MULTISCALE SIMULATIONS: FROM QUANTUM MECHANICS TO COARSE GRAINED MOLECULAR DYNAMICS

by

FRANK CLEMENCE PICKARD IV

(Under the Direction of Henry F. Schaefer III)

ABSTRACT

In chapter 2 the electronic ground state ($\tilde{X}\,^1A_1$) of formaldehyde and its barrier to molecular elimination were rigorously studied using high order coupled cluster techniques [up through fifth order CCSDTQ(P)]. The barrier to molecular products was determined to be 80.82 kcal mol^{-1}, a result which stands between the two most widely accepted experimental values, and disagrees with the most recent *ab initio* study by ~1 kcal mol^{-1}. In chapter 3 we characterize hydroxymethylene, an elusive high energy conformer of formaldehyde. Calculated vibrational and UV/Vis spectra is in strong agreement with experiment, suggesting this short lived molecule has finally been synthesized. High level theoretical calculations suggest this species decomposes to formaldehyde by tunneling through a barrier of ~30 kcal mol^{-1}. In chapter 4 we investigate the electronic ground states ($\tilde{X}\,^1\Sigma^+$) of HSiN, HNSi, and the transition state connecting the two isomers. We also utilize vibrational perturbation theory to model the effects of anharmonicity upon computed spectroscopic properties. In chapter 5 we discuss the implementation of a Gō-like coarse grained protein model and its application to two fast folding proteins.

INDEX WORDS: Combustion Chemistry, Interstellar Chemistry, Quantum Tunneling, Kinetic Isotope Effect, Thermochemistry, Molecular Dynamics, Coarse Grain Modeling, Multi-scale Modeling, Protein Folding

MULTISCALE SIMULATIONS: FROM QUANTUM MECHANICS TO COARSE GRAINED MOLECULAR DYNAMICS

by

FRANK CLEMENCE PICKARD IV

A.B., Hamilton College, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSPHY

ATHENS, GEORGIA

2011

© 2011

Frank Clemence Pickard IV

All Rights Reserved

MULTISCALE SIMULATIONS: FROM QUANTUM MECHANICS TO COARSE GRAINED MOLECULAR DYNAMICS

by

FRANK CLEMENCE PICKARD IV

Major Professor:	Henry F. Schaefer III
Committee:	Nigel G. Adams
	Gary E. Douberly

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2011

In loving memory of Deke and Pick.

ACKNOWLEDGEMENTS

Faithfully enumerating all of the people in my life who helped me successfully complete my education would be a futile endeavor. That said, there are a special few individuals, whose contributions were so critical towards me getting to this point in my life that I would be derelict in not recognizing them by name.

First and foremost, I would like to thank my parents for their support. Both of them, in their own unique and special ways, have provided love and encouragement for my efforts. Throughout my life, their actions have always paved the way for my successes (the failures, I alone take credit for!) The road was not always an easy one, but in the end, I think the journey was worth it. Both of my grandfathers also deserve special mention. From a very young age, they both fostered my interest in science, computers and learning in general. Needless to say, both of them laid the groundwork for my eventual decision to pursue a degree in computational chemistry.

Next I would like to thank my teachers and mentors throughout the years. From elementary school through graduate school, I have been blessed with the most passionate and talented educators. In particular I would like to thank the following people: Coy Johnston for feeding my curiosity for mathematics at such a young age, Dick Cobb for teaching me how handle failure and learn from it (Latin is my kryptonite), Steven Schieffelin for teaching me what coherent prose should look like, George Shields and Karl Kirschner for introducing me to computational chemistry and encouraging me to pursue a doctorate, Wesley Allen for his patience and teaching me the differences between *theoretical* and *computational* chemistry,

Bernie Brooks for taking a chance on me and paving the way for my transition into the exciting field of molecular modeling; and finally, Fritz Schaefer for making all things possible.

I would also like to thank my colleagues past and present. To my friends at the CCQC, thank you for all of your support over the years. In particular, I would like to thank Linda, for helping me deal with all of the various emergencies over the years, and overall just being an amazing person; Jeremy, for his indefatigable help with Professor Allen's course and his acerbic wit; and Andy, for taking me under his wing, and mentoring me in the ways of quartic force fields and anharmonic corrections. At the NIH I have been similarly blessed by being a part of a group of amazingly talented and enjoyable individuals. In particular I would like to thank Lee, for easing my transition from the world of QM to the world of MM; Rick for knowing *everything* about CHARMM and being patient enough to explain it to me; and Tim, for persevering with me through the coarse grained modeling project, we are almost there!

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION & LITERATURE REVIEW

1.1 Overview

Theoretical chemistry is, in its simplest form, the act of describing chemical systems using mathematical tools and physical insights. Different kinds of chemical systems require different types of descriptions, or models, to accurately describe their properties. A model appropriate for predicting an ensemble of protein conformations would likely be inappropriate for predicting the thermodynamics of various carbene species. This example illustrates the most obvious way of differentiating between models: length and time scales. At the shortest length and time scales, electrons and nuclei are treated separately, and often with a quantum mechanical (QM) description of varying sophistication. Pulling back to longer length and time scales involves ignoring quantum effects entirely in favor of classical descriptions and experimental knowledge, so called molecular mechanics (MM) models. Atoms serve as the primary building blocks of these systems, and bonding information is explicitly specified. The longest length and time scales require even further simplification of the model systems. Using *a priori* biochemical knowledge, groups of atoms or molecules may be lumped together into coarse grained (CG) particles.

Other models may mix the aforementioned techniques in an arbitrary fashion, to strike a finer balance between accuracy and efficiency. The most common form of multi-scale modeling is QM/MM, an approach which spatially partitions the molecule into various subsystems each of which are treated using a different method. The chemically active region of a molecule is treated

quantum mechanically to allow bond formation, and the chemically inactive region, is treated classically. Various wave-function based QM methods may also be mixed in an effort to produce highly accurate molecular property calculations. Basis set extrapolation techniques and highly correlated QM methods may be combined in a focal point to achieve results near the one-particle and *n*-particle limits.

By contrast, *computational chemistry* is the act of solving these theoretical models, and deriving physical and chemical insights from their solutions. As the name suggests, computers are required to solve the underlying equations when these models are applied to all but the most trivial of cases. Just as different types of chemical systems may require different theoretical descriptions, a single theory may be executed to various levels of accuracy as prescribed by the constraints of experimental accuracy and computational horsepower. Approximations underlie both the theory and computation of all chemical phenomena. Understanding their origins and limitations is critical towards choosing the appropriate model for a given system.

In this work we will discuss and explore the theoretical methods and computational techniques for modeling systems at the smallest and largest length scales. Different problems underlie attempts to model these physical, chemical and biological phenomena, and different approaches are required to make accurate experimentally verifiable predictions about them. Quantum mechanics (QM) will be brought to bear on the problem of determining highly accurate thermochemistry and spectroscopic properties of small molecule combustion species. At the other extreme, coarse grained (CG) molecular mechanics (MM) will be employed to study ensembles of fast folding proteins, and to determine thermochemical properties and qualitative properties of fast folding proteins.

1.2 Wavefunction Based Methods

The smallest chemical systems can be treated with the most rigorous *ab initio* theories, which are based upon QM formalism, the simplest of which is Hartree-Fock theory (HF).[1, 2] HF theory, and all subsequent QM theories that use HF as their starting point, rely upon the Born-Oppenheimer (BO) approximation. BO tells us that we can assume that the time scales of nuclear and electronic motion are well separated because the mass of a proton is about two-thousand times greater than the mass of an electron.. This, in turn, allows the separation of the nuclear (ψ_{nuc}) and electronic wavefunctions (ψ_{el}), vastly simplifying the wavefunction calculation by breaking it into successive steps (eq. 1.1).

$$\Psi = \psi_{nuc} \times \psi_{el} \tag{1.1}$$

In this regime, the electrons see nuclei that are effectively clamped in space, and the nuclear kinetic energy (T_{nuc}) has been removed from the Hamiltonian, leaving us to solve the electronic Hamiltonian (eq. 1.2, in atomic units). By imposing the Eckart conditions,[3] the nuclear motion can be further decomposed into its translational, rotational and vibrational components, allowing T_{nuc} to be accounted for at a later time.

$$\hat{H}_{elec} = \sum_{A>B}^{N} \frac{Z_A Z_B}{r_{AB}} - \sum_{i}^{n} \left(\frac{\nabla_i^2}{2} + \sum_{i,A}^{n,N} \frac{Z_A}{r_{iA}} \right) + \sum_{i>j}^{n} \frac{1}{r_{ij}} \tag{1.2}$$

The ground state HF wavefunction, $|\Phi\rangle$, is given as a single slater determinant (eq. 1.3). More rigorous QM techniques which account for dynamic electron interactions will be composed of linear combinations of excited state determinants, $|\Phi_i^a\rangle$, where an electron is promoted from orbital *i* to orbital *a*. HF theory employs the mean field approximation, where electrons of opposite spin do not explicitly interact with each other, but rather with the average position of other electrons. Each electron has an orbital energy given by a one-electron Fock operator ($\hat{f}$),

which accounts for kinetic energy, coulombic repulsion (*J*) and exchange (*K*) for each individual electron(eq. 1.4). The HF energy is not simply the sum of the Fock energies, as this would result in a double counting of electron-electron interactions; instead the double sum of electron interactions is replaced with a restricted sum (eq. 1.5).

$$|\Psi_{HF}\rangle = |\Phi\rangle = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(1)] & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix} \tag{1.3}$$

$$\hat{f}_i = \hat{h}_i + \hat{J}_i - \hat{K}_i \tag{1.4}$$

While this model is unphysical, as it only biases electronic motion based upon radial nuclear distance, HF theory still manages to recover ~99% of the energy of the physical system. The remaining 1% of energy, the correlation energy, is omitted. Remarkably, HF theory provides a very robust qualitative description for most chemical systems at a very affordable cost [O(n^4) scaling, where *n* is the number of orbitals]. The Jahn-Teller distortion[4] serves as the most notable chemical phenomenon that cannot be described qualitatively by HF theory.

$$E_{HF} = V_{nuc} + \sum_{i}^{n} h_{ii} + \sum_{i>j}^{n} \langle ij||ij\rangle \tag{1.5}$$

To describe this phenomenon, and to quantitatively describe most chemical species, electron correlation must be included into the physical model. The cheapest method for doing so is Moller-Plesset perturbation theory (MP*n*).[5] First, HF is used for the zeroth order approximation; and next, a perturbation parameter (λ) is introduced to both the Hamiltonian (eq. 1.6) and the wavefunction (eq. 1.7). Expanding the wavefunction as a power-series in terms of λ gives rise to an arbitrary order energy expression.

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{1.6}$$

$$|\Psi\rangle = |\Psi_0\rangle + \lambda |\Psi_1\rangle + \lambda^2 |\Psi_2\rangle + \dots \tag{1.7}$$

At first order, the HF energy is recovered, and at second order (MP2) most correlation effects are included in the energy. While MP2 is a straightforward method, it adequately includes correlation effects and provides very good results for most systems where electronic degeneracy is not present. Most notably, MP2 can account for electron dispersion effects and van der Waals binding, although it does overestimate non-bonded interactions. MP2 is also a relatively inexpensive [O(n^5) scaling] non-iterative method. A quick glance to the MP2 energy expression (eq. 1.8) will show that problems arise when the denominator of the second order energy contribution is small. This occurs in cases with nearly degenerate orbitals in the active space; such cases must be handled by more sophisticated multi-reference theories.

$$E_{MP2} = E_{HF} + \sum_{\substack{i>j \\ a>b}}^{N} \frac{\langle ij||ab\rangle^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{1.8}$$

When extreme accuracy is sought, or when molecular geometries are far from equilibrium, highly correlated methods are required. Like MP2, configuration interaction (CI) theory[6] and coupled-cluster (CC) theory[7] both include correlation effects by expanding the zeroth-order HF wavefunction as a series of excited state Slater determinants. These theories are categorized by the order of excitation included in their wavefunctions. While the method for inclusion of excited state determinants varies between theories, both CI and CC will recover all correlation energy in the limit where all excitations are included in a given basis et space. CC has two main advantages over CI: it generally converges to the basis set limit more rapidly and it is a size extensive method. CI is variational however, while CC is not.

Both methods are prohibitively expensive for large systems [O(n^6) scaling, at second order], however for very small systems they provide a consistent means to obtain reliably

accurate molecular properties without the use of experimental parameterization. The coupled-cluster with singles, doubles and perturbative triples [CCSD(T)] method in particular,[8, 9] has been repeatedly shown to be the 'gold-standard' of computational quantum chemistry. When combined with sufficiently large basis sets, this method can reproduce experimentally derived thermochemical properties to within 1 kcal mol^{-1}, the so-called threshold of *chemical accuracy*.

1.3 Composite Wavefunction Methods

Unfortunately, the combination of CCSD(T) with a robust basis set is only affordable on the smallest chemical systems. To achieve chemical accuracy on larger systems, or to achieve *sub*-chemical accuracy, a composite approach must be employed. The focal point approach[10, 11] (FPA) is one such method; it systematically increases both the basis set size and the order of the correlation treatment. This allows the systematic elimination of basis set deficiencies using accepted extrapolation techniques,[12, 13] and allows the estimation of the full CI limit. With this method, higher order correlation effects, through CCSDTQ(P), are incorporated into a broader energy calculation at the complete basis set (CBS) limit. Basis set extrapolations are carried out with even tempered basis sets. In this work the cc-pV*n*Z (*n* = 2-6) family of basis sets were employed,[14-18] where *n* denotes the number of contracted basis functions for each valence atomic orbital. The computed energy as a function of *n* is measured and for HF energies a 3-point extrapolation is used (eq. 1.9), while for correlation energies a 2-point extrapolation is used (eq. 1.10).

$$E_{HF} = a + b^{-cn} \tag{1.9}$$

$$E_{corr} = a + bn^{-3} \tag{1.10}$$

Different functional forms are required, because the HF energy is based upon a one-electron operator, whereas the correlation energy is a two-electron phenomenon. Both forms are heuristic

in nature. To further reduce errors, and approach the realm of sub-chemical accuracy, a suite of auxiliary calculations may also be included within the FPA. Corrections for core electron correlation, the BO approximation,[19, 20] and special relativity are computed,[21, 22] and additively included.

1.4 Density Functional Theory

All of the QM methods described so far have utilized an electronic wavefunction, an inherently complicated mathematical construction with 3*n* (where *n* in the number of electrons) degrees of spatial freedom. Density functional theory (DFT) reduces the complexity by a factor of *n*, using the one-electron density (ρ) in place of the electronic wavefunction.[23, 24] This dramatically simplifies the calculations of molecular properties [O(n^3) scaling]; however the exact form of the functional, which leads to the correct calculation of ρ, is unknown. The generic form of the DFT functional is remarkably similar to that of HF, the lone exception being the inclusion of an exchange-correlation term (E_{xc}). Flavors of DFT can be categorized in a scheme known as "Jacob's Ladder.[25]" In this system, the lowest "rung" model is the local density approximation (LDA), where only the electron density is utilized in determining the energy. The next rung incorporates the density gradient (GGA) and the third rung includes kinetic energy into the Hamiltonian. The fourth rung incorporates exact exchange, typically by mixing in contributions from HF theory. The B3LYP functional[26, 27] is one of the most widespread quantum chemical methods in use today, owing to its accuracy for small to medium sized, covalently bonded molecules.[28] Unfortunately B3LYP suffers from poor results when calculating reaction barriers,[29] when non-bonded interactions are important[30] and for increasingly large systems with a roughly linear topology. [31] This functional is a hybrid functional, mixing together corrections from LDA, GGA and HF methods in an empirically determined fashion (eq.

1.11). In general however, the biggest disadvantage when using DFT is the inability to include greater amounts of correlation in a straight forward manner. With wavefunction based methods, high order expansions are always available within each correlation treatment, and extra terms can be included to systematically reduce errors. DFT methods are more of a black box, and no such flexibility exists. This property makes DFT ill-suited to address chemical problems where errors must be quantitatively known. Furthermore, DFT methods are typically developed with a certain class of chemical compounds in mind: organic molecules, transition metals and liquids for example. To reduce errors for the systems which DFT methods are tailored, empirically derived parameters are often included in the density functional. This has the inherent disadvantage of making these methods less transferable. Care must always be taken to validate DFT results against highly correlated wavefunction methods.

$$E_{xc}^{B3}[\rho] = E_{xc}^{LDA} + c_1(E_x^{HF} - E_x^{LDA}) + c_2\Delta E_x^{GGA} + c_3\Delta_c^{GGA} \tag{1.11}$$

1.5 Classical Mechanics Methods

If one wishes to model even larger systems, further simplifications to the physical model become necessary. Modeling macromolecular systems (proteins, lipid membranes and carbohydrates) is typically carried out using classical mechanics. The atoms themselves are modeled as discrete objects, and their motions in Cartesian space are governed by a straightforward Hamiltonian (eq. 1.12). This is built up from different types of chemically intuitive bonded terms and physically motivated non-bonded terms. The bonded terms each correspond to an internal coordinate such as bond stretching, angle bending and dihedral rotations. Auxiliary terms can also be included, such as improper dihedral terms, which are heuristically included to enforce the proper chirality. Non-bonded terms are present as well, to describe electrostatics, core-core repulsion and van der Waals dispersion.

$$E(\vec{R}) = \sum_{bonds} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{impropers} K_\omega(\omega - \omega_0)^2 + \sum_{i>j} \left\{ \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\} \quad (1.12)$$

The force constants, K_x, and the minima X_0, taken together with the various non-bonded constants are collectively known as a *force field*, and are determined either from experimental results or QM calculations. Force fields which are parameterized to model a wide variety of biochemical systems are said to be transferable. Most force fields are parameterized with at least a specific class of macromolecules in mind, such as proteins. Others force fields are built with a specific system in mind and are non-transferable. Some of the most commonly used force fields today are: CHARMM,[32, 33] AMBER,[34] GROMOS[35] and OPLS.[36]

CG models further reduce the complexity of the model, in order to allow simulations on biological time and length scales. The simplest CG models are elastic network models (ENM), where the equilibrium position of a system is represented by a network of beads that are interconnected by harmonic oscillators. These 'springs' account for both bonding and non-bonding interactions, and are able to provide surprising insight given the simplistic nature of the model. Harmonic networks can provide qualitative information about several facets of a protein, including secondary structure,[37] principal modes[38] and domain decompositions.[39] ENMs have also been utilized to enhance sampling[40] and refine low resolution experimental data.[41]

Gō-like models[42] are a class of CG models designed to simulate protein folding, and were originally conceived as a one-site model, where each amino acid is mapped into one interaction center. This model's structure is biased towards the experimentally determined native state using the standard bonded potentials from equation 1.12 and an additional set of non-bonded interactions, each represented by a Lennard- Jones (LJ) potential. Because this model has

dramatically fewer interaction sites than an all-atom model, its potential energy surface (PES) is much smoother, and thus the protein can be expected to fold by satisfying the principle of satisfied frustration.[43] That is to say, there are less local minima on which the protein can misfold in the simulation. Single site Gō models tend to have a potential that is *too* smooth however, and because of this it is impossible to investigate metastable states.

In contrast, a balance between the simplicity of a one-site model and the ruggedness of an all-atom model is achieved with off-lattice models, which employ two bonding centers per amino acid. These models can more robustly investigate conformational properties of proteins in the presence of external stimuli, such as denaturants and high temperature. In the off-lattice Gō model,[44-47] each amino acid is represented by an interaction center at the C_α position, and one at the side chain center of mass position. The standard bonded potentials are employed for bonds, angles, dihedrals and impropers. LJ potentials are used to mimic the effects of hydrogen bonds, which are derived from the secondary structure of the all-atom representation of the model. Another set of LJ potentials are used to mimic the attractions between side-chains which are in contact in the natively folded state. This contact parameter set is derived from experiment, and together with the crystal structure of the native state, serves as the basic inputs from which the full parameter set for the model is derived. More details of this method will follow in chapter 5.

Because biological molecules exist at biological temperatures, we are seldom only concerned with the system's minimal energy structure. Furthermore, finding the global minima is a non-trivial task because of the rugged nature of the protein's PES. Instead, proteins at biological temperatures exist as a thermodynamic ensemble and so we must consider a properly weighted ensemble average if we wish to calculate properties from our simulations. The most straight forward method is molecular dynamics, where a trajectory of a protein's motions is

calculated by assuming a starting conformation and numerically integrating the equations of motion. Ideally, the protein will visit all of its conformations in the proper proportions regardless of the length of the simulations, or its starting conditions. In reality however, one must take care their simulation is non-ergodic, and that their statistics are properly converged.[48] Enhanced sampling methods are a class of techniques devised to correct for this very problem. Replica exchange molecular dynamics[49] (REMD) combines simulations performed at different temperatures to overcome local barriers, and allow sampling of the entire conformational space. REMD has been combined with the off-lattice model to enhance the sampling and obtain accurate thermochemical properties.

1.6 Outline of Research Projects

This work presents three projects related to modeling accurate thermochemistry of small molecules using highly correlated wavefunction techniques, and one project on implementing a generalized framework for building coarse grained macromolecule models. Chapter 2 describes the ground state PES of the formaldehyde molecule, and its various properties computed using FPA at the CCSDTQ(P)/CBS level of theory. Anharmonic corrections to vibrational and rotational spectroscopic properties are presented, alongside updated reaction energetics. This modernization of one of the most well studied molecules should resolve the 2 kcal mol^{-1} discrepancy between the observed experimental reaction barrier and its theoretical counterpart. Chapter 3 further delves into the properties of this system in the 'activated' HCOH conformation.[50] This work describes the novel synthesis of HCOH, its computed and observed vibrational spectra and its rapid decomposition via quantum tunneling. This molecule has broad implications for interstellar chemistry, and serves as a canonical example for quantum tunneling through large barriers near absolute zero temperature. Chapter 4 describes the computed

properties of HSiN, an analogue of the well-studied HCN system which is thought to exist in interstellar space.[51] Anharmonic corrections to computed spectroscopic properties are presented, and this molecule's potential to undergo isomerization via tunneling is discussed. Finally, chapter 5 presents the implementation of a generalized framework for building coarse grained models in the computer package CHARMM. Two small fast folding proteins are investigated: the albumin binding domain (PDB accession code: 1prb) a prototypical alpha helix protein and protein G (PDB accession code: 2qmt) a prototypical beta sheet.

1.7 References

1. D. R. Hartree, Math Proc. Cambridge **24**, 89-110 (1928).
2. C. J. Cramer, *Essentials of Computational Chemistry*. (Wiley, West Sussex, 2004).
3. C. Eckart, Phys. Rev. **46**, 383-387 (1934).
4. H. A. Jahn and E. Teller, P. Roy. Soc. Lond. A **161**, 220-235 (1937).
5. C. Møller and M. S. Plesset, Phys. Rev. **46**, 618-622 (1934).
6. T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny and W. D. Allen, J. Comput. Chem. **28**, 1610-1616 (2007).
7. J. Čížek, J. Chem. Phys. **45**, 4256-4266 (1966).
8. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479-483 (1989).
9. J. F. Stanton, Chem. Phys. **281**, 130-134 (1997).
10. W. D. Allen, A. L. L. East and A. G. Császár, *Structures and Conformations of Non-Rigid Molecules*. (Kluwer, Dordrecht, 1993).

11. M. S. Schuurman, S. R. Muir, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **120**, 11586-11599 (2004).

12. R. A. Kendall, T. H. Dunning and R. J. Harrison, J. Chem. Phys. **96**, 6796-6806 (1992).

13. D. Feller, J. Chem. Phys. **98**, 7059-7071 (1993).

14. T. H. Dunning, J. Chem. Phys. **90**, 1007-1023 (1989).

15. D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358-1371 (1993).

16. D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572-4585 (1995).

17. K. A. Peterson and T. H. Dunning, J. Chem. Phys. **117**, 10548-10560 (2002).

18. A. K. Wilson, T. van Mourik and T. H. Dunning, J. Mol. Struc. **388**, 339-349 (1996).

19. H. Sellers and P. Pulay, Chem. Phys. Lett. **103**, 463-465 (1984).

20. N. C. Handy, Y. Yamaguchi and H. F. Schaefer, J. Chem. Phys. **84**, 4481-4484 (1986).

21. S. A. Perera and R. J. Bartlett, Chem. Phys. Lett. **216**, 606-612 (1993).

22. R. D. Cowan and D. C. Griffin, J. Opt. Soc. Am. **66**, 1010-1014 (1976).

23. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864-B871 (1964).

24. W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133-A1138 (1965).

25. J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, J. Chem. Phys. **123**, 062201-062209 (2005).

26. A. D. Becke, Phys. Rev. A **38**, 3098-3100 (1988).

27. C. Lee, W. Yang and R. G. Parr, Phys. Rev. B **37**, 785-789 (1988).

28. E. C. Sherer, K. N. Kirschner, F. C. Pickard, C. Rein, S. Feldgus and G. C. Shields, J. Phys. Chem. B **112**, 16917-16934 (2008).

29. Y. Zhao and D. Truhlar, Theor. Chim. Acta **120**, 215-241 (2008).

30. Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **109**, 5656-5667 (2005).

31. M. D. Wodrich, C. Corminboeuf and P. v. R. Schleyer, Org. Lett. **8**, 3631-3634 (2006).

32. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, J. Comput. Chem. **4**, 187-217 (1983).

33. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, J. Comput. Chem. **30**, 1545-1614 (2009).

34. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179-5197 (1995).

35. W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren, J. Phys. Chem. A **103**, 3596-3607 (1999).

36. W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657-1666 (1988).

37. Y. Wu, X. Yuan, X. Gao, H. Fang and J. Zi, Phys. Rev. E **67**, 041909 (2003).

38. B. Isin, P. Doruker and I. Bahar, Biophys. J. **82**, 569-581 (2002).

39. S. Kundu, D. C. Sorensen and G. N. Phillips, Proteins **57**, 725-733 (2004).

40. Z. Zhang, Y. Shi and H. Liu, Biophys. J. **84**, 3583-3593 (2003).

41. M. Delarue and P. Dumas, Proc. Natl. Acad. Sci. USA **101**, 6957-6962 (2004).

42. Y. Ueda, H. Taketomi and N. Gō, Biopolymers **17**, 1531-1548 (1978).

43. D. Baker, Nature **405**, 39-42 (2000).

44. J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526-3529 (1990).

45. D. K. Klimov and D. Thirumalai, J. Chem. Phys. **109**, 4119-4125 (1998).

46. D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. USA **97**, 2544-2549 (2000).

47. E. P. O'Brien, G. Ziv, G. Haran, B. R. Brooks and D. Thirumalai, Proc. Natl. Acad. Sci. USA **105**, 13403-13408 (2008).

48. A. Roitberg and C. Simmerling, J. Mol. Graphics **22**, 317-317 (2004).

49. Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141-151 (1999).

50. P. R. Schreiner, H. P. Reisenauer, F. C. Pickard, A. C. Simmonett, W. D. Allen, E. Mátyus and A. G. Császár, Nature **453**, 906-909 (2008).

51. M. C. Lind, F. C. Pickard, J. B. Ingels, A. Paul, Y. Yamaguchi and H. F. Schaefer, J. Chem. Phys. **130**, 104301-104313 (2009).

CHAPTER 2

THE BARRIER TO MOLECULAR DISSOCIATION OF FORMALDEHYDE[1]

[1] Pickard, F. C.; A. C. Simmonett and H. F. Schaefer III. To be submitted to *The Journal of Chemical Physics*.

2.1 Abstract

The electronic ground state ($\tilde{X}\,^1A_1$) of formaldehyde and its barrier to molecular elimination were rigorously studied using high order coupled cluster techniques [up through CCSDTQ(P)]. The correlation-consistent family of polarized valence (cc-pV*n*Z), augmented correlation-consistent polarized valence (aug-cc-pV*n*Z), correlation-consistent polarized core-valence (cc-pCV*n*Z) and augmented correlation-consistent polarized core-valence (aug-cc-pCV*n*Z) (*n* = D, T, Q, 5, 6) basis sets were employed for energy calculations. Via focal point analysis, we confirmed that molecular products hydrogen and carbon monoxide lie below formaldehyde by 1.81 kcal mol^{-1}, a result which agrees well with the best available experimental information, 2.09 kcal mol^{-1} from the active thermochemical tables (ATcT) of Ruscic *et al.* The barrier to molecular products was determined to be 80.82 kcal mol^{-1}, a result which stands between the two most widely accepted experimental values, and disagrees with the most recent *ab initio* study by ~1 kcal mol^{-1}.

2.2 Introduction

Due to its ubiquity, small size, high symmetry and relative stability on its ground electronic state potential energy surface (PES), formaldehyde ($^1A_1$, S_0) presents a uniquely tractable test case for cutting edge experimental and theoretical techniques. Its spectroscopic properties are well known,[1] and its rotational spectra[2] was among the first elucidated for a polyatomic molecule. Formaldehyde also plays key roles in interstellar,[3-5] atmospheric[6] and combustion[7] chemistry. For these reasons, the gas phase unimolecular decomposition of formaldehyde has been extensively studied both experimentally[8-15] and theoretically.[16-24] Experimentalists have studied this reaction using a combination of photolysis and pyrolysis techniques. However because electronically excited states are rapidly converted to the electronic ground state by both radiative (eq. 2.1) and non-radiative (eq. 2.2) decay, these two processes are deeply intertwined. Following an excitation to the first excited singlet state ($\pi^* \leftarrow n$, $^1A_2 \leftarrow {}^1A_1$), the following chemical processes may result:

$$H_2CO(S_1) \rightarrow H_2CO\ (S_0) + h\nu \qquad (2.1)$$

$$H_2CO(S_1) \rightarrow [H_2CO\ (S_0)]^* \qquad (2.2)$$

$$H_2CO(S_1) \rightarrow H + HCO^* \qquad (2.3)$$

At slightly higher photolysis energies, the channel to radical dissociation (eq. 2.3) begins to open. When decay to S_0 occurs non-radiatively, high degrees of vibrational excitation are observed, which often leads to further decomposition:

$$[H_2CO\ (S_0)]^* \rightarrow H_2 + CO \qquad (2.4a)$$

$$[H_2CO\ (S_0)]^* \rightarrow [H \cdots HCO]^* \rightarrow H_2 + CO \qquad (2.4b)$$

$$[H_2CO\ (S_0)]^* \rightarrow H + HCO \qquad (2.5)$$

Reactions 2.4 and 2.5 may be initiated by photolysis as illustrated above, or directly via pyrolysis. Equation 2.4 shows the reaction pathways leading to molecular products, where the

pathway in equation 2.4a is the historically well-known 'molecular elimination' channel and the pathway in equation 2.4b is the 'roaming atom' pathway, which was first proposed 18 years ago,[14] and recently elucidated with a combination of experiment and theory.[23, 25]

There are two experimental methods for determining formaldehyde's barrier to molecular product formation: one based on photolysis and one based on pyrolysis. By utilizing the Stark level-crossing technique, Polik *et al.* produced the most reliable experimental photolysis results to date.[13] This work established the barrier to molecular elimination (eq. 2.4a) at 79.2±0.8 kcal mol^{-1}, a measurement that stood as the uncontested experimental benchmark until recently. Recently, shock tube experiments have been performed, where pyrolysis rates of formaldehyde dissociation are measured directly.[15] From these high temperature kinetics results, a quantum state resolved kinetic model was constructed to reproduced high temperature dissociation rates, and photolysis quantum yields.[26, 27] A barrier to molecular dissociation was then backed out of this model predicting a barrier to molecular dissociation of 81.7±0.5 kcal mol^{-1} to molecular elimination.

Throughout the past 30 years the barrier to molecular products has been extensively modeled by *ab initio* theory.[28] Most recently, CCSD(T)[29, 30] theory was employed by Feller *et al.*[21] along with the aug-cc-pVTZ[31] basis set to model the barrier to molecular products. This calculation was refined by a series of auxiliary calculations in an attempt to correct for deficiencies to both the correlation treatment and basis set. First, a single CCSDT[32] calculation was used to estimate the full configuration interaction (FCI) limit. Second, to estimate the basis set limit, a series of extrapolations was used, which were later shown to give inferior results to now standardized extrapolation techniques.[33, 34] Third, zero-point energy contributions were estimated using the harmonic approximation. Anharmonic effects, which can lead to

discrepancies of up to 50 cm^{-1} between experimentally observed and theoretically calculated normal modes, were not modeled in this study.

The apparent discrepancy in experimental results should be reconcilable via electronic structure theory. At first glance this appears hopeful, as the most sophisticated *ab initio* calculation in the literature[21] estimates the zero-point corrected barrier to molecular products is 81.9±0.3 kcal mol^{-1}. Further inspection of the procedure undertaken to arrive at this barrier reveals that this calculation was state of the art at the time of publication, but it is now showing its age nearly ten years later. We will improve upon this old calculation using the focal point approach (FPA) of Allen *et al.*,[35-39] in conjunction with high order coupled-cluster (CC) theories that can efficiently and accurately approach both the basis set and correlation limits. For small molecules such as formaldehyde, is it also possible to incorporate anharmonic effects into calculated molecular properties. Using second-order vibrational perturbation theory (VPT2) in conjunction with a high quality quartic force-field (QFF) expanded about the equilibrium structures and transition state (**TS**), anharmonic effects are reasonably accounted for, and incorporated into computed properties such as vibrational frequencies and rotational constants. Corrections to the Born-Oppenheimer approximation, treatment of core-electron correlation and modeling of relativistic effects can further improve *ab initio* calculations to the realm of 'sub-chemical' accuracy (±0.25 kcal mol^{-1}). The use of this composite approach should improve the current electronic structure calculation, and perhaps resolve the inconsistency between the two experimental measurements.

2.3 Methods

In an attempt to resolve the disparity between various experimental deductions of formaldehyde's barrier to molecular dissociation a set of state of the art, rigorous *ab initio*

calculations was undertaken. Equilibrium geometries on the ground-state singlet molecular dissociation channel potential energy surface (PES) were optimized with the highly accurate all-electron (AE) coupled cluster theory including the single, double and perturbative triple excitations [CCSD(T)] level of theory,[29, 30] in conjunction with the robust cc-pCVQZ basis set.[40]

To obtain highly accurate energetics for the three stationary points along the reaction coordinate, we employed the FPA.[35-39] This method seeks to systematically eliminate the errors associated with deficiencies to both the one-particle basis set and the correlation treatment. However, unlike other popular composite approaches such as HEAT[41] or W4,[42] the FPA treatment for a given system is not explicitly prescribed. Instead, the computational chemist monitors convergence to both the one-particle and many-particle limits, and then tailors the FPA to an acceptable accuracy within the allotted computational resources.

A series of FPA calculations was completed using the various families of correlation consistent polarized valence basis sets[31, 40, 43, 44] [(aug)-cc-p(C)V*n*Z] (*n* = D, T, Q, 5, 6) and the coupled cluster (CC) hierarchy of electron correlation methods. The complete basis set (CBS) limit was estimated using established basis set extrapolation formulae. The SCF and correlation contributions to the absolute energies of each molecule were extrapolated separately. The HF/CBS energy is estimated via a three-point exponential extrapolation (*n*= Q, 5, 6),[33] while the CCSD(T) limit is estimated by a two-point cubic power fit (*n* = 5, 6).[34] Because of the small system size, it is computationally tractable to compute correlation effects beyond CCSD(T), namely CCSDT,[32] CCSDT(Q),[45, 46] CCSDTQ[47, 48] and CCSDTQ(P).[49] These correlation energies are incorporated into the FPA energy additively.

Finally, to approach the realm of sub-chemical accuracy, a suite of auxiliary corrections was employed. Because the primary focal point extrapolations employed the frozen-core

approximation, the effects of core-electron correlation (Δ_{core}) were incorporated into the final FPA results by taking the difference of the all-electron and the frozen-core CCSD(T)/cc-pCVQZ computations. To correct for hydrogen atom motion during formaldehyde's molecular dissociation, the diagonal Born-Oppenheimer correction[50, 51] (Δ_{DBOC}) was evaluated at the HF/aug-cc-pVTZ level of theory, as prescribed by the HEAT protocol.[41] Special relativistic effects (Δ_{rel}) were also accounted for at first order via the calculation of the mass-velocity and one-electron Darwin terms.[52, 53]

Harmonic frequency computations for all structures involved in the molecular dissociation were computed for optimized geometries to confirm them as true stationary points. Furthermore, quartic force fields were calculated for these structures. This explicitly quantifies the effects of anharmonicity upon predicted molecular properties, such as the zero-point vibrational energy (ZPVE) and the fundamental vibrational frequencies, as well as to modernize the force field values present in the literature.[20] The force fields are computed via finite differences of energies at the AE-CCSD(T)/cc-pCVQZ level of theory. We employed standard displacement sizes of 0.01 Å for bond stretch coordinates and of 0.02 radians for linear and out-of-plane bending coordinates. To maintain the numerical integrity of the differentiations, energies computed at displaced geometries must be very tightly converged (10^{-12} E_h), lest numerical artifacts may be introduced to the calculated force constants. After transformation to reduced normal coordinates, second order vibrational perturbation theory (VPT2)[54-59] was applied to the force fields to obtain anharmonic properties and spectroscopic constants. In cases where type 1 ($\omega_i \approx 2\ \omega_j$) or type 2 ($\omega_i \approx \omega_j + \omega_k$) Fermi resonances are present, the explicit diagonalization of the harmonic coupling term was required to accurately calculate the anharmonic coupling constants χ_{ij}.

Most electronic structure calculations present in this work were carried out using the 2006 version of MOLPRO.[60] The PSI3[61] and ACESII[62, 63] (Mainz-Austin-Budapest version) suites of quantum chemistry programs were used for DBOC and relativistic corrections, respectively. Higher order CC calculations required the use of the MRCC[46, 64, 65] code of Kállay as interfaced to ACESII. The Mathematica program INTDIF2005[66] generated the displaced geometries and computed force constants in symmetrized internal coordinates. The INTDER2005[67-69] code performed the non-linear coordinate transformations of force constants between symmetry and Cartesian coordinates. Finally, the VPT2 analysis was carried out upon the Cartesian force constants via the ANHARM program.[70]

2.4 Results and Discussion

2.4.1 Geometries

Optimized structures at the CCSD(T)/cc-pCVQZ level of theory are shown in Figure 2.1. The importance of core-correlation and the associated basis functions is very clear from a comparison with the structures found in Feller's paper. Upon the inclusion of core-correlation, bond lengths contract as electron density populates the core functions. The inclusion of core-correlation gives our predicted *r*(C-O) distances more favorable agreement with experiment (~0.001 Å) than the most recent *ab initio* calculation in the literature (~0.01 Å). Our better bond distances will, then in turn, manifest themselves as more reliably computed properties, such as vibrational modes and rotational constants. This trend continues with the **TS** structure, as there is a significant difference (~0.01 Å) between our *r*(C-O) value and that from the most recent theory. We conclude that our superior agreement with available experimental results on the equilibrium structures corroborates our choice of theoretical method for the ensuing transition state calculations.

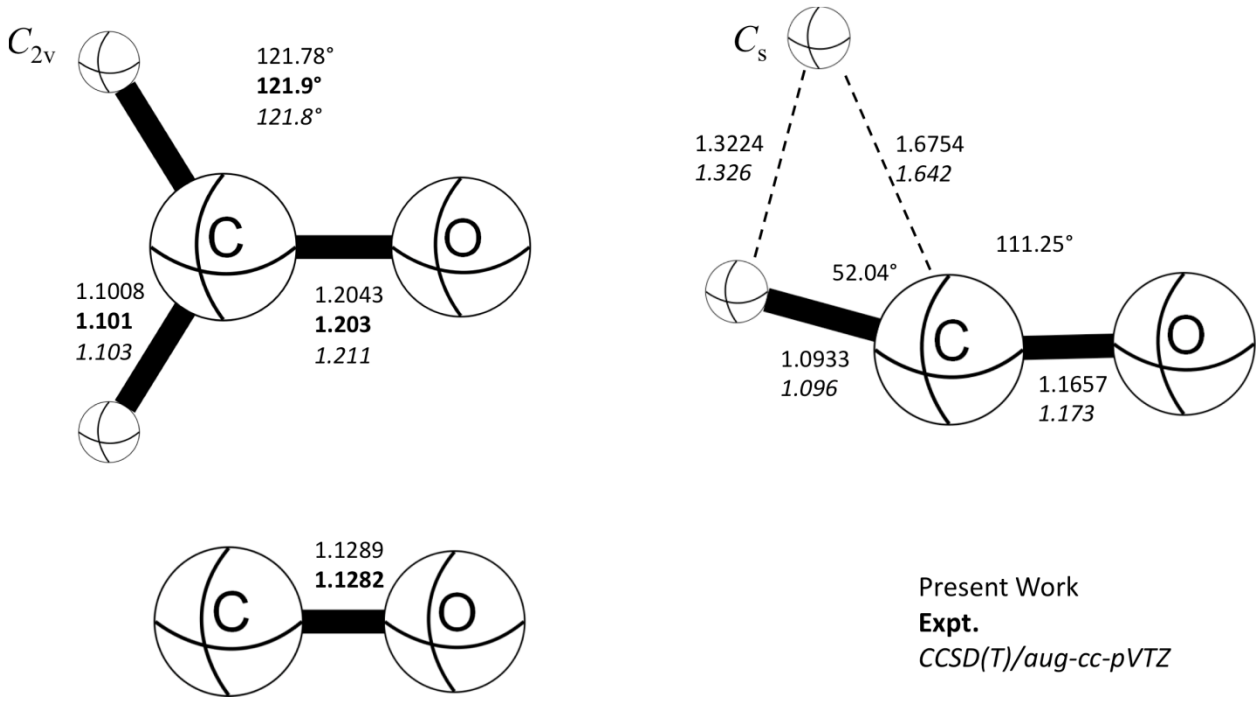

Figure 2.1 Geometries optimized at the CCSD(T)/cc-pCVQZ level of theory, compared with experiment (formaldehyde[71] and carbon monoxide[72]) and the previous best *ab initio* results[21] [CCSD(T)/aug-cc-pVTZ]. Units are in angstroms and degrees.

2.4.2 Basis Set Effects

To compute accurate energies and barriers of reaction, three different FPA extrapolation schemes were employed. Each of these schemes employed a different member of the core-valence family of basis sets (cc-pV*n*Z, aug-cc-pV*n*Z and cc-pCV*n*Z, respectively) where *n* ran from double-ζ through sextuple-ζ. From these extrapolations, we were able to infer that each of the different basis sets arrived at the same CCSD(T)/CBS limit (within 0.05 kcal mol^{-1}) if one accounts for core-correlation. This is an important result, as the computational expense associated with diffuse functions and core-correlation becomes a problem when going beyond CCSD(T). Thus, we can extrapolate to the CCSDTQ(P)/CBS limit using the inexpensive core-valence basis sets.

2.4.3 Energetics

Our CCSD(T) based FPA is consistent with previous *ab initio* calculations[21] of both reaction energy and reaction barrier. Upon incorporation of higher order excitations to the extrapolated energy, the computed reaction energy to molecular products changes by less than 0.05 kcal mol^{-1}, demonstrating the literature values are well converged with respect to basis set and correlation treatment. However an additional ~0.5 kcal mol^{-1} is recovered upon inclusion of CCSDT and CCSDT(Q) energies into the FPA. Further inclusion of CCSDTQ and CCSDTQ(P) effects has a minimal effect upon the computed barrier energy (~0.05 kcal mol^{-1}), demonstrating the CCSDT(Q)/CBS level to be the "Pauling Point" for this barrier. These results are not entirely surprising, as CCSD(T) has been shown to reasonably accurately reproduce experimental thermochemistry for single reference systems near equilibrium. A more rigorous correlation treatment is required, however, for **TS**, as its distorted bonds introduce some multi-reference character into the molecular wavefunction. The convergence toward the correlation limit is best illustrated using an incremental FPA (Tables 2.1 and 2.2). In the incremental FPA tables, each column beyond RHF shows the computed energy difference between the current and previous correlation treatments. In this way, the scientist gains explicit understanding of differential cost versus benefit of more rigorous methods.

2.4.4 Vibrational Frequencies

Computed harmonic and anharmonic frequencies for formaldehyde and formaldehyde-d_2 are presented in Table 2.3, alongside their corresponding experimental values determined by FT-IR.[73] As expected, harmonic frequencies of formaldehyde are substantially higher than the experimental fundamentals. Upon incorporation of anharmonic corrections from VPT2 theory, the computed fundamentals correct considerably with respect to experiment (RMSD < 25 cm^{-1}).

	RHF	+δ MP2	+δ CCSD	+δ CCSD(T)	+δ CCSDT	+δ CCSDT(Q)	+δ CCSDTQ	+δ CCSDTQ(P)	Σ
cc-pVDZ	-1.01	1.92	-0.11	-0.51	0.01	-0.04	0.01	0.01	0.27
cc-pVTZ	-0.86	5.28	-0.40	-0.26	0.02	-0.06	0.01	0.01	3.74
cc-pVQZ	-0.85	6.05	-0.47	-0.14	0.02	-0.06	0.01	0.01	4.56
cc-pV5Z	-0.66	6.23	-0.51	-0.10	0.02	-0.06	0.01	0.01	4.93
cc-pV6Z	-0.64	6.29	-0.52	-0.09	0.02	-0.06	0.01	0.01	5.01
CBS	-0.64	6.38	-0.53	-0.08	0.02	-0.06	0.01	0.01	**5.11**

Δ ZPVE (harmonic) =		-7.388	Δ ZPVE (anharmonic) =	-7.175	
Δ (core correlation) =		0.339			
Δ (relativity) =		-0.176			ATcT[a]
Δ (DBOC) =		0.044			(experiment)
Final ΔE_0 =	(harmonic)	-2.073	(anharmonic)	-1.860	-2.09

Table 2.1 Incremental FPA for ΔE_0 [H_2 + CO] − H_2CO, with all energies in kcal mol^{-1}. Extrapolated FPA values have a grey background. Final FPA value from this work is in bold.
[a] Ref. [42] and [74]

	RHF	+δ MP2	+δ CCSD	+δ CCSD(T)	+δ CCSDT	+δ CCSDT(Q)	+δ CCSDTQ	+δ CCSDTQ(P)	Σ
cc-pVDZ	101.70	-12.81	-0.13	-3.00	-0.27	-0.26	0.03	0.02	85.25
cc-pVTZ	101.42	-11.37	-0.08	-3.23	-0.19	-0.34	0.03	0.02	86.26
cc-pVQZ	101.22	-11.18	-0.02	-3.28	-0.16	-0.34	0.03	0.02	86.28
cc-pV5Z	101.17	-11.10	0.00	-3.30	-0.16	-0.34	0.03	0.02	86.32
cc-pV6Z	101.11	-11.09	0.03	-3.30	-0.16	-0.34	0.03	0.02	86.29
CBS	101.09	-11.07	0.06	-3.32	-0.16	-0.34	0.03	0.02	**86.32**

Δ ZPVE (harmonic) =	-5.370	Δ ZPVE (anharmonic) =		-5.291		
Δ (core correlation) =	0.143					
Δ (relativity) =	-0.010					
Δ (DBOC) =	-0.341					
Final ΔE_0 =	80.741	(anharmonic)	80.819			

Feller *et al.* [a]	Troe[b]	Polik *et al.* [c]
(theory)	(experiment)	(experiment)
81.9 ± 0.3	81.7 ± 0.5	79.2 ± 0.8

Table 2.2 Incremental FPA for ΔE_0 **TS** − H_2CO, with all energies in kcal mol^{-1}. Extrapolated FPA values have a grey background. Final FPA value from this work is in bold.

[a] Ref. 21

[b] Ref. 26 and 27

[c] Ref. 13

ω_n (cm^{-1})	CCSD(T)/ cc-pCVQZ	CCSD(T)/ aug-cc-pVTZ[a]	ν_n (cm^{-1})	CCSD(T)/ cc-pCVQZ	Expt.[b]
ω_1 (a_1)	2938.7	2934.4	ν_1 (a_1)	2782.9	2782
ω_2 (a_1)	1787.1	1767.4	ν_2 (a_1)	1755.8	1746
ω_3 (a_1)	1541.1	1529.9	ν_3 (a_1)	1505.8	1500
ω_4 (b_1)	1192.8	1180.8	ν_4 (b_1)	1172.2	1167
ω_5 (b_2)	3009.0	3002.3	ν_5 (b_2)	2798.6	2843
ω_6 (b_2)	1276.5	1261.3	ν_6 (b_2)	1253.1	1249

Table 2.3 Computed harmonic and anharmonic frequencies of H_2CO from the AE-CCSD(T)/cc-pCVQZ level compared with experimental and theoretical literature values, all frequencies in cm^{-1}.
[a] Ref. 21
[b] Ref. [73]

The difference between the VPT2 result and experiment is surprisingly large for mode ν_5 (~50 cm^{-1}). This is the result of a Fermi resonance between the combination band $\nu_2 + \nu_6$ and ν_5, VPT2 is known to handle such cases of quasi-degeneracy poorly. To improve the accuracy of the computed value for ν_5, the Fermi resonance cutoffs used in the VPT2 calculation should be increased to 100 cm^{-1}, which will exclude erroneously large contributions to the anharmonic correction caused by the near degeneracy. If the potentially spurious value for ν_5 is omitted, the RMSD between the VPT2 and experimental fundamentals falls to 5.9 cm^{-1}.

For **TS**, we expect our computed anharmonic values to be of similar high accuracy, and thus provide a superior ZPE correction to the computed barrier than in the literature. As is similar to the formaldehyde case, there is a potential Fermi resonance in the TS structure, as $\omega_2 + \omega_3 \approx \omega_1$.

2.5 Conclusions

Despite the potential presence of Fermi resonances, our computed anharmonic ZPVE corrections should be considered of very high quality. These ZPVE corrections, along with the other auxiliary corrections present in the FPA analysis establish our formaldehyde disassociation

ω_n (cm^{-1})	CCSD(T)/ cc-pCVQZ	CCSD(T)/ aug-cc-pVTZ[a]	ν_n (cm^{-1})	CCSD(T)/ cc-pCVQZ
ω_1 (a)	3160.3	3140.0	ν_1 (a_1)	3031.6
ω_2 (a)	1862.0	1838.3	ν_2 (a_1)	1823.2
ω_3 (a)	1299.0	1279.5	ν_3 (a_1)	1229.1
ω_4 (a)	792.0	782.7	ν_4 (b_1)	757.6
ω_5 (a)	1840.2*i*	1831.6*i*	ν_5 (b_2)	
ω_6 (a')	875.9	858.8	ν_6 (b_2)	851.8

Table 2.4 Computed harmonic and anharmonic frequencies of **TS** from the AE-CCSD(T)/cc-pCVQZ level, compared with theoretical literature values, all frequencies in cm^{-1}. [a] Ref. 21

barrier as the new benchmark in the literature. A systematic breakdown of the energy contributions as compared to the previous best calculation[21] in the literature is presented in Table 2.5. The main differences between the barrier computed in the present work and that from the literature primarily lays in the inclusion of higher order correlation effects (HLC) which account for 0.8 kcal mol^{-1} difference. The DBOC also provides a substantial correction to the computed barrier, contributing 0.3 kcal mol^{-1} difference. Anharmonic contributions to the ZPE provide a modest correction of less than 0.1 kcal mol^{-1}. Other contributions account for less than 0.1 kcal mol^{-1} difference between the two calculations, and are minor. This updated benchmark value of formaldehyde's barrier to molecular products should prove useful. Formaldehyde has served as an excellent test bed for novel experimental and theoretical methods for the past half century, and this friendly rivalry will certainly continue.

ΔE^{TS}	Feller *et al.*		Present Work	
fc-CCSD(T)/CBS		86.7		86.77
+δ CCSDT/				
aug-cc-pVTZ[†]	+δ(0.6)	87.1		
cc-pVTZ			–δ(0.20)	86.57
HLC	+δ(0)	87.1	–δ(0.25)	86.32
+δ (core)	+δ(0.2)	87.3	+δ(0.14)	86.46
+δ (relativity)	–δ(0.1)	87.2	–δ(0.01)	86.45
+δ (DBOC)	+δ(0)	87.2[a]	–δ(0.34)	86.11
+δ (ZPE)	–δ(5.4)[b]	81.8[c]	–δ(5.37)	80.74
+δ (Anh ZPE)	+δ(0)	81.8	+δ(0.08)	80.82

Table 2.5 An itemized accounting of all the energy contributions to the H_2CO barrier to dissociation to molecular products from this work and the best theoretical study in the literature.[21] All energies in kcal mol^{-1}.

[†] Citing computational cost, the authors removed diffuse *f*-functions on the carbon and oxygen and the diffuse *d*-functions on the hydrogen.
[a] The authors cite 87.4 kcal mol^{-1} as their vibrationless barrier, 0.2 kcal mol^{-1} is unaccounted for.
[b] This value for +δ(ZPE) was arrived at directly from the harmonic frequencies reported in Feller *et al.* at the CCSD(T)/aug-cc-pVTZ level of theory. Taking the difference between their final ΔE_0^{TS} and the vibrationless barrier yields a +δ(ZPE) of –5.5 kcal mol^{-1}.
[c] The final value cited in Feller *et al.* for ΔE_0^{TS} is –81.9 kcal mol^{-1}.

2.6 References

1. D. C. Moule and A. D. Walsh, Chem. Rev. **75**, 67-84 (2002).
2. G. H. Dieke and G. B. Kistiakowsky, Phys. Rev. **45**, 4-28 (1934).
3. L. E. Snyder, D. Buhl, B. Zuckerman and P. Palmer, Phys. Rev. Lett. **22**, 679-681 (1969).
4. D. M. Rank, C. H. Townes and W. J. Welch, Science **174**, 1083-1101 (1971).
5. P. A. Gerakines, W. A. Schutte and P. Ehrenfreund, Astron. Astrophys. **312**, 289-305 (1996).
6. R. Meller and G. K. Moortgat, J. Geophys.-Res. Atmos. **105**, 7089-7101 (2000).
7. T. Carrington, L. M. Hubbard, H. F. Schaefer and W. H. Miller, J. Chem. Phys. **80**, 4347-4354 (1984).

8. B. A. DeGraff and J. G. Calvert, J. Am. Chem. Soc. **89**, 2247-2253 (1967).

9. J. C. Weisshaar, A. P. Baronavski, A. Cabello and C. B. Moore, J. Chem. Phys. **69**, 4720-4731 (1978).

10. J. C. Weisshaar and C. B. Moore, J. Chem. Phys. **70**, 5135-5146 (1979).

11. J. C. Weisshaar and C. B. Moore, J. Chem. Phys. **72**, 2875-2876 (1980).

12. P. Ho, D. J. Bamford, R. J. Buss, Y. T. Lee and C. B. Moore, J. Chem. Phys. **76**, 3630-3636 (1982).

13. W. F. Polik, D. R. Guyer and C. B. Moore, J. Chem. Phys. **92**, 3453-3470 (1990).

14. R. D. van Zee, M. F. Foltz and C. B. Moore, J. Chem. Phys. **99**, 1664-1673 (1993).

15. V. Vasudevan, D. F. Davidson, R. K. Hanson, C. T. Bowman and D. M. Golden, P. Combust. Inst. **31**, 175-183 (2007).

16. J. D. Goddard and H. F. Schaefer, J. Chem. Phys. **70**, 5117-5134 (1979).

17. J. D. Goddard, Y. Yamaguchi and H. F. Schaefer, J. Chem. Phys. **75**, 3459-3465 (1981).

18. M. Dupuis, W. A. Lester, B. H. Lengsfield and B. Liu, J. Chem. Phys. **79**, 6167-6173 (1983).

19. G. E. Scuseria and H. F. Schaefer, J. Chem. Phys. **90**, 3629-3636 (1989).

20. S. Carter, N. Pinnavaia and N. C. Handy, Chem. Phys. Lett. **240**, 400-408 (1995).

21. D. Feller, M. Dupuis and B. C. Garrett, J. Chem. Phys. **113**, 218-226 (2000).

22. X. B. Zhang, S. L. Zou, L. B. Harding and J. M. Bowman, J. Phys. Chem. A **108**, 8980-8986 (2004).

23. D. Townsend, S. A. Lahankar, S. K. Lee, S. D. Chambreau, A. G. Suits, X. Zhang, J. Rheinecker, L. B. Harding and J. M. Bowman, Science **306**, 1158-1161 (2004).

24. S. A. Lahankar, V. Goncharov, F. Suits, J. D. Farnum, J. M. Bowman and A. G. Suits, Chem. Phys. **347**, 288-299 (2008).

25. A. G. Suits, Acc. Chem. Res. **41**, 873-881 (2008).

26. J. Troe, J. Phys. Chem. A **109**, 8320-8328 (2005).

27. J. Troe, J. Phys. Chem. A **111**, 3862-3867 (2007).

28. S. K. Gray, W. H. Miller, Y. Yamaguchi and H. F. Schaefer, J. Am. Chem. Soc. **103**, 1900-1904 (1981).

29. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479-483 (1989).

30. J. F. Stanton, Chem. Phys. **281**, 130-134 (1997).

31. R. A. Kendall, T. H. Dunning and R. J. Harrison, J. Chem. Phys. **96**, 6796-6806 (1992).

32. Y. S. Lee, S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **81**, 5906-5912 (1984).

33. T. Helgaker, W. Klopper, H. Koch and J. Noga, J. Chem. Phys. **106**, 9639-9646 (1997).

34. A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper and J. Olsen, Chem. Phys. Lett. **302**, 437-446 (1999).

35. A. L. L. East and W. D. Allen, J. Chem. Phys. **99**, 4638-4650 (1993).

36. A. G. Császár, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **108**, 9751-9764 (1998).

37. A. G. Császár, G. Tarczay, M. L. Leininger, O. L. Polyansky, J. Tennyson and W. D. Allen, in *Spectroscopy from Space*, edited by J. Demaison, K. Sarka and E. A. Cohen (Kluwer, Dordrecht, 2001), Vol. 20, pp. 317.

38. J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Császár and G. Tarczay, Chem. Eur. J. **9**, 2173-2192 (2003).

39. M. S. Schuurman, S. R. Muir, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **120**, 11586-11599 (2004).

40. D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572-4585 (1995).

41. A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, J. Chem. Phys. **121**, 11599-11613 (2004).

42. A. Karton, E. Rabinovich, J. M. L. Martin and R. Branko, J. Chem. Phys. **125**, 144108 (2006).

43. T. H. Dunning, J. Chem. Phys. **90**, 1007-1023 (1989).

44. A. K. Wilson, T. van Mourik and T. H. Dunning, J. Mol. Struc. **388**, 339-349 (1996).

45. Y. J. Bomble, J. F. Stanton, M. Kállay and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

46. M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

47. S. A. Kucharski and R. J. Bartlett, Theor. Chim. Acta **80**, 387-405 (1991).

48. S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **97**, 4282-4288 (1992).

49. M. Musial, S. A. Kucharski and R. J. Bartlett, Chem. Phys. Lett. **320**, 542-548 (2000).

50. H. Sellers and P. Pulay, Chem. Phys. Lett. **103**, 463-465 (1984).

51. N. C. Handy, Y. Yamaguchi and H. F. Schaefer, J. Chem. Phys. **84**, 4481-4484 (1986).

52. R. D. Cowan and D. C. Griffin, J. Opt. Soc. Am. **66**, 1010-1014 (1976).

53. S. A. Perera and R. J. Bartlett, Chem. Phys. Lett. **216**, 606-612 (1993).

54. D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi and H. F. Schaefer, Chem. Phys. **123**, 187-239 (1988).

55. J. K. G. Watson, in *Vibrational Spectra and Structure*, edited by J. R. Durig (Elsevier, Amsterdam, 1977), Vol. 6, pp. 1-89.

56. I. M. Mills, *Molecular Spectroscopy: Modern Research*. (Academic Press, New York, 1972).

57. D. Papoušek and M. R. Aliev, *Molecular Vibrational-Rotational Spectra*. (Elsevier, Amsterdam, 1982).

58. H. H. Nielsen, Rev. Mod. Phys. **23**, 90-136 (1951).

59. K. Aarset, A. G. Császár, E. L. Sibert, W. D. Allen, H. F. Schaefer, W. Klopper and J. Noga, J. Chem. Phys. **112**, 4053-4063 (2000).

60. MOLPRO is a package of *ab initio* programs written by H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, A. Mitrushenkov, G. Rauhut, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, G. Hetzer, T. Hrenar, G. Knizia, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, A. Wolf .

61. T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny and W. D. Allen, J. Comput. Chem. **28**, 1610-1616 (2007).

62. J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale and R. J. Bartlett, Int. J. Quant. Chem. Symp. **26**, 879-894 (1992).

63. S. P. Nolan, *N-Heterocyclic Carbenes in Synthesis*. (Wiley-VCH, Weinheim, 2006).

64. M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945-2954 (2001).

65. H. S. Johnston, *Gas Phase Reaction Rate Theory*. (Ronald Press Inc., New York, 1966).

66. INTDIF2005 is an abstract program written by Wesley D. Allen for MATHEMATICA to perform general numerical differentiations to high orders of electronic structure data.

67. W. D. Allen and A. G. Császár, J. Chem. Phys. **98**, 2983-3015 (1993).

68. W. D. Allen, A. G. Császár, V. Szalay and I. M. Mills, Mol. Phys. **89**, 1213-1221 (1996).

69. K. Sarka and J. Demaison, *Computational Molecular Spectroscopy*. (Wiley, Chichester, 2000).

70. ANHARM is a FORTRAN program for VPT2 analysis written by Y. Yamaguchi and H. F. Schaefer, Center for Computational Chemistry, University of Georgia, Athens, GA. .

71. J. L. Duncan, Mol. Phys. **28**, 1177-1191 (1974).

72. K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules*. (Van Nostrand Reinhold Co., New York, 1979).

73. T. Nakanaga, S. Kondo and S. Saeki, J. Chem. Phys. **76**, 3860-3865 (1982).

74. B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff and A. F. Wagner, J. Phys. Chem. A **108**, 9979-9997 (2004).

CHAPTER 3

CAPTURE OF HYDROXYMETHYLENE & IMPLICATIONS FOR INTERSTELLAR CHEMISTRY

3.1 Introduction

Although singlet carbenes incorporating divalent carbon (R–C–R) have grown from laboratory curiosities to common reagents in the growing field of stable singlet carbene chemistry,[1] several parent systems still await preparation and characterization. Among them, hydroxymethylene (HCOH, hydroxycarbene) may be the most important because it is implicated in the photochemistry of its proton tautomer formaldehyde (H_2CO)[2], and was called the "activated formaldehyde" in formation of simple carbohydrates as early as 1921;[3] theoretical studies identify it as the key structure in the high-energy chemistry of H_2CO.[2, 4-7] Hydroxymethylene is the parent of alkoxycarbenes, which lie at the heart of transition-metal carbene chemistry (e.g., Fischer carbenes), a field that impacts all aspects of synthetic organic chemistry; the first metal-carbene complex, indeed was the alkoxycarbene complex $W(CO)_5(C(CH_3)OCH_3)$,[8] but all attempts to observe alkoxycarbenes from Fischer carbenes failed.[9] Electronic structure theory has indicated that hydroxymethylene should be observable,[10] and exceptionally high H/D kinetic isotope effects were predicted for the above tautomer equilibrium.[11, 12]

For a long time carbenes were considered to be too reactive to be isolable, and proofs for their existence in the condensed phase were mostly indirect. This situation changed dramatically with the advent of donor-substituted carbenes, especially *N*-heterocyclic carbenes (NHCs), which

can be isolated, stored, and used for a large variety of important chemical transformations, such as ligands in metathesis catalysts.[13] Oxygen-donor substituted carbenes however, are far less stable and have received much less attention, apart from Fischer carbine. This is primarily due to a lack of suitable synthetic precursors for this class of compounds. Therefore, it is understandable that a seemingly simple four-atom molecule such as hydroxycarbene (**1**) has not been conclusively identified. Typical synthetic carbene strategies are not effective, because appropriate starting materials are too unstable or too difficult to prepare.

To synthesize **1** and capture it, high-vacuum flash pyrolysis (HVFP) matrix isolation techniques must be employed. Water free glyoxcylic acid (**3**) is prepared by heating commercially available monohydrate under vacuum conditions for several days. The glyoxcylic acid is then evaporated from a heated storage bulb at 70 – 80 °C into the pyrolysis zone. This consists of a 50 mm long quartz tube, wrapped in a Ni/CrNi thermocouple, and heated to temperatures of up to 1000 °C. Inside the pyrolysis tube, **3** undergoes thermal extrusion of CO_2[14, 15] forming a gaseous mixture of the desired product, along with several other by products. Immediately after leaving the tube, the pyrolysis products are co-condensed with a large excess of either: argon, krypton, xenon or nitrogen on the surface of the 11 K matrix window, and then probed for their spectroscopic properties. The decarboxylation reaction of **3** does not complete, and produces a variety of products, not just the desired species (**1**). Large quantities of formaldehyde (**2**) and carbon monoxide are also synthesized in the process. In an Ar matrix at 11 K, the ratio of **1**:**2** was observed to be 1:5.5. This synthetic process can be repeated using mono-deuterated glyoxcylic acid in D_2O to produce mono-deuterated HCOD ([D_1]-**1**).

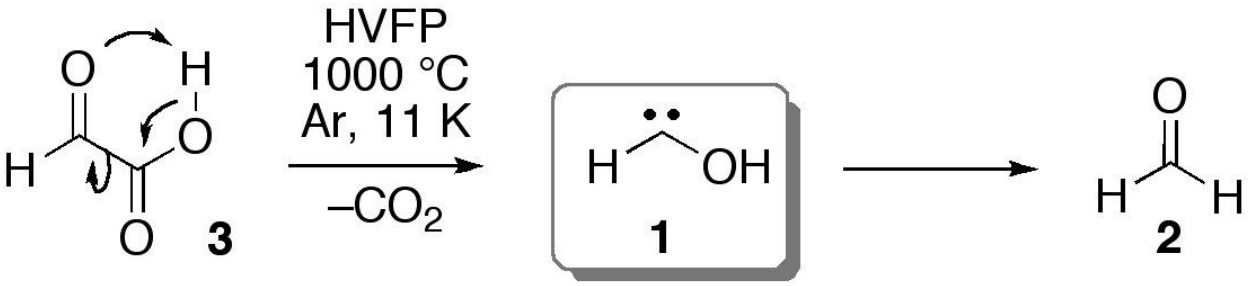

Figure 3.1 Generation of hydroxycarbene (**1**). Thermal extrusion of CO_2 through high-vacuum flash pyrolysis (HVFP) from glyoxcylic acid (**3**) and rearrangement of **1** to formaldehyde (**2**).

Despite the low product concentrations in the noble gas matrix, and the high degree of byproduct contamination, is it still possible to elaborate the IR and UV/Vis spectral properties of **1.** This is accomplished by irradiating the sample at 470 nm, causing **1** to rapidly rearrange to **2**, and then calculating the spectral differences of the irradiated and non-irradiated matrices. Using this method, the kinetics of the thermal rearrangement of **1** to **2** at 11 K (lowest achievable temperature), 15 K and 20 K may also be investigated. By repeatedly taking IR spectra in this manner, while keeping the matrix thermally and optically isolated, the extreme kinetic isotope are elucidated. Decrease of the two most intense peaks of **1** at ca. 1050 and 1300 cm^{-1} clearly followed first-order kinetics with a half-life around 2 h in Ar and Kr, while the lifetime of **1** in an N_2 matrix was prolonged by a factor of larger than three. The bands of *trans*-HCOD did not change under identical conditions for extended periods of time.

3.2 Methods

Correlation-consistent families [(aug)-cc-p(C)V*X*Z] of atomic-orbital basis sets were employed in all electronic structure computations.[16-21] Most wave functions were determined with single-reference coupled-cluster theory, incorporating all single and double (CCSD) excitations, and with perturbative inclusion of connected triple excitations [CCSD(T)].[22, 23] Higher-order correlation effects were evaluated with the recently formulated CCSDT(Q)

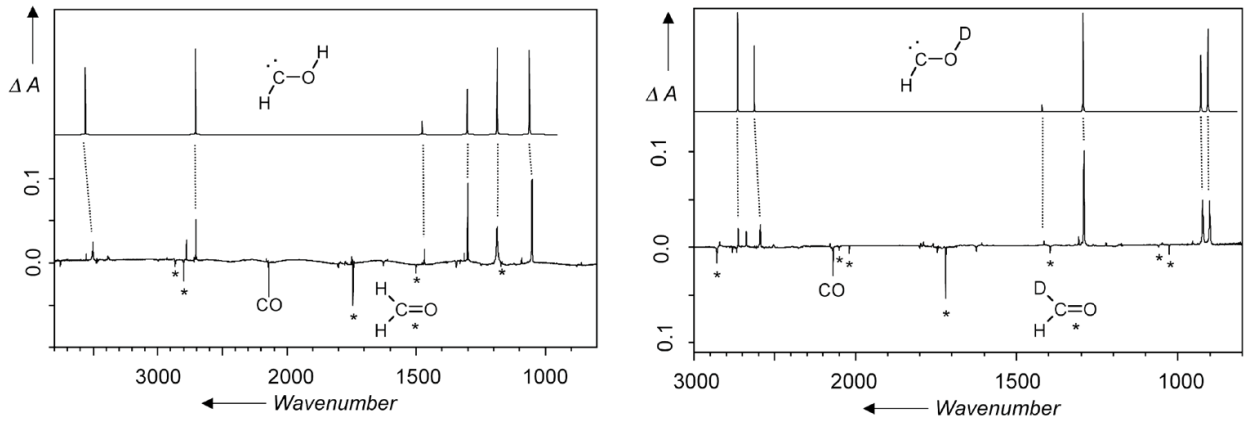

Figure 3.2 Infrared spectra of **1** and [D_1]-**1**. A) Hydroxymethylene (Ar matrix, 11 K) difference IR spectrum obtained by subtracting the spectra of the irradiated (2 min at 470 nm) and unirradiated matrix-isolated pyrolysis products of glyoxcylic acid. Upper trace: Computed variational anharmonic frequencies of *trans*-hydroxycarbene (**1t**) utilizing double-harmonic intensities at AE-CCSD(T)/cc-pCVQZ (cf. Table 3.1). The marked peaks (*) are assigned to **2**. B) [D_1]-Hydroxymethylene (Ar matrix, 11 K) difference IR spectrum obtained by subtracting the spectra of the irradiated (2 min at 470 nm) and unirradiated matrix-isolated pyrolysis products of [D_1]-glyoxcylic acid. Upper trace: Computed vibrational anharmonic frequencies of [D_1]-*trans*-hydroxycarbene utilizing double-anharmonic intensities at AE-CCSD(T)/cc-pCVQZ (Table 3.1).

method.[24, 25] Optimum geometric structures for all species on the ground-state singlet surface (Figure 3.3) were determined with all-electron (AE) CCSD(T) theory using the cc-pCVQZ basis set. Final energetics were determined from valence focal-point extrapolations[26-29] from X = 2 through 6. The focal-point results were appended with core correlation shifts based on RHF, MP2, CCSD, and CCSD(T) computations with the cc-pVXZ series of basis sets determined from all-electron vs. frozen-core cc-pCVQZ CCSD(T) computations, as well as relativistic corrections[30-32] from first-order perturbation theory applied to the one-electron mass-velocity and Darwin terms at the AE-CCSD(T)/cc-pCVQZ level. Anharmonic zero-point vibrational energies (ZPVEs) of all minima on the ground-state singlet surface were computed by applying vibrational perturbation theory (VPT2)[33-38] to complete AE-CCSD(T)/cc-pCVQZ quartic force

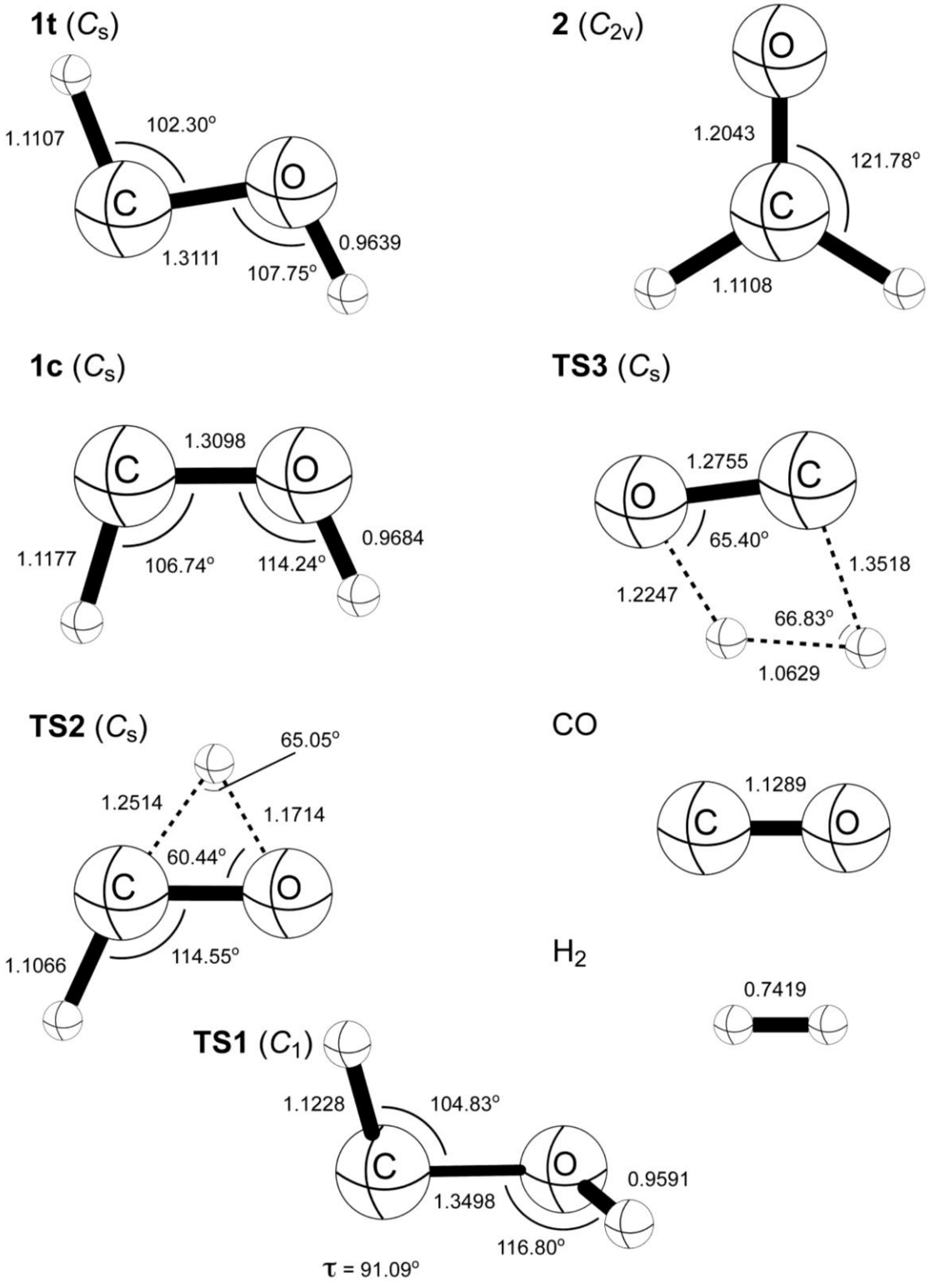

Figure 3.3 Optimized AE-CCSD(T)/cc-pCVQZ geometric structures (Å, deg).

fields, including the G_0 terms.[39] ZPVEs of all transition states were evaluated from harmonic frequencies at the AE-CCSD(T)/cc-pCVQZ level.

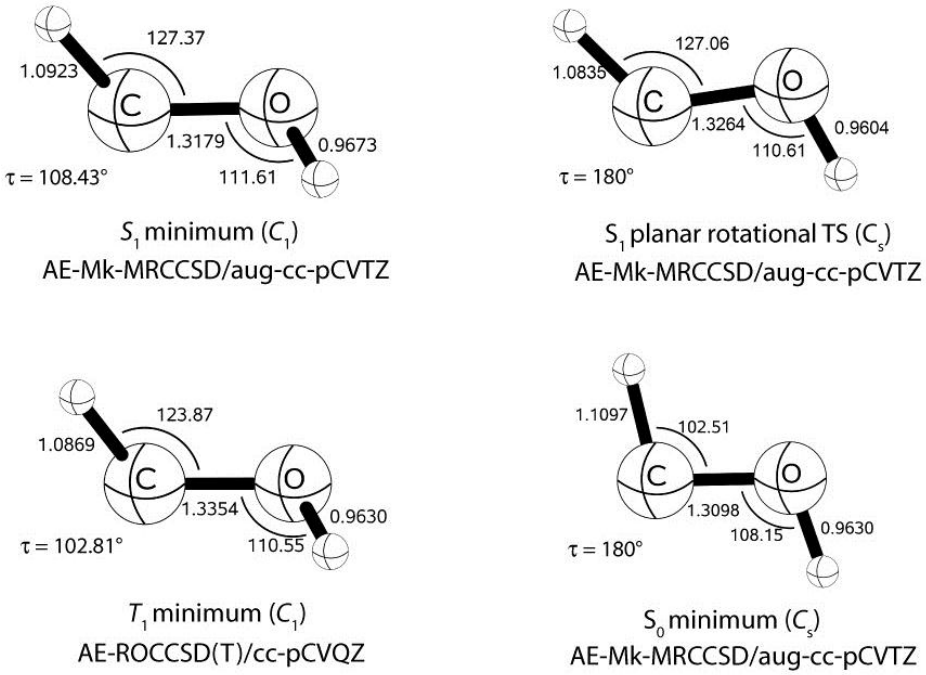

Figure 3.4 Comparative optimized geometric structures (Å, deg) of the three lowest-lying electronic states of hydroxymethylene.

The first open-shell singlet excited electronic state (S_1) of H–C–OH was investigated by the recently developed multireference coupled cluster (Mk-MRCCSD) methods. For planar geometries, the S_1 state is of $^1A''$ symmetry, arising from the $7a' \rightarrow 2a''$ molecular orbital excitation. The reference determinants for the Mk-MRCCSD computations included all possible distributions of two electrons in the ($7a'$, $2a''$) HOMO/LUMO active space. For planar geometries only two references are of the correct symmetry, but four references must be incorporated when the S_1 state twists out of plane. The optimum (twisted, C_1) structure of the S_1 state, as well as the planar transition state for internal rotation, is shown in Figure 3.3, as given by our four-reference, all-electron Mk-MRCCSD computations with the aug-cc-pCVTZ basis

set. Comparative structures for the ground-state singlet (S_0) and analogous triplet (T_1) states are also given in Figure 3.4. All multireference coupled cluster computations for the S_1 state were carried out using ROHF orbitals optimized on the corresponding triplet state (T_1).

Vibrational band origins (VBOs) were computed using two different methods, first with VPT2, and later with a variational approach called DEWE. DEWE constructs the discrete variable representation of the full Eckart-Watson Hamiltonian[40] expressed in normal coordinates and includes exactly a potential energy surface expressed in an arbitrary set of internal coordinates via an analytical transformation expression.[41] The potential energy surfaces of the *trans* conformations of HCOH and HCOD were represented by a complete quartic vibrational force field determined at the AE-CCSD(T)/cc-pCVQZ level from a grid of 263 tightly-converged energy points using established finite-difference techniques.[42, 43] This anharmonic force field was transformed to an internal coordinate set containing Simons-Parr-Finlan (SPF) variables to provide an optimal representation of the potential energy surface for vibrations in the hydroxymethylene well.[44] In the variational vibrational computations, 531441 grid points were used to construct the Hamiltonian matrix, which allowed convergence of the eigenvalues to better than 0.1 cm^{-1}. To give a quantitative interpretation of the computed vibrational levels, overlaps of the numerically exact wave functions were evaluated over multidimensional normal-mode harmonic oscillator basis functions. The squares of these coefficients comprise the normal mode distributions (NMDs) that describe the vibrational states in Table 3.1. Convergence of the NMDs presented in Table 3.1 is better than 1%.

3.3 Results and Discussion

HCOH and HCOD have been characterized through their measured IR difference spectra (Figure 3.2) and compared (Table 3.1) with our decisive electronic structure and variational

Vibrational description[a]	ω (I_{rel})	ν	Expt. (I_{rel})
trans-H–C–OH (1t)			
78% ν_1(OH str.) + 8% [ν_3 + 2 ν_6]	3765.4 (57)	3561.6	3500.6 (43)
84% [ν_4+ ν_5+ ν_6]		3520.8	
83% [3 ν_5]		3516.0	
72% [ν_3 + ν_4] + 15% ν_2(CH str.) + 14% [ν_3 + ν_5]		2785.5	2776.2 (16)
55% ν_2(CH str.) + 18% [ν_3 + ν_4]	2876.6 (99)	2706.5	2703.3 (37)
96% ν_3(HOC def. + HCO def.)	1513.7 (13)	1475.1	1465.5 (10)
92% ν_4(CO str.)	1334.1 (39)	1300.5	1297.1 (53)
97% ν_5(HOC def. – HCO def.)	1220.0 (100)[b]	1183.5	1183.2 (100)
97% ν_6(twist)	1094.3 (82)	1058.9	1048.5 (88)
trans-H–C–OD ([D_1]-1t)			
87% [2 ν_3] + 7% ν_1		2852.6	2841.3 (8)
51% [ν_3 + ν_4]+ 33% ν_1(CH str.) + 5% [2 ν_3]	2878.9 (100)[c]	2729.5	2726.1 (30)
45% ν_1(CH str.) + 37% [ν_3 + ν_4] + 5% ν_2		2682.8	2675.9 (20)
83% ν_2(OD str.)	2739.5 (57)	2626.8	2588.1 (58)
78% [2 ν_4] + 5% ν_4		2566.4	
97% ν_3(HCO def.)	1451.0 (8)	1420.8	1414.7 (4)
92% ν_4(CO str.)	1326.4 (84)	1294.1	1290.8 (100)
98% ν_5(DOC def.)	953.6 (58)	928.7	923.1 (66)
98% ν_6(twist)	933.8 (70)	907.1	901.6 (59)

Table 3.1 Comparison of theoretical [AE-CCSD(T)/cc-pCVQZ] harmonic (ω) and variational anharmonic (ν) vibrational band origins (in cm^{-1}) to experimental bands (Ar matrix, 11 K) for *trans*-hydroxymethylene and its mono-deuterated isotopologue. Relative IR absorption intensities (I_{rel}, in %) from the double-harmonic approximation are included from CCSD(T)/cc-pVTZ computations.

[a] Distribution of the converged vibrational wave functions over the normal mode basis states; all contributions larger than 5% are listed.

[b] Absolute intensity = 142 km mol^{-1}.

[c] Absolute intensity = 116 km mol^{-1}.

nuclear motion computations, the latter utilizing a high-quality quartic vibrational force field.

The remarkable agreement displayed in Table 3.1 between the measured and predicted vibrational band origins (VBOs) provides convincing evidence for the successful preparation of **1**. For 13 of the 15 assigned bands of *trans*-HCOH and HCOD, the (mean, std. dev.) of the residual between theory and experiment is only (6.0, 3.3) cm^{-1}, which is well within the expected

range of matrix shifts. We stress that no empirical adjustments were applied to the variationally computed VBOs. The related harmonic frequencies also displayed in Table 3.1 demonstrate that the full inclusion of vibrational anharmonicity is essential for achieving such a high level of agreement. Particularly impressive is the match for four combination and overtone levels: ν_3 + ν_4 of HCOH, 2 ν_3 of HCOD, and the Fermi resonance pair (ν_3+ ν_4, ν_1) of HCOD. The only notable disparities in Table 3.1 occur for ν_1(OH str.) of **1** and ν_2(OD str.) of [D_1]-**1**, for which the experimental assignments lie 61 and 39 cm^{-1}, respectively, below the theoretical values. Matrix shifts of this magnitude for ground state vibrational fundamentals are not unprecedented.[45]

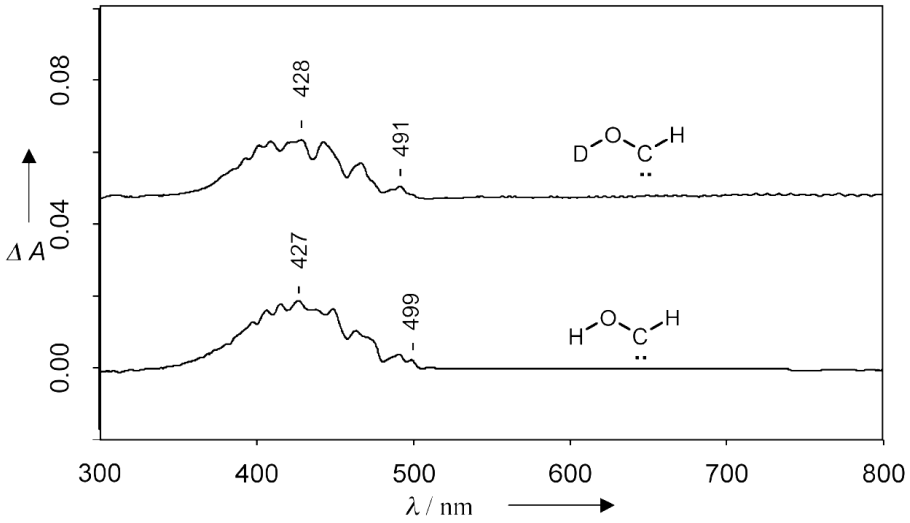

Figure 3.5 UV/Vis spectrum of **1** and D_1-**1**. Difference UV/Vis spectra (unirradiated – irradiated) of **1** and D_1-**1** in an Argon matrix at 11 K.

The recorded UV/Vis spectrum of **1** displays one very weak absorption band between 500 and 380 nm (λ_{max} = 427 nm) with a distinctive vibrational fine structure (Figure 3.3). Irradiation of the matrix with monochromatic light corresponding to this spectral range (435, 470, and 500 nm) causes rapid rearrangement of **1** to **2** and causes partial fragmentation into CO and H_2. Rigorous multireference coupled cluster computations based on the Mk-MRCCSD formalism[46, 47] conclusively show that the signals in Figure 3.2 are attributable to the expected

lowest-lying open-shell singlet excited electronic state (S_1) of carbene **1**. Upon geometry optimization, the S_1 state of **1** relaxes to a nonplanar structure with a widened ∠(H–C–O) = 127.4° and a dihedral angle of 108.4°, consistent with the extensive vibrational progression observed in the electronic absorption spectrum. Our best theoretical vertical and adiabatic excitation energies are 2.99 eV (415 nm) and 2.40 eV (516 nm), respectively, in complete accord with experiment.

To characterize **1** further by theory, we optimized all singlet structures surrounding **1** at the rigorous AE-CCSD(T)/cc-pCVQZ level and pinpointed the energetics to around 0.1 kcal mol^{-1} by means of exhaustive focal-point analyses (FPA)[29]. The potential energy surface[48] surrounding **1** is shown schematically in Figure 3.6. Structure **1t** exhibits a *trans*-planar equilibrium structure with an H–C–O angle of 102.3°. The C–O bond length of 1.311 Å is much shorter than the corresponding 1.427 Å distance in methanol[49], indicative of the expected π-type stabilization of the electron deficient carbene center through the adjacent oxygen *p* lone pair. Our FPA computations place the lowest triplet state of HCOH much higher in energy (T_0 = 28.0 kcal mol^{-1}) than closed-shell singlet 1t. Moreover, **1t** is 4.4 kcal mol^{-1} lower than *cis*-**1** (**1c**), which we do not observe owing to the high barrier (**TS**1, 26.8 kcal mol^{-1}) for internal rotation, and because the CO_2 extrusion process is likely to give **1t** exclusively; similar processes yielding specific conformers have been reported before.[50] The even higher computed barrier (**TS2**, 29.7 kcal mol^{-1}) for the rearrangement of **1t** to **2** emphasizes the theoretical prediction that hydroxymethylene should indeed be observable under matrix isolation conditions.

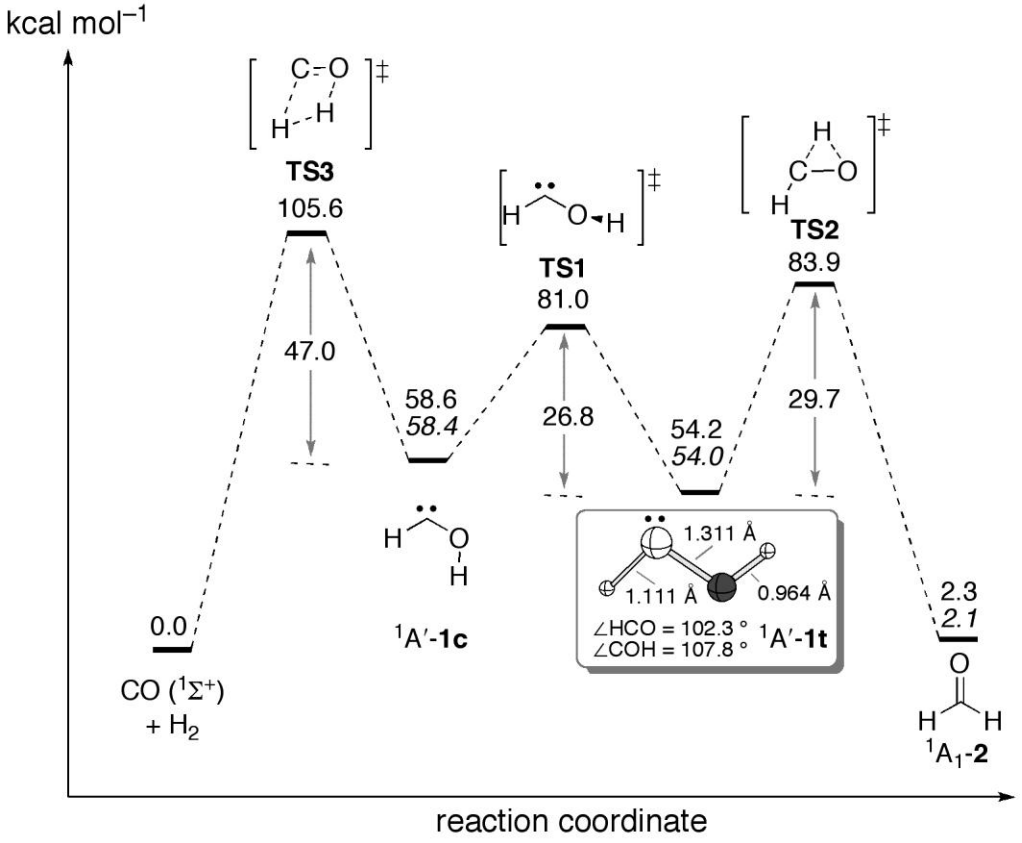

Figure 3.6 Potential energy hyper surface of **1**. Key features of the computed energy diagram [high-level coupled cluster focal point analyses + harmonic zero-point vibrational energies (ZVPE); values in *italics* employ anharmonic ZPVEs] for the rearrangement of singlet **1** (**1c** = *cis* form; **1t** = *trans* form) to formaldehyde and CO + H_2.

Despite the high surrounding enthalpic barriers, matrix-isolated **1** disappears quickly with a half-life ($t_{1/2}$) of ca. 2 h at 11 K in Ar, Kr, and Xe matrices (Table 3.2). However, the mono-deuterated species ([D_1]-**1**) is completely stable under the same conditions. Thermal rearrangement through **TS1** is not conceivable at such low temperatures, and thus other mechanisms must be considered. Owing to the low concentration of **1** and the very limited mobility of molecules in noble-gas matrices, bimolecular reactions are rather unlikely; furthermore, $t_{1/2}$ of **1** is practically unchanged in the 11–20 K temperature interval (Table 3.2). Therefore, quantum mechanical tunneling appears to be the most viable explanation for the rapid

disappearance of *trans*-HCOH together with the persistence of HCOD. Strong tunneling effects in unimolecular reactions on the potential energy surface shown in Figure 3.6 have been noted in a number of previous experimental and theoretical studies.[11, 12]

Temperature/matrix	Ar	Kr	Xe	N_2
11 K	1.8	2.0	2.3	7.0
15 K	1.9	Not determined	Not determined	Not determined
20 K	1.7	2.0	Not determined	Not determined

Table 3.2 Half-life (in h) of matrix isolated 1t in matrices at different temperatures

To gauge the viability of tunneling from **1t** to **2**, we fitted an asymmetric Eckart potential[51, 52] to the **TS2** barrier frequency ($\omega^* = 2174i$ cm^{-1}) as well as the ZPVE-corrected reaction energy (–52.05 kcal mol^{-1}) and barrier height (31.60 kcal mol^{-1}). The vibrational "reaction" mode of 1t that leads toward TS2 has a harmonic frequency of $\omega_0 = 1220$ cm^{-1}. An estimate of the tunneling lifetime of *trans*-H–C–OH is obtained by ascribing a "collision" energy (ε) equivalent to the ZPVE ($\omega_0/2$) of the reaction mode, evaluating the analytic Eckart transmission coefficient at this energy [$\kappa(\varepsilon)$][12, 51], and multiplying by the classical rate (ω_0) at which the reactant hits the barrier. Remarkably, this simple analysis yields a half-life of 24 h for 1t in its ground vibrational level, and scaling the barrier frequency by only 1.06 reproduces the $t_{1/2} \approx 2$ h observed in the Ar and Kr matrices. The same computational procedure estimates a half-life of *trans*-H–C–OD of over 4500 years, consistent with our observations. At the suggestion of referees, a more rigorous quantum tunneling analysis was performed. The intrinsic reaction path (IRP) connecting **TS2** to **1** and **2** was mapped out using the AE-CCSD/cc-CVTZ level of theory. Along this steepest descent route, harmonic vibrational frequencies were calculated and along with a high-quality AE-CCSD(T)/cc-pCVQZ energy points a one-dimensional potential energy curve was computed for the isomerization. The quantum dynamics

were treated with a simple reaction path Hamiltonian model[12, 53] with tunneling probabilities given by the standard WKB (Wentzel-Kramers-Brillouin) formula in terms of barrier penetration integrals, which we computed numerically from our electronic structure data. We then utilized the same procedure as with the Eckart barrier analysis to arrive at a half-life of 2.1 h for the 1t in its ground vibrational state! This result is in amazing agreement with the experimentally observed half-life of ~ 2 h. Furthermore, when this same analysis is applied to trans-HCOD, a half-life of over 1200 years is predicted. Pending arduous full-dimensional quantum dynamics simulations on a semi-global potential energy surface, we conclude that the disappearance of **1t** is attributed to pure quantum mechanical tunneling under a large barrier. The occurrence of such an event near 0 K on a tangible time scale of a few hours is a unique chemical phenomenon with little precedent[54-57].

3.4 Conclusions

These rigorous electronic structure calculations have allowed the highly accurate calculation of vibrational and electronic spectra, both of which are in very strong agreement with spectra taken from recent matrix isolation studies. Data from the experimental and theoretical analysis, when taken together, suggest that the long sought after and highly ephemeral compound, hydroxymethylene has finally been caught. Furthermore, two separate theoretical methods for determining quantum tunneling lifetimes establish the reason this molecule has been so difficult to detect. The WKB approach is in particularly good quantitative agreement with the experimentally observed lifetime of 2 h at 11 K in an argon matrix.

Finally, much previous experimental and theoretical effort has been expended upon this molecule as a potential precursor towards formation of interstellar glycoaldehyde.[58] Both formaldehyde[59] and glycoaldehyde have been observed in the interstellar medium, and it has

been postulated that hydroxycarbene is the missing ingredient (Figure 3.7). Our tunneling analysis would seem to cast a doubt on this proposed mechanism. The short half-life of **1t** toward unimolecular decay and the low collision frequencies in interstellar environments make the detection of non-deuterated **1t** in interstellar space unlikely.

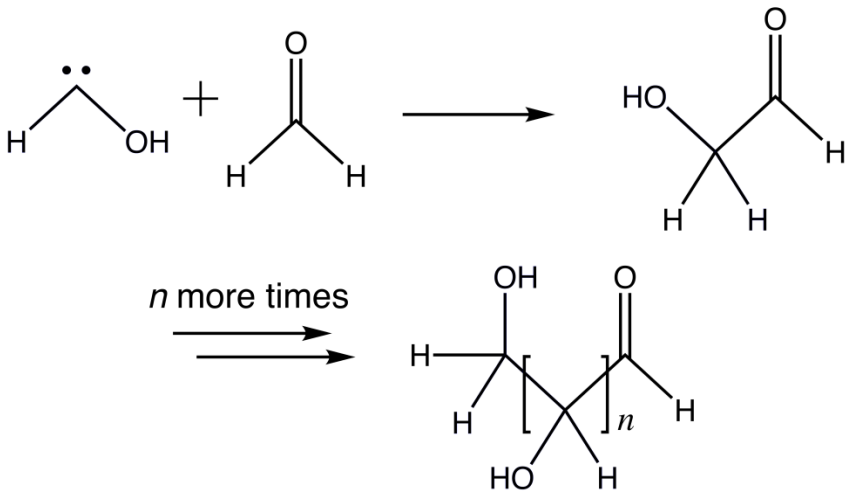

Figure 3.7 Proposed chemical pathway for formation of interstellar glycoaldehyde, and higher order sugars from hydroxymethylene and formaldehyde.

3.5 References

1. D. Bourissou, O. Guerret, F. P. Gabbaï and G. Bertrand, Chem. Rev. **100**, 39-91 (2000).
2. M. J. H. Kemper, J. M. F. Vandijk and H. M. Buck, J. Am. Chem. Soc. **100**, 7841-7846 (1978).
3. E. C. C. Baly, I. M. Heilbron and W. F. Barker, J. Chem. Soc. Faraday T. **119**, 1025-1035 (1921).
4. R. R. Lucchese and H. F. Schaefer, J. Am. Chem. Soc. **100**, 298-299 (1978).
5. M. R. Hoffmann and H. F. Schaefer, Astrophys. J. **249**, 563-565 (1981).
6. D. L. Reid, J. Hernandez-Trujillo and J. Warkentin, J. Phys. Chem. A **104**, 3398-3405 (2000).

7. J. D. Goddard and H. F. Schaefer, J. Chem. Phys. **70**, 5117-5134 (1979).

8. E. O. Fischer and A. Maasbol, Angew. Chem. Int. Ed. Engl. **3**, 580-581 (1964).

9. M. A. Sierra, Chem. Rev. **100**, 3591-3637 (2000).

10. C. F. Pau and W. J. Hehre, J. Phys. Chem. **86**, 1252-1253 (1982).

11. J. R. Sodeau and E. K. C. Lee, Chem. Phys. Lett. **57**, 71-74 (1978).

12. W. H. Miller, J. Am. Chem. Soc. **101**, 6810-6814 (1979).

13. S. P. Nolan, *N-Heterocyclic Carbenes in Synthesis*. (Wiley-VCH, Weinheim, 2006).

14. B. R. Weiner and R. N. Rosenfeld, J. Org. Chem. **48**, 5362-5364 (1983).

15. R. N. Rosenfeld and B. Weiner, J. Am. Chem. Soc. **105**, 3485-3488 (1983).

16. T. H. Dunning, J. Chem. Phys. **90**, 1007-1023 (1989).

17. R. A. Kendall, T. H. Dunning and R. J. Harrison, J. Chem. Phys. **96**, 6796-6806 (1992).

18. D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358-1371 (1993).

19. D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572-4585 (1995).

20. A. K. Wilson, T. van Mourik and T. H. Dunning, J. Mol. Struc. **388**, 339-349 (1996).

21. K. A. Peterson and T. H. Dunning, J. Chem. Phys. **117**, 10548-10560 (2002).

22. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479-483 (1989).

23. J. F. Stanton, Chem. Phys. **281**, 130-134 (1997).

24. Y. J. Bomble, J. F. Stanton, M. Kállay and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

25. M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

26. A. L. L. East and W. D. Allen, J. Chem. Phys. **99**, 4638-4650 (1993).

27. W. D. Allen, A. L. L. East and A. G. Császár, *Structures and Conformations of Non-Rigid Molecules*. (Kluwer, Dordrecht, 1993).

28. A. G. Császár, G. Tarczay, M. L. Leininger, O. L. Polyansky, J. Tennyson and W. D. Allen, in *Spectroscopy from Space*, edited by J. Demaison, K. Sarka and E. A. Cohen (Kluwer, Dordrecht, 2001), Vol. 20, pp. 317.

29. A. G. Császár, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **108**, 9751-9764 (1998).

30. S. A. Perera and R. J. Bartlett, Chem. Phys. Lett. **216**, 606-612 (1993).

31. K. Balasubramanian, *Relativistic Effects in Chemistry: Parts A and B, Theory and Techniques*. (Wiley, New York, 1997).

32. R. D. Cowan and D. C. Griffin, J. Opt. Soc. Am. **66**, 1010-1014 (1976).

33. D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi and H. F. Schaefer, Chem. Phys. **123**, 187-239 (1988).

34. J. K. G. Watson, *Vibrational Spectra and Structure*. (Elsevier, Amsterdam, 1977).

35. I. M. Mills, *Molecular Spectroscopy: Modern Research*. (Academic Press, New York, 1972).

36. D. Papoušek and M. R. Aliev, *Molecular Vibrational-Rotational Spectra*. (Elsevier, Amsterdam, 1982).

37. H. H. Nielsen, Rev. Mod. Phys. **23**, 90-136 (1951).

38. W. D. Allen, Y. Yukio, A. Császár, G., D. A. Clabo, R. B. Remington and H. F. Schaefer, Chem. Phys. **145**, 427-466 (1990).

39. M. S. Schuurman, W. D. Allen and H. F. Schaefer, J. Comput. Chem. **26**, 1106-1112 (2005).

40. J. K. G. Watson, Mol. Phys. **15**, 479-490 (1968).

41. E. Mátyus, G. Czakó, B. T. Sutcliffe and A. G. Császár, J. Chem. Phys. **127**, 084102 (2007).

42. INTDIF2005 is an abstract program written by Wesley D. Allen for MATHEMATICA to perform general numerical differentiations to high orders of electronic structure data.

43. R. L. DeKock, M. J. McGuire, P. Piecuch, W. D. Allen, H. F. Schaefer, K. Kowalski, S. A. Kucharski, M. Musial, A. R. Bonner, S. A. Spronk, D. B. Lawson and S. L. Laursen, J. Phys. Chem. A **108**, 2893-2903 (2004).

44. G. Czakó, T. Furtenbacher, A. G. Császár and V. Szalay, Mol. Phys. **102**, 2411-2423 (2004).

45. M. E. Jacox, Chem. Soc. Rev. **31**, 108-115 (2002).

46. F. A. Evangelista, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **127**, 024102 (2007).

47. F. A. Evangelista, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **125**, 154113 (2006).

48. P. R. Schreiner and H. P. Reisenauer, ChemPhysChem **7**, 880-885 (2006).

49. P. Venkateswarlu and W. Gordy, J. Chem. Phys. **23**, 1200-1202 (1955).

50. H. P. Reisenauer, J. Romanski, G. Mloston and P. R. Schreiner, Eur. J. Org. Chem. **2006**, 4813-4818 (2006).

51. H. S. Johnston, *Gas Phase Reaction Rate Theory*. (Ronald Press Inc., New York, 1966).

52. C. Eckart, Phys. Rev. **35**, 1303-1309 (1930).

53. W. H. Miller, N. C. Handy and J. E. Adams, J. Chem. Phys. **72**, 99-112 (1980).

54. R. J. McMahon and O. L. Chapman, J. Am. Chem. Soc. **109**, 683-692 (1987).

55. P. S. Zuev and R. S. Sheridan, J. Am. Chem. Soc. **116**, 4123-4124 (1994).

56. M. Pettersson, E. M. S. Maçôas, L. Khriachtchev, J. Lundell, R. Fausto and M. Räsänen, J. Chem. Phys. **117**, 9095-9098 (2002).

57. E. M. S. Maçôas, L. Khriachtchev, M. Pettersson, R. Fausto and M. Räsänen, J. Chem. Phys. **121**, 1331-1338 (2004).

58. J. M. Hollis, F. J. Lovas and P. R. Jewell, Astrophys. J. Lett. **540**, L107-L110 (2000).

59. L. E. Snyder, D. Buhl, B. Zuckerman and P. Palmer, Phys. Rev. Lett. **22**, 679-681 (1969).

CHAPTER 4

CHARACTERIZATION OF THE HSIN–HNSI SYSTEM IN ITS ELECTRONIC GROUND STATE

4.1 Introduction

The HSiN-HNSi system is isovalent to the exhaustively studied HCN-HNC system with ten valence electrons but interestingly, exhibits the opposite trend in relative stability of the isomers. The isomers are of astrophysical interest and also may be present in reactions of silane with ammonia in chemical vapor deposition. The HSiN and HNSi isomers have been the subject of experimental[1-8] and theoretical[9-20] studies addressing their formation and existence in interstellar space, and spectroscopic detection in laboratories. Additionally, a number of pioneering studies have focused on characterization of the silicon containing bonds, relative stability, and chemical reactivities.[21-28]

In 1966 Ogilvie and Cradock[1] investigated the infrared (IR) absorption spectra of the silyl and trideuterosilyl azides (SiH_3N_3 and SiD_3N_3) and their photodecomposition products in argon matrices near 4 K. They observed vibrational bands at 3583, 523, and 1198 cm^{-1} for the HNSi isotopologue and 2669, 395, and 1166 cm^{-1} for the DNSi isotopologue. After analyzing the force constants of the three vibrations, Ogilvie and Cradock concluded that HNSi represents the first detected discrete stable molecule with silicon multiply bonded to another atom.

In 1991 Elhanine, Farrenq, and Guelachvili reported the first spectroscopic observation of the HNSi molecule in the gas phase.[2] The fundamental vibration-rotation ν_1 band (NH stretch) of HNSi was observed near 2.7 μm (3584 cm^{-1}) by high resolution Fourier transform spectroscopy

from a mixture of N_2 + SiH_4 excited in a radio frequency discharge. Also in 1991, Bogey, Demuynck, Destombes, and Walters investigated the rotational spectrum of HNSi in the 150-460 GHz frequency range.[3] The observation of nine lines led to the determination of the rotational constant B_0 and the centrifugal distortion constant D_0: B_0=19018.8 MHz and D_0=20.63 KHz. The identification of the HNSi molecule was confirmed by the observation of the ^{29}Si and ^{30}Si isotopologues.

Botschwina *et al.*[4] deduced the equilibrium geometry of HNSi from the vibration-rotation term energies calculated variationally and perturbatively using the CEPA-1 potential and the experimental rotational constants B_0 of HN^{28}Si, HN^{29}Si, and HN^{30}Si. Their equilibrium bond lengths based on the variational rovibrational method are r_e(NH)=1.0005 and r_e(SiN)=1.5482 Å and those based on the rovibrational perturbation theory are r_e(NH)=0.9998 and r_e(SiN)=1.5480 Å.

In 1993 Goldberg, Iraqi, Hrušák, and Schwarz reported the generation and identification of neutral and cationic HNSi by neutralization-reionization mass spectrometry and Gaussian-1 *ab initio* molecular orbital (MO) calculations.[5] Mass spectrometric studies demonstrated that electron bombardment of a mixture of N_2/SiH_3I resulted in the formation of $HNSi^+$ which can be neutralized to the HNSi species. Both theory and experiment pointed to the formation of $HNSi^+(^2\Sigma^+)$/$HNSi(^1\Sigma^+)$ rather than the isomeric forms $HSiN^+(^2\Sigma^+)$/$HSiN$ ($^1\Sigma^+$). In the same year, Elhanine, Hanoune, and Guelachvili observed four hot bands of HNSi, $2\nu_1-\nu_1$, $\nu_1+\nu_3-\nu_3$, $2\nu_1+\nu_3-(\nu_1+\nu_3)$, and $\nu_1+\nu_2-\nu_2$ in emission from a radio frequency excited plasma with a high resolution Fourier transform interferometer.[6] Utilizing the four hot band frequencies, they determined the equilibrium rotational and vibrational parameters of HNSi for the first time.

Maier and Glatthaar reported a matrix isolation study of HSiN and its adduct with hydrogen in 1994.[7] For the HSiN isomer, IR absorptions were observed at 2151, 2149 (SiH stretching) and 1163, 1161 (SiN stretching) cm^{-1}, whereas for the HNSi isomer IR absorptions were observed at 3585, 3580 (NH stretching), 1202, 1200 (SiN stretching) and 522 (bending) cm^{-1}. They stated that the HSiN molecule is the first compound with a formal Si≡N triple bond of the nitrile type to be identified spectroscopically. Additional structural evidence for HSiN was derived from the fact that it loses a hydrogen atom on prolonged irradiation at 193 nm, and is transformed into the well-known SiN radical.

In the earlier theoretical *ab initio* studies carried out by two independent groups[9, 10] it was predicted that HSiN should be considerably less stable than HNSi, in contrast to the situation for the isovalent HCN-HNC system, where HCN is the more stable isomer. Luke *et al.* reported a theoretical survey of unsaturated or multiply bonded and divalent silicon compounds using *ab initio* HF and MP4SDTQ methods.[21] At the MP4SDTQ/6-31G* level of theory they predicted that HCN is 16.8 kcal mol^{-1} lower in energy than HNC, whereas HSiN with a lone pair on nitrogen is 55.0 kcal mol^{-1} higher in energy than the carbene-like HNSi. Their rationalization was based on the argument that silicon prefers to have nonbonding electrons in atomic orbitals with a high percentage of s-character.[23-25]

In 1995 Apeloig and Albrecht theoretically studied the relative stabilities and energy barriers separating silanitriles (RSiN) and silaisonitriles (RNSi).[11] They found that in general, substituents which are more electropositive than Si increase the energy differences between RSiN and RNSi (relative to R=H), while the more electronegative substituents decrease the energy gap. The activation energies for the forward isomerization (RSiN → RNSi) are relatively

small for R=H (~11 kcal mol^{-1}), however when R=F and OH the forward barrier heights grow as RSiN is preferentially stabilized relative to RNSi (49.3 and 35.8 kcal mol^{-1} respectively).

In 1996 and 1997 Parisel, Hanus, and Ellinger published a series of theoretical papers on interstellar silicon-nitrogen chemistry.[12, 13] In the first part of the series they reported the microwave and infrared signatures of the HSiN, HNSi, $HSiNH_2$, $HNSiH_2$, and $HSiNH^+$ species. A number of comparisons with the available rotational and vibrational experimental spectra led to the determination of accurate scaling factors used to calibrate the original *ab initio* results. They pointed out that non-dynamic (static) correlation effects are of particular importance in investigating the HSiN isomer.

In 2003 Hu, Wang, Wang, Chu, and Liu reported a theoretical study on gas-phase reactions between silane (SiH_4) and ammonia (NH_3) using *ab initio* methods at the CCSD(T)/6-311++G**//MP2/6-31+G* level of theory.[14] Within a 180 kcal mol^{-1} energy range, they located 34 equilibria and 23 transition states for various products of the two molecules. In their study, the HSiN isomer was found to be 67.5 kcal mol^{-1} [64.7 kcal mol^{-1} with the zero-point vibrational energy (ZPVE) corrections] higher in energy than the HNSi isomer. The barrier height for the forward isomerization reaction (HSiN $\rightarrow$ HNSi) was predicted to be 11.1 kcal mol^{-1} (10.0 kcal mol^{-1} with the ZPVE corrections).

In the present research, we bring state of the art electronic structure theory methods to bear on this challenging problem. Additionally, the effects of anharmonicity are modeled and incorporated into our predictions. We expect these computed energetic and physical properties to be the most extensive and highly accurate to date.

4.2 Electronic Structure Considerations

The electronic structure of the linear $\tilde{X}\,^1\Sigma^+$ HSiN molecule may be described as:

$$[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi)^4 \quad \tilde{X}\,^1\Sigma^+, \tag{4.1}$$

where [core] denotes the six core orbitals (Si: 1s-, 2s-, 2p-like; N:1s-like). Here the 5σ and 6σ orbitals describe the SiN and SiH σ bonds, respectively. The 7σ orbital is related to the lone pair (non-bonding) orbital on the N atom. The 2π orbital has a SiN π bonding nature. The two highest-lying occupied MOs (7σ and 2π) and the 3π lowest unoccupied orbital of HSiN at the CASSCF/cc-pVTZ level of theory are depicted in Figure 4.1. The electronic structure of the linear $\tilde{X}\,^1\Sigma^+$ HNSi isomer may likewise be expressed as:

$$[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi)^4 \quad \tilde{X}\,^1\Sigma^+. \tag{4.2}$$

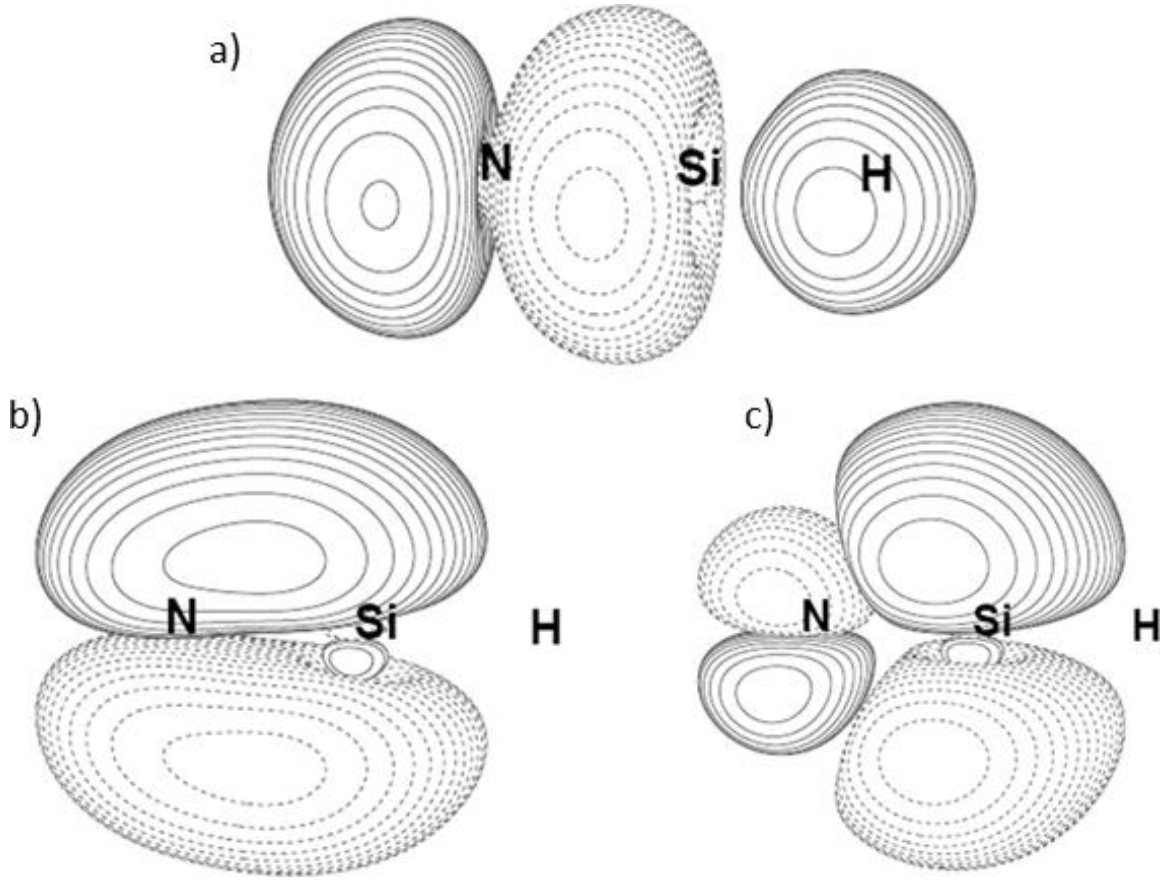

Figure 4.1 The 7σ (a), 2π (b), and 3π (c) molecular orbitals of HSiN at the CASSCF/cc-pVTZ level of theory.

The 5σ and 6σ orbitals are associated with the NH and SiN σ bonds. The 7σ orbital is related to the lone pair (non-bonding) orbital on the Si atom. Similar to HSiN, the 2π orbital of HNSi has SiN π bonding character as well. The two highest-lying occupied molecular orbitals (7σ and 2π) of HNSi at the CASSCF/cc-pVTZ level of theory are depicted in Figure 4.2. For reference, the 3π LUMO is also included in Figure 4.2. The electronic structure at the isomerization reaction transition state (**TS**) may be written as:

$$[\text{core}](6a')^2(7a')^2(8a')^2(2a'')^2(9a')^2 \quad \tilde{X}\,^1A'. \tag{4.3}$$

For bent configurations, the in-plane π orbital (of the linear configuration) will interact with the σ orbitals and form C_s symmetry admixtures of the σ and π bonds.

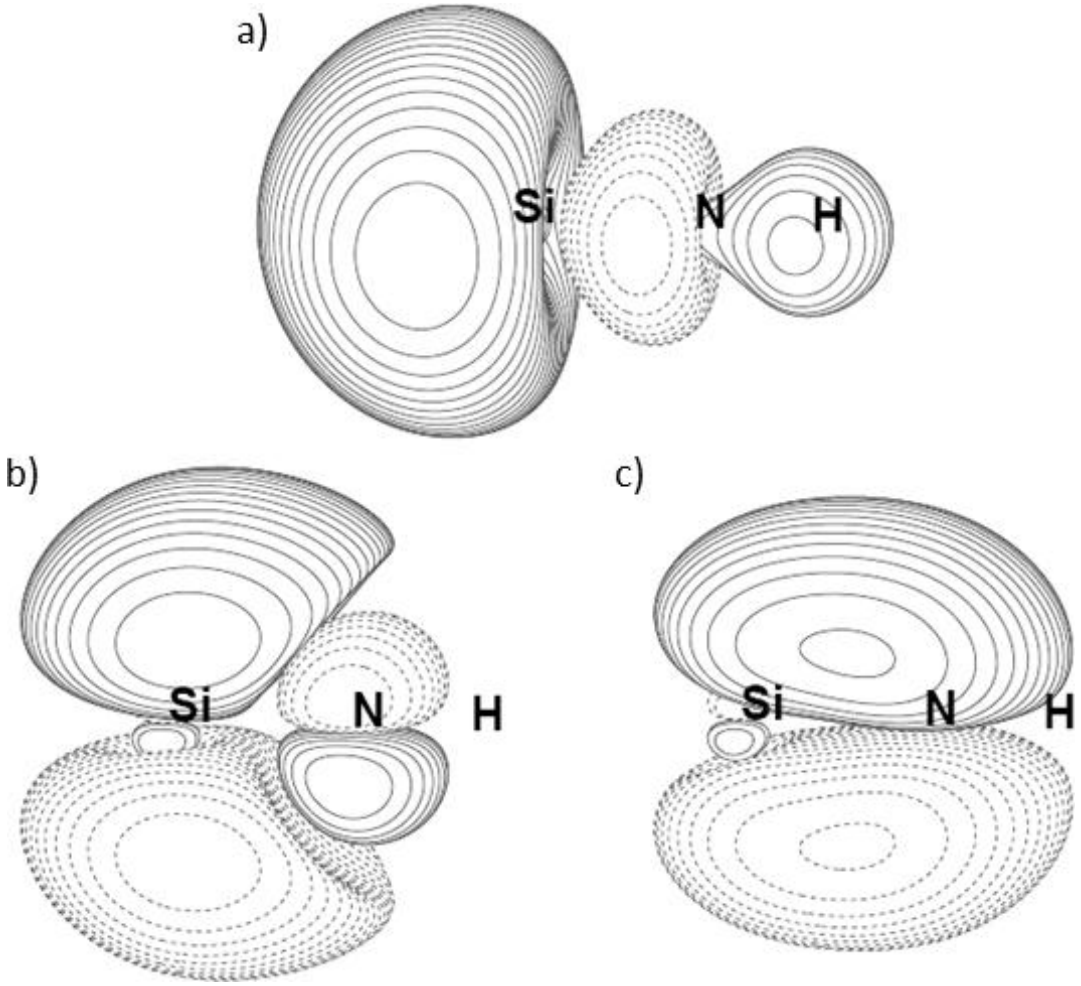

Figure 4.2 The 7σ (a), 2π (b), and 3π (c) molecular orbitals of HNSi at the CASSCF/cc-pVTZ level of theory.

4.3 Methods

The correlation-consistent family of basis sets (aug)-cc-pV*n*Z [*n* = D, T, Q, 5, 6] and (aug)-cc-pCV*n*Z [*n* = T, Q] developed by Dunning and coworkers[29-33] were used for all electronic structure calculations in the present work. Zeroth-order descriptions of all stationary points were obtained using restricted Hartree-Fock (RHF) self-consistent field theory (SCF). Non-dynamic (static) correlation effects were included using complete active space SCF (CASSCF) method,[34-36] while dynamic correlation effects were accounted for using configuration interaction with single and double excitations (CISD), internally contracted multireference configuration interaction (ICMRCI) methods,[37-39] and the coupled cluster hierarchy of methods, including single, double, triple and quadruple excitations, as well as a perturbative treatment of connected pentuple excitations [CCSD,[40, 41] CCSD(T),[42, 43] CCSDT,[44] CCSDT(Q),[45, 46] CCSDTQ,[47, 48] and CCSDTQ(P)[49]]. A full valence (10 e^-/9 MO) active space was chosen for the CASSCF wavefunctions. The CASSCF wavefunctions consist of 1436 configuration state functions (CSFs) for the linear HSiN and HNSi isomers (in C_{2v} point group symmetry) and 2744 CSFs for the bent isomerization reaction transition state (in C_s point group symmetry). The correlated wavefunctions were constructed by freezing the six core orbitals (Si: 1s, 2s, 2p-like; N: 1s-like) for the (aug)-cc-pVXZ basis sets, and freezing only the one core orbital (Si: 1s-like) for the (aug)-cc-pCVYZ basis sets. The structures of the three stationary points were initially optimized at the SCF and CASSCF levels using analytic derivative methods.[50-52] Harmonic vibrational frequencies at the SCF and CASSCF levels were evaluated analytically. The CISD, CCSD, CCSD(T), and ICMRCI geometries and harmonic vibrational frequencies were determined via numerical differentiation of the total energies. The dipole

moments were evaluated by numerical differentiation of the total energies with respect to finite external electric fields.

To obtain highly accurate energetics for the three stationary points along the reaction coordinate, the FPA of Allen *et al.* was employed.[53-57] Like other highly accurate model chemistry approaches such as W4[58] or HEAT,[59] the FPA seeks to systematically eliminate errors associated with deficiencies to both the one-particle basis set and the correlation treatment. Unlike other composite approaches, the FPA treatment for a given system is not explicitly prescribed; rather, the chemical physicist monitors the contributions to both the complete basis set (CBS) and configuration interaction (FCI) limits. The specific FPA is then tailored to the needs of the chemical system of interest given available computational resources and desired accuracy. Other auxiliary corrections are often incorporated into FPA calculations. These account for ZPVE, core correlation, non-Born-Oppenheimer effects and relativistic effects.

FPA energetics were computed here using equilibrium geometries optimized at the all-electron (AE) CCSD(T)/cc-pCVQZ level of theory. The CBS limit was estimated using established basis set extrapolation formulae. [60, 61] The SCF and correlation contributions to the absolute energies of each molecule are extrapolated separately. Higher order correlation effects [through CCSDTQ(P)] were incorporated into the final FPA energy additively. That is, the effects of quadruple excitations are incorporated into the FPA as the difference between CCSDT(Q)/cc-pVTZ and CCSD(T)/cc-pVTZ, while the effects of pentuple excitations are incorporated by the difference between CCSDTQ(P)/cc-pVDZ and CCSDTQ(P)/cc-pVDZ. The inclusion of higher order excitations has been shown to be vital for capturing the effects of static correlation when using single reference theories upon multi-reference systems.[58, 59]

Finally, to approach the realm of sub-chemical accuracy, a suite of auxiliary corrections was calculated. The effects of core electron correlation were incorporated into the valence FPA results by taking the difference of all-electron and frozen-core CCSD(T)/cc-pCVQZ computations. To correct for the Born-Oppenheimer approximation, where the electronic and nuclear motion are uncoupled, the diagonal Born-Oppenheimer correction (DBOC) was evaluated.[62, 63] As per the HEAT protocol,[59] the DBOC was calculated at the HF/aug-cc-pVTZ level of theory. Special relativistic effects were also accounted for at first order via the calculation of mass-velocity and one-electron Darwin terms.[64, 65] These corrections were computed at the AE-CCSD(T)/cc-pCVQZ level of theory.

All three structures along the reaction coordinate were confirmed to be stationary points by the evaluation of harmonic vibrational frequencies. Furthermore, quartic force fields were computed for both the HSiN and HNSi species. This was done to explicitly quantify the effects of anharmonicity on the fundamental vibrational frequencies and other spectroscopic properties. The force fields were computed via finite differences of energies at the AE-CCSD(T)/cc-pCVQZ level of theory. To maintain the numerical integrity of the differentiations, energies computed at displaced geometries were very tightly converged (10^{-12} E_h). VPT2[66-72] was applied to these force fields to compute anharmonic properties.

The majority of electronic structure computations were carried out using the 2002 and 2006 versions of MOLPRO.[73] Other electronic structure calculations were also carried out using the PSI2[74] and ACESII[75, 76] (Mainz-Austin-Budapest version) suites of quantum chemistry programs. Higher order CC calculations were performed with the MRCC code of Kállay[46, 77, 78] as interfaced to ACESII. The Mathematica program INTDIF2005[79] was used to generate displaced geometries and to compute force constants in internal coordinates. The

INTDER2005[80, 81] code of Allen was used to perform non-linear coordinate transformations of force constants between internal and Cartesian coordinates. VPT2 analysis was then performed upon the Cartesian force constants by the ANHARM[82] program.

	HSiN		HNSi		TS			
Method	r_e(HSi)	r_e(SiN)	r_e(HN)	r_e(NSi)	r_e(NSi)	r_e(HSi)	r_e(HN)	θ(NSiH)
CCSD(T) /								
cc-pVQZ	1.4849	1.5826	0.9999	1.5575	1.6198	1.5151	2.0999	84.04
aug-cc-pVQZ	1.4849	1.5831	1.0004	1.5587	1.6204	1.5161	2.1008	84.03
cc-pV5Z	1.4841	1.5790	0.9999	1.5543	1.6156	1.5138	2.0996	84.21
aug-cc-pV5Z	1.4842	1.5792	1.0001	1.5547	1.6159	1.5142	2.1001	84.21
cc-pCVQZ	1.4810	1.5744	0.9989	1.5495	1.6109	1.5088	2.1018	84.64
aug-cc-pCVQZ	1.4810	1.5749	0.9994	1.5508	1.6116	1.5096	2.1026	84.63
B3LYP /								
6-311G+(d,p)[a]	1.482	1.565	1.003	1.559				
CEPA[b]			0.99924	1.55095				
Expt.[b]			0.99983	1.54803				

Table 4.1 Predicted CCSD(T) geometries for the three stationary points on the ground state PES of the HSiN–HNSi system. Bond lengths in angstroms and bond angles are in degrees.
[a]Ref 8
[b]Ref 4

4.4 Results and Discussion

4.4.1 CASSCF Wavefunctions

It is beneficial to analyze the CASSCF wavefunctions for the three stationary structures along the reaction coordinate. The dominant CI determinants of the CASSCF/cc-pVTZ wavefunction for the linear HSiN isomer at its equilibrium geometry are:

$$\begin{aligned}
&+0.937[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^2(2\pi_y)^2\\
&-0.134[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_y)^2(3\pi_x)^2\\
&-0.134[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^2(3\pi_y)^2\\
&-0.087[\text{core}]\ (5\sigma)^2(6\sigma)^2(8\sigma)^2(2\pi_x)^2(2\pi_y)^2\\
&+0.081[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^{\alpha}(2\pi_y)^{\beta}(3\pi_x)^{\beta}(3\pi_y)^{\alpha}\\
&+0.081[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^{\beta}(2\pi_y)^{\alpha}(3\pi_x)^{\alpha}(3\pi_y)^{\beta}. \qquad (4.4)
\end{aligned}$$

The HSiN CASSCF wavefunction involves major contributions from the $2\pi^2 \rightarrow 3\pi^2$ and $7\sigma^2 \rightarrow 8\sigma^2$ double excitations. The electron occupation numbers for the valence CASSCF natural orbitals (NOs) are $n(5\sigma)$=1.997, $n(6\sigma)$=1.983, $n(7\sigma)$=1.936, $n(8\sigma)$=0.065 and $n(9\sigma)$=0.013 for the σ orbitals, $n(2\pi)$=1.911 and $n(3\pi)$=0.092 for the π orbitals. The numbers of electrons excited from the occupied MOs (in terms of RHF) to the virtual MOs are 0.078 for the σ space and 0.184 for the π space. The CI vector shows the reference determinant is less dominant than for the HNSi wavefunction (shown below), indicative of a more significantly multireference system. This is corroborated by the T_1 diagnostic[83] with a value of 0.031.

The dominant CI determinants of the CASSCF/cc-pVTZ wavefunction in terms of the CASSCF (NOs) for the linear HNSi isomer at the equilibrium geometry are:

$$
\begin{aligned}
&+0.964[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^2(2\pi_y)^2 \\
&-0.104[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_y)^2(3\pi_x)^2 \\
&-0.104[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^2(3\pi_y)^2 \\
&+0.069[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^{\alpha}(2\pi_y)^{\beta}(3\pi_x)^{\beta}(3\pi_y)^{\alpha} \\
&+0.069[\text{core}](5\sigma)^2(6\sigma)^2(7\sigma)^2(2\pi_x)^{\beta}\,(2\pi_y)^{\alpha}(3\pi_x)^{\alpha}(3\pi_y)^{\beta} \\
&-0.052[\text{core}](5\sigma)^2(6\sigma)^2(9\sigma)^2(2\pi_x)^2(2\pi_y)^2, \qquad (4.5)
\end{aligned}
$$

where π_x and π_y are the π orbitals perpendicular to the molecular axis (z axis). The CASSCF wavefunction involves the contributions from the $2\pi^2 \rightarrow 3\pi^2$ and $7\sigma^2 \rightarrow 9\sigma^2$ double excitations. The electron occupation numbers for the valence CASSCF NOs are $n(5\sigma)$=1.997, $n(6\sigma)$=1.981, $n(7\sigma)$=1.976, $n(8\sigma)$=0.023 and $n(9\sigma)$=0.022 for the σ orbitals, $n(2\pi)$=1.949 and $n(3\pi)$=0.052 for the π orbitals. The numbers of electrons shifted from the occupied MOs (in terms of RHF) to the virtual MOs are 0.045 for the σ space and 0.104 for the π space. The CI vector indicates this wavefunction is predominantly composed of a single reference configuration.

The CASSCF/cc-pVTZ wavefunction for **TS** at the optimized geometry consists of the following dominant determinants:

$$+0.926[\text{core}](6a')^2(7a')^2(8a')^2(9a')^2(2a'')^2$$
$$-0.181[\text{core}](6a')^2(7a')^2(8a')^2(10a')^2(2a'')^2$$
$$-0.135[\text{core}](6a')^2(7a')^2(8a')^2(9a')^2(3a'')^2$$
$$-0.094[\text{core}](6a')^2(7a')^2(8a')^2(9a')^{\alpha}(10a')^{\beta}(2a'')^{\beta}(3a'')^{\alpha}$$
$$-0.094[\text{core}](6a')^2(7a')^2(8a')^2(9a')^{\beta}(10a')^{\alpha}(2a'')^{\alpha}(3a'')^{\beta}$$
$$-0.061[\text{core}](6a')^2(7a')^2(9a')^2(10a')^2(2a'')^2. \qquad (4.6)$$

The CASSCF wavefunction involves important contributions from $2a''^2 \rightarrow 3a''^2$, $9a'^2 \rightarrow 10a'^2$, and $8a'^2 \rightarrow 10a'^2$ types of double excitations. The electron occupation numbers for the valence CASSCF NOs are $n(6a')=1.990$, $n(7a')=1.977$, $n(8a')=1.952$, $n(9a')=1.866$, $n(10a')=0.150$, $n(11a')=0.036$, $n(12a')=0.025$ for the a′ orbitals and $n(2a'')=1.908$ and $n(3a'')=0.095$ for the a″ orbitals. The numbers of electrons shifted from the occupied MOs (in terms of RHF) to the virtual MOs are 0.211 for the a′ space and 0.095 for the a″ space. As is the case with many transition state structures, the wavefunction is more multireference than those of the corresponding reactants and products. Both the CI vector and the T_1 diagnostic (T_1=0.085) suggest a sophisticated treatment of electron correlation will be crucial to capture the static electron correlation effects.

4.4.2 Geometries

Optimized structures for the three ground state stationary points at the most rigorous levels of theory are presented in Table 4.1. For the HSiN isomer, increasingly sophisticated correlation treatments lengthen both the SiH and SiN bond distances; between the SCF and CCSD(T) methods with the aug-cc-pV5Z basis, these elongations are 0.025 and 0.059 Å, respectively. The core-valence basis sets (cc-pCV*n*Z and aug-cc-pCV*n*Z) provide shorter SiH and SiN bond distances (by 0.004 and 0.014 Å, respectively) compared to the corresponding polarized-valence basis sets (cc-pV*n*Z and aug-cc-pV*n*Z). We observe that the SiN bond in the HSiN isomer is more sensitive to correlation effects than the SiN bond in the HNSi isomer. This feature may be associated with the fact that the CI coefficient (0.937) of the reference

configuration for the HSiN isomer is smaller than the corresponding CI coefficient (0.964) for the HNSi isomer. Similarly, the CI coefficients (0.134) of the 2π (Figure 4.1b) → 3π (Figure 4.1c) double excitations for the HSiN isomer in eq. 4.4 are larger than those (0.104) for the 2π (Figure 4.2b) → 3π (Figure 4.2c) double excitations for the HNSi isomer in eq. 4.5. As mentioned in section 4.5.1, the numbers of electrons shifted from the occupied MOs to the virtual MOs are significantly larger for the HSiN isomer ($0.26e^-$) than the HNSi isomer ($0.15e^-$). At the CCSD(T) level of theory, the predicted structure of the HSiN isomer is r_e(SiH)=1.484 and r_e(SiN)=1.579 Å with the aug-cc-pV5Z basis set, and r_e(SiH)=1.481 and r_e(SiN)=1.575 Å with the aug-cc-pCVQZ basis set.

For the HNSi isomer, an increasingly robust correlation treatment results in bond length elongation. The increase in the NH bond length between the HF and CCSD(T) levels of theory is 0.016 Å and that of the SiN bond length is 0.034 Å using the aug-cc-pV5Z basis set. Within a given level of correlation, basis sets augmented with diffuse functions (aug-cc-*pVnZ* and aug-cc-pCVYZ) consistently predict slightly longer r_e(NH) and r_e(SiN) values, than their non-augmented counterparts (cc-pV*n*Z and cc-pCV*n*Z). The inclusion of core correlation with tight core functions found in the core-valence basis sets (cc-pCV*n*Z and aug-cc-pCV*n*Z) predict shorter r_e(NH) and r_e(SiN) bond distances relative to the analogous polarized-valence basis sets (cc-pV*n*Z and aug-cc-pV*n*Z). It is evident that the HF wavefunctions overestimate the multiple bonding character of the SiN bond, while the CASSCF wavefunctions overestimate the NH bond length. The ICMRCI and CCSD(T) methods produce very similar geometries. The larger elongation of the SiN bond distance due to correlation effects may be attributed to the double excitations from the bonding 2π orbital (Figure 4.2b) to the anti-bonding 3π orbital (Figure 4.2c) as demonstrated in eq. 4.5. The CCSD(T) level of theory predicts the geometry of the HNSi

isomer to be r_e(NH)=1.000 and r_e(SiN)=1.555 Å with the aug-cc-pV5Z basis set, and r_e(NH)=0.999 and r_e(SiN)=1.551 Å with the aug-cc-pCVQZ basis set. These structures are in good agreement with the equilibrium geometry deduced from a combination of CEPA potential and experimental B_0 values by Botschwina *et al.*;[4] r_e(NH)=1.0005 and r_e(SiN)=1.5482 Å (based on variational methods), and r_e(NH)=0.9998 and r_e(SiN)=1.5480 Å (based on perturbation methods). CCSD(T) also predicts the bond distance of the diatomic $X\ ^2\Sigma^+$ SiN molecule to be 1.579 Å with the aug-cc-pV5Z basis set and 1.575 Å with the aug-cc-pCVQZ basis set. Therefore, the SiN bond length of the HNSi isomer is about 0.024 Å shorter than the diatomic SiN, whereas the SiN bond length of the HSiN isomer is similar to that of the diatomic SiN. These observations are in accord with Kuzelnigg's argument[24] that second row atoms do not favor the formation of multiple bonds compared to the analogous first row atoms. According to the bond order analysis based on the vibrational force constants by Maier and Glathaar,[7] the SiN bond order is 2.3 for HNSi and 2.0 for HSiN. Our theoretical SiN bond distances for the two isomers appear to be consistent with their analysis.

For the transition state, at the SCF level, the HSiN bond angle is predicted to be ~140˚. However with the inclusion of correlation, this bond angle decreases dramatically to 92-84˚. The SiN bond distance increases with a more sophisticated treatment of correlation effects, and it is longer than the SiN distances of the HNSi and HSiN isomers at all levels of theory. This phenomenon may be explained by the double excitations in eq. 4.6 which have a strong tendency to weaken the SiN multiple bond. In fact, the numbers of electrons shifted from the occupied MOs to the virtual MOs for the transition state (0.31 e^-) is the largest among the three stationary points. The transition state structures predicted from the ICMRCI method are similar to those from the CCSD(T) method.

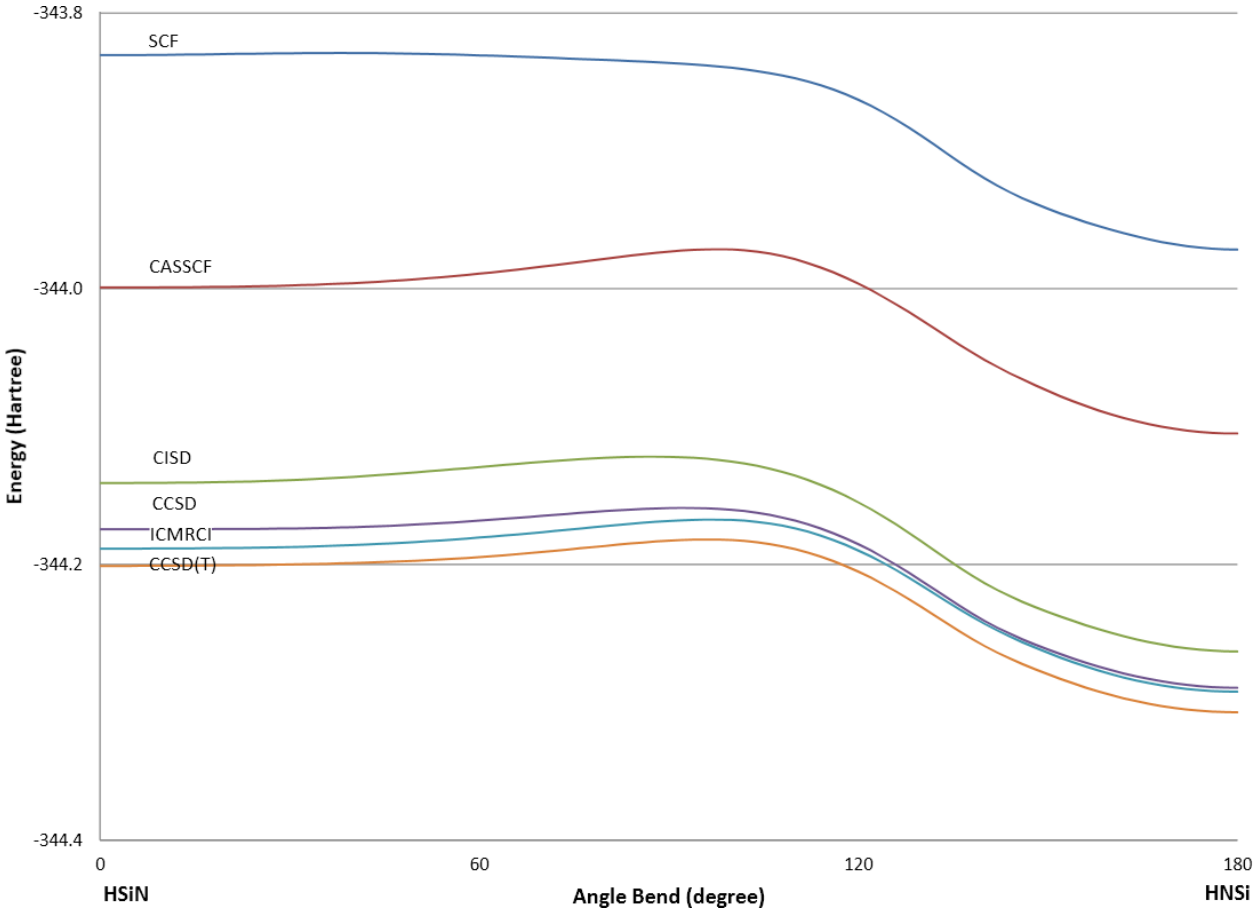

Figure 4.3 Potential energy surfaces for the HSiN-HNSi system at six levels of theory with the cc-pVTZ basis set.

4.4.3 Potential Energy Surface

Six different one-dimensional representations of the PES for the HSiN-HNSi system are depicted in Figure 4.3. Each PES was computed as a series of 91 constrained geometry optimizations, where the H-N-Si angle was held fixed at two degree increments from 0° to 180°, and the two bond lengths were optimized. Clearly the linear HNSi isomer is the global minimum on the ground state PES for all levels of theory. The **TS** structure is energetically similar to the HSiN isomer with the HF method. However, inclusion of correlation effects dramatically shifts the transition state toward the HNSi isomer and decreases the barrier height for the reverse isomerization reaction (HNSi $\rightarrow$ HSiN). The barrier height for the forward isomerization

reaction (HSiN → HNSi) is small (~1 kcal mol^{-1}) with the HF method, whereas it is considerable (~10 kcal mol^{-1}) with the correlated methods. The CCSD, CCSD(T), and ICMRCI methods all present qualitatively similar potential energy curves. Figure 4.3 illustrates the extreme flatness of the PES near the HSiN equilibrium structure, the consequences of which shall be discussed later.

4.4.4 Energetics

In agreement with previous *ab initio*[9-11, 13-18] and DFT[8, 15, 19] studies, we confirmed the HNSi isomer as the global minimum on the ground state potential energy surface. The relative energies and isomerization barriers for our most rigorous levels of theory [CCSD(T), ICMRCI and FPA] are presented in Table 4.2. The valence focal point energies are well converged with respect to both basis set and correlation treatment. We expect these predicted values to be accurate to beyond chemical accuracy. The forward isomerization reaction (HSiN → HNSi) is an exothermic process and the transition state structure is close to the HSiN (reactant) isomer. This feature appears to follow Hammond's postulate,[84] which states that the transition state structure resembles the reactant for an exothermic reaction and the transition state geometry is close to that of the product for an endothermic reaction.

The classical energy separation between the HNSi and HSiN isomers is computed to be 89.1(SCF), 67.4(CASSCF), 77.9(CISD), 73.5(CCSD), 68.0 [CCSD(T)] and 66.4 (ICMRCI) kcal mol^{-1} using the aug-cc-pV5Z basis set. The inclusion of correlation effects preferentially stabilizes the higher energy HSiN isomer relative to the HNSi isomer by as much as ~21 kcal mol^{-1}. This is likely due to the greater multireference character of the HSiN isomer in eq. 4.4 compared to the HNSi isomer in eq. 4.5. Furthermore, the diatomic $X\ ^2\Sigma^+$ SiN molecule has a large dipole moment (2.63 Debye) with the diretion of $^+SiN^-$. The attachment of the

electropositive H atom to the Si end (to form the HSiN isomer) is much less preferable to the attachment to the N end (to form the HNSi isomer). The inclusion of ZPVE corrections changes the predicted energy separations significantly, though not qualitatively. The ZPVE corrected splittings are 64.7 kcal mol^{-1} (22620 cm^{-1}, 2.80 eV) with the CCSD(T)/aug-cc-pV5Z method and 63.0 kcal mol^{-1} (22030 cm^{-1}, 2.73 eV) with the ICMRCI/aug-cc-pV5Z method (with a Davidson correction).[85] The FPA, which includes ZPVE corrections, predicts the HSiN isomer to lie 64.7 kcal mol^{-1} (22640 cm^{-1}, 2.81 eV) above the HNSi isomer on the PES. The remarkable agreement between the CCSD(T) and FPA results stems from error cancellation, as the energy contribution from the inclusion of core correlation negates the energy contribution from higher excitations and auxiliary corrections.

	HSiN – HNSi		**TS** – HSiN		**TS** – HNSi	
Method	ΔE_e	ΔE_0	$\Delta E_e^{\neq}$	$\Delta E_0^{\neq}$	$\Delta E_e^{\neq}$	$\Delta E_0^{\neq}$
CCSD(T) /						
cc-pVQZ	67.61	64.21	11.48	10.77	79.09	74.98
aug-cc-pVQZ	67.77	64.42	11.34	10.59	79.11	75.01
cc-pV5Z	67.98	64.61	11.30	10.55	79.28	75.16
aug-cc-pV5Z	68.01	64.66	11.26	10.51	79.27	75.17
cc-pCVTZ	67.73	64.28	11.51	10.81	79.24	75.09
aug-cc-pCVTZ	68.08	64.78	11.02	10.26	79.10	75.04
cc-pCVQZ	68.63	65.18	10.99	10.32	79.62	75.50
aug-cc-pCVQZ	68.81	65.41	10.83	10.13	79.64	75.54
ICMRCI /						
cc-pVQZ	66.03	62.58	11.97	11.30	78.00	73.88
aug-cc-pVQZ	66.16	62.77	12.64	12.01	78.80	74.78
cc-pV5Z	66.34	62.93	11.84	11.13	78.18	74.06
aug-cc-pV5Z	66.37	62.98	12.54	11.90	78.91	74.88
FPA	67.33	64.74	10.82	9.65	78.15	74.39

Table 4.2 Relative energies of the ground state PES of the HSiN–HNSi system using the CCSD(T), ICMRCI and FPA (kcal mol^{-1}).

With the SCF method, the vibrationless barrier for the forward isomerization (HSiN → HNSi) reaction is computed to be only about 1 kcal mol^{-1}. However, inclusion of correlation effects increases this barrier height by as much as ~10 kcal mol^{-1}. The transition state wavefunction in eq. 4.6 displays stronger mixing with excited configurations than that of the HNSi isomer in eq. 4.5. The bare electronic and ZPVE corrected barrier heights for the forward isomerization reaction are 11.3 kcal mol^{-1} (3940 cm^{-1}, 0.49 eV) and 10.5 kcal mol^{-1} (3680 cm^{-1}, 0.46 eV) with the CCSD(T)/aug-cc-pV5Z method, 12.5 kcal mol^{-1} (4390 cm^{-1}, 0.54 eV) and 11.9 kcal mol^{-1} (4160 cm^{-1}, 0.52 eV) with the ICMRCI/aug-cc-pV5Z method. The FPA predicts the forward isomerization barrier to be 9.7 kcal mol^{-1} (3375 cm^{-1}, 0.42 eV).

The classical barrier height for the reverse isomerization reaction (HNSi → HSiN) is determined to be 90.1(SCF), 84.2(CASSCF), 89.8(CISD), 82.8(CCSD), 79.3 [CCSD(T)] and 78.9 (ICMRCI) kcal mol^{-1} with the aug-cc-pV5Z basis set. Correlation effects stabilize **TS** relative to the linear HNSi isomer by as much as 11 kcal mol^{-1}. The barrier height for the reverse isomerization reaction (HNSi → HSiN) is predicted to be 75.2 kcal mol^{-1} (26290 cm^{-1}, 3.26 eV) with the ZPVE corrections using the CCSD(T)/aug-cc-pV5Z method and 74.9 kcal mol^{-1} (26050 cm^{-1}, 3.23 eV) with the ICMRCI/aug-cc-pV5Z method. The FPA method predicts the barrier for the reverse isomerization process to be 74.4 kcal mol^{-1} (26020 cm^{-1}, 3.23 eV). Once the global minimum HNSi isomer is formed, the isomerization reaction to the HSiN isomer is unlikely to occur at low temperatures.

4.4.5 Dipole moments

The dipole moment for the higher energy HSiN isomer (Table 4.3) decreases with advanced treatments of correlation effects. At a given level of theory the dipole moment is predicted to be slightly larger with the augmented (aug-cc-pV*n*Z and aug-cc-pCV*n*Z) basis sets

compared to the standard (cc-pV*n*Z and cc-pCV*n*Z) basis sets. At the CCSD(T) level of theory the dipole moment for the HSiN isomer is predicted to be 4.37 debye with the aug-cc-pV5Z basis set and 4.36 debye using the aug-cc-pCVQZ basis set, with the direction being $^{+}$HSiN^{-}. At the CCSD(T) level of theory the dipole moment of the diatomic $X\ ^2\Sigma^+$ SiN is computed to be 2.63 debye with the aug-cc-pV5Z basis set and 2.62 debye with the aug-cc-pCVQZ basis set, the direction being $^{+}$SiN^{-} (in this study). An attachment of the H atom to the Si end of diatomic SiN significantly increases the magnitude of the dipole moment, because both electropositive atoms are adjacent to each other. Due to its relatively large dipole moment, the HSiN isomer will be observable via microwave spectroscopy.

Method	HSiN	HNSi	TS
CCSD(T) / **X**			
cc-pVQZ	4.261	0.136	2.950
aug-cc-pVQZ	4.348	0.256	3.032
cc-pV5Z	4.339	0.212	3.026
aug-cc-pV5Z	4.373	0.254	3.055
cc-pCVTZ	4.055	0.021	2.795
aug-cc-pCVTZ	4.299	0.269	3.019
cc-pCVQZ	4.271	0.143	2.975
aug-cc-pCVQZ	4.359	0.263	3.056

Table 4.3 Theoretical predictions of the dipole moments for the three stationary points along the HSiN–HNSi PES at the CCSD(T) level of theory (debye).

The dipole moment of the higher energy HNSi isomer generally decreases with improved treatments of correlation effects (except at the CASSCF level). At a given level of theory, the dipole moment is predicted to be slightly larger with the augmented (aug-cc-pV*n*Z and aug-cc-pCV*n*Z) basis sets relative to the standard (cc-pV*n*Z and cc-pCV*n*Z) basis sets. The magnitude of the dipole moment for the HNSi isomer is quite small, since the electropositive H

and Si atoms reside in both ends. With our highest level of theory, CCSD(T), the dipole moment of the HNSi isomer is predicted to be 0.25 debye with the aug-cc-pV5Z basis set and 0.26 debye with the aug-cc-pCVQZ basis set with the direction of $^{-}$HNSi^{+}. The addition of the H atom to the N end of diatomic SiN indeed greatly decreases the magnitude of the dipole moment for the HNSi isomer. Note that the direction of the dipole moment (in terms of the H atom) for the HSiN isomer is opposite to that for the HNSi structure.

The dipole moment for the isomerization reaction transition state is sensitive to the correlation level and basis sets. It is seen that the transition state is significantly more polarized than the HNSi isomer, but it is less polarized than the HSiN isomer.

4.4.6 Vibrational Frequencies

Computed harmonic vibrational frequencies for the three stationary points along the PES are presented in Table 4.4. For the five vibrational modes of the HSiN-HNSi system for which experimental fundamental vibrational frequencies are available, there exists qualitative agreement between theory and experiment.[1, 2, 7] These "well behaved" vibrational modes include both stretching modes in HSiN and HNSi, as well as the bending mode in HNSi. The bending mode of HSiN is uncharacteristically sensitive to both correlation treatment and basis set effects, so this mode will be discussed more in depth later.

The "well behaved" stretching and bending modes of the HSiN-HNSi system are quite sensitive to correlation treatment, as they decrease by approximately 20 and 50 cm^{-1} respectively upon inclusion of higher order correlation effects [CCSD $\rightarrow$ CCSD(T)] reflecting the elongated bond distances. Computations employing CCSD(T) and ICMRCI give comparable results, with the biggest differences coming from predicted hydrogen stretch vibrations ($\Delta\omega_{HN}\approx$11 cm^{-1} and $\Delta\omega_{HSi}\approx$28 cm^{-1}). The choice of basis set also influences the predicted harmonic vibrational

frequencies. While these calculations seem to be well converged with respect to angular momentum functions at the quadruple-ζ level, the inclusion of core-correlation and the associated tight core functions has a noticeable effect upon the predicted vibrational frequencies. The increase of the stretching modes by ~10 cm^{-1} is consistent with the slight bond contraction observed when the effects of core electron correlation are included in the geometry optimizations. The effect of diffuse functions is less pronounced, as predicted harmonic vibrational frequencies changed by no more than 5 cm^{-1}. The predicted harmonic vibrational frequencies for **TS** follow a similar trend to what is observed with the "well behaved" modes of HSiN and HNSi.

	HSiN			HNSi			**TS**		
Method	$\omega_1(\sigma)$	$\omega_2(\pi)$	$\omega_3(\sigma)$	$\omega_1(\sigma)$	$\omega_2(\pi)$	$\omega_3(\sigma)$	$\omega_1(a')$	$\omega_2(a')$	$\omega_3(a')$
CCSD(T) /									
cc-pVQZ	2203.9	138.9	1163.0	3750.9	528.3	1216.7	2054.9	797.3*i*	1096.5
aug-cc-pVQZ	2202.2	151.0	1161.4	3746.5	524.8	1212.4	2049.7	793.0*i*	1094.2
cc-pV5Z	2206.4	152.0	1168.1	3749.8	532.4	1219.6	2056.3	792.3*i*	1100.9
aug-cc-pV5Z	2205.6	153.6	1167.4	3747.1	530.9	1218.0	2054.8	790.9*i*	1100.7
cc-pCVQZ	2205.9	128.7	1173.1	3757.7	535.6	1224.4	2065.6	786.0*i*	1105.5
aug-cc-pCVQZ	2205.1	140.9	1171.6	3752.8	532.7	1220.3	2061.8	781.5*i*	1103.9
ICMRCI /									
cc-pVQZ	2232.1	107.1	1156.2	3739.7	529.6	1215.8			
aug-cc-pVQZ	2231.5	120.3	1155.2	3736.6	526.1	1212.1			
cc-pV5Z	2235.9	121.1	1161.6	3739.3	533.2	1218.7			
aug-cc-pV5Z	2235.5	122.7	1161.2	3737.2	531.6	1217.4			
B3LYP /									
6-311+G(d,p)[a]	2225	260	1236	3743	544	1230			
	$\nu_1(\sigma)$	$\nu_2(\pi)$	$\nu_3(\sigma)$	$\nu_1(\sigma)$	$\nu_2(\pi)$	$\nu_3(\sigma)$	$\nu_1(a')$	$\nu_2(a')$	$\nu_3(a')$
VPT2	2143.5		1151.3	3598.7	523.0	1205.7			
Expt.									
Ref. 1				3583	523	1198			
Ref. 2				3588					
Ref. 7	2152.2		1162.2	3585	522	1202			

Table 4.4 Harmonic and fundamental vibrational frequencies for the three structures on the PES at the CCSD(T), ICMRCI and B3LYP levels of theory (cm^{-1}).

Unlike other computed harmonic vibrational frequencies, the bending mode of HSiN was poorly converged with respect to the cardinality of the basis set, and the inclusion of diffuse functions caused substantial changes to computed harmonic vibrational frequencies (changes of ~25 cm^{-1} were observed for both effects). As was the case with other modes, we observed a decrease of ~10 cm^{-1} with the inclusion of core correlation effects.

The harmonic vibrational frequency of the diatomic $X\ ^2\Sigma^+$ SiN is predicted to be 1148 cm^{-1} at the CCSD(T)/aug-cc-pCVQZ level of theory (present study). The SiN stretching frequencies of the HSiN and HNSi isomers are 24 and 72 cm^{-1} higher than the diatomic SiN, respectively. This result is consistent with the observation that in the HSiN isomer, the hydrogen atom is not strongly bonded to the silicon atom. A significantly lower bending frequency for the HSiN isomer relative to that for the HNSi isomer is consistent with the smaller activation energy for the forward isomerization reaction. Finally, it should be noted that, unlike the HCN-HNC system, we do not expect quantum tunneling to play a significant role in the isomerization of HSiN to HNSi. While the barrier of the forward isomerization process is fairly low (~10 kcal mol^{-1}), the magnitude of the imaginary frequency of **TS** ($\omega_2 \approx 790i$ cm^{-1}) suggests that the barrier is broader than the barrier between the HCN-HNC isomers (~1200i cm^{-1}), and thus less susceptible to quantum tunneling effects.

4.4.7 Anharmonic Effects

After a careful consideration of the effects of basis set and correlation treatment upon the calculated harmonic vibrational frequencies, we chose the CCSD(T)/cc-pCVQZ method to further investigate the incorporation of anharmonic effects into predicted molecular properties. As expected, upon incorporation of anharmonic effects via VPT2, computed frequencies improved dramatically with respect to experiment. The calculation of higher order force

constants also permitted the computation of zero-point corrected rotational constants. These computed molecular constants may facilitate the interstellar detection of both the HSiN and HNSi isomers. The predicted fundamentals of the HNSi isomer are ν_1=3599, ν_2=523 and ν_3=1206 cm^{-1} as shown in Table 4.4. These values are in close agreement with experimental observations (RMSD=8 cm^{-1}).[7] Our computed QFF and molecular properties represent a dramatic improvement over a previous PES computed by Chong *et al.* using DFT.[20]

	DSiN			DNSi		
Method	$\omega_1(\sigma)$	$\omega_2(\pi)$	$\omega_3(\sigma)$	$\omega_1(\sigma)$	$\omega_2(\pi)$	$\omega_3(\sigma)$
CCSD(T) /						
cc-pVQZ	1603.2	106.1	1144.2	2767.2	404.6	1180.2
aug-cc-pVQZ	1602.0	115.3	1142.5	2763.9	402.0	1176.1
cc-pV5Z	1605.1	116.1	1149.1	2766.6	407.9	1182.9
aug-cc-pV5Z	1604.5	117.4	1148.4	2764.5	406.7	1181.4
cc-pCVQZ	1605.0	98.3	1153.9	2772.5	410.3	1187.6
aug-cc-pCVQZ	1604.4	107.6	1152.4	2768.8	408.1	1183.6
ICMRCI /						
cc-pVQZ	1622.6	81.7	1138.2	2758.2	405.6	1179.6
aug-cc-pVQZ	1622.2	91.8	1137.2	2755.9	403.0	1176.1
cc-pV5Z	1625.3	92.5	1143.5	2758.0	408.5	1182.5
aug-cc-pV5Z	1625.0	93.7	1143.2	2756.4	407.2	1181.2
	$\nu_1(\sigma)$	$\nu_2(\pi)$	$\nu_3(\sigma)$	$\nu_1(\sigma)$	$\nu_2(\pi)$	$\nu_3(\sigma)$
VPT2	1570.5		1137.0	2686.5	401.6	1172.2
Expt.						
Ref. 1				2669	395	1166
Ref. 7	1580.5		1145.4			

Table 4.5 Harmonic and fundamental vibrational frequencies for the two deuterated minima on the PES at the CCSD(T) and ICMRCI levels of theory (cm^{-1}).

As is the case with HNSi, our predicted fundamentals for the stretching modes (ν_1=2144 and ν_3=1151 cm^{-1}) of HSiN are in strong agreement with IR difference spectra taken by Maier

and Glatthaar.[7] Unfortunately, due to the large amplitude motion of the bending mode (ω_2≈125 cm^{-1}), VPT2 theory is incapable of making meaningful predictions about ν_2. This is because VPT2 assumes roughly harmonic zeroth-order behavior, a condition which ν_2 obviously does not meet. This failure of VPT2 theory manifests itself in the form of unphysical "predictions" about the bending mode in HSiN. Specifically, VPT2 predicts an anharmonic correction of *plus* 173 cm^{-1} to a predicted harmonic frequency of 129 cm^{-1}, a result which is unreasonable in both its magnitude and direction. Thus, proper treatment of this bending mode requires the application of more sophisticated variational vibrational methods. Despite the shortcomings of VPT2 in treating the bending mode of HSiN, its other predictions are in very good agreement with experiment (RMSD=9 cm^{-1} between experiment and VPT2 for the remaining five modes), and should not be discounted.

4.4.8 Vibrational isotopic shifts

Predicted harmonic and fundamental frequencies for the deuterated species are presented in Table 4.5. Agreement between the fundamentals predicted by VPT2 theory and those observed in experiments remains excellent (RMSD=11 cm^{-1}). To further evaluate our vibrational results versus available experimental data, we decided to compare isotopic shifts. This approach has the advantage of utilizing error cancelation to remove external effects which might influence an experimental result, yet which are not readily modeled by *ab initio* means. An example of one such effect is matrix induced IR shifts, a phenomenon which can account for shifts of up to 50 cm^{-1} from gas phase observations.[86] The isotopic shifts for the vibrational frequencies of both isomers are presented in Table 4.6. The harmonic isotopic shifts from CCSD(T)/cc-pCVQZ and ICMRCI/aug-cc-pV5Z theories are both in near quantitative agreement with each other for all five computed isotopic shifts. Agreement between the

harmonically computed isotopic shifts and experiment is reasonable on the whole, and quite good in cases where hydrogen stretches are not involved. As one would expect, agreement increases significantly when the VPT2 isotopic shifts are compared with experimental shifts (RMSD=3 cm^{-1}).

	HSiN – DSiN			HNSi – DNSi		
Method	$\Delta\omega_1$	$\Delta\omega_2$	$\Delta\omega_3$	$\Delta\omega_1$	$\Delta\omega_2$	$\Delta\omega_3$
CCSD(T)/cc-pCVQZ	600.9	30.4	19.2	985.2	124.6	36.7
ICMRCI/aug-cc-pV5Z	610.5	29.0	18.0	980.8	124.4	36.2
VPT2	573.0		14.3	912.2	121.4	33.5
Expt.[a,b]	571.7		16.8	914	128	32

Table 4.6 Vibrational isotropic shifts for the HNSi/DNSi and HSiN/DSiN systems (cm^{-1}).
[a]Ref. 1
[b]Ref. 7

Table 4.7 contains computed B_0 values for the five most common isotopologues of HNSi. For the three isotopologues of HNSi where experimental[3, 6] data exists, agreement between our VPT2 values and observation is remarkable. In all three cases, our predicted values underestimate the experimental value by approximately 20 MHz. Our computed B_0 values for these isotopologues represent a significant improvement over previous calculations in the literature,[12] and without employing any empirical scaling factors. The currently accepted experimental geometry for HNSi was calculated from both experimentally[3] and theoretically[4] derived B_0 values. These theoretical rotational constants were computed variationally using a CEPA-1 potential. Our B_0 values calculated using a CCSD(T) potential fall into better agreement with experiment, and might be used to determine a more accurate structure.

Isotopic Substitution	HSiN B_e [a]	HNSi VPT2[a]	HNSi CEPA-1[b]	HNSi Expt.[b]	HNSi B_e [a]
(H,N,Si)	18653.46	18996.91	18959.61	19018.80	19093.97
(D,N,Si)	16372.92	16601.56			16658.30
(H,N,^{29}Si)	18500.96	18776.33	18739.45	18797.99	18872.02
(H,N,^{30}Si)	18357.87	18570.75	18534.25	18592.19	18665.17
(H,^{15}N,Si)	17880.58	18341.12			18435.44

Table 4.7 Predicted and experimental rotational constants of the HSiN and HNSi isomers (MHz). The values computed using VPT2 and CEPA-1 include zero-point corrections
a Computed at the AE-CCSD(T)/cc-pCVQZ optimized geometry
b Ref. 4

4.4.9 Dissociation Energies

The dissociation energies for

$$\text{HSiN } (\tilde{X}\,^1\Sigma^+) \rightarrow \text{H } (^2\text{S}) + \text{SiN } (\text{X}\,^2\Sigma^+) \tag{4.7}$$

and

$$\text{HNSi } (\tilde{X}\,^1\Sigma^+) \rightarrow \text{H } (^2\text{S}) + \text{SiN } (\text{X}\,^2\Sigma^+) \tag{4.7}$$

at the CCSD(T) level of theory are presented in Table 4.8.

The dissociation energy of the HSiN isomer increases with correlation effects by as much as 36 kcal mol^{-1}. The H-SiN dissociation energies are predicted to be D_e=63.8 and D_0=60.2 kcal mol^{-1} with the aug-cc-pV5Z basis set, and D_e =63.2 and D_0 =59.6 kcal mol^{-1} with the aug-cc-pCVQZ basis set. The dissociation energy of the HNSi isomer is less sensitive than that for the HSiN isomer and increases with the inclusion of correlation effects by as much as 15 kcal mol^{-1}. With the CCSD(T) method, the dissociation energies for HNSi are predicted to be D_e=131.8 and D_0=124.9 kcal mol^{-1} with the aug-cc-pV5Z basis set, and D_e=132.0 and D_0=125.0 kcal mol^{-1} with the aug-cc-pCVQZ basis set. It is seen that both the HSiN and HNSi isomers are favorably stabilized with correlation effects compared to the diatomic SiN molecule. The difference in the dissociation energies between eq. 4.7 and eq. 4.8 is, of course, equivalent to the

relative energy difference of the HSiN and HNSi isomers. The three stationary points (HSiN, HNSi and isomerization reaction transition state) studied in this research are well below the dissociation channel H + SiN.

	HSiN		HNSi	
Method	D_e	D_0	D_e	D_0
CCSD(T) /				
cc-pVQZ	63.45	59.87	131.06	124.08
aug-cc-pVQZ	63.61	60.00	131.38	124.42
cc-pV5Z	63.76	60.14	131.74	124.75
aug-cc-pV5Z	63.83	60.21	131.84	124.87
cc-pCVTZ	62.11	58.54	129.84	122.82
aug-cc-pCVTZ	62.54	58.92	130.62	123.70
cc-pCVQZ	63.01	59.45	131.65	124.64
aug-cc-pCVQZ	63.15	59.56	131.96	124.97

Table 4.8 The dissociation energies of the HSiN and HNSi isomers going to H(^{2}S) + SiN ($\tilde{X}\,^1\Sigma^+$) at the CCSD(T) level of theory (kcal mol^{-1}).

4.5 Conclusions

The electronic ground state potential energy surface of the HSiN-HNSi system has been characterized by a variety of highly correlated *ab initio* quantum mechanical methods, including coupled cluster, multireference configuration interaction and focal point analysis (FPA). Large correlation consistent type basis sets were employed to yield the most accurate, yet computationally feasible predictions possible today. Our calculations confirmed the HNSi isomer as the global minimum on the potential energy surface. The most sophisticated method, FPA, predicts the HSiN isomer to lay 64.7 kcal mol^{-1} above the global minimum on the ZPVE corrected surface. This same method also yields highly accurate barrier heights; the forward reaction barrier is predicted to be 9.7 kcal mol^{-1}, whereas the reverse reaction barrier is predicted to be 74.4 kcal mol^{-1}. The juxtaposition of the HSiN-HNSi and HCN-HNC systems can help explain the difference in energetic trends. Because silicon's non-bonding electrons typically

have a high percentage of *s*-character,[21] the orbital overlap between hydrogen and silicon in HSiN will be relatively poor compared with the orbital overlap between hydrogen and carbon in HCN. Furthermore, since silicon is more electropositive than carbon, silicon's non-bonding orbital should be higher in energy than the analogous non-bonding orbital of the carbon system. These two effects may disrupt the formation of a hydrogen-silicon bond relative to the formation of a hydrogen-nitrogen bond. We can corroborate this from our predicted geometries and frequencies, which both show the HSi bond in HSiN to be weak.

Because of their possible presence in the interstellar medium, the properties of these molecules are of interest. VPT2 theory predicts a B_0 value of 18997 MHz for the HNSi isomer. Our most rigorous coupled cluster calculations predict the HSiN isomer to have a dipole moment of 4.36 Debye in the direction $^{+}HSiN^{-}$. Similar calculations predict the HNSi isomer to have a very small dipole moment of 0.26 Debye in the opposite direction. These theoretical properties should facilitate the possible identification of HNSi in the interstellar medium. VPT2 was also utilized in the computation of fundamental frequencies. The predicted fundamentals for HSiN are ν_1=2144 and ν_3=1151 cm^{-1}, and the predicted fundamentals for HNSi are ν_1=3599, ν_2=523 and ν_3=1206 cm^{-1}; agreement with available experiments is excellent with an overall RMSD of 9 cm^{-1}. The ZPVE corrected dissociation energy for the HSiN is predicted to be D_0=59.6 kcal mol^{-1}, while the dissociation energy for the HNSi isomer is predicted to be D_0=125.0 kcal mol^{-1}. We hope that the reliable physical properties and energetics for these two isomers will assist in further spectroscopic characterization of the HSiN-HNSi system. Furthermore, our methods should be easily extensible to other iminosilicon containing systems, and should be able to provide valuable theoretical corroboration as more of these systems are synthetically realized.

4.6 References

1. J. F. Ogilvie and S. Cradock, Chem. Commun.364 (1966).
2. M. Elhanine, R. Farrenq and G. Guelachvili, J. Chem. Phys. **94**, 2529-2531 (1991).
3. M. Bogey, C. Demuynck, J. L. Destombes and A. Walters, Astron. Astrophys. **244**, L47-L49 (1991).
4. P. Botschwina, M. Tommek, P. Sebald, M. Bogey, C. Demuynck, J. L. Destombes and A. Walters, J. Chem. Phys. **95**, 7769-7770 (1991).
5. N. Goldberg, M. Iraqi, J. Hrušák and H. Schwarz, Int. J. Mass Spectrom. Ion Processes **125**, 267-270 (1993).
6. M. Elhanine, B. Hanoune and G. Guelachvili, J. Chem. Phys. **99**, 4970-4974 (1993).
7. G. Maier and J. Glatthaar, Angew. Chem. Int. Ed. Engl. **33**, 473-475 (1994).
8. W.-K. Chen, I. C. Lu, C. Chaudhuri, W.-J. Huang and S.-H. Lee, J. Phys. Chem. A **112**, 8479-8486 (2008).
9. J. N. Murrell, H. W. Kroto and M. F. Guest, J. Chem. Soc., Chem. Commun.619-620 (1977).
10. R. Preuss, R. J. Buenker and S. D. Peyerimhoff, J. Mol. Struct. **49**, 171-179 (1978).
11. Y. Apeloig and K. Albrecht, J. Am. Chem. Soc. **117**, 7263-7264 (1995).
12. O. Parisel, M. Hanus and Y. Ellinger, Chem. Phys. **212**, 331-351 (1996).
13. O. Parisel, M. Hanus and Y. Ellinger, J. Phys. Chem. A **101**, 299-309 (1997).
14. S. W. Hu, Y. Wang, X. Y. Wang, T. W. Chu and X. Q. Liu, J. Phys. Chem. A **107**, 9189-9196 (2003).
15. O. Kwon and Y. Kwon, THEOCHEM **460**, 213-220 (1999).

16. D. Danovich, F. Ogliaro, M. Karni, Y. Apeloig, D. L. Cooper and S. Shaik, Angew. Chem. Int. Ed. Engl. **40**, 4023-4026 (2001).

17. D. Sengupta and M. T. Nguyen, Chem. Phys. Lett. **265**, 35-40 (1997).

18. R. Becerra, J. P. Cannady and R. Walsh, J. Phys. Chem. A **107**, 9588-9593 (2003).

19. B. S. Jursic, THEOCHEM **455**, 77-83 (1998).

20. D. P. Chong, D. Papoušek, Y. T. Chen and P. Jensen, J. Chem. Phys. **98**, 1352-1357 (1993).

21. B. T. Luke, J. A. Pople, M. B. Kroghjespersen, Y. Apeloig, J. Chandrasekhar and P. v. R. Schleyer, J. Am. Chem. Soc. **108**, 260-269 (1986).

22. B. T. Luke, J. A. Pople, M. B. Kroghjespersen, Y. Apeloig, M. Karni, J. Chandrasekhar and P. v. R. Schleyer, J. Am. Chem. Soc. **108**, 270-284 (1986).

23. R. Walsh, Acc. Chem. Res. **14**, 246-252 (1981).

24. W. Kutzelnigg, Angew. Chem. Int. Ed. Engl. **23**, 272-295 (1984).

25. G. Olbrich, P. Potzinger, B. Reimann and R. Walsh, Organometallics **3**, 1267-1272 (1984).

26. D. D. S. Mackay, Mon. Not. R. Astron. Soc. **274**, 694-700 (1995).

27. A. Chive, V. Lefevre, A. Systermans, J. L. Ripoll, M. Bogey and A. Walters, Phosphorus Sulfur **91**, 281-284 (1994).

28. J. G. Radziszewski, D. Littmann, V. Balaji, L. Fabry, G. Gross and J. Michl, Organometallics **12**, 4816-4824 (1993).

29. T. H. Dunning, J. Chem. Phys. **90**, 1007-1023 (1989).

30. D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358-1371 (1993).

31. D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572-4585 (1995).

32. K. A. Peterson and T. H. Dunning, J. Chem. Phys. **117**, 10548-10560 (2002).

33. A. K. Wilson, T. van Mourik and T. H. Dunning, J. Mol. Struc. **388**, 339-349 (1996).

34. P. Siegbahn, A. Heiberg, B. Roos and B. Levy, Phys. Scr. **21**, 323-327 (1980).

35. B. O. Roos, Int. J. Quantum Chem. **17**, 175-189 (1980).

36. B. O. Roos and P. R. Taylor, Chem. Phys. **48**, 157-173 (1980).

37. P. J. Knowles and H. J. Werner, Chem. Phys. Lett. **145**, 514-522 (1988).

38. H. J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803-5814 (1988).

39. H. J. Werner and E. A. Reinsch, J. Chem. Phys. **76**, 3144-3156 (1982).

40. G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910-1918 (1982).

41. M. Rittby and R. J. Bartlett, J. Phys. Chem. **92**, 3033-3036 (1988).

42. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479-483 (1989).

43. J. F. Stanton, Chem. Phys. **281**, 130-134 (1997).

44. Y. S. Lee, S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **81**, 5906-5912 (1984).

45. Y. J. Bomble, J. F. Stanton, M. Kállay and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

46. M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

47. S. A. Kucharski and R. J. Bartlett, Theor. Chim. Acta **80**, 387-405 (1991).

48. S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **97**, 4282-4288 (1992).

49. M. Musial, S. A. Kucharski and R. J. Bartlett, Chem. Phys. Lett. **320**, 542-548 (2000).

50. P. Pulay, Mol. Phys. **17**, 197-204 (1969).

51. P. Pulay, edited by H. F. Schaefer (Plenum, New York, 1977), Vol. 4, pp. 153-185.

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

53. A. L. L. East and W. D. Allen, J. Chem. Phys. **99**, 4638-4650 (1993).

54. M. S. Schuurman, S. R. Muir, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **120**, 11586-11599 (2004).

55. J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Császár and G. Tarczay, Chem. Eur. J. **9**, 2173-2192 (2003).

56. A. G. Császár, W. D. Allen and H. F. Schaefer, J. Chem. Phys. **108**, 9751-9764 (1998).

57. A. G. Császár, G. Tarczay, M. L. Leininger, O. L. Polyansky, J. Tennyson and W. D. Allen, in *Spectroscopy from Space*, edited by J. Demaison, K. Sarka and E. A. Cohen (Kluwer, Dordrecht, 2001), Vol. 20, pp. 317.

58. A. Karton, E. Rabinovich, J. M. L. Martin and R. Branko, J. Chem. Phys. **125**, 144108 (2006).

59. A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, J. Chem. Phys. **121**, 11599-11613 (2004).

60. R. A. Kendall, T. H. Dunning and R. J. Harrison, J. Chem. Phys. **96**, 6796-6806 (1992).

61. D. Feller, J. Chem. Phys. **98**, 7059-7071 (1993).

62. H. Sellers and P. Pulay, Chem. Phys. Lett. **103**, 463-465 (1984).

63. N. C. Handy, Y. Yamaguchi and H. F. Schaefer, J. Chem. Phys. **84**, 4481-4484 (1986).

64. S. A. Perera and R. J. Bartlett, Chem. Phys. Lett. **216**, 606-612 (1993).

65. R. D. Cowan and D. C. Griffin, J. Opt. Soc. Am. **66**, 1010-1014 (1976).

66. D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi and H. F. Schaefer, Chem. Phys. **123**, 187-239 (1988).

67. J. K. G. Watson, in *Vibrational Spectra and Structure*, edited by J. R. Durig (Elsevier, Amsterdam, 1977), Vol. 6, pp. 1-89.

68. W. D. Allen, Y. Yukio, A. Császár, G., D. A. Clabo, R. B. Remington and H. F. Schaefer, Chem. Phys. **145**, 427-466 (1990).

69. I. M. Mills, *Molecular Spectroscopy: Modern Research*. (Academic Press, New York, 1972).

70. D. Papoušek and M. R. Aliev, *Molecular Vibrational-Rotational Spectra*. (Elsevier, Amsterdam, 1982).

71. H. H. Nielsen, Rev. Mod. Phys. **23**, 90-136 (1951).

72. K. Aarset, A. G. Császár, E. L. Sibert, W. D. Allen, H. F. Schaefer, W. Klopper and J. Noga, J. Chem. Phys. **112**, 4053-4063 (2000).

73. S. W. Fox and C. R. Windsor, Science **170**, 984-986 (1970).

74. PSI 2.0.

75. J.F. Stanton, J. Gauss, J.D. Watts, P.G. Szalay, R.J. Bartlett with contributions from A.A. Auer, D.E. Bernholdt, O. Christiansen, M.E. Harding, M. Heckert, O. Heun, C. Huber, D. Jonsson, J. Jusélius, W.J. Lauderdale, T. Metzroth, C. Michauk, D.R. Price, K. Ruud, F. Schiffmann, A. Tajti, M.E. Varner, J. Vázquez and the integral packages: MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), and ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen). Current version see http://www.aces2.de.

76. J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale and R. J. Bartlett, Int. J. Quant. Chem. Symp. **26**, 879-894 (1992).

77. M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945-2954 (2001).

78. MRCC, a string-based quantum chemical program suite written by M. Kállay. Current version see www.mrcc.hu.

79. INTDIF2005 is an abstract program written by Wesley D. Allen for MATHEMATICA to perform general numerical differentiations to high orders of electronic structure data.

80. INTDER2005 is a general program written by Wesley D. Allen which performs various vibrational analyses and higher-order nonlinear transformations among force field representations.

81. W. D. Allen, A. G. Császár, V. Szalay and I. M. Mills, Mol. Phys. **89**, 1213-1221 (1996).

82. ANHARM is a FORTRAN program for VPT2 analysis written by Y. Yamaguchi and H. F. Schaefer, Center for Computational Chemistry, University of Georgia, Athens, GA. .

83. T. J. Lee and P. R. Taylor, Int. J. Quant. Chem. Symp. **23**, 199-207 (1989).

84. G. S. Hammond, J. Am. Chem. Soc. **77**, 334-338 (1955).

85. S. R. Langhoff and E. R. Davidson, Int. J. Quantum Chem. **8**, 61-72 (1974).

86. M. E. Jacox, Chem. Soc. Rev. **31**, 108-115 (2002).

CHAPTER 5

IMPLEMENTATION OF A FRAMEWORK FOR COARSE GRAINED MODELING IN CHARMM

5.1 Introduction

To function properly, proteins must fold.[1] Determining structure, and understanding the mechanisms responsible for folding are an active area of biophysical research, as gleaning this information may provide critical insights towards fighting diseases that have been linked to protein structure, such as Alzheimer's.[2, 3] Experimental determinations of protein structure are typically carried out using x-ray diffraction of crystalized proteins or NMR spectroscopy upon aqueous samples. Both techniques provide important information about a proteins native folded structure, yet both methods are not without their drawbacks. The process of crystalizing a protein is labor intensive, and structural information comes from a non-biological environment. NMR studies on the other hand, do come from an aqueous environment; however are time-averaged, and thus dynamical details are often lost or may not be sufficiently resolved. Computer simulations however, do not suffer these consequences, and provide complementary information which is able to confirm experimental findings, and supplement them with dynamical detail.

Thus theoretical studies may lead to a better understanding of experimental results by providing simple models with easily interpreted results. Such theoretical models have assisted with the analysis and refinement of low resolution experimental data.[4] The ultimate promise of computational protein studies is to be able to predict protein structure from sequence alone. This

promise has yet to be fulfilled however, as the number of possible conformations of a protein scales exponentially with system size, and even for average sized proteins, the problem is computationally intractable. A possible solution to this problem was proposed by Levinthal,[5] where he suggested that protein folding was guided by the rapid formation of local interactions, essentially forcing the protein to go through a sequence of intermediate structures before finally settling into the native state. This viewpoint makes molecular dynamics (MD) simulations particularly attractive for protein folding, as the dynamical information is perfectly resolved.

Unfortunately, only small and ultrafast folding proteins are tractably folded using MD techniques.[6] Modeling larger systems or smaller systems for biologically relevant timescales often requires the use of coarse grained (CG) models. CG models of proteins and biomolecular complexes are an important tool for understanding the nature of protein folding, docking, ligand binding, and other biochemical reactions.

Most generally CG models reduce the computational expense versus their all-atom counter parts two separate ways. First, by reducing the number of particles, the number of pairwise non-bonded interactions which must be computed is vastly reduced. Second, because there are less particles contributing to the potential energy surface, it is much smoother than its all-atom counterpart. This, in turn, allows the equations of motion to be evaluated less frequently without error accumulation, and speeds the time evolution of the system by a factor of ten times or more. This effect is present in all CG models, and is a function of the resolution of the CG model. The simplest protein models map an entire amino acid into one-particle. Higher resolution models map a single amino acid into a backbone particle and a side chain center of mass. Higher resolution models still, may map only three heavy atoms into a single particle, or may map hydrogen atoms into the nearest heavy atom. The choice of the CG topology

(resolution) dictates several factors independent of the details of the specific model. Specifically, the simplest one-bead models have such a smooth potential energy surface that they will fold to the native state under all but the most extreme of conditions. Moreover this folding will happen unnaturally fast, such that many dynamical details may be lost in the simulation. Higher resolution models have a rougher potential surface, and thus more intermediate meta-stable states exist. This slows the folding process, and enables more accurate dynamics to be modeled. Also, the presence of local minima on the potential surface, allows the theoretical study of ensembles of misfolded proteins, which is important for the study of diseases linked to protein misfolding[2, 3] and denaturation.[7]

Because of their emerging importance in the modeling of biological phenomena, and the complexity involved in successfully building a CG protein model, we have chosen to extend the capabilities of the CHARMMing[8] (CHARMM interface and graphics) software package by implementing a framework for CG model development, and then utilized this framework to build a Gō-like protein model.[9] In this model, the experimentally determined crystal structure is set as the energetic minimum by design; this guarantees that the protein's minimum energy structure below the melting point is the experimentally determined native structure. Then, because we know that the native state is the most stable, the native contact interactions must be more stable than the non-native interactions. Thus native interaction strengths are attractive, and are based on an experimentally determined contact pair potential,[10] whereas non-native contacts are repulsive.

CHARMMing is a recently developed graphical user interface (GUI) designed to be used with the CHARMM[11, 12] software package. This program is implemented as a web application, and is designed to help all users, both novice and experienced, with a basic package of tools to

assist with setting up, running and analyzing data from various CHARMM calculations. Specifically, CHARMMing is capable of the following: job preparation, job submission and monitoring, visualization of structures and trajectories, and testing and deployment of new methods. Extensions to CHARMMing's capabilities by the present work will be discussed in subsequent sections. CHARMMing is written in Python,[13] and implemented using the Django framework for web application development.[14]

The CHARMMing package is not the first web-based tool created to assist with molecular modeling. Numerous web applications exist to provide very narrowly focused functionality to a specific problem, and are not tied to any specific underlying simulation package. For example, STRIDE[15] is a web server for providing secondary structure assignment to proteins from known atomic coordinates, ProPKA[16] allows for the calculation of protein pK_a values and ProBis[17] calculates structural similarities in related proteins to find conserved regions crucial to protein-protein interactions. Other web tools provide a wider variety of methods, but for a specific class of biomolecules, such as GLYCAM[18] (carbohydrates) or CHARMM-GUI[19] (membranes and membrane bound proteins). CHARMMing differs from these programs by allowing its users access to its tools in a highly flexible manner. Methods may be combined and applied to arbitrary biomolecules. Users are only limited by their own ingenuity and the models themselves.

In this work we will extend the capabilities of CHARMMing to include multiple types of CG models for both proteins and lipids. In the process of implementing these CG models, we also made enhancements to CHARMMing's internal libraries which should facilitate further development of CHARMMing's capabilities by both ourselves and third parties. Finally, to validate the implementation of our CG protein simulations, it was necessary to implement the

replica exchange (REMD) enhanced sampling method within CHARMMing, this will also be discussed.

5.2 Methods

Our CG protein model is a Gō-like model,[9] and is based off of previous work by Thirumulai *et al*.[20, 21] The experimentally determined native state is biased by native contact interactions. Secondary structure information is also determined from the native crystal structure, and is used to influence bonded and non-bonded parameters. For example, the location of hydrogen bonds is inferred from the native crystal structure. Also, the hydrogen bond strength, and dihedral parameter values are decreased in the presence of alpha helical and beta sheet secondary structure. This is done in an effort to further destabilize these structures over the disorganized coil counterparts, and allow a greater degree of conformational sampling. Each amino acid is reductively mapped into two interaction centers. One site at the alpha carbon (C_α) position and the other at the side chain (SC) center of mass position. This intermediate resolution affords a nice compromise between speed and accuracy.

Bonded interactions are present between neighboring amino acid residues (C_α-C_α) and within each non glycine amino acid (C_α-SC). These interactions are described by a weakly harmonic potential (eq. 5.1), with an equilibrium bond distance (r_0) derived from the experimental crystal structure and K_{bond} = 50 kcal mol^{-1} Å^{-2}. K_{bond} is much lower than the value seen in a typical aliphatic bond (~300 kcal mol^{-1} Å^{-2}), this is because this 'bond' is much longer, and is modeling the coupling between two residues, and not two C_α

$$V_{bond}(r) = K_{bond}(r - r_0)^2 \tag{5.1}$$

Chain stiffness is represented by a second weak harmonic potential (eq. 5.2), with an equilibrium bond angle (θ_0) derived from the experimental crystal structure and K_{angle} = 30 kcal mol^{-1} rad^{-2}.

This value for K_{angle} is very much in line with the standard aliphatic bond angle bend parameter (~40 mol^{-1} rad^{-2}), giving the protein backbone slightly less stiffness than a saturated hydrocarbon chain. Also note that as per CHARMM convention, eq. 5.1 and eq. 5.2 does not contain a prefactor of ½, and care must be taken when translating this parameter sets for use in other MD simulation packages that use the traditional functional forms.

$$V_{angle}(\theta) = K_{angle}(\theta - \theta_0)^2 \tag{5.2}$$

Dihedral torsion potentials are also included in the model (eq. 5.3). Two sinusoidal terms are used with multiplicities (n) of 1 and 3. The parameters for the dihedral ($K_{dihedral}$) terms are weakened in the presence of an alpha helix from 0.55 kcal mol^{-1} (n = 1) and 0.275 kcal mol^{-1} (n = 3) to 0.30 kcal mol^{-1} (n = 1) and 0.15 kcal mol^{-1} (n = 3). This dihedral potential is much weaker than a standard aliphatic carbon dihedral potential which has a value of $K_{dihedral}$ of ~3 kcal mol^{-1}.

$$V_{dihedral}(\varphi) = K_{dihedral}(1 + \cos(n\varphi - \delta)) \tag{5.3}$$

Finally, an improper dihedral term (eq. 5.4) is used to enforce proper chirality around each bonding center. This harmonic potential has an equilibrium value (ω_0) determined by the experimental crystal structure, and an improper angle bending constant ($K_{improper}$) derived from the average native contact interaction multiplied by 20. This effectively restrains the protein to maintain the proper chirality.

$$V_{improper}(\omega) = K_{improper}(\omega - \omega_0)^2 \tag{5.4}$$

Nonbonded interactions between sites i and j are described by a 12-6 Lennard-Jones (LJ) potential (eq. 5.5). The ideal nonbonded interaction distance (r_{min}) is derived from the experimental crystal structure whereas the nonbonded interaction strength (ε_{min}) is determined by the type of nonbonded interaction being modeled. Native contacts, non-native contacts and

hydrogen bonds are all described by the LJ potential in this model. Native contact interactions are taken from an experimentally determined contact interaction parameter set, the most ubiquitous one being determined by Miyazawa and Jernigan (MJ).[10] By the assumptions of the Gō model, the non-native contacts are repulsive, and their interaction strength is zero. Finally, hydrogen bonding effects are also accounted for using the LJ potential. The strength of the hydrogen bond is determined by the presence or absence of secondary structure. In an alpha helix, a hydrogen bond has a LJ potential depth of 0.25 kcal mol^{-1}, outside of an alpha helix, this value doubles to 0.50 kcal mol^{-1}. While using a LJ potential to represent a hydrogen bond is not ideal, as the directionality of the bond is completely lost, this method has been shown to produce qualitatively correct diffusion properties in CG water in the literature.[22, 23] Nonbonded interactions between neighboring bonded (1-2, 1-3 and 1-4 interactions) sites are ignored, this is to prevent double counting of interactions in the model Hamiltonian.

Equilibrium simulations were carried out using REMD[24] in conjunction with low friction Langevin dynamics[25] (LD). REMD simulates multiple non-interacting trajectories in parallel. Each independent trajectory (replica) is simulated at a differing temperature, periodically the structure from one trajectory may be swapped with one from a trajectory from a neighboring simulation, as determined by standard acceptance criteria. For equilibrated ensembles that follow the Boltzmann distribution, the exchange probability (ρ) between replicas i and j is given by eq. 5.5 or unity, whichever value is lesser.[24] In this expression T_i and E_i are the temperature and potential energy, respectively, of the i^{th} replica, and k_B is Boltzmann's constant. Our REMD simulations employed 16 to 20 temperature windows between 200 K and 700 K. Our simulations employed 5 fs time steps, and every 5,000 time steps exchanges between the replicas were attempted. A damping coefficient of 1.0 ps^{-1} was used in our LD calculations, and the

canonical ensemble was used in all calculations. Individual trajectories were run out to 1 μs in an effort to recover converged statistics from our simulations. Non-bonded cutoffs were employed for pairwise interactions distances longer than 23 Å, switching functions were used to smooth these cutoffs starting at distances of 18 Å.

$$\rho = \exp\left\{((k_B T_i)^{-1} - (k_B T_j)^{-1})(E_i - E_j)\right\} \tag{5.5}$$

When analyzing MD trajectory data, one important consideration is the choice of reaction coordinate employed in data analysis.[26] In the present study we used three different reaction coordinates: radius of gyration (R_g), fraction of native contacts (Q) and backbone RMSD deviation (Δ). The radius of gyration was computed using CHARMM using eq. 5.6 where r_i and m_i are the position and mass of interaction site i respectively, and r_{CM} is the center of mass of all interaction sites. R_g is typically a good reaction coordinate for small globular proteins, however it can have trouble differentiating between multiple unfolded states, or determining unfolding for larger proteins whose folded and unfolded states have similar R_g values.

$$R_g^2 = \frac{1}{N}\left\langle \sum_{i=1}^{N} (r_i m_i - r_{CM})^2 \right\rangle \tag{5.6}$$

The fraction of native contacts is computed using eq. 5.7. Q_i may be computed for any subset of contacts in the parent model system. This is useful for quantitatively determining interactions between various domains, or measuring the degree to which a secondary structure element is present at a given time in a trajectory. The i denotes the subset of contacts considered during the analysis. C_i is the number of native contacts present in i, this is computed using the all-atom resolution from the experimental crystal structure. Residues which have any heavy atoms within 4.5 Å of each other are said to be native contacts. R_C is the cutoff distance for determining whether a contact is maintained throughout the trajectory of the CG simulation, this is set to 8 Å.

Finally, d_{jk} is the distance between interaction sites j and k, and Θ is the Heaviside step function. A Q value of 1 denotes a conformation in the native state, whereas a value of 0 corresponds to a totally misfolded protein.

$$Q_i = \sum_{j}^{N-4} \sum_{k=j+4}^{N} \frac{\Theta(R_c - d_{jk})}{C_i} \tag{5.7}$$

Backbone RMSD were also calculated using the CHARMM program. Least squares fitting between the initial structure, and the given protein conformation is carried out to maximally align the two structures. Then the RMSD is calculated between the two sets of C_α coordinates. A large RMSD value denotes an unfolded protein, whereas a small value corresponds to the native state.

Figure 5.1 A flowchart demonstrating how the CG model builder in CHARMMing interacts with the underlying CHARMM functionality.

5.3 Implementation

The CHARMMing CG model builder sits between the user and CHARMM. Taking a protein data bank (PDB) crystal structure, along with various user specified CG parameters, the CG model builder then generates a series of CHARMM input files (Figure 5.1). The user may specify any number of CG parameters including: native contact parameter sets, bonding parameters (described in Section 5.2), a global 'nScale' parameter which is used for tuning the strength of the native contacts versus other parameters, secondary structure parameters and van der Waals radii. These parameters, along with the PDB structure is used to determine secondary

structure information with STRIDE.[15] Finally, topology (rtf), parameter (prm) and structure (pdb) files are written in CHARMM compatible format. From here a standard CHARMM job is initiated using all of the tools made available by CHARMMing. To date, the CG model builder has been tested using the CHARMMing interfaces for the following CHARMM functions: normal mode analysis, molecular dynamics, Langevin dynamics, replica exchange, multi-scale modeling and analysis.

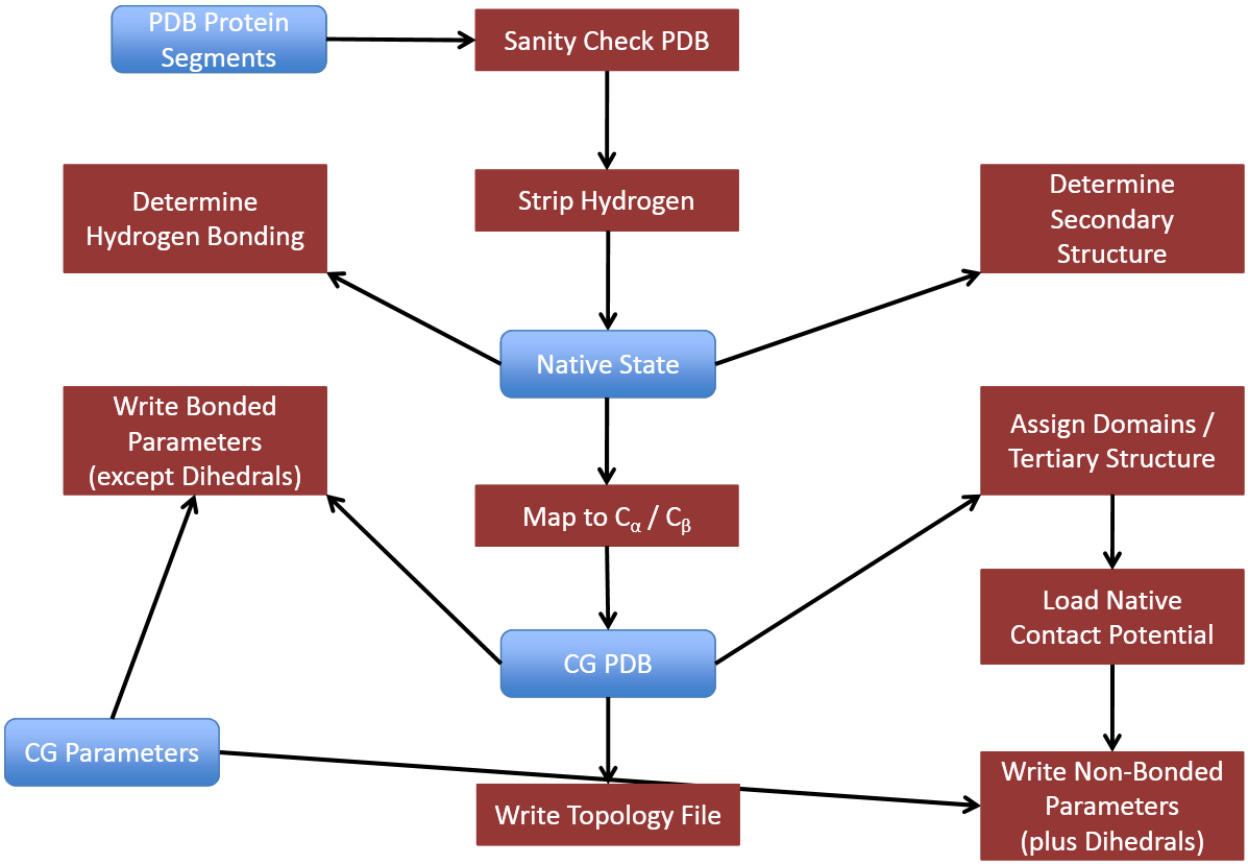

Figure 5.2 A flowchart detailing the algorithm responsible for generating CHARMM parameters from input parameters and native crystal structure.

A flowchart detailing the implementation of the CG model builder is shown in Figure 5.2. After the user specified parameters are input, along with the native crystal structure, the CHARMMing input parser is called, and separates the native structure into segments based upon the type of biomolecule (protein, lipid, carbohydrate, nucleic acid or other) and available

segment information from the original structure file. The protein segments are then input into the CG model builder which first checks them for errors (both typographical and biological) and attempts to correct them. Next, as the model does not require hydrogen atom information, any hydrogen atoms, if present, are discarded. The remaining structure is known as the "native state." The native state is then analyzed to determine the secondary structural elements such as alpha helices, 310 helices, beta sheets and coils. Hydrogen bonding information is also extracted from this structure, and stored for later use. Next, the native state is reductively mapped to a two-center per residue topology as described in Section 5.2. Once this is done, the final CHARMM readable pdb file may be written.

The CG structure is then used together with user specified parameters, and any available user specified domain or tertiary structure information[27] to build the CHARMM readable parameter sets. Unlike in typical dynamics simulations, each individual CG particle is assigned its each unique atom type, this is necessary because interaction potential has its own unique minimum value, determined by the crystal structure. So while the CG parameters and the algorithm for determining the CHARMM parameters is transferable to many different systems, the CHARMM parameters themselves are not transferable at all. This has only one practical implication, and that is that an unusually large amount of memory is required to store the entire set of CHARMM parameters, and it is likely that a custom version of CHARMM will have to be compiled from source to raise the default memory allocation for parameter sets.

Our CG model is written in object oriented Python. This allows us to easily leverage existing features from CHARMMing, such as its PDB processing, sanity checking file writing capabilities. Furthermore, because both CHARMMing and the CG model builder are both written using the same underlying library of utilities, enhancements and features added

specifically for CG model building have been reciprocated back into CHARMMing. Another advantage of using an object oriented implementation is that the Gō model may be extended to include intermolecular interactions from particles not present in the native crystal structure. For example, solvent effects or even protein binding may be accounted for using such a model. CHARMMing also provides us with a robust platform for building a GUI for the CG model builder. Using the GUI (Figure 5.3), it is easy to rapidly reconfigure input CG parameters, run the simulations and view the results. This is an important feature during the early stages of model development, when many simulations must be iteratively run to optimize the weighting between native side chain interactions and all other interactions.

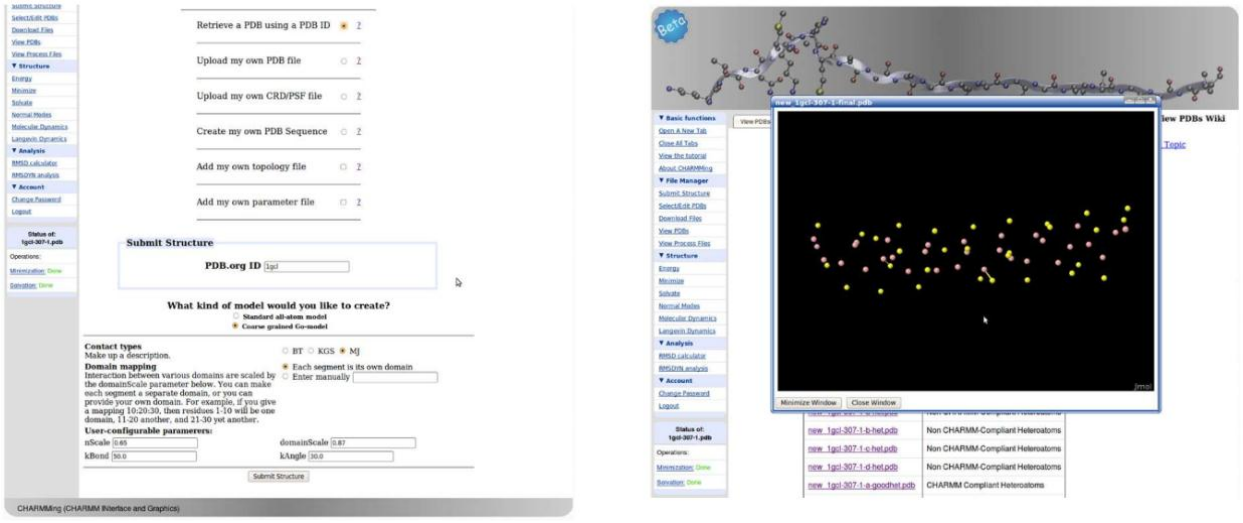

Figure 5.3 Two screen shots taken of the CG model builder GUI from within CHARMMing. On the left, input parameters are specified by the user. On the right, a CG model of an alpha helix is visualized inside of CHARMMing.

5.4 Results

To validate our implementation of the two-site Gō model, we chose to model two small fast folding proteins: the GA module of an albumin binding domain[26, 28, 29] (PDB accession code: 1prb) and the B1 domain of Protein[26, 30, 31] G (PDB accession code: 2qmt). Both of these proteins have been shown by experiment to fold near the semi empirical speed limit of 1 μs,[32]

making them good model proteins for studying new CG methodologies. In solution, 1prb forms a three-helix bundle (Figure 5.4 left), making it an ideal model for an alpha helical system. Likewise, 2qmt provides a suitable test bed for an ideal beta sheet protein (Figure 5.4 right).

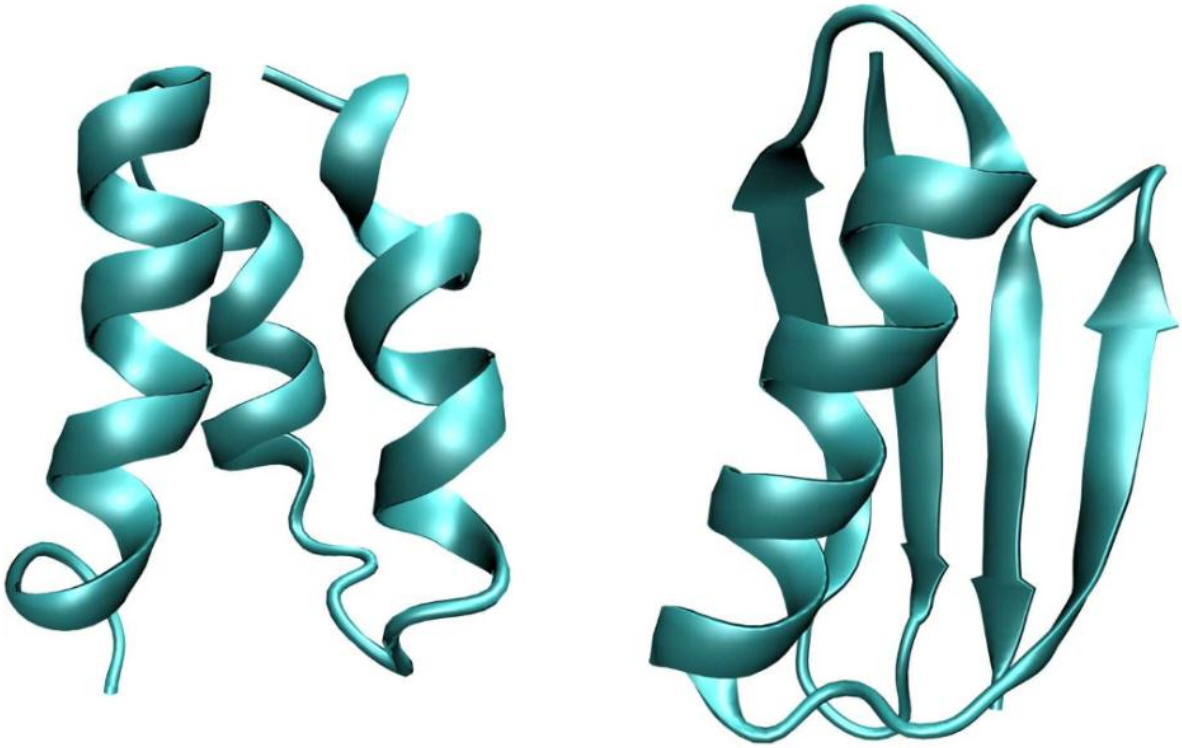

Figure 5.4 Ribbon representations of 1prb, a model alpha helix system (left) and 2qmt, a model beta sheet system (right).

To ensure a proper balance between native side chain interactions, and all other interactions, the 'nScale' parameter must been optimized. Unfortunately, this is a process which must be repeated for distinct model system and cannot be determined *a priori*. Starting from an initial guess of 1, a series of REMD simulations are run. When running REMD simulations, it is important to ensure that there is proper overlap in conformational energies between the separate temperature windows (Figure 5.5). If there is not, the replicas will never exchange, and the REMD will not properly enhance the conformational sampling.

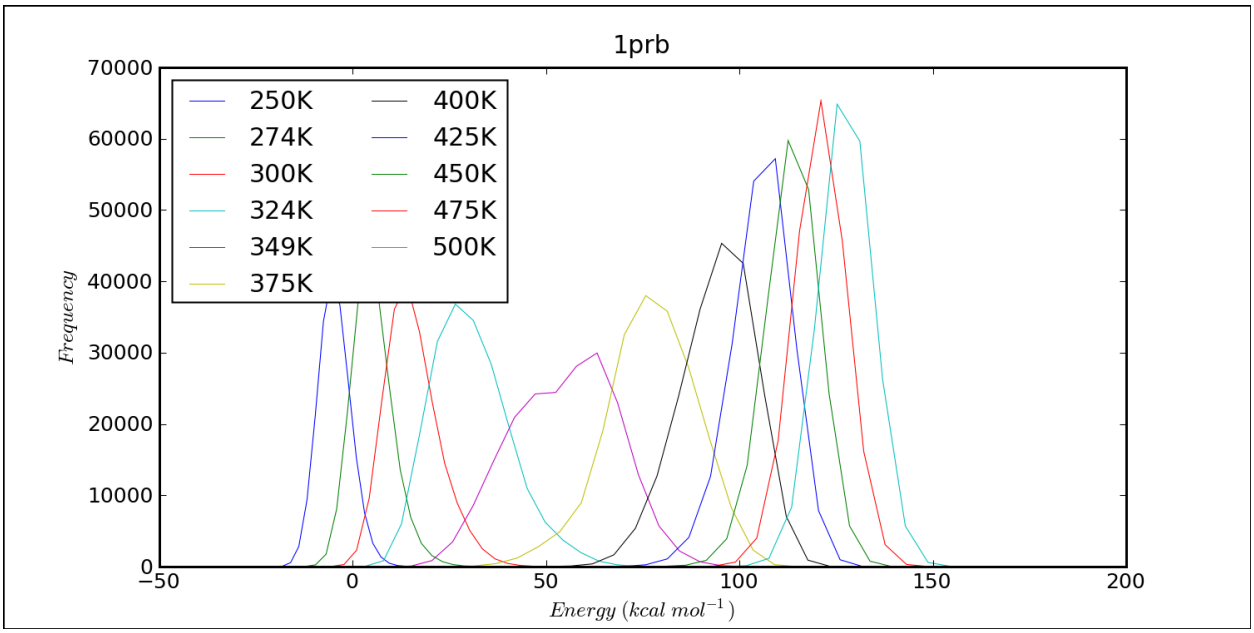

Figure 5.5 A REMD histogram showing the relative occurrence of the potential energy of a given protein conformation. Different replica temperatures appear as different colored curves. From this plot we can clearly see the melting point is between 324 K and 375 K. Also, it is important to note the overlap between adjacent replicas; this overlap indicates that exchange probabilities between replicas will be non-zero.

The first 100 ns of each run are discarded to allow the systems to equilibrate, and the next 900 ns are used to determine the Gibbs free energy between the folded and unfolded states. This problem essentially reduces to a counting problem. Depending upon which reaction coordinate is used and which cutoffs are employed for folded versus unfolded states, the results may differ immaterially. Figure 5.6 shows an example trajectory run for 1prb. On the far right is a histogram which sums up the occurrence of each Q microstate state observed. From this figure we can observe that 1prb may be adequately modeled as a two-state protein. The folded state roughly corresponds to a peak centered about $Q = 0.62$ and the unfolded state corresponds to the peak centered about $Q = 0.37$. A similar, but slightly different plot would follow from a different choice of reaction coordinate such as R_g. When calculating $\Delta G_{folding}$, the choice of cutoffs is somewhat arbitrary, however the histogram in Figure 5.6 illustrates that a cutoff of $Q = 0.55$

might be appropriate for the trajectory depicted, $\Delta G_{folding}$ follows naturally once the relative frequencies of folded and unfolded states are known.

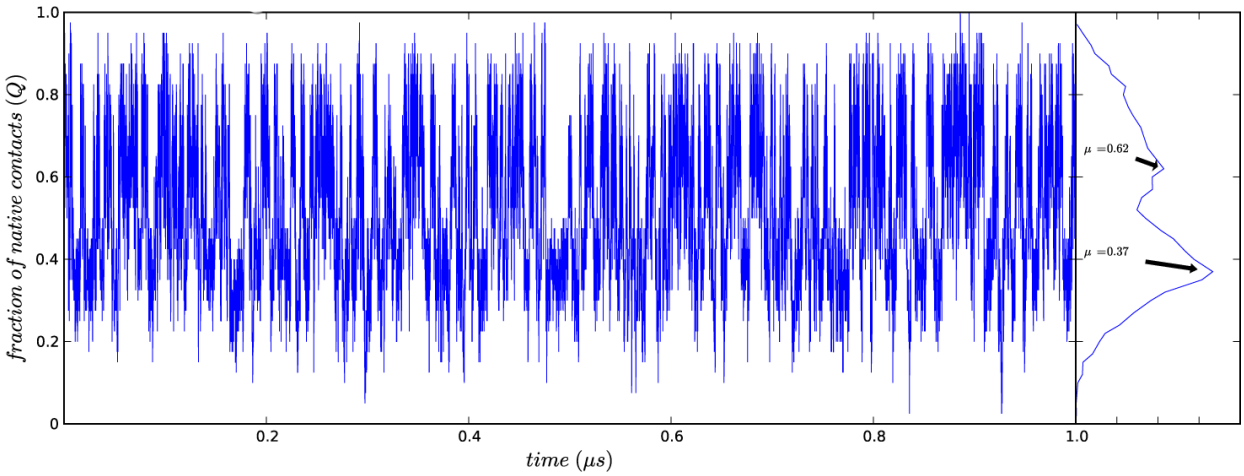

Figure 5.6 The left panel is an example LD trajectory for 1prb, where nScale = 0.95 at 325 K. The right panel is a histogram counting the incidence of fraction of native contacts (100 bins). The folded state peaks about Q = 0.62, and the unfolded state peaks about Q = 0.37.

This procedure is repeated to obtain a melting plot curve as shown in Figure 5.7. When the change in free energy is known as a function of temperature, additional thermodynamic properties may be calculated using the Gibbs-Helmholz equation (eq. 5.8). From this analysis we can determine the protein's melting point (T_m), heat capacity (ΔC_p) and enthalpy of fusion (ΔH_m). Once this analysis is completed, we then iteratively increase or decrease 'nScale' and repeat the analysis until we converge near the physiological value for T_m, near 330 K. For 1prb, the optimal value for 'nScale' is 0.95, which yields T_m = 330 K, ΔH_m = 33.5 kcal mol^{-1} and ΔC_p = 1.13 cal mol^{-1} K^{-1}. For 2qmt the optimal value for 'nScale' is 0.60, which yields T_m = 331 K, ΔH_m = 65.4 kcal mol^{-1} and ΔC_p = 1.77 cal mol^{-1} K^{-1}.

$$\Delta G = [\Delta H_m(1 - \frac{T}{T_m})] - \Delta C_p[(T_m - T) + T\ln(\frac{T}{T_m})] \tag{5.8}$$

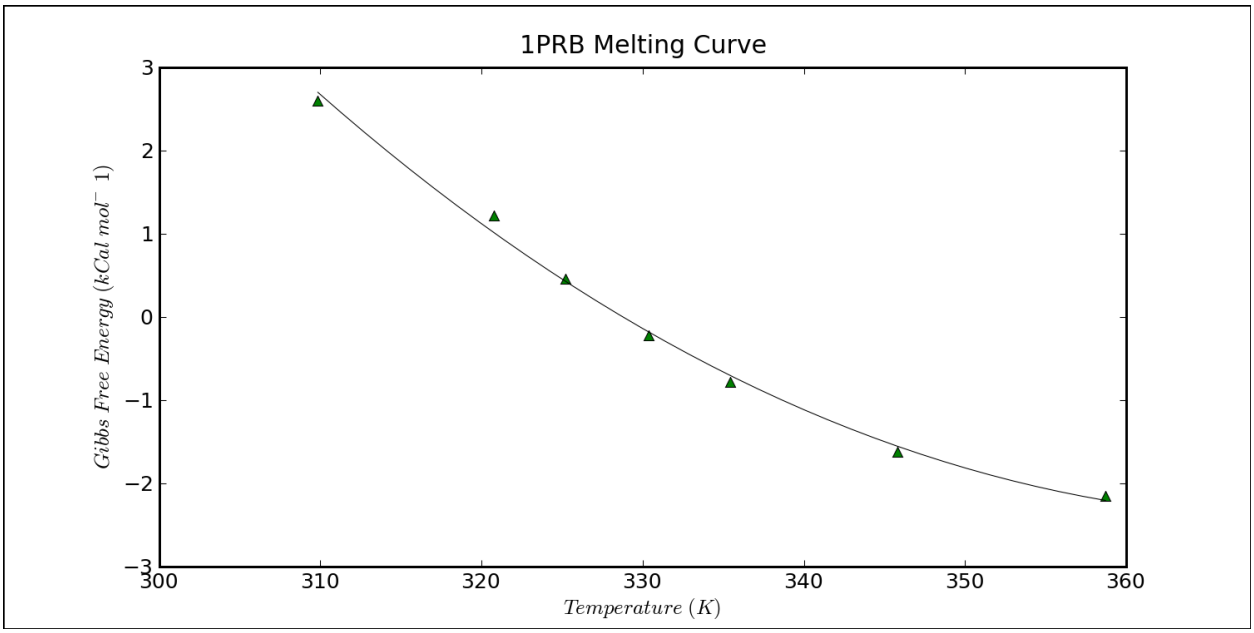

Figure 5.7 Melting point data for the protein 1prb. The data points are fit to eq. 5.8 via least squares, and allow the calculation of several important thermodynamic properties, such as melting point, heat capacity and enthalpy of fusion.

Other techniques may be employed to analyze the formation and likelihood of various secondary structure elements. One way of doing this is by looking at the probability of native contacts between specific side chain pairs. This analysis may be performed for a specific ensemble (such as all folded 1prb structures as determined above) or for the global ensemble. An average contact map (Figure 5.8) is a good way of creating a graphical 'finger print' of a protein ensemble. Similar techniques may also be employed to resolve native contact probabilities as a function of time in the trajectory. In this manner a folding event may be more fully characterized by determining which regions of the protein fold first to nucleate the global folding pathway. These results may be directly comparable to experimental observations.

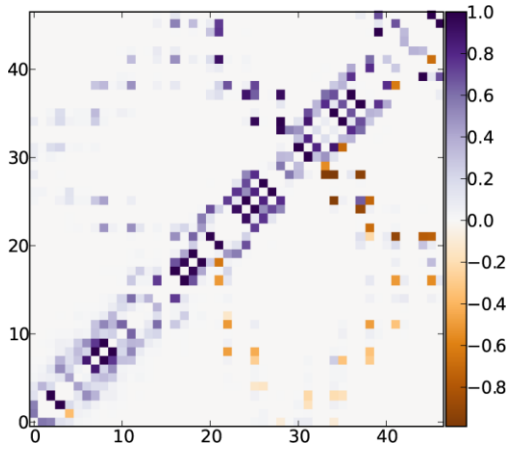

Figure 5.8 The average contact map for the global ensemble of 1prb conformers. The axes denote the side chain residue index as they appear in the PDB crystal structure. This is a symmetric matrix about the diagonal; however the lower diagonal color native contacts orange. Darker purple and orange denotes a higher probability contact formation, whereas lighter colors correspond to more disordered regions of the protein.

5.5 Future Plans

We are currently working to implement additional CG models using our existing framework. The transferable lipid model of Marrink *et al.*[22] has enjoyed great popularity and success in modeling membranes, lipoproteins and diffusion effects. Unfortunately this model is currently only available to users of GROMACS software simulation package.[33] Making this widely used method available to members of the CHARMM community is a task specifically suited to our recently developed framework. We are also interested in extending our Gō-like model to be able to incorporate CG solvent particles and eventually protein-protein and protein lipid interactions. Early implementations of such a model a model have been completed (but not yet validated) using modifications to the Marrink lipid model which were originally proposed by Bond *et al.*[23] Extending our model to be able to incorporate interactions from structures that do not share a crystal structure is critical, as limiting our potential model space to structures only present in the PDB is a major drawback to our current approach. Removing this barrier will make our model applicable to a much broader set of possible applications.

5.6 References

1. A. R. Fersht, *Structure and Mechanism in Protein Science: A Guide Enzyme Catalysis and Protein Folding*, 2nd ed. (Freeman, New York, 1999).
2. F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333-366 (2006).
3. D. Thirumalai, D. K. Klimov and R. I. Dima, Curr. Opin. Struc. Biol. **13**, 146-159 (2003).
4. M. Delarue and P. Dumas, Proc. Natl. Acad. Sci. USA **101**, 6957-6962 (2004).
5. C. Levinthal, J. Chim. Phys. PCB **65**, 44-45 (1968).
6. Y. Duan and P. A. Kollman, Science **282**, 740-744 (1998).
7. E. P. O'Brien, G. Ziv, G. Haran, B. R. Brooks and D. Thirumalai, Proc. Natl. Acad. Sci. USA **105**, 13403-13408 (2008).
8. B. T. Miller, R. P. Singh, J. B. Klauda, M. Hodošček, B. R. Brooks and H. L. Woodcock, J. Chem. Inf. Model **48**, 1920-1929 (2008).
9. Y. Ueda, H. Taketomi and N. Gō, Biopolymers **17**, 1531-1548 (1978).
10. S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 623-644 (1996).
11. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, J. Comput. Chem. **4**, 187-217 (1983).
12. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, J. Comput. Chem. **30**, 1545-1614 (2009).

13. Python is a freely availible computer programming language originally written by Guido van Rossum in the C programming language for the rapid prototyping of computer code. More information is available from www.python.org.

14. Django is a freely availible framework for web based application development. It is written in the Python language, more information is available from www.djangoproject.com.

15. M. Heinig and D. Frishman, Nucleic Acids Res. **32**, W500-W502 (2004).

16. H. Li, A. D. Robertson and J. H. Jensen, Proteins **61**, 704-721 (2005).

17. J. Konc and D. Janežič, Nucleic Acids Res. **38**, W436-W440 (2010).

18. K. N. Kirschner and R. J. Woods, Proc. Natl. Acad. Sci. USA **98**, 10541-10545 (2001).

19. S. Jo, T. Kim, V. G. Iyer and W. Im, J. Comput. Chem. **29**, 1859-1865 (2008).

20. D. K. Klimov and D. Thirumalai, J. Chem. Phys. **109**, 4119-4125 (1998).

21. D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. USA **97**, 2544-2549 (2000).

22. S. J. Marrink, A. H. de Vries and A. E. Mark, J. Phys. Chem. B **108**, 750-760 (2003).

23. P. J. Bond and M. S. P. Sansom, J. Am. Chem. Soc. **128**, 2697-2704 (2006).

24. Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141-151 (1999).

25. T. Veitshans, D. Klimov and D. Thirumalai, Fold. Des. **2**, 1-22 (1997).

26. R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. USA **107**, 1088-1093 (2010).

27. C. C. Mello and D. Barrick, Proc. Natl. Acad. Sci. USA **101**, 14102-14107 (2004).

28. T. Wang, Y. Zhu and F. Gai, J. Phys. Chem. B **108**, 3694-3697 (2004).

29. S. Takada, Proteins **42**, 85-98 (2001).

30. E. L. McCallister, E. Alm and D. Baker, Nat. Struct. Mol. Biol. **7**, 669-673 (2000).

31. S. Kmiecik and A. Kolinski, Biophys. J. **94**, 726-736 (2008).

32. S. J. Hagen, J. Hofrichter, A. Szabo and W. A. Eaton, Proc. Natl. Acad. Sci. USA **93**, 11615-11617 (1996).

33. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701-1718 (2005).

CHAPTER 6

CONCLUDING REMARKS

In chapter 2 we modeled the electronic ground state decoposition of formaldehyde to molecular products using high order coupled cluster techniques [up through fifth order CCSDTQ(P)] in conjunction with the correlation-consistent family of polarized valence (cc-pV*n*Z), augmented correlation-consistent polarized valence (aug-cc-pV*n*Z), correlation-consistent polarized core-valence (cc-pCV*n*Z) and augmented correlation-consistent polarized core-valence (aug-cc-pCV*n*Z) (*n* = D, T, Q, 5, 6) basis sets were employed for energy calculations. Via focal point analysis, we confirmed that molecular products hydrogen and carbon monoxide lie below formaldehyde by 1.81 kcal mol^{-1}, a result which agrees extremely well with the best available experimental information, 2.09 kcal mol^{-1} from the active thermochemical tables (ATcT) of Ruscic *et al.* The barrier to molecular products was determined to be 80.82 kcal mol^{-1}, a result which stands between the two most widely accepted experimental values, and disagrees with the most recent *ab initio* study by ~1.5 kcal mol^{-1}. These are by far the most rigorous electronic structure calculations ever performed on this important model combustion system.

Furthermore, by incorporating anharmonic effects along with corrections for relativity and Born-Oppenheimer assumptions into the computed thermochemistry, we are able to resolve a long standing disparity between the best theoretically determined and most accurately experimentally observed barrier to molecular products. Individually each of these auxiliary calculations contributes a minimal amount to the final computed energy. However, taken

together these corrections shift the best theoretical result by nearly 1 kcal mol^{-1}, a very significant amount for one of combustion chemistry's benchmark systems.

In chapter 3 we report the most accurate theoretical calculations ever performed on the hydroxymethylene system, a recently synthesized, highly unstable carbine. AE-CCSD(T)/cc-pCVQZ was used to generate a highly accurate quartic force field, which in turn was used with the discrete variable representation of the full Eckart-Watson Hamiltonian to variationally compute the vibrational band origins of hydroxymethylene (HCOH). Multireference coupled cluster (Mk-MRCCSD) was also used to model both the vertical and adiabatic excitations from the S_0 electronic state to the S_1 electronic state. Experimentally determined IR and UV/Vis spectra were in strong agreement with their theoretical counterparts, confirming that HCOH had in fact been synthesized and trapped for the first time ever. Finally, the elusive nature of HCOH was determined to come from its propensity to undergo quantum tunneling through a large barrier, even at extremely low temperatures. Using semiclassical theory in conjunction with a zero-point corrected intrinsic reaction path, the half-life of HCOH was predicted to be 2.1 h, in very good agreement with the experimentally observed half-life of 2 h. HCOD however was found to be stable under all conditions.

Much previous experimental and theoretical effort has been expended upon this molecule as a potential precursor towards formation of interstellar glycoaldehyde. Both formaldehyde and glycoaldehyde have been observed in the interstellar medium, and it has been postulated that hydroxycarbene is the missing ingredient. Our tunneling analysis suggests another mechanism is responsible, as the short half-life of HOCH toward unimolecular decay and the low collision frequencies in interstellar environments make the detection of non-deuterated HCOH in interstellar space unlikely.

In chapter 4 we exhaustively characterize the spectroscopic properties of the HSiN–HNSi system. This is a triatomic molecule which has drawn interest as an interstellar molecule. Previous studies have neglected the effects of anharmonicity upon its spectral properties or have utilized density functional theory, a method which is wholly inappropriate for determining highly accurate thermochemical properties. In this work we apply high order coupled cluster methods [CCSD(T) and CCSDT(Q)] in conjunction with large correlation consistent basis sets to accurately model properties such as dipole moments, fundamental vibrational frequencies and rotational constants. The most sophisticated method we use predicts the HSiN isomer to lie 64.7 kcal mol^{-1} above the global minimum on the ZPVE corrected surface. This same method also yields highly accurate barrier heights; the forward reaction barrier is predicted to be 9.7 kcal mol^{-1}, whereas the reverse reaction barrier is predicted to be 74.4 kcal mol^{-1}. Our computed spectral properties are in very good agreement with experiment, and our analysis indicates this molecule is not a good candidate to undergo quantum tunneling.

In chapter 5 we successfully implemented a Gō-like model for the CHARMM molecular dynamics simulation package. CHARMMing was used to provide a graphic front end to facilitate rapid parameterization and refinement of the models. During the course of the implementation, a general framework for CG model building was established, and will be used to extend the functionality of the Gō model and into implement additional models such as those based upon the work of Marrink. To validate our implementation, thermodynamic properties such as melting point, heat capacity and enthalpy of fusion were calculated for two small model proteins. Our results, which are consistent with available experimental data, confirm the successful implementation of the two-cite Gō model.