BENCHMARKING DENSITY FUNCTIONAL THEORY AND ITS APPLICATIONS TO NON-COVALENT INTERACTIONS IN BIOLOGICAL SYSTEMS

by

JACQUELINE CONNIE HARGIS

Under the Direction of Henry F. Schaefer III

ABSTRACT

Density functional theory is used to delve into a variety of biological systems that possess non-covalent interactions. Non-covalent interactions are known to be difficult to examine using theoretical methods. *Ab initio* methods can be used to benchmark density functionals, therefore accuracy can be obtained for systems of a certain chemical nature. Works studied here possess intramolecular hydrogen bonding, intermolecular hydrogen bonding, and π-π interactions. The remarkably short intramolecular hydrogen bond of malonaldehyde, the mechanistic nature of the double proton transfer in the formamide dimer, and the intercalation of a benzo[*a*]pyrene diol epoxide into DNA are examined.

INDEX: Density Functional Theory, Hydrogen Bonding, Benchmarking, π-π Interactions, Non-covalent Interactions, DNA, Malonaldehyde, Formamide Dimer

BENCHMARKING DENSITY FUNCTIONAL THEORY AND ITS APPLICATIONS TO NON-COVALENT INTERACTIONS IN BIOLOGICAL SYSTEMS

by

JACQUELINE CONNIE HARGIS

B. A., New College of Florida, 2007

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

BENCHMARKING DENSITY FUNCTIONAL THEORY AND ITS APPLICATIONS TO NON-COVALENT INTERACTIONS IN BIOLOGICAL SYSTEMS

by

JACQUELINE CONNIE HARGIS

Major Professor: Henry F. Schaefer III

Committee: Nigel G. Adams
Geoffrey D. Smith

Electronic Version Approved:

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

ACKNOWLEDGEMENTS

I would like to acknowledge my family and friends for their everlasting support. I greatly appreciate the guidance and inspiration provided by the members of the Center for Computational Quantum Chemistry. Specifically, Steven E. Wheeler has provided me with valuable direction and leadership that has helped me evolve into a better scientist. Finally, I would like to acknowledge my graduate advisor, Henry F. Schaefer, for the opportunities he has provided me during my graduate research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Non-covalent interactions are involved in a variety of chemical processes that occur in biological systems. These interactions are vital for DNA duplexes, protein folding interactions, ligand-receptor complexes, and protein-substrate binding.[1] Common non-covalent interactions are hydrogen bonding, ionic bonds, van der Waals forces, dipole-dipole interactions, charge transfer intereactions, π-π interactions and hydrophobic interactions.[2] These interactions are prevalent in many biological processes. The wealth of applications involving non-covalent interactions makes this an essential field to explore at the chemical level.

Non-covalent interactions have been a challenge to study both experimentally and theoretically. Computational chemistry has been a valuable tool to study systems that possess these characteristics. The evolution of Density Functional Theory (DFT) has impacted the study

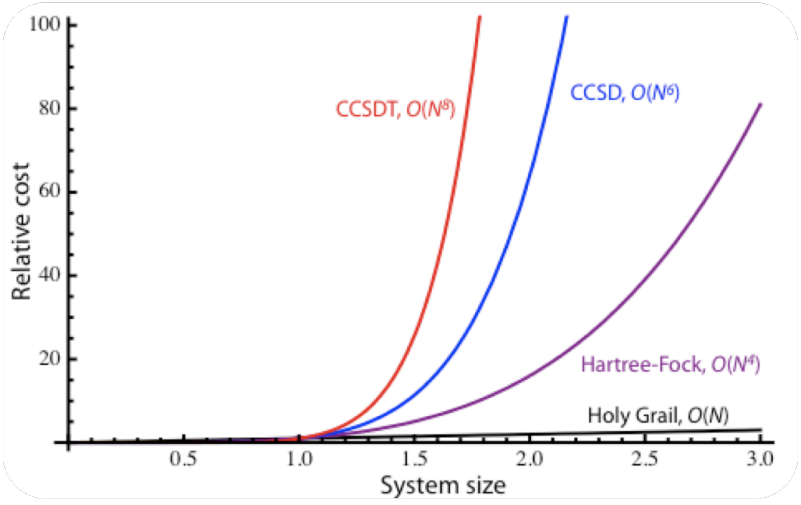


Figure 1.1 Computational scaling for *ab initio* methods is shown. As you increase the computational level, you essentially hit a computational wall. Density functional theory methods often scale as N^3, however depending on the functional computational time can vary.

of biological systems. Challenges arise when using *ab initio* methods to study biological systems because they scale as a function of the number of atoms in the system. This creates a "computational wall." The relative scaling of DFT compared to highly accurate *ab initio* methods make it easier to study larger systems that most often include biological systems.

Over the last thirty years, DFT has become an increasingly popular method in computational chemistry.[3,4] The relatively cheap computational cost has allowed computational chemistry to explore larger systems with increased accuracy. However, DFT functionals aren't developed in a systematic way and often have many parameters. This creates inconsistencies between different functionals. There have been many investigations on DFT functionals to determine their strengths and weaknesses. The nature of the chemical system dictates the strengths and weaknesses of the functionals.

Benchmarking DFT functionals is an essential part to confirm the erratic nature of these methods. By creating a model system, highly accurate *ab initio* methods can be used to understand the quirks of the DFT functional. Being able to predict systematic errors of DFT functionals is a valuable tool. *Ab initio* methods can be used in extrapolation schemes to nail down highly accurate answers for a model system, unlike DFT. *Ab initio* methods have been shown to have systematic basis sets and levels of theory. The systematic nature of *ab initio* methods makes extrapolations a valuable tool in a computational chemist's arsenal. Comparing the highly accurate answers obtained from extrapolation methods to results obtained from DFT allows you to anticipate errors seen in a particular functional.

Computational chemistry can also be used to understand fundamental bonding principles and trends in organic chemistry. Mechanistic information can be determined by different approaches that have been developed to describe the complete picture. Creating systematic models and methods can help clarify a picture to make a multifarious system easier to understand. This can lead to predictions about more complex systems. Creating a good model system proves to be a valid technique when evaluating properties of a chemical system.

The following research focuses on creating an accurate picture of non-covalent interactions by benchmarking systems that have been difficult to study both experimentally and theoretically. Due to computational cost rising rapidly as the size of the system increases, challenges arise when trying to examine biologically relevant systems. However, by combining highly accurate *ab initio* methods with DFT, fascinating results can be obtained for larger systems with biological relevance.

One type of non-covalent interaction of interest is hydrogen bonding. They can be intermolecular or intramolecular in nature. Both intramolecular and intermolecular hydrogen bonding are abundant in biological systems. They are commonly seen in water, DNA, proteins, polymers, and many more applications. The standard textbook definition of a hydrogen bond is an interaction between partially positive hydrogen atoms and partially negative oxygen, fluorine, and nitrogen atoms on adjacent molecules that possess a large dipole moment and stronger-than-average dipole-dipole interactions.[5] This definition has been contested throughout history[6-9] and recently.[10-12]

In 1912, Moore and Winmill[6] described hydrogen bonding as a "weak union." The description "hydrogen bond" appeared after 1930 by Pauling.[8] He considered this interaction to be primarily ionic in nature. Pimentel and McClellan[9] published a more general definition in

1960: “A hydrogen bond exists between the functional group *A*-H, and an atom or a group of atoms, *B*, in the same or different molecules when 1) there is an evidence of bond formation (association, or chelation), 2) there is an evidence that this new bond linking *A*-H and *B* specifically involves a hydrogen atom already bonded to *A*.” This general definition seems suitable due to the variety of phenomena that have been attributed to hydrogen bonding.

The definition of hydrogen bonding seems to keep evolving around a circle of debate. Significant covalent character of a hydrogen bond was determined by an Inelastic Compton X-ray scattering experiment performed by Isaacs and coworkers.[11] The first coordination shell of liquid water was seen in 2004, yet again stirring up the picture of intermolecular bonding in water.[12] Most molecules in liquid water see two hydrogen-bonded configurations, unlike ice that possesses four hydrogen-bonded tetrahedral structures. Recently Zewail described the perplexity of the hydrogen bond in a suitable manner: “Remarkably, this transfer of a small particle appears deceptively simple, but it is in fact complex in nature. For the most part, the dynamics cannot be described by a classical picture and the process involves more than one nuclear motion. For example, the transfer may occur by tunneling through a reaction barrier and a quantum description is necessary; the hydrogen bond is not isolated as it is a part of a chemical bond and in many cases the nature of the bond, “covalent” and/or “ionic” in Pauling’s valence bond description, is difficult to characterize.” Due to the quandaries that still surround the picture of the hydrogen bond, there is opportunity for research to improve our understanding of the marvel.

A complexity in studying hydrogen bonding experimentally lies in the field of x-ray crystallography. X-ray crystallography cannot view hydrogen atoms so their positions are predicted using common valence rules, if at all. Many experimental works characterize the hydrogen bond by the distance between the two non-hydrogen atoms. Computational chemistry

has an advantage in that hydrogen atoms can be explicitly studied. When examining geometries of limited size, computations are inexpensive and accurate. Advantages present themselves when studying hydrogen bonding using computational chemistry.

Accurate determination of thermochemistry has proved to be a successful application of computational chemistry. Energetics of hydrogen transfer can readily be studied using these methods. Both intramolecular and intermolecular hydrogen transfers are of interest in many biological systems. This work examines one of the shortest intramolecular hydrogen bonds. Examining the intramolecular proton transfer of malonaldehyde and its derivatives, allows one to understand some of the factors that can be used to lower the barrier to proton transfer and shorten the intramolecular hydrogen bond. Intermolecular proton transfer is also understood using qualitative and quantitative methods in this research.

An abundant non-covalent interaction in DNA is the π-π interaction, also known as the aromatic interaction. DNA bases are stacked upon each other where the forces between the aromatic rings on top of one another create an interaction allowing the DNA to retain its structural integrity. One model system used to study these interactions is the benzene dimer. There has been a significant amount of benchmarking done on stacked arene dimers[13-17] using various DFT functionals that has resulted in a blueprint for studying larger π stacking compounds.

The first chapter presents a systematic study examining the intramolecular hydrogen bond in malonaldehyde and its derivatives. A challenge was presented[18] to find the shortest intramolecular hydrogen bond. Additionally, the proton transfer in malonaldehyde has a two well symmetrical potential. It is a goal of this work to find a system that eliminates the two well potential into a single well. The DFT functional, B3LYP, is known to produce reasonably

accurate geometries, and it is confirmed in this work[19] once again. However, B3LYP systematically underestimates barrier heights, which is a significant problem when examining barriers as small as the ones for these systems. Another difficulty is that some of the systems are too large to study using *ab initio* methods, therefore DFT is employed to fully explore the derivatives of malonaldehyde. Malonaldehyde has nine atoms; however as you increase the size of the substituents the computational costs rapidly rise. To solve this problem, the highly accurate focal point method is applied on eight of the smallest systems. Incorporating B3LYP and focal point computations together formulates an empirical method. The empirical method is used to predict barriers for systems too large to perform the robust focal point method on. By combining B3LYP geometries with accurately predicted barriers, simple organic trends are confirmed while searching for the shortest intramolecular hydrogen bond.

Another examination of hydrogen bonding in this work relates to the intermolecular double proton transfer in the formamide dimer. A similar approach is taken showing that *ab initio* methods and B3LYP can be used harmoniously to create a cohesive picture of a chemical problem. The intrinsic reaction coordinate is computed for the double proton transfer that occurs in this system. Several locations along the intrinsic reaction coordinate were computed using the focal point method. The accuracy of B3LYP was portrayed in most areas of the reaction, except in the product region. It is fascinating to understand the errors in the B3LYP functional due to the vast quantity of papers published that utilize it. Here the error is investigated in a systematic way that elucidates where the flaw of B3LYP occurs related to the formamide dimer.

This work also presents a unique computational approach that exhibits a detailed understanding of the mechanism involved in the double proton transfer in the formamide dimer. It illustrates the parts of the reaction where structural rearrangement and electronic

rearrangement occur. Information is presented showing changes of bond lengths along the reaction coordinate. Also, the electrostatic potential was computed to create a better understanding of the mechanistic picture.

The final work centers on the importance and relationship of two different non-covalent interactions: hydrogen bonding and π-π interactions. Polycyclic aromatic hydrocarbons (PAH) are known to be tumorogenic. A PAH of interest is benzo[*a*]pyrene due to its importance in biological and atmospheric interactions. When benzo[*a*]pyrene enters the body it gets metabolized into a diol epoxide, which is known to be tumorogenic. Once the diol epoxide is formed, it slides into DNA and attaches via a ring opening mechanism. However, this mechanism at the molecular level is still not understood. This chapter discusses the most likely orientation a benzo[*a*]pyrene diol epoxide slides into DNA and which DNA base it is most likely to stack with. It is heavily dependent on the number of hydrogen bonds present between the benzo[*a*]pyrene diol epoxide and the DNA base pair it is stacked on. The orientation also depends on the amount of aromatic area overlapping which strengthens the overall π-π interactions present in the complex.

Benchmarking has been done on π-π interactions using stacked arene dimers as a model.[13,14,17] Highly accurate coupled cluster calculations have been performed by Sinnokrot and Sherrill[20] on the benzene dimer which has provided a benchmark for other DFT works.[16,21,22] Benchmarking for many DFT functionals has been performed. Traditional functionals such as B3LYP have proven deficient for these systems. It has been found that several Minnesota functionals developed by Donald Truhlar work well for π stacking systems. The M05-2X and M06-2X functionals have been found to be very accurate when used with a computational grid of appropriate size.[23-26]

The strategy of combining DFT with *ab initio* methods, particularly the focal point method can be a successful way to investigate biologically relevant chemical systems. Using a range of computational methods allows for accurate solutions to be obtained. However, as in all science, cost is an important factor. Biological systems can be vast in size, too large and too expensive to compute using *ab initio* methods. It is possible to save costs by benchmarking, therefore sacrificing minimal accuracy.

The importance of non-covalent interactions is well known in chemistry and biology. The following work utilizes several computational chemistry methods to understand these complex interactions. In depth computational investigations allow fascinating results to be obtained for systems with intermolecular hydrogen bonding, intramolecular hydrogen bonding, and π-π interactions. Intriguing mechanistic information is obtained about these relevant non-covalently bonded biological systems.

REFERENCES

(1) Williams, D. H.; Westwell, M. S. *Chem. Soc. Rev.* **1998**, *27*, 57.

(2) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology* Freeman and Company: New York, 2000.

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

(4) Argaman, N.; Makov, G. *Am. J. Phys.* **2000**, *68*, 69.

(5) Gilbert, T. R.; Kirss, R. V.; Davies, G. *Chemistry: The Science in Context*; W. W. Norton & Company Inc.: New York, 2004.

(6) Moore, T. S.; Winmill, T. F. *J. Chem. Soc.* **1912**, *101*, 1635.

(7) Latimer, W. M.; Rodebush, W. H. *J. Am. Chem. Soc.* **1920**, *42*, 1419.

(8) Pauling, L. *J. Am. Chem. Soc.* **1931**, *53*, 1367.

(9) Pimentel, G. C.; McClellan, A. L. *The Hydrogen Bond*; Free-man: San Francisco, 1960.

(10) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293.

(11) Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. *Phys. Rev. Lett.* **1999**, *82*, 600.

(12) Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. Å.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A. *Science* **2004**, *304*, 995.

(13) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

(14) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.

(15) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(16) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *J. Chem. Theory Comput.* **2008**, *4*, 1996.

(17) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.

(18) Belot, J. A.; Clark, J.; Cowan, J. A.; Harbison, G. S.; Kolesnikov, A. I.; Kye, Y. S.; Schultz, A. J.; Silvernail, C.; Zhao, X. G. *J. Phys. Chem. B* **2004**, *108*, 6922.

(19) Hargis, J. C.; Evangelista, F. A.; Ingels, J. B.; Schaefer, H. F. *J. Am. Chem. Soc.* **2008**, *130*, 17471.

(20) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(21) Wheeler, S. E.; McNeil, A. J.; Müller, P.; Swager, T. M.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, *132*, 3304.

(22) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10854.

(23) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.

(24) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*, 034111.

(25) Gräfenstein, J.; Cremer, D. *J. Chem. Phys.* **2007**, *127*, 164113.

(26) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *5*, 2301.

CHAPTER 2

SHORT INTRAMOLECULAR HYDROGEN BONDS: DERIVATIVES OF MALONALDEHYDE WITH SYMMETRICAL SUBSTITUENTS[1]

[1] Reproduced with permission from Hargis, J.C.; Evangelista, F. A.; Ingels, J.B. Schaefer, H.F. *J. Am. Chem. Soc.*, **2008**, *130*, 17471. Copyright 2008 American Chemical Society.

ABSTRACT

A systematic study of various derivatives of malonaldehyde has been carried out to explore very short hydrogen bonds (r_{OO} < 2.450 Å). Various electron withdrawing groups, including CN, NO_2, and BH_2 have been attached to the central carbon atom, C_2. To C_1 and C_3, strong electron donors and/or sterically hindered substituents were used to strengthen the intramolecular hydrogen bond, including but not limited to NH_2, $N(CH_3)_2$, and $C(CH_3)_3$. Seven molecules (Figure 2) were found to have extremely short intramolecular hydrogen bonds.

The chemical systems investigated are intriguing due to their low energetic barriers for the intramolecular hydrogen atom transfers. Energy barriers were predicted using correlated methods including second-order perturbation theory and coupled cluster theory in conjunction with the Dunning hierarchy of correlation consistent basis sets, cc-pV*X*Z (*X*=D, T, Q, 5). Focal point analyses allowed for the barriers to be evaluated at the CBS limit including core correlation and zero-point vibrational energy corrections.

B3LYP energies are benchmarked against highly accurate correlated energies for intramolecular hydrogen bonded systems. The focal point extrapolated method, including coupled cluster full triple excitation contributions, gives a hydrogen transfer barrier for malonaldehyde of ~ 4 kcal mol^{-1}. We describe two compounds with extremely low barriers; nitromalonamide (0.43 kcal mol^{-1}) and 2-borylmalonamide (0.60 kcal mol^{-1}). An empirical relationship was drawn between the B3LYP energetic barriers and the predicted coupled cluster barriers at the CBS limit. By relating these two quantities, barrier heights may be estimated for systems too large to presently use highly correlated electronic structure methods.

INTRODUCTION

Intramolecular hydrogen bonding interactions are essential in biochemical reactions and enzymatic processes.[1,2] Malonaldehyde (MA) has been studied extensively not only because of the system's biological connections, but additionally because it is the prototypical model for intramolecular hydrogen bonding. Malonaldehyde has a symmetrical double well potential energy surface, with two equivalent C_s minima and a C_{2v} transition state (See Figure 1 for atom numbering in MA.).[3-17] Specifically, intramolecular hydrogen bonding is displayed between the two electronegative atoms, which are designated proton donors and proton acceptors.[1] The 2,4-diketones and *cis*-enoicacids are two classes of molecules where some of the shortest intramolecular hydrogen bonding distances occur.[18] The low energy barrier for proton transfer in enolized 2,4-diketones, including malonaldehyde, is usually explained by π delocalization.[19-21] In fact, when hydrogen atom transfer occurs from a equilibrium geometry to a transition state, there is significant π delocalization over the six-membered cyclic transition state geometry.[22] The addition of electron withdrawing and/or elecron donating substituents on the 2,4-diketone affects the delocalized π system, hence lowering or raising the barrier height of the symmetric double well energy potential.

Malonaldehyde and its derivatives have been the focus of numerous experimental[7,13-15,23-30] and theoretical investigations.[22,31-35] In 1985, Frisch *et al.*[8] performed the first modern theoretical computations on malonaldehyde. At the SCF level of theory using a 6-31G** basis set, the hydrogen bonded distance H···O for MA was much too large, 1.88 Å, compared to E.B. Wilson's definitive experimental result[7] of 1.68 Å for the deuterated compound. With the MP2 method using a 6-31G** basis set, a hydrogen bond distance comparable to experiment (1.694

Å) was predicted.[8-10] In 2003, Mil'nikov and coworkers computed the energy barrier at the CCSD(T)/aug-cc-pVTZ level to be 3.8 kcal mol^{-1}, for a time the best *ab initio* value available.[11] Recently, Wang, Braams, Bowman, Carth, and Tew determined an *ab initio* potential energy surface of malonaldehyde at the basis set limit using CCSD(T) electronic energies. They computed the barrier to be 4.1 kcal mol^{-1}, the most reliable result to date.[16] However, the most remkarkable feature of the pioneering work of Bowman and co-workers is their full-dimensional quantum mechnical treatment of the malonaldehyde tunneling splittings.

Many other studies have examined acetyl acetone, which is malonaldehyde with methyl groups substituted on C_1 and C_3 (see Figure 1 for atom labels), over the years via *ab initio* and experimental methods.[33,34,36-40] For example, Dannenberg and Rios reported results on the various conformations of acetyl acetone in 1994.[22] They performed computations at the HF and MP2 levels of theory using basis sets up to D95++**. They confirm in their study that the conjugated enol of acetyl acetone has a double well potential with respect to the hydrogen atom.[22]

Various symmetrically substituted derivatives of malonaldehyde have been investigated.[18,37,41-45] The halogenated, amino, and nitro-substituted derivatives examined by Buemi and Zuccarello[41,42] gave rise to several conclusions reexamined in this research. Electron withdrawing groups substituted on the central atom C_2 shortened the intramolecular hydrogen bond. When electron withdrawing groups, including halogens, are substituted on the symmetrical carbons, C_1 and C_3, the intramolecular hydrogen bond is lengthened. Electron donating groups substituted on the symmetrical carbons also contribute to the shortening of the intramolecular hydrogen bond. Because hydrogen bonds cannot be straightforwardly examined by X-ray crystallography, these bonds are typically characterized the distance between O_4 and O_5. One of

the shortest intramolecular hydrogen bonds known to date is of nitromalonamide, where C_1 and C_3 are substitued with amino groups and C_2 is substituted with a nitro group. For nitromalonamide the O···O separation is predicted to be 2.394 Å using the B3LYP method with the cc-pVTZ basis set.[44]

In this study we initially apply Density Functional Theory (DFT) to examine several symmetrical derivatives of MA and characterize the intramolecular hydrogen bond by the distance between the oxygen atoms. We also applied DFT to study the barrier heights for hydrogen transfer. Highly correlated electronic structure methods are then used for focal point extrapolations nailing down accurate barrier heights of the intramolecular proton transfers for the eight simplest systems. An empirical scheme was developed to approximate the barrier heights at highly correlated levels using only DFT energy barriers, which are predicted to be too low. We investigate whether or not it is possible to have a substituted malonaldehyde system without a barrier.

THEORETICAL METHODS

Energies, optimized structures, transition state structures, and vibrational frequencies were initially determined using the B3LYP generalized gradient approximation (GGA) exchange correlation functional. B3LYP is a combination of Becke's exchange functional, the 3-parameter HF/DFT hybrid exchange function (B3),[46] and the dynamical correlation functional of Lee, Yang, and Parr (LYP).[47] All computations were performed using double-ζ-quality basis sets with polarization and diffuse functions, designated DZP++.

The DZP++ basis sets were constructed by augmenting the Huzinaga-Dunning[48,49] sets of of contracted double- ζ Gaussian functions with one set of *p*-type polarization functions for each H atom and one set of five *d*-type polarization functions for each B, C, N, and O atom [α_p(H) =

1.0, α_d(C) = 0.75, α_d(O) = 0.85, α_d(N) = 0.80, α_d(B) = 0.70]. To complete the DZP++ basis, even-tempered *s* and *p* diffuse functions were centered on each heavy atom. The even-tempered orbital exponents were determined according to the prescription of Lee.[50]

The final DZP++ set contains six functions per H atom (5*s*1*p*/3*s*1*p*) and 19 functions per B, C, N, or O atom (10*s*6*p*1*d*/5*s*3*p*1*d*), yielding a total of 119 contracted functions for the parent molecule, malonaldehyde. All structures were optimized using analytic gradients with tight convergence criteria. Vibrational frequency evaluations were done on all structures, and no scaling factors were applied. Numerical integrations were performed using Q-Chem3.1[51] with a grid consisting of 75 radial shells and 302 angular points.

A focal point analysis using the HF, MP2, CCSD, CCSD(T), and CCSDT levels of theory was performed using the correlation consistent basis sets of Dunning[52] (cc-pVXZ, X = D, T, Q, 5) to yield values extrapolated to the basis set limit. Geometries used for the focal point analysis were generally computed using the aformentioned DFT method. However, for malonaldehyde and its transition state geometries were computed at the far more complete CCSD(T)/cc-pVQZ level of theory to compare the two methods. For nitromalonamide the CCSD(T)/cc-pVQZ computations include 700 basis functions. For the core correlation corrections the correlation-consistent, core-valence polarized, triple zeta (cc-pCVTZ) basis set of Woon and Dunning was used.[53] The total electronic energy extrapolation, was partitioned in two terms. The first term corresponds to the total SCF energy and was fitted to the functional form[54]

$$E_{SCF}(X) = A + Be^{-CX}$$

where *X* is the cardinal number corresponding to the maximum angular momentum of the basis set. The correlation energy was extrapolated using the formula[55]

$$E_{CORR}(X) - E_{SCF}(X) = A + BX^{-3}.$$

Molpro[56-61] version 2006.1 was used for all energies computed in the focal point analyses. For the malonaldehyde transition state, the geometry was optimized using PSI3.[62] The single point energy for malonaldehyde at the CCSDT level was computed using Aces II.[63]

RESULTS AND DISCUSSION

Geometries

In the present research we explore a substantial range of malonaldehyde derivatives targeting the shortest symmetrical intramolecular hydrogen bond. These hydrogen bonds are characterized by their O···O distances, where short and very short O···O connections are considered[64] ≤ 2.500 Å and ≤ 2.450 Å, respectively. X-ray diffraction experiments[65] have shown that one of the derivatives of MA with the shortest hydrogen bond is nitromalonamide (Figure 2 , **VII)**, where amino groups are substituted on C_1 and C_3 and a nitro group attached to C_2. In this work we predict the $O_4\cdots O_5$ distance in nitromalonamide to be 2.380 Å, which is the shortest $O_4\cdots O_5$ distance found here. In the exploration we used nitromalonamide as a model for short intramolecular hydrogen bonds where one substituent, X, was placed on C_1 and C_3 (see Figure 1), and a second substituent, Z, was placed on C_2. In this research, the unique structures **I – VII** are found to have the shortest $O_4\cdots O_5$ distances following nitromalonamide.

Let us first consider the effect of substituents bonded to the central carbon atom C_2. B3LYP/DZP++ predicts that the parent reference structure MA has an O···O distance of 2.546 Å. When an electron withdrawing group is attached to C_2 the hydrogen bond is shortened. A cyano group was substituted in this position, and an O···O distance of 2.526 Å was predicted. The nitro substitued molecule also has a shortened hydrogen bond, with an O···O distance of

2.521 Å. When a BH_2 group is substituted on C_2, the shortest hydrogen bond of this series is obtained, with an O···O distance of 2.499 Å.

There are several trends to consider when substituents X_7 and X_9 replace the malonaldehyde H stoms bound to C_1 and C_3, while the atom Z_8 bonded to C_2 remains hydrogen. The most significant factors playing a role in the hydrogen bond distances are the electron donating properties of substituents and steric hindrance. Our results show that upon placing electron withdrawing groups on C_1 and C_3, the O···O distance increases with respect to the parent MA. Electron donating groups attached to C_1 and C_3 have the desired effect of decreasing the O···O distances. When electron donors are connected to these positions the O···O distances decrease. The hydrogen bond distance decreases as the strength of the electron donor increases. With methoxy groups bonded to C_1 and C_3, the distance of the hydrogen bond decreases, with O···O = 2.498 Å, while an amino group further decreases the O···O distance to 2.474 Å.

To examine the chemical consequences of steric hindrance, we substituted C_1 and C_3 with methyl, isopropyl, and tert-butyl groups. The methyl substituted structure has an O···O distance of 2.511 Å. This is a shorter hydrogen bond than MA (2.545 Å), and can be explained considering the steric hindrance of the methyl group. As the size of the substituent continues to increase to isopropyl and tert-butyl, the O···O distance decreases to 2.491 Å and 2.466 Å, respectively. This series of molecules exemplifies the importance the bulk of the substituent on the hydrogen bond distance.

When considering the shortest, symmetrical, intramolecular hydrogen bond one should attempt to determine the more important factor affecting the distance of the hydrogen bond, namely electron donation or steric hindrance. Here substituents of similar size were compared, one being an electron donor and the other having more steric hindrance surrounding the oxygens

of the parent compound. These results are summarized in Table 1. Methyl and methoxy C_1 and C_3 substituted molecules have similar hydrogen bonding characteristics, with O···O distances of 2.511 Å and 2.498 Å, respectively. When comparing the isopropyl and isopropoxide substituted molecules, a trend still is not clear. However, when increasing the size of the substituent to tert-butyl and tert-butoxide there is a clear discrimination. The tert-butyl group with more steric hindrance has a shorter hydrogen bond than the tert-butoxide substituted compound, the O···O distances being 2.466 Å and 2.485 Å, respectively.

When combining the two factors of steric hindrance and electron donation, the best hydrogen bond shortening substituent for C_1 and C_3 is an excellent electron donating group with significant steric hindrance to surround the oxygen atoms. When considering only substituents on C_1 and C_3, dimethylamino gives the shortest hydrogen bond of this series with an O···O distance of 2.434 Å. This appears to be caused by electron donation to the π orbitals of MA and the steric hindrance of the methyl groups.

By means of the systematic examination we have found that particular combinations of substituents attached to all three carbons in the parent malonaldehyde structure result in some very short symmetrical intramolecular hydrogen bonds. Besides the amino and nitro derivatives of MA, including nitromalonamide, we present several other structures with short hydrogen bonds. A fascinating molecule examined during this inquiry has BH_2 attached to C_2 and NH_2 bonded to C_1 and C_3. This structure, **IV,** (**Figure 2**) has an O···O distance of 2.398 Å. In this respect structure **IV** is similar to nitromalonamide in structure, with a comparable hydrogen bond distance.

Dimethylamino substitued MA leads to the series with some of the shortest intramolecular hydrogen bonds among those considered here. The C_2 cyano derivative (structure

I) displays a short hydrogen bond, with an O···O distance of 2.399 Å. The nitro derivative (structure **II**) has an O···O distance of 2.407 Å, larger than that for structure **I** because the substituent does not lie in the molecular plane due to the steric effects of the substituents attached to C_1 and C_3. With BH_2 substituted on the unique carbon (structure **III**) the O···O distance shortens to 2.394 Å. It is seen that the BH_2 group usually has a more significant electron withdrawing effect compared to the nitro and cyano groups.

There are two other molecules with theoretical O···O distances ≤ 2.400 Å. The first molecule has a tert-butyl group substituted on C_1 and C_3 and a cyano group substituted on C_2 (structure **VI**). Structure **VI** has been studied by neutron diffraction and determined to have O···O distance of 2.393 Å.[18] Here B3LYP with the DZP++ basis set predicts a distance of 2.398 Å. This agrees very well with experiment and confirms the strength of the intramolecular hydrogen bond of **IV**. Another system (structure **V**) with a very short hydrogen bond has isopropyl as the C_1 and C_3 substituents and BH_2 as the C_2 substituent. This structure has an O···O distance of 2.398 Å, presumably due to the steric hindrance of the isopropyl group and the strength of the BH_2 group.

Energy Barrier Studies

Another way to characterize substitued malonaldehyde hydrogen bonds is the energy barrier occuring between the two equivalent minima. B3LYP energy barriers have been established to be underestimated for such systems.[66] However, the study of larger chemical systems at higher levels of theory may be impractical. We determined the energy barrier for intramolecular proton transfer for several derivatives of MA using the B3LYP method with the DZP++ basis set. The B3LYP energies show the correct qualitative trend based upon the O···O distance; however the DFT energies are underestimated. For example, the best estimate of the

barrier from the literature[16] is 4.09 kcal mol^{-1}, while B3LYP predicts a barrier of only 2.1 kcal mol^{-1}. To benchmark the B3LYP energy barriers we performed high level focal point analyses of MA and seven of its derivatives.

In addition, to help validify the method used in this work, the geometry for MA were optimized at the CCSD(T) level using a cc-pVQZ basis set. The energy barrier at this level of theory is 4.06 kcal mol^{-1}. A focal point analysis was performed using this geometry, and the energy barrier is computed to be 4.07 kcal mol^{-1}. A comparison can be drawn between this value, 4.07 kcal mol^{-1}, and 3.92 kcal mol^{-1}, the number computed from the focal point analysis with the B3LYP / DZP++ geometry. The difference between these two methods is 0.15 kcal mol^{-1}, which is a satisfactorily accurate prediction obtained without performing rigorous coupled cluster geometry optimizations.

The focal point analysis for MA is reported in Table II. Our complete basis set extrapolated energy barrier is 3.92 kcal mol^{-1}. We find excellent convergence with respect to correlation and basis set treatments; the CCSD(T)/cc-pVDZ and CCSDT/cc-pVDZ energies are separtated by only 0.04 kcal mol^{-1}, while the CCSD(T) cc-pVQZ and the CCSD(T) cc-pVTZ are separated by only 0.04 kcal mol^{-1}. Core-correlation is estimated to lower the barrier by 0.01 kcal mol^{-1}.

Using the focal point method one may reliably compare the barriers for some of the derivatives to MA. The C_2- cyano substituted molecule has a barrier of 3.56 kcal mol^{-1} which is 0.36 kcal mol^{-1} lower in energy than MA. As the strength of the electron withdrawing substituent on C_2 increased, the barrier height is further decreased to 3.34 kcal mol^{-1} and 2.62 kcal mol^{-1} for the NO_2 and BH_2 substitued molecules, respectively.

The other series of molecules examined with the focal point method include structures with amino (NH_2) groups substituted on both C_1 and C_3 and withdrawing groups substituted on C_2. For the first of the series, where an amino group is attached to C_1 and C_3 and the unique carbon is unsubstituted, the barrier height is 2.02 kcal mol^{-1}. This barrier height decreases as stronger withdrawing groups are substituted on the unique carbon. For the cyano substituted (on atom C_2) system, the barrier height decreases to 1.40 kcal mol^{-1}. The results from the focal point analyses for the systems with stronger substituents bonded to the unique carbon are 0.43 kcal mol^{-1} and 0.60 kcal mol^{-1} for the NO_2 (nitromalonamide, **VII**), and the BH_2 (**IV**), substituted molecules, respectively. This relationship is exactly what is expected from the O···O distances computed in the more qualitative DFT studies. Recall that nitromalonamide is thought[44] to possess one of the smallest intramolecular proton transfer barriers, theoretically estimated to be 0.6 kcal mol^{-1} and 1.2 kcal mol^{-1} using MP2/cc-pVTZ and B3LYP/cc-pVTZ, respectively. In this work, the barrier for nitromalonamide is predicted to be 0.43 kcal mol^{-1} using the focal point method, the lowest barrier we find. This value should be the most reliable barrier reported to date for nitromalonamide, obtained with the aid of the high level focal point analysis. It appears that changing the amino substituent to dimethylamino lowers the barrier when smaller substituents are on C_2, such as cyano. However, as the size of the substituent increases to the NO_2 or BH_2 group, steric hindrance occurs between the C_2 substituents and the dimethylamino groups on C_1 and C_3. When the C_2 substituents are forced to break the plane of the MA system due to steric reasons, they have less of a strengthening affect.

The present study also investigates the relationship between the systematically underestimated B3LYP barriers and the more accurate extrapolated energies computed from the focal point method. The B3LYP energies were plotted versus the extrapolated energies for the

eight chemical systems examined in the focal point extrapolations. A linear fit was performed on this plot to determine the relationship between the energies computed using the two very different methods. We apply the linear equation obtained (in kcal mol^{-1}),

$$\Delta E_b^{FP} = 1.73 \Delta E_b^{B3LYP} + 0.38$$

to our chemical systems. In the above equation ΔE_b^{B3LYP} is the barrier height computed using B3LYP, while ΔE_b^{FP} is the barrier height that is extrapolated using the focal point method. In Figure 3 and Table III, one can see how the projected energy values have good agreement with the extrapolated values. The R^2 value for this linear equation is 0.995.

The above empirical method is now applied to interesting systems that possess very short intramolecular hydrogen bonds that may be too large to study by highly correlated methods such as the focal point analysis. Structure **VI** has a tert-butyl group substituted on C_1 and C_3 and a cyano group attached to the unique carbon. B3LYP predicts the energy of this barrier to be 0.38 kcal mol^{-1}. The projection method is applied to predict the energy barrier to be 1.02 kcal mol^{-1}. Structure **I** is probably too large to study presently by the focal point method, due to the dimethylamino substituents on C_1 and C_3 and the cyano group substitued on the unique carbon. The barrier height for the hydrogen atom transfer is predicted by B3LYP to be 0.20 kcal mol^{-1} and when the projection method is applied the barrier height becomes 0.73 kcal mol^{-1}. This approximation corrects the underestimation that occurs with energies computed by B3LYP.

CONCLUSIONS

The intramolecular hydrogen bond of MA has been explored extensively as the simple molecule that shows a short intramolecular hydrogen bond for which proton transfer occurs between two oxygen atoms. In this study, a series of trends were examined to explore the effect of various substituents on the intramolecular hydrogen bonds. When placing substituents on the

carbons, it was found that bulky electron donators on C_1 and C_3 and strong electron withdrawing groups bonded to the unique carbon created the strongest intramolecular hydrogen bonds. This qualitative conclusion serves as a precusor to the precise quantitative study of the hydrogen transfer barriers.

Via a systematic study, seven compounds were predicted to have very short symmetrical intramolecular hydrogen bonds. These structures all have strong electron withdrawing groups bonded to atom C_2. Structures **I**, **II**, and **III** have short hydrogen bonds because of the steric hindrance and electron donating character the dimethylamino substituent provides. Structure **IV** has a short hydrogen bond due to the significant electron donation the amino substituent provides, and this structure is comparable to nitromalonamide. Structures **V** and **VI** provide much steric hindrance around the oxygen atoms, which are forced together, shortening the hydrogen bond distance.

The proton transfer energy barriers were examined using three different methodologies. B3LYP is the only level of theory applied to all the systems we have examined. However, B3LYP is known to underestimate barrier heights, and this is the case for the malonaldehyde systems. The focal point method using correlated electronic structure methods was applied to the simpler systems to accurately predict the energy required for the proton transfer. A relationship was drawn between the B3LYP energies and the high level theoretical barriers that allows for accurate predictions of the proton transfer barrier heights. This relationship utilizes B3LYP energies to accurately predict the barrier heights that might be expected from highly correlated methods.

The present study reports the first high level focal point analysis for MA and predicts a hydrogen transfer barrier of 3.92 kcal mol^{-1}. Also utilizing the focal point method we predict two

substituted malonaldehyde barriers to be less than 1 kcal mol^{-1}, namely the barriers for nitromalonamide and 2-borylmalonamide. In previous theoretical work,[44] the barrier for nitromalonamide was computed to be 0.6 kcal mol^{-1} using the MP2 method. The latter result may be compared to the value predicted in this work utilizing the focal point analysis, namely 0.43 kcal mol^{-1}. The proton transfer barrier for 2-borylmalonamide is predicted to be 0.60 kcal mol^{-1}. This rigorous approach produces the most reliable results to date for all systems considered.[16]

A remaining goal is the identification of a substituted malonaldehyde with no barrier at all, i.e., a C_{2v} equilibrium geometry. We can speculate that increasing the size of the substituent on C_1 and C_3 may lower the proton transfer barrier for this system. It would be interesting to explore the possibility of $C[C(CH_3)_3]_3$ substituents. Another prospect would be to search for an electron withdrawing group with a more significant effect than the substituents studied in this work.

REFERENCES

(1) Sobczyk, L.; Grabowski, S. J.; Krygowski, T. M. *Chem. Rev.* **2005**, *105*, 3513.

(2) Perrin, C. L.; Nielson, J. B. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511.

(3) Hayashi, T.; Mukamel, S. *J. Phys. Chem. A* **2003**, *107*, 9113.

(4) Tautermann, C. S.; Voegele, A. F.; Loerting, T.; Liedl, K. R. *J. Chem. Phys.* **2002**, *117*, 1962.

(5) Coutinho-Neto, M. D.; Viel, A.; Manthe, U. *J. Chem. Phys.* **2004**, *121*, 9207.

(6) Fillaux, F.; Nicolai, B. *Chem. Phys. Lett.* **2005**, *416*, 357.

(7) Baughcum, S. L.; Duerst, R. W.; Rowe, W. F.; Smith, Z.; Wilson, E. B. *J. Am. Chem. Soc.* **1981**, *103*, 6296.

(8) Frisch, M. J.; Scheiner, A. C.; Schaefer, H. F.; Binkley, J. S. *J. Chem. Phys.* **1985**, *82*, 4194.

(9) Binkley, J. S.; Frisch, M. J.; Schaefer, H. F. *Chem. Phys. Lett.* **1986**, *126*, 1.

(10) Bicerano, J.; Schaefer, H. F.; Miller, W. H. *J. Am. Chem. Soc.* **1983**, *105*, 2550.

(11) Mil'nikov, G. V.; Yagi, K.; Taketsugu, T.; Nakamura, H.; Hirao, K. *J. Chem. Phys.* **2003**, *119*, 10.

(12) Esterbauer, H.; Schaur, R. J.; Zollner, H. *Free Radical Bio. Med.* **1991**, *11*, 81.

(13) Rowe, W. F.; Duerst, R. W.; Wilson, E. B. *J. Am. Chem. Soc.* **1976**, *98*, 4021.

(14) Baughcum, S. L.; Smith, Z.; Wilson, E. B.; Duerst, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 2260.

(15) Turner, P.; Baughcum, S. L.; Coy, S. L.; Smith, Z. *J. Am. Chem. Soc.* **1984**, *106*, 2265.

(16) Wang, Y.; Braams B. J.; Bowman, J. M.; Carter, S.; Tew, D.P. *J. Chem. Phys.* **2008**, *128*, 224314.

(17) Mil'nikov, G. V.; Yagi, K.; Taketsugu, T.; Nakamura, H.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 5036.

(18) Belot, J. A.; Clark, J.; Cowan, J. A.; Harbison, G. S.; Kolesnikov, A. I.; Kye, Y. S.; Schultz, A. J.; Silvernail, C.; Zhao, X. G. *J. Phys. Chem. B* **2004**, *108*, 6922.

(19) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1994**, *116*, 909.

(20) Krokidis, X.; Goncalves, V.; Savin, A. *J. Phys. Chem. A* **1998**, *102*, 5065.

(21) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989**, *111*, 1023.

(22) Dannenberg, J. J.; Rios, R. *J. Phys. Chem.* **1994**, *98*, 6714.

(23) Lowrey, A. H.; George, C.; Dantonio, P.; Karle, J. *J. Am. Chem. Soc.* **1971**, *93*, 6399.

(24) Egan, W.; Gunnarsson, G.; Bull, T. E.; Forsen, S. *J. Am. Chem. Soc.* **1977**, *99*, 4568.

(25) Firth, D. W.; Barbara, P. F.; Trommsdorff, H. P. *Chem. Phys.* **1989**, *136*, 349.

(26) Firth, D. W.; Beyer, K.; Dvorak, M. A.; Reeve, S. W.; Grushow, A.; Leopold, K. R. *J. Chem. Phys.* **1991**, *94*, 1812.

(27) Baba, T.; Tanaka, T.; Morino, I.; Yamada, K. M. T.; Tanaka, K. *J. Chem. Phys.* **1999**, *110*, 4131.

(28) Seliskar, C. J.; Hoffmann, R. E. *J. Mol. Spectrosc.* **1982**, *96*, 146.

(29) Smith, Z.; Wilson, E. B.; Duerst, R. W. *Spectrochim Acta A* **1983**, *39*, 1117.

(30) Chiavassa, T.; Roubin, P.; Pizzala, L.; Verlaque, P.; Allouche, A.; Marinelli, F. *J. Phys. Chem.* **1992**, *96*, 10659.

(31) Ishida, T.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1999**, *110*, 3938.

(32) Bauer, S. H.; Wilcox, C. F. *Chem. Phys. Lett.* **1997**, *279*, 122.

(33) Sharafeddin, O. A.; Hinsen, K.; Carrington, T.; Roux, B. *J. Comput. Chem.* **1997**, *18*, 1760.

(34) Mavri, J.; Grdadolnik, J. *J. Phys. Chem. A* **2001**, *105*, 2039.

(35) Grabowski, S. J. *J. Phys. Org. Chem.* **2003**, *16*, 797.

(36) Delchev, V. B.; Mikosch, H.; St. Nikolov, G. *Monatsh. Chem.* **2001**, *132*, 339.

(37) Delchev, V. B. *Monatsh. Chem.* **2004**, *135*, 371.

(38) Yoon, M. C.; Choi, Y. S.; Kim, S. K. *J. Chem. Phys.* **1999**, *110*, 11850.

(39) Johnson, M. R.; Jones, N. H.; Geis, A.; Horsewill, A. J.; Trommsdorff, H. P. *J. Chem. Phys.* **2002**, *116*, 5694.

(40) Matanovic, I.; Doslic, N. *Int. J. Quantum Chem.* **2006**, *106*, 1367.

(41) Buemi, G. *Chem. Phys.* **2002**, *277*, 241.

(42) Buemi, G.; Zuccarello, F. *Chem. Phys.* **2004**, *306*, 115.

(43) Sebban, M.; Guillard, J.; Palmas, P.; Poullain, D. *Magn. Reson. Chem.* **2005**, *43*, 563.

(44) Madsen, G. K. H.; Wilson, C.; Nymand, T. M.; McIntyre, G. J.; Larsen, F. K. *J. Phys. Chem. A* **1999**, *103*, 8684.

(45) Tayyari, S. F.; Reissi, H.; Milani-Nejad, F.; Butler, I. S. *Vib. Spectrosc.* **2001**, *26*, 187.

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

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

(48) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293.

(49) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823.

(50) Lee, T. J.; Schaefer, H. F. *J. Chem. Phys.* **1985**, *83*, 1784.

(51) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R.

C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys Chem Chem Phys* **2006**, *8*, 3172.

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

(53) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1995**, *103*, 4572.

(54) Feller, D. *J. Chem. Phys.* **1993**, *98*, 7059.

(55) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.

(56) MOLPRO, v., a package of ab initio programs, Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. see http://www.molpro.net.

(57) Lindh, R.; Ryu, U.; Liu, B. *J. Chem. Phys.* **1991**, *95*, 5889.

(58) Eckert, F.; Pulay, P.; Werner, H. J. *J. Comput. Chem.* **1997**, *18*, 1473.

(59) Rauhut, G.; ElAzhary, A.; Eckert, F.; Schumann, U.; Werner, H. J. *Spectrochim Acta A* **1999**, *55*, 647.

(60) Hampel, C.; Peterson, K. A.; Werner, H. J. *Chem. Phys. Lett.* **1992**, *190*, 1.

(61) Deegan, M. J. O.; Knowles, P. J. *Chem. Phys. Lett.* **1994**, *227*, 321.

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

(63) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *Int. J. Quantum Chem.* **1992**, 879.

(64) Hibbert, F. E.; J. *Adv. Phys. Org. Chem.* **1990**, *26*, 255.

(65) Simonsen, O.; Thorup, N. *Acta. Crystallogr. B* **1979**, *35*, 432.

(66) Latajka, Z.; Bouteiller, Y.; Scheiner, S. *Chem. Phys. Lett.* **1995**, *234*, 159.

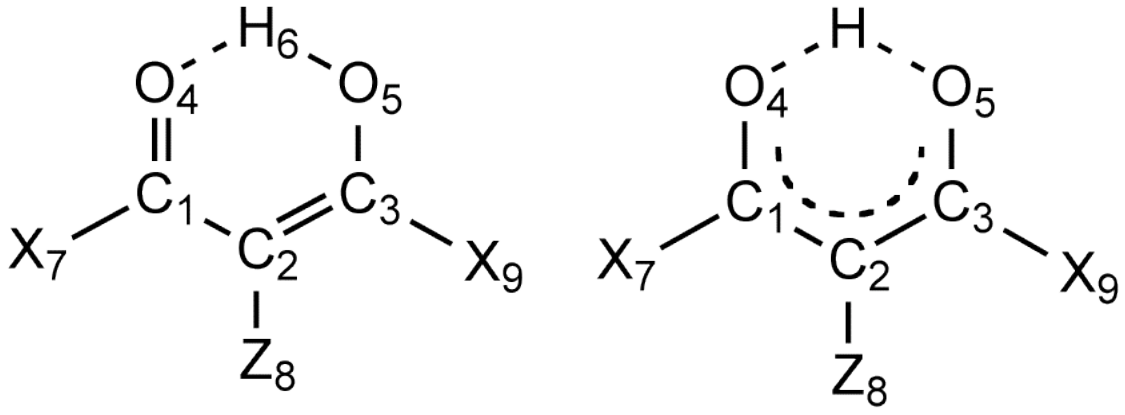
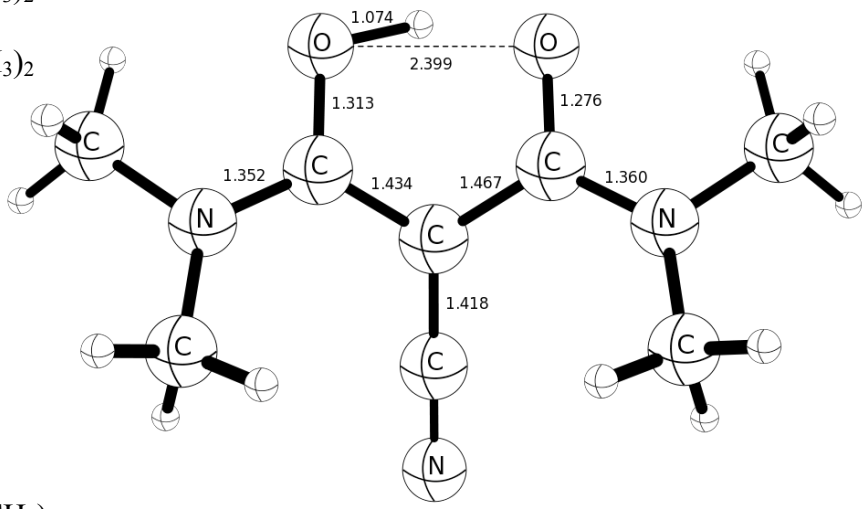
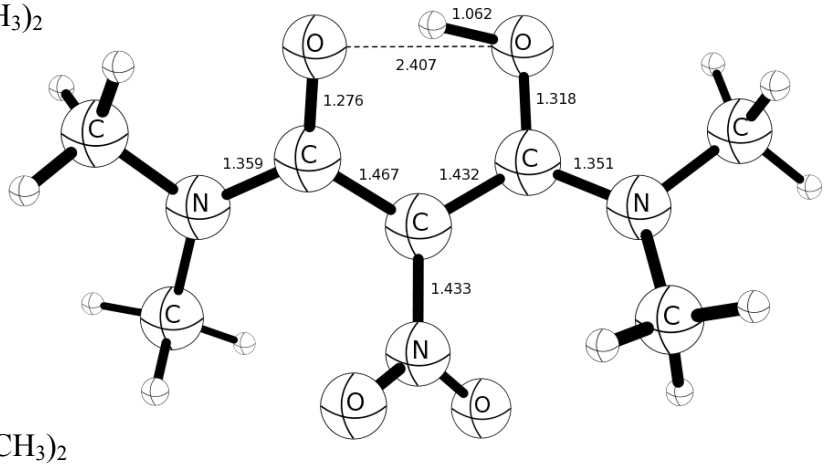


Figure 2.1. The left figure shows the malonaldehyde parent structure at its C_s equilibrium geometry, while the right figure shows the C_{2v} transition state.

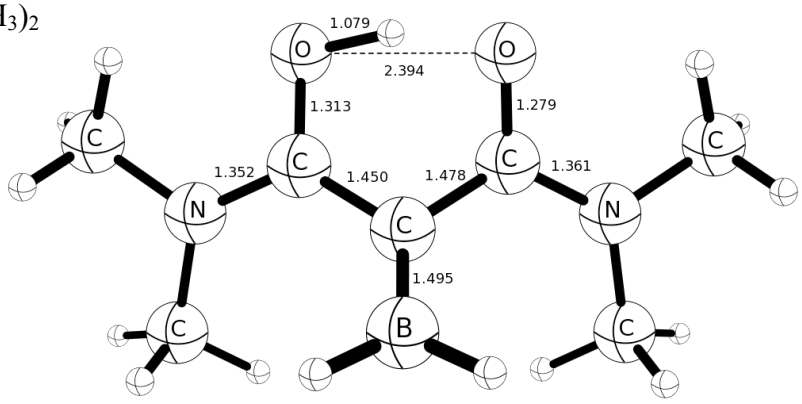
I. 1 – $N(CH_3)_2$
2 – CN
3 – $N(CH_3)_2$



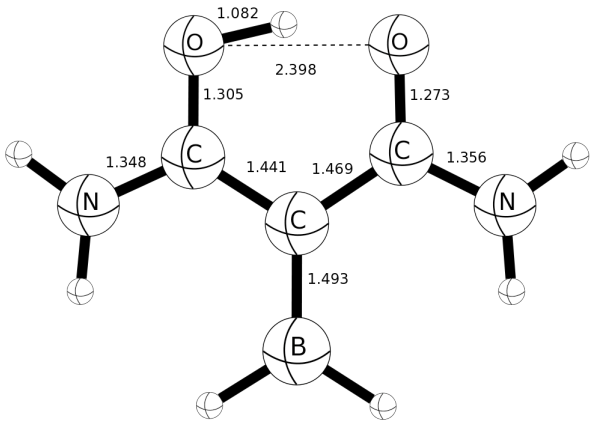
II. 1 – $N(CH_3)_2$
2 – NO_2
3 – $N(CH_3)_2$



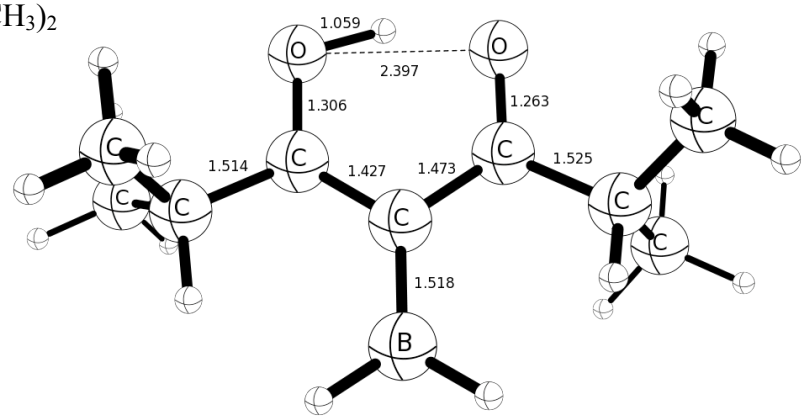
III. 1 – $N(CH_3)_2$
2 – BH_2
3 – $N(CH_3)_2$



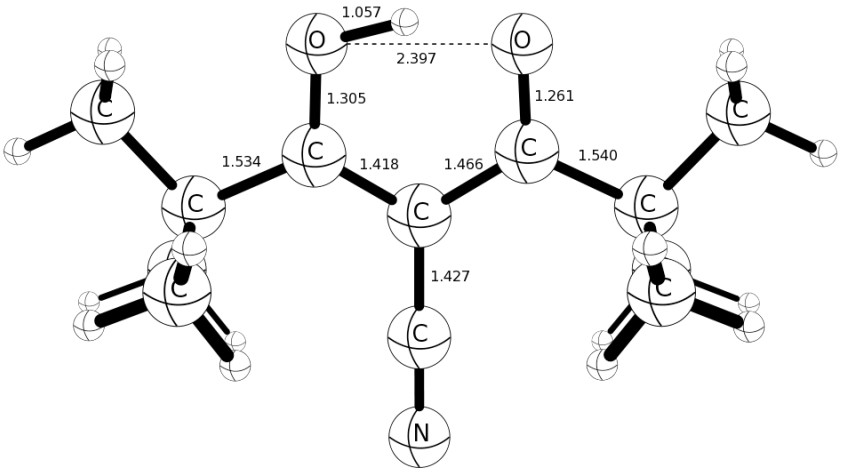
IV. 1 – NH_2
2 – BH_2
3 – NH_2



V. 1 – $CH(CH_3)_2$
2 – BH_2
3 – $CH(CH_3)_2$



VI. 1 – $C(CH_3)_3$
2 – CN
3 – $C(CH_3)_3$



VII. $1 - NH_2$
$2 - NO_2$
$3 - NH_2$

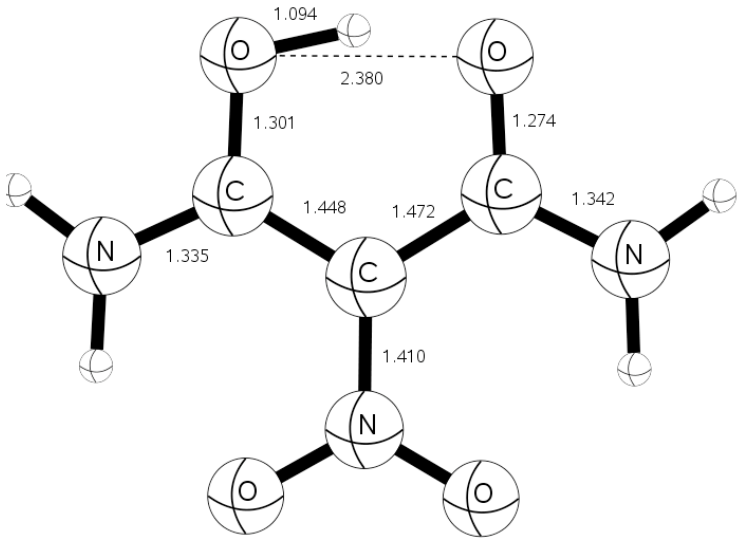


Figure 2.2. Six structures with very short intramolecular hydrogen bonds. Bond distances are reported in Å. The six compounds are named by their substituents. See atom numbering in Figure 1.

Table 2.1. The distance (Å) between atoms O_4 and O_5 (see Figure 2.1) is shown for molecules with varying substituents. This method is useful for characterizing the intramolecular hydrogen bond, because the results may be directly compared to the few available crystal structures, where the precise positions of the H atoms are not known.

		Substituents on the Symmetrical Carbons										
		H	NH_2		$N(CH_3)_2$		CH_3	$CH(CH_3)_2$		$C(CH_3)_3$		OCH_3	$OCH(CH_3)_2$	$OC(CH_3)_3$
Unique Carbon Substituents	H	2.546	2.474		2.434		2.511	2.491		2.466		2.498	2.487	2.485
	CN	2.526	2.448		2.399	**I**	2.471	2.456		2.397	**VI**	2.464	-	-
	NO_2	2.521	2.380		2.407	**II**	2.423	2.425		-		2.442	-	-
	BH_2	2.499	2.398	**IV**	2.394	**III**	2.419	2.397	**V**	-		2.421	-	-

Table 2.2. This focal point analysis shows the barrier height for the proton transfer between the two oxygen atoms in malonaldehyde. MA is optimized at B3LYP / DZP++ in [a] and CCSD(T) / cc-pVQZ in [b]. For the correlated methods, the symbol δ denotes the increment in the relative energies between the preceding level of theory in the hierarchy. The square brackets signify the values extrapolated from the basis sets, where all other entries are computed values. The final predictions are in bold print.

[a]

Basis Set	ΔE_e [RHF]	$+\delta$ [MP2]	$+\delta$ [CCSD]	$+\delta$ [CCSD(T)]	$+\delta$ [CCSDT]	ΔE_e [CCSDT]
cc-pVDZ	+8.03	−5.12	+1.87	−0.96	+0.04	+3.86
cc-pVTZ	+8.11	−5.34	+2.00	−1.08	[+0.04]	[+3.73]
cc-pVQZ	+8.21	−5.31	+2.06	−1.10	[+0.04]	[+3.90]
cc-pV5Z	+8.21	−5.31	+2.08	[−1.10]	[+0.04]	[+3.92]
CBS limit	[+8.21]	[−5.32]	[+2.10]	[−1.10]	[+0.04]	[+3.93]

ΔE_b(final)= ΔE_e[CBS CCSDT]+ Δ_{core}[MP2/cc-pCVTZ]
=3.93-0.01= **3.92** kcal mol^{-1}
$\Delta E_{b,0}$(final)= ΔE_e[CBS CCSDT]+ Δ_{ZPVE}[B3LYP/DZP++] + Δ_{core}[MP2/cc-pCVTZ]
=3.93−2.38−0.01= **1.54** kcal mol^{-1}

Fit	$a+bc^{-eX}$	$a+bX^{-3}$	$a+bX^{-3}$	$a+bX^{-3}$	additive
Points (X)	3,4,5	4,5	4,5	3,4	

[b]

Basis Set	ΔE_e [RHF]	$+\delta$ [MP2]	$+\delta$ [CCSD]	$+\delta$ [CCSD(T)]	$+\delta$ [CCSDT]	ΔE_e [CCSDT]
cc-pVDZ	+9.17	−6.39	+2.34	−1.18	+0.05	+3.99
cc-pVTZ	+9.31	−6.63	+2.49	−1.32	[+0.05]	[+3.90]
cc-pVQZ	+9.44	−6.60	+2.56	−1.33	[+0.05]	[+4.12]
cc-pV5Z	+9.43	−6.60	+2.58	[−1.33]	[+0.05]	[+4.13]
CBS limit	[+9.43]	[−6.60]	[+2.58]	[−1.33]	[+0.05]	[+4.13]

ΔE_b(final)= ΔE_e[CBS CCSDT]+ Δ_{core}[MP2/cc-pCVTZ]
=4.13-0.06= **4.07** kcal mol^{-1}
$\Delta E_{b,0}$(final)= ΔE_e[CBS CCSDT]+ Δ_{ZPVE}[B3LYP/DZP++] + Δ_{core}[MP2/cc-pCVTZ]
=4.13−2.50−0.06= **1.57** kcal mol^{-1}

Fit	$a+bc^{-eX}$	$a+bX^{-3}$	$a+bX^{-3}$	$a+bX^{-3}$	additive
Points (X)	3,4,5	4,5	4,5	3,4	

Table 2.3. Hydrogen atom transfer barrier heights (ΔE_b) (kcal mol^{-1}) for three different approaches applied in this research. These results are shown visually in Figure 2.3.

C_1 and C_3	H			NH_2		
C_2	B3LYP/ DZP++	Focal Point Energy	Linear Fit	B3LYP/ DZP++	Focal Point Energy	Linear Fit
H	2.10	3.92	4.00	0.87	2.02	1.88
CN	1.85	3.56	3.56	0.54	1.40	1.31
NO_2	1.67	3.34	3.25	0.09	0.43	0.53
BH_2	1.32	2.62	2.66	0.17	0.60	0.67

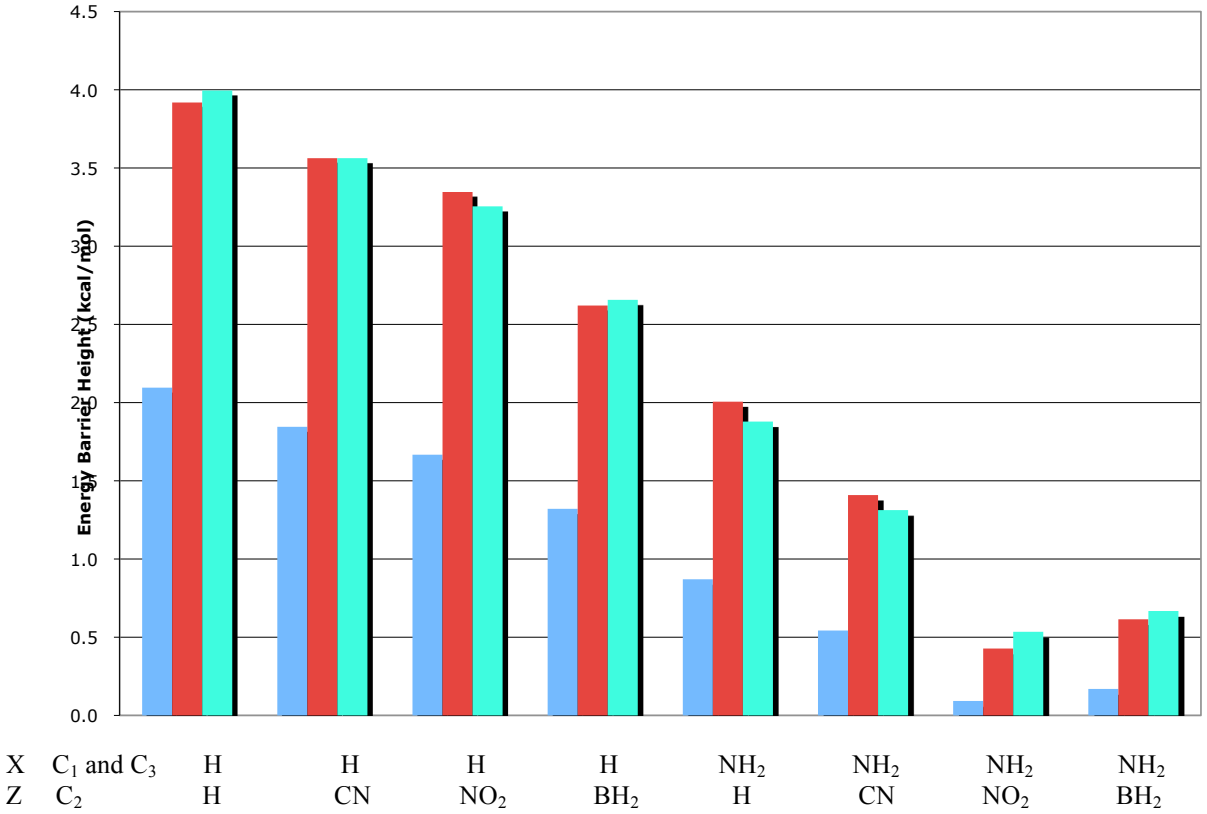


Figure 2.3. For molecular systems 1-8, this figure draws a comparison in the barrier heights between the DFT (left column), the extrapolated focal point energy (middle column), and the energy predicted using the linear fit (right column).

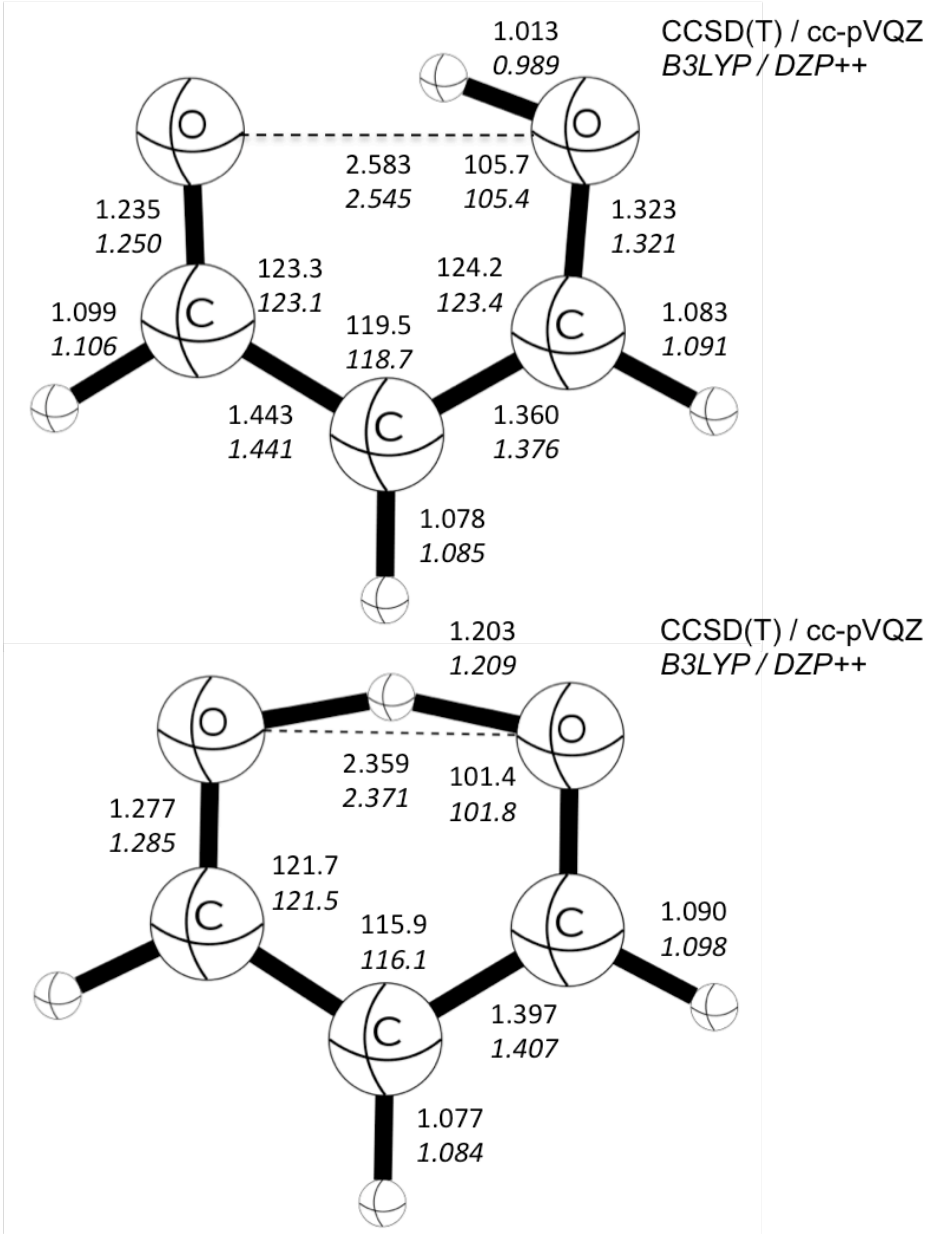


Figure 2.4. Optimized geometries for malonaldehyde using two levels of theory, CCSD(T) / cc-pVQZ and B3LYP / DZP++. This figure illustrates that B3LYP does a satisfactory job of reproducing the geometries computed at higher levels of theory. The B3LYP / DZP++ geometries are in italics. The top figure shows the C_s ground state geometry comparison and the bottom figure shows the C_{2v} transition state geometry comparison.

CHAPTER 3

CHARACTERIZING THE MECHANISM OF THE DOUBLE PROTON TRANSFER IN THE FORMAMIDE DIMER[1]

[1] Reproduced with permission from Hargis, J.C.; Vöhringer-Martinez, E.; Woodcock, H. L.; Toro-Labbé, A.; Schaefer, H. F. *J. Phys. Chem. A (in press)* Copyright 2011 American Chemical Society

ABSTRACT

The double proton transfer in the formamide dimer is characterized computationally by combining density functional theory and *ab initio* methods. The intrinsic reaction coordinate (IRC) is obtained at the B3LYP level of theory. Energies of several points along the IRC are treated by the more rigorous focal point method to test the validity of the B3LYP functional. The reaction mechanism is examined in terms of the energy profile, the reaction force, the chemical potential, and the reaction electronic flux.

The energy profile for the activation process of the formamide dimer to the imino ether product obtained with the B3LYP functional is in agreement with the results of the focal point method. Together with the reaction force analysis and the reaction electronic flux a precise assignment of the structural and electronic contributions to the activation barrier becomes possible. The results show that the reaction starts with a structural rearrangement, where the two dimers approach each other and is followed by electronic changes before the system reaches the transition state. This electronic contribution to the activation barrier steers the activation process. After reaching the transition state, deviations of the B3LYP functional from the more accurate focal point energies become apparent, where the errors may be rationalized in terms of the treatment of exchange. The inconsistency could be assigned to the incapacity of the functional to describe delocalization effects over the whole system.

INTRODUCTION

Proton transfer is one of the most fundamental processes in chemistry and biology.[1-3] Multiple proton transfers, that occur either synchronously or asynchronously, have been found in proton relay systems in enzymes, hydrogen bonded water complexes, and in prototropic tautomerisms. Double proton transfer in DNA base pairs is also a commonly cited example. The

prototropic tautomerism of the formamide dimer is important in proteins and has been used as a model for nucleic bases.[4,5]

The formamide dimer has been studied extensively by theoretical[6-28] and experimental[29-31] methods. Despite the extensive work done on this system, the double proton reaction path has not to date been fully investigated. The study of Grabowski, Sokalski, and Leszcznski[28] examined the dimer at the MP2/6-311++G(d,p) level and concluded that the interaction of the formamide dimer has its largest component in the electrostatic term, while the less stable imino ether product is stabilized by the attractive delocalization term, where the electron correlation becomes more important. The methods in this study will be used to expand on this work exploring additional aspects of this system along the reaction path.

Several studies have examined the intrinsic reaction coordinate (IRC) to determine information about double proton transfer reactions. Cybulski and Sadlej[32] examined this process in the formamide-formic acid and the formamide-formamidine complexes. In that work they utilized the IRC method to monitor the changes in NMR parameters. Research by Toro-Labbé and co-authors[33-35] has utilized IRCs to explain reaction mechanisms examining energetic profiles together with reaction forces, chemical potentials, and reaction electronic flux.

In the present study the above methods will be combined with the focal point analysis scheme, to accurately describe the energy along the reaction path. Combining this with mechanistic investigations, a lucid picture may be painted to show the precise assignment of structural and electronic parts of the activation barrier. In addition, the performance of the B3LYP functional along the reaction coordinate will be analyzed.

THEORETICAL BACKGROUND

A chemical reaction takes place in multidimensional space and the characterization occurs by monitoring changes in geometrical parameters. The intrinsic reaction coordinate[36] (ξ) shows a projection of the multidimensional motion occurring in the chemical reaction. The energy profile along ξ corresponds to the minimum energy path relating reactants and products. Numerical differentiation of $E(\xi)$ results in the reaction force.[37,38]

$$F(\xi) = -\frac{dE}{d\xi}$$

Figure 1 shows a schematic representation of the potential energy along the reaction coordinate ξ for an elementary process and the corresponding reaction force. The reaction force is zero at reactants, transition state and products and displays a minimum (ξ_1) and a maximum (ξ_2). These two extremes divide the reaction in three different regions: reactant region ($\xi_R \leq \xi < \xi_1$), transition state region ($\xi_1 \leq \xi \leq \xi_2$) and the product region ($\xi_2 < \xi \leq \xi_P$).[33-35,37-47] In the reactant region the reactant undergoes mostly structural changes to achieve the reactive conformation at the force minimum. Once this conformation is reached the electronic changes take over in the transition state region, resulting in bond formation and rupture. The region extends until the force maximum where a structural relaxation takes over to reach the product. These three region provide a framework, which allows a detailed analysis of the reaction mechanism, as has been shown for various reactions including intra- and intermolecular proton transfer,[33,34] S_N2[44] and isomerization reactions.[43]

The chemical potential results from the Euler-Lagrange equation of the energy functional in density functional theory (DFT) as a Lagrange multiplier to conform with the stipulation that the electronic density integrates to N, the total number of electrons in the system.[48] For a N-

electron system with total energy E and external potential $v(\vec{r})$, the chemical potential is defined as[48]

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(\vec{r})} = -\chi$$

where χ is the electronegativity.[48-53] With the finite difference approximation the following working expression for the chemical potential is obtained,[48,51,52,54]

$$\mu \approx -\frac{1}{2}(I + A)$$

where I is the first ionization potential and A the electron affinity.

Calculation of the chemical potential for each structure along the reaction coordinate ξ yields $\mu(\xi)$. In previous works we have shown, that variations of the chemical potential $\mu(\xi)$ are associated with electronic reordering in the system, which can be corroborated with changes in bond indices and natural charges.[33,34,43,44] These variations are quantified with the reaction electronic flux, which is defined as [33-35,43,44]

$$J(\xi) = -\frac{d\mu}{d\xi}$$

In analogy to thermodynamics, the reaction electronic flux can be used to describe the spontaneity of a process: positive values characterize a spontaneous change in the electronic density whereas negative ones a non-spontaneous one. The reaction electronic flux has evolved to be a very useful descriptor to identify the regions along the reaction coordinate that are characterized by electronic reordering and transfer. Therefore, it is complementary to the

reaction force and enables a direct access and a better rationalization of the electronic changes along the reaction coordinate.

The focal point analysis is a two-dimensional extrapolation grid used to monitor the basis set and method dependence of the correlation energy in order to track the reaction energy toward the complete basis set full configuration interaction limit. This method, created by Allen and coworkers[55-58] was developed to examine chemical reactions at sub kcal mol^{-1} accuracy. However, in this case DFT geometries are used in conjunction with the focal point method[59] to attain accurate and affordable results appropriate for the current work.

COMPUTATIONAL DETAILS

Energies, structures, and transition states were initially determined using Density Functional Theory employing the B3LYP generalized gradient approximation (GGA) exchange correlation functional. B3LYP combines Becke's description of exchange, the 3-parameter HF/DFT hybrid exchange functional (B3),[60] and the dynamical correlation functional of definition Lee, Yang, and Parr (LYP).[61] The proton transfer reaction was followed with B3LYP and the 6-311G(d,p) basis set along the intrinsic reaction coordinate ξ, using mass-weighted internal coordinates and a step size of 0.01 a_0 $amu^{1/2}$. Reactants, transition state and products were also optimized at the MP2/cc-pVQZ level for comparison. Stationary points were verified by harmonic vibrational analysis at the MP2 and B3LYP levels of theory.

The electronic chemical potential was obtained from single point energy calculations of the cationic and anionic forms of all structures along the reaction coordinate employing the B3LYP/6-311G(d,p) method. The Wiberg bond indeces were obtained from Natural Bond Order

analyses (NBO)[62,63] along the reaction coordinate using the NBO program[64] as implemented in the Gaussian 03[65] package.

A focal point analysis using the HF, MP2, CCSD, and CCSD(T) levels of theory was executed using the correlation consistent basis sets of Dunning[66] (cc-pV*X*Z, *X* = D, T, Q, 5) to yield values extrapolated to the basis set limit. This extrapolation method has proven effective in combination with DFT geometries;[59] therefore, the aforementioned B3LYP and MP2 geometries were utilized. The total energy extrapolation was partitioned in two terms (SCF and correlation energies). The first term, total SCF energy, was fitted to the functional form[67]

$$E_{SCF}(X) = A + Be^{-CX}$$

where *X* is the cardinal number corresponding to the maximum angular momentum of the basis set. *A, B,* and *C* are determined by these equations for the extrapolation of the Hartree-Fock energy.

$$A = E_3 - Be^{-CX_3}$$

$$B = \frac{E_3 - E_2}{e^{-CX_2}\left(e^{-C} - 1\right)}$$

$$C = -\ln\frac{E_3 - E_2}{E_2 - E_1}$$

The correlation energy uses a different formula:[68]

$$E_{CORR}(X) - E_{SCF}(X) = A + BX^{-3}$$

where *A* and *B* are determined by these equations:

$$A = E_2 - BX_2^{-3}$$

$$B = \frac{E_2 - E_1}{X_2^{-3} - X_1^{-3}}$$

In the equations above, E_x corresponds to the energy of the largest basis sets computed at that level of theory. E_3 would be the largest basis set and E_2 and E_1 would correspond to the next to largest basis sets in descending order with X still being defined as the principal number in the Dunning correlation-consistent basis set. Hartree-Fock uses a three point extrapolation and for levels of theory with electron correlation, including MP2, CCSD and CCSD(T), a two-point extrapolation is used. The extrapolated values are additive determining the final energy near the CBS limit. Molpro version 2006.1[69-71] was used for all energies computed in the focal point analyses. For the MP2 geometry optimizations Q-Chem 3.2[72] was used.

RESULTS AND DISCUSSION

The double proton transfer of the formamide dimer has been investigated here using several approaches. The most accurate determination of the barrier combined MP2/cc-pVQZ geometries (See Figure 2) and the focal point extrapolation method. The 19.9 kcal mol^{-1} barrier (Table 1) between the reactant and the transition state ($\Delta E_f^{\ddagger}$) shows that considerable energy is necessary for the double proton transfer to occur. Additionally, the product lies 16.8 kcal mol^{-1} higher in energy ($\Delta E_{reaction}$) than the reactant. B3LYP/6-311G(d,p) geometries (See Figure 2) and the focal point extrapolation method yield nearly the same results (see Table 2). Using the 6-311++G(d,p) basis at the MP2 level, the work of Grabowski, Sokalski, and Leszczynski[28] determined the energy difference between the reactant, the transition state, and product to be 20.5 kcal mol^{-1} and 19.6 kcal mol^{-1}, respectively. DFT in the current work has difficulties

yielding the correct energy in the product region, in contrast to the focal point method, which is able to describe the region accurately. There is a smaller discrepancy in the transition state region between the MP2 energies and the focal point energies.

B3LYP and focal point results (Table 2) yield a similar forward barrier ($\Delta E_f^{\ddagger}$), however there is discrepancy in the reverse barrier ($\Delta E_r^{\ddagger}$). It has been shown that B3LYP underestimates reaction barriers,[73,74] so the perplexing behavior of B3LYP in this case is of interest. This behavior could allude to an error in the basis set, 6-311G(d,p). In the past, certain Pople basis sets have been shown to produce errors associated with intra-molecular basis set superposition error,[75,76] however the 6-311G(d,p) was not highlighted as problematic. Current results agree; after computing B3LYP energies with several basis sets including the sizeable cc-pVQZ basis set, we observe no significant differences (See Supplemental Information). Thus a problem with the B3LYP functional is more likely and has been investigated herein.

After eliminating the basis set as a possible source of error, the remaining possibilities are the exchange or correlation used in the B3LYP functional. The diagnosis of this problem can be obtained by comparing the focal point and B3LYP results to Hartree-Fock barriers. To begin the analysis, the correlation is inspected. As expected, the barrier lowers when going from HF to the focal point results, which is attributed purely to correlation. Furthermore, most of the correlation is recovered when going from HF to MP2 as seen in Table 1. To monitor the effects of LYP correlation NBO analyses were employed (i.e. HF vs. HF-LYP). Orbital stabilizations as a function of reaction coordinate (i.e. hydrogen bonding, etc.) remained consistent with and without LYP correlation. This confirms that LYP correlation has a negligible impact on delocalization and non-bond interactions (*vide infra)*.

It has been shown that B3LYP produces an overestimation of non-bonded repulsion in reactions involving rings and cagelike molecules.[77] Additionally, a B3LYP error exists that overestimates delocalization stabilization.[74,77,78] Due to the elimination of a basis set or correlation effect altering the results of the B3LYP/6-311G(d,p) reaction profile, the problem must lie in the B3 exchange. The B3 exchange consists of 20% HF, 8% Slater exchange, and 72% B88. Again using the HF results as a reference, one can monitor the effect of B88 by varying the amount incorporated.[78] The B88 exchange causes a significant drop in the transition state barrier with a drop in the product to a slightly lesser degree (SI information). This is again confirmed using NBO analysis (i.e. HF vs. B3-noLYP[79]) and examining the orbital stabilizations associated with the reaction coordinate; small differences are observed in reactant interactions (~2 kcal mol^{-1}), large changes at the transition state (~45 kcal mol^{-1}), and slightly smaller changes at the product (~10 kcal mol^{-1}). This effect is exaggerated as the percentage of B88 exchange is increased and attributed to the overestimation of non-bonded repulsion. Hence, exchange in combination with delocalization leads to an underestimation of the forward barrier: whereas the non-bond repulsion leads to an increase in the forward barrier and the product energy. These counteracting factors explain the seemingly correct description of the forward barrier by B3LYP, while also explaining the erroneous reaction energy when comparing the focal point to DFT.

To study the reaction mechanism of the double proton transfer in formamide the intrinsic reaction coordinate between reactant and product was followed. Figure 3a shows the potential energy along the reaction coordinate at the B3LYP/6-311G(d,p) level. From the potential energy it can be determined that the double proton transfer occurs in a synchronous manner. The energy computed using B3LYP/6-311G(d,p) was benchmarked using nine points determined from focal

point extrapolations assuming B3LYP geometries (see blue points in Figure 3a). There is significant agreement between the focal point benchmark and the energies computed with the B3LYP method. The primary disagreement appears to lie in the product region. The results of Grabowski, Sokalski, and Leszczynski[28] show the main contribution to the interaction energy in the product (the imino ether form) is attributed to the delocalization energy over the entire dimer.

The reaction force profile leads to intriguing information about the double proton transfer of the formamide dimer. As shown in Figure 3b, the reaction force profile is linear from the reactant until close to the minimum of the force, where a pronounced decrease sets in. This linear decrease is due to the monomers decreasing their intermolecular distance to each other with no changes in the distance of the N-H bond, as shown in Figure 3c. The negative linear decrease of the reaction force implies a repulsion energy, which depends quadratically on the distance between the two monomers suggesting an interaction similar to the one in a covalent chemical bond. The pronounced decrease near the minimum initiates the elongation of the N-H bond, which continues until the transition state and reaches into the product region. The change in the nature of the reaction force near the minimum together with the increasing N-H distance indicate the onset of an electronic redistribution, which will be confirmed by the reaction electronic flux and the bond orders. The reaction force in the transition state region crosses zero and decreases linearly in the product region. The transition state region describes the travel of the hydrogen atom from the donor to acceptor atom keeping the distance between the nitrogen atom as donor and the oxygen atom as acceptor constant. This process is dominated by electronic changes. In the product region the reaction force adopts again a linear behavior with almost fixed O-H distance and the two monomers separate from each other. The much smaller structural

rearrangements needed in the product region go inline with a product like transition state in accordance with the Hammond postulate.

The chemical potential (see Supplemental Information) and the reaction electronic flux allow a more detailed description where electronic reorganization takes place along the reaction coordinate. As shown in Figure 3d, the almost zero electronic flux in the reactant region confirms the structural rearrangements needed and the low electronic activity taking place in this part of the potential energy surface. The point on the reaction coordinate where the chemical potential changes and the electronic flux deviates from zero, matches the pronounced decrease of the reaction force (ξ=-1.0). This point describes the beginning of the electronic reorganization, which reaches into the transition state region. The energy prior to this point amounts to approximately 7.7 kcal/mol and constitutes only 39% of the activation barrier (6.6 kcal/mol or 33% taking the energy of the more rigorous focal point method) and can be safely associated to structural reordering that prepares the reaction. The chemical potential together with the reaction electronic flux shows that electronic reordering is necessary to reach the transition state region. This electronic activity is responsible for 67% of the activation barrier. The most substantial electronic reordering occurs in the transition state region. However, the conclusion of the electronic reordering extends to the product region. This assists in explaining the concerted nature of the reaction, showing that there is significant electronic redistribution along a considerable portion of the reaction coordinate.

The onset of the electronic reorganization characterized by the reaction electronic flux is also reflected in the bond indices and their derivatives (see Figure 4a,b). Bond formation and rupture indicated by positive and negative derivative respectively, starts with the onset of the electronic reorganization reflected in the electronic flux and extends into the product region.

Since the derivatives of the bonds between the carbon and nitrogen, or carbon and oxygen, atoms reach their maximum before the other two bonds, one may conclude that prior to the formation and rupture of the H-bonds between acceptor and donor a redistribution of the π-bond is required.

To visualize the electronic reorganization in the dimer the electrostatic potential is mapped on the isosurface of the electron density for the reactant, the transition state and the product in Figure 5 (at B3LYP/6-311G(d,p) level). Positive regions are shown in red and regions with a negative potential in blue. The lone pairs of the electronegative oxygen in the reactant lead to a negative potential shown as a blue region, which is more negative than the nitrogen lone pair. The carbon atom lies in a positive region due to its cationic character in the carbonyl bond. When reaching the transition state and in the product the potential in the vicinity of the nitrogen atom gets more negative due to the transfer of the protons and is oriented in direction of the oxygen and hydrogen atom of the other monomer.

This accumulation of electron density between the two monomers along the H-bonds is in accordance with smaller bond indices for the covalent donor hydrogen (O-H) and larger ones for the acceptor hydrogen (N-H) bonds in the product with respect to the reactant (see Figure 4a). The bond index of the C-N bond with respect to the C=O bond in the reactant also increases. Additionally, the electron density between the two monomers in the imino ether product and the transition state is in agreement with the results of Grabowski et al,[28] which reflect a larger delocalization energy term in the transition state and product than in the reactant. Hence, the treatment of non-bonded repulsion and electron delocalization are the key effects that govern this reaction. Further, the incorrect description of these leads directly to the discrepancies observed with DFT.

CONCLUSIONS

The combination of reaction force, reaction electronic flux and NBO analysis with the focal point method provide a very accurate description of the energetic requirements for each part of the double proton transfer in the formamide dimer. The activation barrier consists of a smaller contribution characterized by pure structural rearrangements (33%) and a larger part, which involves electronic reorganization (67%). The synchronous proton transfer is initiated through the approach of the two monomers, which is followed by electronic reorganization, activating the transfer of the two protons before reaching the transition state in accordance with the Hammond postulate. The delocalization of the electrons throughout the ring formed by the two monomers in the transition state is enhanced in the product due to the electronegativities of the involved oxygen and nitrogen atoms. This effect may be directly correlated to the error seen in the B3 exchange.

REFERENCES

(1) Bell, R. P. *The Tunnel Effect in Chemistry*; Chapman and Hall: New York, 1980.

(2) Melander, L.; Saunders, W. H. J. *Reaction Rates of Isotopic Molecules*; John Wiley and Sons: New York, 1980.

(3) Bender, M. L. *Mechanisms of Homogeneous Catalysis from Protons to Proteins*; John Wiley and Sons: New York, 1971.

(4) Zielinski, T. J.; Poirier, R. A. *J. Comp. Chem.* **1984**, *5*, 466.

(5) Wang, X.-C.; Nichols, J.; Feyereisen, M.; Gutowski, M.; Boatz, J.; Haymet, A. D. J.; Simons, J. *J. Phys. Chem.* **1991**, *95*, 10419.

(6) Wójcik, M. J.; Hirakawa, A. Y.; Tsuboi, M.; Kato, S.; Morokuma, K. *J. Phys. Chem.* **1991**, *95*, 10419.

(7) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P. *J. Phys. Chem. A* **2001**, *105*, 4963.

(8) Del Re, G.; Adamo, C. *J. Phys. Chem.* **1991**, *95*, 4231.

(9) Hrouda, V.; Florián, J.; Polášek, M.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 4742.

(10) Florián, J.; Leszczynski, J.; Johnson, B. G. *J. Mol. Struct.* **1995**, *349*, 421.

(11) Neuheuser, T.; Hess, B. A.; Reutel, C.; Weber, E. *J. Phys. Chem.* **1994**, *98*, 6459.

(12) Hobza, P.; Šponer, J. *J. Mol. Struct.* **1996**, *388*, 115.

(13) Cabaleiro-Lago, E. M.; Rios, M. A. *J. Chem. Phys.* **1999**, *110*, 6782.

(14) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372.

(15) Colominas, C.; Luque, F. J.; Orozco, M. *J. Phys. Chem. A* **1999**, *103*, 6200.

(16) Gresh, N.; Guo, H.; Salahub, D. R.; Roques, B. P.; Kafafi, S. A. *J. Am. Chem. Soc.* **1999**, *121*, 7885.

(17) Brdarski, S.; Åstrand, P.-O.; Karlström, G. *Theor. Chem. Acc.* **2000**, *105*, 7.

(18) Podolyan, Y.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 12103.

(19) Langley, C. H.; Allinger, N. L. *J. Phys. Chem. A* **2003**, *107*, 5208.

(20) Salvador, P.; Dannenberg, J. J. *J. Phys. Chem. B* **2004**, *108*, 15370.

(21) Bende, A.; Suhai, S. *Int. J. Quantum Chem.* **2005**, *103*, 841.

(22) Wu, Q.; Ayers, P. W.; Zhang, Y. *J. Chem. Phys.* **2009**, *131*, 164112.

(23) Florián, J.; Johnson, B. G. *J. Phys. Chem.* **1995**, *99*, 5899.

(24) Frey, J. A.; Leutwyler, S. *Chimia* **2005**, *59*, 511.

(25) Šponer, J.; Hobza, P. *J. Phys. Chem. A* **2000**, *104*, 4592.

(26) Kim, Y.; Lim, S.; Kim, J.-H.; Kim, Y. *J. Phys. Chem. A* **1999**, *103*, 617.

(27) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *265*, 89.

(28) Grabowski, S. J.; Sokalski, W. A.; Leszczynski, J. *J. Phys. Chem. A* **2006**, *110*, 4772.

(29) Tam, C. N.; Bour, P.; Eckert, J.; Trouw, F. R. *J. Phys. Chem. A* **1997**, *101*, 877.

(30) Ohtaki, H.; Funaki, A.; Rode, B. M.; Reibnegger, G. J. *J. Bull. Chem. Soc. Jpn.* **1983**, *56*, 2116.

(31) Lucas, B.; Grégoire, G.; Lecomte, F.; Reimann, B.; Schermann, J. P.; Desfrançois, C. *Mol. Phys.* **2005**, *103*, 1497.

(32) Cybulski, H.; Sadlej, J. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11232.

(33) Herrera, B.; Toro-Labbé, A. *J. Phys. Chem. A* **2007**, *111*, 5921.

(34) Vöhringer-Martinez, E.; Toro-Labbé, A. *J. Comp. Chem.* **2010**, *31*, 2642.

(35) Lamsabhi, A. M.; Mó, O.; Gutiérrez-Oliva, S.; Pérez, P.; Toro-Labbé, A.; Yáñez, M. *J. Comp. Chem.* **2008**, *30*, 389.

(36) Fukui, K. *J. Mol. Struct.* **1981**, *14*, 363.

(37) Martinez, J.; Toro-Labbé, A. *J. Math. Chem.* **2009**, *45*, 911.

(38) Toro-Labbé, A. *J. Phys. Chem. A* **1999**, *103*, 4398.

(39) Jaque, P.; Toro-Labbé, A. *J. Phys. Chem. A* **2000**, *104*, 995.

(40) Toro-Labbé, A.; Gutiérrez-Oliva, S.; Murray, J. S.; Politzer, P. *Mol. Phys.* **2007**, *105*, 2619.

(41) Burda, J. V.; Toro-Labbé, A.; Gutiérrez-Oliva, S.; Murray, J. S.; Politzer, P. *J. Phys. Chem. A* **2007**, *111*, 2455.

(42) Jaque, P.; Toro-Labbé, A.; Politzer, P.; Geerlings, P. *Chem. Phys. Lett.* **2008**, *456*, 135.

(43) Vogt-Geisse, S.; Toro-Labbé, A. *J. Chem. Phys.* **2009**, *130*, 244308.

(44) Echegaray, E.; Toro-Labbé, A. *J. Phys. Chem. A* **2008**, *112*, 11801.

(45) Politzer, P.; Murray, J. S.; Lane, P.; Toro-Labbé, A. *Int. J. Quantum Chem.* **2007**, *107*, 2153.

(46) Politzer, P.; Toro-Labbé, A.; Gutiérrez-Oliva, S.; Herrera, B.; Jaque, P.; Concha, M. C.; Murray, J. *J. Chem. Sci.* **2005**, *117*, 467.

(47) Gutiérrez-Oliva, S.; Herrera, B.; Toro-Labbé, A.; Chermette, H. *J. Phys. Chem. A* **2005**, *109*, 1748.

(48) Sen, K. J.; Jorgensen, C. K. *Electronegativity: Structure and Bonding*; Springer-Verlag: Berlin, 1987; Vol. 66.

(49) Pearson, R. G. *Coord. Chem. Rev.* **1990**, *100*, 403.

(50) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: New York, 1960.

(51) Pearson, R. G. *Hard and Soft Acids and Bases*; Dowden, Hutchinson, & Ross: Stroudsberg, 1973.

(52) Pearson, R. G. *J. Am. Chem. Soc.* **1985**, *107*, 6801.

(53) Ayers, P. W.; Parr, R. *J. Am. Chem. Soc.* **2000**, *122*, 2010.

(54) Geerlings, P.; Proft, F. D.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793.

(55) East, A. L. L.; Allen, W. D. *J. Chem. Phys.* **1993**, *99*, 4638.

(56) Kenny, J. P.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **2003**, *118*, 7353.

(57) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Tarczay, G.; Császár, A. G. *Chem.-Eur. J.* **2003**, *9*, 2173.

(58) Császár, A. G.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **1998**, *108*, 9751.

(59) Hargis, J. C. E., F. A.; Ingels, J. B.; Schaefer, H. F.; *J. Am. Chem. Soc.* **2008**, in press.

(60) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

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

(62) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.

(63) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899-926.

(64) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F.; 3.1 ed.; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 1994.

(65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., -J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.;

Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; 02 ed. Wallingford, CT, 2004.

(66) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823-&.

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

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

(69) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Clelani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; 2006.1 ed. 2006.

(70) Lindh, R.; Ryu, U.; Liu, B. *J. Chem. Phys.* **1991**, *95*, 5889-5897.

(71) Hampel, C.; Peterson, K. A.; Werner, H. J. *Chem. Phys. Lett.* **1992**, *190*, 1-12.

(72) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan,

R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys Chem Chem Phys* **2006**, *8*, 3172-3191.

(73) Kobayashi, Y.; Kamiya, M.; Hirao, K. *Chem. Phys. Lett.* **2000**, *319*, 695.

(74) Gritsenko, O. V.; Ensing, B.; Schipper, P. R. T.; Baerends, E. J. *J. Phys. Chem. A* **2000**, *104*, 8558.

(75) Larkin, J. D.; Bhat, K. L.; Markham, G. D.; Brooks, B. R.; Schaefer, H. F.; Bock, C. W. *J. Phys. Chem. A* **2006**, *110*, 10633.

(76) Moran, D.; Simmonett, A. C.; Leach, F. E.; Allen, W. D.; Schleyer, P. v. R.; Schaefer, H. F. *J. Am. Chem. Soc.* **2006**, *128*, 9342.

(77) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 204112.

(78) Woodcock, H. L.; Schaefer, H. F.; Schreiner, P. R. *J. Phys. Chem. A* **2002**, *106*, 11923.

(79) B3-noLYP is B3LYP without LYP correlation. It is the B3 exchange term.

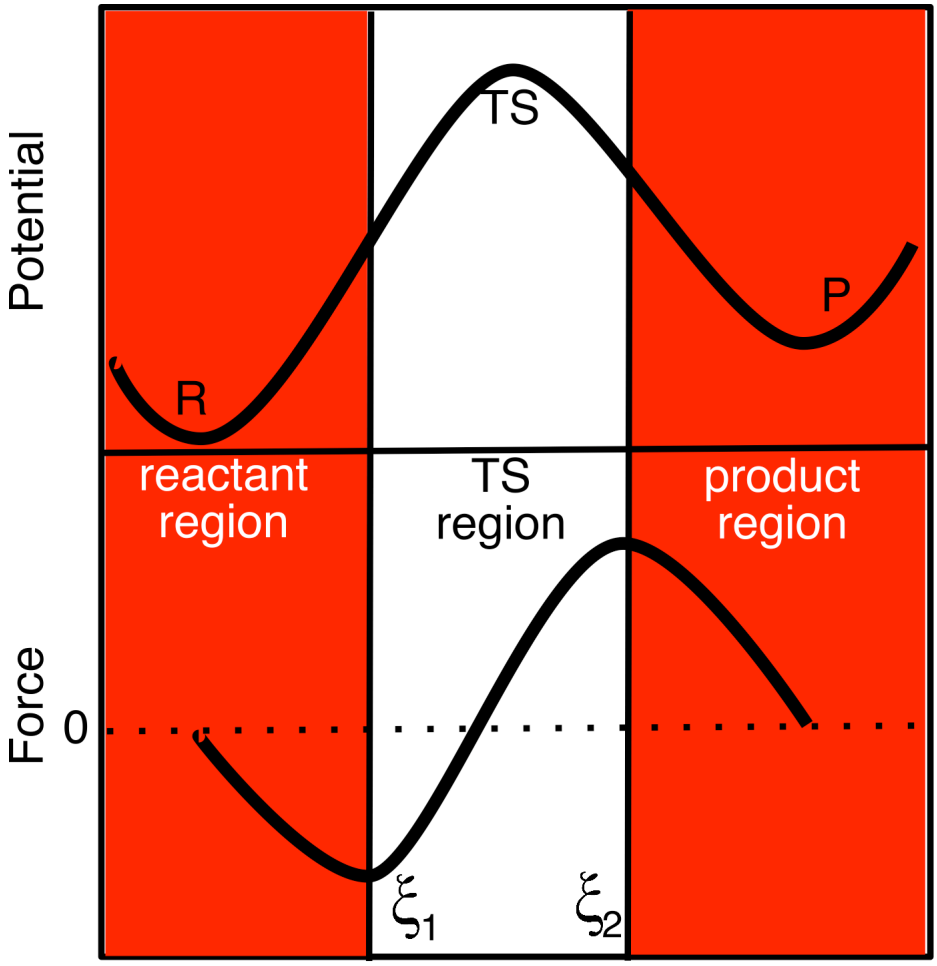


Figure 3.1. Schematic representation of the potential energy along the reaction coordinate for an elementary process and the reaction force providing the limits of characteristic regions of the reaction by its minimum and maximum.

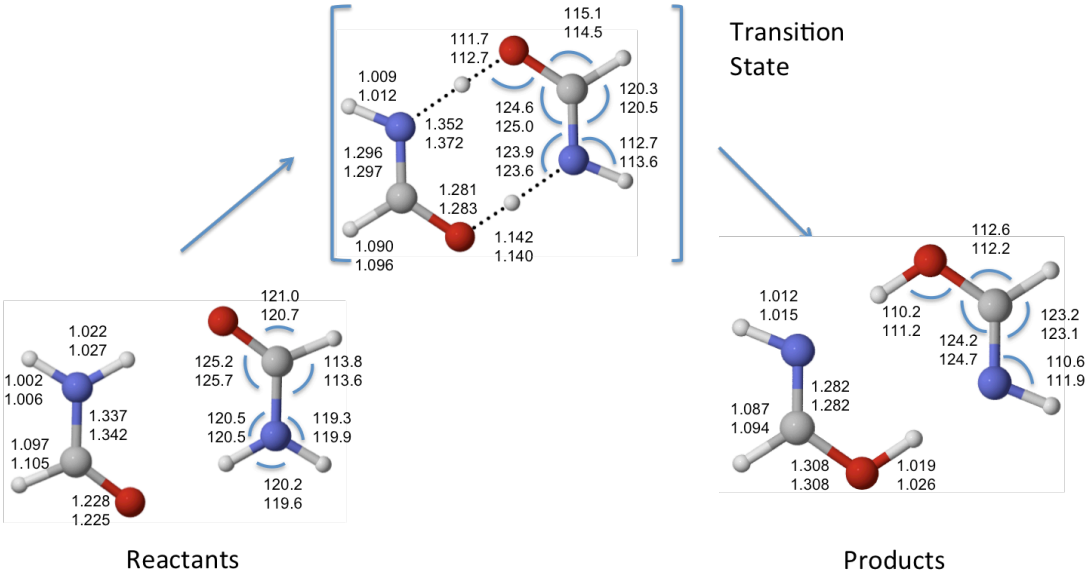


Figure 3.2. Schematic depiction of the formamide dimer undergoing the double proton transfer reaction. The bond distances in Ångstroms and the angles in degrees are shown for the MP2/cc-pVQZ (number on top) and the B3LYP /6-311G(d,p) (number at bottom) optimized geometries for reactants, transition state and products.

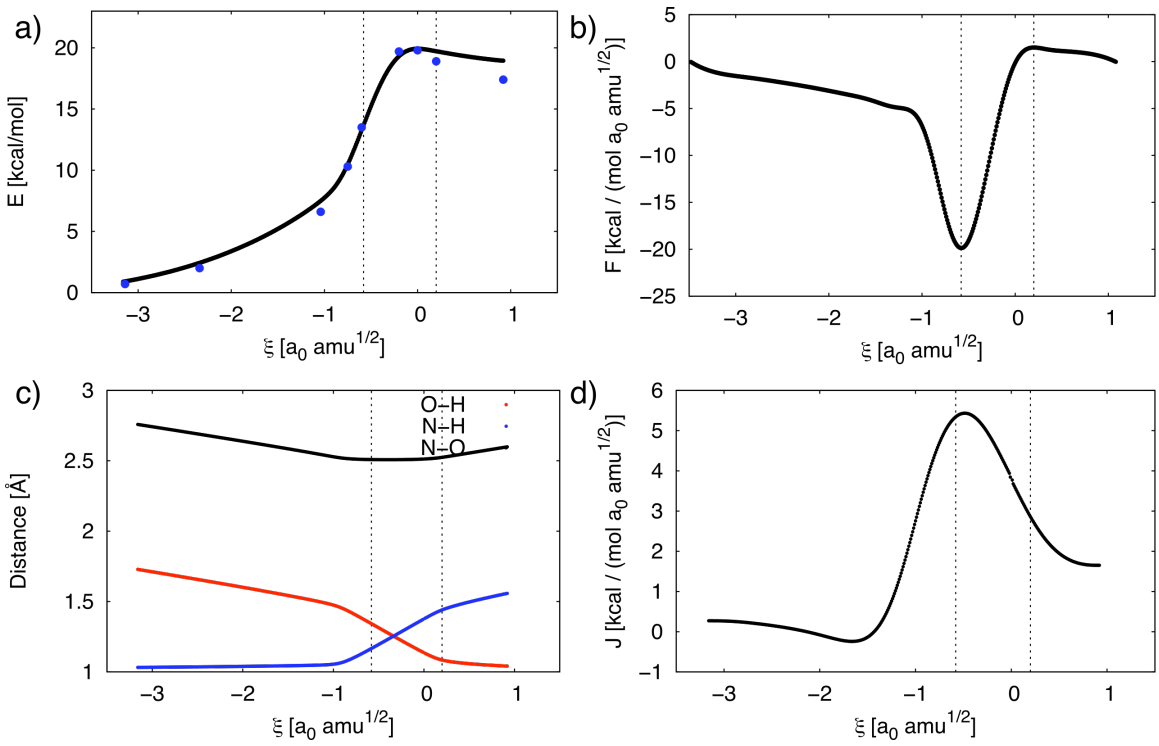


Figure 3.3. a) Potential energy along the intrinsic reaction coordinate computed with the B3LYP/6-311G(d,p) method in comparison to focal point energy analysis of characteristic geometries (blue circles). b) Reaction force and the dashed lines showing the minimum and maximum of the reaction force. c) Distance between the nitrogen or oxygen atom and the transferred hydrogen atom, and the nitrogen-oxygen distance representing the approximation of the two monomers. d) Reaction electronic flux obtained as the derivative of the chemical potential along the reaction path.

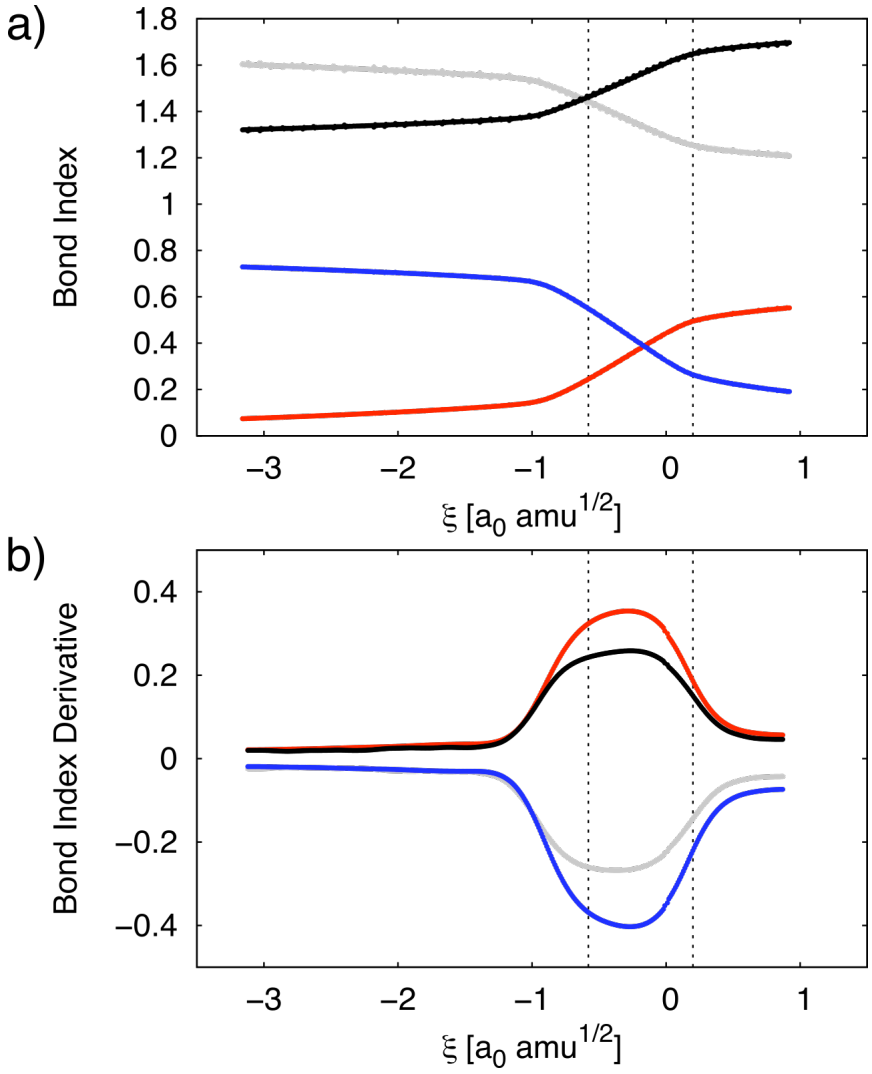


Figure 3.4. a) Wiberg bond indices obtained from a NBO analysis: N-H bond (blue), O-H bond (red), C-N bond (black) and C-O bond (gray). b) Derivative of the bond indices with respect to the reaction coordinate: positive values represent bond formation and negative values bond rupture.

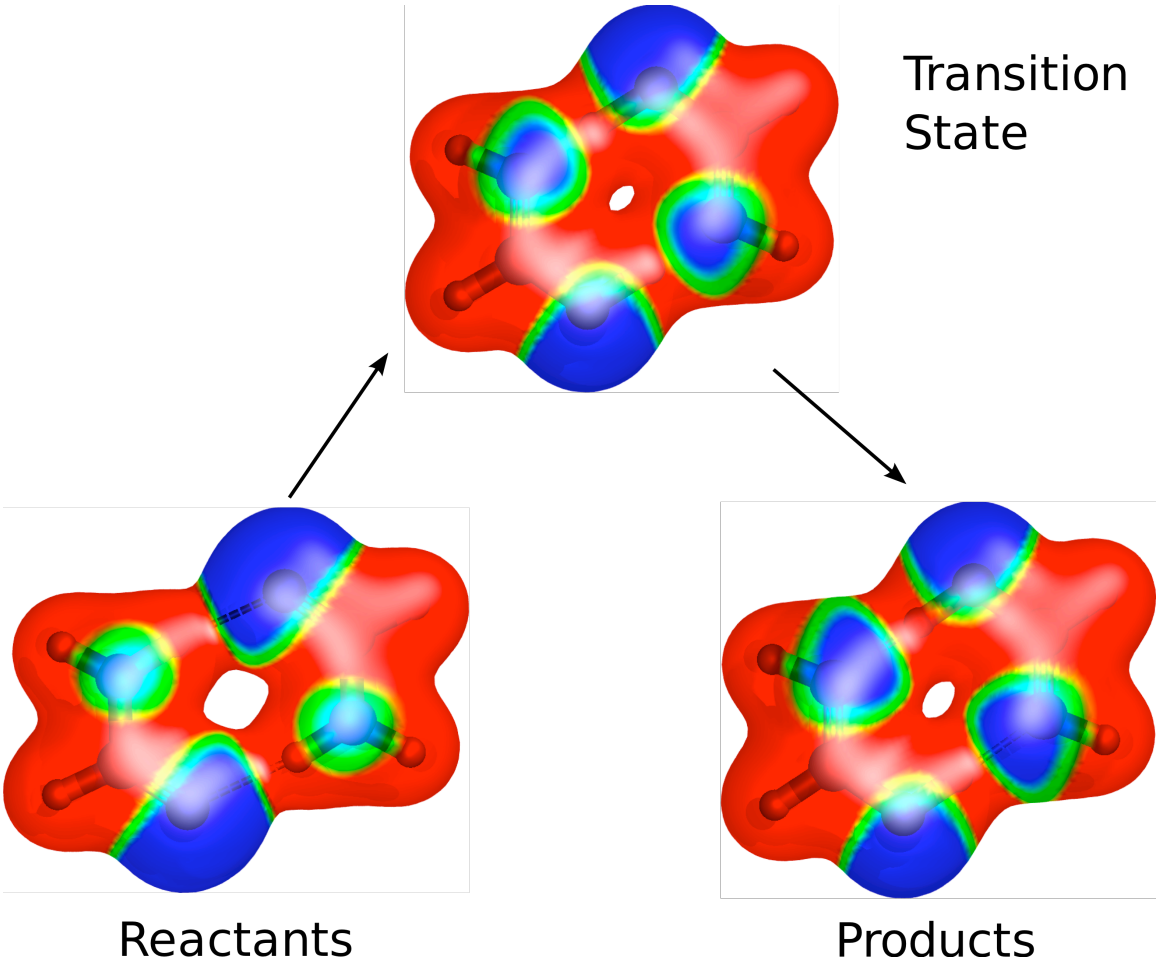


Figure 3.5. Electrostatic potential mapped on the electron density isosurface (isovalue = 0.01 e/a_0^3) for reactant, transition state and product (red = +0.02 e and blue= -0.02 e).

basis set	ΔE_e [RHF]	+δ [MP2]	+δ [CCSD]	+δ [CCSD(T)]	ΔE_e [CCSD(T)]
[a]					
cc-pVDZ	29.44	-11.45	1.97	-1.64	18.32
cc-pVTZ	30.69	-12.45	2.49	-2.01	18.71
cc-pVQZ	31.12	-12.35	2.66	-2.03	19.40
cc-pV5Z	31.19	-12.22	[2.66]	[-2.03]	[19.60]
CBS limit	[31.21]	[-12.08]	[2.78]	[-2.04]	[19.87]
$\Delta E^{\ddagger} = \Delta E_e$[CBS CCSD(T)] = **19.87 kcal mol^{-1}**					
[b]					
cc-pVDZ	21.68	-5.75	-0.28	-0.77	14.88
cc-pVTZ	22.72	-6.27	0.03	-0.96	15.52
cc-pVQZ	23.06	-6.02	0.15	-0.94	16.26
cc-pV5Z	23.12	-5.84	[0.15]	[-0.94]	[16.49]
CBS limit	[23.14]	[-5.65]	[0.24]	[-0.92]	[16.81]
$\Delta E_{reaction} = \Delta E_e$[CBS CCSD(T)] = **16.81 kcal mol^{-1}**					
$\Delta E_{reverse} = \Delta E^{\ddagger} - \Delta E_{reaction}$ = 19.87 kcal mol^{-1} – 16.81 kcal mol^{-1} = **3.06 kcal mol^{-1}**					
Fit	$a+be^{-cX}$	$a+bX^{-3}$	$a+bX^{-3}$	$a+bX^{-3}$	Additive
Points (X)	3, 4, 5	4, 5	3, 4	3, 4	

Table 3.1. [a] The focal point table for the MP2 / cc-pVQZ barrier for formamide double proton transfer. [b] The focal point table for the MP2 / cc-pVQZ reaction energy. The remaining focal point tables will be supplied in the supplemental information.

	MP2 // Focal Point	B3LYP // Focal Point	B3LYP // B3LYP	B3LYP // HF
$\Delta E_f^{\ddagger}$	19.9	19.9	19.9	31.0
$\Delta E_{reaction}$	16.8	16.9	18.7	23.2
$\Delta E_r^{\ddagger}$	3.1	3.0	1.2	7.8

Table 3.2. The table above shows energy barriers for the forward reaction, reaction energy, and the reverse barrier. The HF energies were computed using the B3LYP geometry.

CHAPTER 4

NON-COVALENT INTERACTIONS OF A BENZO[*A*]PYRENE DIOL EPOXIDE WITH DNA BASE PAIRS: INSIGHT INTO THE FORMATION OF ADDUCTS OF (+)-B*A*P DE-2 WITH DNA[1]

[1] Reproduced with permission from Hargis, J.C.; Schaefer, H. F.; Houk, K. N.; Wheeler, S. E. *J. Phys. Chem. A* **2010***, 114*, 2038. Copyright 2010 American Chemical Society.

ABSTRACT

Non-covalent complexes of a tumorigenic benzo[*a*]pyrene diol epoxide with the guanine-cytosine and adenine-thymine base pairs have been examined computationally. (+)-B*a*P DE-2 forms covalent adducts with DNA via nucleophilic attack on the (+)-B*a*P DE-2 epoxide. Computational results predict five thermodynamically accessible complexes of AT with (+)-B*a*P DE-2 that are compatible with intact DNA. Among these, two are expected to lead to adenine adducts. In the lowest energy AT···(+)-B*a*P DE-2 complex, which has a gas-phase interaction energy of –20.9 kcal mol^{-1}, the exocyclic NH_2 of adenine is positioned for backside epoxide attack and formation of a *trans* adduct. The most energetically favorable complex leading to formation of a *cis* ring-opened adduct lies only 0.6 kcal mol^{-1} higher in energy. For GC···(+)-B*a*P DE-2, there are only two thermodynamically accessible complexes. The higher-lying complex, bound in the gas phase by 24.4 kcal mol^{-1} relative to separated GC and (+)-B*a*P DE-2, would lead to a *trans* ring opened N^2-guanine adduct. In the global minimum energy GC···(+)-B*a*P DE-2 complex, bound by 27.3 kcal mol^{-1}, the exocyclic NH_2 group of cytosine is positioned for *cis* epoxide addition. However, adducts of (+)-B*a*P DE-2 with cytosine are rarely observed experimentally. The paucity of cytosine adducts, despite the predicted thermodynamic stability of this GC···(+)-B*a*P DE-2 complex, is attributed to the electrostatic destabilization of the benzylic cation intermediate thought to precede *cis* addition.

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs), a main component of soot, are widely distributed in the environment due primarily to the combustion of biomass and fossil fuels.[1] These PAHs form in flames via well-characterized pathways starting with the formation of an initial aromatic ring (usually benzene),[2-4] and are known to be carcinogenic.[3-7] Similarly,

components of tobacco smoke have been shown to cause cancer, though the complexity of the underlying mechanisms has rendered a complete atomic-level understanding of the process elusive. Among the carcinogenic and tumorigenic components of tobacco smoke, PAHs are known to play a major role.[8] Benzo[*a*]pyrene (B*a*P) is one of the most thoroughly studied carcinogenic PAHs in tobacco smoke, and also one of the most abundant.[9]

In mammals, B*a*P is metabolized to a bay region diol epoxide, B*a*P DE. Four stereoisomers of B*a*P DE are formed (Fig. 1a), of which (+)-B*a*P DE-2[10] is both the most abundant and most tumorigenic.[11] Intercalation of (+)-B*a*P DE-2 into DNA is followed by nucleophilic attack on the epoxide by the exocyclic NH_2 of adenine or guanine.[12] Guanine adducts predominate, although adenine adducts can also be formed in significant quantities.[13-15] Yields of cytosine adducts are much smaller,[13] and the structures of these adducts were not definitively characterized until recently.[16] Nucleophilic attack on the epoxide leads to a covalently bound DNA adduct via either *cis* or *trans* ring opening.[17] *Cis* addition is postulated to occur via a resonance-stabilized benzylic cation intermediate,[18] while *trans* ring-opening can proceed either from this cationic intermediate or by direct backside attack on the epoxide. Covalent DNA adducts inhibit enzymes such as helicase[19] and topoisomerase I.[20] These DNA adducts exhibit diverse, sequence-dependent mutagenic behavior, a feature that has been attributed to the presence of different stable adduct conformations.[21-26] Tobacco-smoke-related cancers are typically due to mutations in the *p53* tumor suppressor gene arising from the preferential addition of (+)-B*a*P DE-2 to specific sequences in this gene.[6]

Most previous investigations of the effects of (+)-B*a*P DE-2 on DNA, both experimental[6,27,28] and theoretical,[29-34] have focused on the covalent adducts with DNA oligomers or single nucleotides. Of particular recent interest has been the characterization of the

conformational behavior of (+)-B*a*P DE-2−DNA adducts. NMR solution data have confirmed the presence of multiple stable conformations depending on the flanking nucleobases.[35,36] In 2004, the first crystal structures of B*a*P DE-2–DNA adducts were published[27,37] providing, along with previously published NMR structures,[37-40] invaluable details regarding the structure of these adducts. A crystal structure of an adduct with B*a*P DE bound to N^2-deoxyguanosine enabled the assignment of absolute configurations to the four optically active B*a*P DE isomers and the eight associated dG adducts resulting from the *cis* and *trans* ring opening of the epoxide.[37] A second crystal structure of a B*a*P DE-2–adenine adduct in a ternary complex with DNA polymerase was published shortly thereafter.[27]

While information about (+)-B*a*P DE-2···DNA complexes can be inferred from the structures of these adducts, a complete understanding of the addition of (+)-B*a*P DE-2 to DNA will require explicit examination of the complexes that precede epoxide attack. Harvey and co-workers[41] studied the formation of non-covalent intercalative complexes of (+)-B*a*P DE-2 and (–)-B*a*P DE-1 with DNA via kinetic flow linear dichroism experiments. Structural differences between these non-covalent complexes were highlighted that presumably underlie the different mutagenic and tumorigenic activities of these diastereomers. Computational methods are ideally suited to provide additional details regarding non-covalent (+)-B*a*P DE-2···DNA complexes, because it is possible to directly probe their structures and thermochemistry and to quantify the role of π-stacking and individual hydrogen bonding interactions. Previously, the inability of popular density functional theory (DFT) functionals[42,43] to accurately describe the π-stacking interactions that drive the intercalation of (+)-B*a*P DE-2 into DNA hampered high-level computational studies of such systems. Recent advances in the development of DFT

functionals[44,45] have led to methods that may accurately describe π-stacking, opening the door for quantum mechanical studies of stacking phenomena in myriad biological contexts.

Several DNA intercalators have previously been examined by *ab initio* and DFT methods,[46-48] with a focus on intercalators that are utilized in antitumor chemotherapy. In 2002, Hobza and co-workers[47] studied the π-stacking interactions of ethidium, daunomycin, ellipticine, and 4,6'-diaminide-2-phenylindole with DNA base pairs. It was shown that in each of these cases the net attraction arises from the competing effects of electrostatic and dispersion interactions and short-range exchange repulsion. In 2006, Leszczynski and co-workers[48] examined the nature of interactions between ethidium and proflavine with DNA bases starting from published crystal structures. A more recent study[46] of the interaction of ellipticine and proflavine with DNA base pairs showed that stabilizing interactions are maximized when the main axis of the intercalator is nearly aligned with the main axis of the bases. In contrast to (+)-B*a*P DE-2, these previously studied intercalating agents interact with DNA solely through π-stacking interactions, rather than a combination of intermolecular hydrogen bonds and π-stacking.

Despite decades of study of the effects of B*a*P DE on DNA, there have been few detailed explorations of the non-covalent (+)-B*a*P DE-2···DNA complexes that precede adduct formation.[41] Consequently, there are details concerning the complexation of (+)-B*a*P DE-2 with DNA base pairs that warrant further exploration. A lingering conundrum concerns the scarcity of observed cytosine adducts,[16] despite the prevalence of adducts of cytosine with diol epoxides derived from dibenz[*a*,*j*]anthracene and benz[*a*]anthracene.[49-51] Presumably, this could arise from disfavored non-covalent interactions between cytosine and (+)-B*a*P DE-2, an insurmountable free energy barrier for epoxide attack by cytosine, or some combination of these two factors.

Similarly, the energetically favored orientation of (+)-B*a*P DE-2 relative to a given base pair prior to adduct formation has not been established. It has been proposed[21-26] that the existence of multiple conformations of (+)-B*a*P DE-2–DNA adducts is responsible for their divergent mutagenic behavior. The geometry of the complex preceding adduct formation could presumably play a role in the formation of these different conformers. Finally, understanding the role of individual π-stacking and hydrogen bonding interactions in these complexes will help further unravel the unique features of (+)-B*a*P DE-2 that lead to its pronounced tumorigenicity. As a first step toward addressing these issues, we have examined the complexes of (+)-B*a*P DE-2 with the AT and GC base pairs to gain insight into the factors governing the addition of (+)-B*a*P DE-2 to DNA.

METHODS

Accurate *ab initio* descriptions of π-stacking interactions require rigorous correlated theoretical methods paired with large basis sets.[52] However, (+)-B*a*P DE-2···DNA base pair complexes, which comprise 66 and 67 atoms with GC and AT, respectively, are too large to treat with such rigorous approaches. Unfortunately, many DFT functionals, which are the most popular methods for computational investigations of systems of this size, fail to accurately describe the dispersion effects that underlie stacking interactions.[42,43] However, the M05-2X and M06-2X functionals have been shown to provide accurate interaction energies for stacked dimers.[44,53-56] Hohenstein, Chill, and Sherrill[57] recently showed that M06-2X in particular performs well for a standard benchmark set of stacked complexes. Also, among DFT functionals including empirical dispersion corrections, PBE-D has been shown[57,58] to yield stacking energies of comparable quality to M06-2X. Similarly, Gu *et al*.[59] recently demonstrated that the M06-2X

functional paired with a double zeta basis set yields interaction energies in good agreement with reliable CCSD(T) results for stacked nucleic acid bases.

Geometry optimizations were performed for two conformers of (+)-B*a*P DE-2 with each of the DNA base pairs using the M05-2X functional. The potential energy surface of each base pair with each conformer of (+)-B*a*P DE-2 was explored by executing optimizations from several initial geometries for each relative orientation of the stacked system. Starting structures were generated by varying the position of (+)-B*a*P DE-2 relative to the base pair in ~2 Å increments while keeping the molecular planes of the two species parallel, to ensure that all low-lying configurations were sampled. Preliminary optimizations were carried out using the M05-2X functional paired with the 3-21G basis set. Structures lying within three kcal mol^{-1} of the predicted global minimum were then further refined using the 6-31+G(d) basis set. To verify the M05-2X optimized structures, M06-2X/6-31+G(d) optimizations were also carried out on the minimum energy and second lowest-lying AT and GC complexes. Differences between the M05-2X and M06-2X geometries were minor. Presented gas-phase interaction energies are M06-2X/6-31+G(d) electronic energies evaluated at M05-2X/6-31+G(d) geometries, and are given relative to separated base pair and conformer **I** of (+)-B*a*P DE-2. PBE-D/aug-cc-pVDZ single point energies[45] were also evaluated at M05-2X geometries for the two lowest lying AT and GC complexes. The PBE-D/aug-cc-pVDZ interaction energies for the AT complexes are similar to the M06-2X/6-31+G(d) data, differing by ±0.6 kcal mol^{-1}. On the other hand, the PBE-D interaction energies for the two lowest-lying GC complexes are smaller than the M06-2X values by 3.5 and 1.9 kcal mol^{-1}. In both cases, the energy ordering of the complexes is unchanged.

All computations were performed using NWChem.[60,61] For the M05-2X and M06-2X computations a fine DFT integration grid was used, consisting of 70 radial shells and 590 angular

points, since these functionals are known to be sensitive to the choice of integration grid.[42,62-64] Standard atomic labels are utilized for the nucleobases, which are depicted in Fig. 2. Relevant atom designations for (+)-B*a*P DE-2 are shown in Fig. 1a and denoted by subscripts in the text.

RESULTS AND DISCUSSION

Conformers of Free (+)-B*a*P DE-2

In the gas phase, there are two low-lying conformers of (+)-B*a*P DE-2 (**I** and **II**, Fig. 3a). In the higher-lying conformer (**II**), the hydroxyl groups occupy pseudo-axial positions, whereas the OH groups in **I** are equatorial. The M06-2X/6-31+G(d)//M05-2X/6-31+G(d) predicted energy difference is 2.1 kcal mol^{-1}, with a conformational barrier of 7.8 kcal mol^{-1}. Associated with each of these conformers are other low-lying minima, connected to **I** and **II** via changes in OH orientations. Both conformers **I** and **II** are stabilized by intramolecular hydrogen bonds that must be broken during the change in ring conformation. Such interactions will be less important in an aqueous environment, and the conformational barriers will be smaller in solution. Most importantly, both conformers **I** and **II** will be thermodynamically accessible and rapidly inter-converting at biologically relevant temperatures.

(+)-B*a*P DE-2···DNA Base Pair Complexes

Gas-phase interaction energy surfaces for (+)-B*a*P DE-2 with both the GC and AT base pairs have been examined as a model for the interaction of (+)-B*a*P DE-2 with DNA. When (+)-B*a*P DE-2 approaches a given base pair,[65] there are four possible relative orientations: the epoxide functionality can be directed toward or away from the base pair and the functionalized end of (+)-B*a*P DE-2 can extend into the minor or major groove of intact DNA. Additional variation arises from the complexation of (+)-B*a*P DE-2 with AT or TA and GC or CG. For each

of these arrangements, multiple relative positions of the centers of mass of the base-pair and (+)-B*a*P DE-2 were considered to ensure that the minimum energy complex has been obtained for each relative orientation. In total, 48 distinct optimizations were carried out for each conformer.

The optimized structures (Figs. 3b and 4) are named according to the relative orientation of (+)-B*a*P DE-2 and the base pair as follows: the orientation of the epoxide towards or away from the base pair is indicated by **E** or **e**, respectively. **M** or **m** indicates that the functionalized end of (+)-B*a*P DE-2 extends into the major or minor groove, respectively. Finally, the ordering of the base pair (*e.g.*: GC versus CG) is explicitly given. For example, the global minimum GC···(+)-B*a*P DE-2 complex is labeled **CG(I-EM)**, signifying that (+)-B*a*P DE-2 is complexed with the epoxide functionality directed towards cytosine-guanine, with the functionalized end extending into the major groove of DNA. It should be noted that complexes in which the epoxide is facing the base pair (denoted by **E**) can only lead to *cis* adducts, while structures in which the epoxide is facing away from the base pair (**e**) are only compatible with *trans* addition.

One of the primary goals of the current work is to gain insight into the non-covalent complexes of (+)-B*a*P DE-2 with DNA that precede covalent adduct formation. Stacking of a single DNA base pair with (+)-B*a*P DE-2 serves as the simplest possible model of the interaction with DNA. Rather than constraining optimizations to be compatible with intact DNA, unconstrained optimizations were executed and final structures incompatible with DNA were eliminated. Since the 2-deoxyribose of DNA was replaced with a hydrogen atom in our model, eliminated structures include five complexes in which the purine N^9 or pyrimidine N^1 act as hydrogen bond donors. Also excluded were two structures in which the base pair undergoes significant distortions that would be improbable in intact DNA due to backbone constraints and the presence of flanking base pairs. One such structure [**AT(I-eM)**] is depicted in Fig. 5. Some of

the excluded structures were otherwise competitive energetically with the global minimum complexes.

Guanine-Cytosine···(+)-B*a*P DE-2 Complexes

The present results indicate that (+)-B*a*P DE-2 will preferentially form complexes with the guanine-cytosine base pair; the global minimum GC···(+)-B*a*P DE-2 complex (Fig. 3b) lies 6.4 kcal mol^{-1} lower in energy than the global minimum energy AT···(+)-B*a*P DE-2 complex (Fig. 4) at the M06-2X/6-31+G(d) level of theory. PBE-D/aug-cc-pVDZ predicts a smaller energy difference of 2.4 kcal mol^{-1}. In the global minimum GC structure [**CG(I-EM)**], conformer **I** of (+)-B*a*P DE-2 is complexed with the epoxide functionality directed towards CG. The gas-phase interaction energy relative to separated GC and conformer **I** of (+)-B*a*P DE-2 is 27.3 kcal mol^{-1}. This complex is stabilized by two somewhat strained hydrogen bonds joining the epoxide oxygen with the exocyclic NH_2 of cytosine and the C_7 hydroxyl group with O^6 on guanine. In order to maintain these intermolecular hydrogen bonds and the favorable stacking interaction of cytosine with the aromatic core of (+)-B*a*P DE-2, the base pair is distorted slightly from planarity. In intact DNA there will be an associated energetic cost due to interactions with flanking base pairs and this structure will lie higher in energy in a more complete DNA model. The oxidized end of (+)-B*a*P DE-2 interacts with the major groove; as a result, formation of a covalent guanine adduct is impossible, since the exocyclic NH_2 of guanine is directed away from the epoxide. Instead, this complex is compatible with formation of an N^4-cytosine adduct via *cis* ring opening of the epoxide, which only form in small quantities.[16]

There is only one additional thermodynamically accessible GC···(+)-B*a*P DE-2 complex [**CG(II-em)**], lying 2.9 kcal mol^{-1} above the global minimum. In this case, it is the higher-lying conformer (**II**) that is complexed with CG, with the epoxide functionality directed away from the

base-pair and the functionalized end in the minor groove. As such, this complex is poised for back-side nucleophilic attack and formation of the frequently observed *trans* N^2-guanine adducts. This complex is stabilized by two cooperative hydrogen bonds involving the C_7 OH group and the exocyclic NH_2 and N^7 of guanine.

Adenine-Thymine···(+)-B*a*P DE-2 Complexes

Even though the minimum-energy AT···(+)-B*a*P DE-2 complex lies 6.4 kcal mol^{-1} higher in energy than the global minimum GC···(+)-B*a*P DE-2 complex, it is important to consider the AT case since covalent adenine adducts frequently form. In contrast to the GC···(+)-B*a*P DE-2 complexes, for which there was only one structure within three kcal mol^{-1} of the global minimum, for AT···(+)-B*a*P DE-2 there are five thermodynamically accessible complexes that are compatible with intact DNA.

The two lowest-lying AT···(+)-B*a*P DE-2 complexes, **AT(II-eM)** and **TA(II-em),** involve conformer **II** of (+)-B*a*P DE-2 with the epoxide functionality directed away from the base pair. These complexes are bound by 20.9 and 20.4 kcal mol^{-1} in the gas phase, respectively. In **AT(II-eM)**, (+)-B*a*P DE-2 is complexed with AT with the functionalized end directed into the major groove, while in **TA(II-em)** the complex involves TA and the functionalized end of (+)-B*a*P DE-2 is directed towards the minor groove. Consequently, in **AT(II-eM)** the exocyclic NH_2 is positioned for backside attack and *trans* adduct formation, while no covalent adduct could form from **TA(II-em)**.

A complex of conformer **I** with AT was also optimized **[AT(I-eM)]**, and it is 0.5 kcal mol^{-1} lower than **AT(II-eM)**. However, as seen in Fig. 5, when complexed with **I**, the AT base pair is drastically distorted from planarity to maintain the cyclic hydrogen bonding arrangement between the C_7 OH group on (+)-B*a*P DE-2 and atoms N^6 and N^7 of adenine and stacking

interaction between thymine and the pyrene. As mentioned above, these extreme distortions of the AT base pair are incompatible with the structure of intact DNA and such complexes are not expected to occur in intact DNA.

There are three other structures that are within 3 kcal mol^{-1} of **AT(II-eM)**, with the most favorable of these [**AT(I-EM)**] higher in energy by only 0.6 kcal mol^{-1}. This is the lowest-lying AT complex with conformer **I** that is compatible with intact DNA. In **AT(I-EM),** (+)-B*a*P DE-2 is oriented with the epoxide facing the base pair and the functionalized end extending into the major groove. In this complex, both N^6 and N^7 are poised for front-side attack on the epoxide, leading to either *cis* N^6-adenine adduct or the less frequently observed N^7 adduct. The next higher-lying structure, **TA(II-EM),** is 2.2 kcal mol^{-1} above the global minimum and features (+)-B*a*P DE-2 bound with the epoxide towards TA and the functionalized end extending into the major groove. This complex is not expected to lead to adduct formation, since N^6 is not near the epoxide carbon. The final structure, **TA(I-Em)**, also involves conformer **I**, this time complexed with TA with the epoxide towards the base pair and the functionalized end in the minor groove. No adduct can be formed from **TA(I-Em)**.

Structures **CG(II-em)** and **AT(II-eM)** provide a demonstration of the differential stacking avidities of (+)-B*a*P DE-2 with the GC and AT base pairs. These two structures exhibit similar hydrogen bonding interactions. Both include a cyclic hydrogen bonding arrangement of the C_7 OH group on (+)-B*a*P DE-2 with the exocyclic NH_2 and ring nitrogen of the purine base. The hydrogen bonds in the GC complex are slightly weaker, yet the GC complex is more strongly bound by 3.5 kcal mol^{-1}.[66] This difference is attributed to a stronger π-stacking interaction with GC over AT, which is consistent with previous findings for proflavin.[46] This difference can be understood qualitatively in terms of simple electrostatic effects. Electrostatic

potential surfaces for conformer **II** of (+)-B*a*P DE-2 and the GC and AT base pairs are shown in Fig. 6. These plots provide a simple tool for understanding the electrostatic component of non-covalent interactions.[67-69] The primary difference in ESPs between the AT and GC base pairs is the sign of the ESP surrounding cytosine-N^4 compared to O^4 on thymine. In **AT(II-eM)** there will be an unfavorable electrostatic interaction between the negative ESP surrounding O^4 on thymine and the negative ESP above the pyrene. The corresponding electrostatic interaction with N^4 on cytosine will be favorable. In other words, the AT···(+)-B*a*P DE-2 complex is destabilized by a direct electrostatic interaction[70] between thymine-O^4 and the pyrene, while the analogous direct interaction with the cytosine NH_2 stabilizes **CG(II-em)**.

Implications for DNA Adduct Formation

The present computations employ a simple model of the interactions of B*a*P DE-2 with DNA consisting of gas-phase complexes between (+)-B*a*P DE-2 and a single base pair. Regardless, some insight into the formation of covalent adducts between (+)-B*a*P DE-2 and DNA can be gleaned from the computed structures. The present results predict that the higher-lying conformer of (+)-B*a*P DE-2 (conformer **II**) leads to formation of *trans* adducts with both guanine and adenine, via complexes **CG(II-em)** and **AT(II-eM)**, respectively. Conformer **I** is predicted to undergo *cis* addition to yield an adenine adduct via structure **AT(I-EM)**. The lowest-lying GC complex leading to *cis* addition of guanine (Fig. 3b) lies 6.4 kcal mol^{-1} above the global minimum, which is isoenergetic with the minimum energy AT complex. The global minimum GC structure [**CG(I-EM)**] will not lead to guanine adduct formation, and is separated from the next lowest-lying structure by 2.9 kcal mol^{-1}. This is in contrast to the AT complexes, for which both the lowest-lying structure [**AT(II-eM)**] and a second complex 0.6 kcal mol^{-1} higher in energy [**AT(I-EM)**] are expected to both lead to covalent adducts.

The prediction that the thermodynamically preferred complex, **CG(I-EM)**, is positioned for formation of an N^4-cytosine adduct is inconsistent with the paucity of observed cytosine adducts.[16] As noted above, in this structure there is some distortion of the base pair from planarity and it would likely lie slightly higher in energy in a more complete model of DNA. Additionally, the effects of solvent, the sugar-phosphate backbone, or the presence of a sandwiching base pair could all alter the predicted energetic ordering in a more realistic model of DNA. On the other hand, there could be some mechanistic origin preventing cytosine attack. One possibility arises from differences in the electrostatic potential surrounding cytosine compared to guanine or adenine. The ESP surrounding cytosine is mostly positive (see Fig. 6), in contrast to the more negative ESP surrounding the purine bases. Front-side epoxide attack is postulated to occur through a fleeting benzylic cation intermediate.[18] The positive ESP surrounding cytosine will destabilize this incipient cation, raising the energy of this intermediate and associated reaction barriers and preventing cytosine addition to (+)-B*a*P DE-2.

The two low-lying conformers of (+)-B*a*P DE-2 lead to qualitatively different intermolecular hydrogen bonds with the GC and AT base pairs. More importantly, the complexes leading to *trans* adduct formation [**CG(II-em)** and **AT(II-eM)**] involve conformer **II**, while conformer **I** is present in the complexes compatible with *cis* adenine and guanine addition [**AT(I-EM)** and **CG(I-Em)**]. These latter complexes feature a hydrogen bond between the exocyclic NH_2 of the purine base and the epoxide oxygen. Additional stabilization is achieved via hydrogen bond contacts with the OH connected to C_7, which are only possible for conformer **I**. Conversely, for structures leading to *trans* adducts, the exocyclic NH_2 is hydrogen bonded to the C_7 hydroxyl group. Forming this hydrogen bond while simultaneously maintaining favorable stacking interactions with the pyrene and a planar base pair arrangement requires conformer **II**.

The *trans* arrangement in (+)-B*a*P DE-2 enables this OH group to position the AT base pair with the NH_2 group ideally oriented for backside attack on the epoxide. Overall, the tendency for formation of a strong hydrogen bond with the this OH group combined with the maximization of π-stacking interactions between the pyrene moiety and the DNA base pairs drive the complexes to adopt arrangements that are pre-organized for epoxide attack and DNA-adduct formation.

CONCLUSIONS

(+)-B*a*P DE-2 is a tumorigenic metabolite of benzo[*a*]pyrene, a polycyclic aromatic hydrocarbon found in soot and tobacco smoke. (+)-B*a*P DE-2 forms covalent DNA adducts by intercalation into DNA followed by the nucleophilic attack of the (+)-B*a*P DE-2 epoxide by the exocyclic NH_2 group of either adenine or guanine. These covalent adducts interfere with key DNA processes, leading to sequence-specific mutations.

Complexes of the AT and GC DNA base pairs with two low-lying conformers of (+)-B*a*P DE-2 have been examined using DFT methods to gain insight into the factors governing adduct formation. Formation of complexes with the GC base pair is favored over AT, due to stronger p-stacking interactions between the pyrene core of (+)-B*a*P DE-2 and guanine-cytosine. Only one GC···(+)-B*a*P DE-2 complex is predicted to lie within 3 kcal mol^{-1} of the global minimum energy GC···(+)-B*a*P DE-2 structure, which is bound in the gas phase by 27.3 kcal mol^{-1}. In this global minimum complex [**CG(I-EM)**], the exocyclic NH_2 group of cytosine is positioned for *cis* addition to the epoxide of (+)-B*a*P DE-2. The scarcity of experimentally observed cytosine adducts arising from this thermodynamically favorable complex are explained based on a simple electrostatic model; the region surrounding cytosine in the GC base pair has a positive electrostatic potential, which will destabilize the cationic intermediate leading to *cis* addition to

the epoxide. The second lowest lying GC complex is compatible with backside attack on the epoxide by guanine, leading to the observed *trans* ring-opened N^2-guanine adducts. We predict five thermodynamically accessible AT···(+)-B*a*P DE-2 complexes that are compatible with intact DNA. The most energetically favored complex [**AT(II-eM)**], lies 6.4 kcal mol^{-1} higher in energy than the global minimum GC···(+)-B*a*P DE-2 complex. Among these low-lying AT complexes, two structures are compatible with adenine adduct formation, accounting for both the *cis* and *trans* adducts.

There are two stable conformers of free (+)-B*a*P DE-2, and when bound to either the AT or GC base pair these are competitive energetically. Complexes involving the lower-lying conformer (**I**) lead to *cis* ring-opened adducts, while *trans* adducts are predicted to form from complexes featuring the higher-lying conformer (**II**).

Stable non-covalent complexes between DNA base pairs and (+)-B*a*P DE-2 arise from a combination of intermolecular hydrogen bonds and π-stacking interactions. Because of the *trans* arrangement of the C_7 OH group and epoxide on (+)-B*a*P DE-2, in many of the low-lying complexes the base pair is oriented with the exocyclic NH_2 group pre-organized for epoxide attack. The energetic tendency to maximize π-stacking interactions further orients the base pair for epoxide attack.

REFERENCES

(1) Edwards, N. T. *J. Environ. Qual.* **1983**, *12*, 427.

(2) Miller, J. A.; Pilling, M. J.; Troe, J. *Proc. Combust. Inst.* **2005**, *30*, 43.

(3) Wheeler, S. E.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **2004**, *121*, 8800.

(4) Wheeler, S. E.; Robertson, K. A.; Allen, W. D.; Schaefer, H. F.; Bomble, Y. J.; Stanton, J. F. *J. Phys. Chem. A* **2007**, *111*, 3819

(5) Durant, J. L.; Busby, W. F.; Lafleur, A. L.; Penman, B. W.; Crespi, C. L. *Mutat. Res.* **1996**, *371*, 123.

(6) Denissenko, M. F.; Pao, A.; Tang, M. S.; Pfeifer, G. P. *Science* **1996**, *274*, 430.

(7) Pullman, A.; Pullman, B. *Adv. Cancer Res.* **1955**, *3*, 117.

(8) Hecht, S. S. *J. Natl. Cancer I.* **1999**, *91*, 1194.

(9) Hoffmann, D.; Djordjevic, M. V.; Hoffmann, I. *Prev. Med.* **1997**, *26*, 427.

(10) Adamson, J. D.; Morter, C. L.; DeSain, J. D.; Glass, G. P.; Curl, R. F. *J. Phys. Chem.* **1996**, *100*, 2125.

(11) Buening, M. K.; Wislocki, P. G.; Levin, W.; Yagi, H.; Thakker, D. R.; Akagi, H.; Koreeda, M.; Jerina, D. M.; Conney, A. H. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 5358.

(12) Szeliga, J.; Dipple, A. *Chem. Res. Toxicol.* **1998**, *11*, 1.

(13) Sayer, J. M.; Chadha, A.; Agarwal, S. K.; Yeh, H. J. C.; Yagi, H.; Jerina, D. M. *J. Org. Chem.* **1991**, *56*, 20.

(14) Wei, S.-J. C.; Chang, R. L.; Wong, C. Q.; Bhachech, N.; Cui, X. X.; Hennig, E.; Yagi, H.; Sayer, J. M.; Jerina, D. M.; Preston, B. D.; Conney, A. H. *Proc. Natl. Acad. Sci.* **1991**, *88*, 11227.

(15) Wei, S.-J. C.; Chang, R. L.; Bhachech, N.; Cui, X. X.; Merkler, K. A.; Wong, C.-Q.; Hennig, E.; Yagi, H.; Jerina, D. M.; Conney, A. H. *Cancer Res.* **1993**, *53*, 3294.

(16) Wolfe, A. R.; Smith, T. J.; Meehan, T. *Chem. Res. Toxicology* **2004**, *17*, 476.

(17) Geacintov, N. E.; Cosman, M.; Hingerty, B. E.; Amin, S.; Broyde, S.; Patel, D. J. *Chem. Res. Toxicol.* **1997**, *10*, 111.

(18) Jerina, D. M.; Chadha, A.; Cheh, A. M.; Schurdak, M. E.; Wood, A. W.; Sayer, J. M. Biological Reactive Intermediates IV. In *Biological Reactive Intermediates IV*; Witmer, C. M., Snyder, R., Jollow, D. J., Kalf, G. F., Kocsis, J. J., Sipes, I. G., Eds.; Plenum Press New York, 1991; pp 533.

(19) Yong, Y.; Romano, L. J. *Chem. Res. Toxicol.* **1996**, *9*, 179.

(20) Pommier, Y.; Laco, G. S.; Kohlhagan, G.; Sayer, J. M.; Kroth, H.; Jerina, D. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10739.

(21) Hanrahan, C. J.; Bacolad, M. D.; Vyas, R. R.; Liu, T.; Geacintov, N. E.; Loechler, E. L.; Basu, A. K. *Chem. Res. Toxicol.* **1997**, *10*, 369.

(22) Alekseyev, Y. O.; Romano, L. J. *Biochemistry* **2002**, *41*, 4467.

(23) Shukla, R.; Liu, T.; Geacintov, N. E.; Loechler, E. L. *Biochemistry* **1997**, *36*, 10256.

(24) Shukla, R.; Geacintov, N. E.; Loechler, E. L. *Carcinogenesis* **1999**, *20*, 261.

(25) Page, J. E.; Zajc, B.; Oh-hara, T.; Lakshman, M. K.; Sayer, J. M.; Jerina, D. M.; Dipple, A. *Biochemistry* **1998**, *37*, 9127.

(26) Seo, K. Y.; Jelinsky, S. A.; Loechler, E. L. *Mutat. Res.* **2000**, *463*, 215.

(27) Ling, H.; Sayer, J. M.; Plosky, B. S.; Yagi, H.; Boudsocq, F.; Woodgate, R.; Jerina, D. M.; Yang, W.; Davies, D. R. *P. Natl. Acad. Sci. USA* **2004**, *101*, 2265.

(28) Chiapperino, D.; Cai, M.; Sayer, J. M.; Yagi, H.; Kroth, H.; Masutani, C.; Hanaoka, F.; Jerina, D. M.; Cheh, A. M. *J. Biol. Chem.* **2005**, *280*, 39684.

(29) Lee, H. M.; Chae, Y.-H.; Kwon, C.; Kim, S. K. *Biophys. Chem.* **2007**, *125*, 151.

(30) Chandani, S.; Lee, C. H.; Loechler, E. L. *Chem. Res. Toxicol.* **2005**, *18*, 1108.

(31) Yan, S. F.; Wu, M.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2004**, *43*, 7750.

(32) Lee, C. H.; Loechler, E. L. *Mutat. Res.* **2003**, *529*, 59.

(33) Rao, S. N.; Lybrand, T.; Michaud, D.; Jerina, D. M.; Kollman, P. A. *Carcinogenesis* **1989**, *10*, 27.

(34) Lee, C. H.; Chandani, S.; Loechler, E. L. *Chem. Res. Toxicol.* **2002**, *15*, 1429.

(35) Fountain, M. A.; Krugh, T. R. *Biochemistry* **1995**, *34*, 3152.

(36) Xu, R.; Mao, B.; Amin, S.; Geacintov, N. E. *Biochemistry* **1998**, *37*, 769.

(37) Karle, I. L.; Yagi, H.; Sayer, J. M.; Jerina, D. M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1433.

(38) Schwartz, J. L.; Rice, J. S.; Luxon, B. A.; Sayer, J. M.; Xie, G.; Yeh, H. J. C.; Liu, X.; Jerina, D. M.; Gorenstein, D. G. *Biochemistry* **1997**, *36*, 11069.

(39) Pradhan, P.; Tirumala, S.; Liu, X.; Sayer, J. M.; Jerina, D. M.; Yeh, H. J. C. *Biochemistry* **2001**, *40*, 5870.

(40) Lakshman, M. K.; Keeler, J. C.; Ngassa, F. N.; Hilmer, J. H.; Pradhan, P.; Zajc, B.; Thomasson, K. A. *J. Am. Chem. Soc.* **2007**, *129*, 68.

(41) Shahbaz, M.; Geacintov, N. E.; Harvey, R. G. *Biochem.* **1986**, *25*, 3290.

(42) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.

(43) Swart, M.; van der Wijst, T.; Guerra, C. F.; Bickelhaupt, F. M. *J. Mol. Model.* **2007**, *13*, 1245.

(44) Zhao, Y.; Truhlar, D. G. *Theo. Chem. Acc.* **2008**, *120*, 215.

(45) Grimme, S. *J. Comp. Chem.* **2006**, *27*, 1787.

(46) Li, S.; Cooper, V. R.; Thonhauser, T.; Lundqvist, B. I.; Langreth, D. C. *J. Phys. Chem. B* **2009**, *113*, 11166.

(47) Řeha, D.; Kabeláč, M.; Ryjáček, F.; Šponer, J.; Šponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.

(48) Langner, K. M.; Kedzierski, P.; Sokalski, W. A.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 9720.

(49) Chadha, A.; Sayer, J. M.; Yeh, H. J. C.; Yagi, H.; Cheh, A. M.; Pannell, L. K.; Jerina, D. M. *J. Am. Chem. Soc.* **1989**, *111*, 5456.

(50) Cheh, A. M.; Chadha, A.; Sayer, J. M.; Yeh, H. J. C.; Yagi, H.; Pannell, L. K.; Jerina, D. M. *J. Org. Chem.* **1993**, *58*, 4013.

(51) Chadha, A.; Sayer, J. M.; Agarwal, S. K.; CHeh, A. M.; Tagi, H.; Yeh, H. J. C.; Jerina, D. M. In *Polynuclear Aromatic Hydrocarbons: Measurements, Means and Metabolism (Eleventh International Symposium)*; Cook, M., Loening, K., Merritt, J., Eds.; Batelle Press: Columbus, OH, 1991.

(52) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(53) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

(54) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.

(55) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.

(56) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(57) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *J. Chem. Theory Comput.* **2008**, *4*, 1996.

(58) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146.

(59) Gu, J.; Wang, J.; Leszczynski, J.; Xie, Y.; Schaefer, H. F. *Chem. Phys. Lett.* **2008**, *459*, 164.

(60) E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, D. Wang, E. Aprà, T. L. Windus, J. Hammond, P. Nichols, S. Hirata, M. T. Hackler, Y. Zhao, P.-D. Fan, R. J. Harrison, M. Dupuis, D. M. A. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, Q. Wu, T. Van Voorhis, A. A. Auer, M. Nooijen, E. Brown, G. Cisneros, G. I. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. A. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, L. Pollack, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z. Zhang, "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1" (2007), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.

(61) Kendall, R. A.; Aprá, E.; Bernholdt, D. E.; Bylaska, E. J.; Dupuis, M.; Fann, G. I.; Harrison, R. J.; Ju, J.; Nichols, J. A.; Nieplocha, J.; Straatasma, T. P.; Windus, T. L.; Wong, A. T. *Computer Phys. Commun.* **2000**, *128*, 260.

(62) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*, 034111.

(63) Gräfenstein, J.; Cremer, D. *J. Chem. Phys.* **2007**, *127*, 164113.

(64) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *Submitted*.

(65) Miller, J. A.; Melius, C. F. *Combust. Flame* **1992**, *91*, 21.

(66) PBE-D predicts a much smaller separation of 1 kcal/mol.

(67) Politzer, P.; Murray, J. S. Computational Medical Chemistry for Drug Discovery. In *Computational Medical Chemistry for Drug Discovery*; Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J. P., Eds.; Marcel Dekker, Inc.: New York, 2004; pp 213.

(68) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *5*, 2301.

(69) Politzer, P.; Murray, J. S. In *Chemical Reactivity Theory: A Density Functional View*; Chattaraj, P. K., Ed.; CRC Press: Boca Raton, FL 2009.

(70) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10854.

(71) Legault, C. Y. CYLview, 1.0b Université de Sherbrooke, 2009.

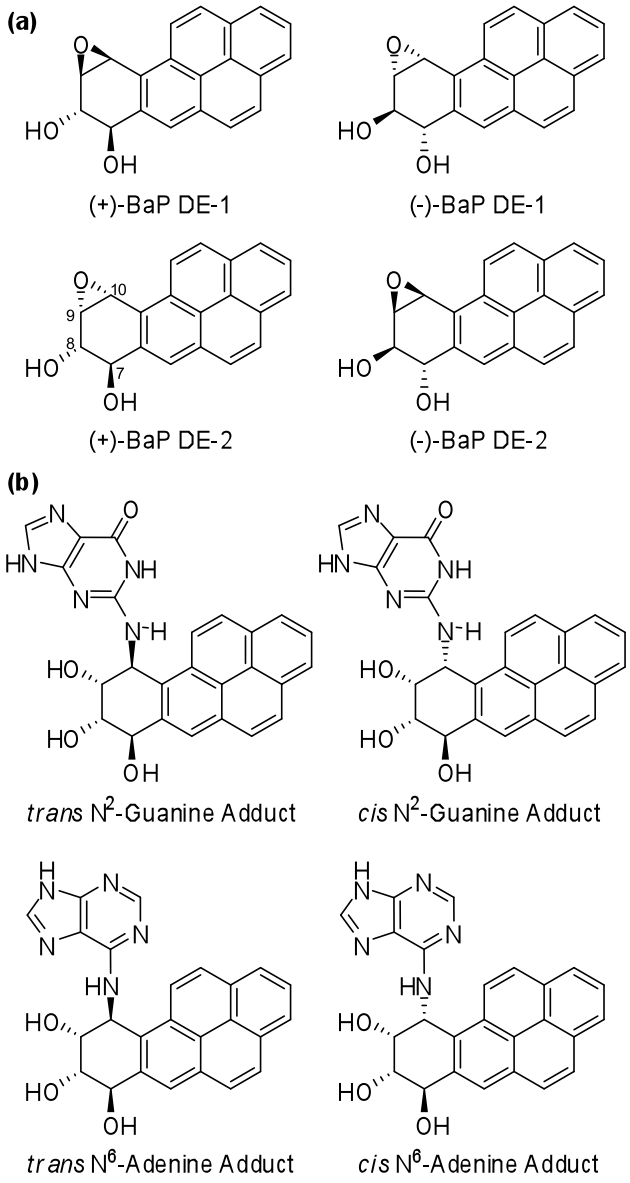


Figure 4.1. (a) Four stereoisomers of the benzo[*a*]pyrene diol epoxide; (b) major *trans* and *cis* guanine and adenine (+)-B*a*P DE-2 adducts

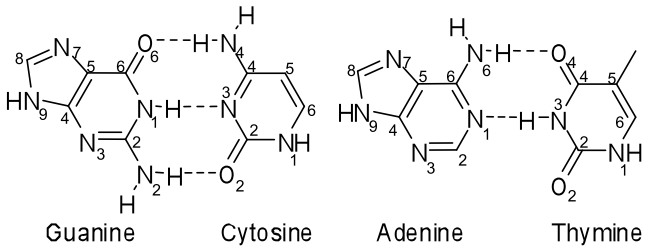


Figure 4.2. Canonical atomic labels for GC and AT base pairs.

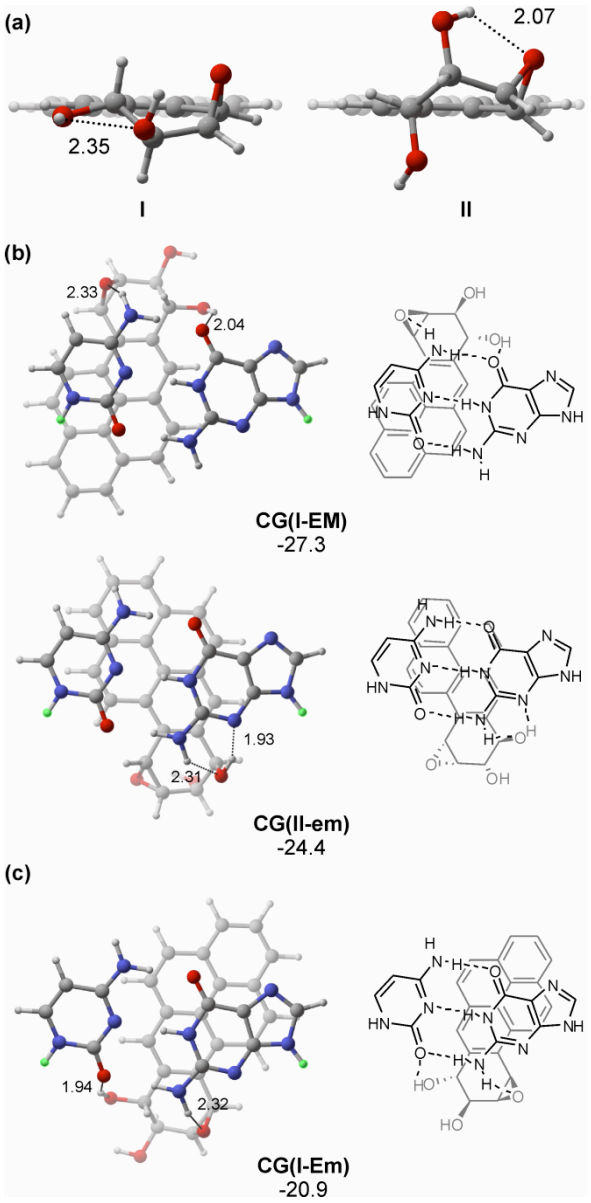


Figure 4.3. (a) Two low-lying gas-phase conformers of (+)-B*a*P DE-2. At the M06-2X/6-31+G(d)//M05-2X/6-31+G(d) level of theory, conformer **II** lies 2.1 kcal mol^{-1} higher than **I**, separated by a conformational barrier of 7.8 kcal mol^{-1}. (b) Low-lying complexes of (+)-B*a*P DE-2 with the GC base pair. (c) Lowest-lying GC complex compatible with *cis* guanine addition. In (b) and (c), gas-phase interaction energies relative to separated GC and conformer **I** of (+)-

B*a*P DE-2 are given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green. Hydrogen bond distances are in Angstroms.

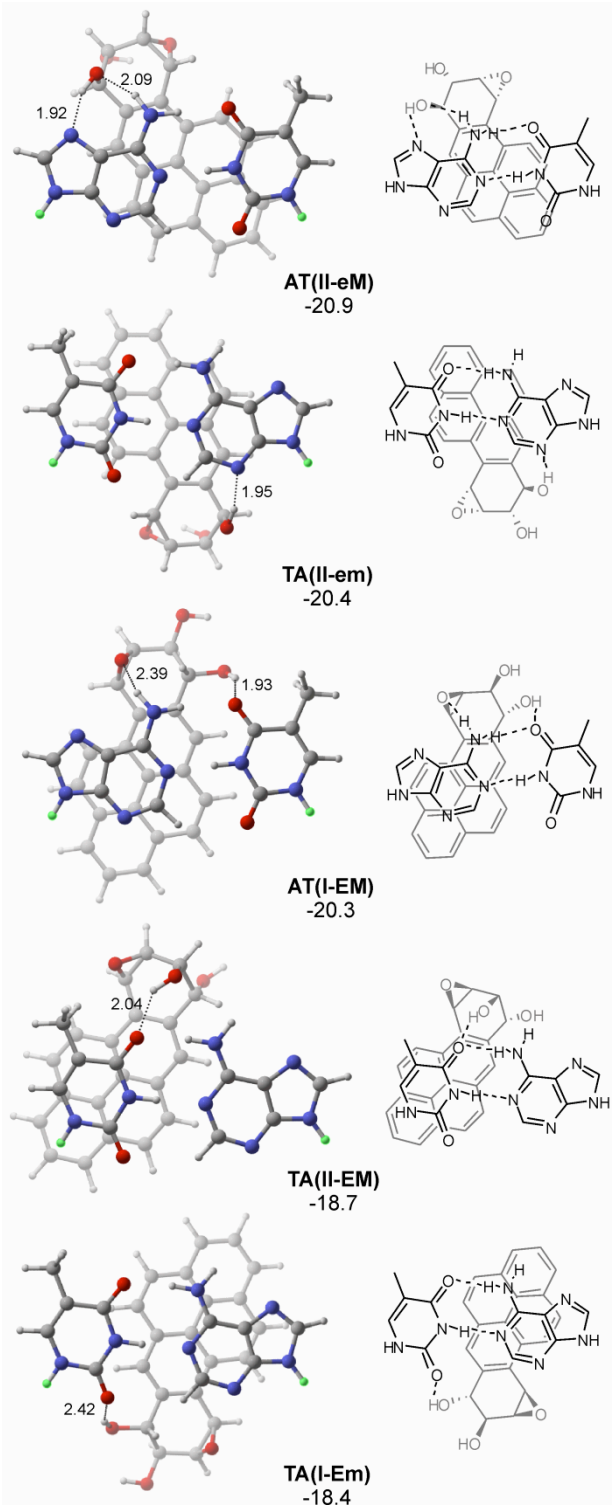


Figure 4.4. Low-lying complexes of (+)-B*a*P DE-2 with the AT base pair. Gas-phase interaction energies relative to separated AT and conformer **I** of (+)-B*a*P DE-2 are given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green. Hydrogen bond distances are in Angstroms.

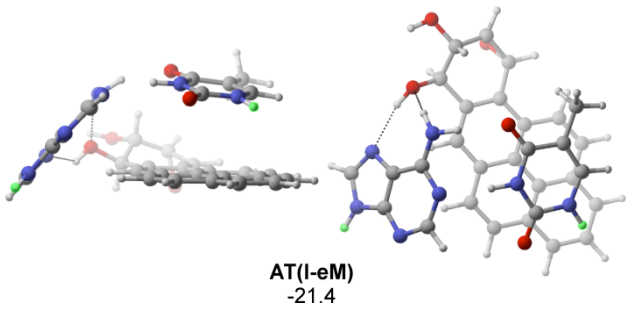


Figure 4.5. Two views of a low-lying complex of conformer **I** of (+)-B*a*P DE-2 with the AT base pair. This structure is not compatible with intact DNA, since the base-pair undergoes significant distortion in order to form a strong hydrogen bond with conformer **I** of (+)-B*a*P DE-2 while maintaining favorable stacking interactions with the pyrene. The gas-phase interaction energy relative to separated AT and (+)-B*a*P DE-2 is given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green.

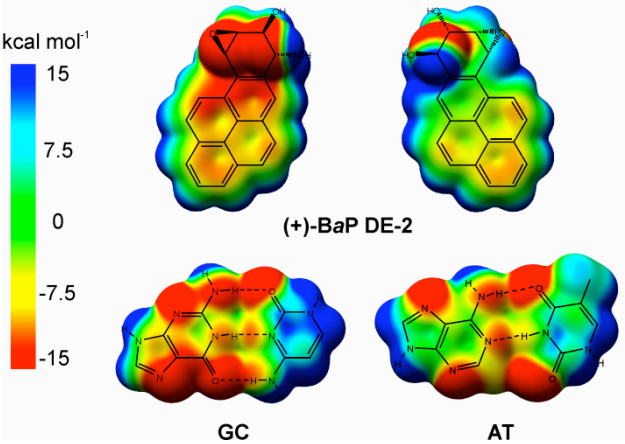


Figure 4.6. Electrostatic potentials of conformer **II** of (+)-B*a*P DE-2, guanine-cytosine, and adenine-thymine, mapped onto electron density isosurfaces ($r = 0.001$ e/au^3).

CHAPTER 5

CONCLUSION

Understanding the nature of non-covalent interactions in biological systems is a formidable task. As shown, some biological systems are too large to gain insight by using highly accurate *ab initio* methods. DFT can be used to make predictions about these systems, however the results can compromise accuracy. Utilizing the focal point method in conjunction with DFT allows for valuable information to be obtained concerning non-covalent interactions.

Creating a model system with the same chemical properties as the problem of interest can be a beneficial tool when benchmarking. The benzene dimer is a common model system when studying π-π interactions. Utilizing benchmarks performed on the benzene dimer made it possible to determine a viable density functional to use for systems that possess π- π interactions. Similarly, malonaldehyde was used as an archetype for the shortest intramolecular hydrogen bond. Using malonaldehyde and seven simple derivatives constituted an in-depth investigation searching for the shortest intramolecular hydrogen bond.

Not only are highly accurate *ab initio* methods employed for benchmarking systems of similar chemical properties, it has been shown they can be a barometer for accuracy related to mechanistic information. In the investigation of the formamide dimer, the focal point analysis was performed to gain insight about the accuracy of the intrinsic reaction coordinate. By comparing energy points along the intrinsic reaction coordinate computed using B3LYP to energy points computed using the focal point analysis, a discrepancy is seen in the product region. This difference is troubling due to the accustomed expectations of the B3LYP functional

that routinely underestimate barrier heights. Here the error is seen in the product region, the examination of the B3LYP functional shows that the error observed in this reaction is due to the B3 exchange.

The advantages of combining these methods were evident in previous chapters. The investigation of the intramolecular hydrogen bond in malonaldehyde and its derivatives was possible due to the combination of *ab initio* and DFT. Geometries were optimized with a DZP++ basis set at the B3LYP level of theory for ~30 structures ranging from nine to 30 atoms. Additionally, transition state optimizations and vibrational frequencies were computed for these structures. The vast amount of computational work is evident. The computational expense is too significant to use *ab initio* methods alone, particularly the expense incurred with analytic first and second derivatives necessary to compute geometry optimizations and frequencies, respectively.

To understand the sacrifice in accuracy made by these approximations, the geometry of malonaldehyde was optimized using cc-pVQZ//CCSD(T) for the ground state and the transition state. The results of this were clear, for systems with similar chemical properties as malonaldehyde the DZP++//B3LYP geometries were suitable. From the reliable B3LYP geometries, it was easy to determine the trends seen when investigating varying substituents placed on malonaldehyde. When placing either bulky substituents or substituents with electron donating properties on the symmetrical carbons, the intramolecular hydrogen bond of the system shortened. The best results were seen when there was a bulky, electron donating substituent on the symmetrical carbons with an electron withdrawing group on the unique carbon. Notably, the extensive study on these geometries would not have been possible without utilizing DFT and the confidence would not have been as prodigious without thorough benchmarking.

Additionally, it was necessary to benchmark the energy barrier of the intramolecular proton transfer in malonaldehyde. Energy barriers computed using B3LYP are known to be too low. When considering barriers that are ≤ 4 kcal mol^{-1}, any error can be problematic when examining trends. To aid this, the highly accurate focal point method was performed on eight of the smallest systems. From these results a correlation was seen between the focal point energies and the B3LYP energies. An empirical method was determined that allowed the energy barrier to be predicted more accurately. The only information necessary for the empirical prediction is the energy barrier computed using B3LYP. Hence, an energy barrier with focal point accuracy can be obtained for systems that are too large to compute the focal point extrapolation on. Using several levels of theory was necessary to optimize the results and computational cost for this investigation.

The formamide dimer was benchmarked in a similar way to malonaldehyde. The geometries were computed using B3LYP and MP2 with a large basis set to confirm that B3LYP was a viable functional for these intermolecular hydrogen bonded systems. The IRC was computed using B3LYP. At each step in the IRC, the geometry was optimized at that point of the reaction coordinate. At several important points along the reaction coordinate, a focal point analysis was computed at the B3LYP geometry to show the accuracy of the B3LYP energy. As previously discussed for malonaldehyde derivatives, B3LYP underestimates barrier heights. Therefore, one would expect minor systematic errors when compared with the focal point energies. The largest deviation between the two methods is seen in the product region of the reaction coordinate. This discrepancy is due to the inaccuracies in the B3 exchange of the B3LYP functional. The imprecision of B3 exchange could be due to a delocalization error in the system.

Due to the limitations and computational costs that incur from computing an IRC via *ab initio* methods, the qualitative mechanistic investigation for the formamide dimer would be difficult. Fascinating mechanistic information about the double proton transfer of the formamide dimer was obtained. The reaction force analysis and the reaction electronic flux are essential to the exploration. The computation of the IRC is an essential stepping-stone to determining these properties. The results of this work determined that the double proton transfer of the formamide dimer begins with structural rearrangement. The structural rearrangement occurs as each monomer moves closer together. Before the transition state is reached, there is considerable electronic rearrangement that contributes to the activation process.

Most of the discussion regarding the strengths of using DFT and *ab initio* methods has centered on hydrogen bonding. However, DFT has been shown to be an effective method of studying π-π interactions in addition to hydrogen bonding. The incongruence from the previous inquisitions shows the B3LYP functional is no longer the DFT functional of choice. Benchmarking π-π interactions has occurred showing shortcomings by traditional DFT functionals. However, M05-2X and M06-2X have proven to work well for π stacking systems. When combining these methods with a reasonably large computational grid, excellent results are obtained for these π stacking systems.

Benzo[*a*]pyrene stacks with DNA base pairs creating a large complex that would be nearly impossible to study using *ab initio* methods. The fascinating complex forms an adduct rapidly when studied experimentally, therefore there are additional advantages to studying this system computationally. From this work it was determined, that guanine-cytosine (GC) complexes are the lowest lying complexes on the global energy surface. However, experiment shows adenine-thymine adducts are more abundant. An electrostatic model gleaned the

limitations of GC complexes to form an adduct as readily as AT complexes. Several low-lying complexes were ruled out due to their incompatibility with DNA. Incompatibilities included bent complexes and complexes that interfere with the DNA backbone.

In this exploration, five thermodynamically accessible structures were found. Two of these are positioned in such a way the adduct can form. One of these structures forms the *cis* adduct and the other forms the *trans* adduct. The overall trend of stability for the complexes results from non-covalent interactions. Hydrogen bonding between the benzo[*a*]pyrene diol epoxide and the DNA bases stabilize the overall complex. Additionally, the amount of π-π interactions between the two aromatic complexes contributes to the stability.

The study of non-covalent interactions on biological systems is an enormous domain due to the panoply of applications. Benchmarking non-covalent interactions in chemical systems is advantageous due to the procured benefits. It allows confidence to be accrued when examining the results and facilitates savings on computational cost. Furthermore, it allows intriguing systems to be investigated that would, otherwise, be insurmountable. Utilizing an armamentarium of computational tools allows for research to be optimized when considering cost, accuracy, and relevant problems.