2008

Modeling ionic liquids, diffusion on surfaces and catalysis: a graduate student looks high and low

Deborah Zorn
Iowa State University

Modeling ionic liquids, diffusion on surfaces and catalysis: A graduate student looks high and low.

by

Deborah Zorn

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 Gordon, Major Professor
Jim Evans
Monica Lamm
Jacob Petrich
Patricia Thiel

Iowa State University

Ames, Iowa

2008

DEDICATION

This thesis is dedicated to my father, Professor Thomas S. Zorn.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

CHAPTER 1. GENERAL INTRODUCTION 1
General Overview 1
Dissertation Organization 1
Theoretical Background 1
References 11

CHAPTER 2. ELECTRONIC STRUCTURE STUDIES OF TETRAZOLIUM BASED IONIC LIQUIDS 16
Abstract 16
Introduction
Computational Methods 20
Results and Discussion 20
Conclusion 28
Acknowledgements 28
References 29

CHAPTER 3. BINDING AND DIFFUSION OF Al ADATOMS AND DIMERS ON THE Si(100)-2x1 RECONSTRUCTED SURFACE: A HYBRID QM/MM EMBEDDED CLUSTER STUDY 52
Abstract 52
Introduction 52
Computational Methods 56
Results and Discussion 57
Conclusions 68
Acknowledgements 71
References 71

CHAPTER 4. COMPARISON OF NITROALDOL REACTION MECHANISMS USING *AB INITIO* CALCULATIONS 94
Abstract 94
Introduction 94
Computational Methods 97
Results and Discussion 98
Conclusions 111
Acknowledgements 112
References 113

CHAPTER 5. INTERACTION OF THE UNIVERSAL FORCE FIELD WITH THE EFFECTIVE FRAGMENT POTENTIAL METHOD 140
Abstract 140
Introduction 140

Interaction terms 149
Energy Gradients 151
EFP-MM Test Calculations 152
Conclusions 157
Acknowledgements 157
References 158

ACKNOWLEDGEMENTS

There are so many people that have supported and encouraged, through all the ups and downs, on my path the completion of this thesis. Listed here are the people who have made this dissertation possible.

First and foremost I would like to thank my advisor, Mark Gordon. Your patience reading draft after draft of these chapters was absolutely incredible. I shall be forever grateful for all I have learned over the past 5+ years and all the opportunities that I have received. I had no idea when I started my graduate career my journey would take me to "the big apple", Europe and even to the other side of the world. I am grateful to my committee members for their guidance through the last five years. I especially enjoyed my interactions with Monica Lamm who mentored me through the preparing future faculty program.

When you are in the Gordon Group, you aren't just a member of a research group, you are a member of a family: To sensei Mike Schmidt – I don't think I ever stepped into your office with out learning something new. To Bosa – I wouldn't have wanted to go through this with anyone else. Your friendship has been invaluable to me. Thanks to all the current and past Gordon group members of the last six years, especially Jamie Rintelman, Christine Aikens, and Heather Netzloff for all of their help in the beginning of my graduate career and their continuing friendship.

While in grad school I have made some of the best friends in the world – To Amie, Jen, Mike, and Brian – you guys were always there to help me forget about bugs, convergence problems, and the difficulties of my long distance relationship.

Although I sometimes have trouble remembering a time before graduate school, there are many people who made my path here possible. I need to especially thank Debbie Sedlacek my "Barn Mom", Mrs. Ep, my high school advanced chemistry teacher, and Dr. Hutchins my undergraduate advisor. I also need to thank Mr. McDaniel, my "unofficial" advisor, for giving me the crazy idea of double majoring in Math and Chemistry.

Lastly, I need to thank the four people who have made all of this worthwhile: I would like to thank my parents Tom and Jo Zorn for their unwavering love, support, and understanding. Mom and Dad - Thank you for the warm, loving, and protected home that I had growing up and the sacrifices you made to make that possible. Thank you for always encouraging me to learn new things and to always try to reach my potential. To my sister Michelle - growing up with you was a great adventure and an absolute blast. I know I am the person I am today because of you. To Jeff, my best friend and the love of my life – What a crazy journey we have been on together, but I know that this was the path that we were both meant to be on, because it brought us together. I can't wait to spend the rest of my life with you.

CHAPTER 1. GENERAL INTRODUCTION

General Overview

Methods currently available to chemists in the field of quantum chemistry make it possible to treat chemical reactions in the gas phase with a high level of accuracy. More challenging is the treatment of reactions taking place in solvents or on surfaces as in the fields of surface chemistry and heterogeneous catalysis. A deep understanding of the chemical processes that take place in these systems is essential for the development of new materials, nano-devices, and reusable catalysts.

Dissertation Organization

The chapters presented in this thesis are as follows: Chapter two covers Electronic Structure Studies of Tetrazolium-Based Ionic Liquids. Chapters three describes a Hybrid QM/MM embedded cluster study of the binding and diffusitno of Al adatoms and dimers on the Si(100)-2x1 Reconstructed Surface. Chapter four is a comparison of competing mechanisms for the nitroaldol reaction using high level *ab initio* calculations. Chapter five illustrates the interface of the effective fragment potential method with molecular mechanics.

Theoretical Background

Quantum Chemistry Methods

A quantum chemistry method is said to be "*ab initio*" if it relies on the basic laws of quantum mechanics without reference to fitted data. Most *ab initio* methods in

computational chemistry have been derived in an attempt to accurately solve the time independent Schrödinger equation:

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

where **H** is the Hamiltonian operator, Ψ is the wave function, and E is the total energy of the system. The total Hamiltonian operator is often written as

$$\boldsymbol{H} = \boldsymbol{T}_n + \boldsymbol{T}_e + \boldsymbol{V}_{ne} + \boldsymbol{V}_{ee} + \boldsymbol{V}_{nn} \tag{2}$$

T and V are the kinetic and potential energy operators and the subscripts n and e refer to the nuclear and electronic coordinates respectively. The Born-Oppenheimer (BO) approximation[1] fixes the nuclei in space and reduces the total Hamiltonian operator to the electronic Hamiltonian:

$$\boldsymbol{H}_e = \boldsymbol{T}_e + \boldsymbol{V}_{ne} + \boldsymbol{V}_{ee} + \boldsymbol{V}_{nn} \tag{3}$$

The electronic Schrödinger equation can be solved self consistently by using Hartree-Fock theory, in which each wave function is represented by a single Slater determinant of one-electron orbitals and each electron interacts with the field (Coulomb and exchange) of all other electrons. When the orbitals are restricted to be doubly occupied we have closed shell restricted Hartree-Fock (RHF), with all electrons paired. There are two general approaches for dealing with open shell systems, restricted Open-shell Hartree-Fock (ROHF) and Unrestricted Hartree-Fock (UHF). In ROHF all orbitals are restricted to be doubly occupied or singly occupied. UHF uses different orbitals for different spins. Unfortunately with this method the wave function is not an eigenfunction of the S^2 operator. As a consequence, the

expectation value of S^2 is incorrect. The difference between the correct and approximate values of S^2 is often referred to as spin contamination.

Hartree-Fock theory does not take into account electron correlation – the instantaneous correlation of the motions of two electrons. There are several ways that the electron correlation can be recovered. Methods that attempt to improve on Hartree-Fock theory are called post-Hartree-Fock methods.

In configuration interaction (CI) Slater determinants representing excitations of one (single), two (double), etc., electrons from the reference state are added to the HF reference wave function:

$$\Psi = \Psi_{HF} + \Psi_{Single} + \Psi_{Double} + \Psi_{Triple} + \ldots \quad (4)$$

If all possible Slater determinants are included, then this is the full CI wave function. This is the exact wave function for a given basis set (Löwdin 1955). A complete CI is impractical for all systems except for very small molecules with modest basis sets. Often the CI wave function is truncated after double excitations giving CI singles and doubles (CISD).

A computationally less expensive method for including electron correlation is perturbation theory. In this method a small perturbation ($\mathbf{V}'$) is added to the reference Hamiltonian ($\mathbf{H}_0$) giving an approximation to the exact Hamiltonian ($\mathbf{H}$). In second order Møller-Plesset Perturbation Theory (MP2)[2, 3] the reference wave function is the Hartree-Fock wave function and the perturbation is defined as

$$\mathbf{V}' \equiv \mathbf{H} - \mathbf{H}_0 \quad (5)$$

The advantage of MP2 is that it is computationally economical for moderately sized systems. Unfortunately MP perturbation methods are not variational and depending on the system, higher orders of MP do not always converge.

Another very popular Post-Hartree-Fock method is Coupled Cluster theory. The coupled cluster with singles, doubles and perturbative triples CCSD(T) method[4, 5] is often called "the gold standard of quantum chemistry."

In some cases, such as for near degeneracies, a single determinant is not a valid representation of the reference wave function. Hence, Hartree-Fock theory and methods that use a Hartree-Fock reference wave function are not valid. In these cases one must employ multi-configuration methods, where the zeroth order starting wave function is described by a linear combination of symmetry adapted Slater determinants. In the fully optimized reaction space multi-configuration self consistent field (FORS-MCSCF) method[6], the orbitals are separated into two regions: the active space and inactive space. The orbitals in the inactive space must be either doubly occupied or empty. Within the active space a full CI is performed in which both the CI and orbital coefficients are optimized. A proper active space should include all the "chemically" significant orbitals. There is no foolproof algorithm for choosing active orbitals and finding the correct active space may take some trial and error. A very nice exposition on this topic is [M.W. Schmidt and M.S. Gordon, "The Construction and Interpretation of MCSCF Wavefunctions", Ann. Rev. Phys. Chem. (INVITED), 49, 233 (1998)]. Once the proper reference wave function has been constructed, the rest of the electron correlation can be properly accounted for using either CI or perturbation theory. There are many options available to do this including multi-reference CI (MRCI) and several

different types of Multi-reference Perturbation Theory (MRPT). The type of MRPT used in this work is MRMP2[7].

A completely different approach to electronic structure theory is that taken in Density Functional Theory (DFT). In DFT the total energy of a molecule is written as a functional of the electron density: E[ρ(x,y,z)]. In quantum chemistry the Kohn Sham formalism is most commonly used to obtain the electron density[8].

The challenge in utilizing density functional theory in quantum chemistry is the development of accurate functionals for exchange and correlation. One popular functional is Becke's three parameter Lee-Yang-Parr (B3LYP) hybrid functional[10-12]. B3LYP is a complicated functional in which the correlation part is taken from Vosko, Wilk and Nusair[9] and the Lee-Yan-Parr functional[11], and the exchange part is taken from the Becke exchange functional with varying amounts (determined by three parameters) of exact exchange from Hartree-Fock added in. The advantage of using B3LYP over other post-Hartree-Fock methods is that is scales on the order of N^4, where N is a measure of the size of the system, rather than N^5 or worse.

Force Field Methods

Force field (FF) methods use classical mechanics to model covalent and non-covalent interactions in a molecule. In force field methods, molecules are modeled essentially as atoms with partial charges held together by springs. This is partially justified by the Born-Oppenheimer approximation, which separates nuclear motion from electronic motion. Each

force field is defined by its functional form and by the parameters it defines for each atom type. The total energy of a molecule represented by a force field can be written as

$$E_{FF} = E_{bond} + E_{angle} + E_{torsion} + E_{electrostatic} + E_{van\ der\ Waals} + E_{other} \qquad (6)$$

The first three terms represent the covalent interactions, the fourth and fifth terms represent the non-covalent interactions and the last term includes other bonded interactions and any cross terms.

The simplest functional form for the bond stretching and angle bending terms is a harmonic oscillator. E_{bond} and E_{angle} can be written as

$$E_{bond}(r) = k_b(r - r_{eq})^2 \qquad (7)$$

and

$$E_{angle}(\theta) = k_\theta(\theta - \theta_{eq})^2 \qquad (8)$$

respectively. In these equations, r is the current bond length, r_{eq} is the equilibrium bond length, k_b is the bond stretching force constant, θ is the current bond angle, θ_{eq} is the equilibrium bond angle, and k_θ is the angle bending force constant. The functional form for a torsion is represented by a Fourier series:

$$E_{torsion}(\phi) = k_\phi \sum_{n=0}^{m} \cos(n\phi) \qquad (9)$$

where ϕ is the torsion angle and k_ϕ is the torsional rotational force constant. Long range nonbonded interactions must also be taken into account. The partial charges Q_i on each atom center are allowed to interact using a Coulombic potential:

$$E_{electrostatics}(R_{ij}) = \frac{Q_i Q_j}{\varepsilon R_{ij}} \tag{10}$$

where Q_i is the partial charge on each atom, R_{ij} is the current distance between the two atoms, and ε is the dielectric constant.

The Van der Waals energy is represented by a Lennard-Jones 6-12 expansion:

$$E_{van\ der\ Walls}(R_{ij}) = \varepsilon\left\{-2\left[\frac{C_0}{R_{ij}}\right]^6 + \left[\frac{C_0}{R_{ij}}\right]^{12}\right\} \tag{11}$$

C_0 is the minimum Van der Waals distance and ε is the well depth parameter.

The advantage of using force field methods is their computational speed, allowing the treatment of large organic and biological systems. A force field that includes noncovalent interactions will typically scale as the square of the number of atoms. Force Field parameters are fitted to experimental or *ab initio* data from a test set of molecules. The parameters usually include equilibrium values and force constants. The force constants are second derivatives of the energy evaluated at the equilibrium values. Unfortunately there is a serious lack of parameters for many chemical systems. Another obvious disadvantage of force field methods is that they do not take into account the electronic nature of a system, making it impossible to study bond-breaking reactions. Some popular force fields include MM3[12], CHARMM[13], the Amber force fields[14] and the Universal Force Field (UFF)[15].

Solvent Methods – EFP

There are two main types of solvent models. One type of method represents the solvent as a continuum and in the other type a solvent shell is created around the solute with explicit solvent molecules. The effective fragment potential method (EFP)[16, 17] is an explicit solvation model, which represents the important non-bonded interactions of solvent molecules with each other and with a solute: Coulomb interactions, induction, exchange repulsion, charge transfer, dispersion, and higher order terms. In EFP, the system is divided into two regions: the quantum (solute) region and the solvent region. The total energy of the system is

$$E_{interaction} = E_{QM-EFP} + E_{EFP-EFP} \tag{12}$$

The interaction energy includes the interactions between the quantum and EFP regions and the interactions between the solvent molecules with other solvent molecules.

The EFP method for water is called EFP1 and has been implemented for three levels of theory: HF, DFT, and MP2. In EFP1 the energy is a sum of three terms: Coulomb, Polarization and a fitted remainder term, which accounts for all interactions not included in the first two terms:

$$E_{Interaction} = E_{Coulomb} + E_{Polarization} + E_{Remainder} \tag{13}$$

In both EFP methods, the electrostatic potential of a molecule is represented by a distributed multipolar analysis (DMA) up to octopoles. The expansion points are the atom centers and bond midpoints. A damping term is used to account for overlapping charge densities and a distance cutoff is used. The polarization of each molecule by the surrounding molecules is treated self consistently using localized molecular orbitals (LMO's).

The general EFP method (EFP2) has no fitted parameters and its energy has the form

$$E_{Interaction} = E_{Coulomb} + E_{Polarization} + E_{Exchange-Repulsion} + E_{Charge-Transfer} + E_{Dispersion} \qquad (14)$$

Because EFP2 has no fitted parameters it is possible to use it with any closed shell molecule.

Hybrid Methods - SIMOMM

An efficient way to model large chemical systems is to use a hybrid method which employs a combination of QM and MM levels of theory. The Surface Integrated Molecular Orbital Molecular Mechanics (SIMOMM) method[18] is a QM/MM method specifically designed for reactions on surfaces. In SIMOMM, the surface is divided into three parts, the bulk model, the reactive site model (RSM), and the "boundary" region. The bulk model is a large cluster model of the reactive surface. The bulk model must be several lattice positions deep in order to properly account for the subsurface displacements from surface rearrangements. Carved out of the bulk model is a smaller cluster where the "action" takes place. This "action region" is called the reactive site model (RSM). The RSM region is treated with some level of quantum mechanics, while the "bulk" region is treated with molecular mechanics. Simply using the small cluster without the surrounding bulk cluster often does not properly represent the bulk behavior of the surface. The RSM and the bulk model are linked in the boundary region. When the RSM is carved out of the bulk model, bonds are broken between the two regions. In the "action" region these bonds are most often capped with hydrogens.

The SIMOMM energy calculation is conceptually very simple. In the first step, the MM energy of the bulk model is calculated with the "action" region interactions zeroed out

to avoid double counting. Next the energy of the RSM, capped with hydrogens, is calculated with an appropriate level of quantum mechanics. Long-range interactions between the RSM and the bulk region are still accounted for with MM. The QM and MM energies are added together to give the total QM/MM energy. This energy only has meaning when comparing it to other QM/MM energies of the system.

During an optimization, the system steps along a potential energy surface according to a hybrid QM-MM gradient. The total gradient with respect to the coordinates of the three regions is:

$$\frac{dE_{Total}}{d\vec{R}_{action}} = \frac{\partial E_{QM}}{d\vec{R}_{action}} + \frac{\partial E_{MM}}{d\vec{R}_{action}} \tag{15}$$

$$\frac{dE_{Total}}{d\vec{R}_{boundary}} = \frac{\partial E_{QM}}{d\vec{R}_{boundary}} \tag{16}$$

$$\frac{dE_{Total}}{d\vec{R}_{bulk}} = \frac{\partial E_{MM}}{d\vec{R}_{bulk}} \tag{17}$$

Hence both regions are fully optimized.

All methods described in this chapter are available in the electronic structure system GAMESS[19,20].

References

(1) Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys. (Leipzig)* **1927**, *84*, 457.

(2) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(3) Fletcher, G. D.; Rendell, A. P.; Sherwood, P. *Mol. Phys.* **1997**, *91*, 431.

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

(5) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2000**, *149*, 71.

(6) Reudenberg, K.; Schmidt, M.W.; Gilbert, M.M.; Elbert, S. T. *Chemical Physics* **1982**, *71*, 41.

(7) Hirao, K. *Chemical Physics Letters* **1992**, *190*, 374.

(8) Hohenberg, P.; Kohn, W. *Phys. Rev.*, **1964**, *136*, B864

(9) Vosko, S. J.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(10) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(11) Lee, C.; Yang, W.; Parr, R.G. Phys. Rev. B **1988**, *37*, 785.

(12) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(13) Brooks, B. R.; Bruccoleri R. E.; Olafson; B. D. States, D. J. Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.

(14) Weiner, S.J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765

(15) A. K. Rappe; C. J. Casewit; K. S. Colwell; W. A. Goddard III; W. M. Skiff *J. Am. Chem. Soc.* **1992,** *114* 10024.

(16) Gordon, M. S.; Freitag, M.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.

(17) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Kraus, M.;

Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

(18) Shoemaker, J.R.; Burggraf, L.W.; Gordon, M.S.; *J. Phys. Chem. A,* **1999,** *103*, 3245.

(19) Schmidt, M.W., et al., *Journal of Computational Chemistry* **1993**, *14*, 1347.

(20) Schmidt, M. W.; Gordon, M. S. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, **2005**.

CHAPTER 2: Electronic Structure Studies of Tetrazolium Based Ionic Liquids

[1]A paper published in the *Journal of Physical Chemistry B*

Deborah D. Zorn, Jerry A. Boatz and Mark S. Gordon

Abstract. New energetic ionic liquids are investigated as potential high energy density materials. Ionic liquids are composed of large, charge-diffuse cations, coupled with various (usually oxygen containing) anions. In this work, calculations have been performed on the tetrazolium cation with a variety of substituents. Density functional theory (DFT) with the B3LYP functional, using the 6-311G(d,p) basis set was used to optimize geometries. Improved treatment of dynamic electron correlation was obtained using second order perturbation theory (MP2). Heats of formation of the cation with different substituent groups were calculated using isodesmic reactions and Gaussian-2 calculations on the reactants. The cation was paired with oxygen rich anions ClO_4^-, NO_3^-, or $N(NO_2)_2^-$ and those structures were optimized using both DFT and MP2. The reaction pathway for proton transfer from the cation to the anion was investigated.

Introduction

There is considerable current interest in ionic liquids as solvents. No other class of solvents offers the versatility that ionic liquids do: They are typically thermally stable, have negligibly low vapor pressure, high density, and large liquid ranges up to 400°C. Because of

[1] Reproduced with permission from *J. Phys. Chem. B*, 2006, **110**, 11110. Copyright 1996 American Chemical Society.

these unique properties, ionic liquids may be used as electrolytes for batteries, extraction media, and catalyst carriers. The popularity of ionic liquids has also been spurred on by their classification as 'green,' due to their negligible vapor pressure, thereby decreasing levels of volatile organic carbons in the environment.[1]

1-Butyl-3-methyl-imidazolium (BMIM) is one of the most widely used and best understood ionic liquid cations, so when beginning to design energetic ionic liquids, this cation is a natural place to start. The unsubstituted imidazolium cation is shown in Figure 1a. The energy content of the ionic liquid cation needs to be raised for high energy applications, and this can be done in several ways.

First, replacing the imidazolium ring with a more nitrogen rich ring can increase the energy content of the ionic liquid. Second, the hydrogen or alkyl side chains can be replaced with high energy groups, such as -CN, $-NH_2$, $-N_3$, and $-NO_2$. Finally, the anion chosen should be oxygen rich to serve as an oxidizer. Interesting anions include nitrate, perchlorate, and dinitramide anions (Figure 2). Two promising nitrogen-rich candidates are the triazolium (Figure 1b) and the tetrazolium (Figure 1c) cations with three and four nitrogens respectively, in the ring. These two cations could present a useful balance between exothermicity and thermal stability. A thorough study of the 1,2,4-triazolium cation has recently been published.[2] The focus of the present study is on the electronic structure of the tetrazolium cation.

The energetic tetrazolium cations that have been synthesized include 1-amino-4,5-dimethyltetrazolium,[3] 2-amino-4,5-dimethyltetrazolium,[3] 2,4,5-trimethyltetrazolium,[4] and 1,5-diamino-4-H-tetrazolium.[5] Melting points for these cations combined with iodide, nitrate, and perchlorate anions range from –59 ºC to 156 ºC and their liquid range can be up

to 229 ºC. A summary of energetic tetrazolium cations that have recently appeared in the literature and their melting points are given in Table 1.[3-6]

Synthesis of triazolium cations is easier than that of tetrazolium cations, hence, there have been many more studies of energetic triazolium than tetrazolium compounds. Triazolium cations have been successfully substituted with azido,[4] nitro,[4] and amino groups.[3-6] The successful synthesis of the triazolium cation is not so surprising due to its similarity to the popular imidazolium cation, but what is surprising is the use of the triazole ring as an ionic liquid anion. Ionic liquids typically contain a large asymmetric organic cation, which causes the ions to be poorly coordinated, but as shown by Katritzky et al.,[7] the anion can be large as well. An ionic liquid, 1-butyl-3-methylimidazolium 3,5-dinitro-1,2,4-triazolate, containing a high energy planar anion, has been synthesized and has a low melting point of 35 ºC. Delocalization of charge on the anion ring is caused by the nitro substituent groups, which also serve to increase the energy of the compound.

Interactions in ionic liquids are more complicated than those in simpler liquids, making them more difficult to understand on a molecular level. Theory can provide an excellent tool for understanding the structure and dynamics of ionic liquids. There have been several molecular dynamics (MD) simulations of ionic liquids based on imidazolium, pyridinium and ammonium cations. These studies modeled bulk properties, such as melting points,[8] diffusion,[8-12] and viscosity.[8,9] Radial distribution functions[8,10,12-16] and densities have also been calculated.[8,10,15,17] Dynamics simulations have revealed that 1-ethyl-3-methylimidazolium nitrate has diffusion properties similar to those of a supercooled liquid.[16] Systematic MD studies of 1-alkyl-3methylimidazolium ionic liquids with several anions have been performed using various classical force fields.[11,17] Three first principles based MD

studies on 1,3-dimethylimidazolium chloride have been performed and compared to classical simulations and experimental neutron scattering experiments.[12-14] There are differences in the local structures predicted by classical vs. first principles based MD. Efforts to improve both classical and first principles based MD simulations of ionic liquids are ongoing in several groups.

The solvent properties of ionic liquids and their propensity to be synthesized depend greatly on their acid-base properties, so the ability to predict these properties would be of great use. The more acidic the cation, the more difficult it will be to protonate the neutral ring to form an ionic liquid. Of the three neutral rings, imidazole, triazole, and tetrazole, the weakest base is tetrazole. Tetrazolium is therefore, the most difficult to synthesize. Acidity constants of triazole and tetrazole have been calculated using several semi-empirical methods.[19]

In order to have low melting ionic liquids, the charge of at least one of the ions must be delocalized. Multi Configuration Self Consistent Field (MCSCF) analysis of the triazolium cation shows that the electrons on the cation are shared between two resonance structures.[2] This study also investigated the effects of more energetic and less energetic substituent groups. For example, the nitrile group was proposed as a better substituent for high energy applications than an azide group. The structures of triazolium dinitramide systems were investigated using dimer pairs. A wide variety of geometries were found and the presence of small barriers for proton transfer from the cation to the anion show that deprotonation may be an important mechanism in decomposition of triazolium-based ionic liquids.

Computational Methods

Initial structures were obtained by performing gas phase density functional theory (DFT) calculations on the isolated ions using the Becke three parameter Lee-Yang-Parr hybrid functional (B3LYP).[20,21] The basis sets used were 6-31G(d,p),[22-24] and 6-311G(d,p).[18,19] The DFT geometries and energies are compared to second order Moller-Plesset perturbation theory (MP2) calculations.[26,27] A MCSCF population analysis[28] using Edmiston-Ruedenberg type localized molecular orbitals (LMO)[29] was carried out to investigate the amount of electron delocalization in the ring. The orbitals included in the MCSCF active space are both π orbitals, their corresponding antibonds, and the lone pair.

Calculations on ion pairs provide information about the fundamental interactions between the cation and the anion. The gas phase ion interactions can provide insight into the bulk liquid structure. Dimer pairs were optimized using DFT and MP2 methods and the 6-31+G(d)[30,31] basis set. Hessians (matrices of energy second derivatives) are used to determine if stationary points are minima or transition states. At the final MP2/6-311G(d,p) geometries, improved relative energies were obtained for some of the tetrazolium cation isomers, using singles and doubles coupled cluster theory with perturbative triples (CCSD(T))[32,33] with the 6-311G(d,p) and cc-pVTZ[34] basis sets.

Intrinsic reaction coordinate (IRC) calculations (also referred to as the minimum energy path) were used to connect transition states with reactants and products.[35,36] The step size used for the IRC calculations was 0.05 $(amu)^{1/2}$ bohr. All calculations were done with GAMESS,[37,38] and all molecules were visualized with MacMolPlt.[39]

Results and Discussion

Relative Energies of Tetrazolium Cations

The tetrazolium cation has four parent isomers (all R=H) labeled I, II, III, and IV (Figure 3). Each isomer has two possible resonance contributors. The relative energies of I, II, III, and IV are shown in Table 2. Isomer I is predicted by MP2 to be the lowest in energy, with isomer II only 1.9 kcal/mol higher in energy. The MP2 relative energy ordering is I<II<IV<III. In general, DFT and MP2 are in good agreement, although DFT slightly reverses the order of isomers I and II.

The relative energies of isomers I and II were also calculated using the CCSD(T) method at the MP2/6-311G(d,p) geometries. Using the 6-311G(d,p) and cc-pVTZ basis sets, isomer II is predicted to be lower in energy than I by 0.2 and 0.5 kcal/mol respectively. The open chain form of the cation, azido formidinium (V), shown in Figure 4, was also compared to the four parent isomers. At the MP2/6-311++G(d,p) geometry this isomer is 0.7 kcal/mol lower in energy than isomer I (Table 3). According to CCSD(T)/cc-PVTZ//MP2/6-31+G(d), isomer V is 1.8 kcal/mol higher in energy than isomer I, and 1.3 kcal/mol higher in energy than isomer II.

The large MP2/6-311G(d,p) barriers for proton transfer reactions between isomer I and isomers II, III, and IV (Figure 5) imply that movement of a proton on the ring is not likely.

Calculated MP2/6-311G(d,p) σ and π bond orders[40] (Figure 6) suggest that the two double bonds in the ring are delocalized. The bond lengths between atoms in the rings do not exhibit any significant changes among the four parent isomers (Table 4). The bond lengths between atoms in the cation rings are consistent with a delocalized ring system (cf., 1.32 Å for a CN double bond and 1.34 Å for a NN double bond).

Bond lengths can suggest which resonance structures are favored for isomers I and III. (The two resonance structures for isomers II and IV are equivalent.) The nearly equal N1-N2 and N2-N3 bond lengths in isomer I (1.31 and 1.29 A, respectively) suggest that resonance structures I and I' make similar contributions to the electronic structure of the ring, with perhaps a slight preference for I over I'. The two resonance structures for isomer III are likewise equally important as indicated by the equal N2-N3 and N3-N4 bond lengths and the nearly equal lengths of the N1-C5 and N4-C5 bonds.

MCSCF π LMO populations and π bond orders (Table 5) can provide a more sophisticated analysis of competing resonance structures.[2] The diagonal density matrix elements give the electron occupancy of the localized molecular orbitals. The off-diagonal elements give the bond orders. A positive bond order indicates a bonding interaction and a negative bond order indicates an antibonding interaction between the LMOs.

For isomer I, the π orbital populations on N1 and N3 are 1.42 and 1.37 respectively. In a resonance structure that was purely I, N1 would have a π population of 2.00 and all other atoms in the ring would have π populations of 1.00 (see Figure 3). If the structure was purely I', then N3 would have a π population of 2.00 and all other atoms in the ring would have π populations of 1.00 (cf., Figure 3). In reality both N1 and N3 have π populations of less than 1.5 electrons. This indicates that the ring is a hybrid between I and I' and is delocalized. The difference in π orbital populations indicate that I is slightly favored over I'. In both I and I' C5-N4 is a double bond. This bond has a π bond order of 0.73, the largest of the five bonds. The N2-N3 π bond order, a double bond in I, is 0.65. The N1-N2 π bond order, a double bond in I', is 0.57. Neither of these bond orders is close to 1.00 indicating once again that the ring is delocalized.

A similar analysis can be done for isomers II-IV showing that all the π electrons are delocalized on the cations. According to the π orbital populations in isomer III, the lone pair is distributed equally on N1 and N2. Similar comments apply to isomer II(IV), in which the lone pair is distributed equally on N1 and N4 (N2 and N3).

The relative energies of the parent isomers with a single substituent on a nitrogen or a carbon are shown in Table 6. The isomers with a substituent R' (on a nitrogen) are labeled I-A though IV-A (Figure 7a), the isomers with a substituent R'' (on the carbon) are labeled I-B though IV-B (Figure 7b), and the isomers with a substituent R''' (on a nitrogen) are labeled I-C through IV-C (Figure 7c). For all substituents, the relative energies remain the same as those of the parent isomers with no substitution.

A $-NO_2$ substituent does not bind well to a nitrogen of the tetrazolium ring: the NN distance is 2.4 Å, compared with a normal single NN bond distance of 1.430 Å as, for example, in N_2H_4.[41] Because of the lack of a strong NN bond, the nitro group easily changes positions on the ring. For example, III-A(NO_2) converts to isomer I-A(NO_2), with no barrier. A cation with a $-NO_2$ substitutent bound to a ring nitrogen, is most likely an ion-dipole complex, in which the cation becomes a neutral tetrazole ring and the nitro group becomes an incipient cation. This is easily verified by examining the charge on the substituent $-NO_2$ group. The average of the Mulliken charges[42] on the $-NO_2$ group for all isomers with $-NO_2$ on N is +0.83. The average MP2/6-311G(d,p) binding energy of NO_2^+ to a tetrazole ring is 22.8 kcal/mol.

A comparison of relative energies (averaged for all four parent isomers) for substitution on a C vs. N is shown in Table 7. Substitution at C is almost always energetically favored over substitution at N. The only exception is for the $-NO_2$ substituent.

The bond lengths within the ring do not change significantly from the parent isomer when a single substituent is present.

Heats of Formation

The energy content of a heterocycle can be raised dramatically by increasing the number of nitrogen atoms in the ring. For example, the experimental heats of formation for imidazole, 1,2,4-triazole, and tetrazole are 14.0, 26.1, and 56.7 kcal/mol respectively.[40]

In this work, heats of formation for the cation rings were calculated using isodesmic reactions and the Gaussian-2 (G2) method,[43,44] a multilevel method designed to obtain accurate thermochemistry. The G2 calculations were performed using Gaussian 94.[45] Isodesmic reactions, in which the number of formal bond types is conserved, minimize the change in correlation energy,[46] thereby reducing the error in a computed heat of formation.

The isodesmic reactions of each resonance structure are shown in Figure 8. For I and III each resonance structure has a different isodesmic reaction. The heats of formation for I and I' (III and III') are similar, and their average will be taken as a heat of formation for the composite I/I' (III/III').[2] The largest difference in heats of formation between resonance structures is 13.0 kcal/mol for the $-NO_2$ substituent; the other differences are all less than 3 kcal/mol.

The heats of formation for the parent isomers are shown in column 2 of Table 8. These compare well to heats of formation that were calculated previously using Gaussian-3.[19] The changes in heats of formation for substituted cations relative to the hydrogen substituted cations are shown in Table 9. For the most part, substitution destabilizes the ring, although, substitution by a fluorine or an amine on a carbon stabilizes the ring by up to 40 kcal/mol.

However, an azide group or a nitrile group can increase the heat of formation of the ring by as much as 112 kcal/mol. Substitution with –F, $–NH_2$, and $–NO_2$ generally increases the heats of formation by about 20 kcal/mol.

Proton Transfer Reactions

One possible reaction pathway for the simple ion pairs is proton transfer from the cation to a partner anion to form a neutral pair. The heats for the reaction $CN_4H_3^+ + Y^- \rightarrow CN_4H_2 + HY$ (where Y=NO_3, ClO_4, or $N(NO_2)_2$) are shown in Table 10. All of the energy differences (ΔE) are exothermic, by 46-144 kcal/mol, and the ΔEs for proton transfer to each of the three anions are similar. The heat of reaction for removal of H"', shown in column three of Table 10, is more exothermic than the removal of H' (see Figure 3) because the 2,5-disubstituted neutral ring is 3.7 kcal/mol lower in energy than the 1,5-disubstitued neutral ring. The 1,3-disubstituted ring is considerably higher in energy. There are three possible positions for proton transfer to the dinitramide anion, but dinitramic acid is most stable with the proton on the center nitrogen, so this is the acid used when calculating the ΔEs.

Ion Pair Interactions

Ionic liquid dimer pairs were optimized with DFT and MP2 methods using the 6-31+G(d) basis set. ZPE corrections at the B3LYP/6-31+G(d) level of theory were used in all cases. The structures of the tetrazolium dimers with perchlorate and nitrate partners are shown in Figures 9 and 10, respectively. For the cation with perchlorate or nitrate, at both levels of theory a proton transfers from the cation to the anion during geometry optimization of the ion pair. Hydrogen bonding and other attractive electrostatic interactions are found to

be important in the dimer structures. Optimizations give geometries in which the number of attractive electrostatic interactions is maximized. For pairs containing HNO_3 or $HClO_4$, there is always one linear O-H---N hydrogen bond. Hydrogen bond lengths vary from 1.795 Å to 1.846 Å for HNO_3 and 1.771 Å to 1.822 Å for $HClO_4$. The majority of structures also contain a non-linear N-H---O hydrogen bond or a C-H---O attractive electrostatic interaction.

Tetrazolium dinitramide (Figure 11) pairs have a much wider variety of structures than tetrazolium nitrate or tetrazolium perchlorate pairs. This is because the dinitramide anion has the ability to bind a proton at three different positions (Figure 12). For $HN(NO_2)_2$ pairs, hydrogen bond lengths vary from 1.721 Å to 1.942 Å. Once again hydrogen bonding and other electrostatic interactions are important for structures containing $HN(NO_2)_2$. However, six out of 19 structures do not contain a linear hydrogen bond at all. These structures usually contain weaker, nonlinear, N-H---N, O-H---N, or N-H---O attractive electrostatic interactions.

The relative energies of the pairs with HNO_3 or $HClO_4$ show that the lowest energy pairs contain the 1,3-H substituted tetrazole isomer. For example, structure 1 in Figure 11 is 2.8 kcal/mol lower in energy than structure 6. This difference in energy can be attributed to the difference in energy of the 1,3-H substituted tetrazole ring (number 1) and the 1,2-H substituted tetrazole ring (number 6). The lowest energy structures with $HN(NO_2)_2$, contain the 1,3-H substituted tetrazole isomer and have the proton transferred to the central nitrogen of $N(NO_2)_2^-$. Also shown in Figures 9, 10, and 11 are the changes in energy for the reaction in which the infinitely separated anion and cation come together to form either an ion pair or a neutral pair by proton transfer from the cation to the anion:

$$[\text{Tetrazolium}]^+ + [\text{Anion}]^- \rightarrow [\text{Tetrazole}][\text{H-Anion}] \text{ or}$$

$$[\text{Tetrazolium}]^+ + [\text{Anion}]^- \rightarrow [\text{Tetrazolium}][\text{Anion}]$$

(Anion = NO_3, ClO_4, or $N(NO_2)_2$).

For pairs in which a proton transfers, the neutral pair is lower in energy than the separated neutral products, due to the stabilization from electrostatic interactions in the neutral dimer products. The average dimer stabilization is 7.3, 9.6, and 8.4 kcal/mol for [HNO_3], [$HClO_4$], and [$HN(NO_2)_2$] respectively.

At the DFT level of theory there are some stable $HN(NO_2)_2^-$-cation pair geometries; however when those DFT geometries are optimized with MP2, the acidic proton usually transfers to the anion. Only one stable anion-cation pair was found at the MP2 level of theory. This ion pair is structure 13 in figure 11. The ion pair is higher in energy by 18.9 kcal/mol than the lowest energy neutral pair at MP2/6-31+G(d). In this structure the N-H bond length is stretched to 1.076 Å as compared to 1.019 Å for the non-acidic hydrogen. The hydrogen bond length is also much shorter at 1.598 Å. The length of the hydrogen bonds show that the hydrogen bond in the ionic structure is much stronger than the hydrogen bond in the various neutral structures. The MP2 barrier for proton transfer from the cation to the anion of structure 13 of Figure 11 is 0.4 kcal/mol (Figure 13), however when ZPE corrections are accounted for, the barrier disappears. For all dimer pairs, proton transfer from a cation to the nitrogen of the dinitramide anion is favored energetically over proton transfer to one of the oxygens of the dinitramide anion.

Gas phase calculations on an ionic pair can, of course, only give an approximation to the true interactions in a crystalized ionic liquid. Proton transfer from the cation to the anion should be less likely in the crystal or in bulk liquid. For example, in studies of ammonium salts, proton transfer occurs in the isolated gas phase ion pair, but stable ion pairs are found

when two anions and two cations are present.[47] MP2/6-31++G(d,p) calculations on 1,2,4-triazolium dinitramide typically do not give stable ion pairs.[2] The most recent MP2 calculations done on three ion pairs of 1,2,4-triazolium dinitramide, show that the six ion cluster is only 1.0 kcal/mol higher in energy than the six neutral cluster.[48] This is caused by an increase in charge balance that occurs as the number of ion pairs present increases. That is, each negative (positive) ion is balanced by an increasing number of positive (negative) ions. The structures and relative energies for multiple ion pairs is still under investigation. Future studies on larger clusters of ionic pairs are needed to more fully understand the interactions in the bulk liquid.

Conclusion

In this work, tetrazolium cations I and II were established to be the lowest energy isomers, for the parent isomers as well as for all substituted isomers. The DFT and MP2 energies are generally in good agreement. The relative energy of the open chain form of the cation is predicted to be only slightly higher in energy than isomers I and II. A MCSCF π orbital analysis indicates that the electrons in the cation ring are delocalized. Calculated heats of formation show that the tetrazolium cation ring has the potential to release large amounts of energy during decomposition and thus has excellent potential as a high energy fuel. This is especially true when the ring is substituted with $-N_3$ or –CN. When a cation is paired with oxygen rich anions, a single gas phase ion pair was not generally found to be stable. A proton transfers without barrier from the cation to the anion to form a neutral pair.

Acknowledgements

This work was supported by a grant from the Air Force Office of Scientific Research. The authors are grateful to Drs. Michael Schmidt and Greg Drake for many helpful discussions.

References

(1) *Ionic Liquids: Industrial Applications to Green Chemistry*; Rogers, R. D., Seddon, K. R., Eds.; ACS Symposium Series 818; American Chemical Society: Washington DC, 2002.

(2) Schmidt, M. W.; Gordon, M. S.; Boatz, J. A. *J. Phys. Chem. A* **2005**, *109*, 7285.

(3) Xue, H.; Arritt, S. W.; Twamley, B.; Shreeve, J. n. M. *Inorg. Chem.* **2004**, *43*, 7972.

(4) Xue, H.; Gao, Y.; Twamley, B.; Shreeve, J. n. M. *Chem. Mater.* **2005**, *17*, 191.

(5) Drake, G. W.; Hawkins, T. W.; Boatz, J.; Hall, L.; Vij, A. *Propellants, Explos., Pyrotech.* **2005**, *30*, 156.

(6) Drake, G.; Hawkins, T.; Brand, A.; Hall, L.; McKay, M. *Propellants, Explos., Pyrotech.* **2003**, *28*, 174.

(7) Katritzky, A. R.; Singh, S.; Kirichenko, K.; Holbrey, J. D.; Smiglak, M.; Reichert, W. M.; Rogers, R. D. *Chem. Comm.* **2005**, 868.

(8) Velardez, G. F.; Alavi, S.; Thompson, D. L. *J. Chem. Phys.* **2003**, *119*, 6698.

(9) Yan, T.; Burnham, C. J.; Popolo, M. G. D.; Voth, G. A. *J. Phys. Chem. B* **2004**, *108*, 11877.

(10) Margulis, C. J.; Stern, H. A.; Berne, B. J. *J. Phys. Chem. B* **2002**, *106*, 46.

(11) Andrade, J. d.; Böes, E. S.; Stassen, H. *J. Phys. Chem. B* **2002**, *106*, 3546.

(12) Bühl, M.; Chaumont, A.; Schurhammer, R.; Wipff, G. *J. Phys. Chem. B* **2005**, *109*, 18591.

(13) Bhargava, B. L.; Balasubramanian, S. *Chem. Phys. Lett.* **2006**, *417*, 486.

(14) Del Pópolo, M. G.; Lynden-Bell, R. M.; Kohanoff, J. *J. Phys. Chem. B* **2005**, *109*, 5895.

(15) Cadena, C.; Zhao, Q.; Snurr, R. Q.; Maginn, E. J. *J. Phys. Chem. B* **2006**, *110*, 2821.

(16) Del Pópolo, M. G.; Voth, G. A. *J. Phys. Chem. B* **2004**, *108*, 1744.

(17) Canongia Lopes, J. N.; Deschamps, J.; Pádua, A. A. H. *J. Phys. Chem. B* **2004**, *108*, 2038.

(18) Urahata, S. M.; Ribeiro, M. C. C. *J. Chem. Phys.* **2004**, *120*, 1855.

(19) Satchell, J. F.; Smith, B. J. *PCCP* **2002**, *4*, 4314.

(20) Vosko, S. J.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(21) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(22) Ditchfield R.; Hehre W. J.; Pople J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(23) Hehre WJ; Ditchfield R; Pople J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(24) Hariharan P. C., Pople J. A. *Theor. Chem. Acta* **1973**, *28*, 213.

(25) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(26) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(27) Fletcher, G. D.; Rendell, A. P.; Sherwood, P. *Mol. Phys.* **1997**, *91*, 431.

(28) Schmidt, M. W.; Gordon, M. S. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233.

(29) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

(30) Gordon M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(31) Clark T.; Chandrasekhar J.; Spitznagel G. W.; Schleyer PvR *J. Comput. Chem.* **1983**, *4*, 294.

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

(33) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2000**, *149*, 71.

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

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

(36) Truhlar, D. G., Gordon, M. S. *Science* **1990**, *249*, 491.

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

(38) Schmidt, M. W.; Gordon, M. S. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, 2005.

(39) Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Modell.* **1999**, *16*, 133.

(40) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.

(41) DeFrees, D. J.; Raghavachari, K.; Schlegel, H. B.; Pople, J. A. *J. Am. Chem. Soc.* **1982**, *104*, 5576.

(42) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833-1840, 1841-1846, 2338-2342, 2343-2346.

(43) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.

(44) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

(45) Frisch, M. J.; G. W. Trucks; H. B. Schlegel; P. M. W. Gill; B. G. Johnson; M. A. Robb; J. R. Cheeseman; T. Keith; G. A. Petersson; J. A. Montgomery; K. Raghavachari; M. A. Al-Laham; V. G. Zakrzewski; J. V. Ortiz; J. B. Foresman; J. Cioslowski; B. B. Stefanov; A. Nanayakkara; M. Challacombe; C. Y. Peng; P. Y. Ayala; W. Chen; M. W. Wong; J. L. Andres; E. S. Replogle; R. Gomperts; R. L. Martin; D. J. Fox; J. S. Binkley; D. J. Defrees; J. Baker; J. P. Stewart; M. Head-Gordon; Gonzalez, C.; Pople, J. A. *Gaussian, Inc., Pittsburgh PA* **1995**.

(46) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796.

(47) Alavi, S.; Thompson, D. L. *J. Chem. Phys.* **2003**, *119*, 4274.

(48) Li, H.; Gordon, M. S. **2006**, *in preparation*.

Table 1: Energetic tetrazolium cations, their melting points (T_m), thermal decomposition temperatures (T_d), and glass transition temperature (T_g).

1-Amino-4,5-dimethyltetrazolium Iodide[a]	T_m 121 ºC	
1-Amino-4,5-dimethyltetrazolium Nitrate[a]	T_g -59 ºC	T_d 170 ºC
1-Amino-4,5-dimethyltetrazolium Perchlorate[a]	T_m 51 ºC	T_d 182 ºC
2-Amino-4,5-dimethyltetrazolium Iodide[a]	T_m 124 ºC	
2-Amino-4,5-dimethyltetrazolium Nitrate[a]	T_m 94 ºC	T_d 173 ºC
2-Amino-4,5-dimethyltetrazolium Perchlorate[a]	T_m 140 ºC	T_d 238 ºC
2,4,5-Trimethyltetrazolium Iodide[b]	T_m 156 ºC	
2,4,5-Trimethyltetrazolium Nitrate[b]	T_m 94 ºC	T_d 193 ºC
2,4,5-Trimethyltetrazolium Perchlorate[b]	T_m 133 ºC	T_d 315 ºC
1,5-Diamino-4-H-Tetrazolium Perchlorate[c]	T_m 125-130 ºC	

[a] from Xue, Arritt, Twamley, and Schreeve[3], [b] from Xue, Gao, Twamley, and Schreeve[4], [c] from Drake, Hawkins, Boatz, Hall, Vij [5]

Table 2: Relative energies (kcal/mol) of Isomers I, II, III, and IV.

Cation	B3LYP/6-31G(d,p)	B3LYP/6-311G(d,p)	MP2/6-31G(d,p)	MP2/6-311G(d,p)
I	0	0	0	0
II	-0.6	-0.4	1.4	1.9
III	16	16.2	19.1	19.2
IV	14.9	14.7	15.2	14.7

Table 3: Relative energies (kcal/mol) of the closed chain isomers I, II, and the open chain isomer, V, using CCSD(T).

	MP2/6-311++G(d,p)	CCSD(T)/cc-PVTZ// MP2/6-311++G(d,p)	CCSD(T)/6-311G(d,p)// MP2/6-311++G(d,p)
I	0.0	0.0	0.0
II	1.9	-0.5	-0.2
V	-0.7	1.8	1.6

Table 4: Bond lengths (Angstroms) in the ring for isomers I through IV.

	C5-N1	N1-N2	N2-N3	N3-N4	N4-C5
I	1.36	1.31	1.29	1.33	1.31
II	1.33	1.35	1.30	1.35	1.33
III	1.35	1.32	1.32	1.32	1.34
IV	1.35	1.31	1.32	1.31	1.35

Table 5: MCSCF π orbital populations and bond orders.

	π orbital populations					
	N1	N2	N3	N4	C5	Total bonding
I	1.42	1.10	1.37	1.12	0.98	1.93
II	1.44	1.09	1.09	1.44	0.94	1.92
III	1.46	1.46	1.04	1.10	0.95	1.91
IV	1.07	1.43	1.43	1.07	0.99	1.94

	Adjacent π bond orders				
	C5-N1	N1-N2	N2-N3	N3-N4	N4-N5
I	0.55	0.57	0.65	0.48	0.73
II	0.65	0.44	0.77	0.44	0.65
III	0.61	0.44	0.60	0.62	0.64
IV	0.64	0.56	0.50	0.56	0.64

	Next neighbor π bond orders (antibonding)				
	C5-N2	N1-N3	N2-N4	N3-C5	N4-N1
I	-0.22	-0.30	-0.27	-0.11	-0.15
II	-0.18	-0.19	-0.19	-0.18	-0.29
III	-0.30	-0.33	-0.22		-0.15
IV	-0.14	-0.34	-0.34	-0.14	

Table 6: MP2/6-311G(d,p) relative energies (kcal/mol) isomers with a single substitution on a nitrogen or a carbon. (See Figure 7 for notation.)

R'=	I-A	II-A	III-A	IV-A
H	0.0	1.9	19.1	14.7
F	0.0	3.4	18.3	12.6
CN	0.0	2.6	18.6	14.4
N_3	0.0	5.3	16.8	16.0
NH_2	0.0	1.8	17.1	13.0
NO_2	0.0	2.7		4.9

R''=	I-B	II-B	III-B	IV-B
H	0.0	1.9	19.1	14.7
F	0.0	5.0	22.6	16.8
CN	0.0	1.3	19.2	14.6
N_3	0.0	2.9	15.4	11.0
NH_2	0.0	3.5	16.3	11.7
NO_2	0.0	2.5	17.8	15.6

R'''=	I-C	II-C	III-C	IV-C
H	0.0	1.9	19.1	14.7
F	0.0	2.3	22.1	14.0
CN	0.0	1.5	19.9	14.9
N_3	0.0	1.6	20.1	9.6
NH_2	0.0	1.0	18.2	9.4
NO_2	0.0	1.1	13.5	3.3

Table 7: MP2/6-311G(d,p) average energy differences between substitution at N and C on the tetrazolium ring.

	Average Values (E_C - E_N):
F	-56.9 kcal/mol
CN	-20.1 kcal/mol
N_3	-30.0 kcal/mol
NH_2	-27.9 kcal/mol
NO_2	+6.0 kcal/mol

Table 8: Heats of formation (kcal/mol), from G2 theory, for parent isomers.

	Calculated G3 Heats of Formation[a]	Calculated G2 Heats of Formation[b]
I	248.0	245.1
II	247.1	246.9
III	263.7	264.3
IV	262.8	259.8

[a] G3 Heats of Formation are taken from Satchell and Smith[19]. [b] From this work.

Table 9: Changes in heats of formation (kcal/mol), from G2 theory, due to substituents, relative to parent compoiunds. Refer to Figure 3 for structures.

R'=	I	I'	average	R''=	I = I'	R'''=	I	I'	average
H	0.0	0.0	0.0	H	0.0	H	0.0	0.0	0.0
F	17.1	17.8	17.5	F	-37.5	F	19.9	20.6	20.2
CN	70.7	71.7	71.2	CN	51.0	CN	70.5	71.5	71.0
N_3	108.6	111.4	110.0	N_3	79.7	N_3	109.9	112.8	111.3
NH_2	18.2	18.9	18.6	NH_2	-10.0	NH_2	20.7	21.4	21.0
NO_2	31.1	18.1	24.6	NO_2	18.0	NO_2	32.8	19.7	26.3

R'=	II	II'	average	R''=	II = II'	R'''=	II	II'	average
H	0.0	0.0	0.0	H	0.0	H	0.0	0.0	0.0
F	20.3	21.0	20.6	F	-36.0	F	20.3	21.0	20.7
CN	70.2	71.2	70.7	CN	51.7	CN	70.1	71.1	70.6
N_3	109.6	112.5	111.0	N_3	83.1	N_3	109.6	112.5	111.1
NH_2	19.9	20.6	20.3	NH_2	-10.0	NH_2	19.8	20.5	20.1
NO_2	32.0	18.9	25.4	NO_2	18.6	NO_2	32.0	18.9	25.4

R'=	III	III'	average	R''=	III = III'	R'''=	III	III'	average
H	0.0	0.0	0.0	H	0.0	H	0.0	0.0	0.0
F	20.5	21.2	20.8	F	-38.4	F	22.8	23.5	23.1
CN	70.8	71.8	71.3	CN	50.3	CN	71.2	72.3	71.8
N_3	104.8	107.6	106.2	N_3	77.3	N_3	110.8	113.7	112.2
NH_2	15.4	16.0	15.7	NH_2	-12.1	NH_2	19.8	20.4	20.1
NO_2				NO_2	16.6	NO_2	27.1	14.0	20.6

R'=	IV	IV'	average	R''=	IV = IV'	R'''=	IV	IV'	average
H	0.0	0.0	0.0	H	0.0	H	0.0	0.0	0.0
F	19.2	19.9	19.5	F	-39.6	F	19.2	19.9	19.6
CN	70.7	71.7	71.2	CN	50.7	CN	70.7	71.7	71.2
N_3	104.9	107.8	106.3	N_3	81.0	N_3	104.8	107.7	106.3
NH_2	15.3	15.9	15.6	NH_2	-11.7	NH_2	15.4	16.1	15.7
NO_2	21.3	8.3	14.8	NO_2	18.9	NO_2	21.4	8.3	14.9

Table 10: Heats of proton transfer reaction (kcal/mol) for anion = NO_3^-, ClO_4^- , and $N(NO_2)_2^-$ for deprotonation at H', H'', & H'''. (See Figure 3)

Cation	anion	H'	H''	H'''
I	NO3-	-124.8	-86.1	-128.5
	ClO4-	-102.6	-63.9	-106.3
	N(NO2)2-	-112.0	-73.3	-115.7
II	NO3-	-126.6	-103.7	same as H'
	ClO4-	-104.4	-81.5	
	N(NO2)2-	-113.9	-90.9	
III	NO3-	-143.7	-85.9	-147.4
	ClO4-	-121.5	-63.7	-125.3
	N(NO2)2-	-131.0	-73.1	-134.7
IV	NO3-	same as H'	-68.9	-142.9
	ClO4-		-46.7	-120.8
	N(NO2)2-		-56.2	-130.2

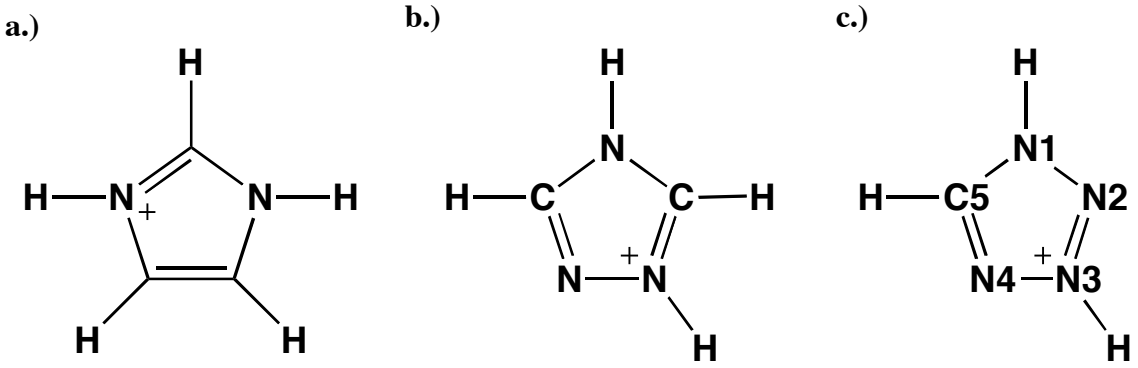


Figure 1: Imidazolium (a), 1,2,4-triazolium (b), and tetrazolium (c) cations.

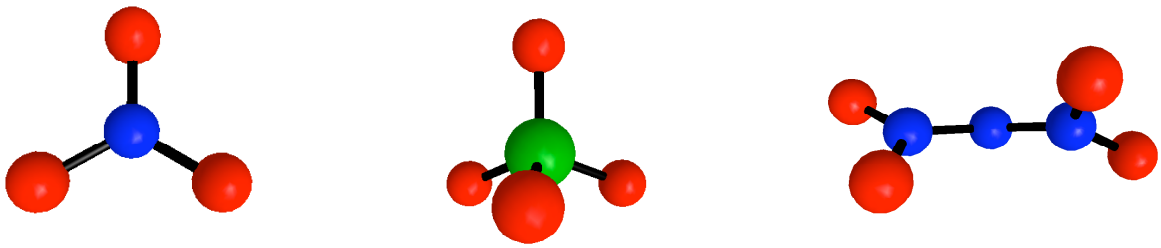


Figure 2: Nitrate, perchlorate and dinitramide anions (Oxygen=red, nitrogen=blue, chlorine=green).

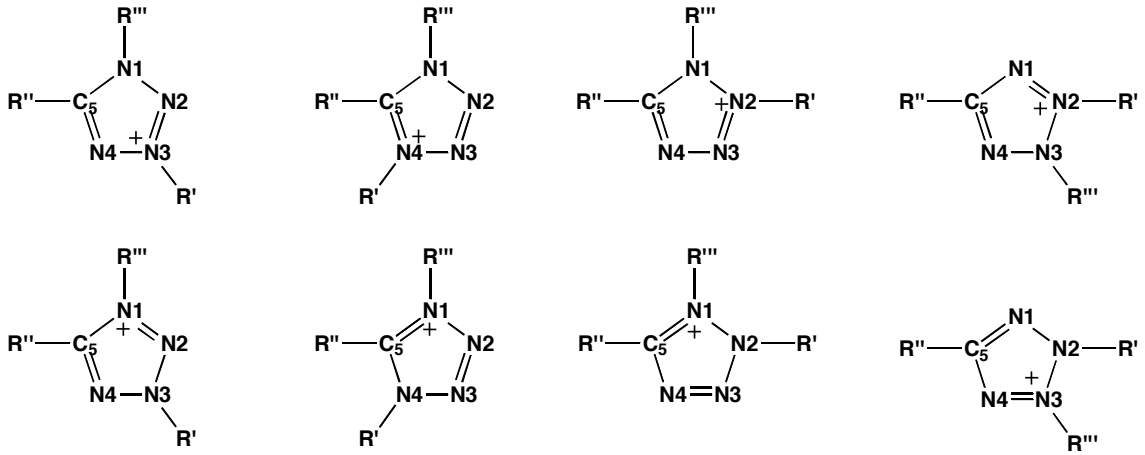


Figure 3: Resonance structures of isomers I-IV.

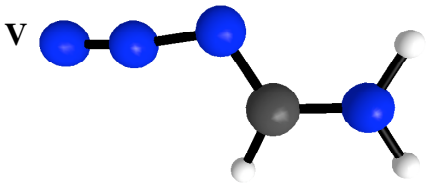


Figure 4: The open chain isomer, azido formidinium, V.

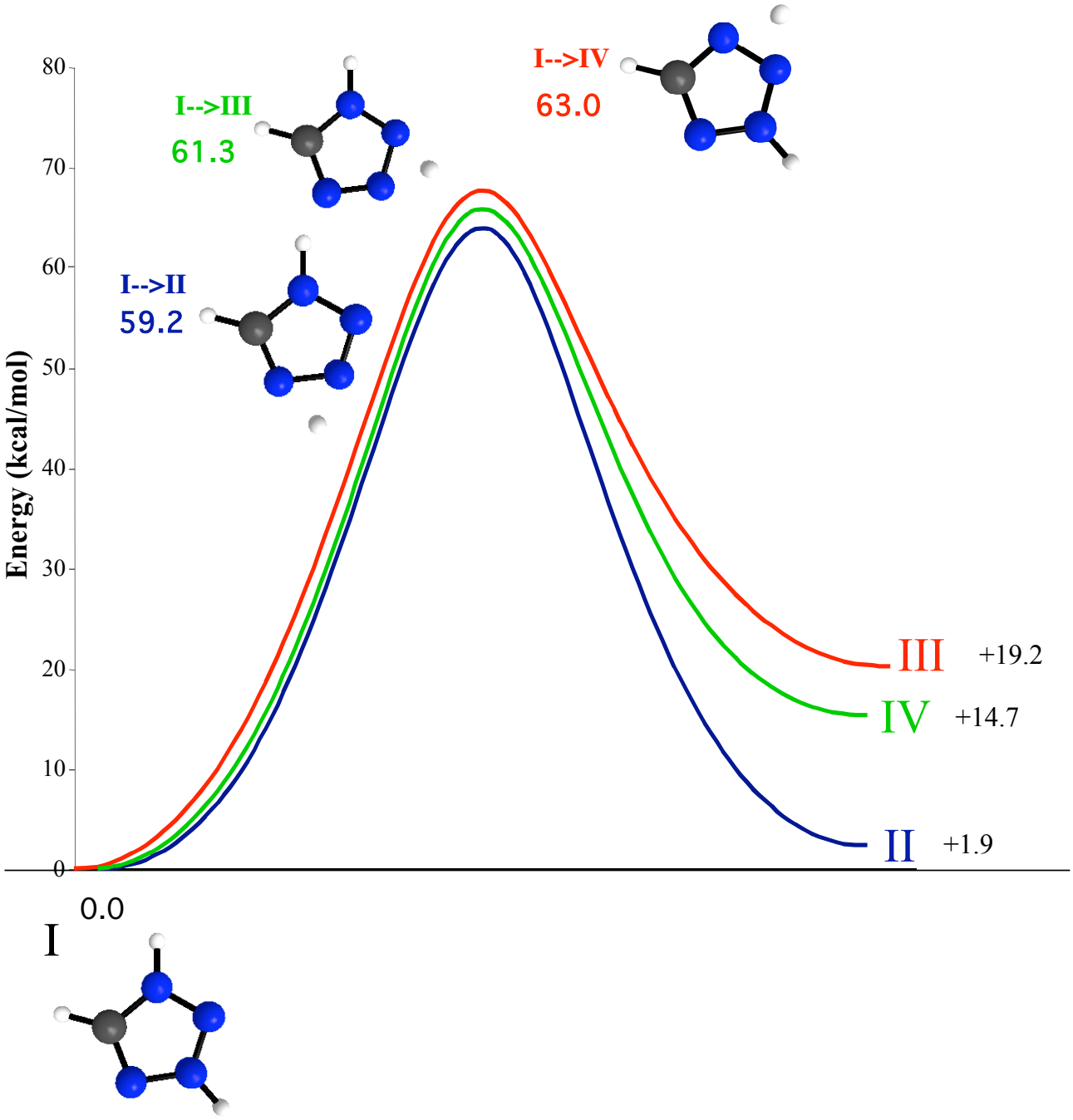


Figure 5: MP2/6-311G(d,p) minimum energy path for proton transfer between isomer I and isomers II, III, and IV, kcal/mol. Structures for the three transition states (I-->II, I-->III, I-->IV) are shown above the curves.

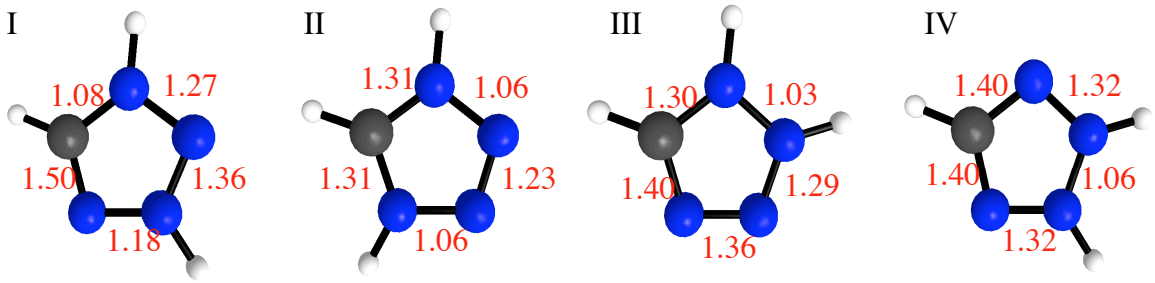


Figure 6: Isomers I-IV with calculated MP2 bond orders in red.

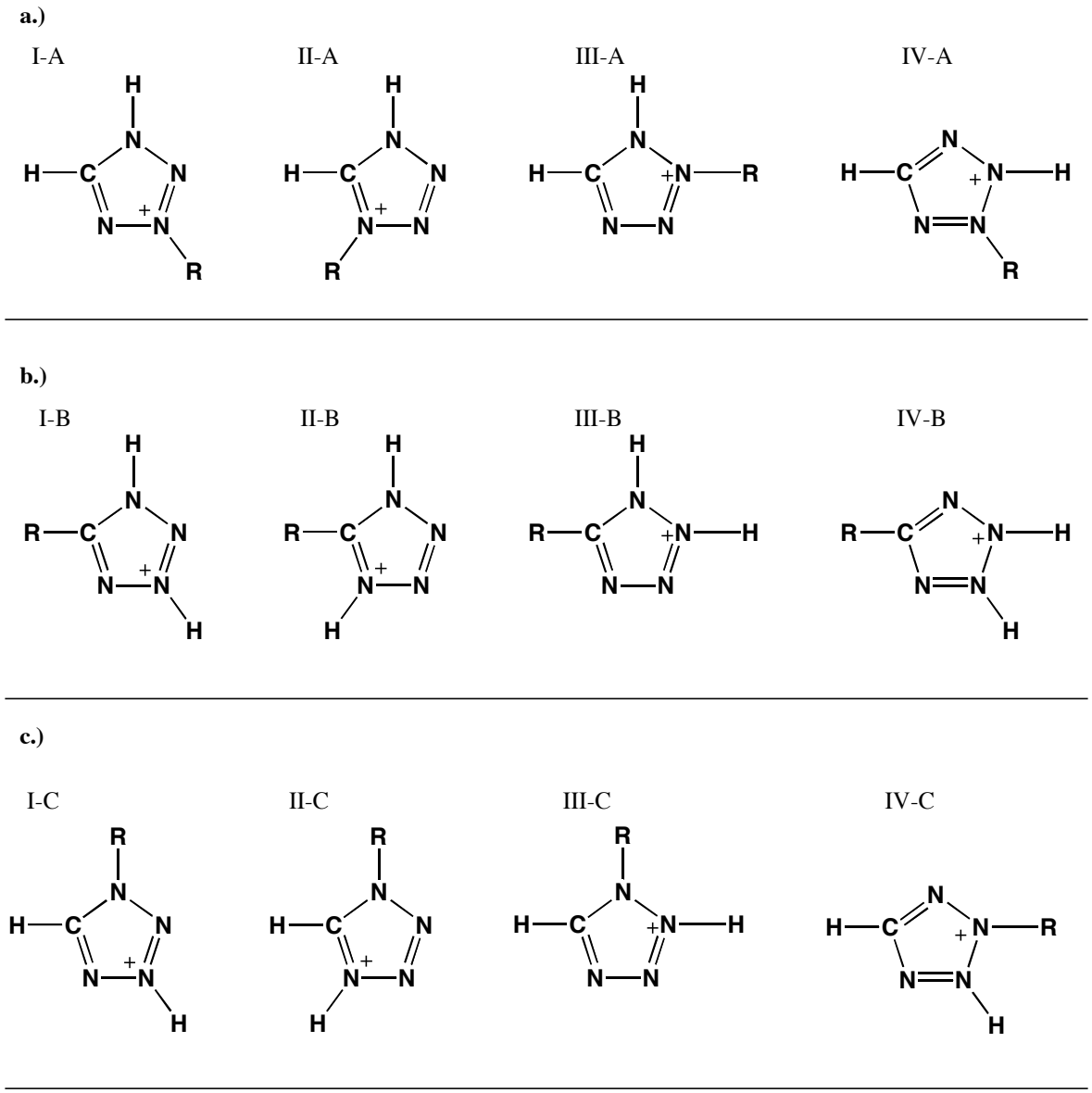


Figure 7: Isomers with a single substitution on a nitrogen or a carbon.

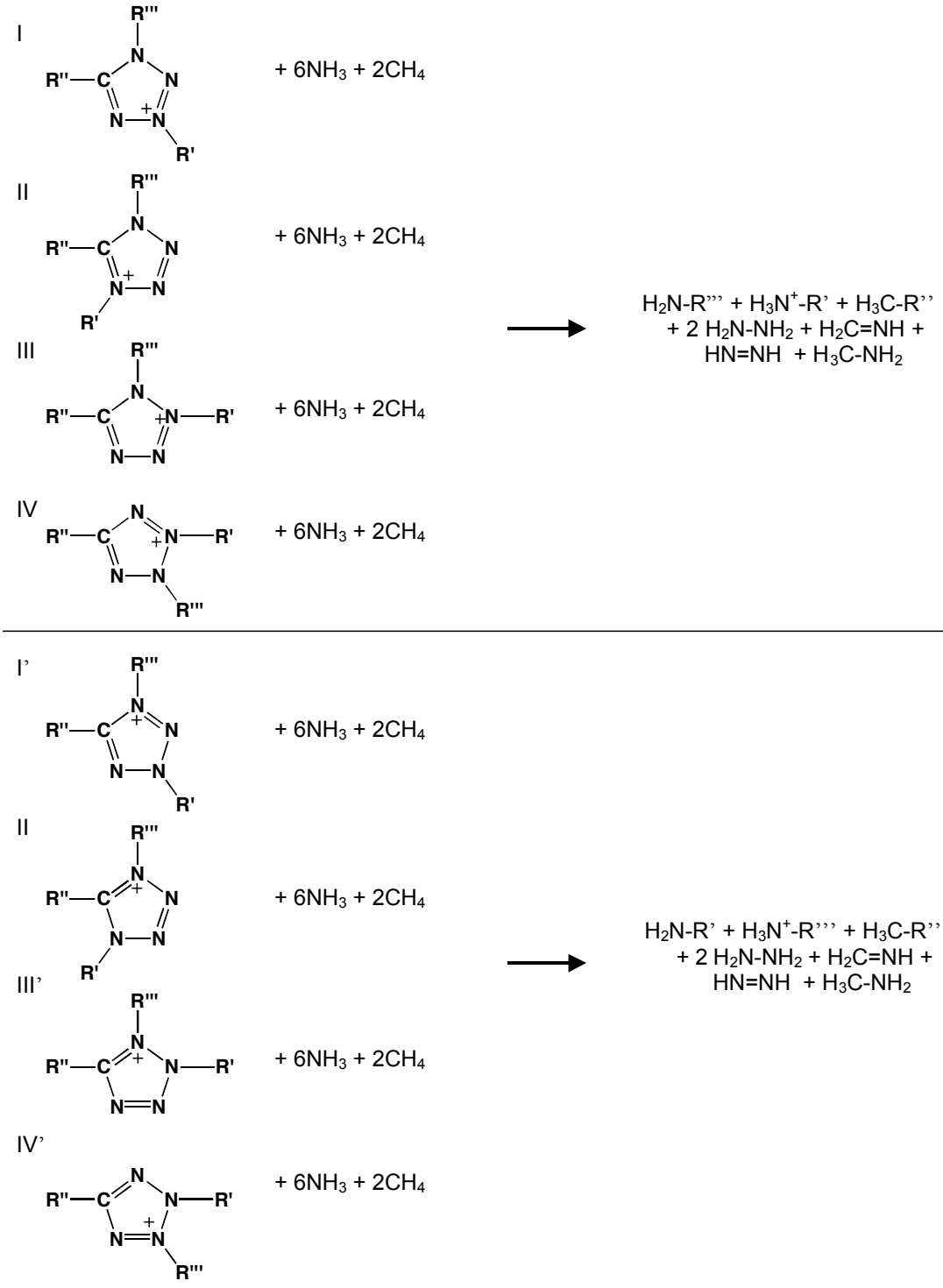


Figure 8: Isodesmic reactions for all resonance structures.

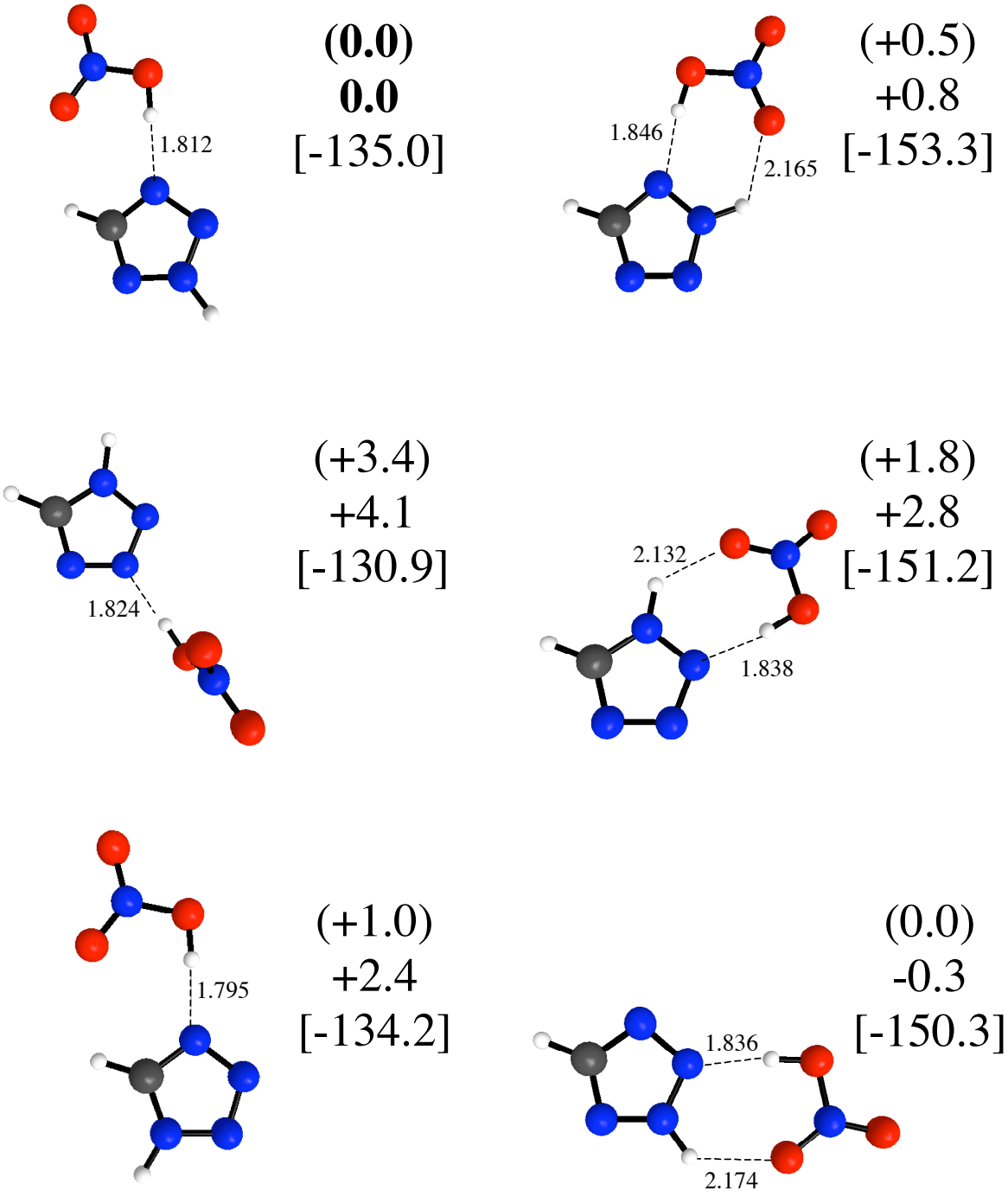


Figure 9: Optimized structures (Å) and MP2 and B3LYP (in parentheses) relative energies (kcal/mol) for proton transferred structures that result from optimization of initial geometries of a tetrazolium cation paired with a NO_3^- anion. MP2 heats of reaction (in brackets) in kcal/mol for the reaction: Tetrazolium + NO_3^- → [Tetrazole][H-NO_3].

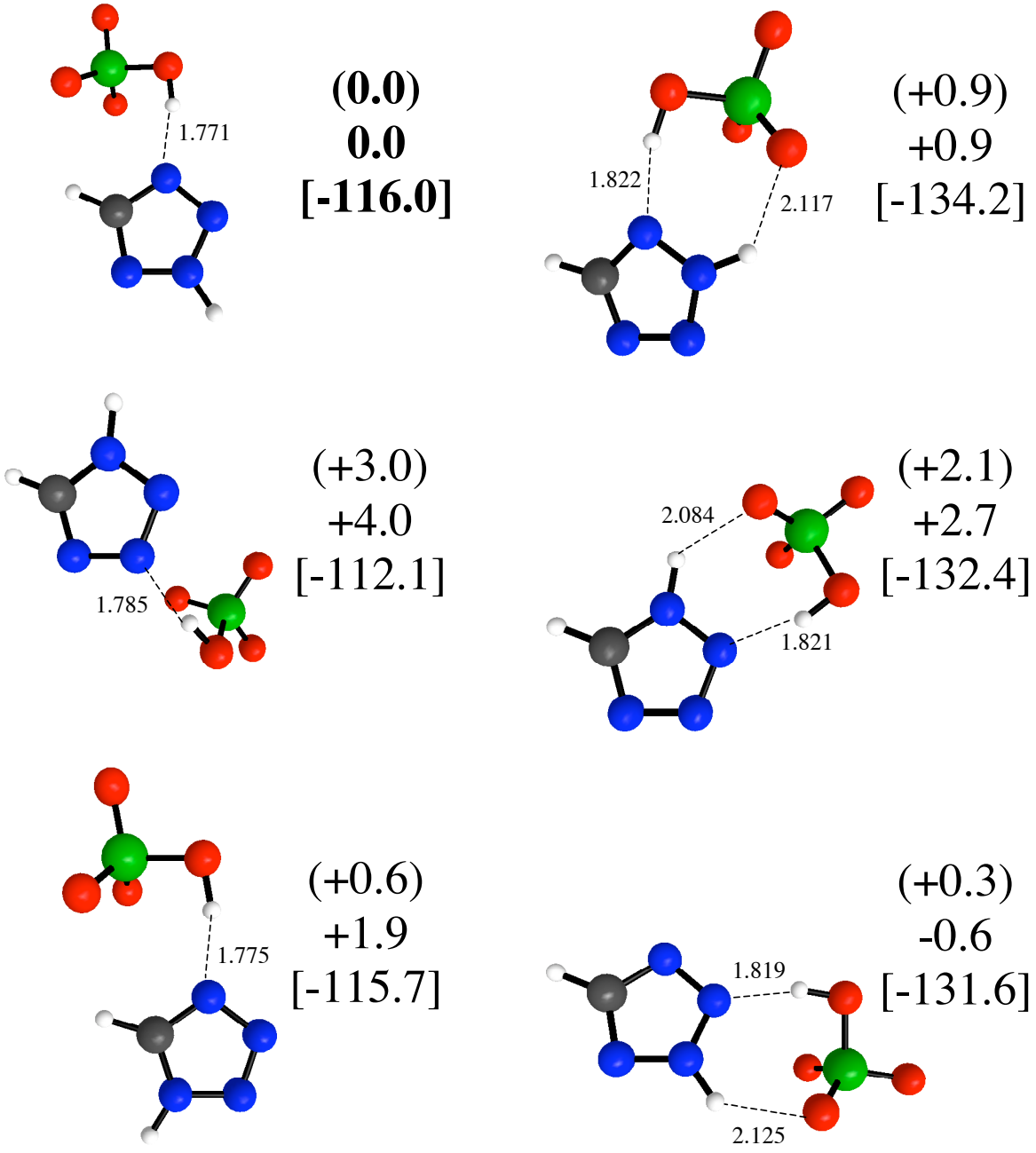


Figure 10: Optimized structures (Å) and MP2 and B3LYP (in parentheses) relative energies (kcal/mol) for proton transferred structures that result from optimization of initial geometries of a tetrazolium cation paired with a ClO_4^- anion. MP2 heats of reaction (in brackets) in kcal/mol for the reaction: Tetrazolium + ClO_4^- → [Tetrazole][H-ClO_4].

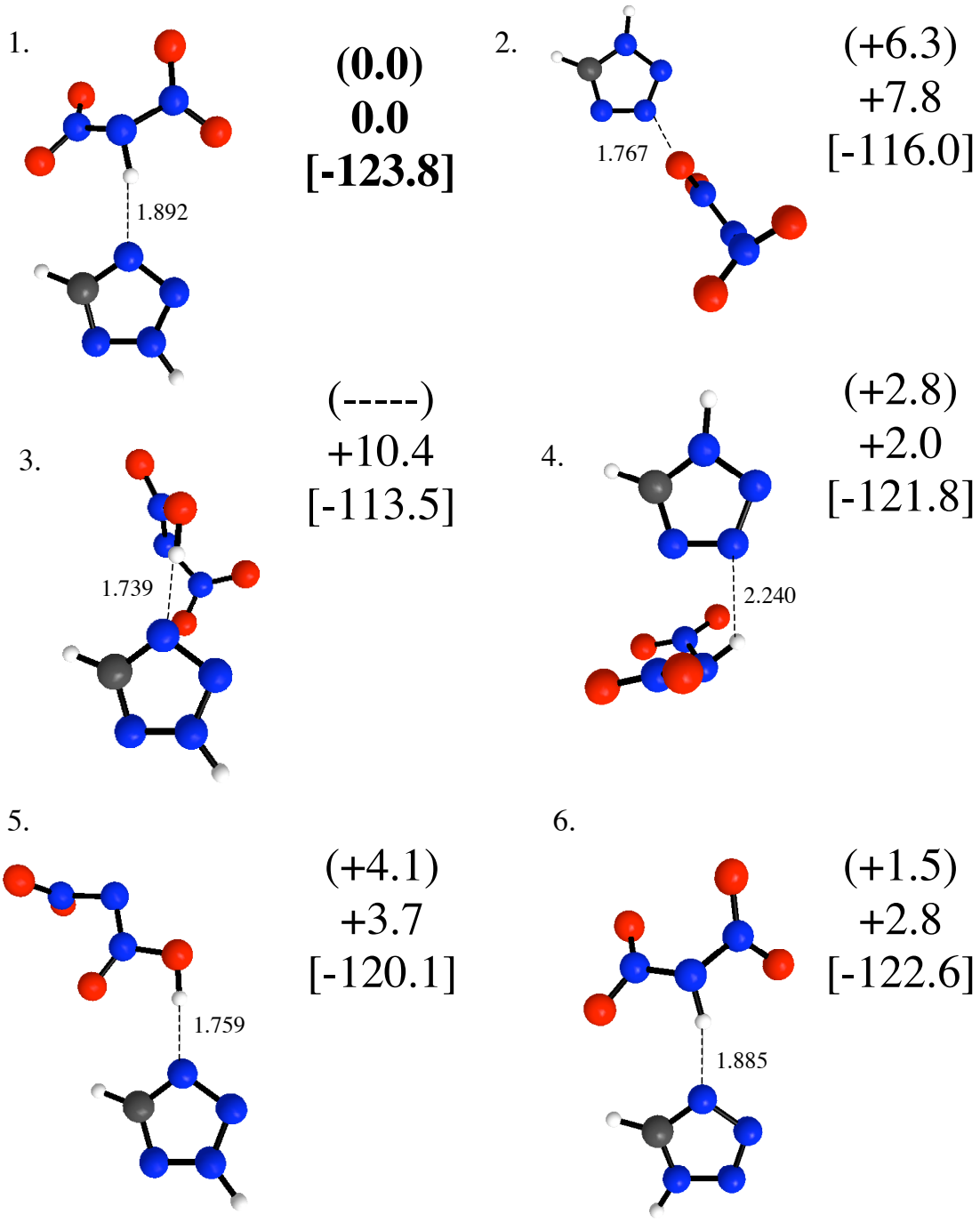


Figure 11: Optimized structures (Å) and MP2 and B3LYP (in parentheses) relative energies (kcal/mol) of the tetrazolium cation paired with a $N(NO_2)_2^-$ anion. MP2 heats of reaction (in brackets) in kcal/mol for the reaction: Tetrazolium + $N(NO_2)_2^-$ → [Tetrazole] $[HN(NO_2)_2]$ or (see text) [Tetrazolium]$[N(NO_2)_2]$.

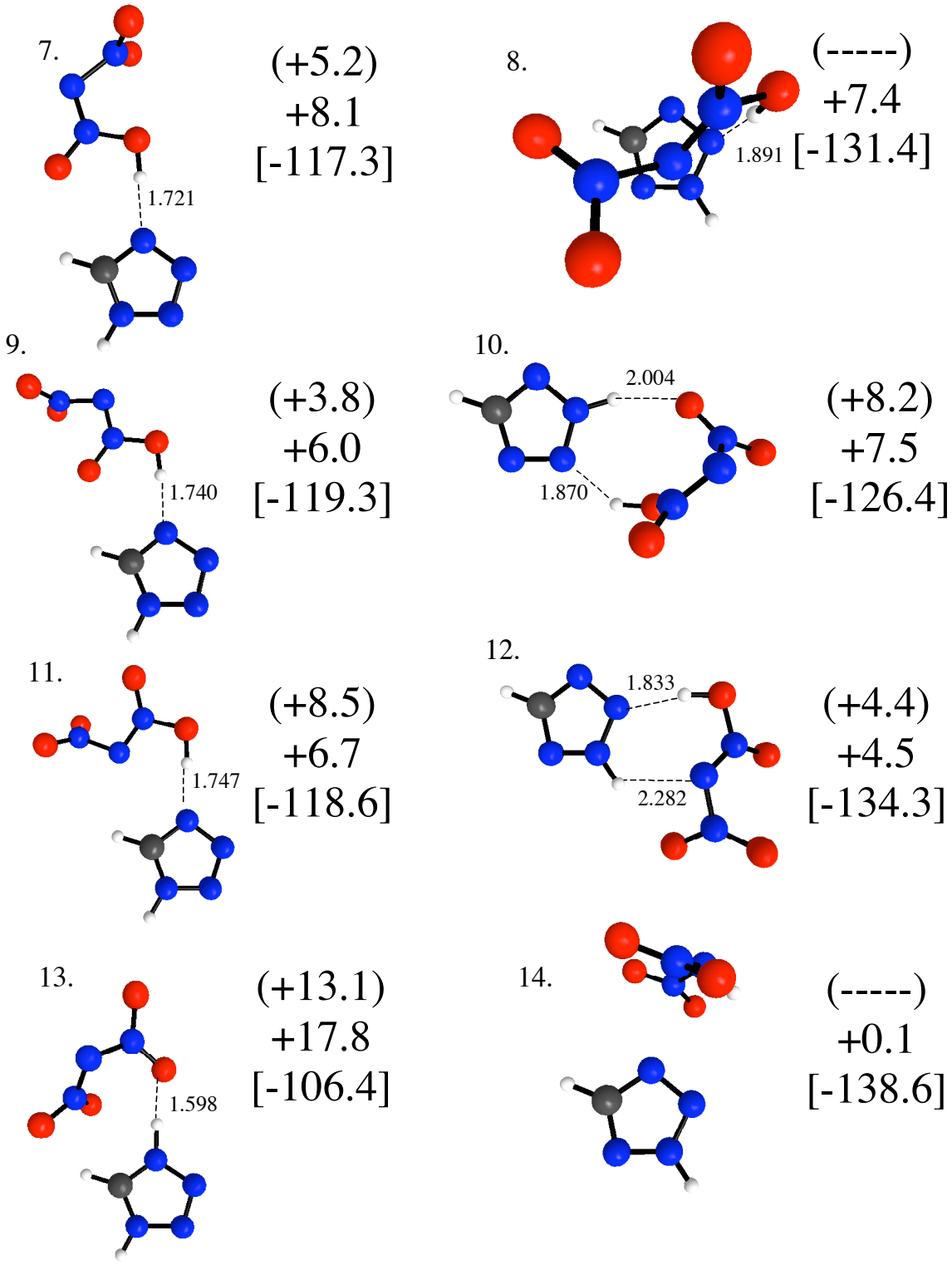


Figure 11: Continued

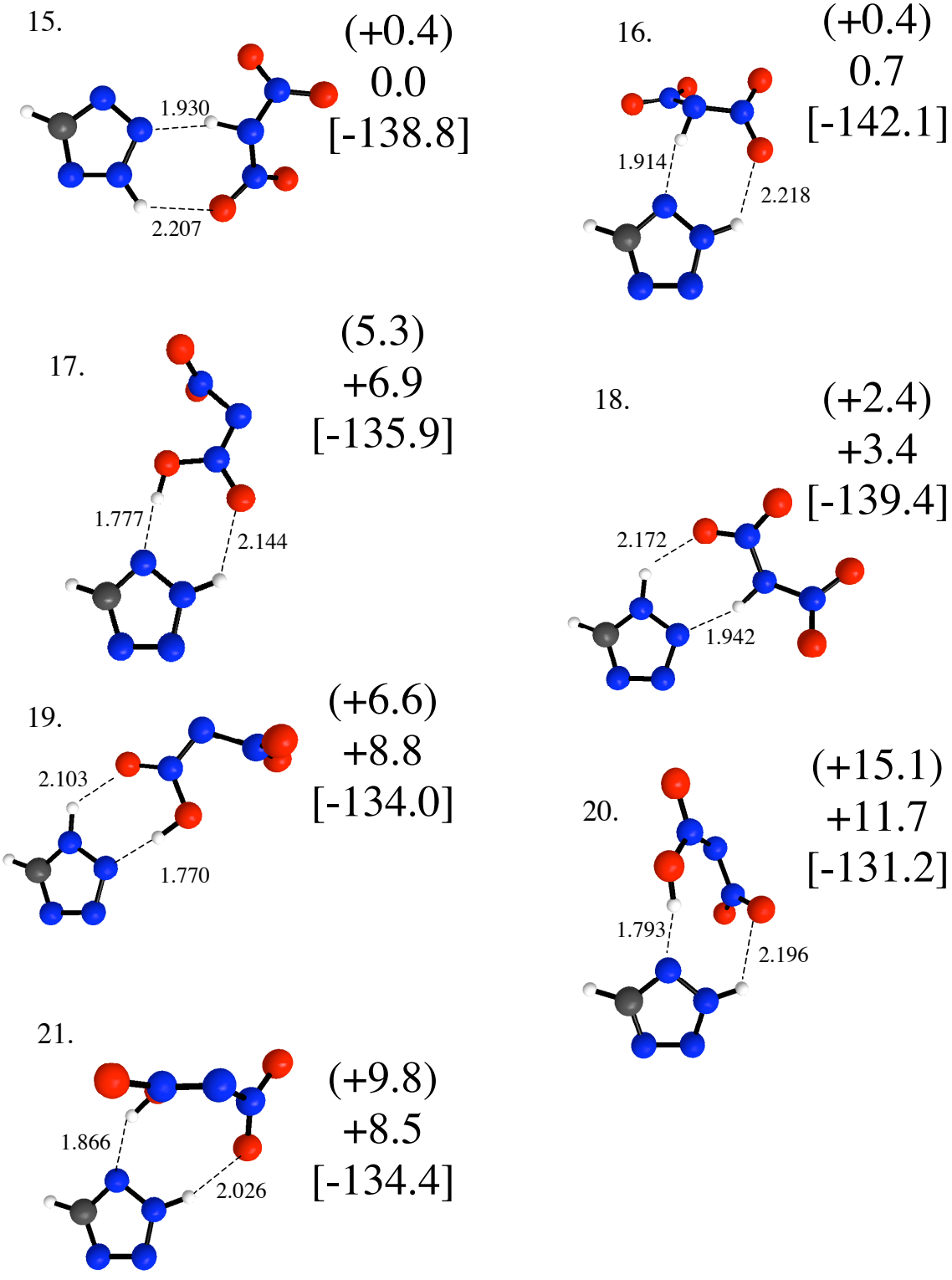


Figure 11: continued.

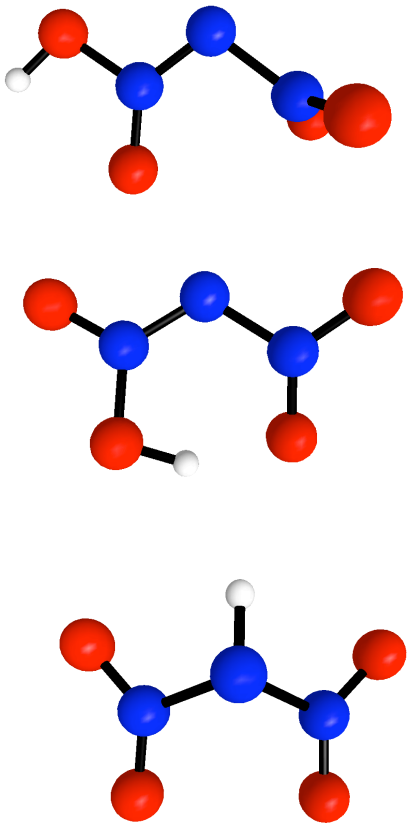


Figure 12: Three different geometries of dinitramic acid.

CHAPTER 3: BINDING AND DIFFUSION OF Al ADATOMS AND DIMERS ON THE Si(100)-2x1 RECONSTRUCTED SURFACE: A HYBRID QM/MM EMBEDDED CLUSTER STUDY

Deborah D. Zorn, Marvin A. Albao, James W. Evans, and Mark S. Gordon

Abstract. When group III metals are deposited onto the Si(100)-(2x1) reconstructed surface they are observed to self assemble into chains of atoms that are one atom high by one atom wide. These chains may have applications as atomic wires. In order to better understand the one-dimensional island growth of these systems, *ab initio* electronic structure calculations on the structures of Al atoms on silicon clusters have been performed. Natural orbital occupation numbers show that these systems display significant diradical character, suggesting that a multi-reference method is needed. A multi-configuration self consistent field (MCSCF) calculation with a 6-31G(d) basis set and effective core potentials was used to optimize geometries. The surface integrated molecular orbital molecular mechanics (SIMOMM) QM/MM embedded cluster method was used to take the surface chemistry into account, as well as the structure of an extended surface region. Potential energy surfaces for binding of Al adatoms and Al-Al dimers on the surface were determined and the former was used to assess the surface diffusion of adatoms. Hessians were calculated to characterize stationary points and improved treatment of dynamic electron correlation was accomplished using multi-reference second order perturbation theory (MRMP2) single point energy calculations. Geometries and relative energies from the MRMP2//MCSCF embedded cluster

calculations are compared with results from embedded cluster unrestricted density functional theory (UDFT) calculations and QM-only cluster calculations.

Introduction

The "holy grail" of nanotechnology is the self-assembly of nano-particles into structures with controlled size separation and electronic properties. The ultimate goal is to utilize these processes to build up molecular-devices from the atomic level. An obvious first step toward the creation of molecular devices is the study of molecular and atomic wires. Wires on the molecular scale can be made by epitaxial growth of metal atoms deposited on a metal or semi conductor surface.[1] One model system of interest is that of Group III and IV metals on the Si(100)-2x1 reconstructed surface. It has been shown that when deposited on the Si(100)–2x1 surface, Al, Ga, and In are able to self assemble into long chains.[2-6] The metal chains run perpendicular to the silicon surface dimer rows and are one atom high by one atom wide. Scanning Tunneling Microscopy (STM) images along with theoretical calculations have shown that the metal rows consist of metal dimers with each metal atom bonded to another metal adatom and two surface silicon atoms. This configuration, called the parallel dimer structure (Figure 1), has been confirmed by total energy calculations,[2, 7-9] a tensor LEED (low energy electron diffraction) study[10] and ion-scattering spectroscopy.[11]

Although self-assembly is a fundamental process, predictive modeling is a challenge, as behavior is often controlled by a combination of system-specific thermodynamics and kinetics. In particular, this is the case for deposition on silicon surfaces. However, theoretical

electronic structure calculations together with atomistic modeling and kinetic Monte Carlo (KMC) simulation can provide insight into the underlying adsorption and diffusion processes. Such KMC modeling has been performed to analyze the formation of atomic rows during deposition of group III metals, in particular Ga, on Si(100).[12,13] In that study the Si(100) surface was represented by a square lattice of adsorption sites. In the, KMC model, when pairs of diffusing atoms meet they irreversibly nucleate new islands. Aggregation was limited by only allowing diffusing metal atoms to bond at sites at the ends of metal rows. Sites adjacent to a metal adatom row were blocked prevent to diffusion. The simulations were able to match the experimental mean island size and reproduce the monotonically decreasing island size distribution using strongly anisotropic barriers for diffusion.

Al on Si(100) has been studied in detail with the Car-Parinello (CP) method[14] by Brocks, Kelly and Car (BKC). The CP study by BKC predicted the global minimum for one Al adatom on the surface to have the Al sitting on a silicon dimer row, directly between two silicon dimers (Figure 2a). This structure resembles an Al atom coordinated to four neighboring Si atoms. Another Car-Parinello study by Takeuchi[15] found two binding sites for Al, an offcenter binding site shown in Figure 2b and a pseudo-threefold binding site shown in Figure 2c. Takeuchi predicted that these two binding sites have the same total energy. So, the previous electronic structure calculations on the structures and energetics of these processes are in serious disagreement with each other. The BKC study predicted barriers for diffusion of an Al adatom on the surface. The CP predicted barrier for an adatom moving in the direction perpendicular to the silicon dimer rows is 2.3 kcal/mol. The barrier

for diffusion in the direction parallel to the silicon dimer rows is predicted to be 6.9 kcal/mol by the CP calculations. The present work employs well correlated *ab initio* electronic structure calculations in order to provide both quantitative and qualitative insight into the behavior of Al on the Si(100) surface.

When studying reactions on surfaces, a model must be used that takes into account the chemistry of the surface atoms as well as the affect of the bulk. In particular, a wave function must be used that adequately accounts for the diradical character of the silicon surface.[16-23] The CASSCF (Complete Active Space SCF) natural orbital occupation numbers (NOONs) indicate that for each surface dimer, nearly one third of an electron resides in an antibonding orbital.[23] Because of the computational cost associated with the use of a multi-reference wave function, a slab model that employs a plane wave basis is not currently feasible for such a wave function, so a cluster model is usually the method of choice. One disadvantage to cluster models is that edge effects can potentially impact the predicted outcomes. On the other hand, an advantage is that high level *ab initio* methods that can provide reliable structures and (especially) relative energies can be used. Advanced embedded cluster models have been developed to treat the reactive part of the surface with accurate electronic structure methods, while still including the bulk effects at some lower level of theory. The most successful form of these methods is a hybrid approach, in which the "action" region is treated with some level of quantum mechanics (QM), while the "bulk" region is treated with molecular mechanics (MM). These QM/MM embedded cluster methods diminish the edge effects of the cluster model[24] by greatly expanding the size of the cluster.

The embedded cluster method utilized in this study is the Surface Integrated Molecular Orbital/Molecular Mechanics (SIMOMM) method.[25] The "link region" that connects the QM atoms with the MM atoms is treated by capping the Si-Si broken bonds in the QM region with hydrogens. The SIMOMM method has been used successfully to study many different adsorbates on the Si(100)-2x1 surface,[26-29] as well as the SiC(100)[30] and diamond(100)[31] surfaces.

Computational Methods

All results, unless specified otherwise, were obtained using the QM/MM SIMOMM method. Most of the figures show only the QM region to save space. The two dimer, $Si_{15}H_{16}$, cluster was used to represent the reactive part of the silicon surface (Figure 3). The combined QM and MM system (Figure 4) has 12 surface dimers, is 11 layers deep, and contains a total of 199 Si atoms. QM structures were optimized using a complete active space self-consistent field (CASSCF) wave function. The orbitals included in the CASSCF wave function are the σ, π, π^*, and σ^* orbitals for each Si-Si dimer (Figure 5) and the 2s, $2p_x$, $2p_y$ and $2p_z$ orbitals on Al. The Stevens-Basch-Krauss-Jasien-Cundari (SBKJC) effective core potential basis set augmented with d polarization functions was used for all Si and Al atoms,[32] and the 6-31G basis set was used for H. Hessians (energy second derivatives) were used to characterize stationary points. Intrinsic reaction coordinate (IRC) calculations[33] were used to connect transition states with reactants and products. The IRC calculations were performed using the second order method developed by Gonzalez and

Schlegel (GS2)[34] using a step size of 0.3 $(amu)^{1/2}$ bohr. At the final MCSCF geometries, improved relative energies were obtained, using second order multi-reference perturbation theory (MRMP2)[35]. QM/MM unrestricted density functional theory (UDFT) calculations using the Becke three parameter Lee-Yang-Parr (B3LYP)[36-38] hybrid functional are compared with the multi-reference geometries and relative energies. The QM calculations were performed using the GAMESS electronic structure program.[39,40] The MM portion of the calculation was completely optimized using the MM3[41-43] parameters in the Tinker program.[44,45] All SIMOMM calculations have been carried out using the GAMESS/Tinker interface and all structures are visualized with MacMolPlt. [46]

Results and Discussion

Adsorption sites and diffusion of one Al adatom

A. Potential Energy Surface. Four adsorption sites are found when one Al is adsorbed on the $Si_{15}H_{16}$ embedded cluster (Figure 6). [Recall that the MM atoms are not shown, for clarity.] The off-center and pseudo-threefold sites are in agreement with the structures found by Takeuchi (Figure 2 b & c), but not with the Car-Parinello study by Brocks, Kelly, and Car: the structure in figure 2a, was not found to be a minimum in the present work.

The energies for the doublet sites are all within 7 kcal/mol of each other. The lowest energy site is the on-dimer site, which is 5.9 kcal/mol lower in energy than the on-top site.

The off-center and pseudo-threefold sites are 6.7 and 6.8 kcal/mol higher in energy than the on-dimer site, respectively. Relative energies for the different binding sites for an Al adatom on the Si(100) surface were not given in the studies by BKC or Takeuchi, so no comparison is possible.

Bond lengths for the four minima are shown in Table 1. The Si-Al bond lengths range from 2.43 Å to 2.78 Å. An Al adatom sitting on the Si-Si dimer bond in the on-dimer structure breaks the weak silicon π-bond, lengthening the Si-Si bond by 0.28 Å. The on-top structure lengthens the Si-Si dimer bond by only 0.15 Å. The off-center position also shortens the distance between the silicon dimers by 0.30 Å. The pseudo-threefold structure lengthens the dimer Si-Si bonds by 0.11 Å and 0.30 Å (Figure 6), relative to an isolated dimer. The pseudo-threefold structure also distorts the surrounding silicon dimers, by shifting the dimer with the longer bond length (Figure 6, pseudo-threefold structure, Si3-Si4) out of the silicon dimer row, and decreases the distance between the Si dimers by 0.32 Å. The length of the Si3-Si4 bond for the on-dimer and pseudo-threefold structures indicates that the σ bond is partially broken. This is an excellent example of why the σ orbitals must be included in the active space. In fact, when the σ-space is not included, the on-dimer position is found to be a transition state connecting the two on-top sites on a Si-Si dimer.

Orbitals at each minimum that displays a large amount of multi-reference character are shown in Figure 7. The associated natural orbital occupation numbers (NOONs) indicate diradical character for this system, similar to the bare silicon surface. The highest amount of diradical character is found in the off-center isomer with leading occupation numbers of 1.366 and 0.629 electrons. The NOONs for the quartet structures do not deviate significantly

from the ROHF occupation numbers of 2,1, and 0, suggesting that a multi-reference treatment is less critical for this state. The singly occupied orbitals of the quartet structures and their corresponding NOONs are shown in Figure 8.

The MCSCF minimum energy path for diffusion among the four minima on the embedded $Si_{15}H_{16}$ embedded cluster is shown in Figure 9. When diffusing from the on-dimer to the on-top site, the Si3-Al bond (see Figure 6 for atom numbering) is broken and the Si3-Si4-Al angle opens up from 59.6 to 113.8°. The Si3-Si4 bond length is also shortened by 0.13 Å as the σ-bond is reformed. A MCSCF transition state connecting the on-dimer and on-top structures could not be found. Constrained optimizations at points along a linear least motion (LLM) path were carried out to approximate the path connecting the two minima. Such a path provides an upper bound to the classical barrier height. Based on this procedure, the MCSCF barrier is less than 1 kcal/mol. MRMP2 single points along the MCSCF path predict diffusion to be continuously up hill from the on-dimer to the on-top site, suggesting that the on-top species may not be a local minimum on the potential energy surface.

An adatom can diffuse between two Si-Si dimers of the $Si_{15}H_{16}$ embedded cluster by moving from the on-top site to an off-center site and then over to the on-top site of the second dimer. MRMP2 energies calculated at the MCSCF geometries indicate that the transition state from on-top to off-center is lower in energy than both the on-top and off-center sites. This implies that MRMP2 potential energy surface (PES) is different from the MCSCF PES. To investigate this further, MRMP2 single point energy calculations were performed at two points along the MCSCF path near the transition state. These MRMP2 calculations suggest that the on-top and off-center structures are not local minima on the PES. The MCSCF transition state connecting the on-top and off-center sites is near a

MRMP2 local minimum. From this minimum there is a very low barrier (less than 1kcal/mol) to the on-dimer site.

Starting at the on-dimer site, the adatom can diffuse to the pseudo-threefold site. The MRMP2 barrier for this process is 0.3 kcal/mol higher in energy than the pseudo-threefold site. The barrier for diffusion between the two Si-Si dimers, from the off-center site to the pseudo-threefold site, is 11.5 kcal/mol higher in energy than the on-dimer site. This indicates an energy savings of approximately 4.5 kcal/mol for diffusion across a Si-Si dimer relative to diffusion between two Si-Si dimers. In order for an adatom to diffuse from one pseudo-threefold site into another pseudo-threefold site it must first diffuse back to either an on-dimer or an off-center site.

The quartet surface for an Al adatom on the $Si_{15}H_{16}$ embedded cluster was investigated in order to determine if there is surface crossing between the doublet and quartet surfaces. MCSCF optimizations to find quartet adsorption sites used the doublet structures as starting geometries. Quartet adsorption sites were found for the on-dimer, on-top and off-center structures. Bond lengths for these structures are given in Table 1. The pseudo-threefold structure does not exist on the quartet surface. The NOONs for the quartet active orbitals are shown in Table 2. These values do not deviate significantly from the ROHF occupation numbers of 2,1, and 0. The quartet singly occupied orbitals and their corresponding NOONs are shown in Figure 8.

Quartet structures are all higher in energy than the corresponding doublet structures by at least 4 kcal/mol. The quartet on-dimer structure is 14.9 kcal/mol above the doublet on-dimer structure; the quartet on-top structure lies 14.7 kcal/mol above the doublet on-top structure; and the off-center quartet lies 4.8 kcal/mol above the off-center doublet. There is

not a large difference in the Si3-Si4 and Si-Al bond lengths when comparing the doublet and quartet structures. For the quartet on-dimer and on-top structures Si1-Si2 is elongated by 0.17 Å due to an increase in electron density in its π^* orbital.

The red curve in Figure 9 illustrates diffusion processes that connect the three quartet minima. The on-top quartet site is 5.7 kcal/mol higher in energy than the on-dimer quartet site. At the MRMP2 level of theory diffusion from the on-dimer quartet site to the on-top quartet site is monotonically up hill. From the quartet on-top site the Al atom can diffuse off the silicon dimer into the off-center position. The off-center quartet site is 9.0 kcal/mol lower in energy than the on-top quartet site. The MCSCF barrier for diffusion from the quartet on-top site to the quartet off-center site is 3.7 kcal/mol higher in energy than the quartet on-top site. Once dynamic correlation is included at the MRMP2 level of theory, the barrier disappears and the transition state is 3.2 kcal/mol lower in energy than the on-top quartet site. As for the doublet case, MRMP2 single point energy calculations were performed at two points along the MCSCF path near the transition state. These energies indicate that the on-top structure is near a transition state and that diffusion is downhill to both the quartet on-dimer and on-top sites.

The curves in Figure 9 correspond to the MCSCF PES surface. To illustrate the difference in the MRMP2 and MCSCF surfaces, an approximates paths for MRMP2 diffusion on the doublet and quartet surfaces are shown in Figure 10. The on-dimer and pseudo-threefold structures are minima on both the MRMP2 and MCSCF doublet surface. On the MRMP2 doublet surface the on-top structure is no longer a minimum and the transition state structure connecting the on-top and off-center structures now appears to be near a minimum on the MRMP2 doublet PES. The on-dimer and off-center quartet

structures are still minima on the MRMP2 quartet PES, which now appear to be connected by a transition state near the on-top structure.

B. Comparison of Methods. Unrestricted density functional theory geometry optimizations with the B3LYP functional (UB3LYP) were carried out on the four doublet minima for one Al on the surface. This method predicts the energy differences between minima to be less than 7.7 kcal/mol, while the MRMP2 energy range is predicted to be less than 6.8 kcal/mol. However, the two methods predict different energy orders for the four minima. The relative MRMP2//MCSCF energies are: on-dimer < on-top < off-center < pusdo-threefold, whereas UB3LYP predicts: off-center < on-dimer < pseudo-threefold < on-top. So, B3LYP does not capture the correct ordering, probably due to the varying multi-reference character of the various species. UB3LYP also consistently underestimates the bond lengths with a RMS error of 0.05 Å when compared with MCSCF bond lengths. Bond lengths for the four UB3LYP minima are shown in Table 2. UB3LYP does an adequate job of reproducing bond angles. The only structures with any significant amounts of distortion are the on-top and the pseudo-threefold structures. The on-top UB3LYP Si3-Si4-Al angle is 8.7 degrees too small and the Si2-Si4-Al angle is 5.2 degrees too small. The pseudo-threefold Si2-Si4-Al angle is 3.1 degrees too small.

Al2 Potential Energy Surface

It is thought that pairs of diffusing Al adatoms will irreversibly nucleate a new island when they meet in the same row. An Al-Al dimer can form in several positions on the

Si(100) surface. The formation, relative energies, and rotation of the Al-Al dimer on the $Si_{15}H_{16}$ embedded cluster are presented in the following sections.

Dimer minima. All minima found for two Al adatoms on the $Si_{15}H_{16}$ embedded cluster are shown in Figure 11. Table 3 lists the bond lengths for these structures. Four singlet minima (Figure 11- **I**, **II**, **III**, and **IV**) are found for an Al-Al dimer on the $Si_{15}H_{16}$ cluster. The "bridge" (**IV**) structure is lowest in energy, with the "between" (**I**) structure 6.6 kcal/mol higher in energy. Structures **I** and **IV** correspond to the structures in Figure 2(c) and 2(d) that were reported by BKC. The relative energies of **I** and **IV** agree qualitatively with the Car-Parinello calculations performed by BKC[7]. The other possible dimer structures presented in BKC were not investigated in this work.

The Al-Al bond lengths for the bridge and between structures are 2.76 Å and 2.64 Å respectively. The Si-Al bond lengths for the between structure are 0.09 Å shorter than for the on-dimer structure. The Si-Si bond distance in the between structure is also 0.07 Å shorter than the on-dimer structure indicating that an Al-Al dimer does not break the σ Si-Si bond to as great an extent as a single Al adatom. The NOONs of the HOMO and LUMO orbitals are given in Figure 12. Both the bridge and between structures are single reference with less than 0.1 electrons in an anti-bonding orbital.

The two cross structures, (Figure 11-**II** and **III**) are 18.5 and 11.6 kcal/mol higher in energy than the bridge (**IV**) structure, respectively. These structures have not been previously reported. The Al-Al bond length of **II** is 2.67 Å and the Al-Al bond length of **III** is 2.65 Å. The Si1-Si2-Al bond angle of structure **II** is 61.7°. The strain due to this small angle causes the decreased stability of **II** relative to **III**. The high energies of **II** and **III** is

most likely due to the inability of the Al-Al dimer to passivate the surface dimers in this position. Their NOONs indicate diradical character similar to the bare silicon surface. Two minima with no bond between the two Al atoms were found on the singlet potential energy surface (Figures 11 - **V** and **VI**). Both of these structures are significantly higher in energy than the bridge structure. **V** is a pure diradical and **VI** has diradical character similar to that of the bare silicon surface, as shown in Figure 12.

Only two Al-Al dimer minima were found on the triplet surface. These are the cross structures (Figures 11-**VII** and **VIII**), which have geometries similar to the singlet cross structures Figures 11-**II** and **III**, but are 9.7 and 6.3 kcal/mol higher in energy respectively. The Al-Al bond length in **VII** is similar to that in the singlet structure, however in **VIII** the Al-Al bond length is elongated by 0.13 Å. When the between (**I**) and the bridge (**III**) structures were optimized using a triplet wave function, their geometries distorted to the cross structure (between, 10-**II**) and a dual-off-center structure (Figure 11-**XI**). The dual-off-center structure is 18.8 kcal/mol higher in energy than the bridge dimer. Two other separated structures were found on the triplet surface (Figures 11-**IX** and **X**). The structure **IX** is degenerate with the singlet structure **V** and structure **X** is 4.1 kcal/mol higher in energy than the singlet structure **VI**. Some triplet structures are lower in energy than the corresponding singlet structures, indicating that a further investigation of the PES will be necessary to determine if singlet-triplet crossings are likely for Al_2 on Si(100).

HOMO and LUMO orbitals for the singlet and triplet minima are shown in Figures 12 and 13, respectively. **I** and **IV** (Figure 12) are both single reference structures, due to the saturation of the surface by the Al-Al dimer, and the formation of the Al-Al bond. The singlet cross structures (Figures 12-**II** and **III**) are multi-reference, with NOONs of about 1.4

and 0.6 electrons. All singlet structures without an Al-Al bond display at least some multi-reference character. The dual-on-top (Figure 12-**V**) structure is almost a pure diradical, hence the singlet and triplet spin states are nearly degenerate. The remaining Al_2 separated structure is also multi-reference. The Al_2 triplet structures are single reference. Although the lowest energy Al_2 dimer structures are single reference, when studying diffusion, it is important to take into account the multi-reference character of significant cross sections of the surface.

Diffusion of two separated adatoms. Figure 14 illustrates how two separated Al adatoms can diffuse on the $Si_{15}H_{16}$ embedded cluster before forming a dimer. The highest energy position for two separated Al adatoms on the doublet surface is **V**, where each Al is in an on-top position. Starting from structure **V** the atom in the on-top position can diffuse to a off-center position giving structure **VI** (Figure 14). This lowers the energy of the system by 11.4 kcal/mol. At this point the two Al atoms are 4.87 Å apart. If the two atoms on the singlet surface move any closer together they will bond to form a dimer. In contrast, on the triplet surface the atom in the on-top position can diffuse to the off-center position to form **XI**. **XI** is 18.8 kcal/mol above **IV** and is the lowest energy structure for separated Al atoms on the triplet surface.

Dimer formation. Figure 15 shows how the adatoms can diffuse to form an Al-Al dimer on the $Si_{15}H_{16}$ embedded cluster. As shown in Figure 15, there is a large thermodynamic advantage for forming a dimer on the surface in the bridge position (Figure 11-**IV**). The binding energy of two radical Al adatoms infinitely separated on the surface is 11.8 kcal/mol. This is a decrease of 5.5 kcal/mol from the experimental binding energy of 17.3 for gas phase Al_2.[47] According to BKC, the effective binding energy of two Al

adatoms, in the center position, forming a dimer in **I** was calculated to be 19.6 kcal/mol. This is 2.3 kcal/mol higher in energy than the gas phase binding energy. The high binding energy in BKC is most likely due to the inability of the CP method to accurately represent the multireference character of the separated Al adatoms.

On the MCSCF surface there is a barrier for diffusion between structures **VI** and **III** (Figure 11). However once dynamic correlation is added via MRMP2, the transition state is 2.9 kcal/mol lower in energy than **VI**. This indicates that the dimer should form easily once two aluminums are on adjacent silicon dimers.

Dimer rotation. Figure 16 illustrates the path for rotation of an Al-Al dimer relative to the $Si_{15}H_{16}$ embedded cluster. The singlet structure **I** can rotate to the lowest energy position **IV** by traveling through local minima **II** and **III** (see Figure 11). The height of the barrier for rotation between **I** and **II** is 28.5 kcal/mol, making this step even more energy intensive than breaking an Al-Al bond, as in Figure 15. In the parallel dimer model (Figure 1) Al-Al dimers are parallel to the Si-Si surface dimers. In order for **I** to rotate into position **IV** the Al-Al bond must either break or pass through the high barriers for rotation to **IV**. Therefore it seems unlikely that an island could be irreversibly nucleated in **I**. On the MCSCF surface, there is a barrier connecting the two triplet cross structures (Figure 11-**VII** and **VIII**). However once dynamic correlation is added via MRMP2, the MCSCF transition structure is 1.7 kcal/mol lower in energy than **VII**.

Importance of Bulk Atoms

In order to determine the importance of including the surrounding bulk (MM) atoms, MCSCF calculations were carried out for QM-only $AlSi_{15}H_{16}$ and $Al_2Si_{15}H_{16}$ clusters. As

discussed below, bulk effects are important for determining relative energies and accurate geometries.

A. $AlSi_{15}H_{16}$. The MRMP2//MCSCF relative energies with no MM bulk atoms are: pseudo-threefold < on-dimer < off-center < on-top < center. The bond lengths for the QM-only cluster geometries are given in Table 4. Including the surrounding bulk is also important for determining accurate geometries. The RMS error in bond lengths for the QM-only geometries is 0.20 Å. The RMS error in angles is 2.8°. The most significantly distorted QM-only geometry is the pseudo-threefold structure (Figure 17a). In the QM/MM pseudo-threefold geometry, the Si1-Si2 dimer is shifted 0.26 Å out of the silicon dimer row, but without the MM part, this shift is 0.61 Å. The distance between Si2 and Si4 is also shortened by 0.10 Å. So, the MM part of the structure imposes some constraints on the movements of the QM atoms, even though all atom positions are optimized in SIMOMM.

The off-center doublet (Figure 17b) and quartet structures are also distorted. In both structures the Si2-Si4 distance is shortened by 0.12 Å. In the on-dimer QM-only structure the Si3-Si4 bond length is too short by 0.03 Å (Figure 17c). The QM-only center structure does not even exist when the silicon bulk is included (Figure 17d).

B. $Al_2Si_{15}H_{16}$. QM-only geometries were found for the bridge and triplet structures. Both QM/MM and QM-only calculations find the singlet bridge structure to be lower in energy than the singlet between structure. QM-only geometries for the doublet structures are in reasonable agreement with the SIMOMM results. The between triplet structure calculated without the surrounding bulk atoms is symmetric, but the between triplet structure distorts to a cross structure when the surrounding bulk silicon is included (Figure 18a). The QM-only

triplet structure has a bond between the two Al atoms, but when the bulk silicon is included in the optimization, the Al-dimer bond disappears, forming a dual off-center structure (Figure 18b). The RMS error in bond lengths for the four Al-Al dimer QM-only structures is 0.38 Å.

Conclusions

Pathways for short-range diffusion and rotation of Al adatoms and dimers have been successfully mapped out using the SIMOMM QM/MM method. The effects of these processes on the structure of the surface and the electronic structure of the system have been discussed. Four minima on the $Si_{15}H_{16}$ embedded cluster have been identified. The four minima are the on-dimer, on-top, off-center and pseudo threefold sites. The lowest energy site for a single Al adatom is the on-dimer structure. No minimum on the MCSCF surface was found for an adatom sitting in the center site on the $Si_{15}H_{16}$ cluster, which was the global minimum in the BKC CP study. Another CP study by Takeuchi found two binding sites for an Al adatom on the surface. These were the off-center and pseudo-threefold sites and in agreement with the multireference calculations in this work, the center structure was not found to be a minimum. Relative energies of the off-center and pseudo-threefold sites were not presented in that work. Neither the Takeuchi nor BKC study addressed the chemistry of the quartet PES.

It is not physically relevant to directly compare the multireference barrier heights found in this work to the CP barrier heights from the BKC study, as the CP barriers pertain to long-range diffusion and are relative to the energy of the center structure, which was not found on the MCSCF surface; however some general comparisons are possible. The BKC study found that the low energy pathway for diffusion perpendicular to the silicon dimer

rows is between two silicon dimer rows. Both the MRMP2 and MCSCF PES in this work indicate that diffusion between the two silicon dimers of the $Si_{15}H_{16}$ cluster is limited by the high barrier to the peudo-threefold site. The MCSCF surface indicates that it is more likely for an adatom to diffuse across a silicon dimer, when attempting to diffuse in the direction perpendicular to the silicon dimer surface. Diffusion across a Si-Si dimer was not specifically addressed in BKC. BKC found the barriers for diffusion parallel to the silicon dimer rows (between two dimer rows) to be higher in energy than diffusion perpendicular to the silicon dimer rows. The MRMP2 barrier for diffusion on the $Si_{15}H_{16}$ cluster, parallel to the silicon dimer rows (on-top to off-center), is approximately 1 kcal/mol higher in energy than the barrier for diffusion across a Si-Si dimer in the direction perpendicular to the dimer rows (on-top to on-top), but much lower than the barriers for diffusion between the two silicon dimers.

The lowest energy position for the Al_2 dimer on the $Si_{15}H_{16}$ embedded cluster is the bridge structure. The lowest energy Al_2 structures are those that can passivate the diradical nature of the surface. The more diradical character a minimum has, the higher in energy it is. Barriers for rotation of an Al-Al dimer from the between structure to the bridge structure are very high, indicating that a dimer formed in the between position could not nucleate an island. When two Al adatoms are placed separately on the $Si_{15}H_{16}$ there is no barrier on the MRMP2 surface for dimer formation, indicating that two Al adatoms will spontaneously self assemble into dimers when they meet on neighboring dimers in the same silicon dimer row. Dimerization occurs even when the adatoms are sitting as far apart as possible on the $Si_{15}H_{16}$ embedded cluster (Figure 11-Structure **V**). In Structure **V**, the two Al adatoms are over 6 Å

apart, indicating the importance of non-covalent interactions between the Al adatoms in the self-assembly process.

Comparisons of QM-only with embedded cluster QM/MM calculations suggest that it is necessary to include the surrounding bulk silicon atoms to obtain correct geometries and relative energies. An unrestricted hybrid DFT method does not predict the correct order of stability for the isomers. This can be explained by the inability of a single reference method such as DFT to account for the varying multi-reference character of the surface. It will be necessary to study larger clusters with more QM dimers to obtain a complete picture of the diffusion of Al adatoms and dimers between the silicon dimer rows. When larger embedded clusters are investigated, long-range diffusion could be investigated using molecular dynamics or dynamic reaction coordinate[48] calculations. The effect the effect of a neighboring QM dimer rows on the on-top and off-center structures should also be investigated. Future studies could also address the effects surface features such as steps and defects on the reactivity of the surface for better comparison with experiment.

Unfortunately MCSCF calculations on larger clusters is currently not manageable due to the large active spaces that are required. To reduce the time and memory requirements of these calculations, while still including all the important configurations, the active space can be subdivided into multiple subspaces using the ORMAS method developed by Ivanic.[49] Calculations on large silicon clusters up to 7 dimers[50] are possible using the ORMAS code in GAMESS. Preliminary studies also show this method to be promising.

Acknowledgements

This work was supported by the SicDAC Computational Chemistry Program and Division of Chemical Sciences, Basic Energy Sciences, US Department of Energy (USDOE). The work was performed at Ames Laboratory, which is operated for the USDOE by Iowa State University under Contract No. DE-AC02-07CH11358. The authors also acknowledge Drs. Mike Schmidt, Jamie Rintelman and Professors Cheol Ho Choi and Tim Dudley for many helpful discussions.

References

(1) Joachim, C.; Roth, S. (eds) 1997 Atomic and Molecular Wires (NATO Adv. Series E: Appl. Sci. vol 341) (Dordrecht: Kluwer)

(2) Nogami, J.; Baski, A.A.; Quate, C.F. *Physical Review B*. **1991**, *44*, 1415.

(3) Nogami, J.; Park, S.-I.; Quate, C. F., *Appl. Phys. Lett*. **1988**, *53*, 2086.

(4) Nogami, J., in Atomic and Molecular Wires, edited by C. Joachim and S. Roth (Kluwer, Dordrecht, **1997**, *341*, 11.

(5) Evans, M. M. R.; Nogami, J. *J. Phys. Rev. B*, **1999**, *59*, 7644.

(6) Itoh H.; Itoh J.; Schmid A.; Ichinokawa, T. *Phys. Rev. B* **1993**, *48*, 14663.

(7) Brocks, G.; Kelly, P.J.; Car R., *Phys. Rev. B* **1993**, *70*, 2786.

(8) Northrup, J.E.; Schabel, M. C.; Karlsson, C. J.; Uhrberg, R.I.G. *Phys. Rev. B* **1991**, *44*, 13799.

(9) Adams, B.G.; Sankey, O.F. *Journal of Vacuum Science & Technology A* **1992**, *10*, 2046.

(10) Sakama, H; Murakami, K.; Nishikata, K.; Kawazu, A. *Phys. Rev. B*, **1994**, *50*,

14977.

(11) Steele, B.E.; Li, L.; Stevens, J.L.; Tsung, I.S.T. *Phys. Rev. B* **1993**, *47*, 9925.

(12) Albao, M. A.; Evans, M. M. R.; Nogami, J.; Zorn, D.; Gordon, M. S.; Evans, J. W. *Phys. Rev. B,* **2005**, *72*, 035426.

(13) Albao, M. A.; Evans, M. M. R.; Nogami, J.; Zorn, D.; Gordon, M. S.; Evans, J. W. *Phys. Rev. B*. **2006**, *74*, 037402.

(14) Car, R.; Parrinello, M. *Phys. Rev. Lett*. **1985** *55*, 2471.

(15) Takeuchi, N. *Phys. Rev. B*. **2000,** *63*, 035311.

(16) Redondo, A.; Goddard III, W. A. *J. Vac. Sci. Tech. B*. **1982**, *21*, 344.

(17) Jung, Y., Shao, Y., Gordon, M.G., Doren, D.J., and Head-Gordon, M., *J. Chem. Phys*., **2003**, *19*, 10917.

(18) Shoemaker, J.; Burggraf, L.W.; Gordon, M.S. *J. Chem. Phys*. **2000**, *112*, 2994.

(19) Gordon, M.S., Shoemaker, J.R.; Burggraf, L.W., *J. Chem. Phys*. **2000,** *113*, 9355.

(20) Paulus, B., *Surf. Sci*. **1998**, *408*, 195.

(21) Choi, C.H.; Gordon, M.S., *J. Am. Chem. Soc*. **1999**, *121*, 11311.

(22) Jung, Y.; Choi, C.H.; Gordon, M.S. *J. Phys. Chem. B* **2001**, *105*, 4039.

(23) Jung, Y.; Akinaga, Y.; Jordan, K.D.; Gordon, M.S. *Theor. Chem. Acc*. **2003**, *109*, 268.

(24) Choi, M.S.; Gordon, M.S. in 'The Chemistry of Organic Silicon Compounds', Vol. 3; Zvi Rappoport, Yitzhak Apeloig; Eds; John Wiley and Sons; New York, **2001**; Ch. 15, pp 821-852

(25) Shoemaker, J.R.; Burggraf, L.W.; Gordon, M.S. *J. Phys. Chem. A*, **1999**, *103,* 3245.

(26) Rintelman, J.M.; Gordon, M.S. *J. Phys. Chem. B* **2004,** *108*, 7820.

(27) Jung, Y.; Gordon, M.S. *J. Am. Chem. Soc*. **2005**, *127*, 3131.

(28) Choi, C.H.; Liu, D., Evans, J.W.; Gordon, M.S. *J. Am. Chem. Soc*. **2002**, *124*, 8730.

(29) Choi, C.H.; Gordon, M.S. *J. Am. Chem. Soc*., **2002,** *124*, 6162.

(30) Tamura, H.; Gordon, M.S. *J. Chem. Phys*. **2003**, *119*, 10318.

(31) Zapol, P.; Curtiss, L.A.; Tamura H.; Gordon, M.S. "Theoretical Studies of growth Reactions on Diamond Surfaces", in Computational Materials Chemistry: Methods and Applications, L.A. Curtiss and M.S. Gordon, Eds. pp. 266-307 (**2004**).

(32) W. J. Stevens, M.K., H. Basch, P.G. Jasien, *Canad. J. Chem*., **1991***, 70*, 612

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

(34) Gonzales, C.; Schelgel, H.B. *J. Chem. Phys*. **1998**, *90*, 2154.

(35) Hirao, K. *Chem. Phys. Lett*. **1992**, *190*, 374.

(36) Becke, A.D. *J. Chem. Phys*. **1993**, *98***,** 5648.

(37) Lee C.; Yang, W; Parr, R.G. *Phys. Rev. B* **1998**, *37*, 785.

(38) Vosko, S.H.; Wilk, L.; Nusair, M. *Can. J. Phys*. **1980**, *58* 1200.

(39) Schmidt, M.W., et al. *Journal of Computational Chemistry* **1993**, *14*, 1347.

(40) Schmidt, M. W.; Gordon, M. S. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, 2005.

(41) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc*. **1989**, *111*, 8566,.

(42) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc*. **1989**, *111*, 8576.

(43) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc*. **1989**, *111,* 8551.

(44) Kundrot, C. E.; Ponder, J.W.; Richards, F.M. *Journal of Computational Chemistry*

1991, *12*, 402.

(45) Ponder, J.W.; Richards, F.M. *Journal of Computational Chemistry* **1987**, *8*, 1016.

(46) Bode, B. M.; Gordon, M. S. *Journal of Molecular Graphics & Modeling*, **1999**, *16*, 133.

(47) Huber, K.P., Herzberg G (1979) Molecular spectra & molecular structure, constants of diatomic molecules. Van Nostrand-Reinhold, NewYork

(48) Stewart, J. J. P.; Davis, L. P.; Burggraf, L. W. *J. Comput. Chem.* **1987**, *8*, 1117.

(49) Ivanic, J. *J. Chem. Phys.* **2003**, *119*, 9364.

(50) Roskop, L., Gordon, M.S., in preparation.

Table 1: MCSCF bond lengths in angstroms for Al doublet and quartet minima. (Structures and atom numbering shown in Figure 6.)

	Si1-Si2	Si3-Si4	Si2-Al	Si3-Al	Si4-Al
On-dimer doublet	2.31	2.60		2.57	2.57
On-top doublet	2.31	2.47			2.63
Off-center doublet	2.46	2.42	2.65		2.70
Pseudo-threefold doublet	2.43	2.62	2.78	2.61	2.43
On-dimer quartet	2.47	2.60		2.57	2.57
On-top quartet	2.47	2.47			2.63
Off-center quartet	2.47	2.47	2.64		2.64

Table 2: UB3LYP bond lengths in angstroms for Al doublet minima. (Atom numbering shown in Figure 6.)

	Si1-Si2	Si3-Si4	Si2-Al	Si3-Al	Si4-Al
On-dimer doublet	2.28	2.56		2.53	2.53
On-top doublet	2.29	2.43			2.60
Off-center doublet	2.42	2.39	2.60		2.62
Pseudo-threefold doublet	2.39	2.58	2.67	2.57	2.41

Table 3: MCSCF bond lengths in angstroms for Al_2 minima. (Structures and atom numbering shown in Figure 11.)

	Si1-Si2	Si3-Si4	Al1-Al2	Si1-Al1	Si1-Al2	Si2-Al1	Si3-Al2	Si4-Al1	Si4-Al2
I	2.53	2.53	2.64	2.44		2.44	2.44		2.44
II	2.47	2.37	2.67	2.54		2.49	2.62		
III	2.44	2.44	2.65		2.59	2.59	2.59	2.59	
IV	2.42	2.37	2.76		2.76	2.52	2.68		
V	2.43	2.43				2.63	2.63		
VI	2.38	2.44				2.66	2.62	2.69	
VII	2.50	2.43	2.68	2.46		2.48	2.53		
VIII	2.42	2.42	2.78		2.73	2.53	2.66		
IX	2.43	2.43				2.63	2.63		
X	2.43	2.45				2.64	2.63	2.67	
XI	2.44	2.44			2.63	2.63	2.63	2.63	

Table 4: QM cluster model MCSCF bond lengths in angstroms for Al doublet and quartet minima. (Atom numbering shown in Figure 6.)

	Si1-Si2	Si3-Si4	Si2-Al	Si3-Al	Si4-Al
On-dimer doublet	2.30	2.57		2.57	2.57
On-top doublet	2.30	2.44			2.63
Off-center doublet	2.43	2.41	2.61		2.62
Pseudo-threefold doublet	2.43	2.64	2.57	2.48	2.44
On-dimer quartet	2.45	2.58		2.57	2.57
On-top quartet	2.45	2.44			2.63
Off-center quartet	2.44	2.44	2.59		2.59

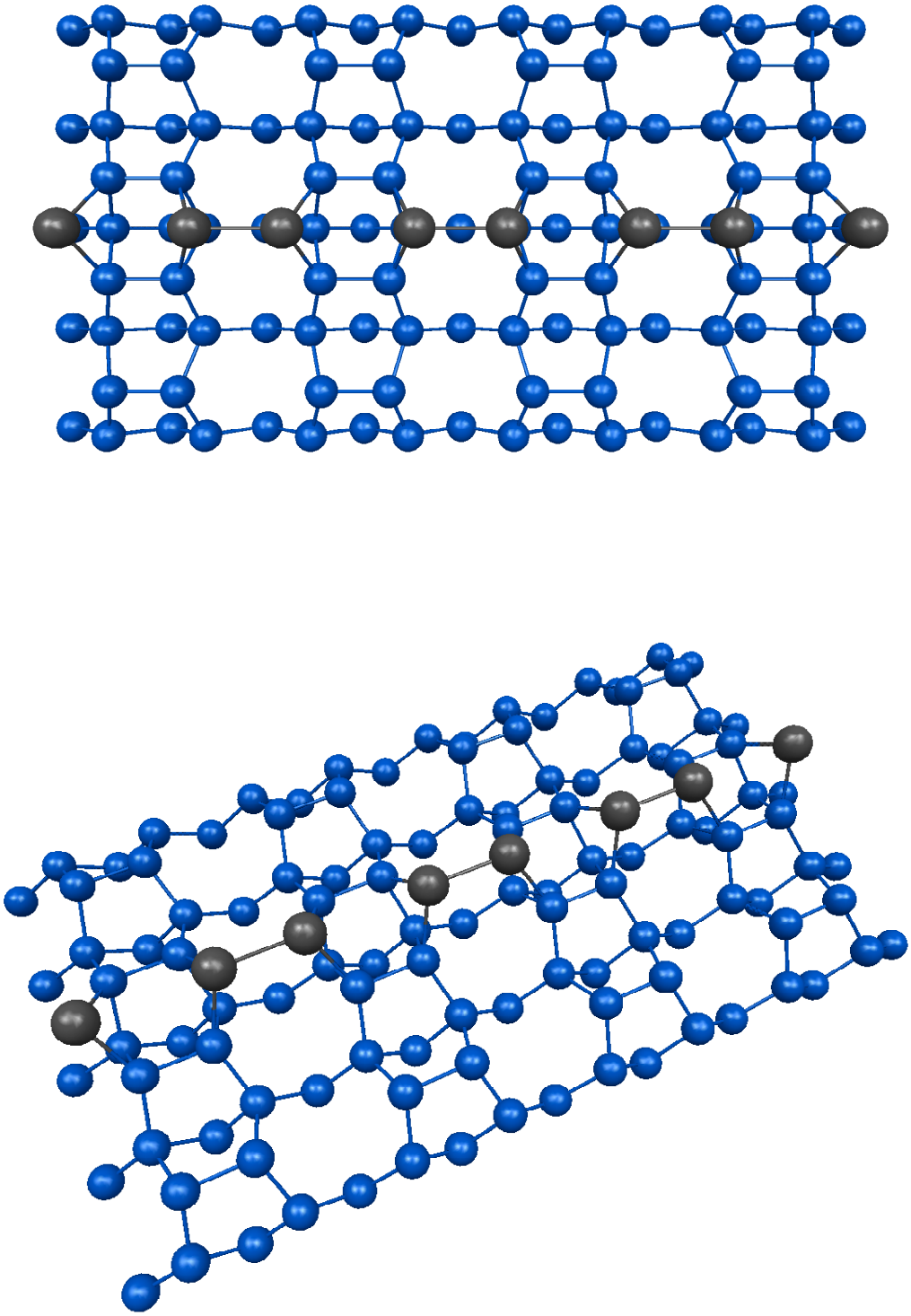


Figure 1. Schematics of the parallel dimer structure.

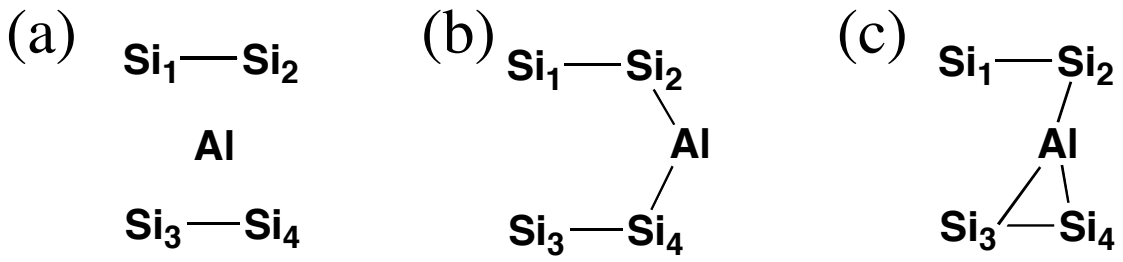


Figure 2. (a)global minimum for one Al on the Si(100) surface predicted by BKC. (b) and (c) binding sites for one Al predicted by Takeuchi.

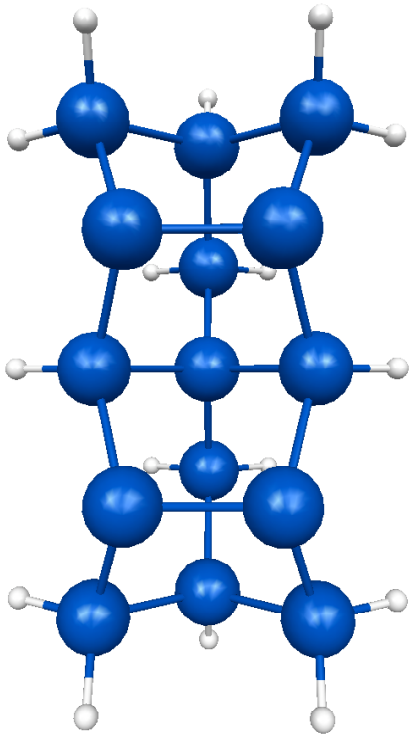


Figure 3. $Si_{15}H_{16}$ QM cluster.

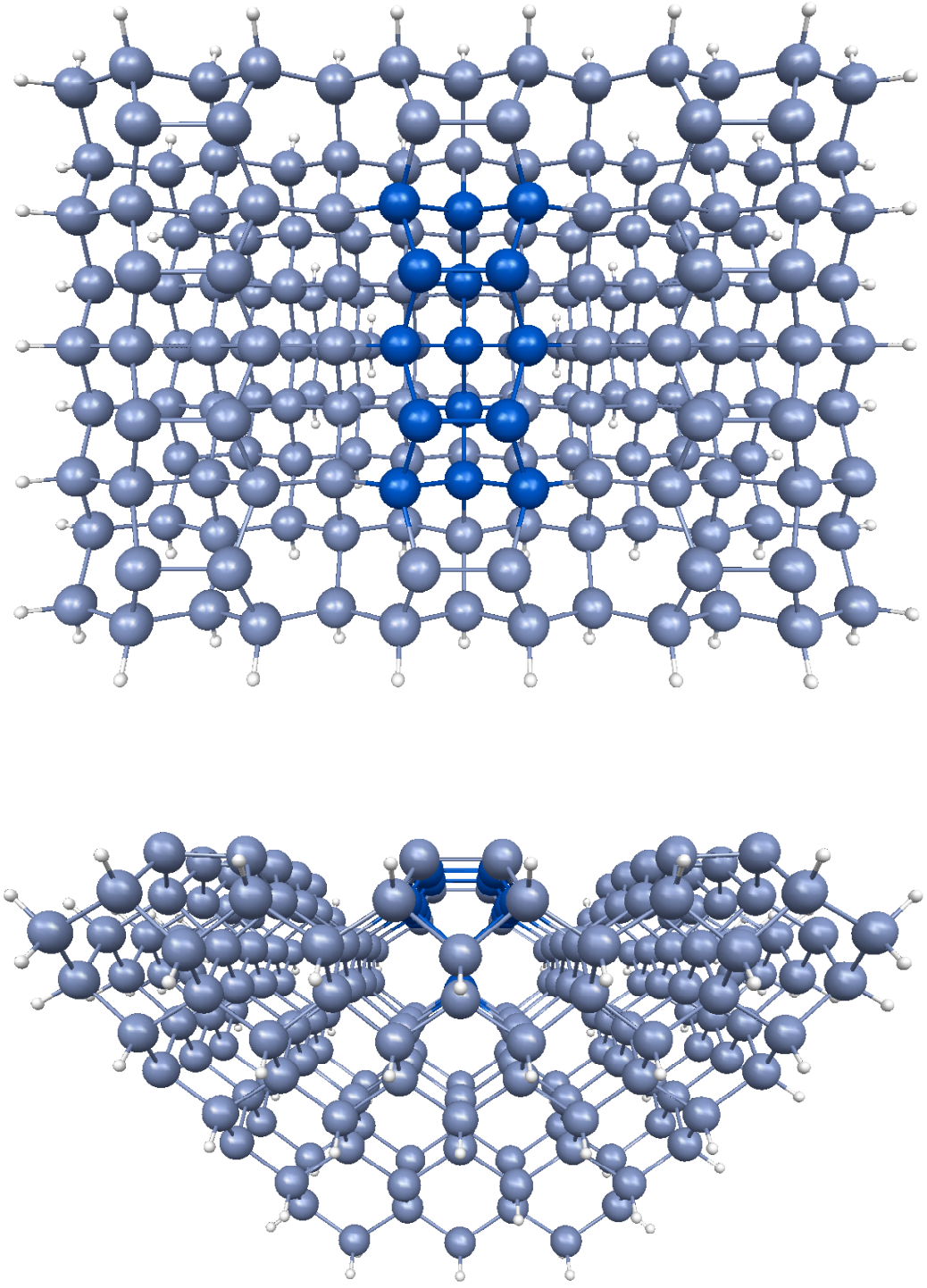


Figure 4. Bulk Silicon Cluster top and side views.

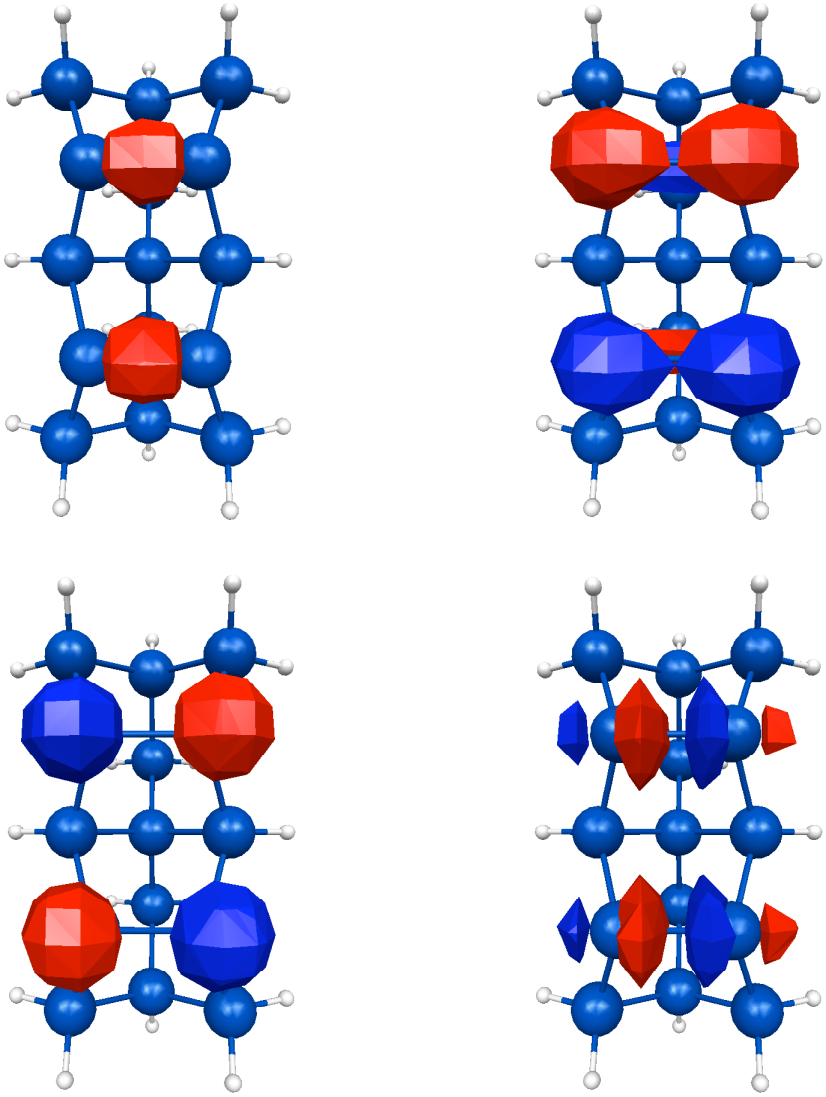


Figure 5. σ, π, π*, and σ* orbitals for each Si-Si dimer.

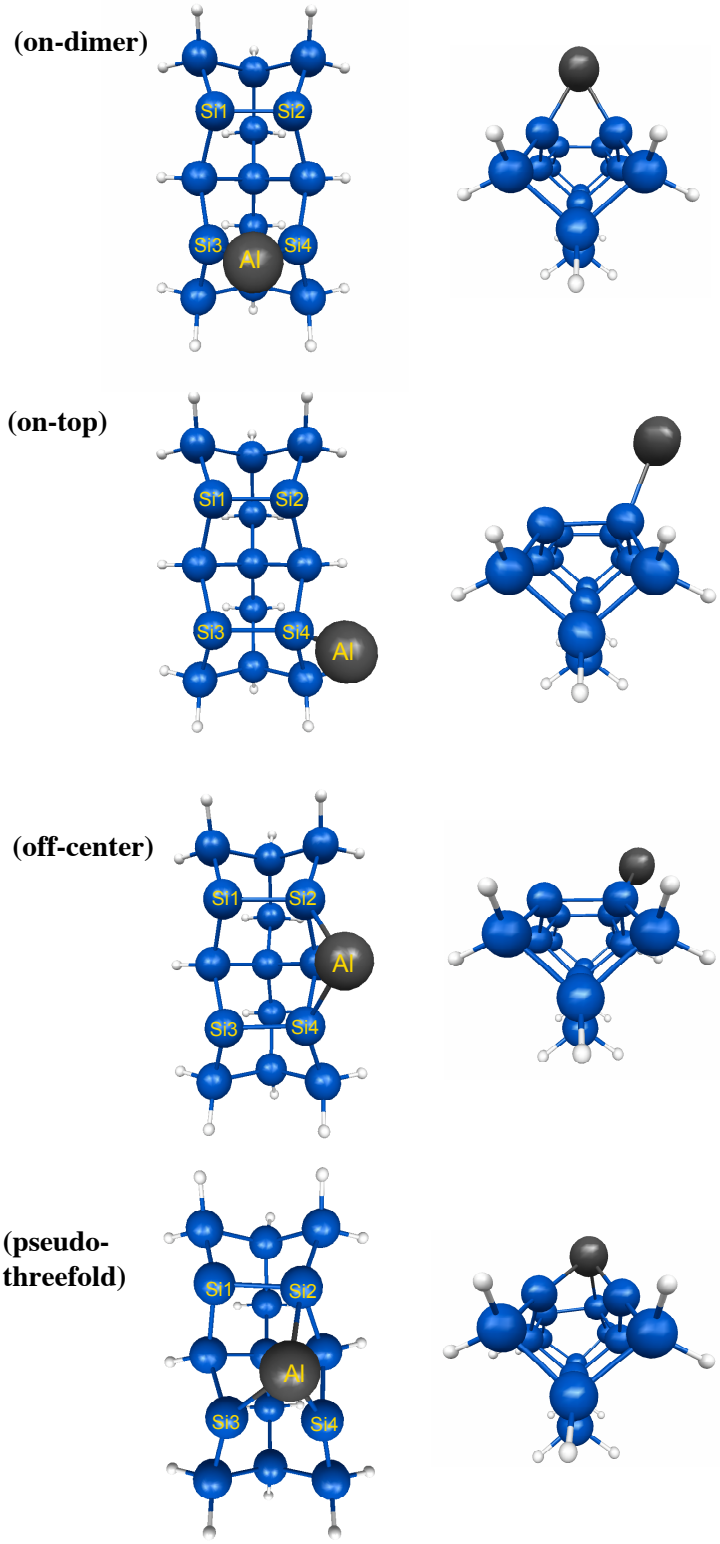


Figure 6. Binding sites for 1 Al. The QM region of the cluster is shown without the surrounding bulk cluster. Bond lengths are given in Table 1.

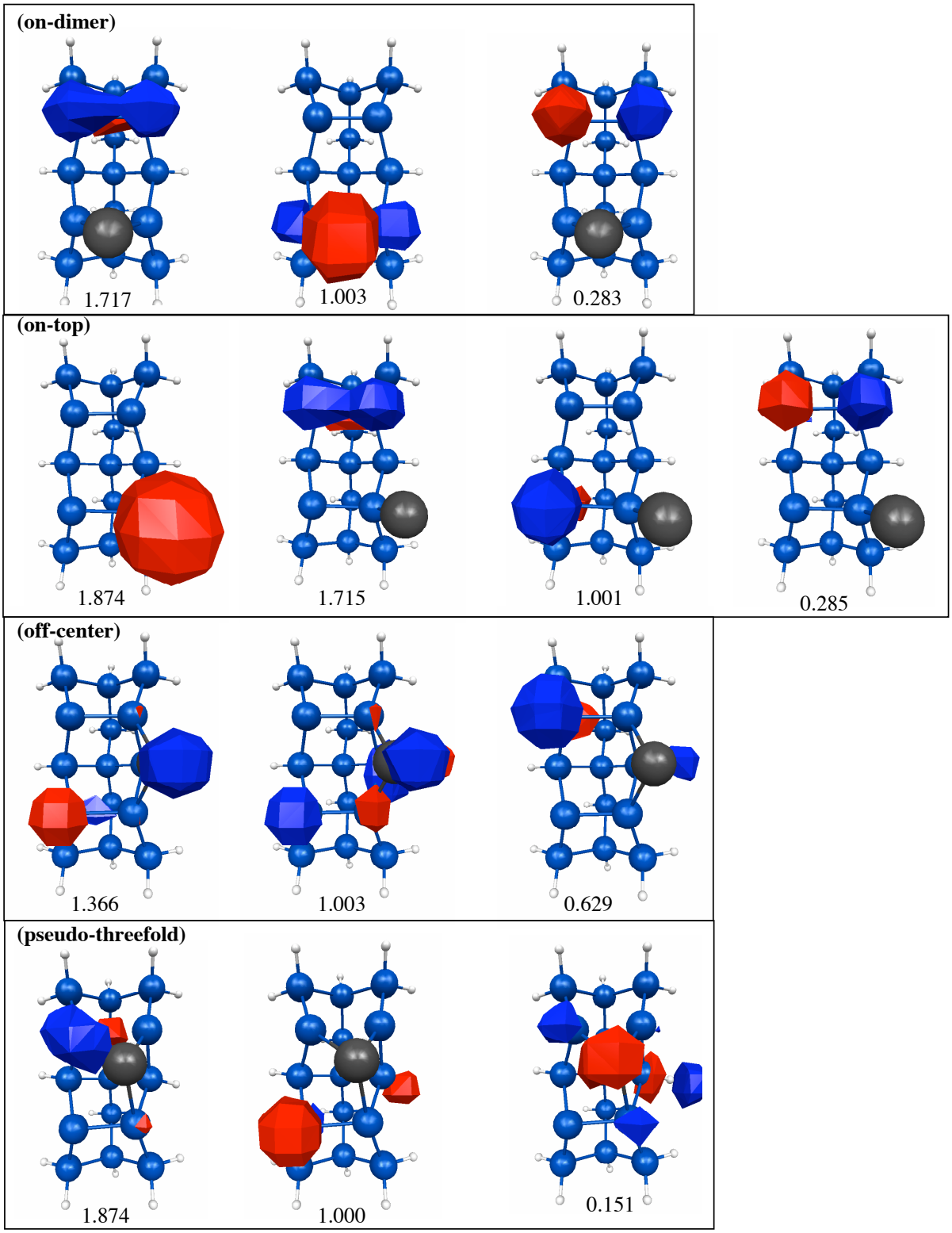


Figure 7. Multireference orbitals for doublet structures. NOON are given below each structure.

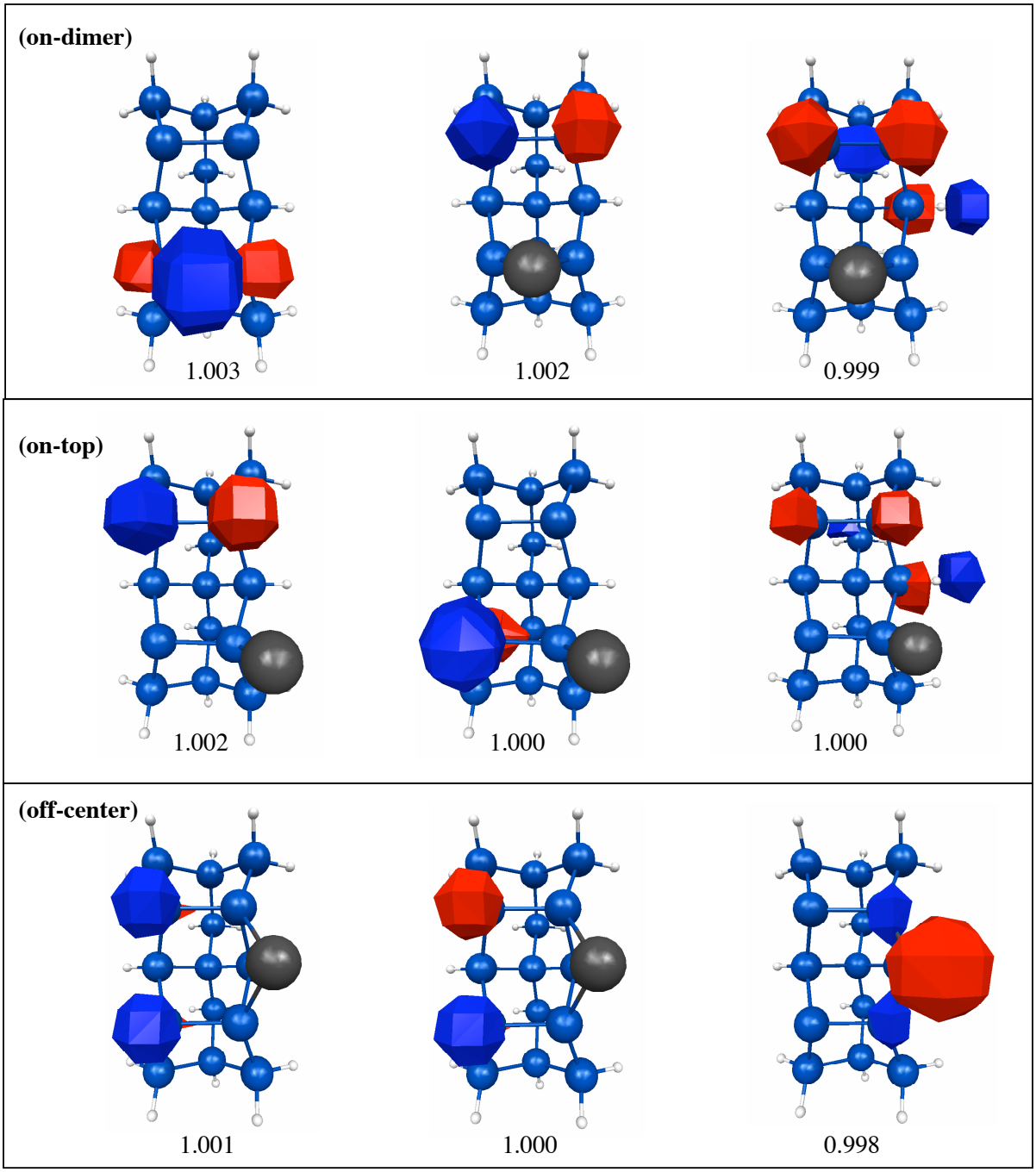


Figure 8. Singly occupied orbitals for quartet structures. NOON are given below each structure.

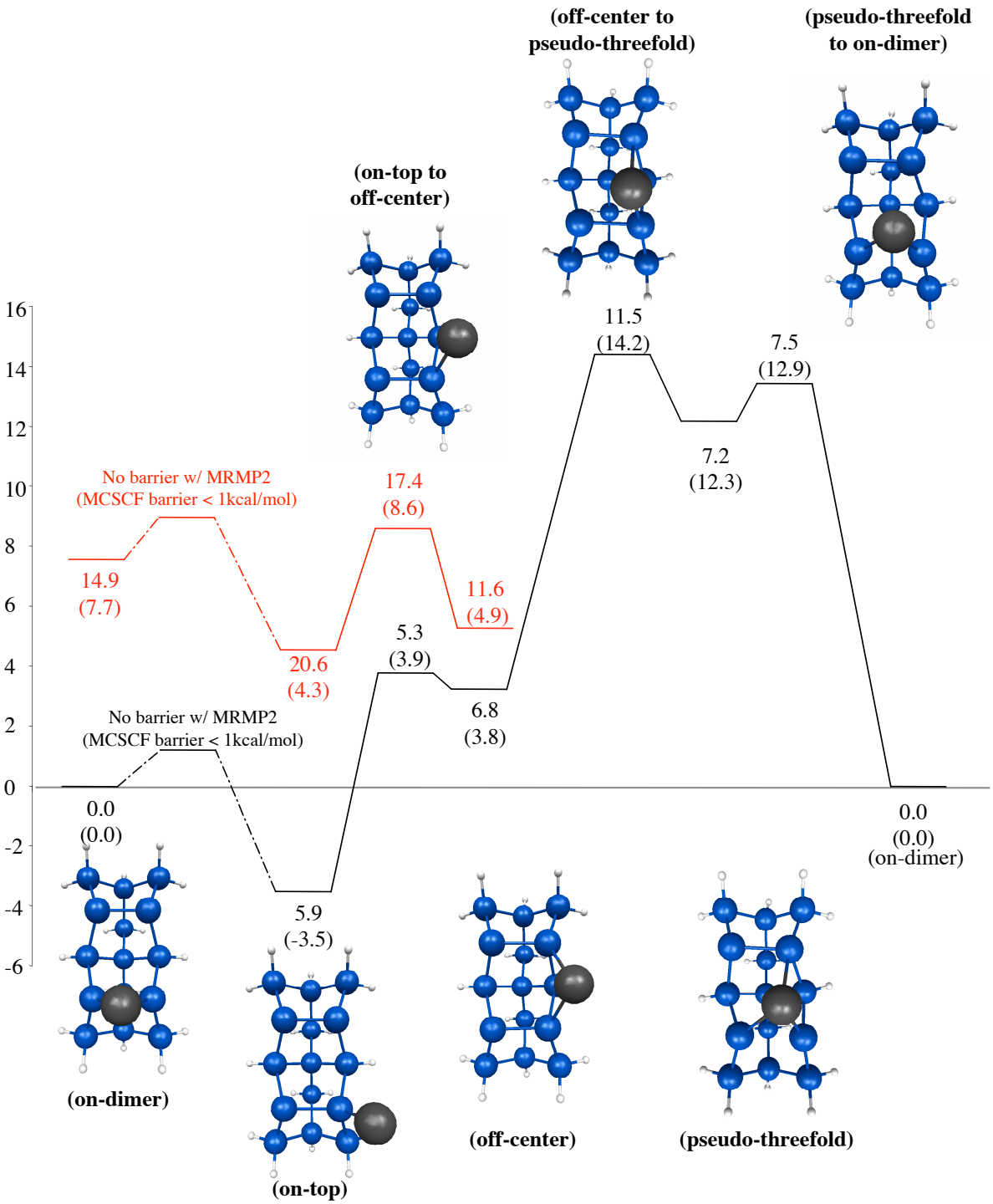


Figure 9. MRMP2//MCSCF diffusion of a single Al adatom. Doublet MCSCF minimum energy path in black, quartet MCSCF minimum energy pathway in red. MCSCF energies in parentheses. Energies in kcal/mol. Structures shown in figure are doublet minima and transition states.

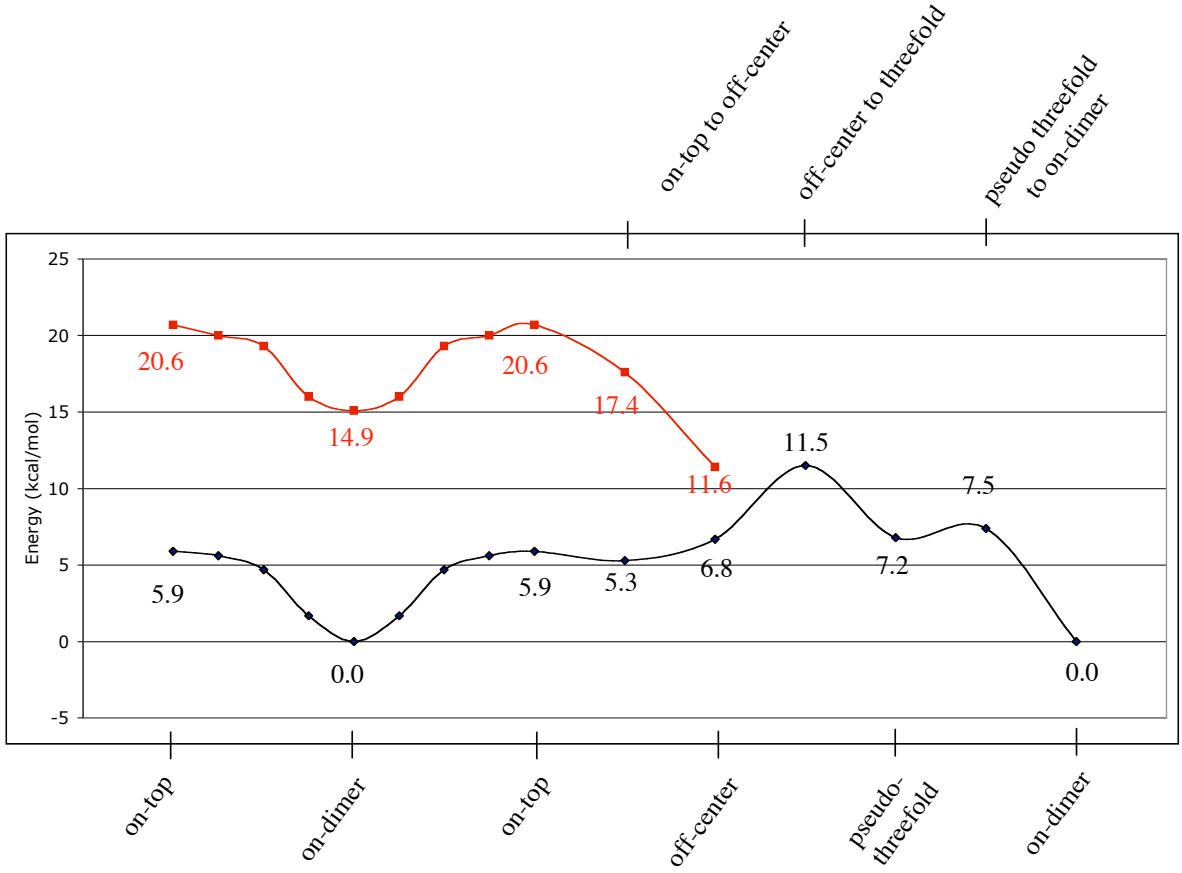


Figure 10. MRMP2//MCSCF diffusion of a single Al adatom. Approximate doublet MRMP2 minimum energy path in black, approximate quartet MRMP2 minimum energy path in red. Energies in kcal/mol.

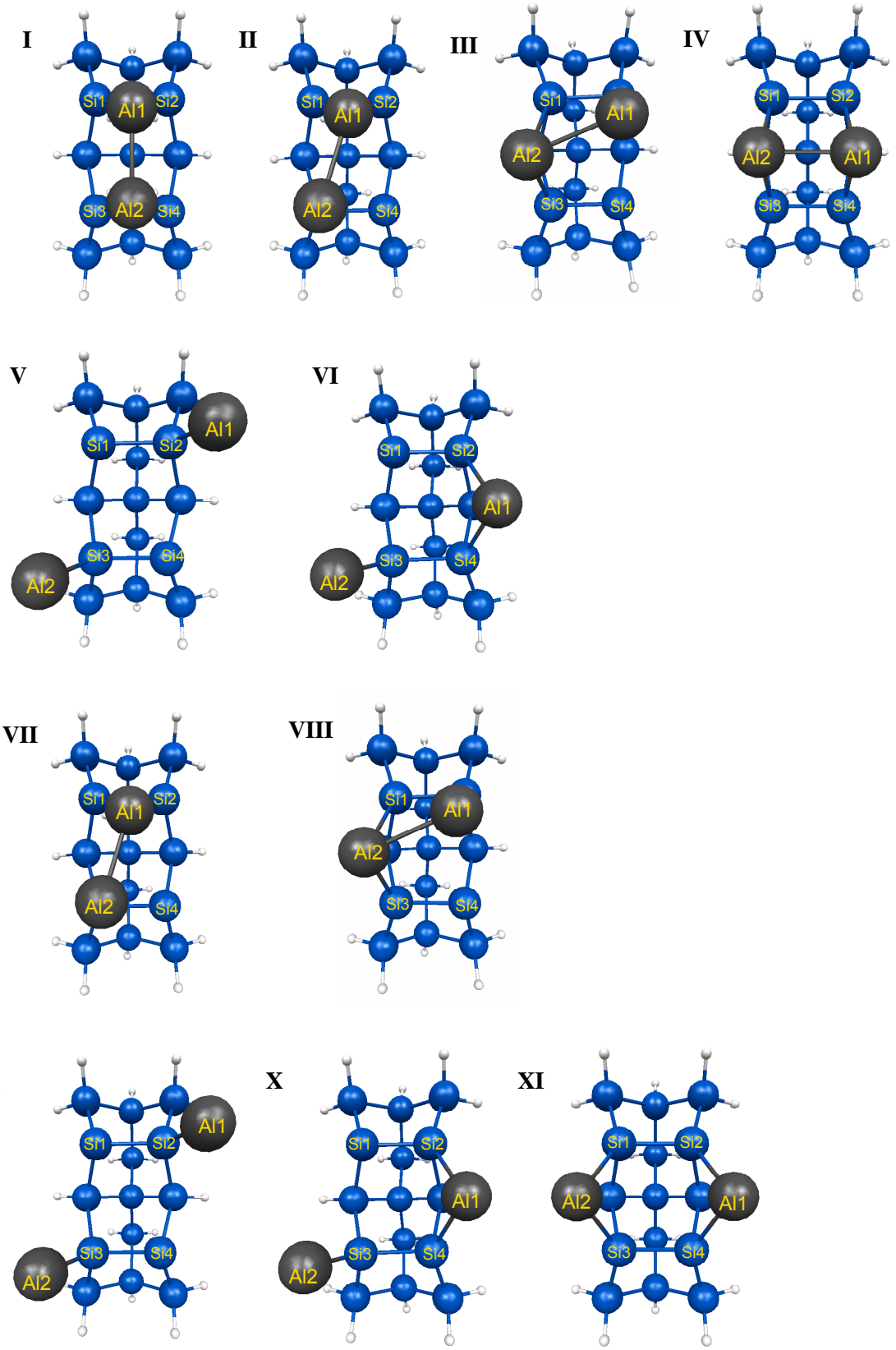


Figure 11. Al_2 minima. I-VI singlet minima, VII-XI Triplet minima.

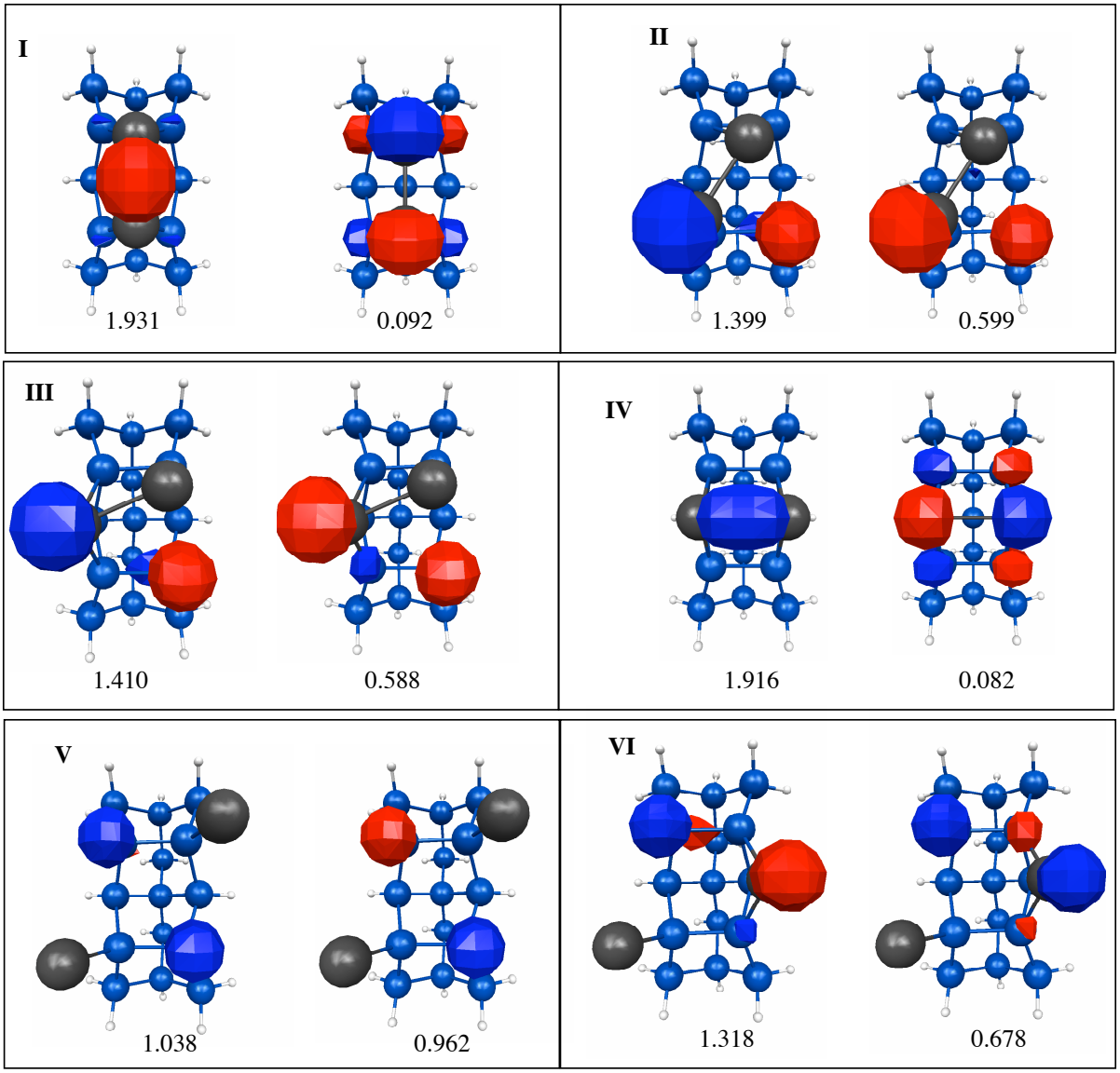


Figure 12. HOMO and LUMO orbitals for singlet structures. NOON given below each structure.

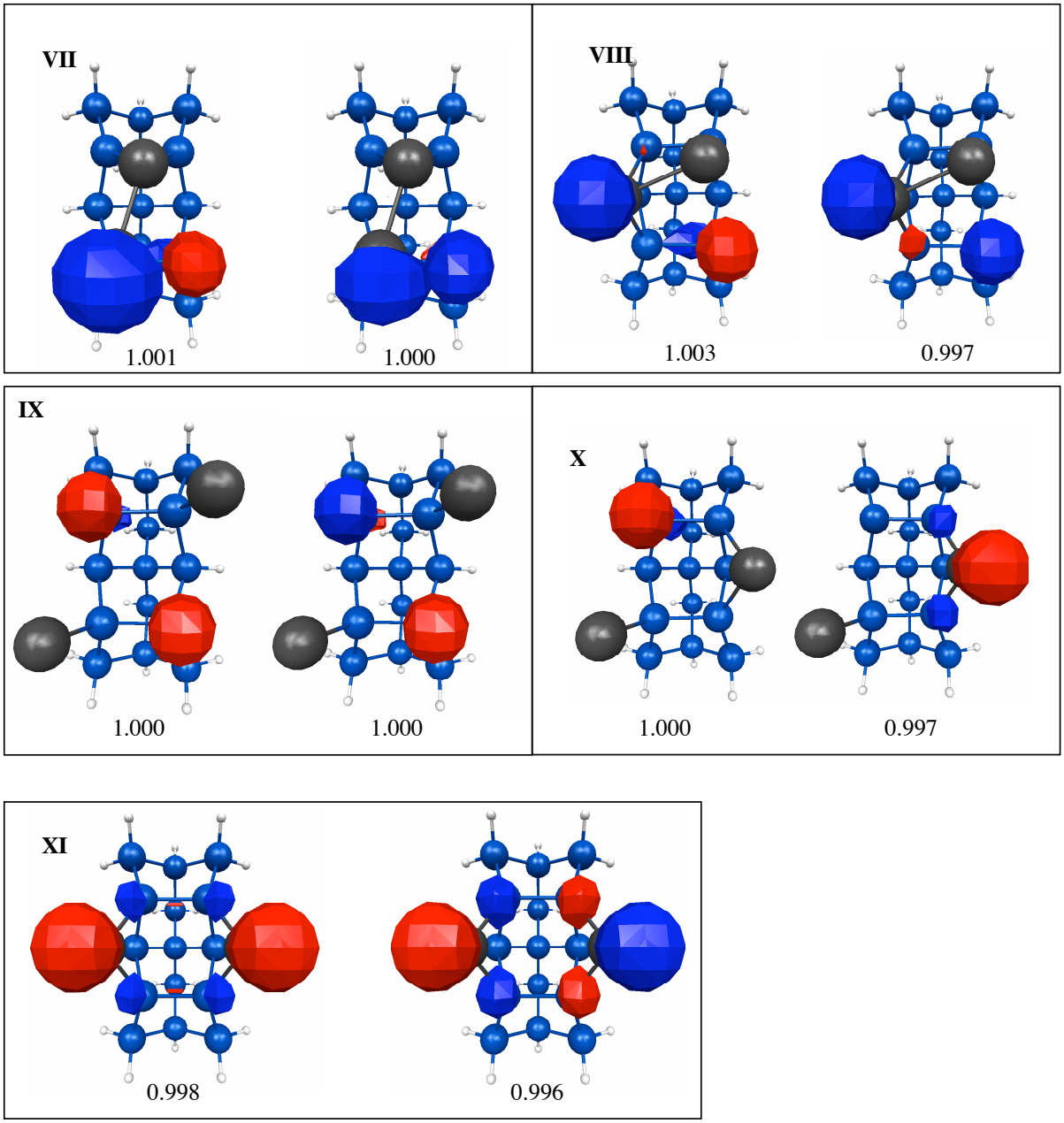


Figure 13. Singly occupied orbitals for triplet structures. NOON given below each structure.

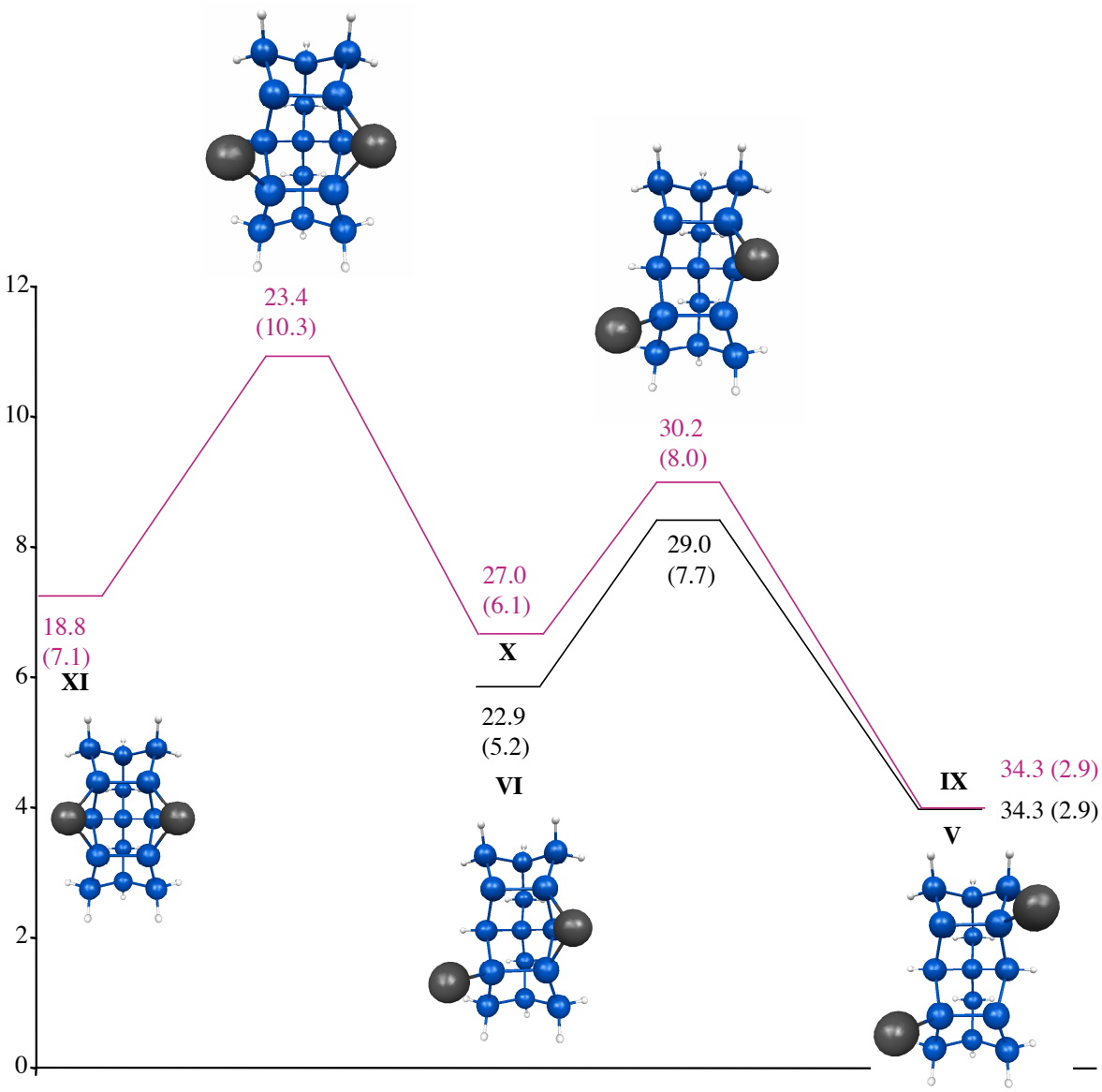


Figure 14. MRMP2//MCSCF diffusion of 2 separated Al adatoms. Singlet MCSCF minimum energy path in black, triplet MCSCF minimum energy path in red. MCSCF energies in parentheses. Energies in kcal/mol. The zero of energy corresponds to structure **IV** (Figure 10).

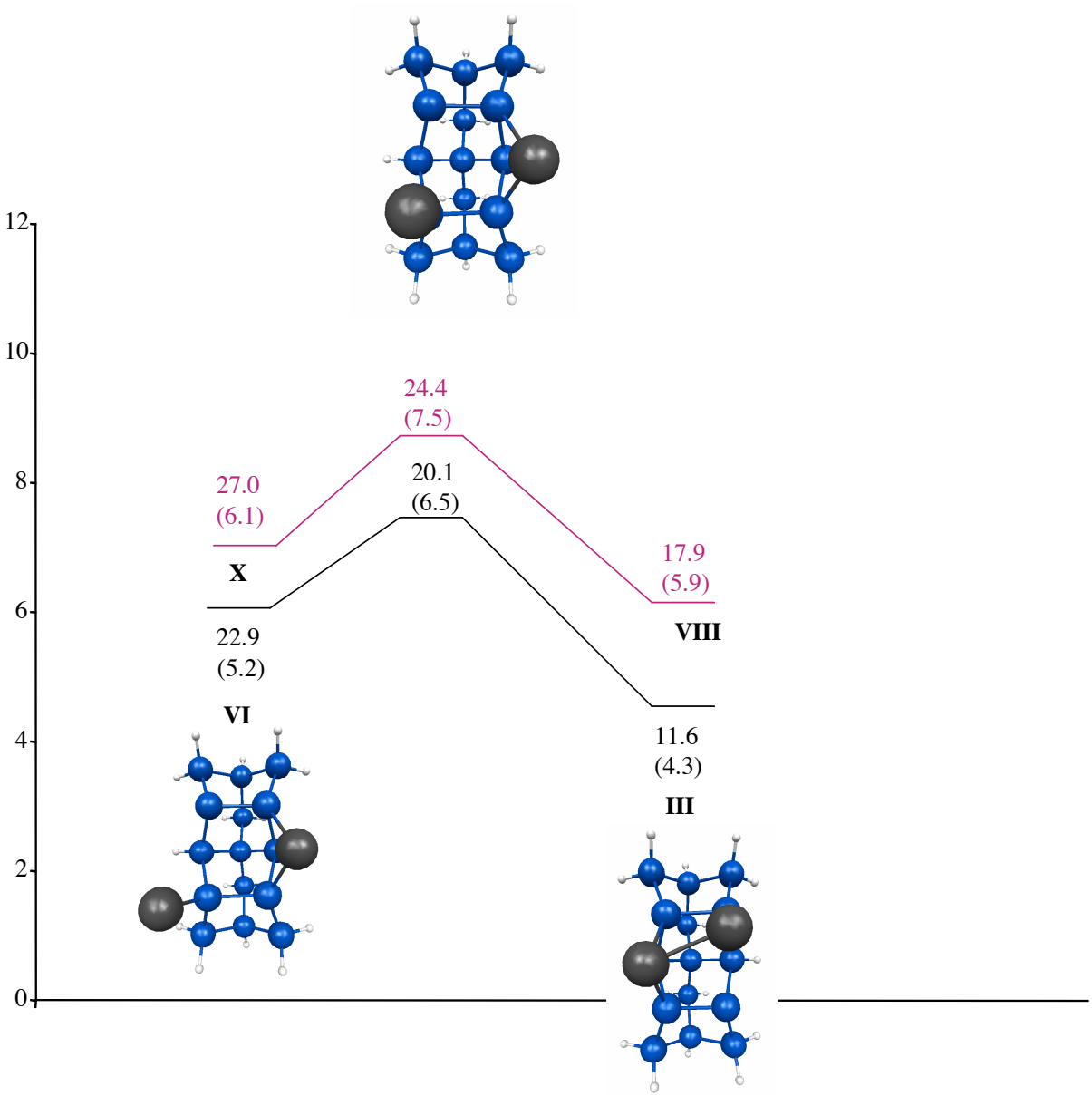


Figure 15. MRMP2//MCSCF formation of an Al-Al dimer. Singlet MCSCF minimum energy path in black, triplet MCSCF minimum energy path in red. MCSCF energies in parentheses. Energies in kcal/mol. The zero of energy corresponds to structure **IV** (Figure 10).

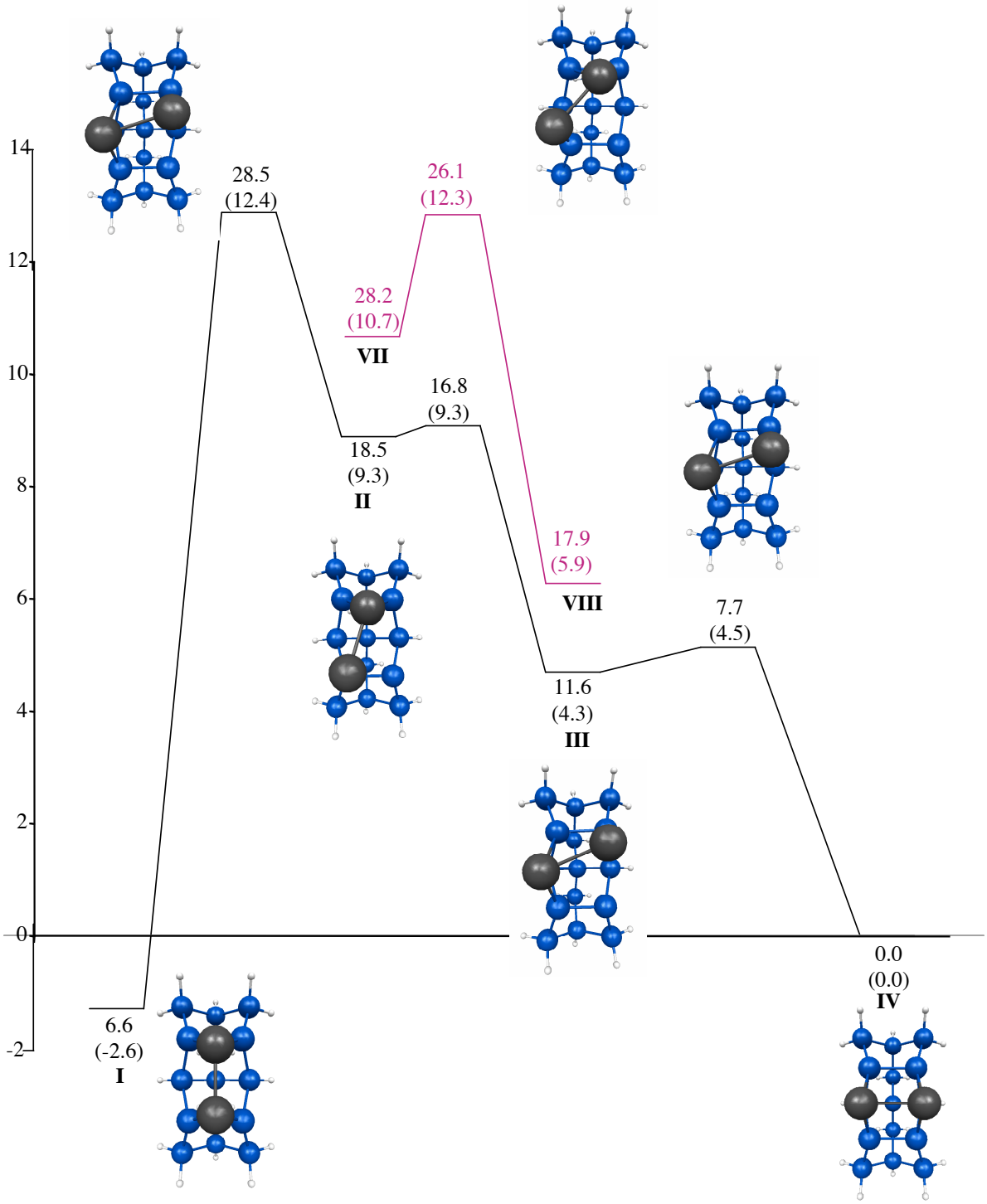


Figure 16. MRMP2//MCSCF rotation of an Al-Al dimer. Singlet MCSCF minimum energy path in black, triplet MCSCF minimum energy path in red. MCSCF energies in parentheses. Energies in kcal/mol.

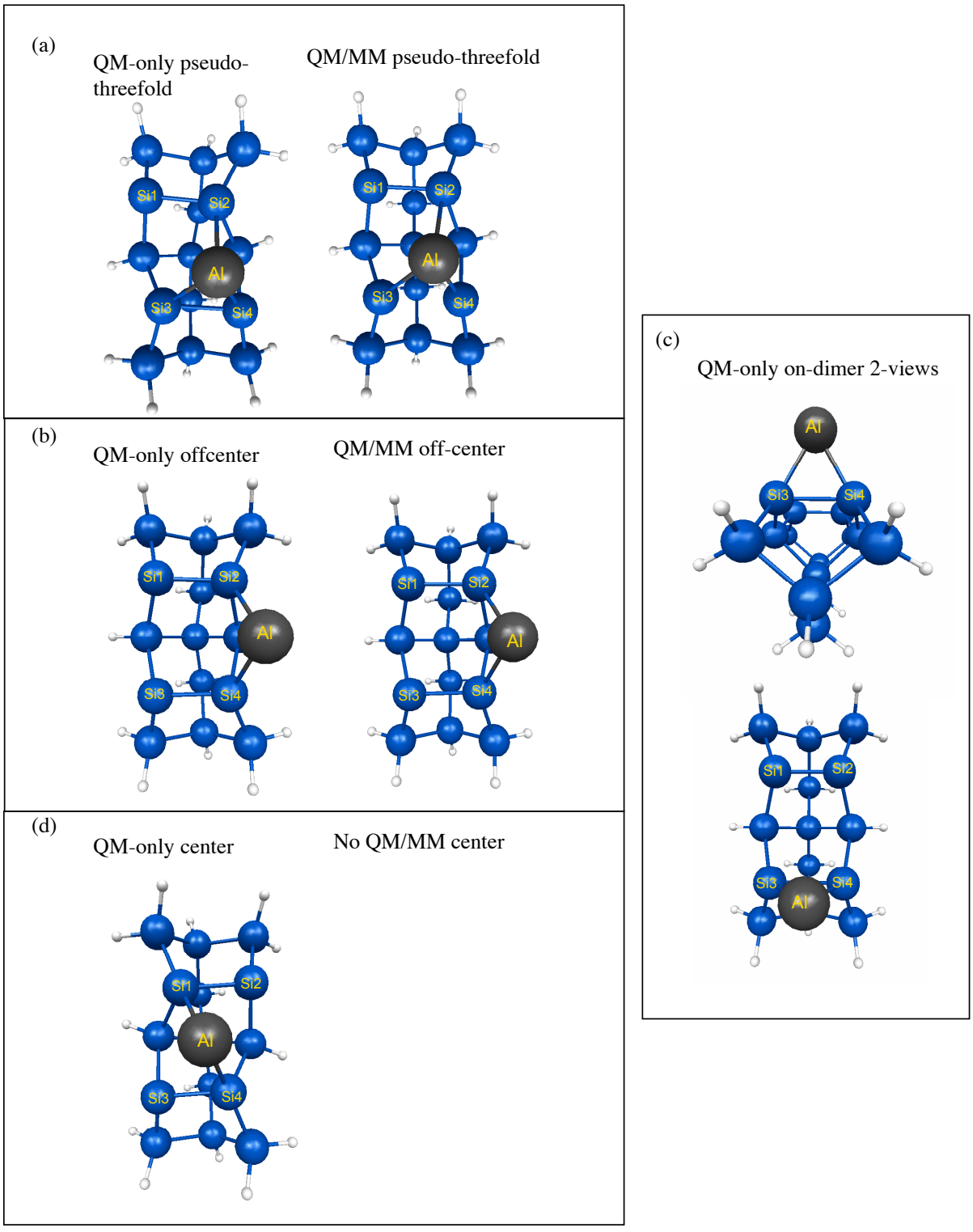
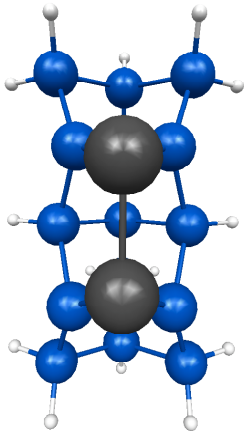


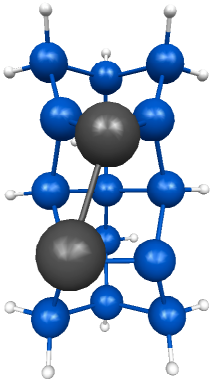
Figure 17. Comparison of distorted QM-only geometries with QM/MM geometries for singlet Al on Si(100).

(a)

QM-only between triplet

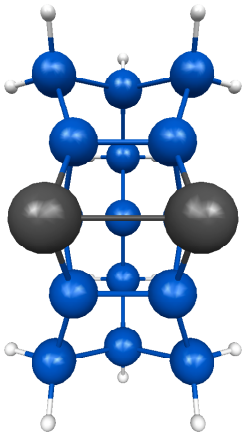


QM/MM between triplet



(b)

QM-only bridge triplet



QM/MM bridge triplet

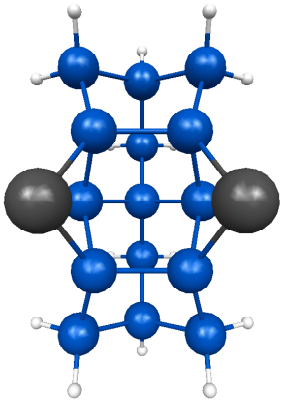


Figure 18. Comparison of distorted QM-only geometries with QM/MM geometries for triplet Al_2 on Si(100).

CHAPTER 4. COMPARISON OF NITROALDOL REACTION MECHANISMS USING ACURATE *AB INITIO* CALCULATIONS

Deborah Zorn, Victor S.-Y. Lin, Marek Pruski, and Mark Gordon

Abstract. In the nitroaldol reaction, condensation between a nitroalkane and an aldehyde yields a nitroalcohol that can undergo dehydration to yield a nitroalkene. Amine functionalized, MCM-41 Mesoporous Silica Nanosphere (MSN) materials, have been shown to selectively catalyze this reaction. Gas phase reaction paths for the several competing mechanisms for the nitroaldol reaction have been mapped out using second order perturbation theory (MP2). Improved relative energies were determined using singles and doubles coupled cluster theory with perturbative triples, CCSD(T). The mechanism in the absence of a catalyst was used to provide a baseline against which to assess the impact of the catalyst on both the mechanism and the related energetics. Catalyzed mechanisms can either pass through a nitroalcohol intermediate as in the uncatalyzed mechanism or an imine intermediate.

I. Introduction

The nitroaldol (or Henry) reaction (Scheme 1) is a base catalyzed reaction between a nitrostabilized carbanion and an aldehyde or ketone. The reaction product is a nitroalcohol, which can undergo elimination of water to give a nitroalkene product.[1,2] Mesoporous silica nanosphere (MSN) catalysts have been found to selectively catalyze the nitroaldol reaction.[3]

These MSN catalysts have been synthesized by co-condensation in order to immobilize multiple functional groups on the inside of the silica pores. A primary amine functionalized group catalyzes the nitroaldol reaction, and secondary groups control the selectivity. The secondary groups are called "gate keepers" because they prevent unwanted reactants from entering the catalyst pore by non-covalent (e.g., hydrophobic or hydrophilic) interactions.[3] In addition to their selectivity, advantages of these new MSN catalysts include their inert stationary phase, large surface area, and tunable pore size. A schematic of a multi-functionalized system is shown in Figure 1. In this example, the gatekeeper groups only allow reactant A to enter the functionalized pore, yielding product A selectively.

Demicheli et al.[4] proposed a mechanism, shown in Scheme 2, for the reaction of benzaldhyde with nitromethane in an amine functionalized MSN catalyst yielding nitrostyrene. The first step in this mechanism is the condensation of the supported amine with benzaldehyde, yielding a supported imine. The deprotonated nitromethane nitronate anion ($(NO_2CH_2)^-$) then adds to the carbon of the imine carbon-nitrogen double bond to give a beta-nitroamine. In the final step beta-scission gives nitrostyrene and regenerates the catalyst. The experimental evidence for this mechanism was derived from the FT-IR spectrum of the product, showing the formation of a C=N stretch, which disappeared upon further addition of nitromethane. This evidence cannot rule out the classical mechanism shown in Scheme 1, suggesting that further study of this system is necessary.

Computational chemistry can be particularly helpful in elucidating reaction mechanisms. In work reported by Lecea et al.[5] fourth order perturbation theory (MP4) calculations excluding triples, MP4SDQ,[6] were performed on five model nitroaldol reactions

to study the stereochemical control of the reactions. Lecea et al. only presented the barriers for the first step of the uncatalyzed nitroaldol reaction: addition of the $((NO_2CH_2)^-$ and an aldehyde. No barriers were given for subsequent steps such as the formation of the nitroalcohol and the dehydration reaction to give a nitroalkene.

The deprotonation of nitromethane by $(OH)^-\cdot nH_2O$ (n=0,2) clusters was studied by Beksic et al.[7] Hartree Fock and second order perturbation theory (MP2) calculations with the 6-31+G(d,p) basis set were performed to determine the geometries and energetics of the systems. The energy barrier for proton transfer from nitromethane to hydroxide with two waters was found to be only 4 kcal/mol above the reactant complex. The proton transfer reaction was found to be exothermic by 6.7 kcal/mol.

The catalyzed nitroaldol mechanism can proceed through an imine intermediate. Imine formation was studied with *ab inito* molecular orbtital calculations by Hall and Smith.[8] In that work the Gaussian-2(MP2,SVP) level of theory[9] was employed. The barrier for carbinolamine formation was predicted to be 112.3 kJ/mol (28.9 kcal/mol) in the gas phase. With the addition of two water molecules, the formation of the carbinolamine procedes via a zwitterionic intermediate, and the barrier is reduced to 14.7 kJ/mol (3.5 kcal/mol). Imine formation without water proceeds by a 4-center transition state, which is 231.4 kJ/mol (55.3 kcal/mol) higher in energy than the carbinolamine. Addition of one water lowers this barrier by 92.4 kJ/mol (22.1 kcal/mol). Addition of a second waters lowers the barrier by an additional 27.1 kJ/mol (6.5 kcal/mol) to 111.9 kJ/mol (26.7 kcal/mol). Aqueous free energies and acid-base equilibrium constants were also calculated.

The first step in understanding the mechanism for the amine functionalized MSN catalyzed reaction is to study the gas phase reactions. The present work will compare several possible mechanisms (see Schemes 4, 5 and 6) for the amine catalyzed nitroaldol reaction using accurate *ab initio* electronic structure calculations. These mechanisms will be discussed in detail in section III. Figure 2 summarizes most of the structures presented in Schemes 1, 3, 4, 5 and 6. The structure numbers (**1** through **8**) used throughout the text refer to the numbers given in Figure 2. The uncatalyzed mechanism will be used as a baseline for comparison.

II. Computational Methods

Structures were obtained by performing gas-phase MP2 calculations,[10,11] using the 6-31+G(d) basis set.[12-15] Hessians (second order derivatives of the energy) were used to characterize stationary points. Intrinsic reaction coordinate (IRC) calculations with the Gonzalez-Schlegel second order method[16,17] were used to connect transition states with reactants and products. The step size used for the IRC calculations was 0.1 $(amu)^{1/2}$ bohr.[18,19] At the final MP2/6-31+G(d) geometries, improved relative energies and barriers were determined using singles and doubles coupled cluster theory with perturbative triples (CCSD(T))[20,21], using the aug-cc-pVDZ[22] basis set. Partial charges on the optimized geometries were found using a Mulliken Population analysis.[23] Solvent effects were taken into account with the Polarizable Continuum Model (PCM) [24] using a solvent radius of 21.55

Å and a dielectric constant of 38.2 for nitromethane. PCM calculations were performed in two ways: in the first, PCM-MP2 single point energies were performed at the MP2 gas phase structures and in the second PCM-MP2 single point energies were performed at the optimized geometries from PCM-HF. These two methods are denoted MP2-PCM/6-31+G(d)//MP2/6-31+G(d) and PCM-[MP2/6-31+G(d)//HF/6-31+G(d)] respectively. The relative energies of all minima and transition include zero point energy (ZPE) corrections, calculated from the MP2 frequencies. All calculations were done with GAMESS[25,26] and all molecules were visualized with MacMolPlt[27].

III. Results and Discussion

The analysis of the nitroaldol reaction is presented in several sub-sections. In section III.1, the uncatalyzed reaction is investigated by studying the mechanism for addition of $(NO_2CH_2)^-$ to formaldehyde to yield a nitroalcohol **1** (Figure 2). A schematic of this mechanism is shown in Scheme 1. In section III.2, solvent effects are investigated in two ways: first by using the PCM continuum method and second by including an *ab initio* solvent molecule. In section III.3, the role of the amine catalyst is considered. The amine catalyst used by Huh et al. was an immobilized 3-[2-(aminoethylamino)ethylamino]propyl (AEP) group. Methylamine is used in the present study as a model amine catalyst. The non-bonded effects of the methylamine catalyst were investigated by studying a mechanism in which a methylamine molecule is present as an observer; that is, this additional methylamine does not form covalent bonds to the reactants. In Section III.4, three catalyzed mechanisms are investigated and compared; in these mechanisms covalent bonds are formed between the

amine catalyst and formaldehyde (see Scheme 5). Structures are labeled in Figure 2. In Section III.5 the effect of adding a second catalytic group to path D (Scheme 2) is considered. All mechanisms are compared to determine the most likely pathway(s) for the formation of the nitroalkene products in the gas phase.

III.1: Uncatalyzed mechanism

There are two steps in the uncatalyzed mechanism (whose minimum energy path is shown in Figure 3): first, $(NO_2CH_2)^-$ adds to formaldehyde to form 2-nitroethoxide; in the second step a proton transfers to the carbonyl oxygen to form the 2-nitroethanol anion. In the following discussion, relative energies are quoted at the highest level of theory used: CCSD(T)/aug-cc-pVDZ//MP2/6-31+G(d). In Figure 3, structure **I**, the reactants formaldehyde and $(NO_2CH_2)^-$ form a Van der Waals complex with a C-C bond length of 3.166 Å. In the first transition state (Figure 3, Structure **TS1)** $(NO_2CH_2)^-$ attacks the carbonyl carbon forming a C-C bond giving 2-nitroethoxide (Figure 3, Structure **II**). **TS1** has a stretched C-C distance of 2.291 Å and is 3.2 kcal/mol above complex **I**. Structure **II** is 1.4 kcal/mol lower in energy than the starting complex, **I** and has a C-C bond length of 1.600 Å and an O-C-C angle of 109.3 degrees.

Formation of 2-nitroethanol anion can proceed through a 4-center transition state (Figure 3, Structure **TS2**) in which a proton transfers from the carbon that is bonded to the NO_2 group to the carbonyl oxygen. **TS2** is 23.3 kcal/mol above the starting complex. This large barrier is due to the strain in the 4-center transition state. In **TS2** The O-C-C angle is only 97.7°. The geometry of **5** is shown in Figure 3 structure **III**, which is 13.0 kcal/mol lower in energy than the starting complex and is a pseudo-cyclic compound with a hydrogen

bond between the hydroxy group and an oxygen from nitromethane. The O-C-C bond angle has now opened to 113.8°.

MP2 Mulliken charges on the starting complex suggest that the nitromethane carbon has a charge of –0.53, and the nitro group caries a net charge of –0.78. The formaldehyde carbonyl carbon carries a small net negative charge of –0.04 and the carbonyl oxygen has a charge of –0.40. The charges do not change significantly from **I** to **II**. In **II**, the carbonyl oxygen has a net charge of –0.40 and the net negative charge on the nitro group is reduced slightly to –0.75. The Mulliken charges on **III** are significantly different from those in **I** or **II**. In **III** the nitromethane carbon has a charge of only –0.27, the carbonyl carbon has a charge of –0.22, the carbonyl oxygen has a charge of –0.74, and the nitro group has a net charge of –0.82. These values suggest significant charge delocalization in **III**.

The MP2/6-31+G(d) barrier at **TS1** reproduces the full core MP2 barrier determined by Lecea et al.[5] However, the MP3 and MP4SDQ barriers quoted by Lecea et al. are approximately twice as high as the CCSD(T)/aug-cc-pVDZ//MP2/6-31+G(d) values, indicating the unreliability of the MP3 and MP4SDQ energies. Lecea et al. did not report barriers for **TS2**. All MP2 relative energies and barrier heights in Figure 3 compare very well with those from CCSD(T)//MP2 single point energy calculations.

III.2: Solvent effects

Solvent effects for the uncatalyzed mechanism were taken into account using the PCM continuum approach[24] for the nitromethane solvent, as illustrated in Figure 4. The dielectric constant chosen for the nitromethane solvent is the default GAMESS value of

38.2[25,26]. Figure 4 shows a comparison of the MP2-PCM//MP2 and MP2-PCM//HF-PCM minimum energy paths. Both solvent approaches produce an increase in the **TS1** and **TS2** barrier heights relative to the gas phase MP2 barriers. The **TS1** barrier is increased over the MP2 gas phase barrier by 8.6 and 14.4 kcal/mol with MP2-PCM//MP2 and MP2-PCM//HF-PCM, respectively. The **TS2** barrier is raised by at least 20 kcal/mol with both PCM methods. PCM also slightly increases the energy of structure **II** relative to the starting structure as compared to the gas phase. When the PCM solvent is present, the Mulliken charge distribution is less delocalized than it is in the gas phase species. For example, PCM increases the negative charge on the carbonyl oxygen in all species and decreases the amount of negative charge on the carbonyl carbon.

Figure 5 demonstrates the effect of the addition of one *ab initio* solvent molecule to the uncatalyzed mechanism. All resulting minima and transition states are shown in Figure 5a. The relative energies are presented in Figure 5b. A schematic of this mechanism is shown in Scheme 3. The first step of this mechanism is the addition of $(NO_2CH_2)^-$ to formaldehyde to form 2-nitroethoxide. Structure **5** is formed in two steps: First, a proton transfers from nitromethane to the carbonyl oxygen to form **4** and $(NO_2CH_2)^-$; then a proton transfers from the carbon bonded to the nitro group of **1** back to $(NO_2CH_2)^-$ to form nitromethane and **5**. **5** can now eliminate water to form the nitroethene products. The barrier height (2.9 kcal/mol) for addition of $(NO_2CH_2)^-$ to formaldehyde (Figure 5a, Structure **TS3**) is virtually unchanged from the gas phase value. Addition of an *ab initio* nitromethane solvent molecule can decrease the barrier for formation of 2-nitroethanol (cf., **TS2** in Figure 3) by using a 2-step mechanism (Figure 5b). In the first step (Figure 5a, Structure **TS4**) a proton transfers from

nitromethane to the carbonyl oxygen of 2-nitroethoxide in structure **V** forming 2-nitroethanol and $(NO_2CH_2)^-$ (Figure 5a, Structure **VI-a**). **TS4** is 6.8 kcal/mol higher in energy than **V**. In the second step of the nitroalcohol formation a proton on the carbon bonded to the nitro group of **1** needs to transfer to $(NO_2CH_2)^-$ to form 2-nitroethanol anion and nitromethane, but first isomer **VI-a** must convert to isomer **VI-b** (see Figure 5a). The barrier for this step (Figure 5b, $\widetilde{\mathbf{TS}}$ **5**) is estimated to be less than 1.3 kcal/mol higher in energy than **VI-a** by a series of constrained optimizations along a linear least motion (LLM) path. The barrier for proton transfer to form 2-nitroethanol anion (Figure 5a, Structure **TS6**) is 9.3 kcal/mol higher in energy than **VI-b**, but still 6 kcal/mol below the starting reactants. The complex of **5** with nitromethane (Figure 5a, Structure **VII**) is 14.0 kcal/mol lower in energy than the starting complex (see Figure 5b). The net energy requirement for formation of **5** in this mechanism is 2.9 kcal/mol, which is significantly lower than the net energy requirement in the uncatalyzed mechanism (23.6 kcal/mol).

A transition state was not found for elimination of water from **VII** by proton transfer from a nitromethane solvent molecule. A series of constrained optimizations along a LLM path found an approximate upper bound of this barrier to be 37.8 kcal/mol higher in energy than the starting complex. The elimination product (Figure 2, Structure **2** is 4.5 kcal/mol lower in energy than the starting complex and its geometry is shown in (Figure 5a, Structure **VIII**).

III.3: Amine "Assisted" Mechanism

In the amine "assisted" mechanism (Figure 6), the nonbonded (environmental) affects of the methylamine catalyst were investigated. A schematic of this mechanism is shown in Scheme 4. In the first step $(NO_2CH_2)^-$ attacks the carbonyl carbon forming a C-C bond, producing **4**. A proton then transfers to the carbonyl oxygen to form **5**.

The minima and transition states for this mechanism are shown in Figure 6a, and the corresponding minimum energy path (MEP) is depicted in Figure 6b. **I-A** is a complex between formaldehyde, methylamine and $(NO_2CH_2)^-$. In **I-A** there is no hydrogen bond between the complex of formaldehyde and $(NO_2CH_2)^-$. In **TS1-A**, $(NO_2CH_2)^-$ attacks the carbonyl carbon forming a C-C bond, leading to 2-nitroethoxide (Figure 6a, Structure **II-A-a**). **TS1-A** is 3.1 kcal/mol higher in energy than the reactant complex (**I-A**), as shown in Figure 6b. **TS1-A** has a structure that is almost identical to the structure of **TS1** (see Figure 3) and their barrier heights are almost identical as well. In **TS1-A** there are no hydrogen bonds between CH_3NH_2 and the reacting molecules.

The next step in the mechanism is conversion from **II-A-a** to **II-A-b** (Figure 6b, $\widetilde{\text{TS}}$ **2-A**). In both **II-A-a** and structure **II-A-b,** there is a hydrogen bond between CH_3NH_2, and the carbonyl carbon, with H-bond lengths of 1.859 Å and 1.845 Å, respectively. The barrier for conversion from **II-A-a** to **II-A-b** (approximated by a series of optimizations along the LLM path.) is less than 1 kcal/mol above **II-A-a**.

In the final step of this mechanism, Structure **III-A** (Figure 6a) is formed. The transition state structure connecting **II-A-b** and **III-A** (Figure 6a, Structure **TS3-A**) is a 4-center transition state in which a proton transfers to the carbonyl oxygen. **TS3-A**, is 22.3 kcal/mol higher in energy than the reactant complex. **TS3-A** has nearly the same geometry

as the transition state in the uncatalyzed mechanism, (Figure 3, Structure **TS2**) and its barrier is only 1 kcal/mol lower in energy. The presence of CH_3NH_2 in a hydrogen-bonded arrangement causes no significant change in the bond lengths of **TS2** or **III** (figure 3) with the addition of CH_3NH_2. The hydrogen bond length involving CH_3NH_2 in **III-A** is 2.154 Å. To summarize, the presence of CH_3NH_2 does not significantly affect the barrier heights of the two transition states leading to the formation of 2-nitroethanol. Interaction with an *ab intio* solvent molecule as discussed in section III.2 is a much more effective way to lower the barrier height of the second transition state (Figure 3, Structure **TS2**).

III.4: Catalysis Mechanisms

The mechanisms in which a covalent bond is formed between the catalyst and the reactants will now be discussed. A schematic of three possible catalyzed mechanisms is shown in Scheme 5. The first step in all catalyzed mechanisms is addition of methylamine to formaldehyde to form **3** (see Scheme 5a). Three possible pathways, arbitrarily labeled B, C, and D, are shown in Scheme 5b.

In path B, **3** undergoes a S_N2 reaction with $(NO_2CH_2)^-$ to form **4** and CH_3NH_2. A proton then transfers to the carbonyl oxygen to form **5**. Once a proton is added to the system, **5** can undergo elimination to form the nitroethene product, **2**. In Figure 3 path C, **3** (Figure 2) undergoes a S_N2 reaction with $(NO_2CH_2)^-$ to form **6**. In this path the leaving group is an hydroxide ion, rather than $(CH_3NH)^-$. A proton then transfers from the amine nitrogen of **6** to the hydroxyide ion to from **7** and water. With the addition of a proton, the catalyst is regenerated and the nitroethene product, **2**, is formed. In Scheme 5 path D, water is eliminated from **3** by proton transfer from the amine nitrogen to the alcohol oxygen forming

8 and water. $(NO_2CH_2)^-$ can then add to the carbon of the imine double bond to form **7**. Once a proton is added to the system, the catalyst is regenerated and the nitroethene product, **2**, is formed.

Formation of 3. The first step in the three catalyzed mechanisms is the formation of **3** (Scheme 5a). The minima and transition states for this step are shown in Figure 7a and the corresponding MEP is shown in Figure 7b. In the first step, CH_3NH_2 adds to formaldehyde in **I-A** forming **B-I (Figure 7b, B-** $\widetilde{\textbf{TS}}$ **1**). A series of constrained optimizations along a linear least motion path was used to estimate the barrier height for this step. The upper bound for this barrier was found to be less than 1.3 kcal/mol higher in energy **I-A**. In the next step (Figure 7a, Structure **B-TS2**) a proton transfers from the amine nitrogen to $(NO_2CH_2)^-$ to form isomer **B-II-a**. **B-TS2** is 5.6 kcal/mol above **B-I,** and **B-II-a** is 5.5 kcal/mol higher in energy than the reactants. After inversion of the amine nitrogen in isomer **B-II-a** to form isomer **B-II-b** (Figure 7b, **B-** $\widetilde{\textbf{TS}}$ **3**) a proton can now transfer to the carbonyl carbon (Figure 7a, Structure **B-TS4**) to form **B-III-a** (Figure 7a. The barrier for inversion of the amine and rotation of the hydroxy group was estimated by a series of constrained optimizations on a linear least motion path. The transition state structure for the proton transfer is **B-TS4**. **B-TS4** was found to be –0.8 kcal/mol lower in energy than **B-II-b** after ZPE was included.

Catalyzed Mechanism - Path B. The minimum energy path for path B (see Scheme 5b and associated discussion) is shown in Figure 8. The first step in this mechanism is the conversion of **B-III-a** to **B-III-b**. This step is simply a rotation of methylaminomethanol with respect to $(NO_2CH_2)^-$. The barrier was approximated by a series of constrained optimizations along a LLM path. The upper bound for this step is 9.7 kcal/mol above **I-A**

(Figure 8, **$\tilde{\text{TS}}$ 5**). The second step in this catalyzed mechanism path is attack of the carbonyl carbon in structure **B-III-b** by $(NO_2CH_2)^-$ to form the nitroalkoxide and regenerate the catalyst (Figure 8a, Structure **B-IV)**. The transition state for this step is **B-TS6**, which is 51.5 kcal/mol higher in energy than **I-A**. In this step $(NO_2CH_2)^-$ first attacks the carbonyl carbon eliminating $(CH_3NH)^-$, along with a simultaneous proton transfer from the alcohol to the amine nitrogen. **B-V** is 14.0 kcal/mol below the reactants.

In order to form **B-V** from **B-IV**, a proton must transfer from the carbon bonded to the NO_2 group to the carbonyl oxygen. During the investigation of the uncatalyzed mechanism, the MP2/6-31+G(d) barrier for this step with no methylamine catalyst group (Figure 3, Structure **TS2**) was found to be 23.6 kcal/mol higher in energy than the reactant complex (Figure 3, structure **I**). When the methylamine catalyst forms covalent bonds with the reactants the barrier for proton transfer drops to only 11.3 kcal/mol (Figure 8a, Structure **B-TS7**) above the reactant complex. **B-TS7** is 10.3 kcal/mol lower in energy than **TS3-A**. This is because adding a methylamine transforms the 4-center transition state (Figure 6a, Structure **TS3-A**), to a 6-center transition state (Figure 8a, Structure **B-TS7)**. In the four-center transition state the proton transfers directly from the carbon bonded to the nitro group to the carbonyl oxygen. In the 6-center transition state (**B-TS7**), a proton transfers from CH_3NH_2 to the carbonyl alcohol and then in the same step, a proton transfers from the carbon bonded to NO_2 back to CH_3NH. Such mechanism modifications are well known, especially when water molecules are present. This was especially true for the synthesis of three and four membered cyclosiloxanes,[28] where the potential energy barriers are reduced nearly to

zero in the presence of a water molecule. The O-C-C bond angle in **B-TS7** is 108.4˚, which is much less strained than in **TS2** (Figure 3).

After a proton has been added to the system **B-V** becomes **IX** (Figure 9). **IX** can undergo dehydration as shown in Figure 9 to give the final nitroalkene product (Figure 2, Structure **2**). This step occurs by a concerted reaction in which a protonated amine donates a proton to the hydroxy group in **IX** (Figure 9). The transition state for elimination of water (Figure 9, Structure **TS7**) is 20.6 kcal/mol higher in energy than the reactant complex of $CH_2O + CH_3NO_2 + CH_3NH_2$. The product (Figure 9, Structure **X**) is -0.9 kcal/mol lower in energy than the reactant complex. In **TS7**, the distance between the alcohol carbon and the alcohol oxygen elongates to 1.864 Å as a proton transfers to the carbonyl oxygen. The distance between the proton transferring from $CH_3NH_3^+$ and the alcohol oxygen is 1.252 Å.

Catalyzed Mechanism - Path C. The second option for the catalyzed mechanism is depicted in Scheme 5b path C. The minima and transition states in path C are shown in Figure 10a and the corresponding MEP is shown in Figure 10b. The starting complex for path B is structure **C-I** (Figure 10a), which is a complex of methylaminemethanol with $(NO_2CH_2)^-$. The transition state for the first step along this path is structure **C-TS1** (Figure 10a). **C-TS1** is 37.8 kcal/mol higher in energy than **I-A** (Figure 10b). **C-II** is 0.9 kcal/mol lower in energy than the reactant complex, **I-A**. **C-TS1** is a pentacoordinated transition state, in which $(NO_2CH_2)^-$ adds as the nucleophile and HO^- is the leaving group, in a S_N2-like process. In Structure **C-II**, the amine hydrogen is hydrogen bonded to the OH oxygen.

A proton can now transfer from the amine nitrogen of **C-II** OH^- to form **C-III-a** (Figure 10a), in which water is H-bonded to the amine N. The transition state for this step is structure **C-TS2** (Figure 10a). Before ZPE is accounted for, this transition state is 1.7

kcal/mol higher in energy than **C-II**, however with ZPE included, the energy of C-TS2 is 1.4 kcal/mol *lower* in energy than **C-II**, indicating that **C-TS2** is not a true TS on the PES. **C-III-a** is 2.1 kcal/mol lower in energy than **C-I**. The barrier for the conversion of **C-III-a** to **C-III-b** (~2.6 kcal/mol for rotation of the water molecule) was approximated by a series of constrained optimizations along a LLM path.

From **C-III-b,** a proton on the carbon bonded to the nitro group must transfer to the amine nitrogen. This is a two-step process (Figure 10a **C-TS4** and **C-TS5**) in which a proton is first transferred from the water molecule to the amine nitrogen and then a second proton is transferred from the carbon bonded to the nitro group back to the hydroxide ion. Including ZPE, the barriers for both **C-TS4** and **C-TS5** are lower in energy than **C-III-a**. At this point, structure **C-IV** is 23.5 kcal/mol lower in energy than the reactants. In the final step of path C (Figure 11) a proton is added to the system and nitromethane is regenerated forming **2** (Figure 11, **C-VII**). The transition state for this step is C**-TS5** has a barrier 11.7 kcal/mol above the net-neutral complex of the reactants.

Catalyzed Mechanism - Path D. The third pathway for the catalyzed mechanism is shown in Scheme 5b, path D. The reactants and transition states along pathway D are shown in Figure 12a, and the corresponding MEP is shown in Figure 12b. In the first step of this mechanism (Figure 12a, Structure **D-TS1**), a proton transfers from the amine nitrogen of **D-I** to the carbon to form nitromethane **D-II-a** (Figure 12a,b). **D-TS1** is 10.9 kcal/mol higher in energy than **D-I**. **D-II-a** must now convert to **D-II-b**. The barrier for this step (Figure 12b, **D-$\widetilde{\text{TS}}$2**) was approximated by a series of constrained optimizations along a LLM path to be less than 2.0 kcal/mol.

In the next step (Figure 12a, Structure **D-TS3**) a proton transfers from nitromethane to the alcohol oxygen and a hydroxy group is eliminated forming **D-III**. **D-TS3** is higher in energy than **D-II-b** before zero point energy (ZPE) is accounted for. When the ZPE is included, the energy of the transition state is 0.4 kcal/mol lower in energy than **D-II-b**. Water is formed when a proton is transferred from nitromethane to the hydroxy group (Figure 12a, Structure **D-TS4**). **D-TS4** is higher in energy than **D-III** before the ZPE is included, but after the ZPE is added in, the transition state is 1.3 kcal/mol below **D-III**.

Now that water has been eliminated, nitromethane anion can add to the carbon of the C-N double bond to form a C-C single bond. The transition state for this step is **D-$\tilde{\text{TS}}$5** (Figure 12b). The barrier for this step is simply the barrier for moving the water molecule out of the way, making the imine available to attack by $(NO_2CH_2)^-$. This barrier is approximated by a series of constrained optimizations on a LLM path to be less than 2 kcal/mol. A proton can now transfer to form the β-nitroamine in the same way as in mechanism C (See **C-TS4** and **C-TS5** in Figure 10a).

In summary, only path D has a net energy requirement (7 kcal/mol) less than that of the uncatalyzed mechanism. Paths B and C can be eliminated because of their high barriers compared to those in the amine "assisted" mechanism and path D of the catalyzed mechanism. The S_N2 reaction in these paths would be especially difficult if a more complex and possibly sterically hindered aldehyde was used. Catalyzed Mechanism D has the lowest barriers of any mechanism that was investigated here.

III.5: Multiple amine molecules

In the mechanism proposed by Demicheli et al. (Scheme 2), the intermediates formed are the same as in Scheme 5, path D, except that multiple amine catalyst groups are proposed to be involved. This could be an important affect. To investigate this, a second methylamine molecule was added to the reaction of formaldehyde with methylamine. A schematic of this mechanism is shown in Scheme 6. The first step is addition of methylamine to formaldehyde to eliminate water and form **8**. $(NO_2CH_2)^-$ then adds to the carbon of the imine C-N double bond to form **7**. In the final step acid is added and the catalyst is regenerated forming the product, **2**.

The affect of adding a second methylamine molecule will be investigated in two steps: first, the energy requirement for imine formation will be presented; then, the net energy requirement for addition of $(NO_2CH_2)^-$ to the imine and the regeneration of the methylamine catalyst will be explored.

Imine Formation. The structures of the minima and transition states for imine formation are shown in Figure 13a, and the corresponding MEP is shown in Figure 13b. The starting complex in this mechanism is formaldehyde plus two methylamine molecules (Figure 13a, Structure **E-I**). The first step in this mechanism is the addition of nitromethane to formaldehyde (Figure 13a, Structure **E-TS1**). In this step, the amine attacks the carbonyl carbon, and at the same time a proton transfers to the to the second amine. Then, a proton transfers from the second amine to the carbonyl oxygen forming (**E-II-a**). **E-TS1** has a barrier of 13.9 kcal/mol (Figure 13a). In **E-II-a**, there is a hydrogen bond between the alcohol and the methylamine.

Before water can be eliminated, the amine nitrogen in **E-II-a** (Figure 13a) must undergo inversion to make the proton available to CH_3NH_2. This also breaks the O-H…N hydrogen bond. The barrier for conversion from **E-II-a** to **E-II-b** (Figure 13, **E-$\tilde{\text{TS}}$2**) is estimated to be less than 6.4 kcal/mol higher in energy than **E-II-a**. Water can now be eliminated with a barrier of 25.3 kcal/mol to form the imine (Figure 13a, Structure **E-III**). The net energy requirement of 25.3 kcal/mol for this first step is already much higher than that for path D discussed above, so this mechanism that involves two methylamine molecules is not likely to be competitive. Although the second part of this mechanism, addition of NO_2CH_2 and regeneration of the catalyst, has been explored in detail, it is not presented here to save space.

Conclusions

Several pathways for the nitroaldol reaction have been compared to determine the energetically most favorable mechanism. To form the final nitroalkene products, the reaction must pass through either a 2-nitroethanol or an imine intermediate. The highest barrier **(23.6 kcal/mol)** in the uncatalyzed mechanism is for the formation of 2-nitroethanol (Figure 3, Structure **III**). Addition of solvent effects represented by PCM increases this barrier by 20 kcal/mol. When an *ab initio* solvent molecule is added, the net energy requirement of nitroalkene formation is reduced to less than 2.9 kcal/mol. The decrease in the barrier height is due to a change from a one step mechanism with a strained 4-center transition state to a two-step mechanism that involves a 6-center transition state. Effectively, the solvent molecule serves a role of a catalyst in this manner.

Barriers for formation of nitroalcohol are lowest in the presence of an explicit solvent molecule. Once the nitroalcohol is formed it can undergo elimination by forming a 6-center transition state with the methylamine catalyst. The barrier for elimination is very high unless the amine catalyst is present. The net energy requirement for this "classical" mechanism is 20.6 kcal/mol. In an alternative mechanism: first methylaminomethanol is formed, then water is eliminated forming an imine intermediate and finally the catalyst is regenerated. Then net energy requirement for this mechanism is 17.0 kcal/mol, which is only 3.6 kcal/mol lower in energy than the "classical" mechanism, indicating that neither mechanism can be eliminated as a possibility. When a second amine is added to the system the net energy requirement actually increase, demonstrating that a second amine molecule is not necessary to facilitate the reaction. Solvent effects were shown to be very important for nitroalcohol formation, however PCM does not give a realistic representation of the implicit solvent effects. Future work will investigate the effect of the silica pore and explicit observer solvent molecules.

Acknowledgements. This research was supported by a grant from the U.S. Department of Energy to the Ames Laboratory, office of BES, under contract DE-AC02-07CH11358.

References

(1) Henry, L., C.R. *Hebd. Seances Acad. Sci.*, **1895**, *120*, 1265

(2) Rosini, G., *Comprehensive Organic Synthesis*, ed. B. M. Trost, Pergamon, Oxford, **1999**, *1*, 321.

(3) Huh, S., Chen, H-T., Wiench, J.W, Pruski, M., and Lin, V.S.-Y. *J. Am. Chem. Soc.,* **2004** *126*, 1010.

(4) Demicheli, G., Maggi, R., Mazzacani, A., Righi, P., Sartori, G., and Bigi, F., *Tetrahedron Letters* **2001,** *42*, 2401

(5) Lecea, B., Arrieta A., Morao, I., Cossío, F.P., *Chem. Eur. J.*, **1997**, *3*, 20.

(6) Krishnan, R. and Pople, J.A, *Int. J. Quantum Chem.* **1978**, *14*, 91.

(7) Beksic, D., Bertran, J., and Lluch, J.M. *J. Phys. Chem. A* **1998**, *102,* 3977.

(8) Hall, N.E.; Smith, B.J., *J. Phys. Chem. A* **1998**, *102*, 4930.

(9) Curtiss, J.A.; Ragahavachari, K.; Pople, J.A. *J. Chem. Phys.* **1993**, *98*, 1293.

(10) Møller, C.; Plesset, M. S. *Phys. Re*V. **1934**, *46*, 618.

(11) Fletcher, G. D.; Rendell, A. P.; Sherwood, P. *Mol. Phys.* **1997**, *91*, 431.

(12) Hehre W.J..; Ditchfield R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(13) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acta* **1973**, *28*, 213.

(14) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(15) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.

(16) Gonzalez, C.; Schlegel, H.B., *J. Phys. Chem.* **1990**, *94*, 5523

(17) Gonzalez, C.; Schlegel, H.B., *J. Phys. Chem.* **1991**, *95*, 5853.

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

(19) Truhlar, D. G.; Gordon, M. S. *Science* **1990**, *249*, 491.

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

(21) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2000**, *149*, 71.

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

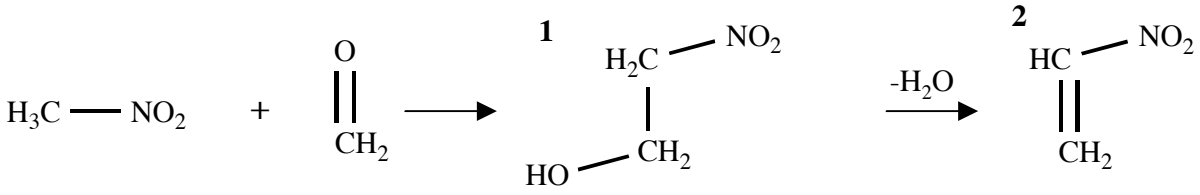
(23) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833-1840, 1841-1846, 2338-2342, 2343-2346.

(24) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

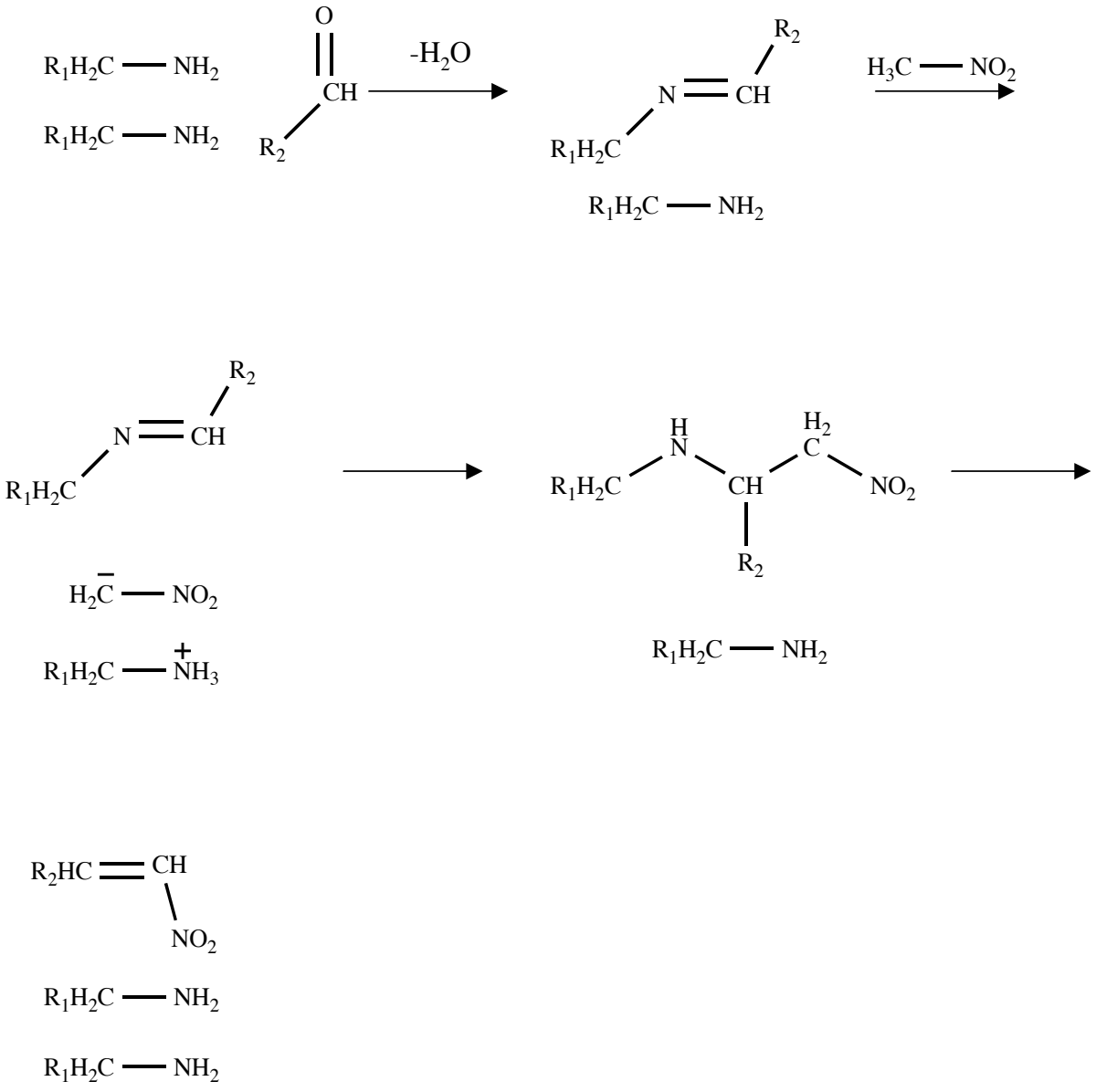
(25) Schmidt, M.W., et al., *Journal of Computational Chemistry*. **1993**, *14*, 1347.

(26) Schmidt, M. W., Gordon, M. S. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, **2005**.

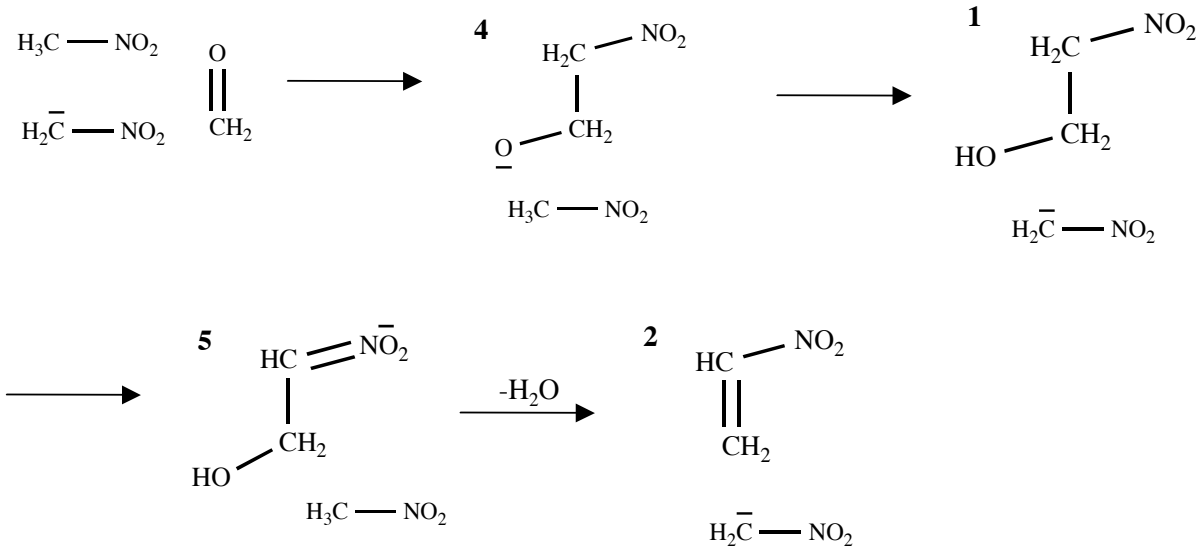
(27) Kudo, T.; Gordon, M. S. J. Phys. Chem. A 2000, 104, 4058.



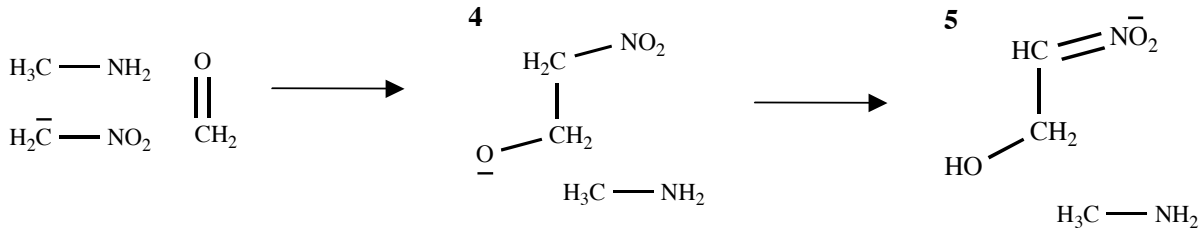
Scheme 1. The nitroaldol (Henry) reaction.



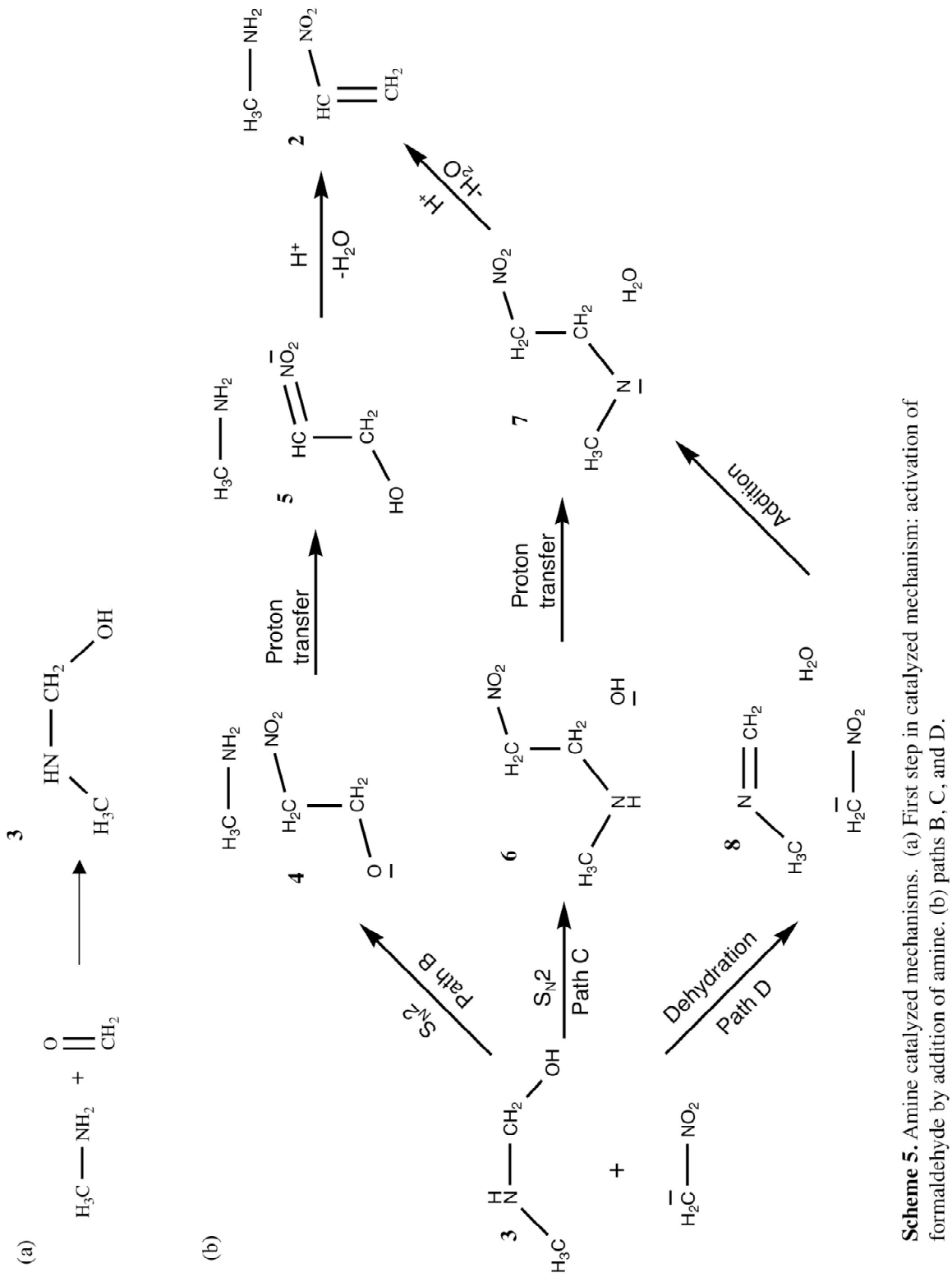
Scheme 2. Proposed mechanism for MSN catalyzed nitroaldol reaction. R_1=$(CH_2)_2$-MCM-41, R2=C_6H_6.



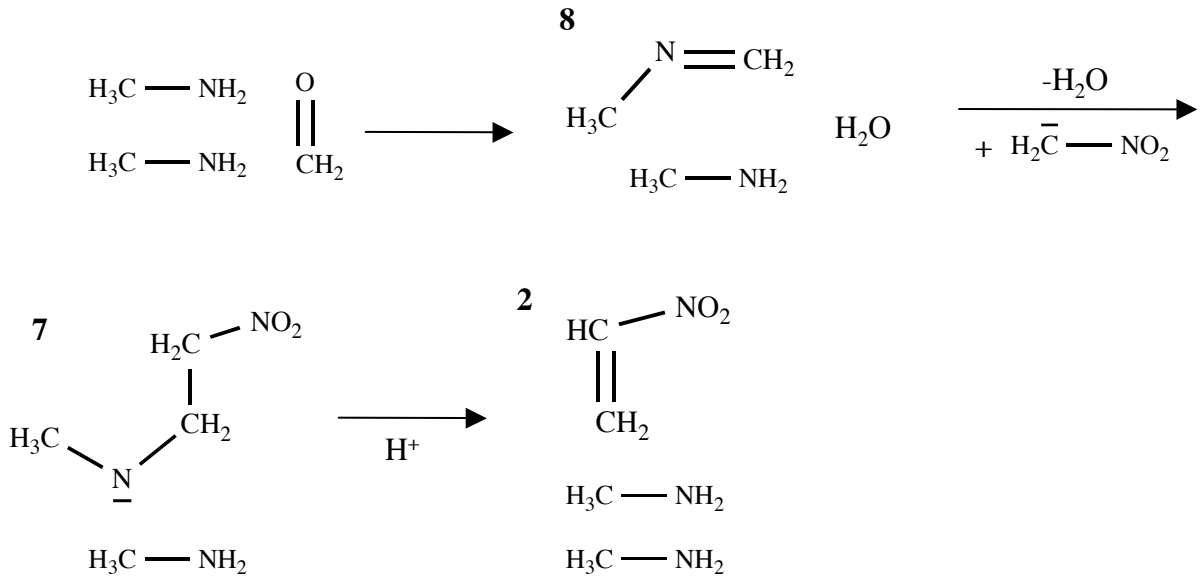
Scheme 3. Uncatalyzed mechanism with an *ab intitio* nitromethane solvent molecule.



Scheme 4. Amine "assisted" mechanism of the nitroaldol reaction.



Scheme 5. Amine catalyzed mechanisms. (a) First step in catalyzed mechanism: activation of formaldehyde by addition of amine. (b) paths B, C, and D.



Scheme 6. Catalyzed mechanism pathway D with an additional methylamine molecule.

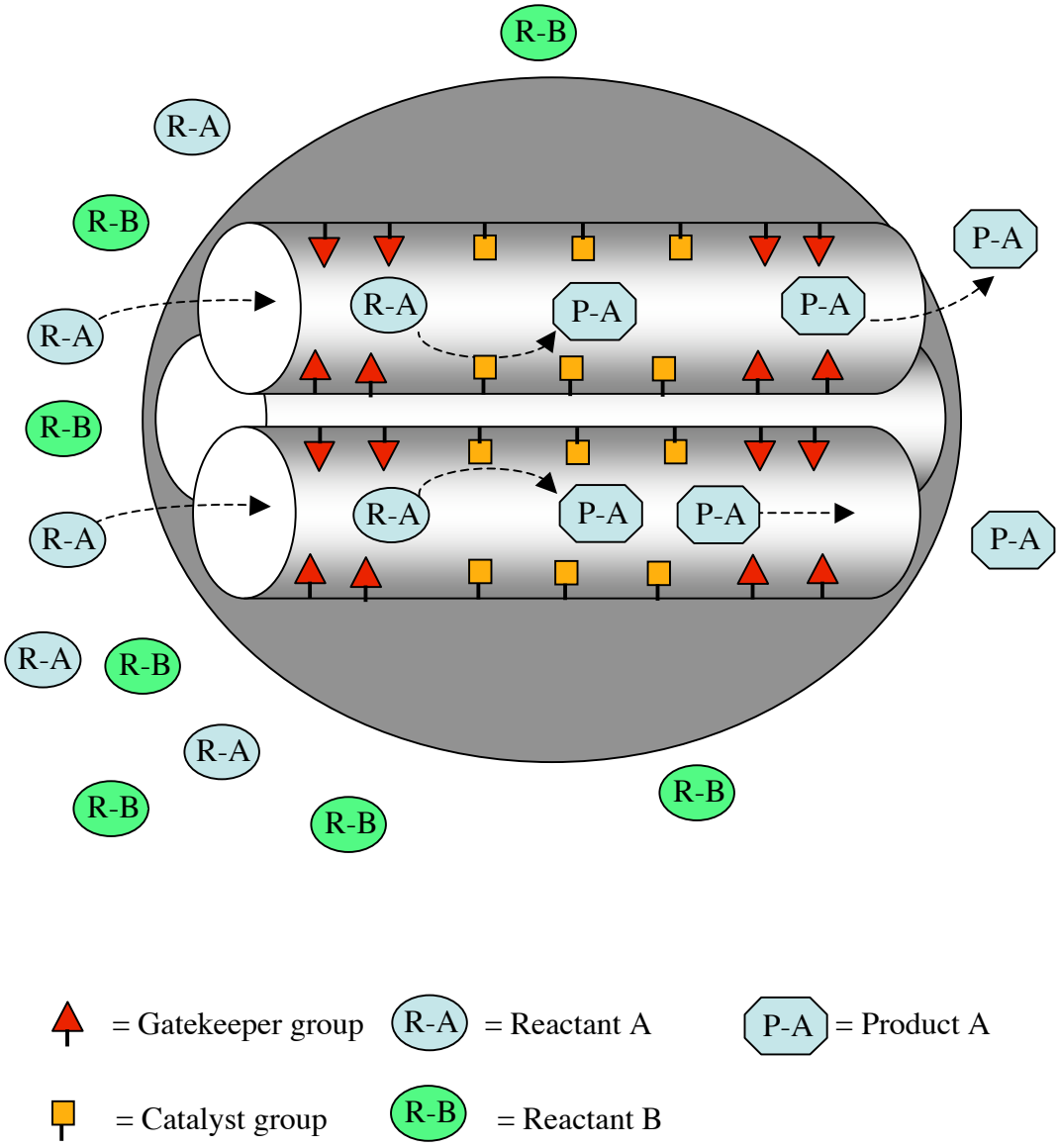
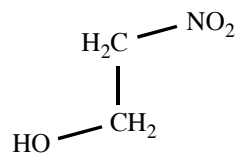


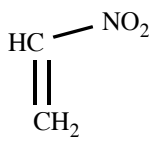
Figure 1. Schematic of a multi-functionalized mesoporous silica system.

1



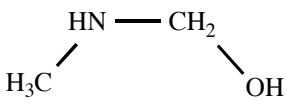
2-nitroethanol

2



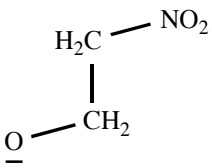
nitroethene

3



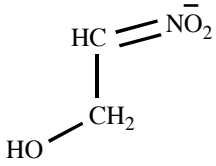
methylaminomethanol

4



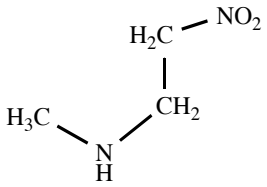
2-nitroethoxide

5



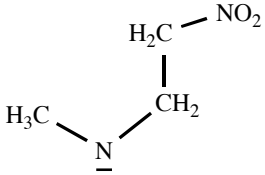
2-nitroethanol nitronate

6



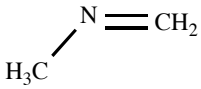
N-methyl-2-nitroethanamine

7



N-methyl-2-nitroethanamine anion

8



N-methylmethanimine

Figure 2. Summary of reaction intermediates and products shown in Schemes 1, 3 and 4.

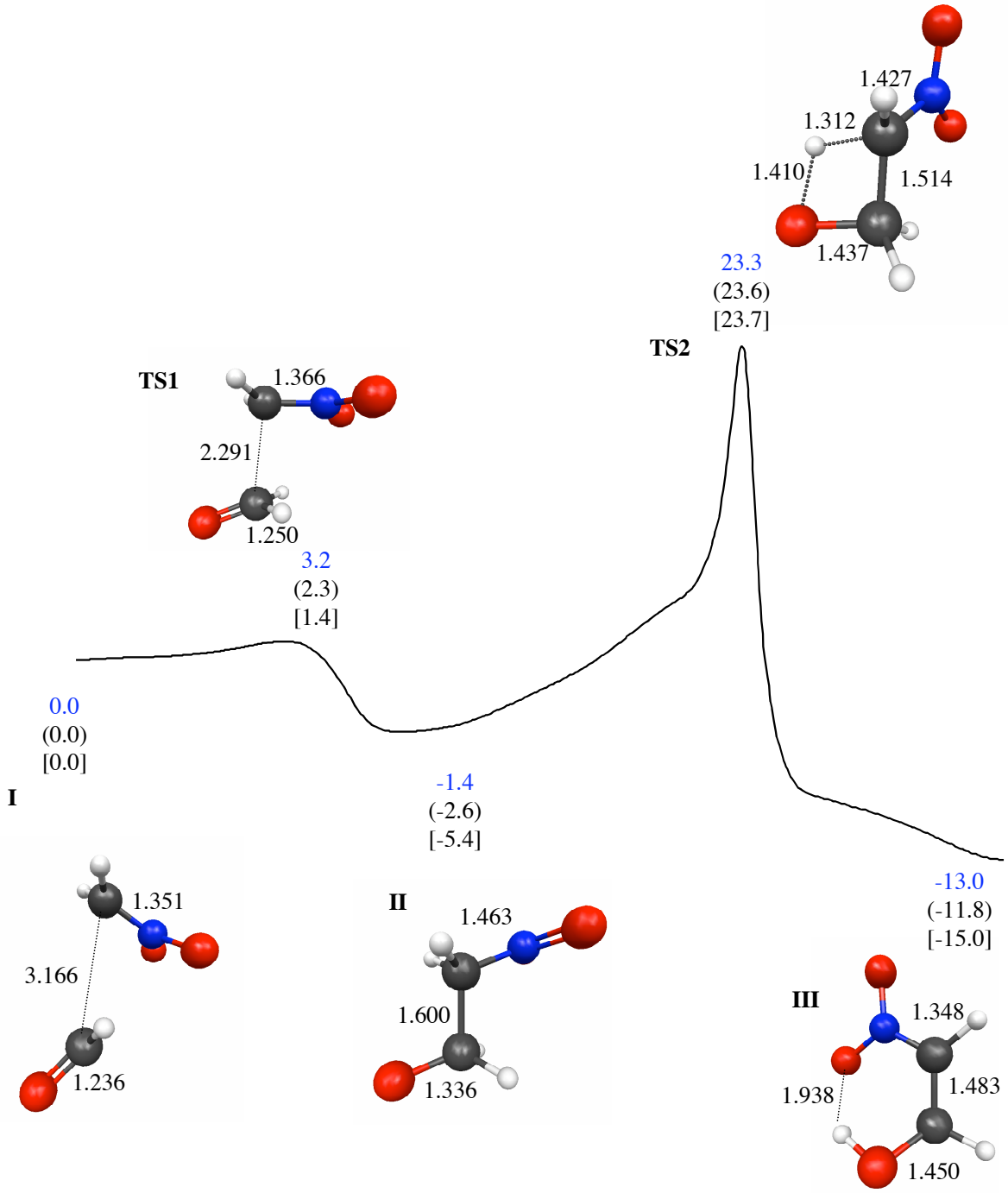


Figure 3. MP2/6-31+G(d) minimum energy path for the uncatalyzed nitroaldol reaction. CCSD(T)/aug-cc-pVDZ//MP2/6-31+G(d) single point energy calculations in blue. MP2/6-31+G(d) energies in parentheses. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

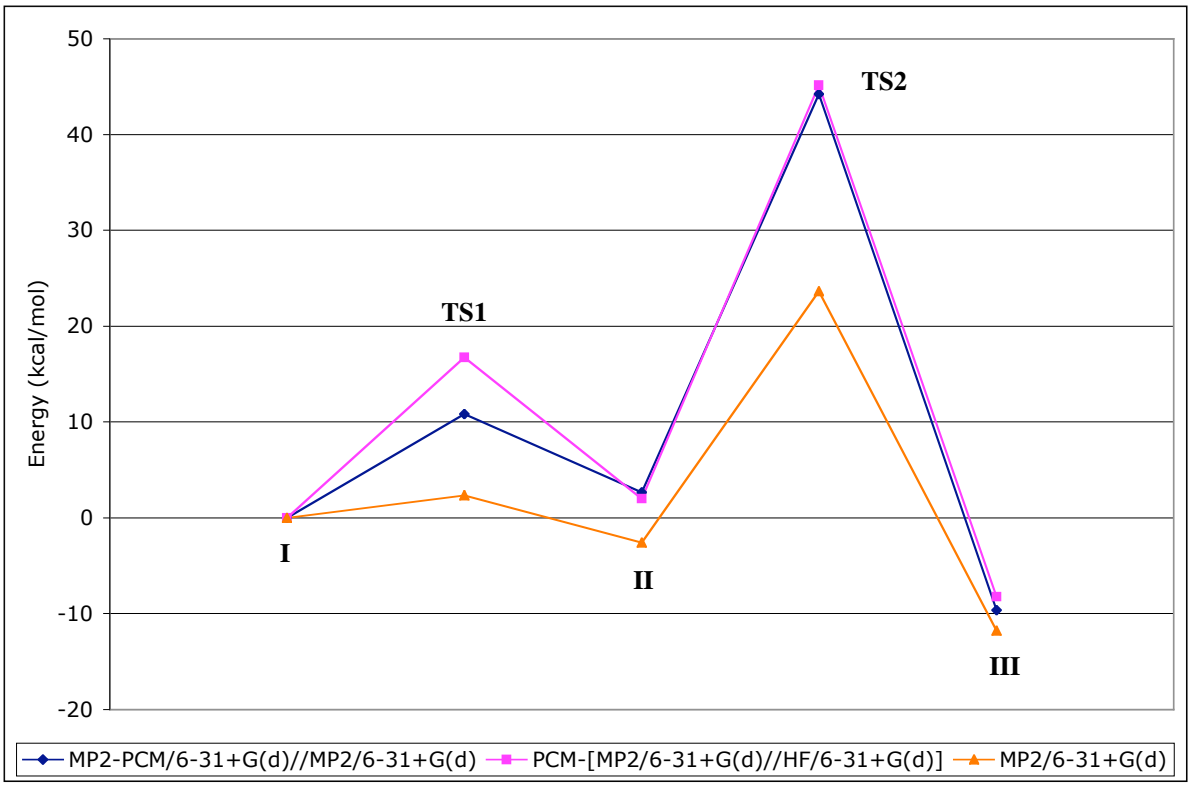


Figure 4. Solvation effects with PCM for uncatalyzed mechanism.

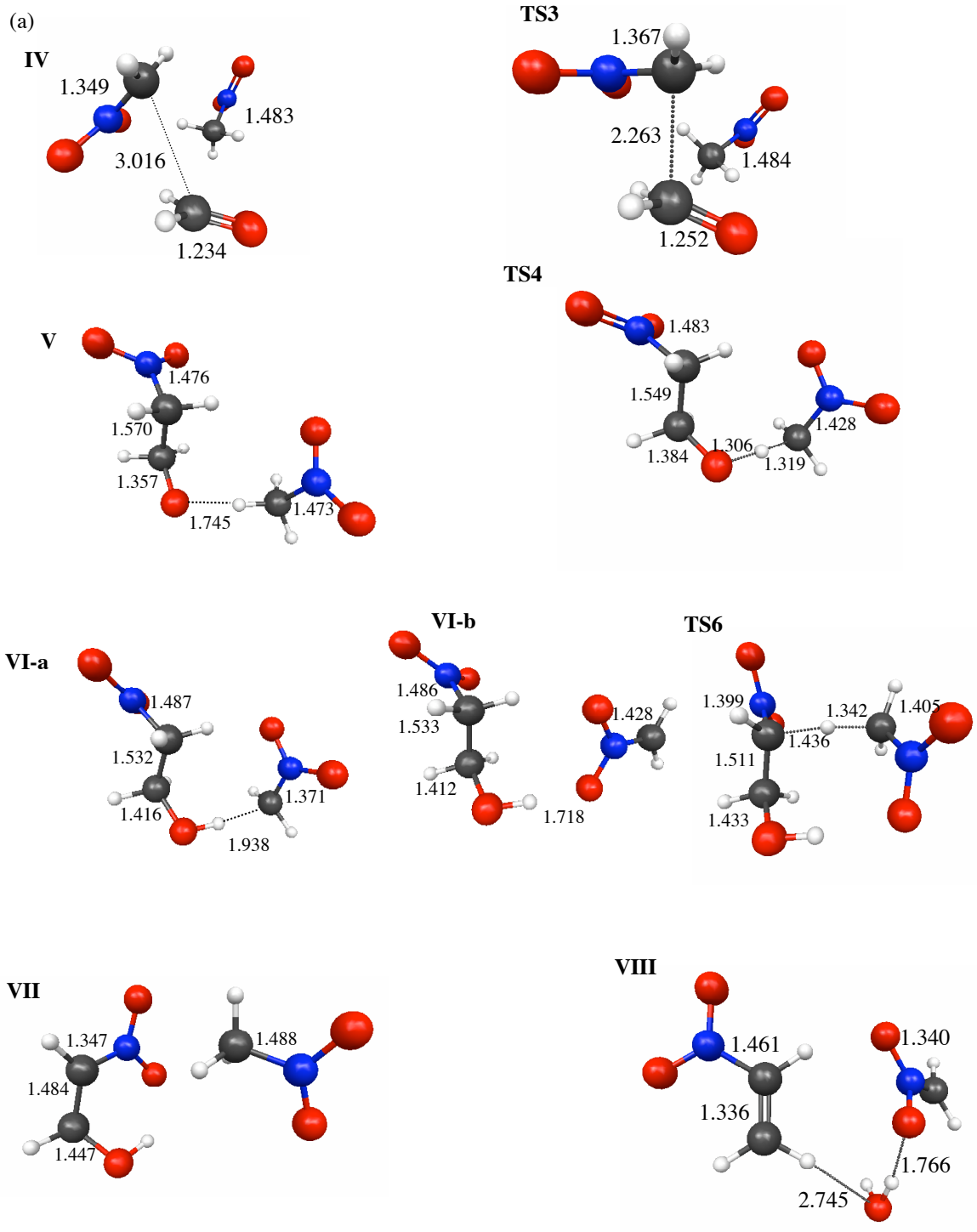


Figure 5. (a) MP2/6-31+G(d) reactants and products for uncatalyzed nitroaldol reaction with *ab initio* solvent molecule. (b) Uncatalyzed MP2/6-31+G(d) MEP. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

(b)

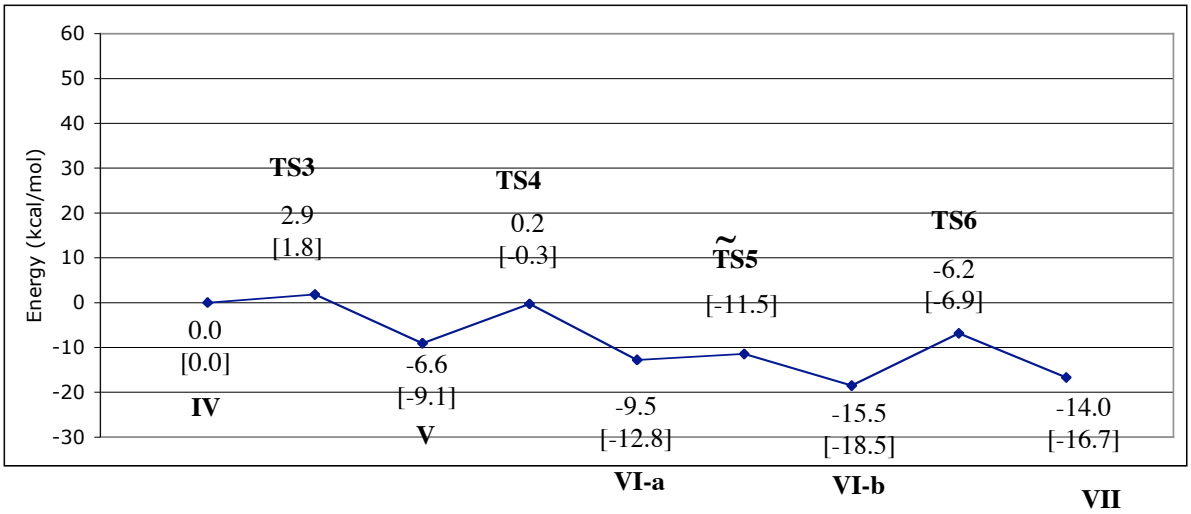


Figure 5. (continued)

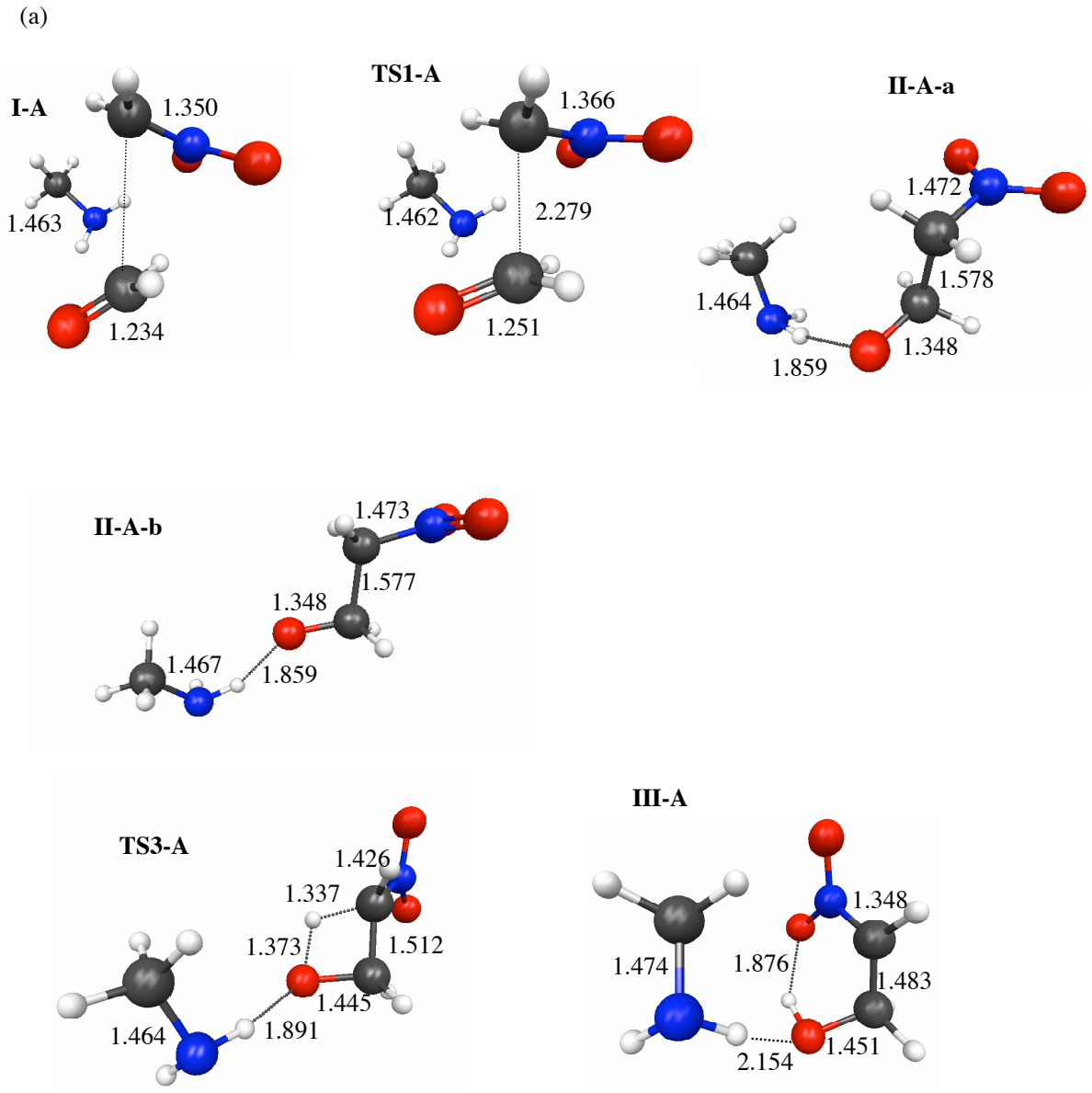


Figure 6. (a) MP2/6-31+G(d) minima and transition state structures for amine "assisted" mechanism. (b) MP2/6-31+G(d) minimum energy path. Relative energies without ZPE in brackets. Energies in kcal/mol. MP2/6-31+G(d) energies in parenthesis. Improved relative energies from CCSD(T)/aug-cc-pVDZ//MP2/6-31+G(d) single point energy calculations. Bond lengths in Å.

(b)

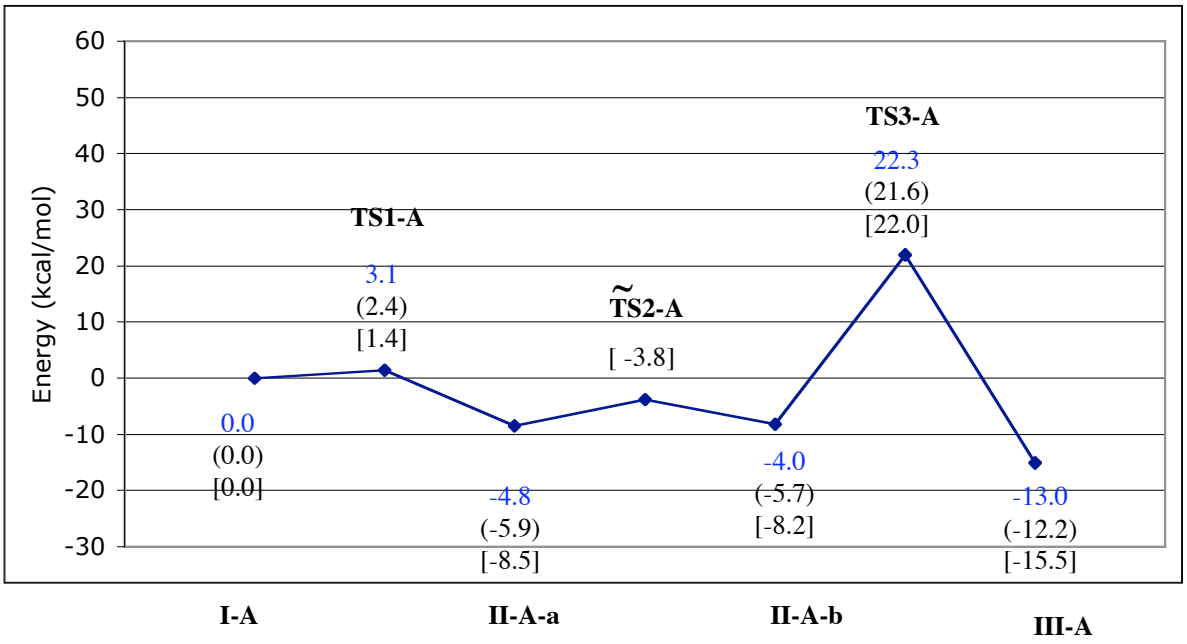


Figure 6. (continued)

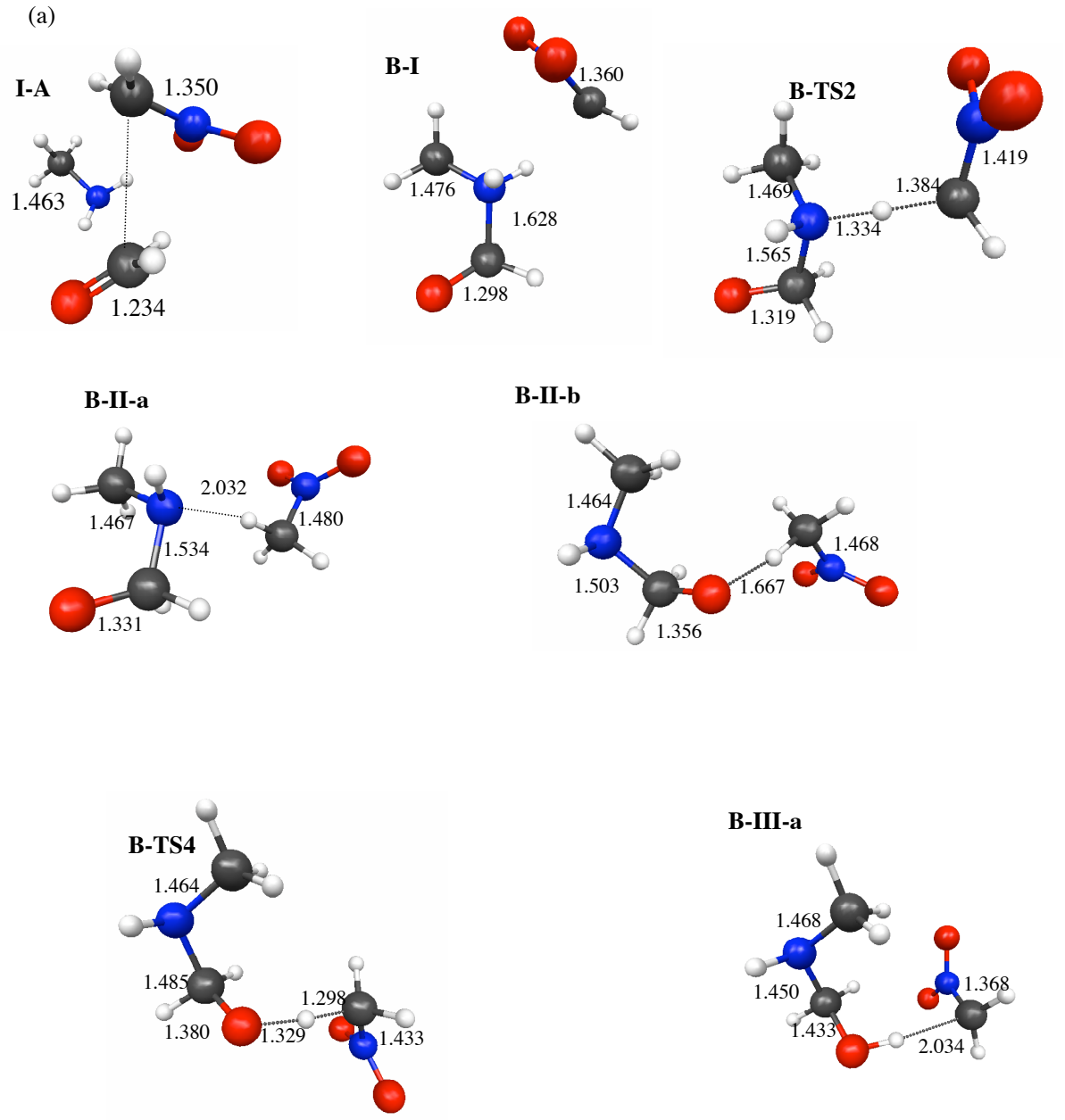


Figure 7. (a) MP2/6-31+G(d) minima and transition state structures for formation of **3**. (b) MP2/6-31+G(d) MEP. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

(b)

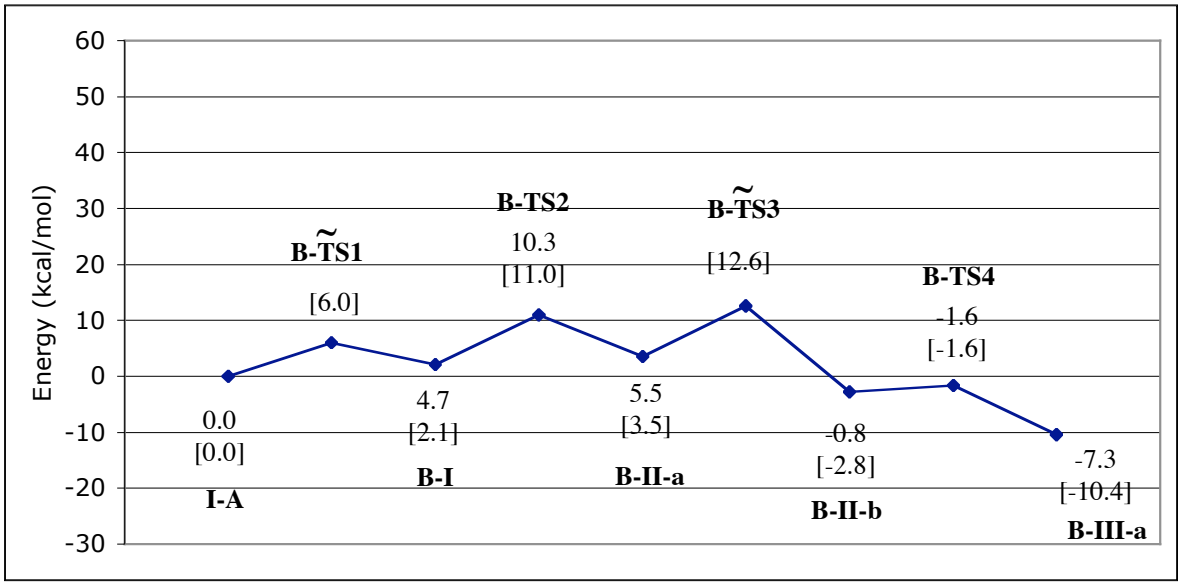


Figure 7. (Continued)

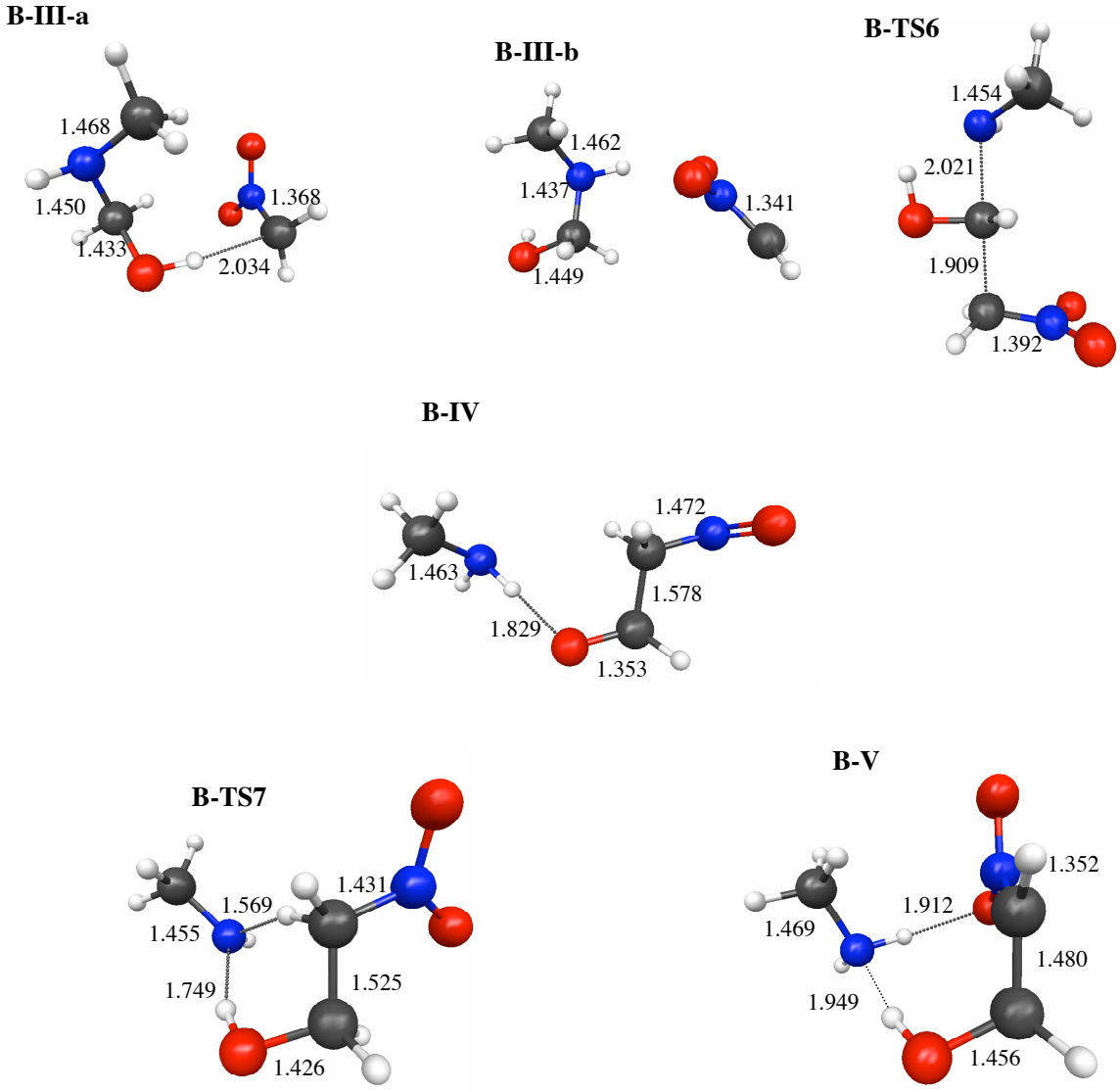


Figure 8. (a) MP2/6-31+G(d) minima and transition state structures for catalyzed mechanism pathway B. (b) MP2/6-31+G(d) MEP. The zero of energy corresponds to structure **I-A**. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

(b)

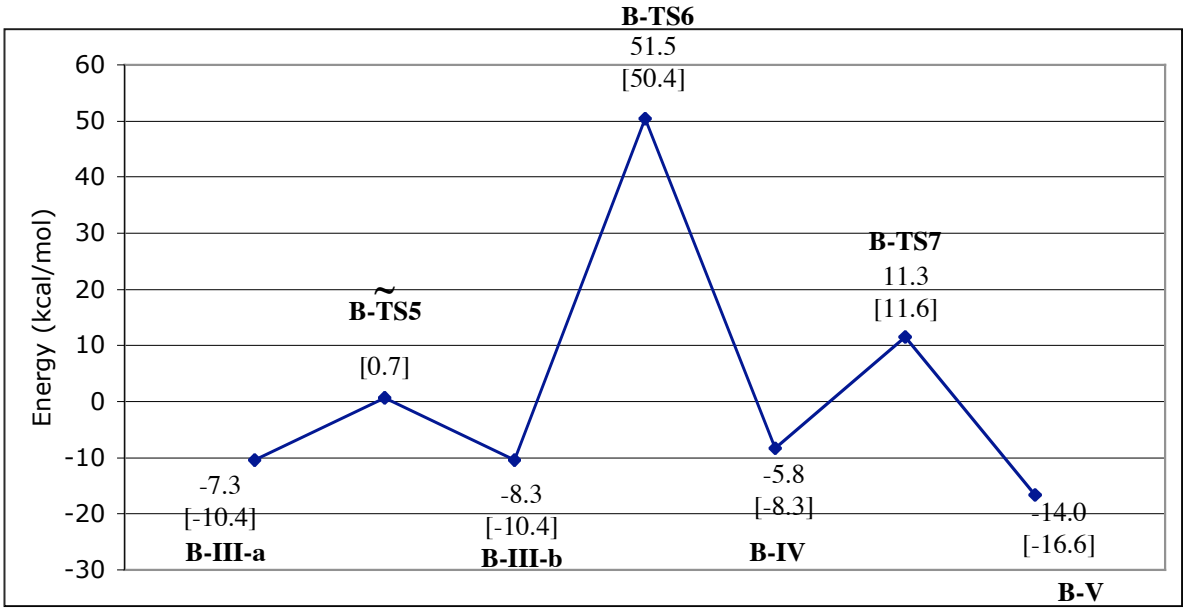
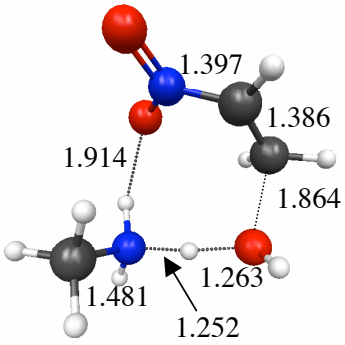
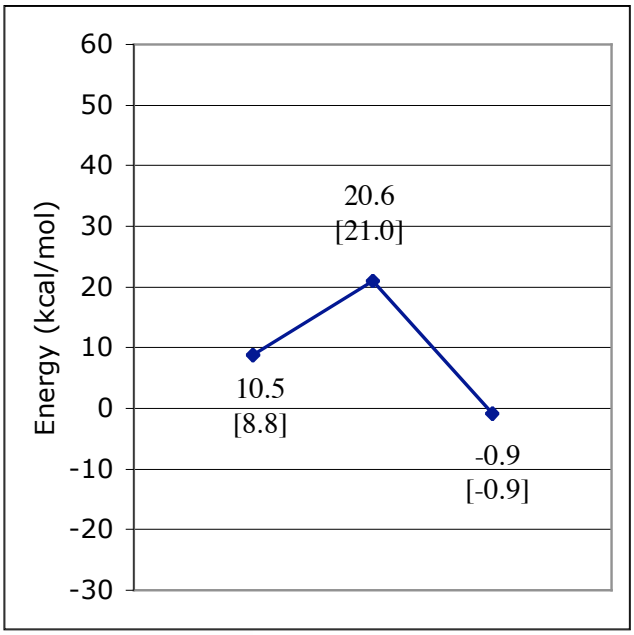


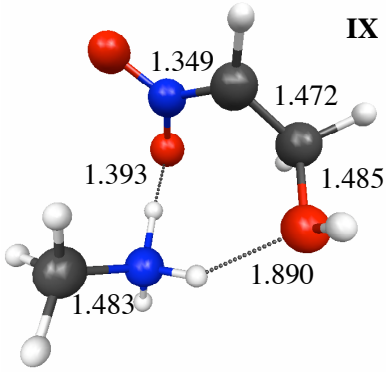
Figure 8. (continued)



TS7



IX



X

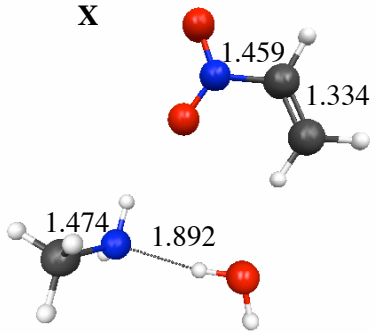


Figure 9. MP2/6-31+G(d) MEP for elimination of a nitroalcohol to give a nitroalkene. The zero of energy corresponds to the reactant complex of CH_2O + CH_3NO_2 + CH_3NH_2. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

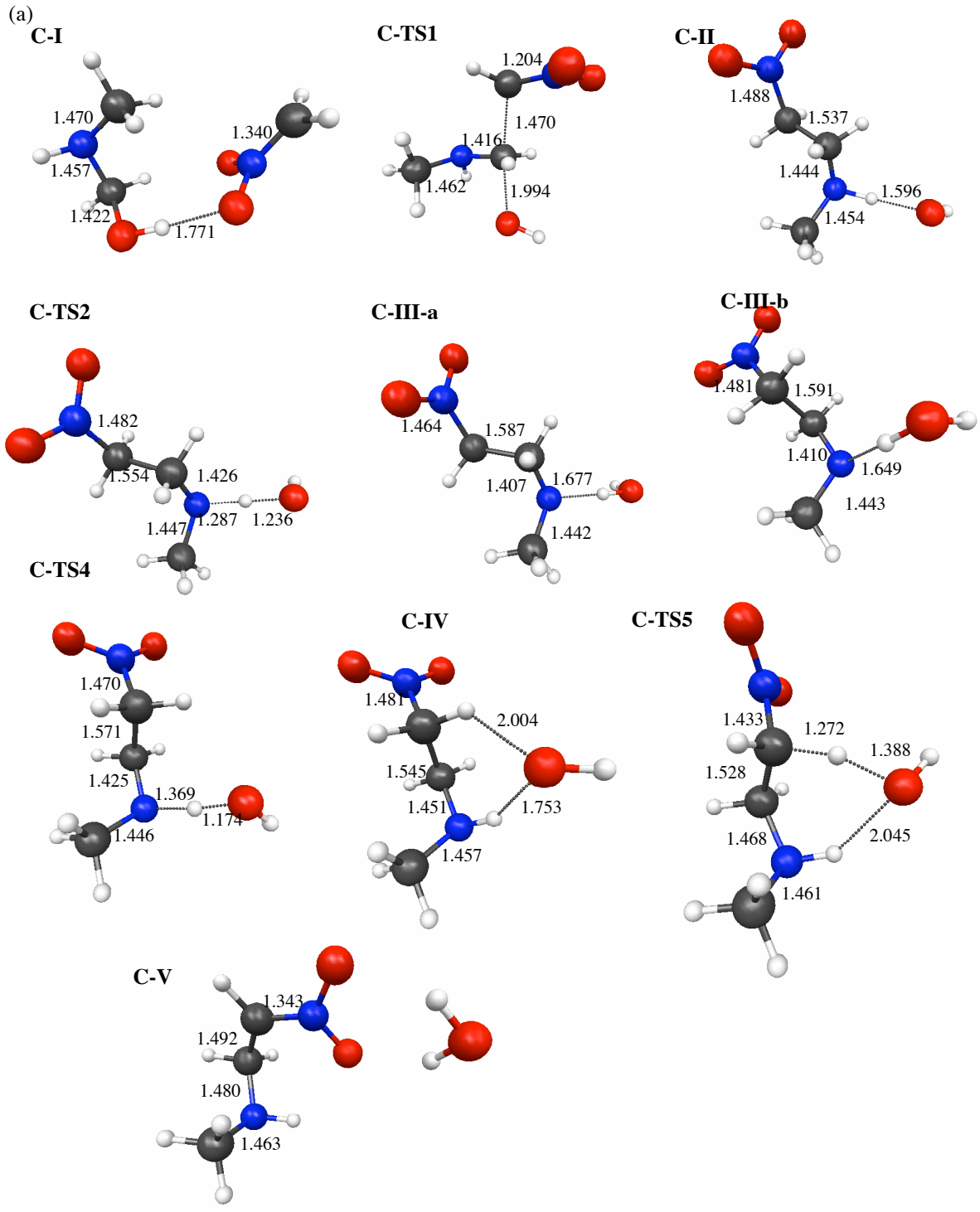


Figure 10. (a) MP2/6-31+G(d) minima and transition states in catalyzed mechanism pathway C. (b) MP2/6-31+G(d) MEP. The zero of energy corresponds to structure **I-A**. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

(b)

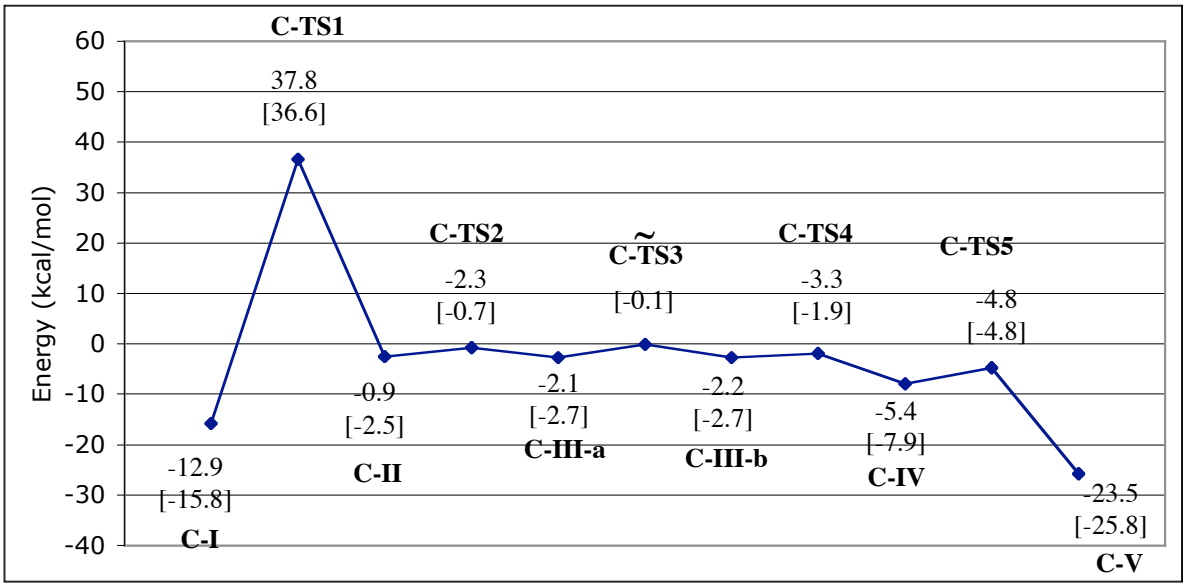


Figure 10. (continued)

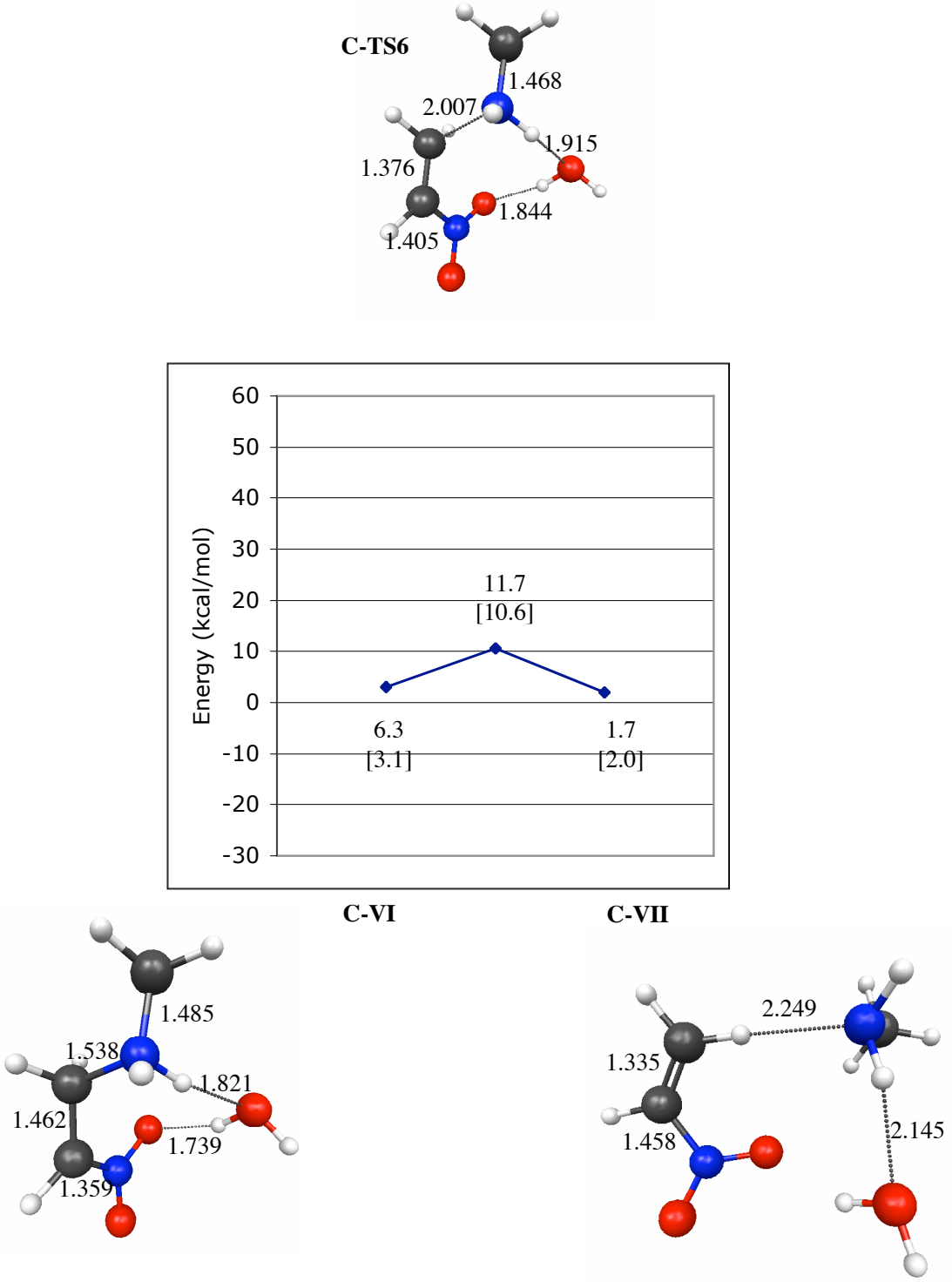


Figure 11. MP2/6-31+G(d) minimum energy path for regeneration of amine catalyst in catalyzed mechanism pathway D. The zero of energy corresponds to the reactant complex of $CH_2O + CH_3NO_2 + CH_3NH_2$. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

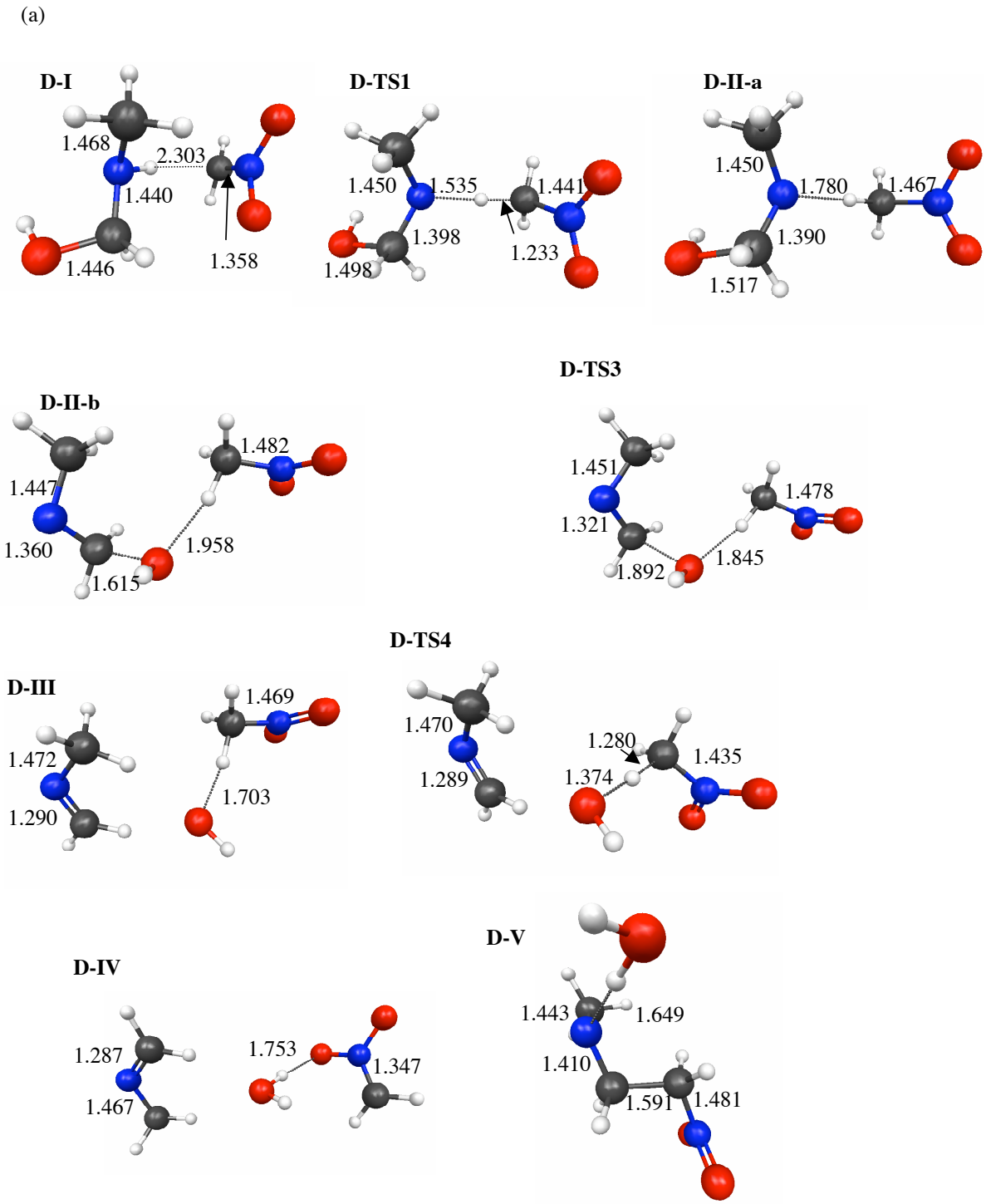


Figure 12. (a) MP2/6-31+G(d) minima and transition states for catalyzed mechanism pathway D. (b) MP2/6-31+G(d) MEP. The zero of energy corresponds to structure **I-A**. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

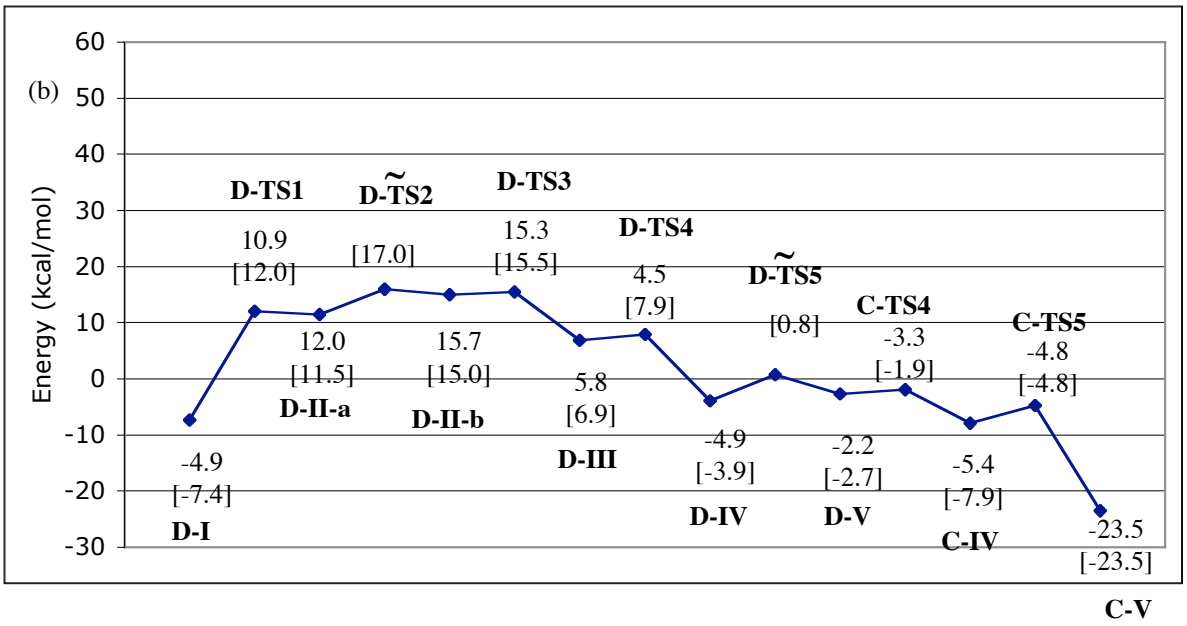


Figure 12. (continued)

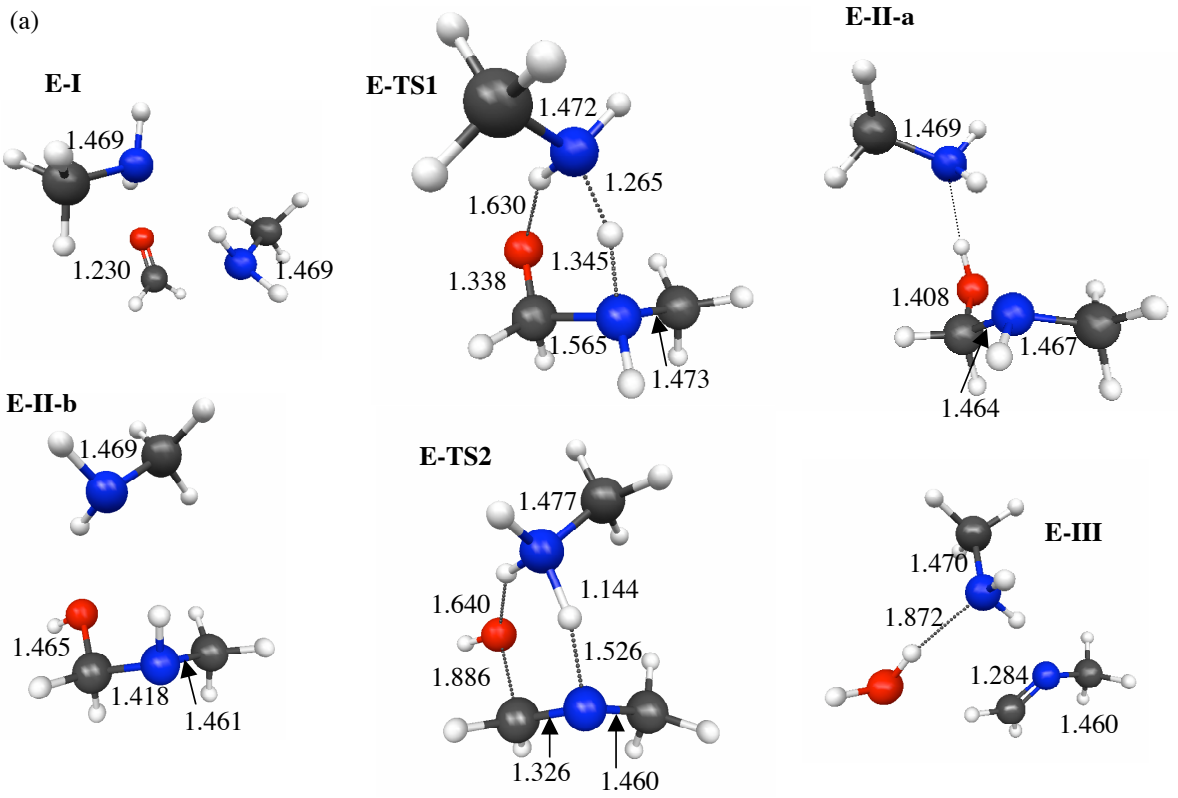


Figure 13. (a) MP2/6-31+G(d) minima and transition states for imine formation with an additional methylamine, (b) MP2/6-31+G(d) minimum energy path. Relative energies without ZPE in brackets. Energies in kcal/mol. Bond lengths in Å.

CHAPTER 5. INTERACTION OF THE UNIVERSAL FORCE FIELD WITH THE EFFECTIVE FRAGMENT POTENTIAL METHOD

Deborah Zorn, Victor S.-Y. Lin, Marek Pruski and Mark S. Gordon

Abstract. In order to properly describe reactions in heterogeneous catalyst systems, the reactants, solvent, and bulk effects of the surface must be taken into account. Embedded cluster QM(quantum mechanics)/MM (molecular mechanics) methods can treat reactions on surfaces (the gas-surface interface), and the effective fragment potential method (EFP) can accurately treat the solvent effects on reactions (the gas-liquid interface). In order to create a QM/MM/EFP hybrid method for treatment of heterogeneous catalytic systems in the presence of a solvent (the liquid-surface interface), an EFP-MM interaction potential has been developed. Example calculations on small clusters of silica and water have been carrier out.

I. Introduction

Functionalized mesoporous silica nanosphere (MSN) based catalysts have been found to selectively catalyze many different types of reactions.[1-6] In these systems the silica is not simply an inert support with size/shape sieving selectivity. Rather, the selectivity is determined by covalent and non-covalent interactions between reactants and functional groups immobilized on the inside of the silica pores. In a bifunctionalized MSN system there are two different functional groups: the first group catalyzes the reaction and the secondary groups control the selectivity. The secondary groups are called “gate keepers” because they

prevent unwanted reactants from entering the catalyst pore by non-covalent (e.g., via hydrophobic or hydrophilic) interactions. In addition to their selectivity, advantages of these new MSN catalysts include their inert stationary phase, large surface area, and tunable pore size. A schematic of a multi-functionalized system is shown in Figure 1. In this example, the gatekeeper groups only allow reactant A to enter the functionalized pore, yielding product A selectively. In a paper by Huh et al.[1] a bifunctionalzed MSN system was reported to selectively catalyze the nitroaldol reaction, in which condensation between a nitroalkane and an aldehyde yields a nitroalcohol that can undergo dehydration to yield a nitroalkene. In this system the catalytic group is a 3-[2-(aminoethylamino)ethylamino]propyl group, and the secondary groups are ureidopropyl, mercaptopropyl and allyl groups. Accurate gas phase calculations have been carried out on a nitroaldol reaction by Zorn et al. [7] with methylamine used as a model catalyst. To properly treat the entire system, including the catalyst, the pore and the solvent, a computational method must properly account for the electronic structure of the reactants, the effects of the silica and the effects of the surrounding solvent.

To efficiently and accurately treat these MSN heterogeneous catalysis systems, a hybrid approach that employs both quantum mechanics (QM) and model potentials (for the solvent and the non-reactive part of the functionalized pore) might provide an effective model. The electronic structure of the reacting species and the immobilized catalyst groups must be treated with quantum mechanics. The important nonbonded solvent-substrate interactions can usually be reasonably accounted for using an explicit solvent model, and the largely non-interacting bulk silica support can be treated with molecular mechanics (MM).

Several of the computational components that are needed for the study of heterogeneous catalysis in the presence of a solvent are already available in or interfaced with the GAMESS (General Atomic and Molecular Electronic Structure System) [8,9] computer code. These are the surface integrated molecular orbital molecular mechanics (SIMOMM) embedded cluster method[10] that was designed for QM/MM calculations on surfaces, and the effective fragment potential method (EFP)[11,12] method that was developed for investigations of explicit solvent effects. The goal of this work is to combine the SIMOMM and EFP methods, thereby creating a QM/MM/EFP method. The QM/MM/EFP energy can be written as:

$$E_{TOTAL} = E_{QM} + E_{MM} + E_{EFP} + E_{QM\text{-}EFP} + E_{QM\text{-}MM} + E_{MM\text{-}EFP} \qquad (1)$$

In Eq. (1) E_{QM}, E_{MM}, and E_{EFP} are the internal energies of the QM, MM, and EFP regions of a composite system, respectively, while the last three terms are the corresponding interaction energies. All except the last term in Eq. (1) have previously been derived and coded. [10-12] The last term is the focus of the current work.

In the SIMOMM embedded cluster method, a surface to be modeled is divided into two regions: the bulk region and the "action" region. The bulk region is a large cluster that models the surface of interest. Carved out of the center of the bulk model of the surface is a smaller cluster where the "action" (chemistry) takes place. In SIMOMM the action region is treated with quantum mechanics, and the bulk region is treated with an MM force field.

The effective fragment potential (EFP) method is an explicit solvent model, which represents the important non-bonded interactions of solvent molecules with each other and

with a QM solute. In the most general EFP model, these interactions include: Coulomb, induction, exchange repulsion, charge transfer, and dispersion interactions. In EFP the system is divided into two regions: the quantum (solute) region and the EFP solvent region. The total energy of the QM-EFP system is

$$E_{interaction} = E_{QM-EFP} + E_{EFP-EFP} \quad (2)$$

The interaction energy includes the interactions between the quantum and EFP regions and the interactions between the solvent molecules and other solvent molecules.

The original EFP method, called EFP1[12], was designed specifically for water and has been implemented for three levels of theory: Hartree-Fock (HF), density functional theory (DFT), and second order perturbation theory (MP2). In EFP1 the energy is a sum of three terms: Electrostatic, polarization and a fitted remainder term, which accounts for all interactions not included in the first two terms.

$$E_{Interaction} = E_{Electrostatic} + E_{Polarization} + E_{Remainder} \quad (3)$$

The electrostatic term is represented by a distributed multipolar analysis (DMA), in which the multpoles are expanded up to octopoles. The expansion points are the atom centers and bond midpoints. A damping term is used to account for overlapping charge densities at small intermolecular distances, and a distance cutoff is used for this damping term[13,14]. The polarization of each molecule by the surrounding molecules is obtained using a finite field model and iterated to self consistency using localized molecular orbital (LMO) polarizability tensors. The remainder term is fitted to a functional form,[12] in which the fitted parameters are obtained by subtracting the first two terms in Eq. (3) from the water dimer interaction

potential at many points on the water dimer potential energy surface. For the HF implementation of EFP1, $E_{remainder}$ contains contributions from exchange repulsion E_{exrep} and charge transfer E_{ct}. For EFP1/DFT[efpdft], $E_{remainder}$ includes some short-range correlation, and for EFP1/MP2, there is a separately fitted dispersion term[efpmp2].

The general EFP method (EFP2) has no fitted parameters; its interaction energy can be expressed as

$$E_{Interaction} = E_{Elec} + E_{Pol} + E_{ExRep} + E_{CT} + E_{Disp} \quad (4)$$

Because there are no empirically fitted parameters, an EFP2 can be generated for any molecule. The internal geometries are fixed (no intra-fragment vibrations) in both EFP1 and EFP2.

Due to the internal rigidity of fragments, Nemukhin et al.[15] interfaced the EFP1 method with MM force field methods in the molecular mechanics package Tinker[16,17] to facilitate the modeling of conformational changes in biological molecules, represented by dipeptides, that are solvated by water. In their method, fragment-fragment interactions were replaced by force field interactions calculated by one of the molecular mechanics options in Tinker, creating a new flexible EFP/MM scheme. The force field and parameters used were from the OPLS-AA force field[18]. The authors took two approaches to modeling the dipeptide water system. In the first approach the dipeptide was treated at an *ab inito* level of theory and the waters were represented with the flexible EFP/MM scheme. In the second approach the dipeptide was decomposed into 8 fragments, which were modeled with the EFP/MM scheme and the waters were treated with an *ab inito* level of theory. Both of these

models were able to correctly describe the conformational changes of a didpeptide in the presence of water.

Although the approach taken in the method by Nemukhin et al. can properly account for the interactions between the dipepetide and each solvent molecule, the MM treatment of the EFPs cannot adequately account for the interactions between solvent molecules due to the lack of an accurate intermolecular potential for water, and because of the inherent limitations of the MM method. The importance of the structure of the surrounding solvent molecules on the electronic structure of the solute has been demonstrated for systems such as solvated glycine[19] and alanine[20], for which the structure of the surrounding water has a significant impact on the relative stabilities of the nonionic and zwitterionic species. That method also only has two regions: a QM region and an EFP region. Applications to reactions on surfaces surrounded by a solvent require three regions: a MM region for the bulk, a QM region for the "action" region, and a EFP region for the solvent molecules.

Currently the implementation of SIMOMM uses an interface with the molecular mechanics package TINKER. There are several choices of force field potentials in TINKER. These are mainly designed to treat biological and organic systems, so parameters are primarily limited to atoms that commonly appear in such species. In order to model metal oxide surfaces, such as silica and titania the Universal Force Field was implemented directly into GAMESS. The universal force field (UFF)[21] is a general all atom force field that has been applied to organic molecules,[22] metallic complexes,[23] and main group compounds.[24] It is therefore applicable to a broad range of interesting problems related to heterogeneous catalysis.

In traditional force fields large sets of parameters are necessary, in order to account for all possible combinations of atoms that could be involved in a bond, angle, or torsion. UFF replaces these large sets of parameters with a smaller set of parameters for each atom type. Currently there are 127 atom types available in UFF, based on hybridization and oxidation state. Force field parameters can be generated for every possible combination of atom types based on the connectivity of the atoms. The UFF energy is given in Eq. (5).

$$E_{UFF} = E_{bond} + E_{angle} + E_{tors} + E_{invers} + E_{elec} + E_{vdw} \quad (5)$$

The UFF describes the bond stretching term, E_{bond}, as a harmonic oscillator:

$$E_{bond} = k_{IJ}(r - r_{IJ})^2 \quad (6)$$

or a Morse function:

$$E_{bond} = D_{IJ}\left[e^{-\alpha(r-r_{IJ})^2} - 1\right]^2 \quad (7)$$

The user may choose the functional form of E_{bond} that is best for their application.

In the bond stretching functions, k_{IJ} is the stretching force constant in kcal mol^{-1} $Å^{-2}$, r_{IJ} is the equilibrium bond length, and D_{IJ} is the bond dissociation energy. The parameter α is obtained from k_{IJ} and D_{IJ}:

$$\alpha = [k_{IJ} / 2\, D_{IJ}]^{1/2} \quad (8)$$

The equilibrium bond length is the sum of bond radii parameters of the two atoms, plus a bond order correction and an electronegativity correction. The bond stretching force constants come from Badger's rules.[25]

The angle stretching contribution to the energy, E_{angle}, is a truncated Fourier expansion, with the equilibrium angle defined by the atom type of the central atom. For linear, trigonal-planar, and octahedral molecules, the expression is:

$$E_{angle} = \frac{K_{IJK}}{n^2}\left[1 \pm \cos(n\theta)\right] \tag{9}$$

and for general nonlinear case E_{angle} is:

$$E_{angle} = K_{IJK}\left[C_0 + C_1\cos(\theta) + C_2\cos(2\theta)\right] \tag{10}$$

where $C_2 = 1/(4\sin^2(\theta_0))$, $C_1 = -4C_2\cos(\theta_2)$, and $C_0 = C_2(2\cos^2(\theta_0)+1)$. The angular force constant is defined by the equilibrium angle and its connectivity.

The torsional energy contribution, E_{tors}, is represented with a truncated cosine Fourier expansion:

$$E_{tors}(\phi) = K_{IJKL}\sum_{n=0}^{m} C_n\cos(n\phi_{IJKL}) \tag{11}$$

where K_{IJKL} and the C_n coefficients are determined by a torsional barrier parameter, the periodicity of the torsion and the equilibrium torsion angle.

The inversion contributions are described by a one or two term cosine Fourier expansion:

$$E_{invers}(\omega) = K_{IJKL}(C_0 + C_1\cos(\omega_{IJKL}) + C_2\cos(2\omega_{IJKL})) \tag{12}$$

where K_{IJKL} is the force constant for inversion and ω_{IJKL} is the angle between the IL axis and the IJK plane.

The last two terms in Eq. (5) represent non-bonded interactions. A Coulomb potential is used to describe the electrostatic interactions:

$$E_{elec} = Q_i Q_j / \varepsilon R_{ij} \qquad (13)$$

In this term ε is the dielectric constant (set to 1), Q_i and Q_j are the partial charges on the atom centers and R_{ij} is the distance between an MM atom center and an EFP atom center or bond midpoint. A Lennard Jones 6-12 expression is employed for the Van der Waals (vdw) interactions:

$$E_{vdw} = D_{IJ}\left\{-2\left[\frac{x_{IJ}}{x}\right]^6 + \left[\frac{x_{IJ}}{x}\right]^{12}\right\} \qquad (14)$$

In Eq. (14), x is the distance between MM atom centers and EFP atom centers, x_{ij} is the Van der Waals bond length parameter and D_{ij} is the well depth parameter. D_{ij} is obtained from geometric combination rules of atomic Van der Waals energies, D_i:

$$D_{ij} = (D_i * D_j)^{1/2} \qquad (15)$$

x_{ij} is obtained from sums of Van der Waals radii, x_i:

$$x_{ij} = (x_i * x_j)^{1/2} \qquad (16)$$

In the UFF the non-bonded terms are excluded for 1,3 and 1,4 interactions.

II. Interaction terms

Electrostatics

In the combined EFP-MM method, the MM partial charges interact with partial charges on the EFP expansion points. This interaction term is modeled with a Coulomb potential as in the UFF:

$$E_{elec} = Q_i Q_j / \varepsilon R_{ij} \qquad (17)$$

In this term ε is the dielectric constant (set to 1), Q_i is the partial charge on the MM atom centers and Q_j is the partial charge on the atom centers or bond midpoints of the EFP fragment, and R_{ij} is the distance between an MM atom center and an EFP atom center or bond midpoint.

Partial charges on the EFP atom centers and bond midpoints are obtained from the DMA[ref]. To maintain consistency with EFPs, partial charges on MM atom centers are also obtained from the DMA. No distance cutoff is used between the MM atoms and EFPs.

The UFF bond stretching, angle bending, torsion and inversion parameters were determined without partial charges. In the original UFF implementation, partial charges were obtained using the Charge Equilibration (*QEq*) method proposed by Rappé and Goddard.[26] When QEq partial charges are included in the force field, the relative energies predicted by UFF are not in good agreement with the experimentally determined energies.[27] The *QEq* method was not implemented in GAMESS; instead, it is left up to the user to decide whether and how to obtain partial charges. Two excellent options for obtaining charges are from electrostatic fitting[28] or the DMA.[29,30]

Dispersion

Van der Waals interactions between the EFP atoms and the MM atoms can be treated by following the approach used in the UFF force field; that is, by using a Lennard-Jones 6-12 potential:

$$E_{vdw} = D_{IJ}\left\{-2\left[\frac{x_{IJ}}{x}\right]^{6} + \left[\frac{x_{IJ}}{x}\right]^{12}\right\} \qquad (18)$$

In this term, x is the distance between MM atom centers and EFP atom centers, x_{ij} is the Van der Waals bond length parameter and D_{ij} is the well depth parameter. D_{ij} and x_{ij} are obtained in the same was as in UFF (see equations 15 and 16).

For all atoms except for those involved in hydrogen bonds between the MM and EFP regions, the MM and EFP well depth parameters and vdw radii parameters for the interaction term are taken from UFF. The parameters from UFF were developed to model systems of metal oxides, consequently they perform very poorly for hydrogen bonded systems as will be demonstrated for water dimer in the EFP-MM test calculation section. Hydrogen bonding parameters for the EFP-MM interaction term were obtained from the DREIDING force field, which has same functional form as UFF for nonbonded interactions. These parameters were developed specifically for atoms involved in hydrogen bonds.

Other Terms

The two terms discussed above take into account electrostatic and dispersion interactions between the EFP solvent molecules and the MM region. Other terms to consider

include polarization, charge-transfer and exchange-repulsion. The most serious problem with the intermolecular interactions in typical force fields, such as UFF, is their fixed charge formalism. This formalism makes it difficult to treat the Dipole-Induced-Dipole (or Polarization) effects of polar solvents. Polarization effects are important for an accurate description of liquids,[31] and polarization can contribute as much as 20% of the interaction energy of hydrogen bonding interactions.[12] There are two main possibilities for how to model polarization of the EFP atoms by the MM region. In the first possibility, atomic partial charges are allowed to change as the geometry changes throughout the simulation, and in the second possibility, multipoles are included and allowed to polarize each other. The first option is employed in the QEq method[26]. The second option has been used successfully in EFP and classical force fields,[12,32] and this approach will be implemented in future versions of the method discussed here.

The exchange repulsion is a purely quantum mechanical interaction that arises from the overlap of wave functions on different molecules. The short-range repulsion is taken into account to some extent by the R^{-12} term of the Lennard-Jones potential. Although there is no fundamental theoretical justification for this term, it does describe repulsion at short range.

III. Energy Gradients

The combined total EFP-MM interaction energy gradient was derived with respect to the coordinates of the EFP and MM regions:

$$\frac{dE_{Total}}{d\vec{R}_{solvent}} = \frac{\partial E_{EFP}}{d\vec{R}_{solvent}} + \frac{\partial E_{I}}{d\vec{R}_{solvent}} \qquad (19)$$

$$\frac{dE_{Total}}{d\vec{R}_{bulk}} = \frac{\partial E_{MM}}{d\vec{R}_{bulk}} + \frac{\partial E_{I}}{d\vec{R}_{bulk}} \qquad (20)$$

where E_{Total} is the total energy of the combined MM and EFP system, E_{EFP} is the energy of the EFP region, E_{MM} is the energy of the MM region and E_I is the interaction energy between the EFP region and the MM region. $R_{solvent}$ and R_{bulk} refer to the atomic coordinates of the EFP atoms and the MM atoms respectively. The internal geometries of the EFPs are fixed, so in a geometry optimization, the EFPs move according to a net force on the center of mass COM) of each fragment and a net torque around the center of mass of each fragment.[12] The net force on each fragment is obtained by summing the forces on each expansion point. The torque on a fragment is the cross product of the position vector from the point of rotation to the COM of the fragment and the vector of the net force acting on the fragment. Optimizations described in the next section were performed using the search method in GAMESS.

IV. EFP-MM Test Calculations

Water dimer

The ability of the EFP-MM method to treat the water dimer was investigated by comparison with full MP2,[33] HF, and EFP structures. The basis set used for the *ab intio* waters was the Dunning Hay basis set with d and p polarization functions.[34] All EFP waters are modeled with the HF based EFP1 method. In EFP1, the geometry of each EFP fragment is chosen to have an OH bond length of 0.9572 Å and an HOH bond angle of 104.52˚. Each EFP has five expansion points located on the atom centers and bond midpoints.

The EFP-MM method was first tested on the water dimer to gauge the ability of the method to treat hydrogen bonding. As a baseline for comparison, full MP2/DH(d,p) and HF/DH(d,p) optimizations were performed on the water dimer. Their geometries are shown in Figure 2. The MP2 structure in Figure 2(a) has a H-bond length of 1.945 Å and an O-H···O bond angle of 174.2˚. The HF geometry in Figure 2(b) has an H-bond length of 2.040 Å and an O-H···O bond angle of 177.8˚. Figure 3 shows the structure of an all-EFP water dimer. The H-bond length for this method is 2.047 Å, and its bond angle is 176.6˚, in good agreement with the HF results.

When a full MM optimization of water dimer was performed using vdw parameters from UFF, the shape of the water dimer is badly distorted (Figure 4(a)). The H-bond length is 2.503 Å, which is elongated by 0.558 Å over the MP2 value. The H-bond angle is also distorted by 62.3˚. The vdw parameters for O and H in UFF were parameterized for metal oxide systems, so it is not surprising that they perform poorly for hydrogen bonding systems. The precursor to UFF was the DREIDING force field (DFF),[35] and the functional form for its non-bonded terms is the same as that in the UFF. The DFF has special vdw parameters for atoms involved in hydrogen bonds, and when these are used, the UFF is able to more accurately represent the geometry of the water dimer, as shown in Figure 4b and 4c. While the H-bond angle of 175.4˚ in Figure 4b is reasonable, the hydrogen bond length is still not acceptable, as it is 0.294 Å shorter than the MP2 H-bond length. This occurs because the vdw parameters in the Dreiding FF were parameterized with Gasteiger charges[36] and the structure in Figure 4b used the more accurate, but more expensive charges from *ab intio* calculations. When a more consistent set of parameters is employed, using the Gasteiger

charges, the structure is shown in Figure 4c is obtained, with a reasonable H-bond length of 1.795 Å.

Since the two water molecules in the water dimer are not equivalent, the mixed EFP-MM method must be tested for two cases: In the first case the water acting as the H-bond donor is replaced by an EFP water, and in the second case the water acting as the H-bond acceptor is replaced by an EFP water. Charges on the MM region were obtained from DMA. The resulting geometries in these two cases are shown in Figure 5. In Figure 5a the EFP water is the H-bond donor and the MM water is the H-bond acceptor and vice-versa for Figure 5(b). In both of these cases the structure of the water dimer is qualitatively reproduced. In Figure 5(a) the H-bond length is 0.062 Å shorter and the bond angle is only 3.6 degrees larger than the MP2 values. In Figure 5(b) the H-bond length is 0.096 Å shorter and the angle is 1.6 degrees larger than the MP2 values. Not surprisingly, the EFP-MM H-bond length is the difference between the H-bond lengths in the full MM structure (Figure 4(b)) and the full EFP structure (Figure 3).

SiH_3OH and $Si(OH)_4$

The EFP-MM method was next tested on hydrogen bonded complexes between water and SiH_3OH and $Si(OH)_4$. The silica clusters were treated with MM, and the waters were treated with the HF based EFP1 method. Charges on the MM atoms were obtained from the DMA. Because UFF was not parameterized with charges included, the electrostatic interactions between MM partial charges were not included.

The MM-EFP geometries were compared to full ab initio MP2 calculations with the 6-31G(d) basis set.[37-41] The geometries are also compared to *ab inito* HF calculations on

SiH_3OH and $Si(OH)_4$ with one EFP1/HF water molecule. The basis set used for the HF calculations was also 6-31G(d). Density functional theory (DFT) calculations by Thompson and Margey (TM),[42] which used Becke's three parameter Lee-Yang-Parr hybrid functional (B3LYP)[43-45] and the 6-311+G(d) basis set,[46-48] were used as the starting structure for all calculations.

The MP2, HF and MM geometries of SiH_3OH and $Si(OH)_4$ are shown in Figures 6 and 7, respectively. Both HF and UFF are able to reasonably reproduce the MP2 bond lengths and angles in these molecules. The geometries of SiH_3OH plus one water molecule are shown in Figures 8 and 9. When the water molecule acts as the hydrogen bond donor, the MM-EFP hydrogen bond length for the MM-EFP method is 1.913 Å (Figure 8c), in good agreement with the MP2 values. The O-H···O bond angle is 177.1˚, 23.3 degrees larger than the MP2 value (Figure 8a), but only 2.8˚ larger than the HF value with an EFP water (Figure 8b). The latter is a more appropriate comparison, since the EFP1/HF method is derived from HF.

When the silanol acts as the H-bond donor (Figure 9) the H-bond for the EFP-MM length is 0.062 Å longer than the H-bond length obtained with the HF silanol - EFP water combination (Figure 9b). The MM-EFP H-bond angle (Figure 9c) is somewhat smaller than those predicted by HF-EFP (Figure 8b) and MP2 (Figure 8a).

The structures of $Si(OH)_4$ plus one water are shown in Figure 10. In this case the the $Si(OH)_4$ molecule acts as both a hydrogen bond donor and a hydrogen bond acceptor forming a pseudo 6-center ring with the water molecule. Neither H-bond is close to being linear. The MM-EFP method gives H-bond distances that are somewhat elongated compared with the

MP2 structures. The angles are in reasonable agreement. Although the agreement between the EFP-MM and the other methods is not quantitative, it is reasonable.

Interaction energies

The interaction energies for the two types of MM-EFP water dimers are compared in Table 1 with interaction energies from full HF[12] and MP2 calculations, as well as calculations with one HF water and one EFP water[12]. The EFP-MM interaction energies for water dimer are both within 0.5 kcal/mol of the HF baseline interaction energies.

Table 2 gives the interaction energies for $Si(OH)_4$ and SiH_3OH with an EFP water molecule at the equilibrium geometries. For SiH_3OH, with the water molecule acting as the H-bond acceptor, the interaction energy is underestimated by 0.7 kcal/mol when compared with the HF-EFP value. For SiH_3OH, with the water molecule acting as the H-bond donor, the interaction energy is overestimated by 1.9 kcal/mol compared with the HF-EFP result. Not surprisingly, MP2 predicts mush stronger binding than do the HF-based methods. The EFP-MM method does reproduce the MP2 relative energies of the two SiH_3OH isomers, whereas the HF-EFP method does not.

The last row of Table 2 gives the interaction energies of $Si(OH)_4$ with an EFP water. The EFP-MM and HF-EFP interaction energies are in good agreement with each other (within 1.0 kcal/mol), while MP2 again predicts much stronger binding. It is worth noting that MP2 with a modest basis set will suffer from significant basis set superposition error (BSSE), leading to over-binding, whereas neither MM nor EFP are subject to BSSE problems.

V. Conclusions

A method for modeling the interaction between EFP solvent molecules and atoms described by a molecular mechanics force field has been defined and implemented into the GAMESS electronic structure program. The interaction terms are similar to those in the universal force field, with partial charges from Stone's DMA. This method was able to reproduce the geometry of water dimer as well as the geometry of hydrogen bonded systems of SiH_3OH and $Si(OH)_4$ with an EFP water. Interaction energies from the EFP-MM method are within 2 kcal/mol of the interaction energies predicted by the HF-EFP method. The MP2 interaction energies are not well reproduced, but this is due in part from MP2 over-binding due to expected BSSE error. In order to improve the accuracy of the EMP-MM interaction energies, the functional form of the interaction term needs to be improved. The first step to doing this is to add polarization to the system to account for the dipole-induced-dipole effects of polar solvent molecules.

Acknowledgements

This work has been supported by a grant from the U.S. Department of Energy, administered by the Ames Laboratory. The authors also acknowledge Drs. Mike Schmidt and Lyuda Slipchenko for many helpful discussions.

References

(1) Huh, S., Chen, H-T., Wiench, J.W, Pruski, M., and Lin, V.S.-Y. *J. Am. Chem. Soc.,* **2004** 126, 1010.

(2) Mbaraka, I.K.; Radu, D.R.; Lin, V.S.-Y.; Shanks, B.H. *J. Catalysis*, **2003**, *219*, 329.

(3) Lin, VS.-Y.; Radu, D.R.; Han, Mi-K.; Deng, W.; Kuroki, S.; Shanks, B.H.; Pruski, M. *J. Am. Chem. Soc*. **2002**, *124*, 9040.

(4) Chen, H-T.; Huh, S.; Wiench, J.W; Pruski, M.; Lin, VS.-Y.; *J. Am. Chem. Soc*., **2005**, *127* 13305.

(5) Lin, VS.-Y.; Nieweg, J.A.; Kern, C.; Trewyn, B.G.; Wiench, J.W; Pruski, M.; *Prepr. Symp. Am. Chem. Soc., Div. Fuel Chem*., **2006**, *51*, 426.

(6) Y. Cai, R. Kumar, W. Huang, B.G. Trewyn, J.W. Wiench, M. Pruski, and V. S.-Y. Lin, *J. Phys. Chem. C*, **2007**, 111, 1480-6.

(7) Zorn, D.; Lin, V.S.-Y.; Pruski, M.; Gordon, M. in preparation.

(8) Schmidt, M.W., et al., *J. of Comput. Chem.* **1993**, *14*, 1347.

(9) Schmidt, M. W.; Gordon, M. S. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, **2005**.

(10) Shoemaker, J.R.; Burggraf, L.W.; Gordon, M.S. *J. Phys. Chem. A*, **1999**, *103,* 3245.

(11) Gordon, M. S.; Freitag, M.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.

(12) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Kraus, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

(13) Freitag, M.A.; Gordon, M.S.; Jensen, J.H.; Stevens W.A. J. Chem. Phys. **2000**, *112*,

7300.

(14) Slipchenko, L.V.; Gordon, M.S. *J. Comput. Chem.* **2007**, *28*, 276.

(15) Nemukhin, A.V.; Grigorenko, B.L.; Bochenkova, A.V.; Topol, I.A.; Burt, S.K. *Journal of Molecular Strucutre*: THEOCHEM, **2002**, *581*, 167.

(16) Kundrot, C. E.; Ponder, J.W.; Richards, F.M. *J. Comput. Chem.* **1991**, *12*, 402.

(17) Ponder, J.W.; Richards, F.M. *J. Comput. Chem.* **1987**, *8*, 1016.

(18) Kaminski, G. A.; Friesner, R. A. Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105* 6474-6487.

(19) Aikens, C.M.; Gordon, M.S. *J. Am. Chem. Soc.*, **2006**, *128*, 12835.

(20) Mullin, J.M.; Gordon, M.G. In preparation.

(21) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III; W.A.; Skiff; W.M. *J. Am. Chem. Soc.* **1992**, *114* 10024.

(22) C. J. Casewit, K. S. Colwell, and A. K. Rappe'. *J. Am. Chem. Soc.*, **1992,** *114*, 10035.

(23) Rappe, A. K.; Casewit, C. J.; Casewit. C. J. *Inorg. Chem.* **1993,** *32*, 3438.

(24) Casewit, C. J.; Colwell, K. S.; Rappe, A. K. *J. Am. Chem. Soc.* **1992,** *114*, 10046.

(25) (a) Badger, R. M. *J. Chem. Phys.* **1934**, *2*, 2128, (b) Badger, R. M. *J. Chem. Phys.* **1935**, *3*, 710, (c) Pearson, R. G. *J. Am. Chem. Soc.* **1977**, *99*, 4869, (d) Ohwada, K. *J. Chem. Phys.* **1980**, *72*, 1, (e) Ohwada, K. *J. Chem. Phys.***1980**, *72*, 3663, (f) Ohwada, K. *J. Chem. Phys.***1980**, *72*, 3663, (g) Ohwada, K. *J. Chem. Phys.* **1981**, *75*, 1309, (h) Chang, C.-A. *J. Phys. Chem.* **1983**, *87*, 1694, (i) Barbiric, D. A.; Castro, E. A.; Fernandez, F. M. *J. Chem. Phys.* **1984**, *80*, 289 (j) Ohwada, K. *J. Chem. Phys.* **1984**, *80*, 1556, (k) Halgren, T. A. *J. Am. Chem. Soc.* **1990**, *112*, 4710.

(26) Rappé. A.K.; Goddard III, W.A. *J. Phys. Chem.* **1991**, *95*, 3358.

(27) Gundertofte, K.; Liljefors, T.; Norrby, P-O.; Pettersson, I. *Journal of Computational Chemistry*, **1996**, *17*, 4, 429.

(28) Spackman, M.A. *J. Comp. Chem.,* **1996**, *17*, 1.

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

(30) Stone, A. *J Chem Phys Lett* **1981**, *83*, 233.

(31) Iuchi, S.; Izvekov, S.; Voth, G.A. *J. Chem. Phys.* **2007**, *126*, 124505.

(32) Rasmussen, T.D.; Ren, P.; Ponder, J.W.; Jensen, F. *International Journal of Quantum Chemistry* **2007**, *107*, 1390.

(33) Møller, C.; Plesset, M. S. *Phys. Re*V. **1934**, *46*, 618.

(34) Dunning, Jr., T. H. ; Hay, P. J. in *Methods of Electronic Structure Theory*, edited by H. F. Shaefer III (Plenum, NY 1977), pp. 1–27.

(35) Mayo, S.L.; Olafson, B.D.; Goddard III, W.A. *J. Phys. Chem.* **1990,** *94*, 8897-8909

(36) Gasteiger, J.; Marsili, M. Tetrahedron **1980**, *36*, 3219.

(37) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(38) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.

(39) Hariharan, P.C.; Pople, J.A Theoret. Chimica Acta **1973**, *28*, 213.

(40) Clark T.; Chandrasekhar J.; Spitznagel G. W.; Schleyer PvR *J. Comput. Chem.* **1983**, *4*, 294.

(41) Frisch, M.J.; Pople, J.A.; Binkley, J.S. *J. Chem. Phys.* **1984**, *80*, 3265.

(42) Thompson, K.C.; Margey, P. *Phys. Chem. Chem. Phys.*, **2003**, *5*, 2970

(43) Becke, A.D. *J. Chem. Phys.* **1993**, 98**,** 5648.

(44) Lee C.; Yang, W; Parr, R.G. *Phys. Rev. B* **1998**, 37, 785.

(45) Vosko, S.H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58* 1200.

(46) R. Krishnan, J.S. Binkley, R. Seeger and J.A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

(47) McLean, A.D.; Chandler, G.S. *J. Chem. Phys.* **1980,** *72*, 5639.

(48) Clark, T.; Chandrasekhar, J.; Schleyer, P.V.R. *J. Comp. Chem.* **1983,** *4*, 294.

Table 1. Interaction energies in kcal/mol for water dimer.

	MP2/DH(d,p)	HF/DH(d,p)[a]	HF/DH(d,p) with 1 EFP[a]	EFP-MM
Ab inito	-4.1	-5.0		
Water dimer - A			-4.7	-5.3
Water dimer - D			-4.5	-5.5

[a]From Day et al.

Table 2. Interaction energies in kcal/mol for hydrogen bonded complexes.

	MP2/6-31G(d)	HF/6-31G(d) w/ 1 EFP	EFP-MM
SiH_3OH - A	-4.9	-4.3	-3.6
SiH_3OH - D	-7.3	-2.7	-4.6
$Si(OH)_4$	-10.4	-5.2	-6.2

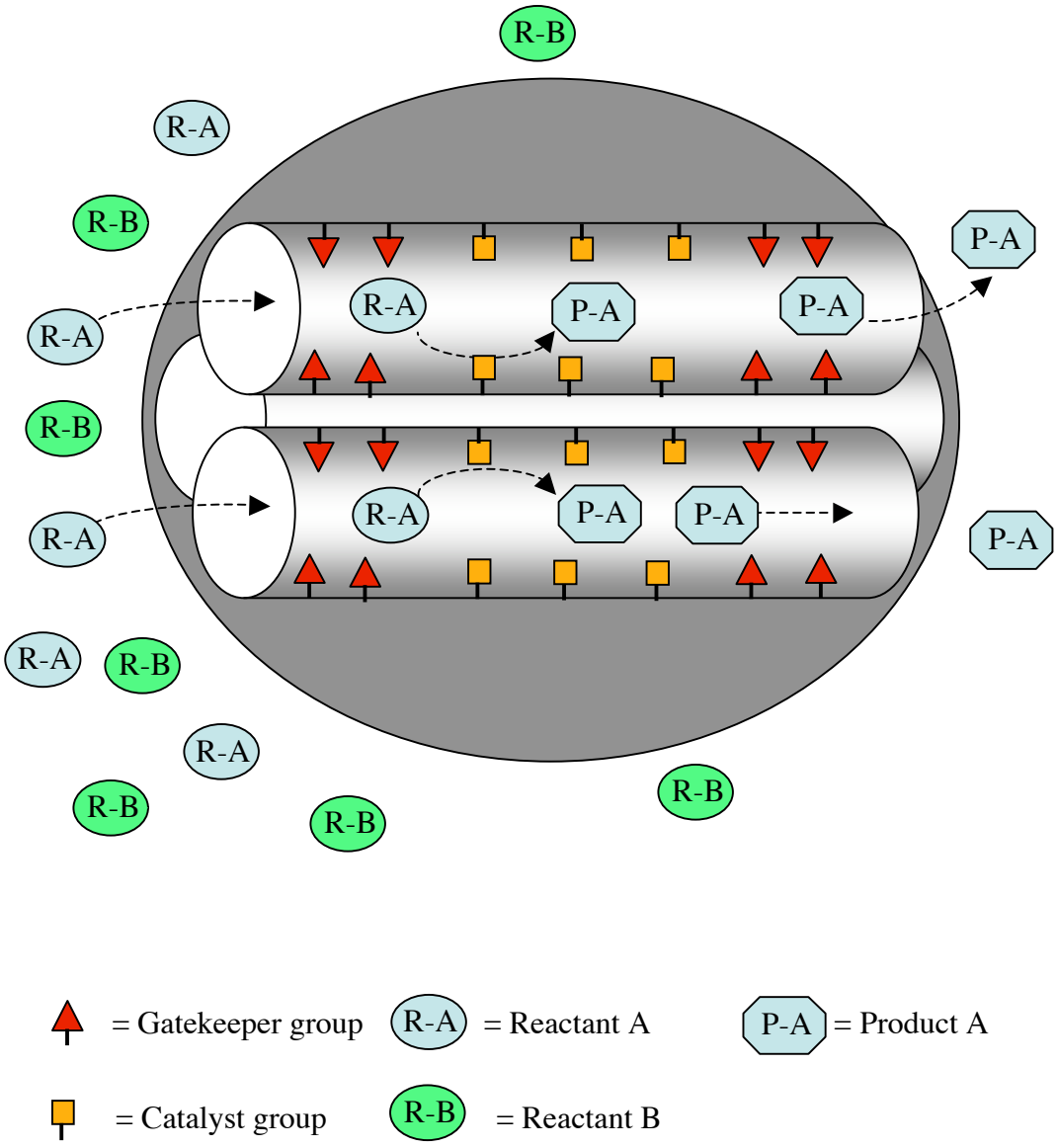


Figure 1: Schematic of a multi-functionalized mesoporous silica system.

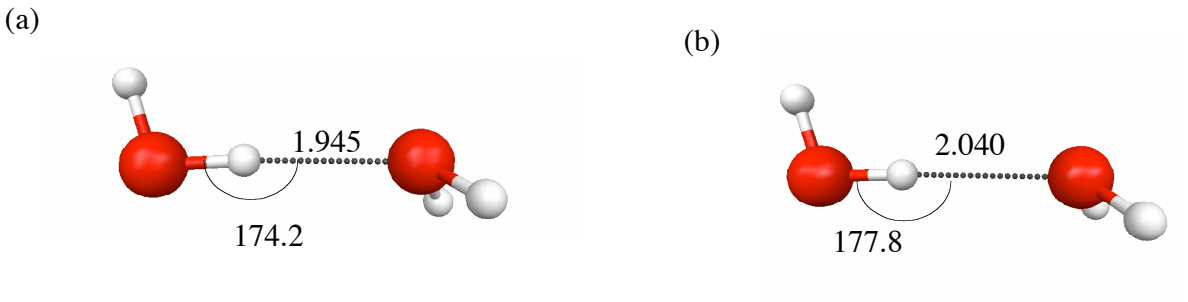


Figure 2. *Ab initio* structures of water dimer: (a) MP2/DH(d,p), (b) HF/6-DH(d,p). Hydrogen-bond lengths given in Å.

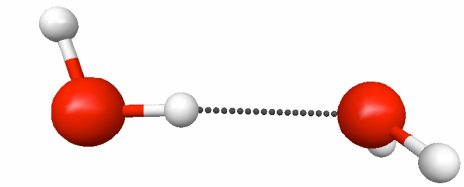


Figure 3. Full EFP structure of water dimer. Hydrogen-bond lengths given in Å.

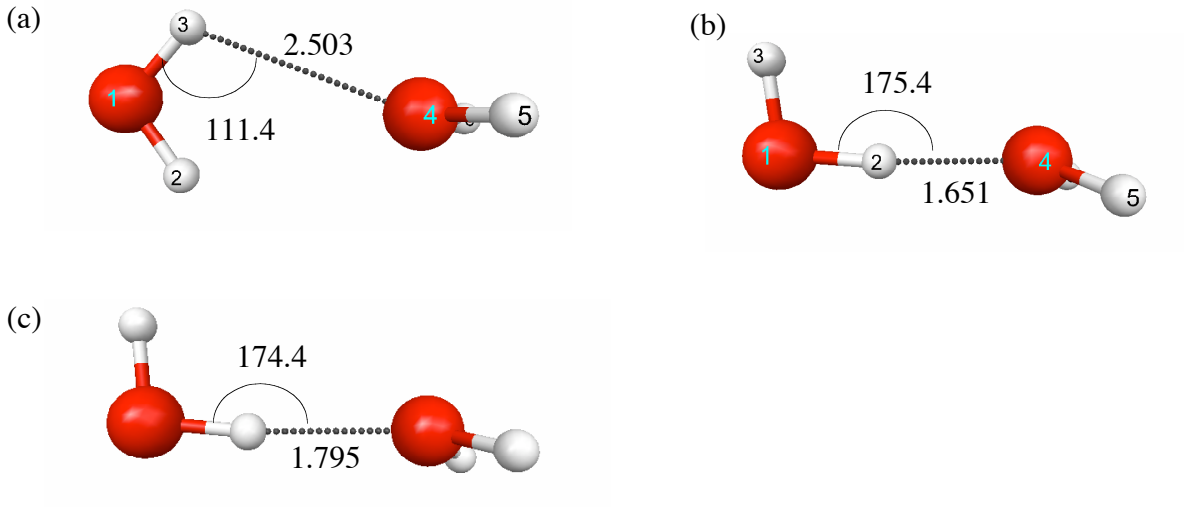


Figure 4. MM structures of water dimer: (a) With parameters from UFF (b) with parameters from DREIDING and charges from DMA (c) with parameters from DREIDING and Gasteiger charges. Hydrogen-bond lengths given in Å.

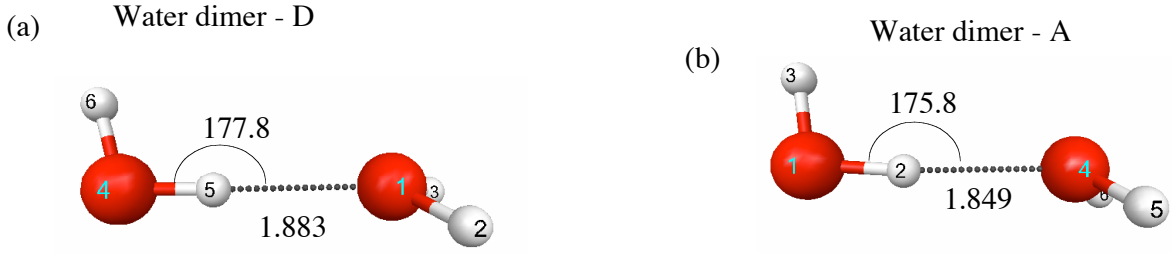
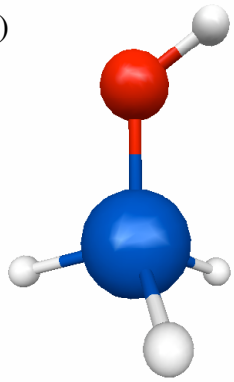


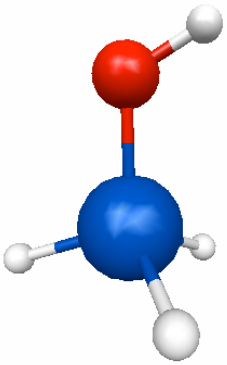
Figure 5. EFP/MM structures of water dimer: (a) EFP water is H-bond donor and MM water is H-bond acceptor (Water dimer - D) and (b) MM water is H-bond donor and EFP water is H-bond acceptor (Water dimer - A). Hydrogen-bond lengths given in Å.

(a)



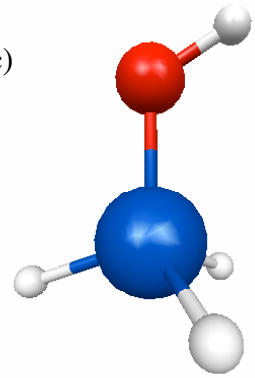
O-H: 0.969
Si-O: 1.678
Si-H: 1.483
H-O-Si: 115.5
O-Si-H: 105.7

(b)



O-H: 0.946
Si-O: 1.663
Si-H: 1.483
H-O-Si: 118.2
O-Si-H: 106.8

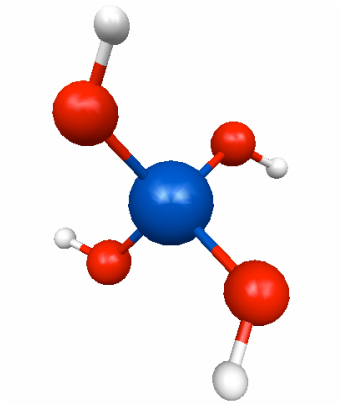
(c)



O-H: 0.961
Si-O: 1.679
Si-H: 1.463
H-O-Si: 115.4
O-Si-H: 106.5

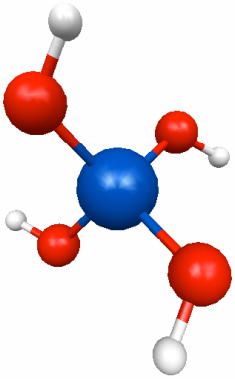
Figure 6. Structures of Si_3OH with (a) MP2/6-31(d), (b) HF/6-31(d), and (c) UFF. Bond lengths and angles given below each structure. Bond lengths in Å.

(a)



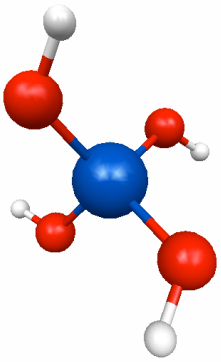
O-H: 0.970
Si-O: 1.658
H-O-Si: 113.4
O-Si-O: 105.8

(b)



O-H: 0.947
Si-O: 1.633
H-O-Si: 116.5
O-Si-O: 106.4

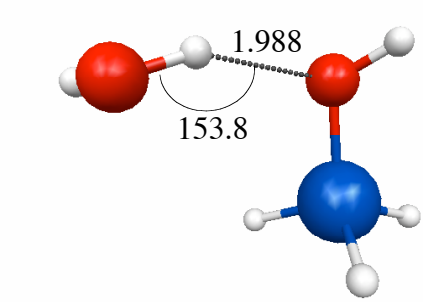
(c)



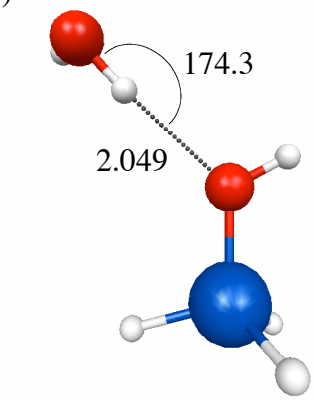
O-H: 0.958
Si-O: 1.618
H-O-Si:107.1
O-Si-O: 109.1

Figure 7. Structures of $Si(OH)_4$ (a) MP2/6-31(d), (b) HF/6-31(d) and (c) UFF. Bond lengths and angles given adjacent each structure. Bond lengths in Å.

(a)



(b)



(c)

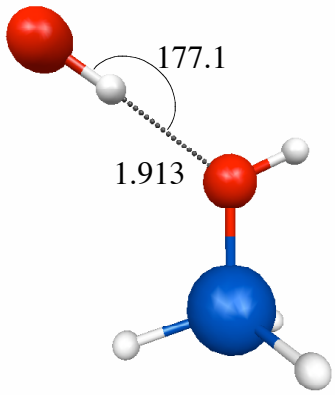


Figure 8. Structures of SiH_3OH with 1 water, where the water acts as the H-bond donor (SiH_3OH - D): (a) MP2/6-31G(d), (b) HF for $Si(OH)_4$ and EFP water, and (c) UFF for $Si(OH)_4$ and EFP water. Hydrogen-bond lengths in Å.

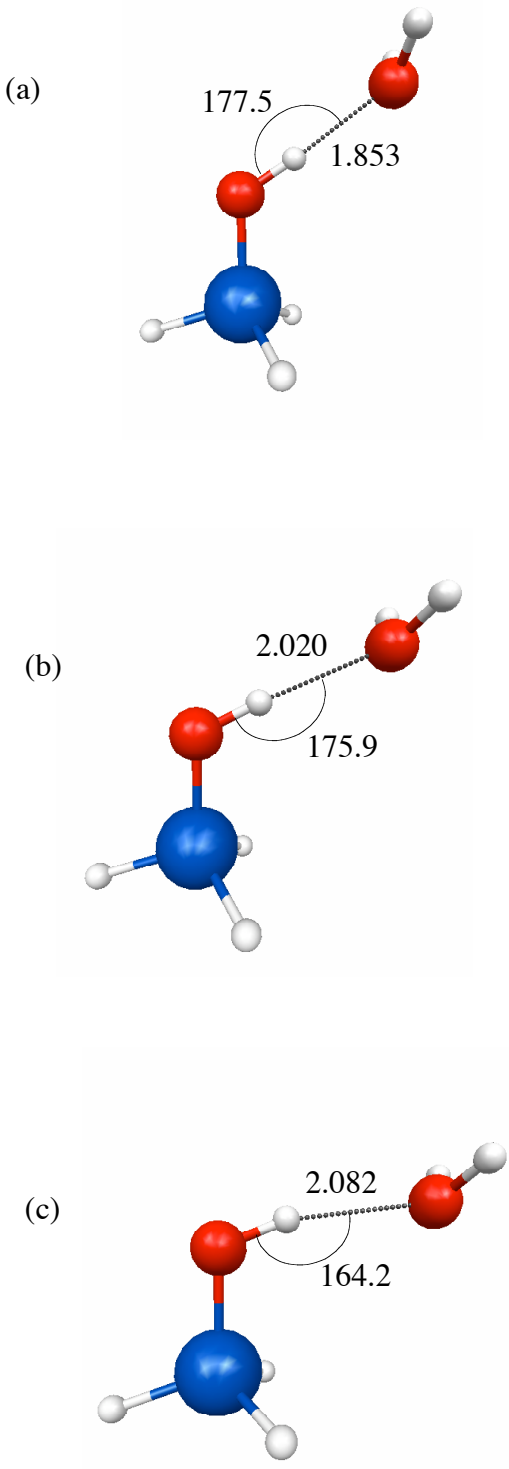
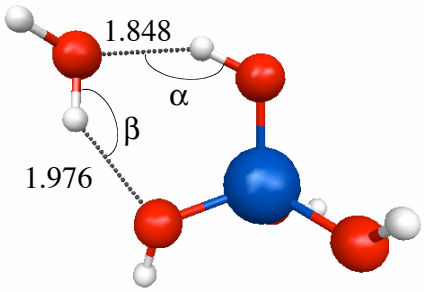


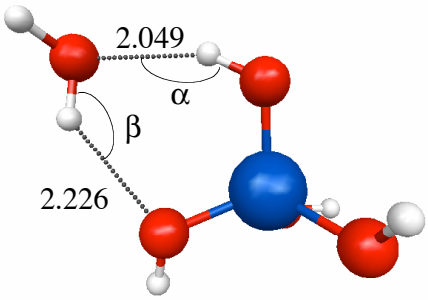
Figure 9. Structures of SiH_3OH with 1 water, where the water acts as the H-bond acceptor (SiH_3OH - A): (a) MP2/6-31G(d), (b) HF for $Si(OH)_4$ and EFP water, and (c) UFF for $Si(OH)_4$ and EFP water. Hydrogen-bond lengths in Å.

(a)



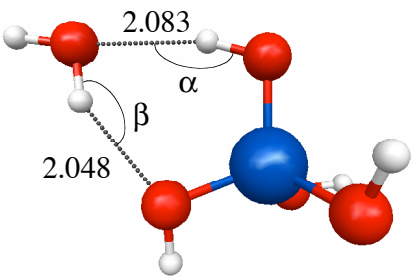
α = 154.6
β = 145.1

(b)



α = 149.7
β = 136.5

(c)



α = 159.7
β = 148.7

Figure 10. Structures of $Si(OH)_4$ with 1 water: (a) MP2/6-31G(d), (b) HF for $Si(OH)_4$ and EFP water, and (c) UFF for $Si(OH)_4$ and EFP water. Hydrogen-bond angles given below each structure. Hydrogen-bond lengths in Å