.01 fs; the permitted error in energy conservation was set at 4.0×10^{-7} hartrees. A Markov walk is used to generate a micro-canonical distribution of initial atomic positions and velocities for the classical trajectories.[33] The number of Markov chain steps per

trajectory and Markov step length were set at 500 and 0.15 Å, respectively, for all fragments. If a saddle point is used as the initial reference geometry, the Markov walk will wander downhill to products or reactants unless constrained to remain in the saddle point vicinity. The X-H bond is held fixed during the PES construction to avoid this scenario.

In order to decrease time-consuming calculations, a neighbor list scheme is employed. Inner and outer neighbor lists are defined with weight cutoffs, as well as the number of time steps between which both lists must be updated[4]. The calculations described below set the outer neighbor list to be evaluated every 10 time steps and used a cutoff of 10^{-6}. The inner neighbor list was evaluated every 5 time steps with a cutoff of 10^{-4}. These values are appropriate for the time step of 0.01 fs. The growth cycles were discontinued when the PES data file contained approximately 300-500 points in accord with the number of points needed to describe the PES with this process in many previous studies of small systems.[3,5,14] Many large scale trajectories were also run on the data sets to check if the same frequency and type (within ± 5 per type of product, depending on simulation duration) of products were attained in all cases.

C. $C + H_2$

Reactant, product, and transition state bond lengths for the triplet surface are given in Tables 1 and 2; energies and barrier heights are given in Tables 3 and 4. The initial data supplied to the PES growing process was calculated at the ROHF level of theory (as mentioned above), but in order to test the MCSCF PES growing procedure, calculations switched back to the FORS-MCSCF level of theory. This means that the MCSCF energies were calculated for each initial ROHF data point geometry before any new PES data points were calculated and that any additional data points and classical simulations were done at the MCSCF level of theory. It should be pointed out that the TS calculated at the ROHF level is not a TS at the MCSCF level of theory, but this approach (ROHF stationary point calculations and MCSCF growth/dynamics) is sensible as the interpolation process has the ability to sample configuration space and only requires a rough estimate of reaction path data. MCSCF energies are supplied for reactants and products in Table 3 only for reference. Table 2 shows that the abstraction TS has considerable CH (product) character since the C-H bond is nearly formed. This is reinforced by Table 4 in that the forward barrier to reach the

abstraction saddle point is much larger than the reverse barrier. The energy profile along the abstraction IRC pathway (Figure 2) also illustrates this feature.

The initial internal energies of the reference fragments and the relative translational energy of the system for the classical trajectories are shown in Table 5. Initial internal energies were chosen based on the zero point energy of each fragment; low translational energies were supplied to the system. If the abstraction saddle point was used as the reference geometry, the C-H bond was held fixed at 1.3 Å during the growing process.

Energy and atom distance distributions (Figure 3) are reported for a data set that contains 349 points. This data set defines the PES for this reaction. In order to get the energy distribution of data points, all energies were collected and "binned" in intervals of 0.01 au. These values were then plotted in histogram form; thus, the histogram with energy between –55.46 au and –55.45 au shows the frequency of data points with energies that fall in this range. The energy values of the reactants, products, and TSs are indicated on the graph for reference. The energy distribution of the PES data points (Figure 3a) is concentrated around the abstraction reaction, but, in sampling other parts of space, Grow has found the CH_2 well (indicated by points with much lower energies). The atom distance distribution plot is shown in Figure 3b. The H1-H2 and H1-C3 distances are calculated for each data point; data point geometries are then plotted as a function of the H1-H2 vs. H1-C3 (the system is defined as H1-H2-C3); the initial IRC data points are indicated on the plot. The data points are most likely shifted from the IRC data due to the switch from ROHF to MCSCF (MCSCF energies for the reactants and products are lower as indicated in Table 3). The energy contour plot is shown in Figure 4. Energy has been plotted as a function of the two CH bond lengths, finding the minimum in the HH bond length for each CH, CH pair. Zero is defined as the lowest energy in the PES data file, in this case, equilibrium CH_2. The deep CH_2 well is readily apparent in the lower left-hand corner of the plot. As indicated in the initial QM calculations, no TS is found in the exchange reaction; reactions directly go downhill to CH_2. The abstraction saddle point, indicated by an X, is shown in the "ridge" of higher energy separating C + H_2 and H + CH (or HC + H).

D. $N + H_2$

Reactant, product, and transition state bond lengths for both abstraction and exchange reactions are given in Tables 1 and 2; energies and barrier heights are given in Tables 3 and 4. Again, Table 2 shows that the N-H bond is almost formed by the time that the abstraction TS is reached. The barrier to go from $N + H_2$ to this abstraction TS is even larger than that for $C + H_2$ (see Table 4). The energy profiles along both IRC pathways are shown in Figure 5.

The initial internal energies of the reference fragments and the relative translational energy of the system for the classical trajectories are shown in Table 6. Larger translational energies were supplied to the system (as compared with the $C + H_2$) system. If the abstraction saddle point was used as the reference geometry, the N-H bond was held fixed at 1.2 Å during the growing process.

Energy and atom distance distributions and a PES contour are reported for a data set that contains 585 points. The energy distribution of the PES data points (Figure 6a) is mainly concentrated around the abstraction reaction TS and product, H + NH, but data points at other energies areas are also sampled. The atom distance distribution plot (Figure 6b) gives another view of the sampling of data points. Data point geometries are plotted as a function of the H1-H2 distance vs. the H1-N3 distance (the system is defined as H1-H2-N3); the initial IRC data points are indicated on the plot. While Grow mainly generates points around the IRC data (the Shepard interpolation is well-defined in this area), it also samples configurations in other regions of space as seen by the "tails" extending in both directions from the initial IRC data. The energy contour plot is shown in Figure 7. Energy has been plotted as a function of the two NH bond lengths, finding the minimum in the HH bond length for each NH, NH pair. Zero is defined as the lowest energy in the PES data file. The abstraction saddle point is located on the "curve" between the $N + H_2$ and NH + H and is indicated on the plot. There is no deep well with the $N + H_2$ reaction, as there was for $C + H_2$. The exchange saddle point is located in the horseshoe-shaped region in the lower left-hand corner. This also agrees with the equal N-H distances in the geometry of the exchange saddle point.

E. $O + H_2$

Reactant, product, and transition state bond lengths are given in Tables 1 and 2; energies and barrier heights are given in Tables 3 and 4. The abstraction reaction TS also shows O-H bond character (Table 2), and the forward barrier to the abstraction reaction is again larger than the reverse. The energy profile along the abstraction IRC is shown in Figure 8.

The initial internal energies of the reference fragments and the relative translational energy of the system for the classical trajectories are shown in Table 7. Large translational energies were supplied to the system since smaller energies gave no reaction and added no new points to the surface. If the abstraction saddle point was used as the reference geometry, the O-H bond was held fixed at 1.2 Å during the growing process.

Energy and atom distance distributions and a PES contour are reported for a data set that contains 430 points. The energy distribution of the PES data points (Figure 9a) shows that configuration space around the abstraction reaction has been sampled the most but also that the other areas have also been explored and sampling was not exclusive to the reaction path data. The atom distance distribution plot is shown in Figure 9b. Data point geometries are plotted as a function of the H1-H2 distance vs. the H1-O3 distance (the system is defined as H1-H2-O3), and the initial IRC data points are indicated on the plot. Again, sampling was focused around the reaction path, but other configurations are also included in the PES. For example, the "tail" on the right side of the graph places data points in regions not accounted for in the static IRC pathway.

The contour energy plot is shown in Figure 10. Energy has been plotted as a function of the two OH bond lengths, finding the minimum in the HH bond length for each OH, OH pair. Zero is defined as the lowest energy in the PES data file. This plot is very similar to the NH_2 system plot in that the abstraction saddle point is located on the "curve" between $O + H_2$ and OH + H. Even though no exchange reaction information was supplied to the MCSCF PES growing process, there appears to be an exchange saddle point in the horseshoe-shaped region in the lower left-hand corner.

IV. Conclusions

The new MCSCF PES growing process samples a appropriate amount of configuration space to describe the X + H_2 reactions. Even though C, N, and O are neighbors on the periodic table, the PES studies show that they have very different behaviors in terms of a global, ground state PES of the separated reactants. The triplet C + H_2 surface exhibits the deep CH_2 well, while both N + H_2 (quartet) and O + H_2 (triplet) surfaces indicate that a barrier exists in the exchange reaction. These surfaces were characterized with very few data points and QM calculations, making the process very desirable. The new MCSCF process also makes it possible to build potential energy surfaces using multi-configurational wavefunctions. This opens up a wide range of new, important applications that cannot be studied with single determinant methods. However, many reactions do not take place on a single adiabatic surface when multiple surfaces are in close proximity. The dynamics of reactions that take place on multiple electronic states are governed by multiple PESs. These descriptions are very important in regions of surface crossings and conical intersections, as well as in chemical reactions that take place in excited electronic states and proceed via multiple electronic states. The extension to treating multi-surface problems is an important tool for a vast number of scientific problems, ranging from atmospheric chemistry to photochemical reaction mechanisms in organic and inorganic chemistry to fundamental biological phenomena such as photosynthesis. The calculation of accurate PES to be used for dynamic simulations is very important to our understanding of the chemical world at a molecular level. A central aspect of chemistry is unraveling complex chemical mechanisms, including multiple surfaces and competition among the many competing reactions. Accurately representing the PES with a small amount of data, based on accurate, high level *ab initio* quantum chemistry, allows an efficient and effective approach to understanding these reactions and processes.

Acknowledgements

This work was supported by a National Science Foundation Foreign Travel Grant and a Fulbright Senior Scholar Award (to MSG). HMN was supported by a Department of Energy Computational Science Graduate Fellowship. The calculations described here were performed on both local Iowa State University/DOE Ames Laboratory workstations, as well

as the HP SC system at the Australian National University (part of the Australian Partnership for Advanced Computing).

References

1 J. N. Murrell and et al., *Molecular Potential Energy Functions*. (Wiley, Chichester, 1984).

2 A. Frishman, D. K. Hoffman, and D. J. Kouri, Journal of Chemical Physics **107**, 804 (1997).

3 M. A. Collins, Advances in Chemical Physics **93** (New Methods in Computational Quantum Mechanics), 389 (1996).

4 J. Ischtwan and M. A. Collins, Journal of Chemical Physics **100** (11), 8080 (1994).

5 M. J. T. Jordan, K. C. Thompson, and M. A. Collins, Journal of Chemical Physics **102** (14), 5647 (1995).

6 M. J. T. Jordan, K. C. Thompson, and M. A. Collins, Journal of Chemical Physics **103** (22), 9669 (1995).

7 K. C. Thompson and M. A. Collins, Journal of the Chemical Society, Faraday Transactions **93** (5), 871 (1997).

8 K. A. Nguyen, I. Rossi, and D. G. Truhlar, Journal of Chemical Physics **103**, 5522 (1995).

9 T. Ishida and G. C. Schatz, Journal of Chemical Physics **107**, 3558 (1997).

10 J. J. P. Stewart, L. P. Davis, and L. W. Burggraf, Journal of Computational Chemistry **8**, 1117 (1987).

11 T. Taketsuga and M. S. Gordon, Journal of Physical Chemistry A **99**, 8462 (1995).

12 T. N. Truong and et al., Computer Physics Communications **75**, 143 (1992).

13 M. J. T. Jordan and M. A. Collins, Journal of Chemical Physics **104** (12), 4600 (1996).

14 K. C. Thompson, M. J. T. Jordan, and M. A. Collins, Journal of Chemical Physics **108** (20), 8302 (1998).

15 R. P. A. Bettens and M. A. Collins, Journal of Chemical Physics **111** (3), 816 (1999).

16 R. P. A. Bettens and M. A. Collins, Journal of Chemical Physics **109** (22), 9728 (1998).

[17] R. P. A. Bettens and M. A. Collins, Journal of Chemical Physics **108** (6), 2424 (1998).

[18] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Journal of Computational Chemistry **14**, 1347 (1993).

[19] R. Farwig, in *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox (Clarendon, Oxford, 1987), pp. 194.

[20] P. Lancaster and K. Salkauskas, *Curve and Surface Fitting, An Introduction.* (Academic, London, 1986).

[21] M. A. Collins and K. C. Thompson, in *Chemical Group Theory: Techniques and Applications*, edited by D. Bonchev and D. H. Rouvray (Gordon and Breach, Reading, 1995), pp. 191.

[22] B. C. Garrett and et al., Journal of Physical Chemistry A **92**, 1476 (1988).

[23] M. Koshi and et al., Journal of Chemical Physics **93**, 8703 (1990).

[24] Z. F. Xu, D. C. Fang, and X. Y. Fu, Journal of Physical Chemistry A **101**, 4432 (1997).

[25] L. A. Pederson and et al., Journal of Chemical Physics **110**, 9091 (1999).

[26] K. Prasad, R. A. Yetter, and M. D. Smooke, Combust. Sci. Technol. **124**, 35 (1997).

[27] D. F. Davidson and et al., Int. J. Chem. Kinet. **22**, 513 (1990).

[28] R. Z. Pascual, G. C. Schatz, G. Lendvay, and D. Troya, Journal of Physical Chemistry A **106** (16), 4125 (2002).

[29] P. Knowles, N. C. Handy, and S. Carter, Mol. Phys. **49**, 681 (1983).

[30] L. B. Harding, R. Guadagnini, and G. C. Schatz, Journal of Physical Chemistry A **97**, 5472 (1993).

[31] B. Bussery-Honvault, P. Honvault, and J. M. Launay, Journal of Chemical Physics **115**, 10701 (2001).

[32] G. C. Schatz and et al., Journal of Chemical Physics **107** (7), 2340 (1997).

[33] H. W. Schranz, S. Nordholm, and G. Nyman, Journal of Chemical Physics **94** (2), 1487 (1991).

Table 1. Optimized bond distances (FORS-MCSCF(x,y)/6-311++G(d,p)). Values in parentheses indicate the number of active orbitals and electrons in the MCSCF calculation.

Diatom	R (Å)
H_2 singlet (2,2)	0.757
CH quartet (5,5)	1.094
NH triplet (6,5)	1.050
OH doublet (7,5)	0.975

Table 2. Geometric parameters of transition state structures (all structures are co-linear).

Transition State	R_{12}(Å)	R_{13}(Å)
Abs. C (triplet): H1-H2-C3	1.033	1.265
Abs. N (quartet): H1-H2-N3	1.164	1.157
Ex. N (quartet): H1-N2-H3	1.292	1.292
Abs. O (triplet): H1-H2-O3	0.969	1.183

*Abs. and Ex. indicate abstraction and exchange reactions.

Table 3. Energies (au) of reactants, products, and transition states (FORS-MCSCF). ROHF energies are supplied for the C + H_2 system; MCSCF values are given in parentheses for reactants and products.

Species	Energy
H_2+C	-38.817746 (-38.855563)
H+CH	-38.785135 (-38.803445)
H_2+N	-55.546407
H+NH	-55.496814
H_2+O	-75.953192
H+OH	-75.935269
Abs. C TS	-38.745659
Abs. N TS	-55.484306
Ex. N TS	-55.446333
Abs. O TS	-75.908853

*Abs. and Ex. Indicated abstraction and exchange reactions.

Table 4. Potential energy barriers and energies of reaction (kcal/mol); zero point energy corrections are included (FORS-MCSCF). ROHF energies are supplied for the C + H_2 system.

Reaction	Forward Barrier	Reverse Barrier	Reaction Energy
H_2+C⇔HC+H	41.9	23.6	18.3
H_2+N⇔HN+H	36.4	7.4	29.0
H+NH⇔HN+H	32.6	32.6	0.0
H_2+O⇔HO+H	25.4	15.5	9.9

Table 5. Reaction conditions for the C + H_2 system.

Vibrational energy of H_2	24 kJ/mol
Vibrational energy of CH	24 kJ/mol
Vibrational energy of ab. TS, H--H--C	26 kJ/mol
Relative translational energy, H_2 + C or H + CH	26 kJ/mol

Table 6. Reaction conditions for N + H_2 system.

Vibrational energy of H_2	24 kJ/mol
Vibrational energy of NH	53 kJ/mol
Vibrational energy of ab. TS, H--H--N	26 kJ/mol
Relative translational energy, H_2 + N or H + NH	158 kJ/mol

Table 7. Reaction conditions for the O + H_2 system.

Vibrational energy of H_2	24 kJ/mol
Vibrational energy of OH	53 kJ/mol
Vibrational energy of ab. TS, H--H--O	26 kJ/mol
Relative translational energy, H_2 + O or H + OH	210 kJ/mol

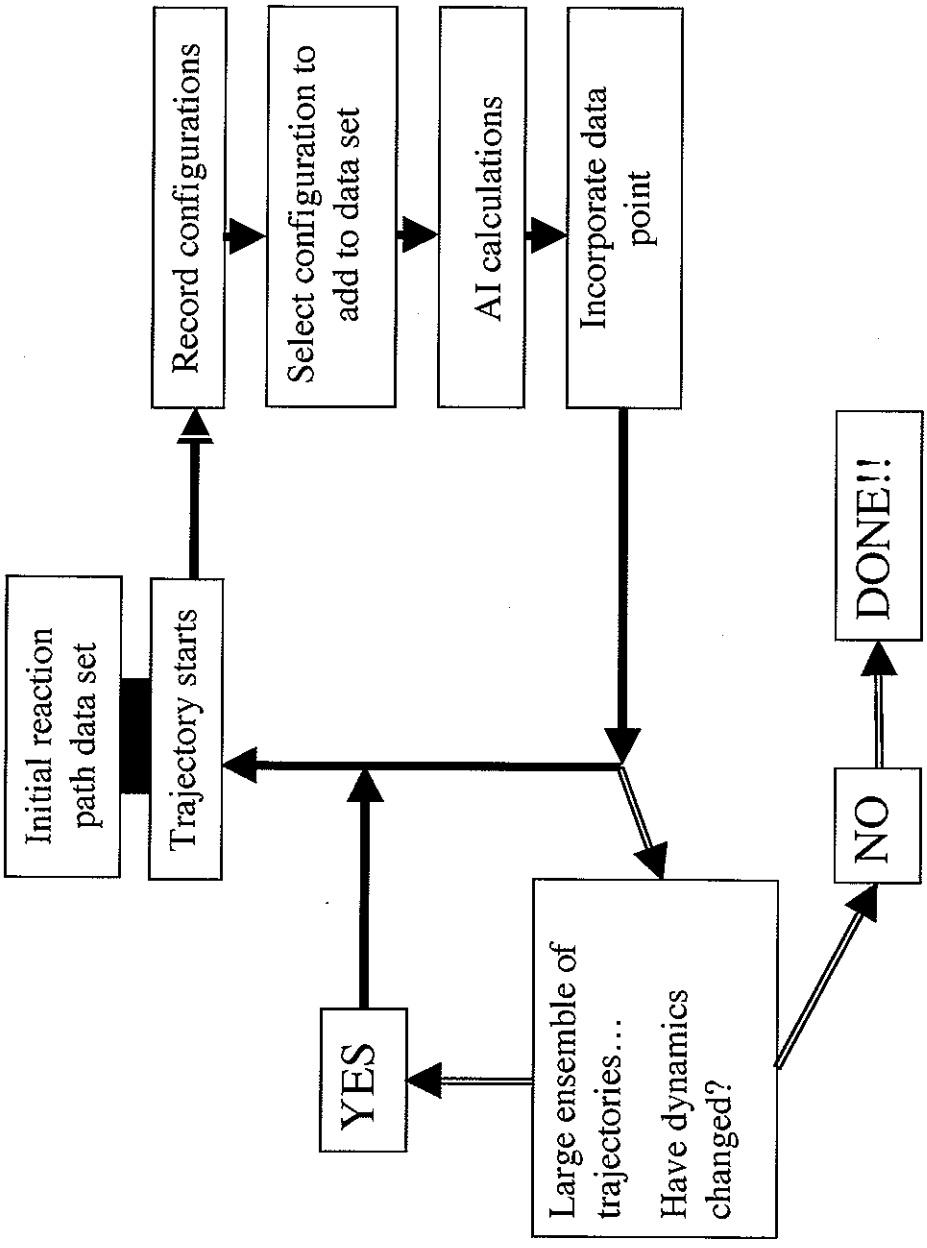


Figure 1. Schematic for the iterative procedure in Grow.

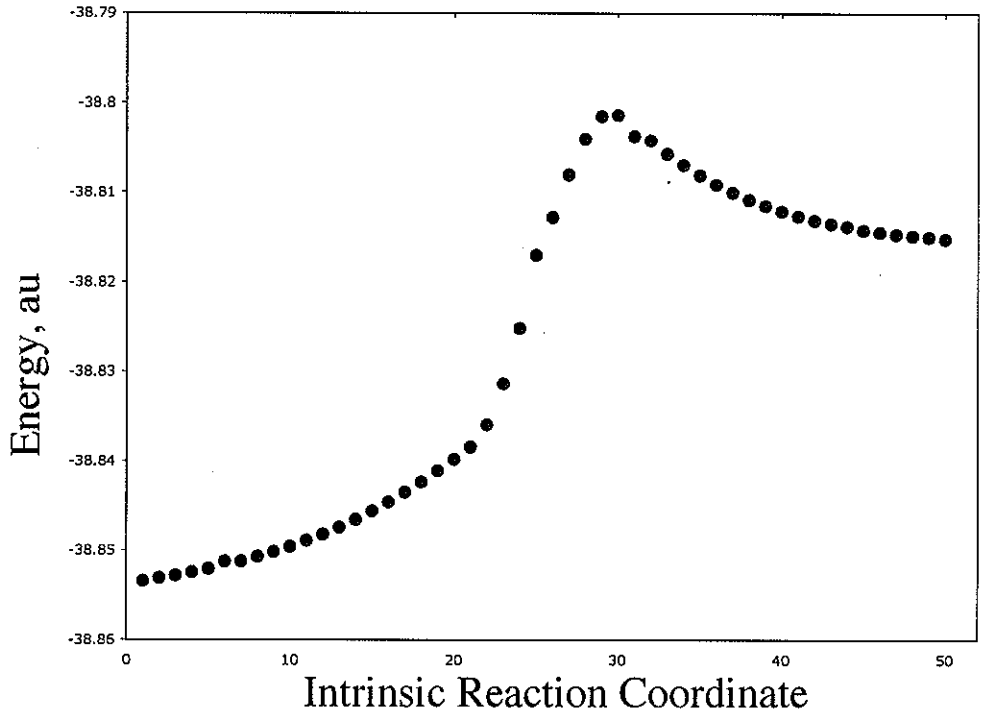
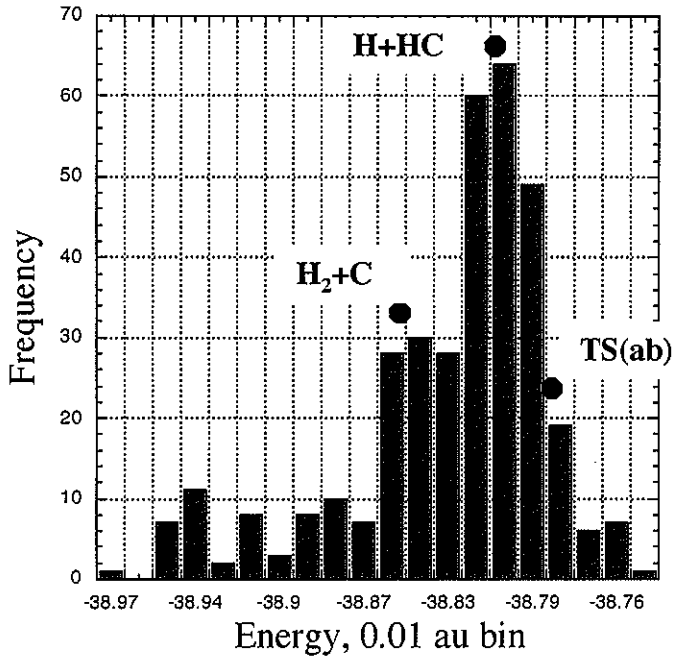
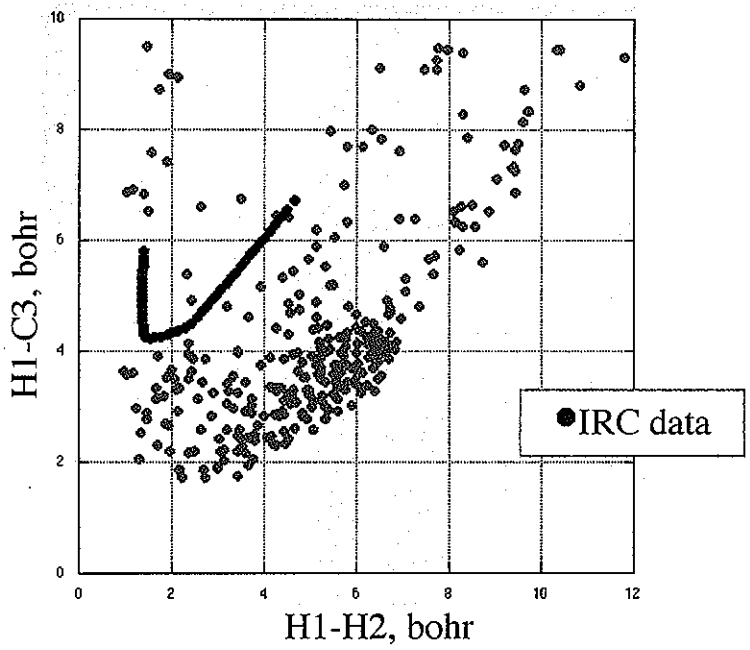


Figure 2. Energy profile along the abstraction IRC for C + H_2 (the IRC pathway was calculated at the ROHF level, but MCSCF single point energy calculations were done at the initial points for the PES growing process; the MCSCF energies are shown here).



3a.



3b.

Figure 3. C + H_2: (a) Energy distribution (dots represent energies of reactants, products, and TS; this shows where these energies would be placed in the histogram). (b) Atom distance distribution (dots indicate initial IRC data).

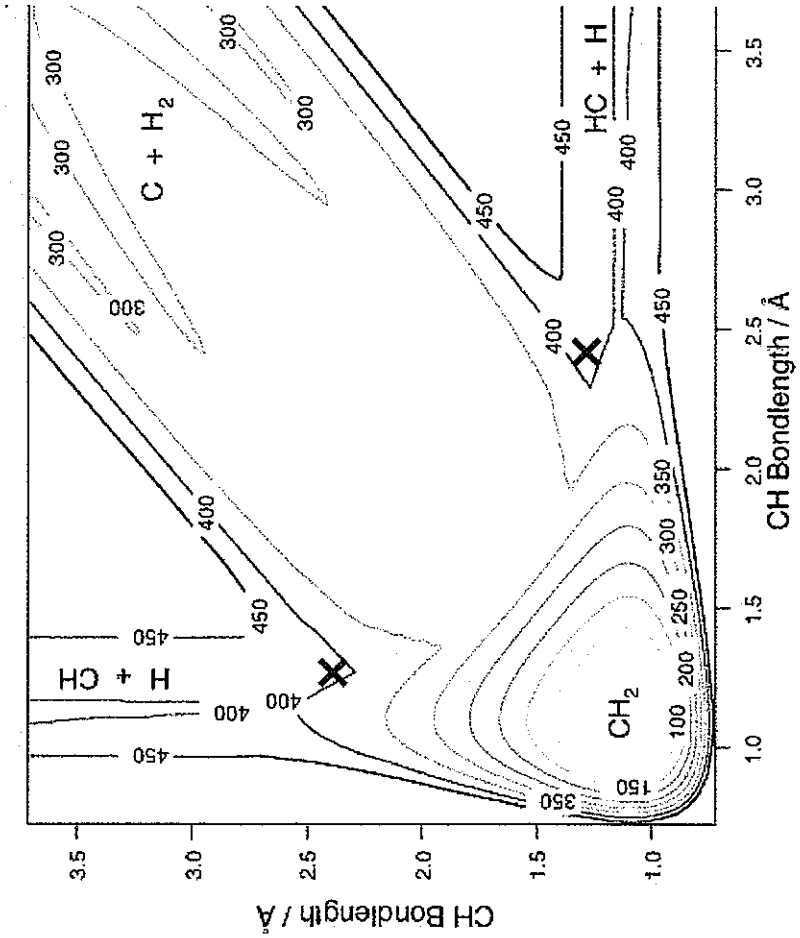
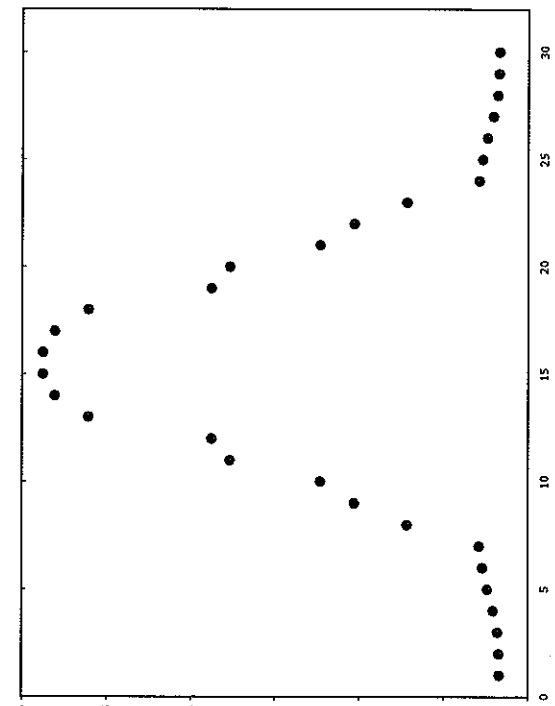
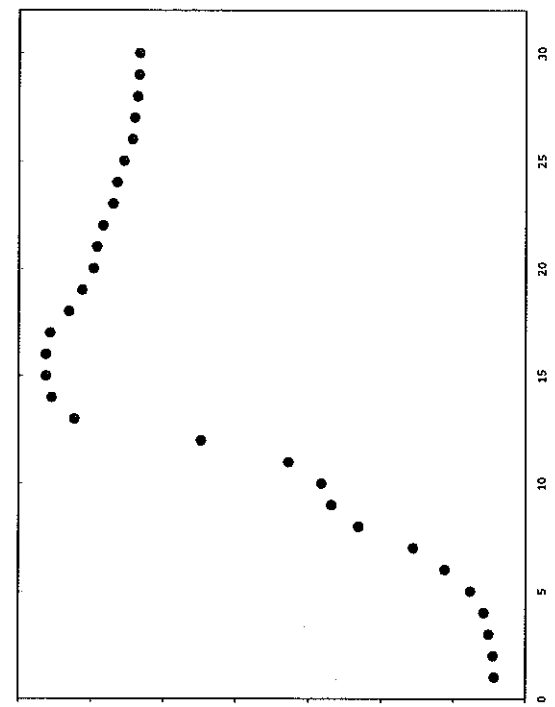


Figure 4. Energy (kJ/mol) contour for the C + H_2 PES (zero is defined as the lowest energy in the PES data file, equilibrium CH_2). The abstraction TS is indicated with an 'X'.

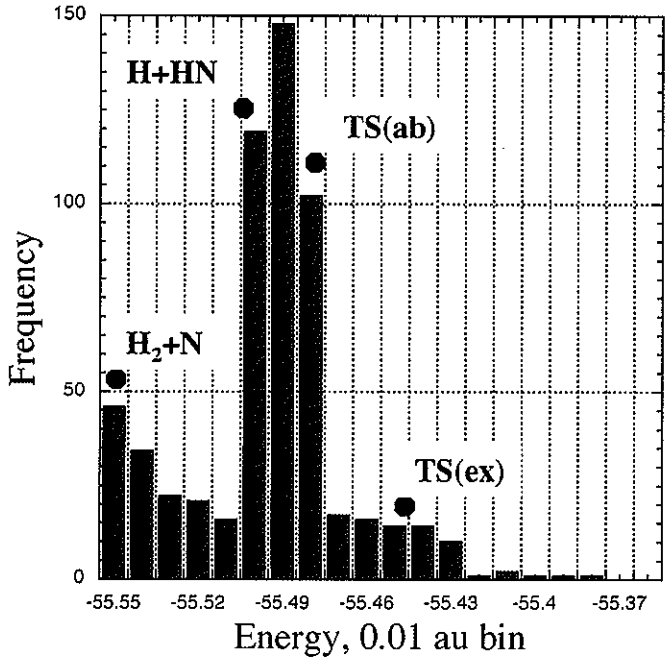


5b.

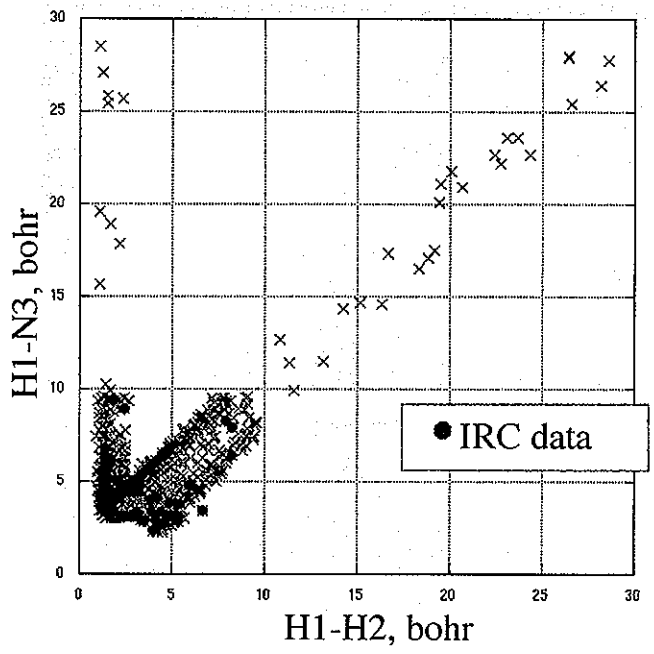


5a.

Figure 5. Energy profile along the (a) abstraction and (b) exchange IRCs for the N + H_2 system.



6a.



6b.

Figure 6. N + H_2: (a) Energy distribution (dots represent energies of reactants, products, and TSs; this shows where these energies would be placed in the histogram). (b) Atom distance distribution (dots indicate initial IRC data).

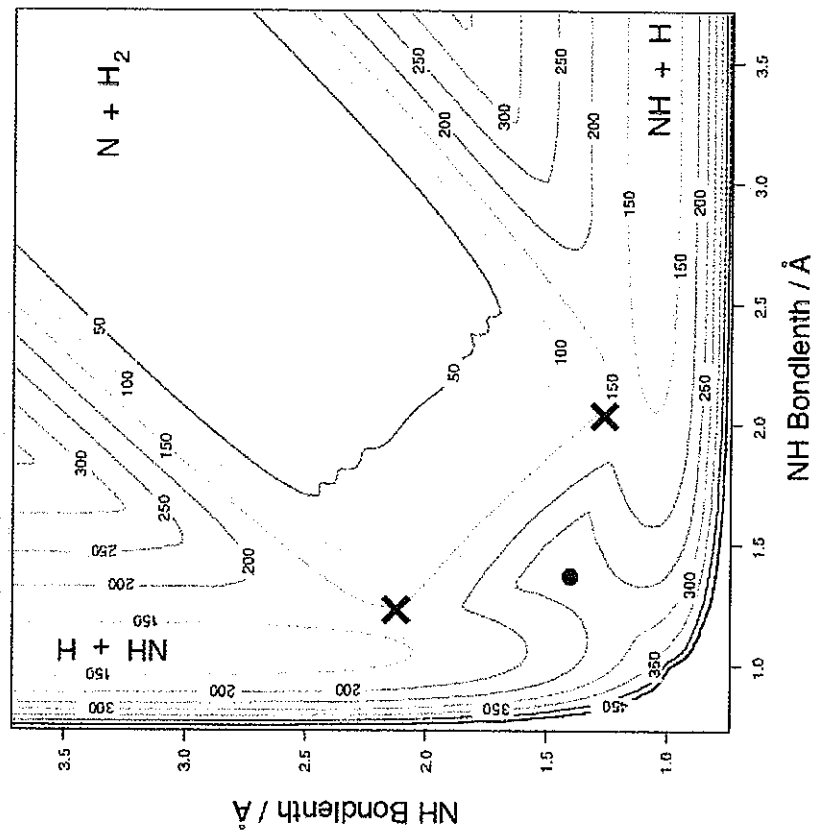


Figure 7. Energy (kJ/mol) contour for the N + H_2 PES (zero is defined as the lowest energy in the PES data file). The abstraction and exchange TSs are indicated with an 'X' and dot, respectively.

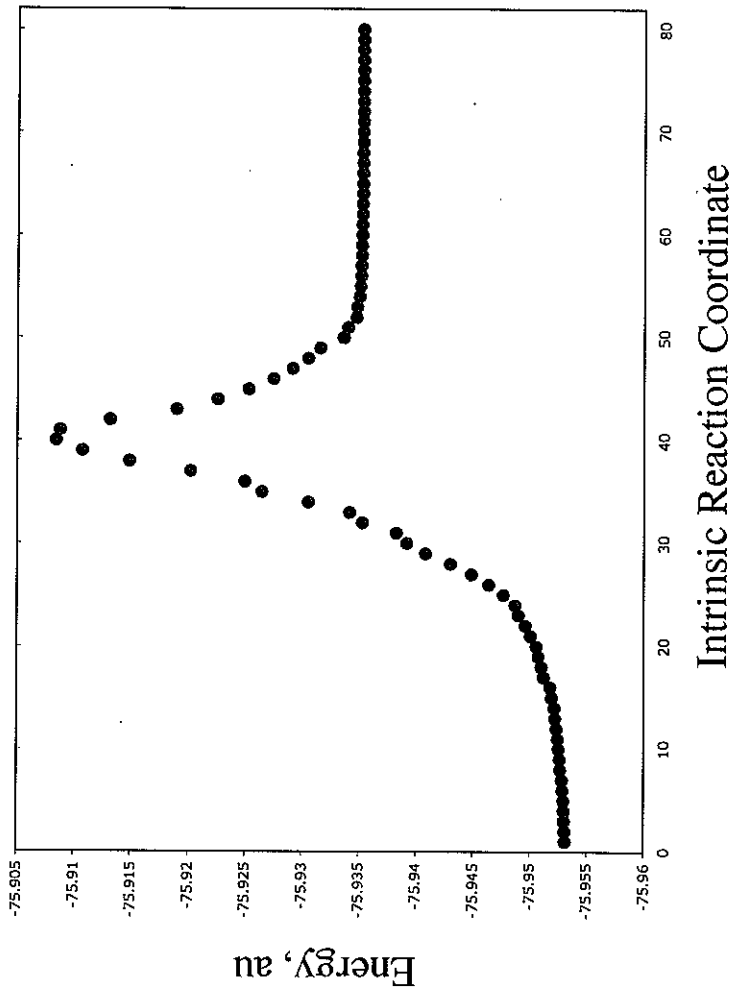
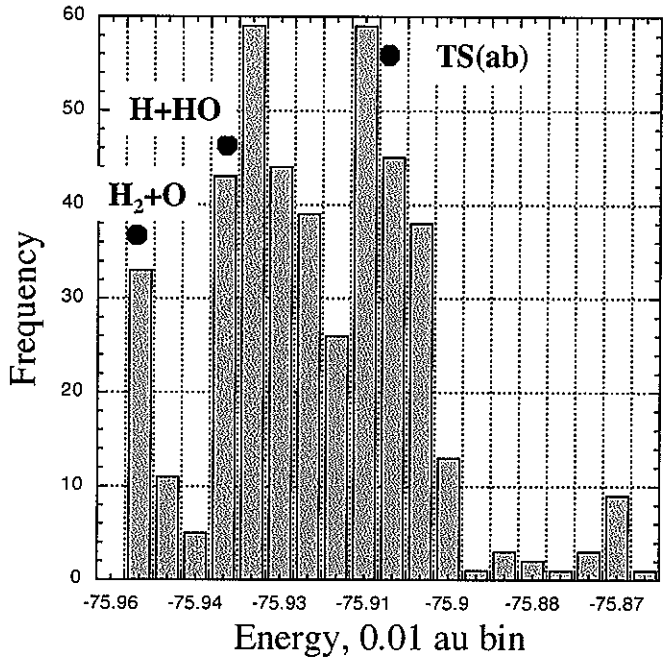
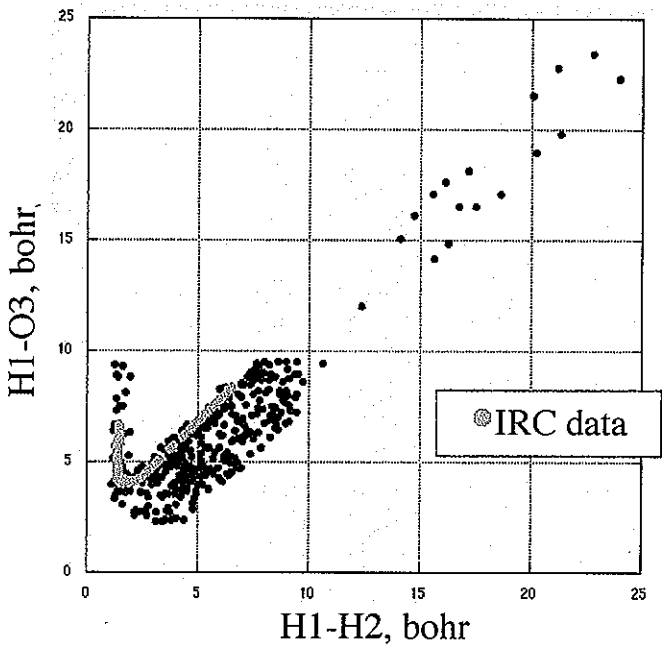


Figure 8. Energy profile along the abstraction IRC for O + H_2.



9a.



9b.

Figure 9. O + H_2: (a) Energy distribution (dots represent energies of reactants, products, and TS; this shows where these energies would be placed in the histogram). (b) Atom distance distribution (dots indicate initial IRC data).

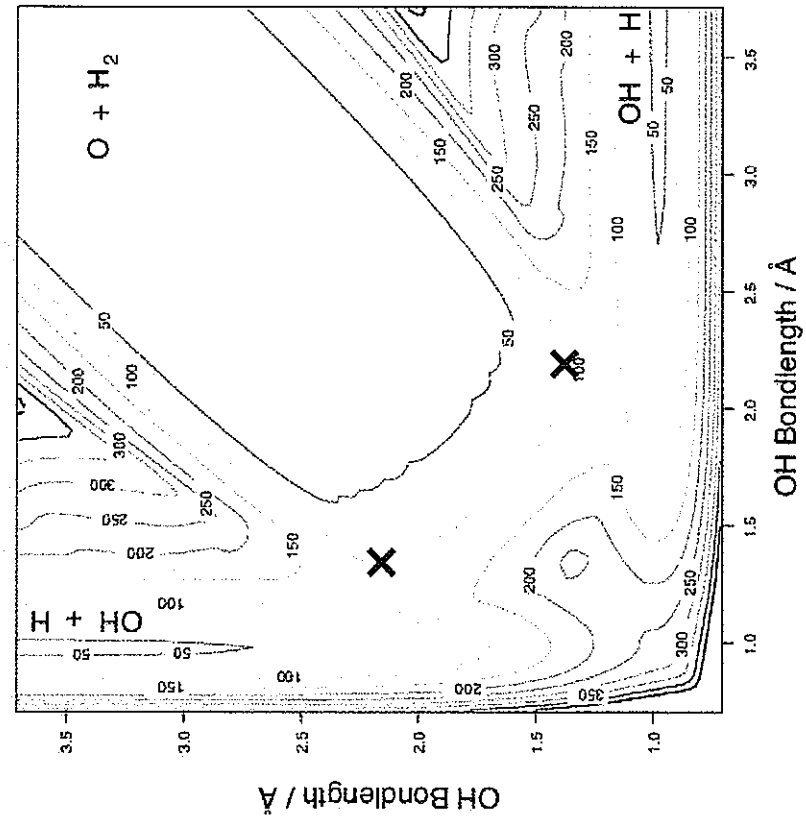


Figure 10. Energy (kJ/mol) contour for the O + H_2 PES (zero is defined as the lowest energy in the PES data file). The abstraction TS is indicated with an 'X'.

CHAPTER 6: GENERAL CONCLUSIONS

When using dynamics methods, careful consideration of the system under study, as well as the type of information desired in performing the dynamics, is vital. Chapter 2 showed that high-level, "on-the-fly" dynamics, along with intrinsic reaction coordinate calculations of the static PES, can provide important evidence about the behavior of small polynitrogen systems. Coupled with reaction dynamics theories, stability and lifetime predictions can be made and confirmed by experiments. Chapter 3 showed that in order to use high-level methods for solvation, in treating large systems both statically and dynamically, the Effective Fragment Potential (EFP) used to describe solvation must be computationally efficient, i.e. parallel. Based on quantum mechanics (QM), the EFP method has been shown to reproduce corresponding pure QM results for small clusters. Chapter 4 showed the use of molecular dynamic simulations with the EFP method in predicting bulk behavior. Finally, Chapter 5 showed the interface of the Grow and GAMESS programs to provide an *ab initio*-based molecular potential energy surface (PES). A multi-reference capability was also included in the interface; this provides the opportunity to study a wide array of problems. The developed PES has the potential to be utilized for classical or quantum dynamics.

The EFP molecular dynamics simulations have several levels of expansion. First, testing and developing other bulk properties of the EFP method is of interest. Both thermodynamic (for example, heat capacity, density, and compressibility) and kinetic (for example, diffusion coefficient and time-correlation functions) properties can be evaluated within the scope of molecular dynamics. These properties are important to test the ability of the EFP method to predict experimental bulk behavior. Next, the development of EFP2 molecular dynamics, for any type of solvent, is very desirable. Applications such as anion solvation and the treatment of supercritical properties would benefit from these further developments.

The use of Grow to calculate multi-reference PESs in itself opens up a wide range of new, important applications that cannot be studied with single determinant methods. However, many reactions do not take place on a single adiabatic surface when multiple surfaces are in close proximity. The dynamics of reactions that take place on multiple

electronic states are governed by multiple PESs and "nonadiabatic coupling surfaces" that allow molecules to move from one PES to another. The further development of GROW to provide the global potential energy surfaces and coupling terms that arise when a chemical reaction involves multiple electronic states would have a broad range of applications, from to atmospheric chemistry to fundamental biological phenomena such as photosynthesis.

Overall conclusions…

Life is not a black box: users have the responsibility to have some level of knowledge of the chemical system and type of calculation.

Life is not static: there is drama in dynamics.

ACKNOWLEDGMENTS

The United States Government has assigned the DOE Report number IS-T 1930 to this thesis. Notice: This document has been authored by the Iowa State University of Science and Technology under Contract No. W-7405-ENG-82 with the U.S. Department of Energy. The U. S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this document, or allow others to do so, for U. S. Government purposes.

I would first like to thank my family and friends for their support, encouragement, and prayers during this graduate school "journey." Most importantly, I would like to thank my parents, Charles and Roberta, and brother, Michael, for their patience and love but most of all for encouraging me to be all that I can be, yet not pressuring me to be anything other than myself.

I would also like to thank my undergraduate research advisor, Dr. David Gano. I would never have dreamed of research in chemistry, much less graduate school, without his guidance.

Many chemistry colleagues (and friends) have come across my path during this time, sharing insights on chemistry, mathematics, physics, computer science, and life in general. They are too numerous to name (and I do not want to leave anyone out!); I thank them all, especially my "buddy", Ivana Adamovic.

This page would not be complete without thanking my graduate school advisor, Dr. Mark Gordon. I do not have the words to express my appreciation and respect. The glass is half full, inanimate objects cannot possess anything, and North Dakota is unfit for human habitation (what??) will always stick in my mind.

Finally, none of this would have been possible without Christ. This journey has been possible through HIM alone...

"I can do all things through CHRIST who strengthens me."—Philippians 4:13