Theoretical Studies of Intra- and Intermolecular Interactions

by
Nadezhda (Nadya) Kobko

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirement for the degree of Doctor of Philosophy, the Graduate School and University Center of the City University of New York

2005

This Manuscript has been read and accepted for the Graduate faculty in Chemistry in satisfaction of the dissertation requirements for the Degree of Doctor of Philosophy

22 April 2005
Date

Prof. Joseph J. Dannenberg
Chair of Examining Committee

April 25, 2005
Date

Gerald Koeppl (ada)
Prof. Gerald W. Koeppl
Executive Officer

Prof. Charles M. Drain

Prof. Richard W. Franck

Prof. Alexander Greer

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Acknowledgments

To my mentor Prof. Joseph J. Dannenber - for his kind help, guidance, support, and patience during my doctoral thesis studies.

To all the great teachers who I met during my studies at the City University of New York, in particular, Prof. Richard D. Pizer, Prof. Jerome M. Schulman, Prof. Arthur D. Baker, Prof. Richard W. Franck, Prof. Pamela A. Mills, Prof. Gerald W. Koeppl, Prof. Max Diem, Prof. John R. Lombardi, Prof. Garry Quigley, Prof. Richard W. Frank, Prof. Malgorzata Ciszkowska - for broadening my knowledge and understanding of modern chemistry.

To my Executive Committee members, Prof. Jerome M. Schulman, Prof. Charles M. Drain, Prof. Richard W. Franck, and Prof. Alexander Greer, - for their expertise and many helpful suggestions.

To Prof. Klaus Grohmann, the Graduate Student Advisor at the Chemistry Department of Hunter College, - for all his aid and support during all these years.

To my friend Dr. Artëm Masunov - for introducing me to the field of Quantum Chemistry and for all his assistance, encouragement, and advises.

To Waldemar Cieniewizc - for keeping all the computers in our lab running smoothly.

To Prof. Miguel Duran and Dr. Pedro Salvador from University of Gerona, Spain - for designing the computer code used for the CP-calculations employed in Chapter 6.

To Prof. Amparo Asensio from University of Valencia, Spain - for helping to complete the Chapter 5 research and for being an incredible New York exploration companion.

To all my friends, in the United States and in Russia, and my family - for their faith in me and their constant support.

Dedicated to
my father,
my brother Geshka,
my grandmother Valia.
You are always on my mind...

Table of Contents

List of Tables

Chapter 2

Chapter 4

Chapter 5

Chapter 6

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix 1

Abbreviations

AM1	Austin Model 1, semi-empirical method developed by Dewar's group
B3LYP	hybrid functional; combines Becke's 3-parameter functional with the non-local correlation, provided by the functional of Lee, Yang and Parr
B3PW91	hybrid functional; combines Becke's 3-paramter functional with the non-local correlation provided by the Perdew-Wang expression
BSSE	basic set superposition error
CP	counter-poise correction for basic set superposition error (BSSE)
CP-opt	optimized on the potential energy surface corrected for BSSE by the means of counterpoise correction
CPU	central processor unit
DA	Diels-Alder (reaction)
DFT	density functional theory
FA	formamide molecule
Gly	glycine
HF	Hartree-Fock
IR	infrared
LUMO	lowest unoccupied molecular orbital
MC	Monte Carlo method
MD	molecular dynamics
MM	molecular mechanics
MO	molecular orbital

MP2	the second order Moller-Plesset perturbation theory
NMA	N-methylacetamide
PBE	exchange functional by Perdew, Burke and Ernserhof
PES	potential energy surface
QCI	quadratic configuration interaction
RAHB	resonance assisted hydrogen bond
TS	transition state
*	asterisk in the name of the basis set indicates that this basis set includes polarization functions
+	plus in the name of the basis set indicate that the basis set include very diffuse functions
//	double-slash shows that the energy was computed at the theory level that is different from the level at which the geometry was computed (e.g. B3LYP/D95**//AM1 means the B3LYP/D95** energy calculations of the structure computed at the AM1 theory level)

1.

INTRODUCTION

Understanding the different aspects of inter- and intra-molecular interactions is vital to the studies of chemical reactions and structures. In the past half century, together with the diverse experimental methods, molecular orbital (MO) calculations have become an essential tool in investigations of all kinds of chemical systems ranging from the simplest gas phase reactions to the processes of biological importance. Yet MO methods, required to accurately describe large biological systems that include thousands of atoms (like proteins and DNAs), are still too computationally expensive. One of the important examples of processes involving such a system is protein folding.[1] There now exist two major, rapidly developing, and relatively low in computational costs approaches that are used to study such processes: Molecular Dynamics (MD) and Monte Carlo (MC)

[1] *Protein Structure: Determination, Analysis, and Applications for Drug Discovery*; Chasman, D. E., Ed.; Marcel Dekker, Inc.: New York and Basel, 2003.

simulations[2,3] - both employing Molecular Mechanics (MM) methods, also known as Force Fields (for example, AMBER[4] and CHARMM[5]). Said force fields are based on empirical sets of covalent and van der Waals parameters, fixed atomic charges (electrostatics), and including only pairwise interactions in their analytical form. Although these force fields have been shown to give reasonable results in some cases,[6] the importance of many-body effects has been demonstrated for various systems – specifically, those that include hydrogen bonding interactions.[7,8,9] These many-body effects manifest themselves in cooperativity of H-bonds, making pairwise additive approximation a poor model. Taking the cooperative terms into account could significantly improve the description of biological systems and may radically alter the field of computational biophysics.[3,10,11] (The MO methods account for many-body interactions, but, as we pointed out before, their high computational costs make them inefficient for the all-atoms treatments of large systems.)

---

[2] Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868.

[3] Friesner, R. A.; Gunn, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 315.

[4] Weiner, S.; Kollman, P.; Nguyen, D.; Case, D. *J. Comput. Chem.* **1986**, *7*, 230; McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743.

[5] Brooks, C. R.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187; Mackerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

[6] Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins* **2000**, *39*, 252; Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474; Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849.

[7] Dykstra, C. E. *Theochem* **1996**, *362*, 1.

[8] Van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M., Jr. *J. Phys. Chem. B* **2000**, *104*, 9554.

[9] Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.

[10] Guo, H.; Gresh, N. Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.

[11] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

When planning my research the H-bonding cooperativity (pairwise nonadditivity of the interaction energies) became widely recognized.[12] The substantial cooperative effects for hydrogen bonding chains within molecular crystals were already reported in the other studies by our group. The crystals considered included those of acetic acid,[13] urea,[14] nitroanilines,[15] and the enol of 1,3-cyclohexanedione.[16,17] The latter studies indicate that cooperativity brings a significant contribution to the stability of the H-bonded complexes. Part I of this thesis explores the effects of H-bonding cooperativity: in the systems that model the secondary structures of proteins (Chapters 2-4), and in the base pairs of normal DNA (Chapter 5) using *ab initio* and Density Functional Theory (DFT) calculations.

In Chapter 2, Part I, the study of cooperativity in the H-bonded chains of formamide molecules is presented. These relatively simple systems model the hydrogen bonding in α-helices and β-strands of proteins but yet are small enough so that the long H-bonded chains can be studied (see Figure 1.1). (Truncation of the systems from α-helices and β-strands to the H-bonded chains of formamides seems to be acceptable for our purposes. For example: Kang reports dimerization energies and H-bonding distances at various levels of theory for formamide, acetamide and N-methylacetamide (NMA) dimers that are very similar to each other.[18]) I present density functional theory (DFT)

---

[12] Scheiner, S. *Hydrogen Bonding*; Oxford University Press: New York, 1997.

[13] Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714.

[14] Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806.

[15] Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1996**, *100*, 9638.

[16] Turi, L.; Dannenberg, J. J. *Chem. Mater* **1994**, *6*, 1313.

[17] Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1992**, *96*, 5819.

[18] Kang, Y. K. J. *Phys. Chem. B* **2000**, *104*, 8321; for NMA see also Aleman, C. *J. Phys. Chem. A* **2001**, *105*, 6717.

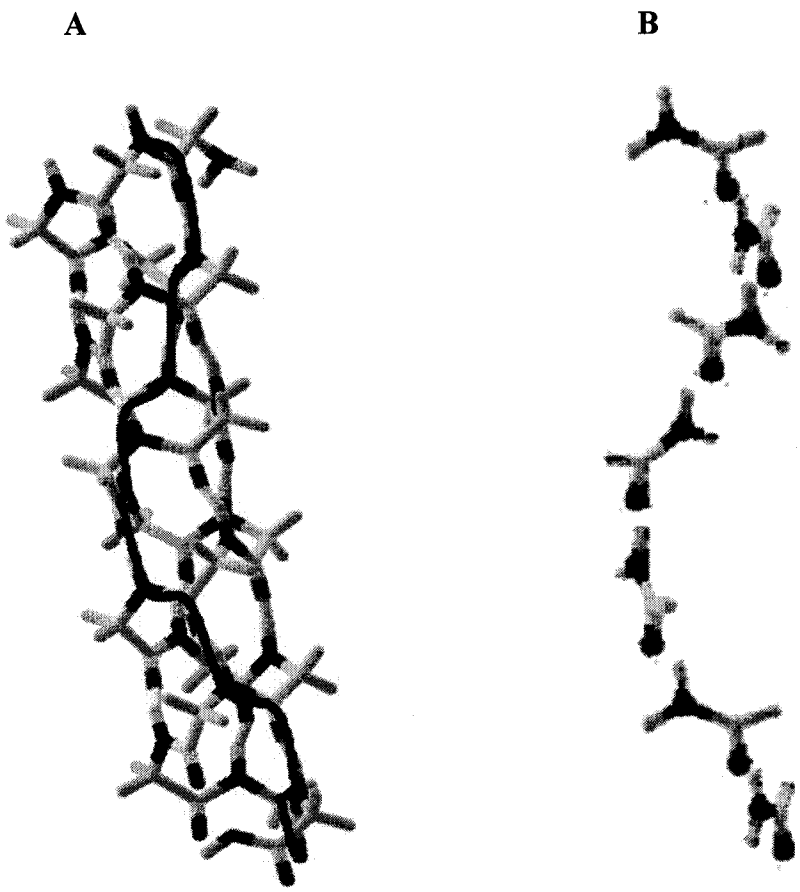


Figure 1.1. In figure A the dark line represents one of the three hydrogen-bonding chains of the α-helix. Figure B shows a chain of H-bonded formamide molecules that models the geometry of this chain.

calculations of hydrogen bonds in chains of up to 15 formamide molecules. I shall also introduce an empirical formula that relates the strength of the hydrogen bond with the length of the H-bonded chain and its position in a chain. I believe that this equation may help to improve the description of the hydrogen bonding for modeling the protein folding.

In Chapter 3, Part I, I discuss the relative stabilities at the B3LYP/D95**//AM1level of theory for the four major conformers of polyglycine: (1) C5 or, using another notation, 1_5 fully extended ribbon (β-strand), (2) C7 or 2_7-ribbon (repeated γ-turns), (3) C10 or 3_{10}-helix, (4) C13 or 4_{13}-helix, also known as α-helix.[19] The relative stability of the conformations plays an essential role in such processes of biological importance such as protein folding and transitions between active and inactive conformations of enzymes. While discussing the effects of H-bonding interactions, I will touch upon some other factors (like dipole-dipole interactions) on the relative stabilities of the polyglycine conformers.

While I was working on the previous two projects, I had to examine the normal modes of the systems – the H-bonded chains of formamides and the conformers of polyglycine. I then noticed that the cooperativity of the H-bonding interactions have a particular signature in calculated infrared (IR) spectra. Vibrational spectra have been used to determine the secondary structures in proteins. Both the frequency shift and the intensity of the absorption are sensitive to the coupling of the vibrational modes. The latter can result from the interactions of the modes through covalent bonds, through space (via classical dipole-dipole interactions), and through hydrogen bonds or some

---

[19] Creighton, T. E. *Proteins: Structure and Molecular Properties*; W. H. Freeman and Co.: New York, 1993.

combinations of these. Chapter 4, Part I, presents the analysis of the effects of hydrogen bonding cooperativity on vibrational spectra of proteins. In the study we used DFT molecular orbital (MO) calculations to compare the coupling of several vibrations through the covalent bonds and through the hydrogen bonds. I compare the coupling of vibrational modes in fully extended single β-strands of polyglycines of varying sizes, with their coupling in H-bonded chains of formamide molecules containing the equivalent number of the interacting groups. The single β-strands form only weak, cyclic H-bonds which are insulated from each other through CH_2 groups and exhibit no cooperativity. Thus all coupling of vibrational modes in these systems should come through the covalent bonds or through space. The H-bonded formamide chains, on the other hand, have no covalent linkages between the formamide molecules. Therefore all coupling of vibrational modes there should come through hydrogen bonding or through space. I chose to study the three most intense characteristic bands of proteins: (1) the C=O stretching vibrations (amide I infrared absorption band), (2) the combined N-H binding and C-N stretching vibrations (amide II band) and (3) the N-H stretching vibrations (amide II band). Also discussed are some of the effects of isotopic substitutions on vibrational spectra of proteins.

In Chapter 5, the last chapter of Part I, I present the evaluation of the cooperative contribution to the H-bonding interactions in the base pairs of normal DNA: adenine-thymine (A-T) and guanine-cytosine (G-C). The energies of the individual bonds in each base pair were ascertained by using model structures. One H-bond at a time was kept intact by rotating one base, with respect to the other, around the axis of each H-bond to form structures with the bases perpendicular to each other. The energies of the individual

H-bonds calculated in this way are compared with those of the planar base pairs. Optimized geometries were obtained using *ab initio* molecular orbital theory with electron correlation (MP2/D95**), and with density functional theory (B3LYP/D95**). To the extent with which the findings in this chapter can be generalized, they may be important in the design of materials that employ H-bonding motifs in self-assembly.

Part II of this thesis (Chapter 6) deals with a methodological subject. Of all the projects discussed herein, this is the earliest. It addresses the problem of the basis set superposition error (BSSE), which arises from the incompleteness of basis sets used in quantum mechanical calculations. The counterpoise (CP) correction for BSSE that was proposed by Boys and Bernardy in 1970[20] is considered to be one of the most popular tools to account for this error.[21] When I joined Prof. Dannenberg's group, he, and his colleagues in Spain, were working on designing the script for the optimization on the CP-corrected potential energy surfaces (PESs).[22] Using the latter script we decided to apply CP-optimization to the calculations of the transition states (TS) of the several organic reactions (concerted Diels-Alder, 1,2-H-atom shift in ethyl radical, and H-atom transfers from methane and propene to methyl radical). Used were a variety of methods (HF, MP2, DFT) and basis sets to optimize these TSs on the potential energy surfaces that included CP-correction. In doing this I thus corrected not only enthalpies of the TSs (done earlier) but, for the first time, their geometries as well. Later, I used the script for CP-optimization in my other projects (Part I). The script for the optimization on the potential

---

[20] Boys, S. F.; Bernardi, F *Molecular Physics*, **1970**, *19*, 553.

[21] van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

[22] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

energy surface that includes CP-correction for the BSSE is now implemented in the Gaussian 03.[23]

Lastly, I would like to briefly discuss the choice of the computational methods used in the studies of hydrogen bonding. In most cases the hybrid density functional theory (DFT) level was used; it was shown that accurate description of hydrogen bonding is achieved at MP2 theory level with very large basis sets, while HF description is inadequate.[24] I chose the hybrid DFT method B3LYP because it is less basis-set-dependent and reproduces the MP2 results well, while giving reliable geometries and energies in other studies of the H-bonded complexes.[25] B3LYP combines Becke's 3-parameter functional[26] with the non-local correlation provided by correlation functional of Lee, Yang and Parr.[27]

For all calculations I employed the suite of computational programs developed by the Gaussian Inc. In most cases the revisions A10 and A11 of Gaussian 98[28] were used.

---

[23] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewsk, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; J. Tomasi; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03*, Gaussian, Inc.: Pittsburgh PA, 2003.

[24] Tsuzuky, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906.

[25] (a) Lozynski, M; Rusinska-Roszak, D. *J. Phys. Chem.* **1998**, *102*, 2899. (b) Tsuzuly, S.; Luthai, P. H. *J. Chem. Phys.* **2001**, *114*, 11906. (c) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705.

[26] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

[27] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

[28] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewsk, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; J. Tomasi; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.;

For the treatment of large systems I employed the cluster comprised of Pentium 3, Pentium 4 and AMD Athlon computers parallelized using LINDA[29] at Hunter College of the City University of New York.

Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople J. A. *Gaussian 98*, Gaussian, Inc.: Pittsburgh PA, 1998-2001.

[29] LINDA. Scientific Computing Associates: New Haven.

Part I

HYDROGEN BONDING COOPERATIVITY

2.
*COOPERATIVITY IN AMIDE HYDROGEN BONDING CHAINS. THE RELATIONSHIP BETWEEN ENERGY OF H-BOND, ITS POSITION AND H-BONDING CHAIN LENGTH IN PEPTIDES AND PROTEIN FOLDING**

2.1. Introduction

Protein folding is still relatively poorly understood process. Many reviews[1] including an entire issue of Accounts of Chemical Research[2] have appeared. As it was pointed out in Chapter 1, the dynamics of the folding process is typically studied using molecular dynamics or Monte Carlo techniques based on sets of the fixed potentials that are parameterized from experimental data and quantum mechanical calculations. While

---

* Results discussed in this Chapter are published in Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348 and Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 10389. This work was supported (in part) by grants from the National Institutes of Health (S06GM60654) and PSC-CUNY.

[1] (a) Dill, K. A. *Biochem.* **1990**, *29*, 7133. (b) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868. (c) Schuster, P.; Wolschann, P. *Monatsh. Chem.* **1999**, *130*, 947. (d) Kaya, H.; Chan, H. S. *Phys. Rev. Lett.* **2000**, *85*, 4823. (e) Skolnick, J.; Kolinski, A. *Comput. in Sci. Eng.* **2001**, *3*, 40. (f) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499. (g) Mirny, L.; Shakhnovich, E. *Annu. Rev. of Biophys. Biomol. Struct.* **2001**, *30*, 361. (h) Chasse, G. A.; Rodriguez, A. M.; Mak, M. L.; Deretey, E.; Perczel, A.; Sosa, C. P.; Enriz, R. D.; Csizmadia, I. G. *Theochem* **2001**, *537*, 319. (i) Arteca, G. A.; Reimann, C. T.; Tapia, O. *Mass Spectrom. Rev.* **2002**, *20*, 402.

[2] *Acc. Chem. Res.* **1998**, *31*, entire issue.

success in using effective two-body potentials to simulate a many-body problem has been reported,[3] use of pairwise contact potentials has also been criticized as inadequate.[4,5] Refining those potentials and accounting for many-body effects, such as cooperativity, are the two major ways of improving force fields.

Taking into account many-body effects is especially important for the description of protein folding since hydrogen bonding in peptides is shown to be highly cooperative in many cases. For example, Kemp suggested that H-bond cooperativity contributes strongly to the formation of α-helices,[6] while Wu has reported hydrogen bonding cooperativity in both α and 3_{10}-helices[7] and in sheets of β-polypeptides (constructed of peptides of β-amino acids)[8] but little or none in β-sheets of α-polyglycines.[9] In recent studies by our group some inter-strand H-bonding cooperativity was observed in sheets of γ-polypeptides but again none in β-sheets of α-polyglycines.[10] On the other hand, recent NMR evidence suggests that the hydrogen bonds towards the interior of an α-helix are shorter than those closer to the ends.[11] It was also noted in the recent theoretical study of formation of various α-helices by Wieczorek and Dannenberg that H-bonding

---

[3] van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M. J. *J. Phys. Chem. B* **2000**, *104*, 9554.

[4] King, B. F.; Weinhold, F. *J. Chem. Phys.* **1995**, *103*, 333.

[5] Vendruscolo, M.; Domany, E. *J. Chem. Phys.* **1998**, *109*, 11101.

[6] Kennedy, R. J.; Tsang, K.-Y.; Kemp, D. S. . *J. Am. Chem. Soc.* **2002**, *124*, 934.

[7] Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, 123, 5313.

[8] Lin, J.-Q.; Juo, S.-W.; Wu, Lin, J.-Q.; Juo, S.-W.; Wu, Y.-D. *J. Comput. Chem.* **2002**, *23*,1551.

[9] Zhao, Y.-L.; Wu, Y.-D. . *J. Am. Chem. Soc.* **2002**, *124*, 1570.

[10] Viswanathan, R.; Asensio, A.; Dannnenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205.

[11] Jaravine, V. A.; Alexandrescu, A. T.; Grzesiek, S. *Protein Sci.* **2001**, *10*, 943.

cooperativity is crucial for α-helix to become more stable than a corresponding β-strand.[12]

I should also point out that substantial cooperative effects for hydrogen bonding chains within molecular crystals have been previously reported in the other studies by the group of Prof. Dannenberg (see Ref. 13-17, Chapter 1). There have been also several reports of theoretical studies on cooperative H-bonding in amides involving small numbers of molecules[13] or using periodic calculations for infinite chains.[14]

In this chapter, I will present the results on *ab initio* (HF) and density functional theory (DFT) calculations of hydrogen bonding in chains of formamide molecules (Figure 2.1; also see Chapter 1 on the choice of this system). First, I studied the chains of up to 10 formamides at the HF and DFT levels using the constraint that while the intermolecular parameters were allowed to vary all the formamide molecules were kept geometrically equivalent and coplanar with each other. I then continued with the study of hydrogen bonding in completely optimized planar chains of up to 15 formamides at the DFT level. I originally sought to study planar chains as well as chains that retained the average dihedral angles that the formamide entities would have in an α-helix (Figure

---

[12] Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.

[13] (a) Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1748.(b) Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451. (c) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273. (d) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104. (e) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746. (f) Ludwig, R. *J. Mol. Liq.* **2000**, *84*, 65.(g) Cabaleiro-Lago, E. M.; Otero, J. R. *J. Chem. Phys.* **2002**, *117*, 1621.

[14] Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3950.

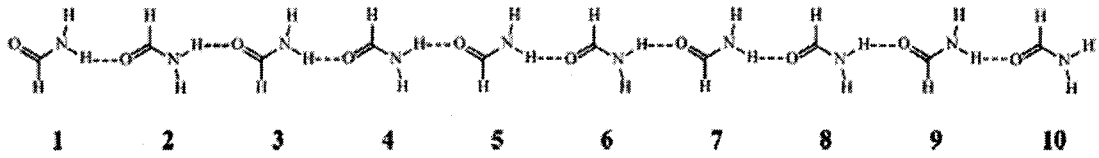


Figure 2.1. Chains of formamides indicating the numbering convention used in the text. Note that the translational repeating unit for this system contains two formamide molecules.

1.1B, Chapter 1). The data on the chains, that were constrained to have the dihedrals appropriate for an α-helix, did not differ substantially from those for the planar chains (Figure 2.2). For this reason, and because the vibrational analysis could not be performed on these helical systems (since they are not minima on the potential energy surface), I did not pursue the study of the helically arranged formamides any further. I shall also present an empirical formula that relates the energy of the individual hydrogen bond to the length of the H-bonding chain and the position of the hydrogen bond within the chain.

2.2. Computational Details

Molecular orbital calculations were performed at the HF/D95** and B3LYP/D95** level using the Gaussian 98 suite (Ref. 28, Chapter 1). At first, I used the constraint that all the formamides in each chain be identical in geometry. Results of these calculations are discussed in Section 2.3.1. Section 2.3.2 presents the results for the completely optimized formamide chains. In both cases all the structures were restrained to be of C_S symmetry (all atoms are coplanar). I used our cluster of Intel and AMD powered computers that are parallelized using LINDA (Ref. 29, Chapter 1) for these calculations. The number of nodes used for each calculation varied with the sizes of the systems studied. Vibrational frequencies were calculated for the fully optimized planar structures, using the normal harmonic approximations employed in the Gaussian 98 program, to verify the stationary points and calculate the enthalpies of the various species. All frequencies were real except for some very low frequency imaginary

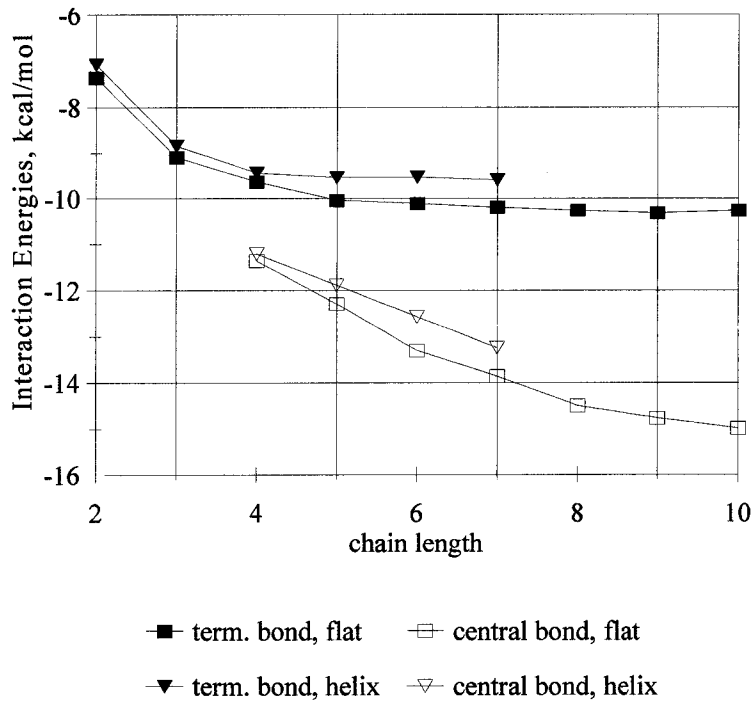


Figure 2.2. Comparison of the H-bonding Energies for the planar and helical chains of formamide molecules at the B3LYP/D95** level.

vibrations that involved out of plane twists between pairs of formamides in chains of five formamides and longer (less than 25 cm^{-1}). I do not believe that the small imaginary frequencies compromise the results that are presented below.

In section 2.3.1 I used the values of zero point energy correction and CP-correction for the BSSE (basis set superposition error)[15] for the CP-optimized dimer as corrections for the larger chains. I used the assumption that both of these corrections could be treated as additive. In this manner, nine times the vibrational and CP corrections for the dimer (one H-bond) are used to correct the enthalpy of the decamer (nine H-bonds). The validity of this assumption was tested at the B3LYP/D95** level on fully optimized small aggregates containing from two to seven formamides. The enthalpy correction per H-bond remained within 0.02 kcal/mol and the CP-correction per H-bond within 0.1 kcal/mol of their respective average values (see Figure 2.3 and Table 2.1). The same assumption was used before by Masunov in calculations of the interaction energies for the H-bonded aggregates of urea molecules (Ref. 14, Chapter 1).

The enthalpies of H-bonds for the completely optimized formamide chains (Section 2.3.2) include correction for basis set superposition error (BSSE) and the appropriate vibrational corrections to for 298 K. In order to obtain the BSSE correction per H-bond for my system at the B3LYP/D95** I used results from the complete CP-optimization of chains containing up to seven formamides. I then took the average of per H-bond corrections (see Table 2.1) and used it as the correction in calculations of the interaction hydrogen bonding energies for the longer chains. The vibrational corrections

---

[15] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024. See also Chapter 6 on BSSE.

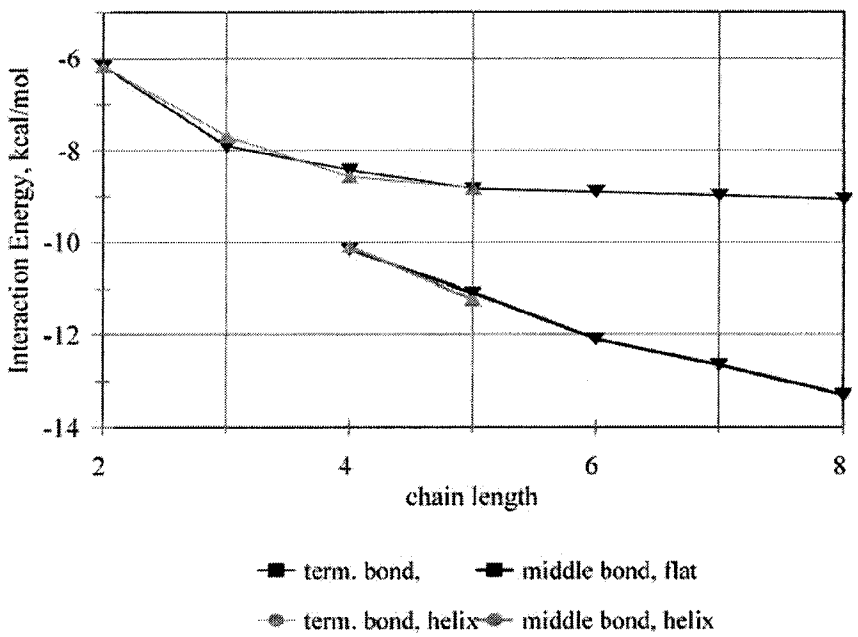


Figure 2.3. Comparison of the optimization plus the correction (1.2 kcal/mol per H-bond) and CP-optimization (optimization on the PES included the counterpoise (CP) correction for the BSSE) for the H-bonded chains of formamides at the B3LYP/D95** level.

Table 2.1: Comparison of the Normal Optimization with the CP-Optimization for the H-bonded Chains of Formamides

Chain length	2FA	3FA	4FA	5FA	7FA[a]
Opt , hartree	*-339.8755*	*-509.8218*	*-679.7690*	*-849.7169*	*-1189.6130*
CP-opt , hartree	*-339.8736*	*-509.8177*	*-679.7632*	*-849.7091*	*-1189.6016*
Diff., hartree	*0.00192*	*0.00418*	*0.00585*	*0.00778*	*0.01133*
Diff., kcal/mol	*1.203*	*2.620*	*3.671*	*4.879*	*7.109*
Average diff. per bond	*1.203*	*1.310*	*1.224*	*1.220*	*1.185*
Overall average diff. per bond					***1.218***

[a] Note, that I could not achieve convergence for the CP-optimization of the chain of six formamides.

needed to obtain the enthalpies in this case were calculated using a complete vibrational analysis for each chain.

The energies of the individual hydrogen bonds are calculated by simply taking the difference between a chain containing n formamides and subtracting the energies of the two smaller chains which remain after the individual H-bond is broken (of i and j formamides where $i + j = n$). Breaking hydrogen bonds that are the same distance from either end of the chain to form the same two smaller fragments will obviously require the same amount of energy (for example, a hexamer can be broken into a dimer and a tetramer by breaking the second or the fourth hydrogen bond).

2.3 Results and Discussion

2.3.1. Constrained Chains of up to 10 Formamides

The calculated terminal and central H-bonding enthalpies are plotted in Figure 2.4 as a function of the number of formamide molecules in the H-bonding chain. Selected energetic values for the dimer, hexamer and decamer are presented in Table 2.2. The data clearly illustrate the strong cooperative nature of the amide H-bond interactions. Thus, the strongest H-bond (the central interaction in the decamer) at the DFT level is predicted to stabilize by 12.87 kcal/mol, approximately 2.5 times as much as that in the formamide dimer (5.07 kcal/mol). The DFT calculations predict slightly stronger H-bonds throughout the series. Interestingly, the cooperativity of the H-bonds in formamide resemble that of the enol of 1,3-cyclohexanedione (Ref. 16 and 17,

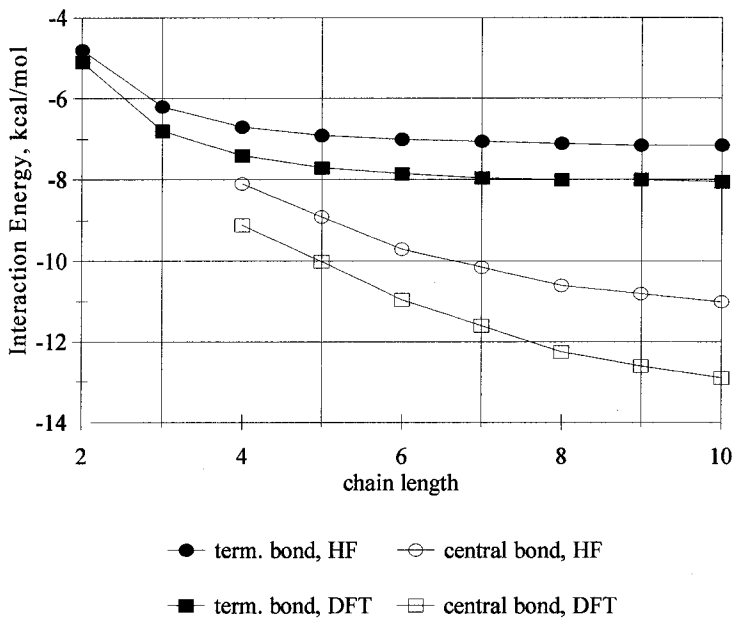


Figure 2.4. Energies of terminal and most central hydrogen bonds in H-bonding chains containing from two to ten formamide molecules (all formamides in each chain were kept geometrically equivalent and coplanar with each other, only intermolecular parameters were completely optimized). Note that chains containing odd numbers of formamides have two energetically equivalent central H-bonds.

Table 2.2: Bonding Energetics of Selected H-bonds in kcal/mol at the B3LYP/D95** Level for the Constrained Formamide Chains[b]

H-bond	*ΔE*	*ΔH(CP-corrected)*
dimer	-7.31	-5.07
hexamer terminal	-10.04	-7.80
hexamer central	-13.20	-10.96
decamer terminal	-10.27	-8.03
decamer central	-15.11	-12.87

[b] All formamides in each chain are geometrically equivalent and coplanar with each other, only intermolecular parameters are optimized completely.

Chapter 1) much more than the less cooperative interactions of the H-bonds of (the structurally more similar) urea (Ref. 14, Chapter 1).

I can also point out that calculated H-bonding energy values for the dimer at the B3LYP/D95** and HF/D95** levels (-7.31 and -6.36 kcal/mol) are very close to the results reported by Kang[16] at the B3LYP/6-31G** and HF/6-31G** levels (-7.27 kcal/mol and -6.34 kcal/mol).

From this preliminary study one could clearly see that the energies of the individual H-bonds within an H-bonded chain depend on both the length of the chain and the position of the individual H-bond in that chain. The large variation of the H-bonding energy within formamide chains strongly suggests that pairwise interactions between individual hydrogen-bond donors and acceptors that are modeled solely upon interaction of one donor and one acceptor will be inadequate to evaluate the H-bonding contributions to the energies of different conformers of polypeptides.

2.3.2. Completely Optimized Chains of up to 15 Formamides

The data in Table 2.3 and Figure 2.5 present the results of DFT calculations on chains containing from 2 to 15 formamides. I shall refer to the H-bonds at different positions from either end of the chain (up to the middle of the chain) as different *types* of H-bonds (k) depending upon their position in a chain. First H-bonds (k=1) exist in all chains starting with the dimer. H-bonds of type two (k=2) occur in chains containing four or more formamides. Third H-bonds (k=3) occur in chains of six or more, etc.

---

[16] Kang, Y. K. *J. Phys. Chem. B* **2000**, *104*, 8321.

Table 2.3: B3LYP/D95** Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) for Completely Optimized Formamide Chains

	k						
n	1	2	3	4	5	6	7
2	-4.49						
3	-6.23						
4	-6.75	-8.49					
5	-7.16	-9.42					
6	-7.23	-9.90	-10.42				
7	-7.31	-10.05	-10.98				
8	-7.39	-10.20	-11.20	-11.62			
9	-7.44	-10.33	-11.42	-11.90			
10	-7.38	-10.33	-11.49	-12.05	-12.12		
11	-7.50	-10.38	-11.60	-12.23	-12.38		
12	-7.42	-10.42	-11.57	-12.27	-12.49	-12.57	
13	-7.48	-10.41	-11.67	-12.31	-12.59	-12.74	
14	-7.42	-10.41	-11.60	-12.35	-12.57	-12.78	-12.85
15	-7.52	-10.45	-11.71	-12.37	-12.71	-12.86	-12.99

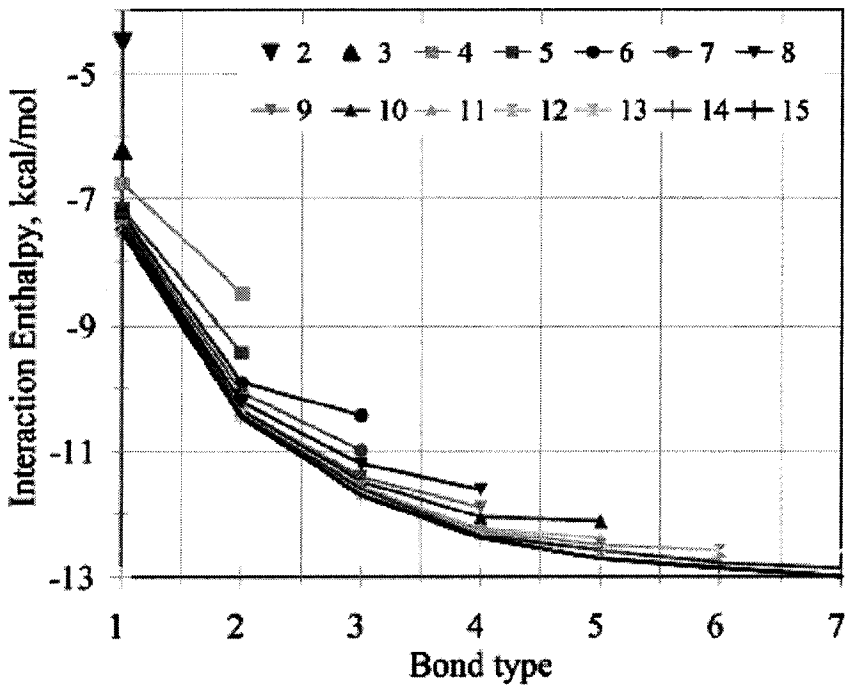


Figure 2.5. Interaction enthalpies for H-bonds organized by H-bond type (k) for chains of the lengths indicated by the symbols. Note that the enthalpies for the first and last H-bonds are the same as are the second and penultimate H-bonds, etc., so that H-bond type varies from 1 to 7.

For reasons explained above, Table 2.3 and Figure 2.5 only contain entries for breaking hydrogen bonds starting at one end, up to and including the most central hydrogen bond. One can immediately see that the hydrogen bonds become stronger as the chain becomes larger. Furthermore, hydrogen bonds close to the interior of the chain are stronger, than those near the ends. The weakest hydrogen bond is that in the dimer (-4.49 kcal/mol), while the strongest bond in the set studied is the central bond in the chain containing 15 formamides (-12.99 kcal/mol). Thus, the strongest bond is almost 2.9 times stronger than that of the dimer. The data in Table 2.1 indicate that the enthalpy of the central H-bond has not quite reached an asymptotic limit at 15 formamides. *These results indicate that a cooperative effect of almost 200% of the dimer H-bonding energy operates in long H-bonding formamide chains.* Clearly the strength of the H-bond depends upon the length of the chain and the position of the H-bond in the chain.

2.3.3. Relationship between Energy of H-bond, H-bonding Chain Length and Position of the H-bond in this Chain

When the energetic data are presented as in Figure 2.6, where each type of H-bond is plotted against chain length, one clearly sees that each type of H-bond asymptotically approaches a distinct limiting enthalpy as the chain length increases. It should also be noted that while the pairs of H-bonds that have the same type have the same energy, they do not necessarily have the same H-bond length, as the two termini of the chains are not equivalent. The linear H-bonding chains of formamides contain a repeating unit of two molecules (see Figure 2.1). The data in Table 2.3 and Figure 2.6

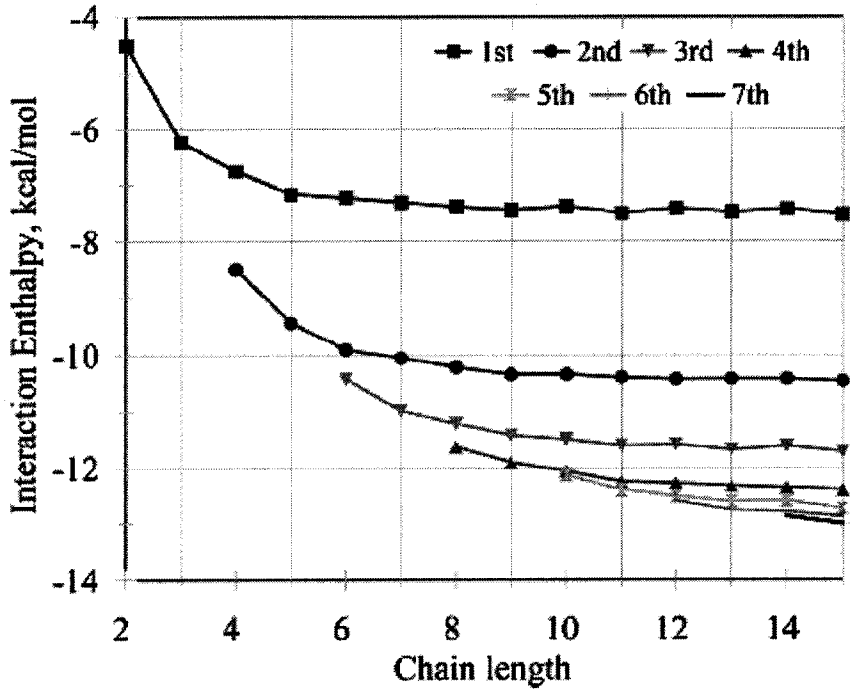


Figure 2.6. Interaction enthalpies for H-bonds organized by chain length for H-bond types (k) indicated by the symbols. Note that the enthalpies for the first and last H-bonds are the same as are for the second and penultimate H-bonds, etc., so that H-bond type varies from 1 to 7.

show a slight oscillation of H-bond enthalpies between chains containing even and odd numbers of formamides consistent with the two molecules repeating unit. This behavior is most apparent for H-bond types that have approached the asymptotic limit for that type. However, due to the small magnitude of this oscillation it has not been taken into account in the further analyses.

The cooperative component of the terminal H-bonds is defined as the additional stabilization over that of the dimer as the chain becomes longer. Based upon the asymptotic relationship for the terminal H-bonds (bond of the first type as displayed in Figure 2.6), one can imagine that an inverse exponential relationship between the cooperativity and the chain length (n) should be appropriate. I fitted the parameters, a_1 and b_1, of Equation 2.1 to get the best fit for the terminal H-bonds.

$$E(n,1) = a_1 + b_1 (1-e^{(2-n)}) \qquad (2.1)$$

In this equation, n represents the chain length. When $n = 2$, the calculated energy should be that of the dimer or a_1 should be set to the H-bond energy of the dimer, E_{dim}. That leads to Equation 2.2 for the terminal (type 1) H-bonds:

$$E(n,1) = E_{dim} + b_1 (1-e^{(2-n)}) \qquad (2.2)$$

The parameter b_1 should be the difference in energy of the H-bond of the dimer and the asymptotic limit (since $exp(-x)=0$, when $x=\infty$). Thus b_1 represents the limiting cooperative effect of the terminal H-bonds (Equation 2.1). From Figure 2.6, one clearly sees that the curvature of the exponential behavior diminishes as the H-bond becomes more removed from the end of the chain. One can consider an equation similar to

Equation 2.1 for each type of H-bond

$$E(n,k) = a_k + b_k\,(1\text{-}e^{(m-n)}) \qquad (2.3)$$

In each case, the first term, a_k, should be the energy of the first H-bond for type k. The limiting cooperative effect for each type would be represented by b_k. The exponent for each equation would change to $(m\text{-}n)$, where m represents the number of monomers in the smallest aggregate that contains that type of bond (for k=3, m=6, since the smallest aggregate with the H-bond of the 3rd type is a hexamer or in general m=2k). I fit a unique equation similar to Equation 2.1 from the data in Table 2.3 for each type of H-bond. This leads to a set of similar equations (like Equation 2.3), one for each value of k. The first term was always taken to be the energy of the first H-bond of the appropriate type. Only b_ks were varied to best fit the data. As expected, b_ks become smaller in magnitude as the H-bond type becomes farther removed from the termini of the chains. The equations fit the data quite well as can be seen from Figure 2.7, which compares the B3LYP/D95** data with those obtained from the empirical equations.

Using this series of equations, one can calculate the energy of any H-bond in formamide chains of this type. However, one needs to know two parameters, a_k and b_k, for each equation. Thus, 14 parameters would be needed to calculate all the H-bonds in Table 2.3. This seems altogether too complex to be useful. However, from Figure 2.8, which presents the values of a_k and b_k for each type, one can see that these values also approach asymptotic limits. Thus, the values of a_k and b_k can be fitted to other functions in order to reduce the number of independent parameters needed to fit all the H-bond

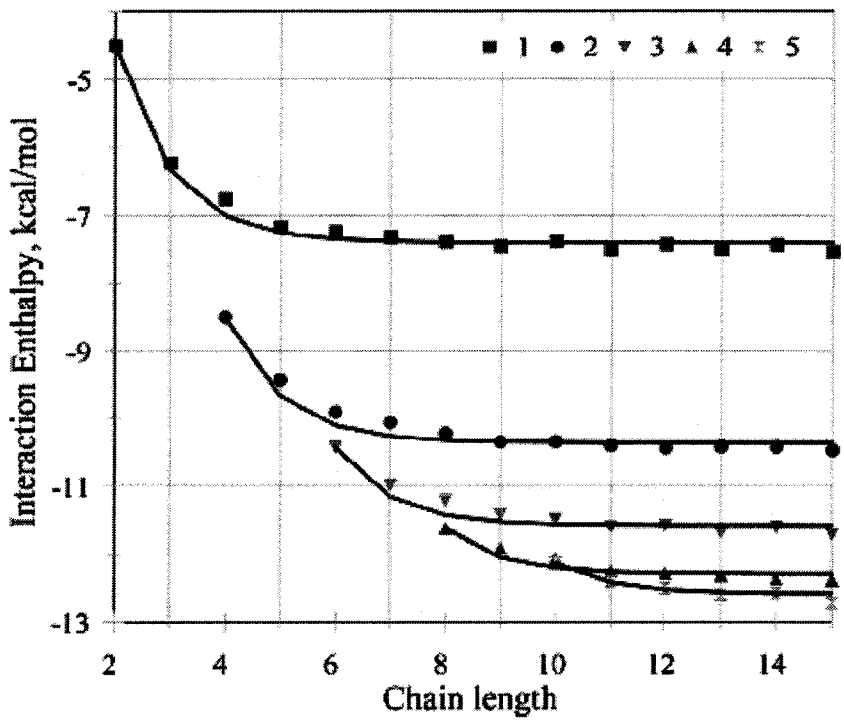


Figure 2.7. B3LYP/D95** H-bond enthalpies results (markers only) compared to those fitted by equation (solid line) for the first five H-bond types (*k*).

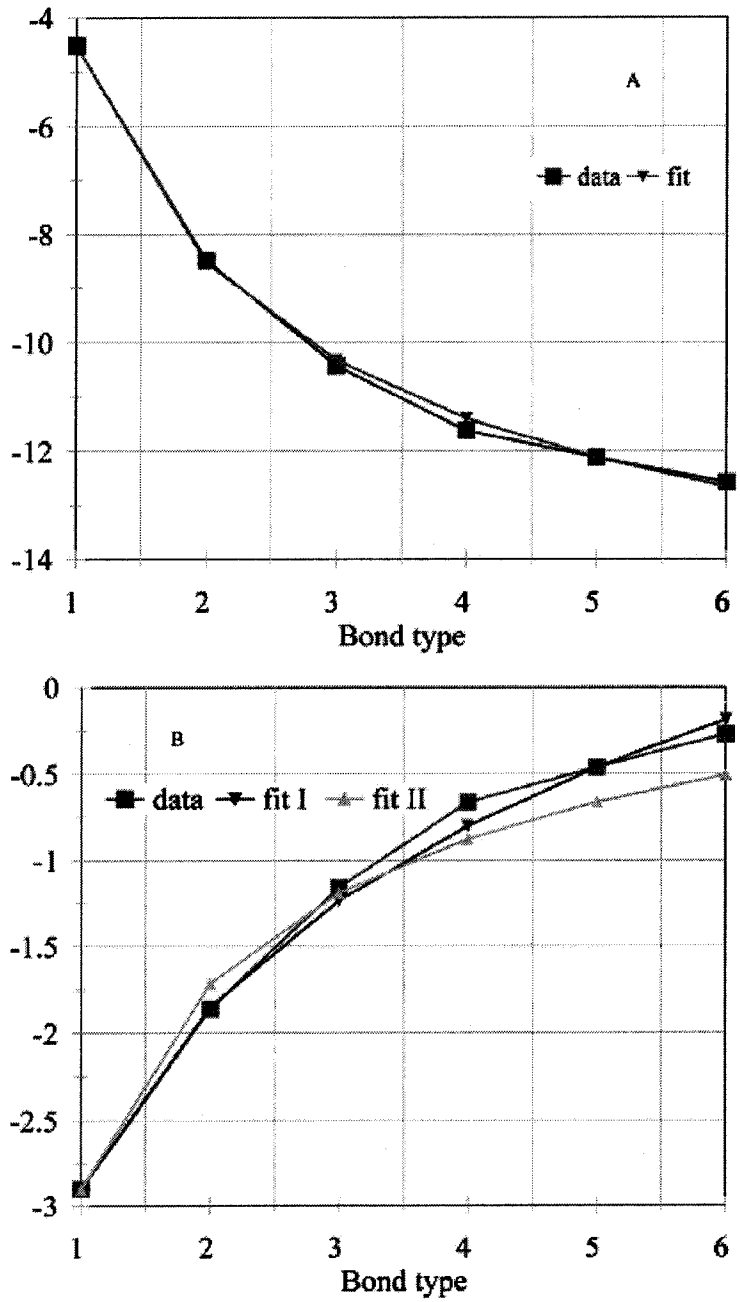


Figure 2.8. Comparison of calculated results with fitted parameters: (A) a_k and (B) b_k. Fits I and II obtained using the equation 2.6 and 2.5 respectively.

energies. I fitted those sets of parameters with similar equations (Equations 2.4 and 2.5). For b_ks though one can also get a slightly better fit if Equation 2.6 is used:

$$a_k = a_1 + c\ (1/sqrt(k)-1) \tag{2.4}$$

$$b_k = b_1 + d\ (1/sqrt(k)-1) \tag{2.5}$$

$$b_k = b_1 + f \ln k \tag{2.6}$$

Substituting for a_k and b_k in Equation 2.3 from Equations 2.4 and 2.5 and taking into account that $m=2k$, one obtains Equation 2.7. If one uses Equation 2.6 rather than 2.5, he obtains Equation 2.8 (where a_1 and b_1 are defined as in Equation 2.1).

$$E(n,k) = a_1 + c\ (1/sqrt(k)-1)) + (b_1 + d\ (1/sqrt(k)-1))\ (1-e^{(2k-n)}) \tag{2.7}$$

$$E(n,k) = a_1 + c\ (1/sqrt(k)-1)) + (b_1 + f \ln k)(1-e^{(2k-n)}) \tag{2.8}$$

Using the data of Table 2.3 to get the best fit to Equation 2.7, I obtain Equation 2.9, while the best fit to Equation 2.8 leads to Equation 2.10.

$$E(n,k) = -4.493 + 13.789(1/sqrt(k)-1)) + (-2.900 - 4.050\ (1/sqrt(k)-1))(1-e^{(2k-n)}) \tag{2.9}$$

$$E(n,k) = -4.493 + 13.789(1/sqrt(k)-1)) + (-2.900 + 1.516 \ln k)(1-e^{(2k-n)}) \tag{2.10}$$

Tables 2.4 and 2.5 present the fitted data from Equations 2.9 and 2.10 respectively, while Figure 2.9 compares the fitted data to the calculated data of Table

Table 2.4: Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) Calculated Using Equation 2.9

	k						
n	1	2	3	4	5	6	7
2	-4.49						
3	-6.33						
4	-7.00	-8.53					
5	-7.25	-9.61					
6	-7.34	-10.01	-10.32				
7	-7.37	-10.16	-11.07				
8	-7.39	-10.21	-11.35	-11.39			
9	-7.39	-10.23	-11.45	-11.94			
10	-7.39	-10.24	-11.49	-12.14	-12.12		
11	-7.39	-10.24	-11.50	-12.22	-12.53		
12	-7.39	-10.24	-11.51	-12.25	-12.69	-12.65	
13	-7.39	-10.25	-11.51	-12.26	-12.74	-12.97	
14	-7.39	-10.25	-11.51	-12.26	-12.76	-13.09	-13.07
15	-7.39	-10.25	-11.51	-12.26	-12.77	-13.13	-13.31

Table 2.5: Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) Calculated Using Equation 2.10

	k						
n	1	2	3	4	5	6	7
2	-4.49						
3	-6.33						
4	-7.00	-8.53					
5	-7.25	-9.70					
6	-7.34	-10.13	-10.32				
7	-7.37	-10.29	-11.10				
8	-7.39	-10.35	-11.39	-11.39			
9	-7.39	-10.37	-11.49	-11.89			
10	-7.39	-10.38	-11.53	-12.08	-12.12		
11	-7.39	-10.38	-11.55	-12.15	-12.41		
12	-7.39	-10.38	-11.55	-12.17	-12.51	-12.65	
13	-7.39	-10.38	-11.55	-12.18	-12.55	-12.77	
14	-7.39	-10.38	-11.56	-12.18	-12.57	-12.81	-13.07
15	-7.39	-10.38	-11.56	-12.19	-12.57	-12.83	-13.04

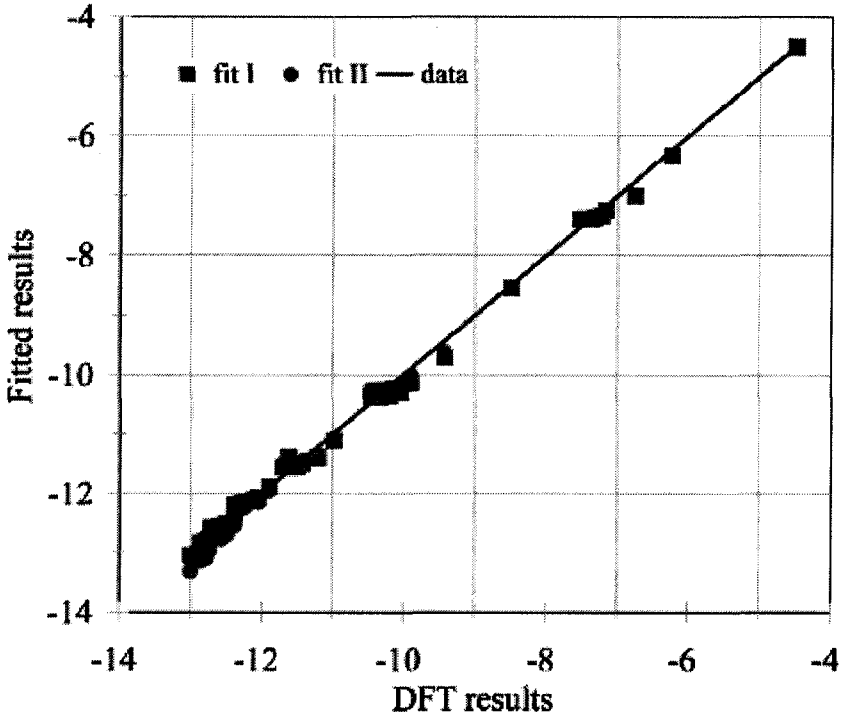


Figure 2.9. Comparison of calculated (DFT) and fitted H-bonding enthalpies.

2.3. In these equations, the best fit for a_k is found to be the hyperbola, inverse square root of k. The relationship $f \ln k$ used in Equation 2.6, 2.8 and 2.10 was obtained purely from empirical curve fitting. The parameters c, d, and f come from the empirical fittings discussed above. While Equations 2.6, 2.8 and 2.10 provide a slightly better fit to the data that I have, they do not extrapolate properly as k goes to infinity. Using these equations would predict that b_k would become infinitely negative (as would the limiting H-bonding enthalpy). Clearly, b_k (and the H-bonding energy) should asymptotically approach some limiting value. Using Equations 2.5, 2.7 and 2.9 preserves this physical principle and using Equation 2.10 does not. But the last equation is still good for practical reasons since the largest H-bonding chain in proteins is rarely longer than 15 H-bonds.[17] So the physical model implied by Equations 2.5, 2.7 and 2.9 is partially sacrificed to preserve a slightly better fit for the shorter chains. I provide both sets of equations as one provides a better fit, the other a somewhat better behaved physical model.

The H-bond lengths in the optimized structures are presented in Table 2.6 and Figure 2.10. As expected, the stronger hydrogen bonds tend to be shorter than the weaker ones. Since the formamide chains are not symmetrical (one end terminates with an amino group in the other with the carboxyl group), the H bond lengths for the pairs of hydrogen bonds that have the same strength within a chain are not exactly the same. Figure 2.11 shows the relationship between the H-bond strengths and the O..H distances.

---

[17] Very long α-helices (almost 100 residues long) have been recently observed. That means that the long H-bonded chains containing more than 30 hydrogen bonds exist. See, Hoeflich, K. P.; Tsukita, S.; Hicks, L.; Kay, C. M.; Tsukita, Sh.; Ikura, M. *Biochemistry* **2003**, *42*, 11634.

Table 2.6: H-Bond Lengths from B3LYP/D95** Optimized Geometries as a Function of Chain Length (n) and Position of H-Bond in Chain (Numbering as in Figure 2.1)

							n							
bond	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.920	1.873	1.859	1.853	1.850	1.850	1.849	1.850	1.847	1.847	1.847	1.847	1.847	1.847
2		1.878	1.835	1.813	1.810	1.808	1.809	1.805	1.805	1.807	1.805	1.806	1.805	1.806
3			1.869	1.819	1.806	1.797	1.793	1.792	1.792	1.791	1.790	1.790	1.790	1.790
4				1.862	1.814	1.797	1.791	1.786	1.789	1.785	1.785	1.785	1.785	1.785
5					1.861	1.810	1.794	1.786	1.786	1.783	1.780	1.781	1.781	1.781
6						1.860	1.808	1.793	1.787	1.780	1.781	1.782	1.781	1.782
7							1.859	1.807	1.794	1.784	1.780	1.781	1.781	1.782
8								1.859	1.807	1.794	1.784	1.780	1.780	1.781
9									1.858	1.805	1.790	1.784	1.780	1.781
10										1.858	1.807	1.791	1.784	1.780
11											1.857	1.806	1.791	1.784
12												1.858	1.806	1.791
13													1.857	1.806
14														1.858

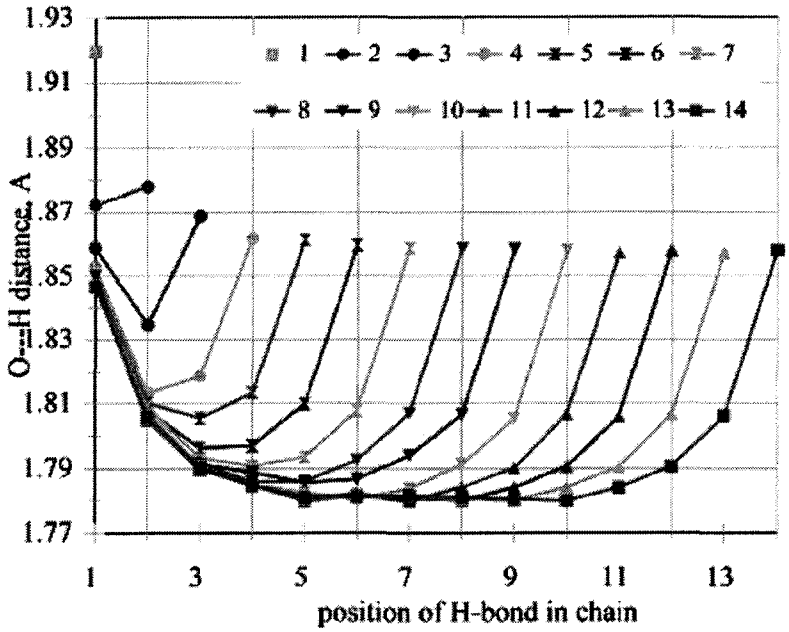


Figure 2.10. Calculated B3LYP/D95** H-bond lengths organized by bond position in chains.

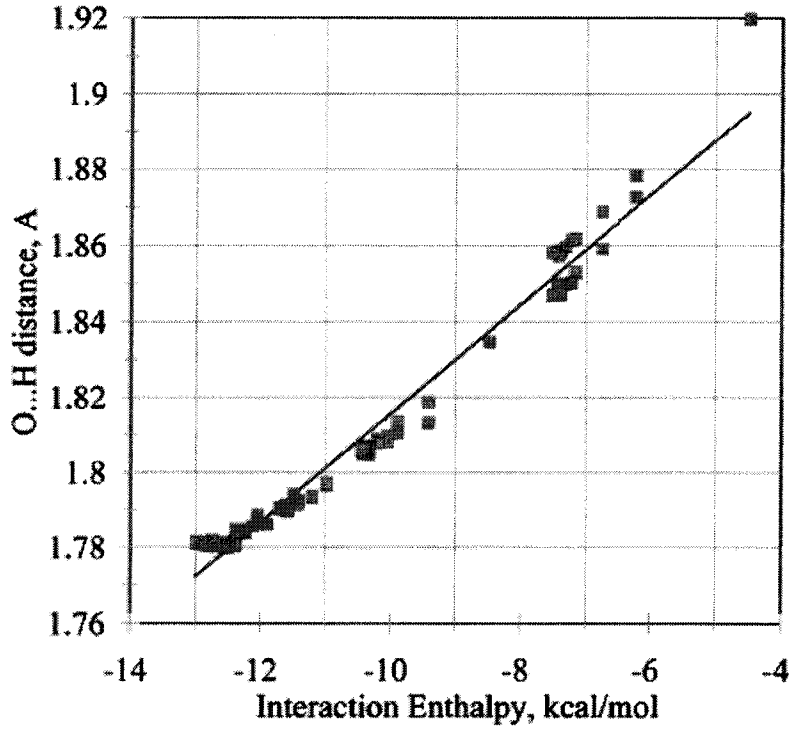


Figure 2.11. Relation between H-bond length and interaction enthalpies for all H-bonds calculated. The best linear fit is plotted: *H-bond Length (Å) = 0.0147 x Interaction Enthalpy (kcal/mol) + 1.9619* ($R^2 = 0.9765$).

H-bonding distance is often taken as a linear function of the energies. While this might be approximately true within a certain range of normal H-bond lengths, it cannot hold in general. Clearly, at very long H-bonding distances the interaction energies must approach zero. Equation 2.11, where D is the O...H distance in Å and E is the H-bond interaction enthalpy,

$$D = 0.0147\,E + 1.9619 \tag{2.11}$$

gives the best linear fit (R^2=.9765) within the range of the data that I have. Of course, the relationship cannot be linear for complete variation of D, as the interaction enthalpy must go to zero for large values of D and to infinity as D approaches zero. The present data are insufficient for defining the expected curvature in function that expresses the relationship of D with interaction enthalpy for a larger range of D. Despite the fact that two energetically equivalent H bonds within the same chain have different lengths, the correlation between bond strengths and lengths remains reasonably good. Thus, if one can calculate the H-bond energy using the relationship described above, one can obtain a reasonable estimate of the H-bond distance from Equation 2.11.

2.4. Origin of the Cooperativity

The unusually large cooperative interactions between H-bonds in formamide chains result form a combination of different kinds of physical interactions. Pair-wise interactions suffice to physically model electrostatic interactions. Three-body interactions are necessary to account for polarization, while many-body interactions are

required to properly describe mutual polarization and covalent interactions.[18] The electrostatic component can be considered as composed of dipole-dipole interactions. These are proportional to $1/R^3$, where R is the distance between the dipoles. If *N* dipoles be linearly aligned and evenly spaced, the 1-3 interactions would be 1/8th of the 1-2 interactions as the 1-3 distances are twice the 1-2 distances. Similarly, the 1-4 interactions would be 1/27th, the 1-5 interactions 1/64th of the 1-2 interactions, etc. The maximum cooperative electrostatic interaction would be a sum of all such pair-wise interactions. For the strongest central bond in the chain of 15 formamides, this interaction would be 1.50 times that of a dimer, while the H-bond in an infinite chain would be about 1.61 times that of a dimer. The data that have been presented above clearly show this electrostatic model to be inadequate to describe the major component of cooperativity. As the H-bonds studied readily become close to three times (200% larger) in magnitude than the prototypical N-H...O interaction of the formamide dimer, the calculated electrostatic cooperative effect of 50% represents only 1/4 of the total cooperativity. The results presented agree with the interpretation of Weinhold[4] (for HCN H-bonding chains) rather than with that of Wu, who prefers an electrostatic model.[7]

One must look elsewhere for physical contributions to the cooperativity. Examination of the relationship between the C=O and C-N bond distances (Figures 2.12 and 2.13) as a function of the H-bonding chain length and the position of the formamide in the chain clearly illustrates a structural response by each formamide unit to the H-bonding within its chain. As the chains become longer and the formamide monomer

[18] Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339.

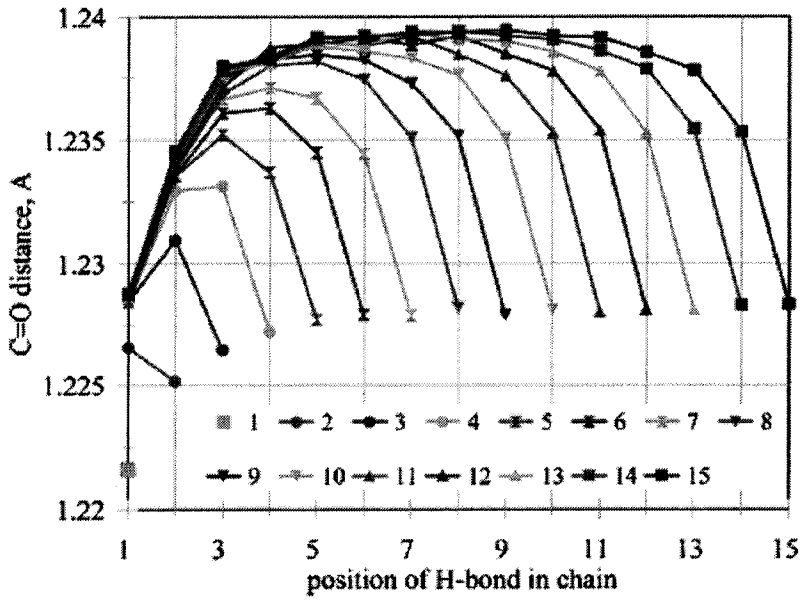


Figure 2.12. Calculated B3LYP/D95** C=O lengths organized by bond position in chains of lengths indicated.

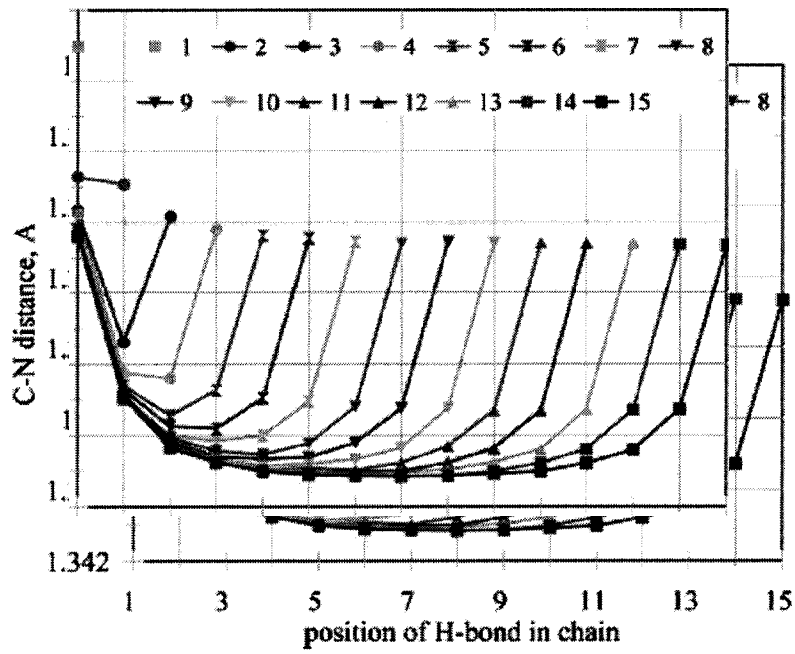


Figure 2.13. Calculated B3LYP/D95** C-N lengths organized by bond position in chains of lengths indicated.

becomes more central, the C=O distance increases, while the C-N distance decreases. These relationships further indicate that the cooperativity cannot be purely electrostatic in origin. However, they could be the result of a combination of polarization, mutual polarization and covalent interactions. Gilli has emphasized the importance of resonance assisted hydrogen bonds (RAHBs) in several crystal structures.[19] This question was addressed in a model study where the response of two molecules that form extensive H-bonding chains in their crystal structures was compared to an electric field that reproduces the interaction energy of the molecule in the H-bonding environment.[20] Since only electrostatic interactions and polarization (not mutual polarization or covalent interactions such as resonance assistance) can occur in the model of a molecule in the electric field, the relative importance of the two pairs of effects could be addressed. The H-bonding environment of one molecule (urea) was somewhat reasonably described by the applied electric field. That of the other molecule (the enol of 1,3-cyclohxanedione), which Gilli considers a model of RAHB, was extremely poorly described by the electric field. H-bonding chains of 1,3-cyclohexane clearly afford a more favorable situation for RAHBs as forming a covalent O-H in place of the H-bond O...H and vice versa would create an identical H-bonding chain if done for each H-bonding interaction. Thus, the second resonance structure would be very close to the first in energy. They would have the same energy if the O...H and O-H distances were equal. For formamide chains, one

---

[19] (a) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Acta Crystallogr., Sect. B: Struct. Sci.* **1998**, *B54*, 50. (b) Gilli, G.; Bertolasi, V.; Ferretti, V. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *B49*, 564. (c) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Chem.-Eur. J.* **1996**, *2*, 925.

[20] Dannenberg, J. J.; Haskamp, L.; Masunov, A. *J. Phys. Chem. A* **1999**, *103*, 7083.

needs to form an O-H in place of an O...H and an H...N in place of an H-N. This would provide a second resonance structure of clearly higher energy than the first. RAHB is expected to be an important (but not the only) contributor to the cooperativity that was observed. The effects of the H-bonding cooperativity upon the vibrational modes of these formamide chains are addressed in detail in Chapter 4.[21]

The nonlinear increase in the dipole moment of the formamide chains as they increase in length provides yet another indication of the inadequacy of an electrostatic explanation to the observed cooperativity. Figure 2.14 illustrates the average dipole moment of a formamide molecule in chains of increasing length. This value increases with chain length approaching asymptotic limits that are slightly different for chains containing even and odd numbers of formamides. In an electrostatic model, the dipole moment of N formamides would simply be the vector sum of the individual N dipoles. One cannot determine from the data of Figure 2.14 how much of the increase in average dipole moment is due to polarization, mutual polarization and covalent (or RAHB) interactions. All of these effects would tend to increase the dipole moments as observed. This behavior is qualitatively similar to the increasing dipole moments per amino acid residue of α-helical peptides as the helix increases in length.[22] One should note that calculation of the dipole-dipole electrostatic interactions (see above) using the average dipole moment of a formamide in a long chain will give approximately the correct H-bonding interaction for the strongest H-bond in a 15-mer. However, it should be noted

---

[21] Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A*, **2003**, *107*, 6688.

[22] Applequist, J.; Mahr, T. G. *J. Am. Chem. Soc* **1966**, *88*, 5419.

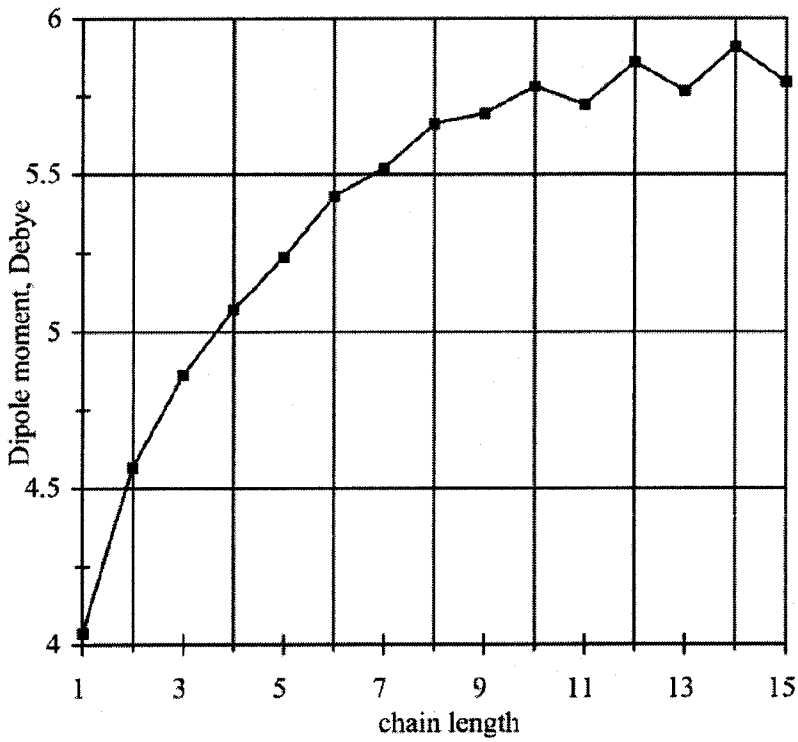


Figure 2.14. Dipole moments per formamide monomer as a function of chain length. The repeating unit of two formamides in the linear periodic structure gives rise to an alternation of dipoles for chains of even and odd numbers of formamide.

that an electrostatic model cannot account for the increase in average dipole moment as the chain increases.

2.5. Application to Protein Models

Recent reports of trans-H-bond ^{13}C-^{15}N couplings in the NMR spectra of α-helical peptides confirm that the H-bonds closer to the middle of the helices are shorter[18] (thus stronger) than those near the ends. Each α-helix contains three H-bonding chains that are similar to the formamide chains discussed in this chapter. Kemp recently explained his experimental demonstration of an enthalpic component to the cooperativity of α-helical peptides as evidence for H-bond cooperativity in these structures.[6] These results are consistent with calculations reported by Wu on helical polyglycines,[7] as well as the recent calculations on five completely optimized α-helical peptides containing 17 amino acid residues by our group.[12] Curiously, Wu has not observed similar cooperativity in β-sheets of α-polyglycines,[9] but he reported H-bonding cooperativity in sheets of β-polypeptides (constructed of peptides of β-amino acids).[8] In recent studies by our group some inter-strand H-bonding cooperativity was observed in sheets of γ-polypeptides but again none in β-sheets of α-polyglycines.[10]

Tailoring the relationships between chain length and H-bond type and position in the chain that have been presented above to modeling protein structure does not pose major conceptual or calculational problems. The determination of the energy of each H-bond using either Equation 2.9 or 2.10 is straightforward and economical with respect to computer time. Nevertheless, some modifications will be necessary for this (or similar)

equations to be directly useful for peptide modeling. Crystallographic H-bond distances in α-helices tend to be longer than those in β-sheets, which are longer than those calculated here for formamide chains.[23,24] These differences are likely to be due to the strain required to form the secondary peptide structures and the fact that most of the C=Os will form another H-bond to water or another H-bonding donor,[23,24] thereby weakening and lengthening the peptide H-bond. The formamide chains are strain free. In principle, the effect of the strain on the H-bond strength can be parameterized in improved peptide models that include the relationships derived here. The H-bonding interaction energies reported in this study were lately used by Wieczorek and Dannenberg to estimate the steric strain of formation of the α-helical structure.[12] In order to get this estimate (in the polyalanine of 17 peptide unites) the authors took the difference between the interaction enthalpies of three five-H-bond formamide chains obtained in our study and the difference in energies between the α-helical and open β-strand forms of the polyalanine. The estimate of the value of the strain was found to be 6.6 kcal/mol per H-bond at the B3LYP/D95** level. Comparing this estimated value with the value of the H-bonding interaction energy in the formamide dimer (-4.5 kcal/mol) they concluded that H-bonding chains should inherit significant cooperativity in order to overcome the strain.

H-bonds are the most important interactions between amino acid residues in peptides, as these interactions provide the stability for secondary structures, such as

---

[23] Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Molec. Biol.* **1987**, *44*, 97.

[24] Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.

helices or sheets. Most methods currently used to model protein structure treat the H-bonds as pair-wise interactions between nearest neighbors. Empirically fitted pair-wise potentials do not necessarily represent purely electrostatic interactions. Rather they use whatever functions chosen to reproduce a set of data by appropriately fitting the parameters in the defined functions. Since the energy of an individual H-bond in a formamide chain is seen to vary with chain length and position within the chain, the same pair-wise potential cannot be appropriate to all such interactions. However, fitted pair-wise interactions could be scaled to account for the position in and size of the H-bonding chain using functions similar to those described in this Chapter.

The geometries of hydrogen-bonded chains reported in this work were used in another the study by our group of the trans H-bond ^{13}C-^{15}N 3-scalar J-couplings for peptide models.[25]

---

[25] Salvador, P.; Kobko, N.; Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14190.

* * *

Finally, I should mention that after the results described in this Chapter were published I designed yet another equation that fits the data and employs only three parameters. When a_k is defined as in Equation 2.4 and

$$b_k = 5b_1/(k^2 + 4) \tag{2.12}$$

then

$$E(n,k) = -4.493 + 13.789 * (1/sqtr(k) - 1) - 2.900 * (1-e^{2k-n}) * 5/(k^2 + 4) \tag{2.13}$$

This equation also gives clear meaning to the second parameter ($+13.789$). It is the absolute difference between the energy of the weakest H-bond (in the dimer) and the asymptotic limit for H-bond strength. This equation should work for H-bonding formamide chains of any length and give the reasonable H-bond strengths even for very long chains. Thus, when k is infinitely large, the last term in the sum becomes zero and the multiple in parenthesis in the second term becomes -1. That mean that $E = -4.493 - 13.789 = -18.282$ *kcal/mol*, which is the limit for the H-bond strength in this model. Table 2.7 displays the hydrogen bonding energies for the chains of up to 15 hydrogen bonded formamide molecules calculated using Equation 2.13. To get energies of the hydrogen bonds in any system using this equation one needs to know the energy of the H-bond in the dimer (*-4.493 kcal/cal* in our case) and the asymptotic value of the energy of the bond of the first type in the corresponding H-*bonded* chain (one can get it from the calculations of the relatively long chain; in our case this energy is *-7. 393*

Table 2.7: Enthalpies (kcal/mol) of Individual H-Bonds as a Function of Chain Length (n) and H-Bond Type (k) Calculated Using Equation 2.13

	k						
n	1	2	3	4	5	6	7
2	-4.49						
3	-6.33						
4	-7.00	-8.53					
5	-7.25	-9.68					
6	-7.34	-10.10	-10.32				
7	-7.37	-10.25	-11.03				
8	-7.39	-10.31	-11.29	-11.39			
9	-7.39	-10.33	-11.38	-11.85			
10	-7.39	-10.34	-11.42	-12.01	-12.12		
11	-7.39	-10.34	-11.43	-12.08	-12.43		
12	-7.39	-10.34	-11.43	-12.10	-12.55	-12.65	
13	-7.39	-10.34	-11.44	-12.11	-12.59	-12.88	
14	-7.39	-10.34	-11.44	-12.11	-12.61	-12.97	-13.07
15	-7.39	-10.34	-11.44	-12.11	-12.61	-13.00	-13.24

kcal/mol for the chains of at least eight formamides). One also needs to estimate the overall strongest energy for the H-bonding in the system. Again this can be done by performing calculations on relatively large cluster. I tried to get this estimate for my system from the Periodic Boundary Condition (PBC) Calculations (as implemented in Gaussian 03, see Ref. 23, Chapter 1) on the infinite H-bonded formamide chain. However, as Table 2.8 indicates the PBC calculations effectively underestimate the cooperativity. The H-bonds obtained with PBC when only two formamide molecules are included in the unit cell (Table 2.7, ***A***) are weaker and the overall changes in bond's lengths of the formamide molecule are much smaller than that for the central bond in the long hydrogen-bonded formamide chain optimized without CP-correction. Similar results are obtained for the unit cell with 4 independent molecules (Table 2.7, ***B***). This similarity indicates convergence of numerical integration on the number of points of reciprocal space. The significant difference between PBC and cluster results may indicate the deficiency in multipole approximation for integral evaluation used in PBC.

Table 2.8: H-Bonding Energy and Geometric Parameters for the Infinite H-Bonding Chains of Formamides Obtained by the Periodic Boundary Condition Calculations (PBC) with Two Formamides in Unit Cell (A) and Four Formamides in Unit Cell (B) Compared to that for the Central Formamide in the Chain of 15 Formamides Obtained by the Normal Optimization (Opt)

A

	PBC	Opt	(Opt–PBC)
H-bond Energy, kcal/mol	-10.23	-15.86	-5.63
Geometry			
O...H	1.7779	1.7810	0.0031
OC	1.2359	1.2393	0.0034
CN	1.3450	1.3437	-0.0013
NH (H-bond)	1.0243	1.0287	0.0043
NH	1.0122	1.0123	0.0001
CH	1.1043	1.1042	0.0001

B

	PBC	Opt	(Opt –PBC)
H-bond Energy, kcal/mol	-10.35	-15.86	-5.51
Geometry			
O...H	1.7793	1.7810	0.0017
OC	1.2386	1.2393	0.0007
CN	1.3438	1.3437	-0.0001
NH (H-bond)	1.0269	1.0287	0.0014
NH	1.0122	1.0123	0.0001
CH	1.1041	1.1042	-0.0001

3.
HYDROGEN BONDING AND RELATIVE STABILITY OF POLYGLYCINE CONFORMERS

3.1. Introduction

Protein folding is increasingly recognized as an important field with wide relevance to a number of biochemical processes.[1] Although the nature of the residues forming the polypeptide plays essential role in the final geometry and functioning[2] of protein, both inter- (such as between the backbone and the solvent, for example) and intra-molecular hydrogen-bonding is considered to be among the driving forces in the chain of transformations from one-dimensional amino-acid sequence into three-dimensional folded protein structure.[1b,3] For example, the H-bonding interactions between

---

[1] (a) Dill, K. A., *Biochem.* **1990**, *29*, 7133. (b) Friesner, R. A.; Gunn, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 315. (c) Dobson, C. M.; Šali, A.; Karplus, M., *Angew. Chem. Int. Ed.* **1998**, *37*, 868. (d) *Accounts of Chemical Research* **1998**, *31 (11)*, entire issue.

[2] See, for example, (a) Tobias, D. J.; Brooks, C. L., III *Biochemistry* **1991**, *30*, 6059. (b) Wang, Y.; Kuczera, K. *J. Phys. Chem. B* **1997**, *101*, 5205. (c) Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974. (d) Hudgins, R. R.; Jarrold, M. F. *J. Phys. Chem. B* **2000**, *104*, 2154. (e) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 14065.

[3] See, for example, (a) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605. (b) Kim, K.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 12952.

amide groups in different secondary structures of proteins define (although, not exclusively[4,5,6,7]) their relative stabilities.[5,8] Therefore gaining a comprehension of the nature and main characteristic features of H-bonding and other factors that influence the secondary structures of polypeptides will aid in building better models for protein folding. That in turn will not only help to further understand and model many processes of biological importance, but will also facilitate in design and syntheses of new drugs and other bioactive materials. And since some factors that govern the behavior of the biological molecules have cooperative or long-range character (e.g., H-bonding, dipole-dipole[9] interactions), their description requires computational study of relatively large systems[6,10,11] at sufficient level of theory[12].

There are four major conformations of polypeptides that constitute their secondary structure: (a) C5 or, using another notation, 1_5 fully extended ribbon (β-strand), (b) C7 or 2_7-ribbon (repeated γ-turns), (c) C10 or 3_{10}-helix, (d) C13 or 4_{13}-helix,

---

[4] Viguera, A. R.; Serrano, L. *Biochemistry* **1995**, *34*, 8771.

[5] Iprota, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Am.Chem. Soc.* **2001**, *123*, 3311.

[6] Park, C.; Goddard, W. A., III *J. Phys. Chem. B* **2000**, *104*, 7784.

[7] Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.

[8] See, for example, (a) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriat, I.; Muir, T.; Gellman, S. H. *PNAS* **2001**, *98 (21)*, 12015. (b) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124. (c) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198.

[9] Dipole-dipole interactions are proportional to $1/r^3$ in solids (long-range) and to $1/r^6$ in liquids (short-range), where r is the distance between the dipoles. In the case of proteins the secondary structures are relatively rigid and behave more like solids, so one can assume that dipole-dipole interactions are proportional to $1/r^3$ and therefore they are long-range.

[10] Schafer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272.

[11] Iprota, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *114 (6)*, 2541.

[12] See for example, Sirois, S.; Proynov, E. I.; Nguyen, D. T.; Salahub, D. R. *J. Chem. Phys.* **1997**, *107 (17)*, 6770.

also known as α-helix[13] (see Figure 3.1). In recent years the subject of the relative stabilities of the polypeptides conformations has been studied extensively at different levels.[5,7,8c,11,14] Also there are a few reports (both theoretical[2a,2d,7,15,16] and experimental[17,18,19]) that focus specifically on the relations of the helical conformers of various polypeptides. Most of the studies report similar stability trends for the conformers of particular polypeptides. However, Wu *et al.* reported the relative stabilities of the conformers of polyglycines of up to 14 residues to vary with the level of theory used[12] and Improta *et al.* showed that the relative stabilities per residue of the conformers of infinite polyglycine depend strongly on the DFT functional used.[11] The stability order for the same length polypeptides may also differ depending on their sequence.[5]

In this chapter I performed a study of the relative stabilities of the conformers of polyglycine containing up to 19 residues at the B3LYP/D95**//AM1[20] level of theory. I will show that there is much more cooperativity in formation of the helical conformers.

---

[13] Creighton, T. E. *Proteins: Structure and Molecular Properties*; W. H. Freeman and Co.: New York, 1993.

[14] Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Freisner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.

[15] Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. *J. Am. Chem. Soc.* **1993**, *115*, 11590.

[16] Topol, I. A.; Stanley, K. B.; Tang, T.-H.; Perczel, A.; Rashin, A.; Csizmadia, I. G. *J. Am. Chem. Soc.* **2001**, *123*, 6054.

[17] (a) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873. (b) Hanson, P.; Martinez, G.; Millhauser, G.; Formaggio, F.; Crisma, M.; Toniolo, C.; Vita, C. *J. Am. Chem. Soc.* **1996**, *118*, 271. (c) Bolin, K. A.; Millhauser, G. L. *Acc. Chem. Res.* **1999**, *32*, 1027.

[18] Long, H. W. Tycko, R. *J. Am. Chem. Soc.* **1998**, *120*, 7039.

[19] Sudha, R.; Kohtani, M.; Breaux, G. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2004**, *126*, 2777.

[20] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3909.

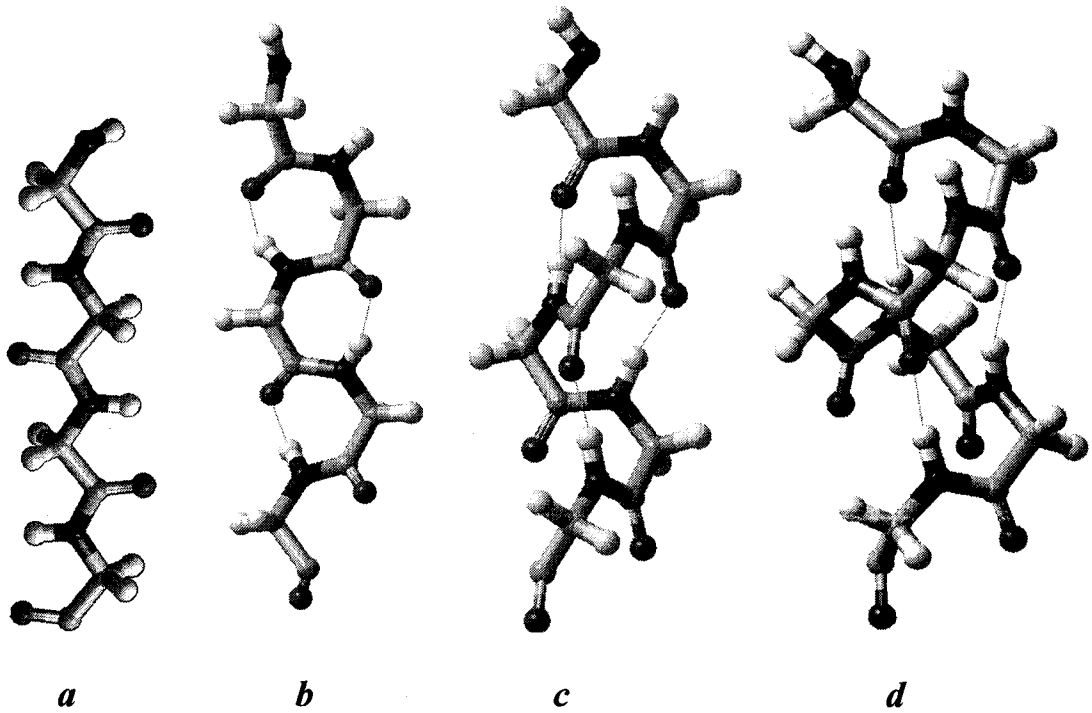


Figure 3.1. Schematic drawings of different conformers of polyglycines (a) C5; (b) C7; (c) 3_{10} helix; (d) α-helix. Dotted lines represent H-bonds.

3.2. Computational Details

Semi-empirical (AM1) geometrical optimization of the different conformers of polyglycines was performed with constraint that all the residues within each conformer be identical in structure. I then carried out single point DFT calculations at the B3LYP/D95** level on the AM1-optimized geometries. All calculations were executed using Gaussian 98 suite of computer programs (see Ref. 28, Chapter 1) on the cluster of Intel and AMD powered computers that are parallelized using LINDA (see Ref. 29, Chapter 1). The number of nodes used in each calculation varied depending on the size of the polyglycine. In this study I considered the four major conformers of polyglycine (constructed of two to nineteen glycine residues) mentioned above.

3.3. Results and Discussion

I began this study by comparing polycondensation energies for the four conformers of polyglycines. These energies ΔE^{n}_{polyc} were calculated according to the following equation:

$$\Delta E^{n}_{polyc} = E^{n}_{chain} - nE^{1} + (n-1)E_{H_2O} \qquad (3.1)$$

where n represents the length of the polyglycine (from 2 to 19), E^{n}_{chain} is the energy of the polypeptide chain of n residues, E^{1} is the energy of the single glycine molecule (for the most stable glycine conformation without the H-bond), and E_{H_2O} is the energy of the water molecule. As can be seen from the Figure 3.2 in order to break the polyglycine into separate glycine molecules one would need to add to the system one water molecule for each bond that is being broken to complete the resulting glycines.

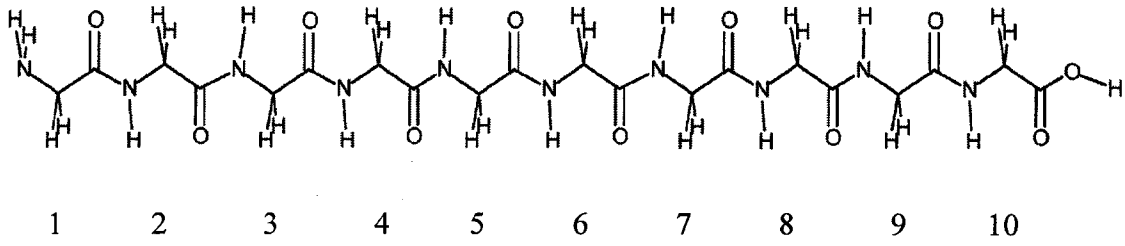


Figure 3.2. Example of the system used in this study. Numbers denote residues.

Figure 3.3 shows the relationship between polycondensation energies of different polyglycine conformers as a function of the length of the polypeptide. The stabilization trend for the short polyglycine chains (up to 7 residues) in this study is:

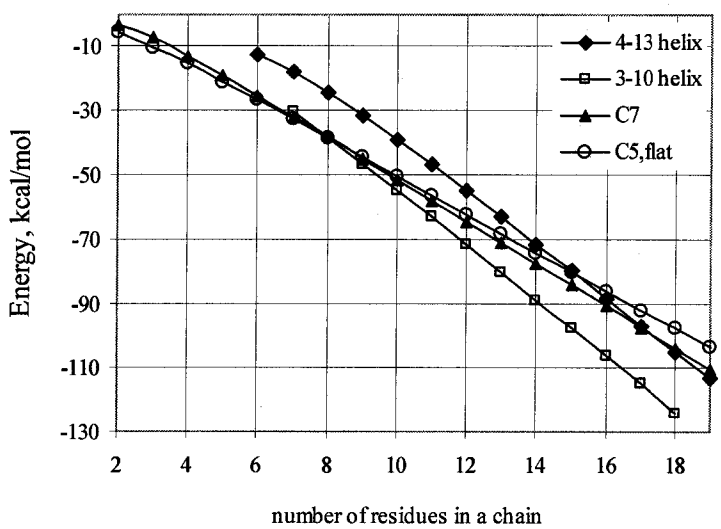
C5 > C7 > 3_{10} helix > α-helix.

For the chains of 8 residues long, the stability of C5, C7 and 3_{10}-helical conformations is close (within 0.3 kcal/mol) and by ~14 kcal/mol exceeds stability of α-helix. For the chains longer than 8 residues the stability of 3_{10}-helix surpasses that of the C5 conformation (β-strand, which exhibits only intraresidual H-bonds) and of the C7 conformation (where the H-bond is formed between each residue and its next nearest neighbor). My data predict the α-helix to be the least stable conformation in the chains of up to 15 residues. But this conformation becomes lower in energy relatively to the C5 and C7 chains when containing more than 15 and 17 residues respectively. The stability order for the chains of more than 17 glycines is the following:

3_{10} helix > α-helix > C7 > C5.

This is in a partial agreement with the trend reported in Ref. 11 for the infinite glycine (and alanine) homopolypeptides that was obtained at the PBE/6-31G* level with the periodic boundary conditions (PBC). According to this paper the α-helix is the most stable conformation followed by the 3_{10}-helix, C7, and C5 conformers. On the other hand, my study shows that the polycondensation energies per residue for the C5 conformer and, to a slightly lesser degree, for the C7 conformer almost reached their asymptotic values (Figure 3.3B). For the two helical conformers these energies are still far from their respective asymptotic values, indicating that the formation of the helices is more cooperative. I should also point out that the general trends presented in Figure 3.3A

A



B

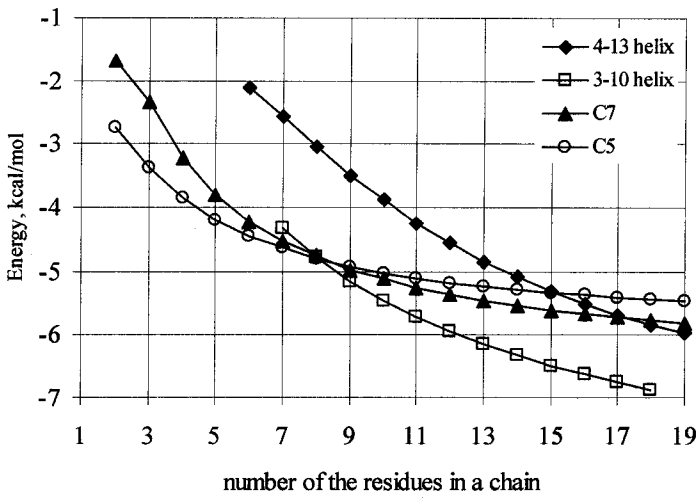


Figure 3.3. Polycondensation energies (A) and polycondensation energies per residue (B) for the conformers of polyglycines containing up to 19 residues at the B3LYP/D95**//AM1 level with the constraint that all the residues within each conformer be identical in structure. Note that I could not achieve optimization at the AM1 level for the short 3_{10} and α-helices.

are in a relatively good agreement with the data reported by Wu and Zhao (see Ref. 7, Figure 2). In their study they calculated the energies of the four polyglycine conformers at the HF/6-31G* and B3LYP/6-31G* levels using repeating unit approach where all amino acid residues in all chains of the same conformation have the same fixed geometry. This method is somewhat similar to the approximation that was used in my study. Though their general results are comparable to the results reported here there are a few important distinctions. (1) Contrary to my data they report that at the B3LYP/6-31G* level the energy of the C5 conformer increases linearly with respect to the C7 (the C7 conformer of 14 residues is more stable than C5 by 9 kcal/mol) and at the HF/6-31G* level the C5 conformation is always slightly more stable than C7 and the difference in their energies is almost independent of the length of the polypeptide chain (around -1.5 kcal/mol). In the study presented here the C5 conformer is the most stable one for the chains of up to 8 glycines. After that it becomes less stable than the C7 conformer, but the difference in their energies reaches only 6.9 kcal/mol for the chains of 18 residues. (2) The two helical structures in the study by Wu and Zhao become more stable than the C5 and C7 conformers for the much shorter chains (especially at the HF/6-31G* level). (3) Their data clearly indicate that the α-helix has lower energy than the 3_{10}-helix and therefore becomes the most stable of the four conformations when the chains are relatively long (longer than 20 and 24 residues at the B3LYP/6-31G* and HF/6-31G* levels respectively). The α-helical polyalanine is also reported to become lower in energy relatively to the 3_{10}-helcal one when the chains are 20 or more alanines long in the study by Wieczorek and Dannenberg.[15] Their results are for the capped polyalanines optimized at the DFT(B3LYP/D95**)/AM1 ONIOM level. Though in my study both helical

structures exhibit considerable cooperativity, the 3_{10}-helix is still more stable than the α-helix by almost 19 kcal/mol for the chains of 18 glycines (Figure 3.3A). Furthermore, my data seem to suggest almost equal rates of stabilization with the size of the system for the two helical structures (see Figure 3.3B), while the value of the polycondensation energy per residue for the 3_{10}-helix is still more negative than that for the α-helix by more than 1 kcal/mol for the longest chains studied. These facts appear to suggest that the cooperativity in the 3_{10}-helix is more favorable. However I should point out that glycine residue is generally known to destabilize α-helix.[8b] In addition one has to remember that the 3_{10}-helices, when compared to the α-helices of the same length, (1) have one extra hydrogen bond; (2) have only two H-bonding chains (vs. three in the latter) which are longer. This means that each 3_{10}-helix longer than 5 glycines has 4 terminal H-bonds and each α-helix of more than 8 glycines has 6 of these bonds which are also more diverse. As I show in Chapter 2 the central hydrogen bonds in the H-bonding chains of formamides is much stronger than the terminal ones (in 1.25 times in tetramer, in 1.44 times in hexamer, in 1.57 times in octamer, etc.). The constraint used in this study introduces averaging over all residues in each polyglycine. Therefore the termini of the helical conformers will have more effect on the average values in the α-helices than in the 3_{10}-helices. This may make the H-bonding interaction in the α-helices appear to be less cooperative.

In order to further illustrate the hydrogen bonding cooperativity in the helical chains of polyglycines I plotted the H-bonding distances (at the AM1 theory level) for the four conformers as a function of the chain length (Figure 3.4). Since the length of the hydrogen bond is generally considered to be an indication of its strength, the changes in

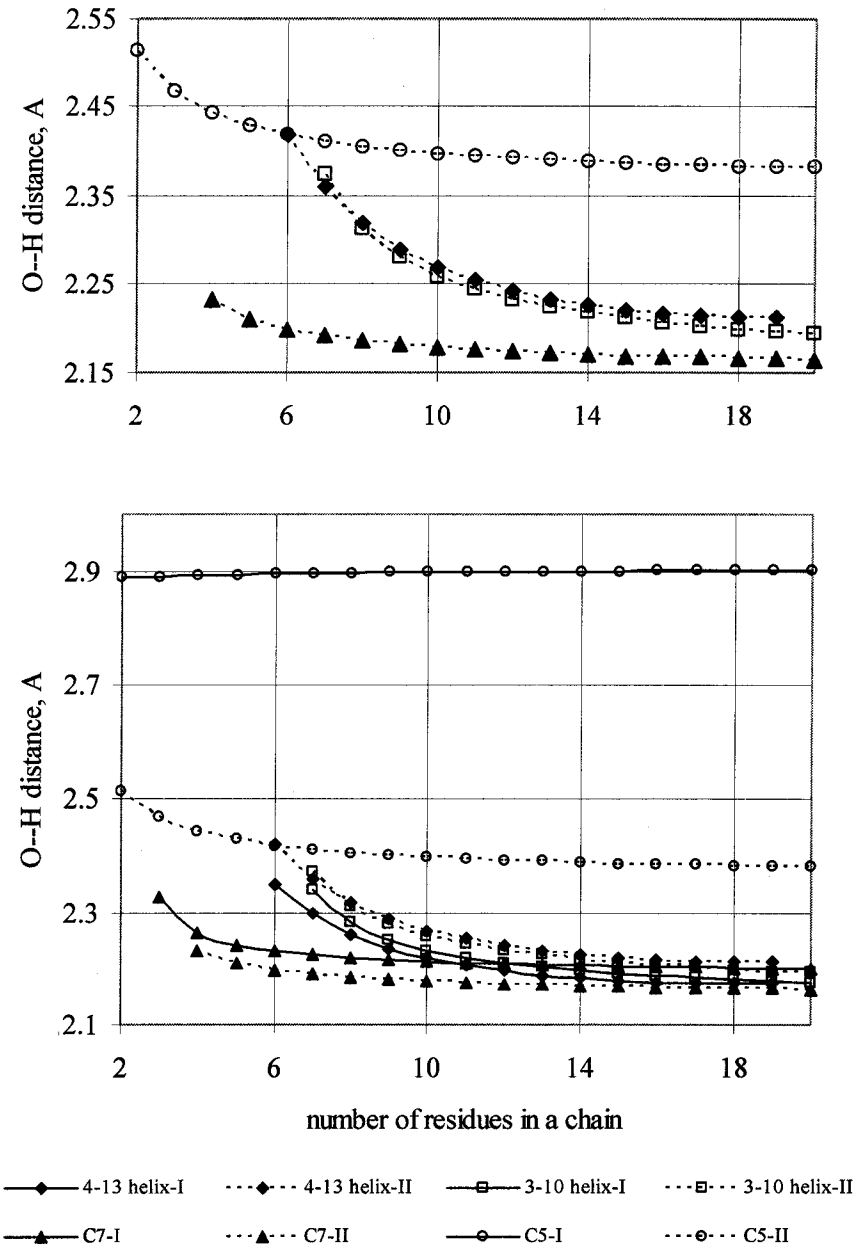
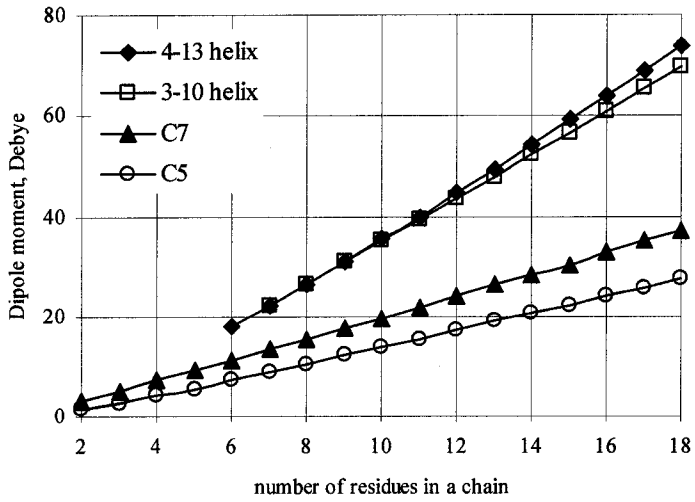


Figure 3.4. H-bonding distances for the conformers of polyglycines of up to 20 residues at the AM1 level. Due to constant used all the H-bonds (II, dashed lines) in each polyglycine but the first one from the N-terminus (I, solid lines) are the same. The H-bonds I in C5 are between the C=O and the two hydrogens of the NH_2 group. These H-bonds stay almost constant at 2.9 Å independent of the chain length.

the hydrogen bonding distances with the change of the size of the system can serve as an indication of the hydrogen bonding cooperativity. Notice that because of the constraint used in this study (that all of the residues within each conformer be identical in their geometry) all the H-bonds in each polyglycine in Figure 3.4 are the same (type II), except for the first from the N-terminus (type I). As it was expected, for all conformers the H-bonding distances become shorter as the length of polypeptide chains increases. Again one can see that O...H distances in the two helices decrease more rapidly than in C5 and C7 conformers. This results from the cooperative character of hydrogen bonding in the helical polyglycines. The differences in the type II H-bonding distances for the conformers with 2 and 14 H-bonds are 0.205 Å (8.5%), 0.177 Å (7.6%), 0.065 Å (3.0%), and 0.125 Å (5.0%) for the α, 3_{10}, C7 and C5 conformers respectively. Of the two helices 3_{10} has longer first hydrogen bond, but the rest of the H-bonds in it are slightly shorter than in the α-helix in the chains longer than 8 glycines (by around 0.012 Å). Improta and coworkers also reported the 3_{10}-helix to have shorter H-bonding distances than the α-helix for the infinite polyglycines (the PBC approach with LSDA, PBE, VSXC density functionals in combination with the 6-31G* basis set)[11] and for the infinite polyalanine (at the PBE/6-31G* level by the PBC method).[5] However in the paper by Wieczorek and Dannenberg[15] at the B3LYP/D95** level the α-helical polyalanines have shorter average and minimum O...H distances than the 3_{10}-helical one for the chains of more than 17 and 11 alanines respectively.

Finally, Figure 3.5 represents the dipole moments and the dipole moments per residue as a function of the chain length for the four polyglycine conformers of polyglycine. One can see that the total dipoles for both the helical structures grow much

A



B

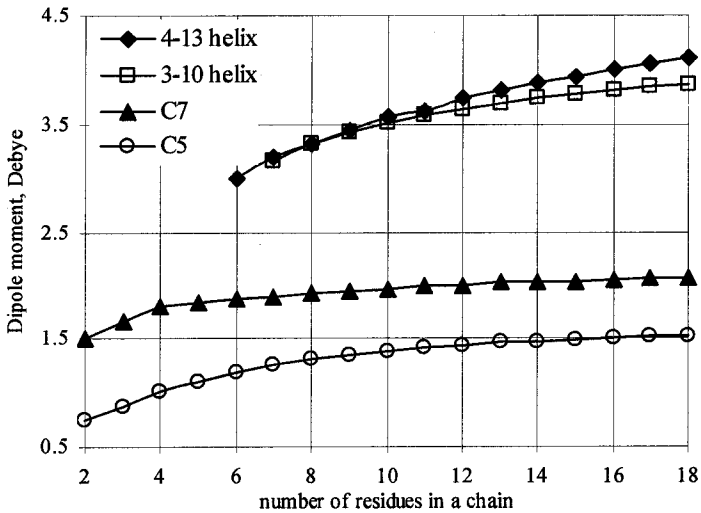


Figure 3.5. Total dipole moments (A) and total dipole moments per residue (B) for the conformers of polyglycine containing up to 18 residues.

more rapidly than for C5 and C7. This indicates that the helical structures have more favorable positioning of the dipoles of each peptide unit. Similar to the interaction energies per residue, the dipole moments per residue for the C5 and C7 conformers reach their asymptotic values when the chains are around 16 residues long (about 1.5 D and 2 D respectively). The general trends in Figure 3.5B are again in a good agreement with the data reported by Wu and Zhao (Ref. 7, Figure 4). However, in my study the dipole moments for the C7 and C5 conformers are much lower and the relative increase in dipole moments with the chain length for 3_{10} and α-helical conformations are more pronounced. When the chain length increases from 7 to 18 glycines, the dipole moment per residue increases by ~29% in the α-helix, and by 22.5% in the 3_{10}-helix, indicating that there is more cooperativity in the former. Much weaker increase for the dipole moments reported by Wu and Zhao are likely due to frozen repeating unit model they used. As a result, part of the cooperativity due to geometric relaxation is lost.

3.4. Conclusions

In this model study at the B3LYP/D95**//AM1 level of theory the relative stabilization for the longer polyglycine chains (up to 18 residues) was found to be:

3_{10} helix > α-helix > C7 > C5

The two helical conformers are predicted to be clearly more stable than the C5 and C7 conformers due to the cooperativity of the hydrogen bonding. Prediction of the relative stability for 3_{10} and α-helices for longer chains is more delicate. My results indicate that slope of stabilization curves, as well as the trend in H-bond lengths both predict 3_{10}-

helices to remain the most stable. However, the trends in per residue dipole moments demonstrate faster increase for the α-helices.

Some researchers believe that the long-distance electrostatic interactions may favor the α-helical conformation which has stronger dipole moments,[5,6,7] making it more stable than the 3_{10}-helix for long chains. However, other studies have shown[15] that dipole-dipole interaction on itself can not explain the extraordinary degree of cooperativity in H-bonding chains, but may serve as its indicator.

In my opinion, the reason why the 3_{10}-helix seems to be more stable is that the α-helix does not reach its asymptotic value even at 18 residues. Two longer H-bonding chains in the 3_{10}-helix are more stable than three shorter chains in the α-helix, where the terminal effects are still important. Thus, the studies of the longer chains are needed for the final conclusion. I can also point out that though the α-helix is considered to be the most common conformation for the long helical polypeptides the glycine residue is not among the residues that favor the formation of the α-helices.[8b]

My results demonstrate that even at the relatively low level of theory that was used in this work one can acquire reasonable stabilization trends for conformers of the polypeptides. This study may be helpful in the modeling of the protein folding.

4.
*COMPARISON BETWEEN VIBRATIONAL COUPLING THROUH HYDROGEN BONDS AND COVALENT BONDS. IMPLICATIONS FOR PEPTIDE VIBRATIONAL SPECTRA**

4.1. Introduction

Vibrational spectra have been extensively used to characterize the secondary structure of proteins. There are few major modes that are important for the spectroscopy of peptides and proteins: (i) the amide A is associated with the N-H stretch; (ii) the amide I is primarily the C=O stretch; and (iii) the amide II is a combination of the N-H stretch and the C-N bend. In particular, the coupling of the frequencies and intensities of the amide I infrared absorption has been used as an indication of the nature of a secondary structure.[1-5]

---

* The results reported in this Chapter are published in Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6688. This work was supported (in part) by grants from the National Institutes of Health (S06GM60654) and PSC-CUNY.

[1] Yoder, G.; Pancoska, P.; Keiderling, T. A. *Biochemistry* **1997**, *36*, 15123.

[2] Krimm, S.; Reisdorf, W. C., Jr. *Faraday Discuss.* **1995**, *Volume Date 1994*, *99*, 181.

Both the frequency shift and the intensity of the absorption are sensitive to the coupling of the carbonyl stretches in the polypeptide chain. These couplings result from the interaction between the amide carbonyls either through covalent bonds, through space (via classical dipole-dipole interactions), hydrogen bonds or some combination of these. Spectroscopists often use a model that assumes the through bond or through space dipole-dipole coupling to be dominant,[6] although some recent reports have found some fault with this model.[7,8] However, we have demonstrated the hydrogen bond cooperativity in amide H-bonding chains to be unusually extensive[9,10] (see Chapter 2). The central hydrogen bond in a chain of 15 H-bonding formamides is approximately 2.9 times as strong as that of a dimer. Coupling of the amide vibrations of H-bonding amides in solution has been reported by Fillaux over 30 years ago.[11]

In this chapter, I examine the separate effects of several vibrational couplings through the covalent bonds and through the hydrogen bonds. To do this, I shall compare the carbonyl stretching frequencies in single β-strands of oligoglycines of varying sizes with chains of hydrogen-bonding formamides containing equivalent numbers of interacting

---

[3] Yoder, G.; Keiderling, T. A.; Formaggio, F.; Crisma, M.; Toniolo, C. *Biopolymers* **1995**, *35*, 103.

[4] Manas, E. S.; Getahun, Z.; Wright, W. W.; GeGrado, W. F.; Vanderkooi, F. M. *J. Am. Chem. Soc.* **2000**, *122*, 9883.

[5] Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628.

[6] (a) Chirgadze, Y. N.; Nevskaya, N. A. *Dokl. Akad. Nauk SSSR* **1973**, *208*, 447. (b) Chirgadze, Y. N.; Nevskaya, N. A. *Biopolymers* **1976**, *15*, 627.

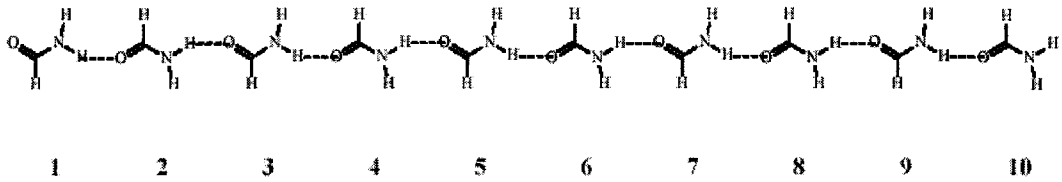
[7] Cha, S.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 740.

[8] Rubtsov, I. V.; Hochstrasser, R. M. *J. Phys. Chem. B* **2002**, *106*, 9165.

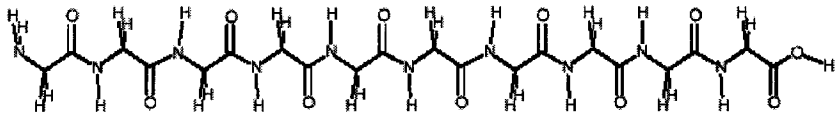
[9] Kobko, N.; Dannenberg, J. J. *J. Phys. Chem.A* **2003**, *107*, 10389.

[10] Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348.

[11] Fillaux, F.; De Loze, C. *J. Chim. Phys. Physicochim. Biol.* **1972**, *69*, 36.



A



B

Figure 4.1. Chains of formamides with the numbering convention used in the text (A) and extended single β-strand (B).

carbonyl groups (Figure 4.1) using density functional theory (DFT) molecular orbital (MO) calculations. The extended single β-strands can only form weak, cyclic H-bonds, which are insulated from cooperativity through the π-system by a CH_2 group. Thus, all coupling between the carbonyl groups must come through the covalent bonds or through space. On the other hand, the H-bonding chains have no covalent linkages between the formamide molecules. Thus, all coupling between these carbonyl groups must come through the hydrogen bonds or through space. In this study I present results for the coupled C=O (amide I), H-bonding N-H stretching (amide A) and H-C stretching/N-H bending (amide II) vibrations. Ab initio[12-18] and DFT[19] studies on the coupling of these vibrations have previously been restricted to relatively small oligopeptides.

Isotopic substitutions by individual heavy nuclei of H, C, N, and O provide added insights into the interpretation of the vibrational spectra by partially or (almost) completely attenuating the vibrational coupling.

---

[12] Mehler, E. L. *J. Am. Chem. Soc.* **1980**, *102*, 4051.

[13] Bour, P.; Keiderling, T. A. *J. Am. Chem. Soc.* **1993**, *115*, 9602.

[14] Aleman, C.; Navas, J. J.; Munoz-Guerra, S. *J. Phys. Chem.* **1995**, *99*, 17653.

[15] Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283.

[16] Torii, H.; Tatsumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309.

[17] Torii, H.; Tasumi, M. *AIP Conf. Proc.* **1998**, *430*, 719.

[18] Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142.

[19] Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2000**, *53*, 380.

4.2. Computational Details

Molecular orbital calculations were performed using a hybrid DFT method at the B3LYP/D95** level (see Ref.26 and 27, Chapter 1) using the Gaussian 98 suite of programs (see Ref. 28, Chapter 1). The geometries were completely optimized with the sole constraint that all the structures were of C_s symmetry (all atoms are coplanar except for the hydrogens of the terminal NH_2 and of CH_2 groups in the polyglycines, which are symmetrically placed above and below the plane). For these calculations I used the cluster of Intel and AMD powered computers that are parallelized using LINDA (see Ref.29, Chapter 1). The number of nodes used for each calculation varied with the sizes of the systems studied. The vibrational frequencies were calculated using the normal harmonic approximations employed in the Gaussian 98 program. All frequencies were real except for some very low imaginary frequencies that involved out of plane twists between pairs of formamides in some of the longer formamide chains and inversion of the terminal groups in polyglycines. All these imaginary frequencies were less than 21 cm^{-1} and 64 cm^{-1} respectively. The frequencies reported here have not been modified by any scaling factor. B3LYP frequency calculations are usually scaled by a factor of 0.96. When scaled by this factor, the calculated frequencies for the formamide monomer are in good agreement with the experimental IR spectra in the gas phase.[20]

---

[20] Evans, J. C. *J. Chem. Phys.* **1954**, *22*, 1228.

4.3. Results and Discussion

I shall first present and discuss the results for the H-bonding chains and the β-strand separately and then compare them.

4.3.1. H-Bonding Chains

The results for the hydrogen bonding formamide chains are gathered in Tables 4.1-4.4 and Figures 4.2-4.10.

The most significant vibrations for the H-bonding chains are the carbonyl stretch and the N-H stretch that is associated with the H-bond (see Table 4.1). We have previously noted that the H-bonding interactions near the center of the formamide chains increase dramatically in strength as the H-bonding chains increase in length[9,10] (see also Chapter 2). This behavior is reflected in the N-H stretching frequencies associated with the H-bonds. There are one fewer H-bonds than formamide units in the H-bonding chain. The NH_2 group at one end of the chain will not H-bond. The N-H stretching frequency for the hydrogen bond in the dimer is calculated to be 3549 cm^{-1} (see Table 4.1). As the chain becomes longer the lowest frequency vibration decreases markedly to 3337 cm^{-1} in the chain of ten formamides, a red shift of 212 cm^{-1} from that of the dimer and 269 cm^{-1} from the monomer. This lowest vibration is composed of a linear combination of N-H stretches that are all in phase. As might be expected from a positive combination of all, this vibration has the highest intensity of the N-H stretches (see Figure 4.2 and Table 4.1). It grows in intensity as the chain lengthens from 421 km/mol for the dimer to 7150 km/mol for the decamer. The

Table 4.1: Vibrational Frequencies for Formamide Monomer (1FA), Dimer (2FA) and Decamer (10FA)

	N-H stretch Frequency cm-1	Intensity km/mol	C=O stretch Frequency cm-1	Intensity km/mol	C-N stretch/ N-H bend Frequency cm-1	Intensity km/mol
1FA	3608	41	1813	406	1608	64
2FA	3549	418	1798	824	1614	79
	3604	52	1812	202	1628	33
10FA	3337	7150	1757	5747	1615	68
	3356	24	1765	0	1634	64
	3369	1207	1773	799	1635	73
	3381	16	1779	4	1639	29
	3384	457	1784	307	1640	142
	3410	652	1787	41	1642	31
	3416	695	1790	219	1643	4
	3488	591	1791	92	1644	4
	3495	556	1802	210	1645	13
	3603	59	1803	251	1645	10

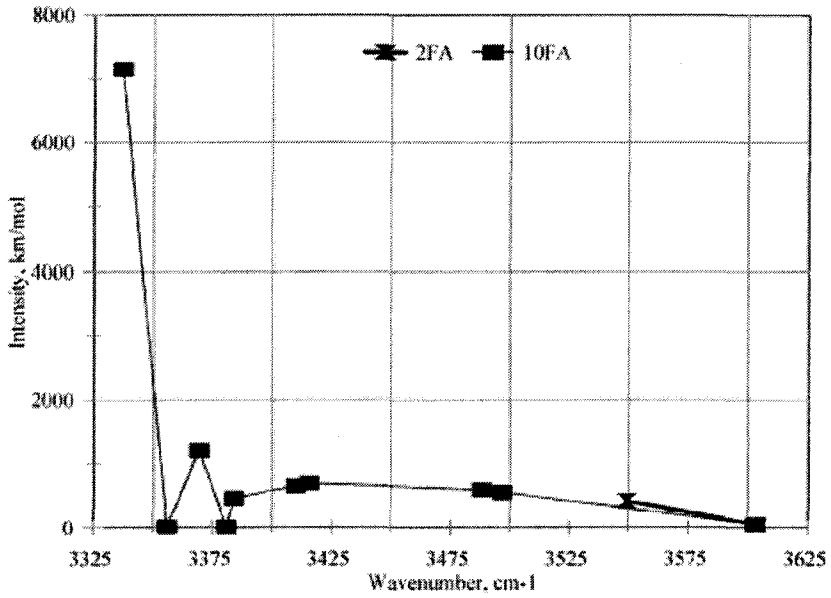


Figure 4.2. Variation of frequencies and intensities for the H-bonding N-H stretches in formamide chains.

second lowest frequency in this group has the half of the N-H stretches of one end of the chain in-phase with each other but out of-phase with those at the other side. Thus, its intensity should be quite small, but not zero, as there is no symmetry element that relates one end of the chain to the other. These two vibrations clearly can be thought of as the lowest two wavefunctions of the coupled N-H stretches. The amplitudes of the N-H stretches for each formamide in a chain of ten were approximated by the changes of the individual N-H bonds as determined from the relative displacements of the N and H atoms. These are shown in Figure 4.3 for each of the coupled vibrations. The lowest has no nodes, while the second has one node. One might expect this pattern to continue for all the linear combinations. However, this is not the case. The central H-bonds become increasingly stronger than the terminal ones as the chains become longer. As a result, the coupling between the central and terminal N-H stretches decreases with increasing chain length. Thus, the higher frequency N-H stretches become localized at each end of the longer chains (see Figure 4.3).

The C=O stretching frequencies behave in a manner qualitatively similar to those of the N-Hs. However, since the H-bond strength affects the C=O less than the N-H, the effects are somewhat attenuated compared to those of the N-H stretches. For example, the lowest frequency C=O stretch decreases by only 41 cm^{-1} from 1798 to 1757 upon going from the dimer (and 55 cm^{-1} from the monomer) to the decamer chain while the corresponding N-H stretch decreases by 212 cm^{-1} (see Table 4.1, Figure 4.4). The intensity pattern for the C=Os (see Figure 4.5) is similar to that for the N-Hs. Once again, the linear combinations of amplitudes of the individual C=O stretches for the lowest frequencies

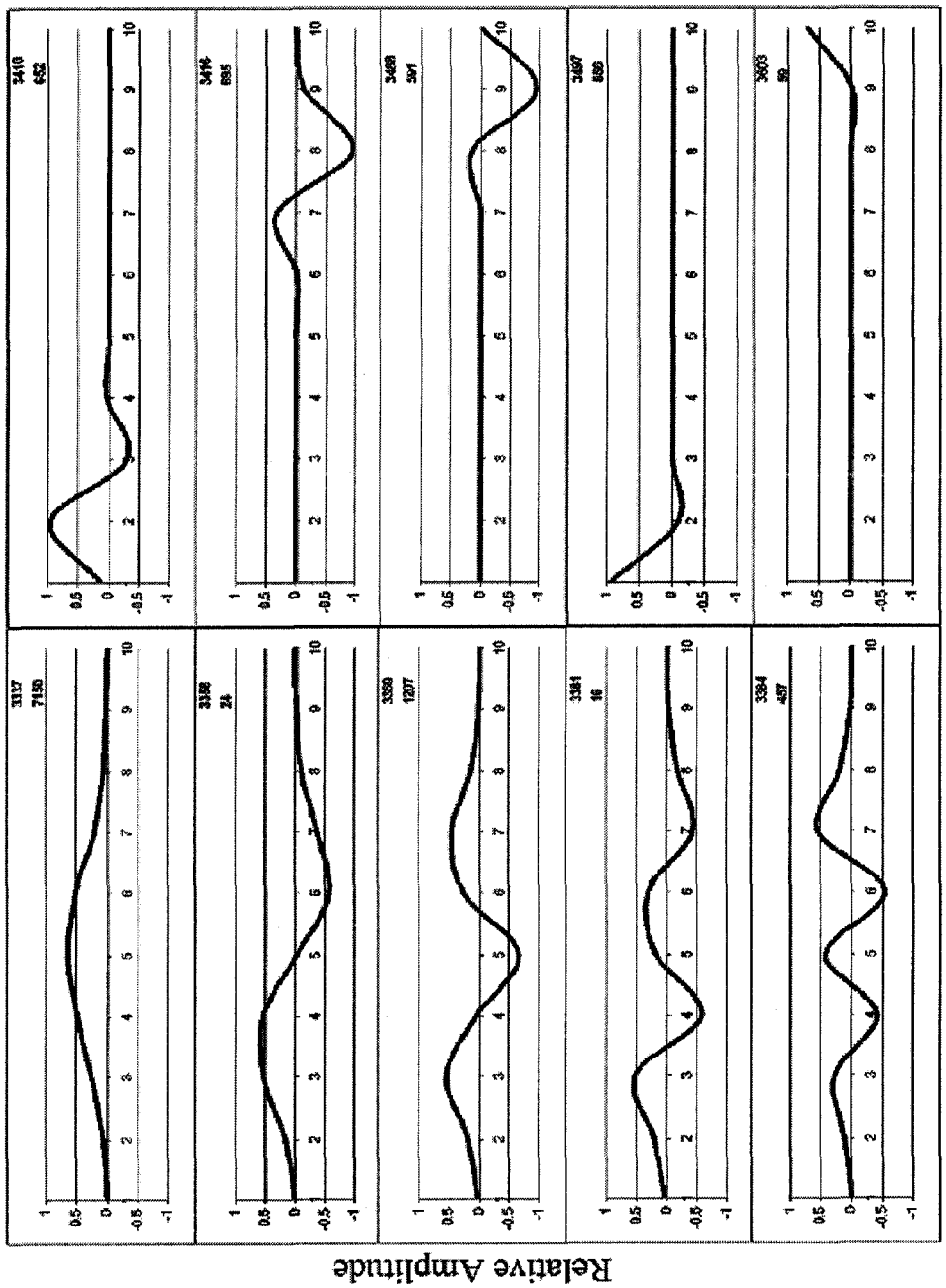


Figure 4.3. Amplitudes for each formamide in the coupled N-H stretches. See text for explanation; see Figure 4.1 for numbering.

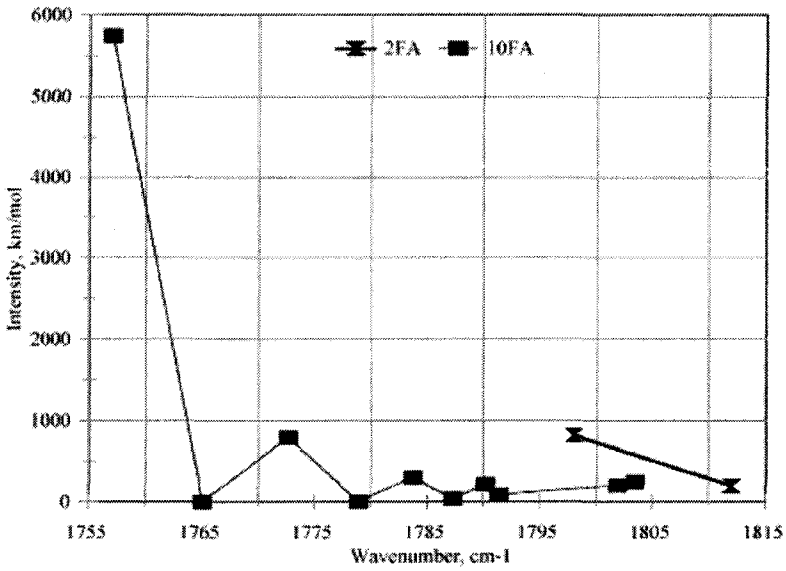


Figure 4.4. Variation of frequencies and intensities for the C=O (amide I) stretches in formamide chains.

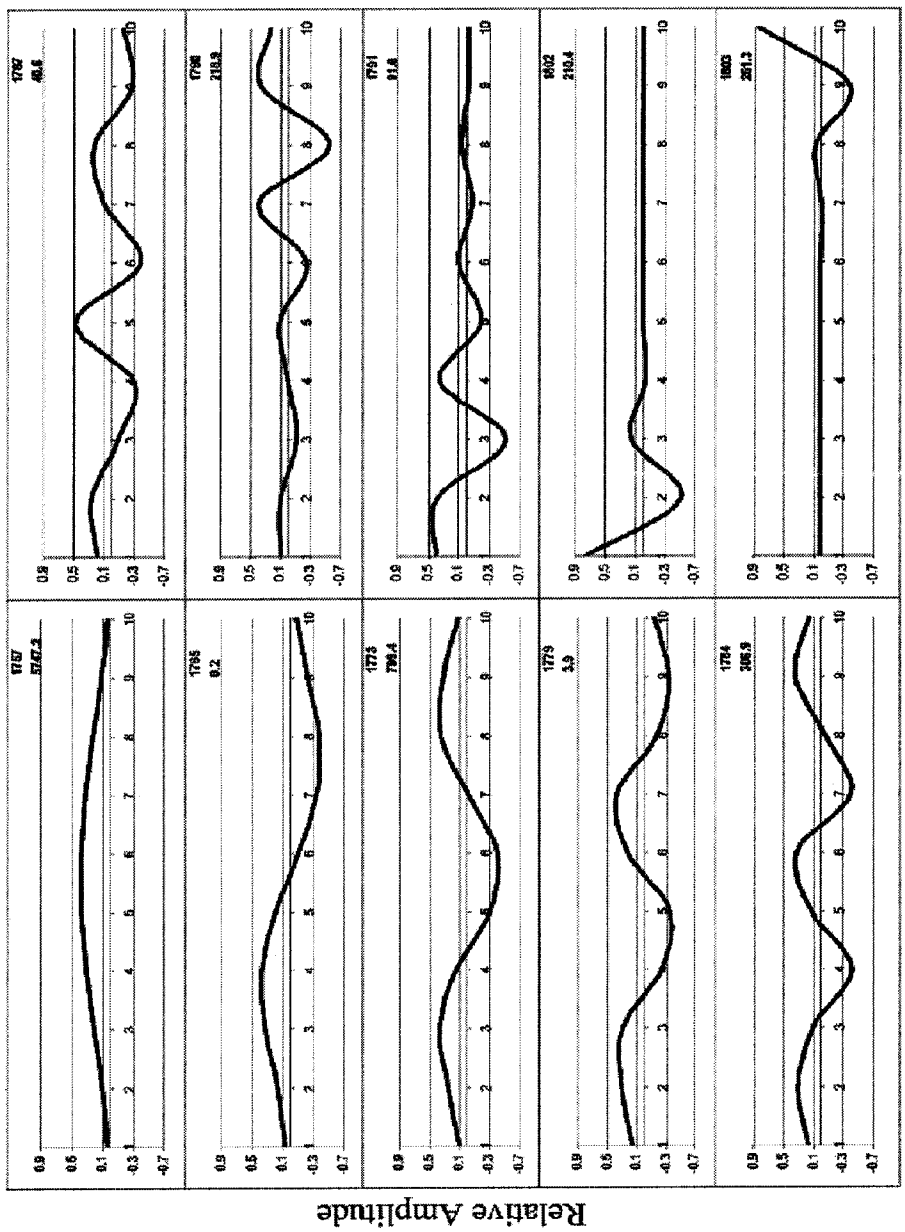


Figure 4.5. Amplitudes for each formamide in the coupled C=O stretches (amide I). See text for explanation; see Figure 4.1 for numbering.

resemble the lowest wavefunction for the coupled C=Os. Like the N-H stretches, in long chains, the higher frequency C=Os tend to be localized at the ends (Figure 4.5).

The intensities of the lowest frequency C=O stretches for the short chains are larger than those for the N-H stretches. However, the intensities of the N-H stretches increase more with increasing chain length than do the C=Os (see Table 4.1). Chains with six or more formamides have more intense low frequency, delocalized, N-H than C=O stretches.

The formamide chains also have deformations that resemble the amide II deformations of peptides. They involve the C-N bond stretches and N-H bond bends. These vibrations also couple, but to a lesser extent than the N-H and C=O stretches. In the decamer, the deformations involved in the nine H-bonds vary from 1634 to1645 cm^{-1} compared with 1628 cm^{-1} for the dimer. The intensities of these vibrations, which are much less than those of the N-H and C=O stretches, do not change substantially with increasing chain length (see Figure 4.6). In contrast to the C=O and N-H stretches, the frequencies with amplitudes of the individual amides concentrated near the center of the decamer are more blue shifted than those near the ends (the individual amplitudes are taken from the C-N displacements) (see Figure 4.7). Likewise, the lowest frequencies of this coupled set are more localized near the ends of the formamide chains. Unlike the N–H and C=O coupled stretches, the terminal deformation is largely unperturbed upon going from the dimer to the decamer as it is blue shifted by only 1 cm^{-1}.

I noted before that the N-H and C=O stretches undergo substantial red shifts (in the case of N-H very substantial). In fact, all the coupled transitions for both of these

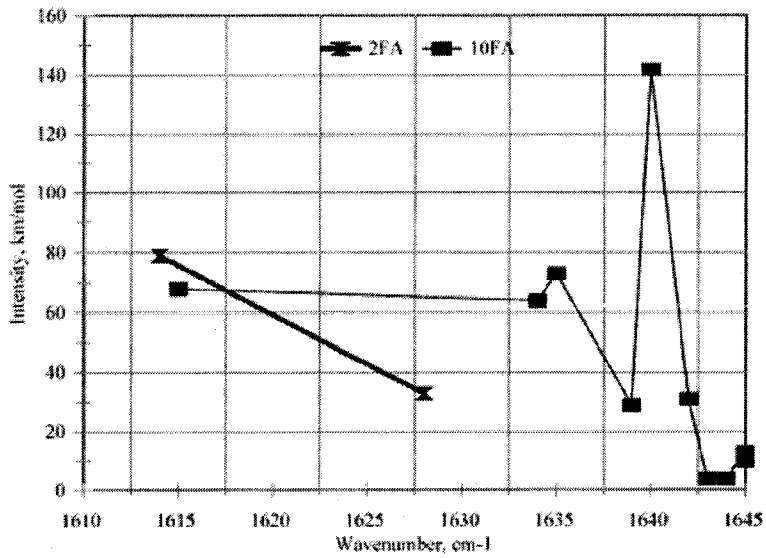


Figure 4.6. Variation of frequencies and intensities for the C-N stretch/N-H bend (amide II) vibration in formamide chains.

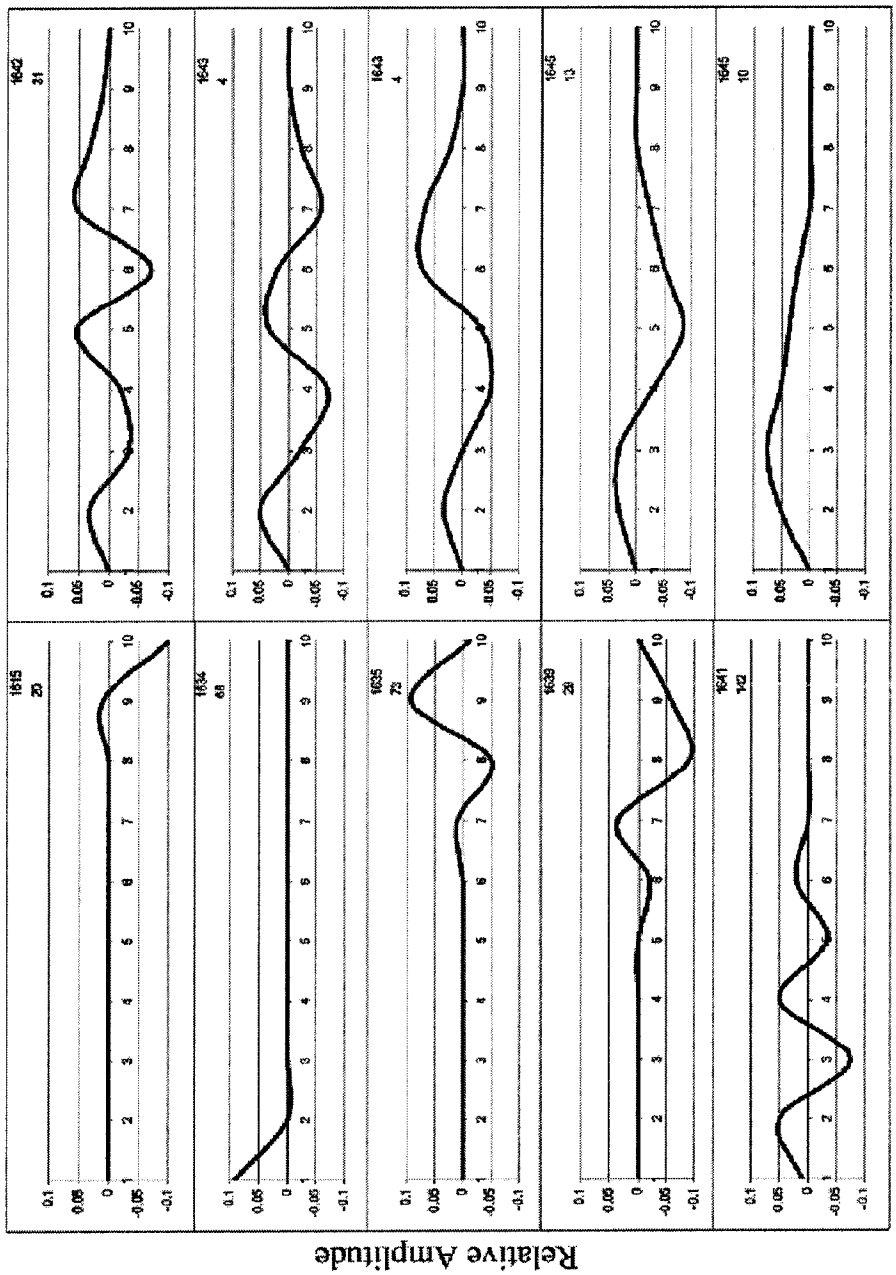


Figure 4.7. Amplitudes for each formamide in the coupled C-N stretch/N-H bend (amide II). See text for explanation; see Figure 4.1 for numbering.

modes are red-shifted from the corresponding modes in monomer and dimer. On the other hand, the modes that resemble the amide II vibration undergo blue shifts as the chain length increases. This reflects the cooperative stabilization of the H-bonds which causes both the N-H and C=O bonds to lengthen and the C-N bond to shorten with increasing chain length (see Table 4.2). These effects are larger near the center then near the ends of the chains. This observation reflects the polarization of the π-system of the formamides that accompanies the cooperative interactions of the H-bonds. As the C=O bonds lengthen they lose some of their π-character, while the C-N bonds contract as they acquire more π-character. I can also point out that similar effects of the H-bonding cooperativity on the geometry and vibrational spectra have been reported for the H-bonded clusters of up to five N-methylacetamides.[21]

4.3.1.1 Isotopic Substitutions

I explored the effects of isotopic substitutions of individual nuclei in the various chains of H-bonding formamides. Several studies have recently appeared describing the effect of ^{13}C substitution at the C=O upon the amide I absorptions of peptides and proteins.[16, 18, 19, 22-28] For purpose of illustration, I shall use the data for the decamer. I considered substitution of ^{14}C, ^{15}N, ^{18}O, and ^{2}H (deuterium) individually at each possible

---

[21] Ludwit, R.; Reis, O.; Winter, R. *J. Phys. Chem. B* **1998**, *102*, 9312.

[22] Eberhardt, E. S.; Raines, R. T. *J. Am. Chem. Soc.* **1994**, *116*, 2149.

[23] Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T. *Biopolymers* **2001**, *59*, 396.

[24] Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727.

[25] Brauner, J. W.; Dugan, C.; Mendelsohn, R. *J. Am. Chem. Soc.* **2000**, *122*, 677.

Table 4.2: Variation of Bond Lengths (Å) for Formamide Monomer (1FA), Dimer (2 FA), and Decamer (10FA). The Bond Lengths for the Dimer for Those Most Involved in the H-Bond, for the Decamer, for Those Most Involved in the Central H-Bond

	1FA	2FA	10FA, center
C=O	1.222	1.225	1.239
N-H	1.010	1.016	1.028
C-N	1.368	1.360	1.344
O...H	--	1.923	1.782

position in the decamer (only the H-bonding N-Hs were substituted by deuterium). The ^{14}C substitution reduces the C=O frequency of monomer formamide by 81 cm^{-1} from 1813 to 1732 cm^{-1} (see Table 4.3). The H-bonding within the decamer chain effectively changes the environments of the individual C=Os. However, these changes upon their respective native C=O stretching frequencies are masked by the extensive coupling between them. ^{14}C substitutions effectively decouple the ^{14}C=O stretches from the corresponding ^{12}C=O values due to the isotopic shifts. Comparison of the ^{14}C=O stretching vibrations as a function of the position of isotopic substitution in the decamer allows to assess the effects of H-bonding and H-bonding cooperativity upon the C=O stretches of individual carbonyls. Figure 4.8 displays the shift in ^{14}C=O and ^{13}C=O frequencies from the monomer as a function of the ^{14}C and ^{13}C positions in the decamer (see also Table 4.4). The frequency shifts are largest when the isotopic substitution is near the middle of the H-bonded chain. I have already noted that the C=O bonds lengthen more for those carbonyls near the center of the H-bonding chain. The ^{14}C=O and ^{13}C=O stretching frequencies at different positions in the decamer correlate linearly with the corresponding calculated C=O bond lengths (see Figure 4.9).

---

[26] Fang, C.; Wang, J.; Kim, Y. S.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B., III; Decatur, S. M.; Hochstrasser, R. M. *J. Phys. Chem. B* **2004**, *108*, 10415.

[27] Barber-Armstrong, W.; Donaldson, T.; Wijesooriya, H.; Silva, R. A. G. D.; Decatur, S. M. *J. Am. Chem. Soc.* **2004**, *126*, 2339.

[28] Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R. A. G. D.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.

Table 4.3: Comparison of Vibrational Frequencies and Absorption Intensities for Formamide with Different Isotopic Substitutions

	N-H stretch		C=O stretch		C-N stretch N-H bend	
	Frequency cm^{-1}	Intensity Km/mol	Frequency cm^{-1}	Intensity Km/mol	Frequency cm^{-1}	Intensity km/mol
Normal	3608	41	1813	406	1608	64
^{13}C	3608	40	1770	388	1607	56
^{14}C	3608	40	1732	375	1605	48
^{2}H	2695	37	1811	399	1469	30
^{15}N	3603	39	1812	400	1600	61
^{18}O	3608	41	1785	401	1606	55

Table 4.4: Frequencies of the Highest Intensity Absorptions for the Coupled C=O Vibrations (cm^{-1})

Position Of isotope	Highest intensity ^{14}C	Highest intensity ^{13}C
None	5747	5747
1	5480	4368
2	4982	4270
3	4440	3906
4	3945	3360
5	3782	2865
6	4199	2919
7	3781	3122
8	4224	3694
9	4772	4186
10	5263	3421[a]

[a]This vibration is coupled with another.

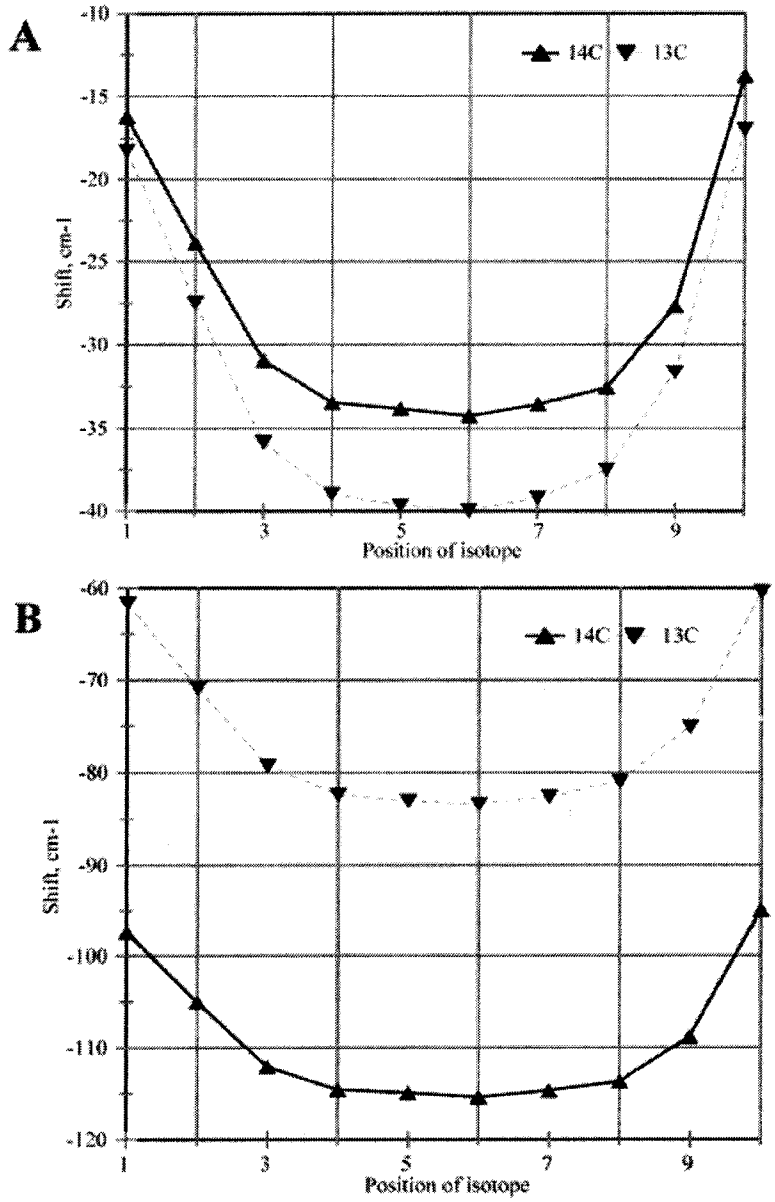


Figure 4.8. Frequency shift for ^{13}C and ^{14}C from isotopically substituted (A) and unsubstituted (B) formamide for monosubstituted decamers with the same isotopic substitute at indicated formamide. See Figure 4.1 for numbering.

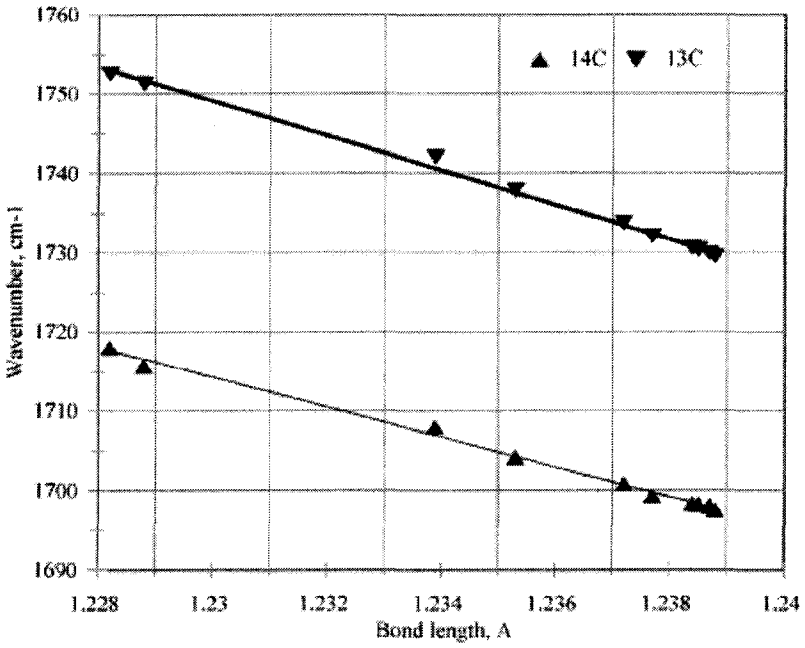


Figure 4.9. Correlation between C=O bond length and stretching frequencies for formamide decamers individually ^{13}C and ^{14}C substituted at various positions.

The ^{14}C=O also acts as a barrier to coupling of the other C=Os. Their couplings are effectively constrained to the C=Os on either side of the ^{14}C=O as illustrated for the decameric chain with ^{14}C=O in the fourth formamide (see Figure 4.10). Coupling through the ^{14}C=O is minimal. (Compare Figure 4.10 with the respective frequencies of the normal decamer in Figure 4.5). Furthermore, intensities decrease upon ^{14}C-substitution. These decreases are greatest when the substitution is near the center of the H-bonding chain as exemplified by the data for ^{14}C-substitution at position 5 (Figure 4.11). Since the coupled vibrations are essentially constrained to one side or the other of the ^{14}C-substituted amide (as noted above, see Figure 4.10) the most delocalized of these vibrations involve fewer C=Os when the substitution is near the center.

Isotopic substitution by ^{18}O causes a smaller shift (-28 cm^{-1}) in monomeric formamide than does ^{14}C (see Table 4.3). Because of this, substituting individual C=^{18}Os in the decamer does not decouple the C=O stretches to the same extent as ^{14}C (Figure 4.12). In fact, when C=^{18}O is substituted in formamide number 2, the shift caused by the isotope moves its 'natural' frequency closer to those C=Os near the middle of the H-bonding chain. Consequently, the coupling increases between this C=O and the more central C=Os (Figure 4.13).

Substitution of the H-bonding N-H by D (^{2}H) causes the lowest N-H stretch (now a localized N-D stretch) to be substantially red shifted by 760-860 cm^{-1} depending upon the position of substitution. Once again, the greatest shifts are observed when D is substituted at the middle of the chain. Unlike the ^{14}C-substitutions, the D substitutions have little effect upon the coupling of the N-H and C=O vibrations on either side of the substitution.

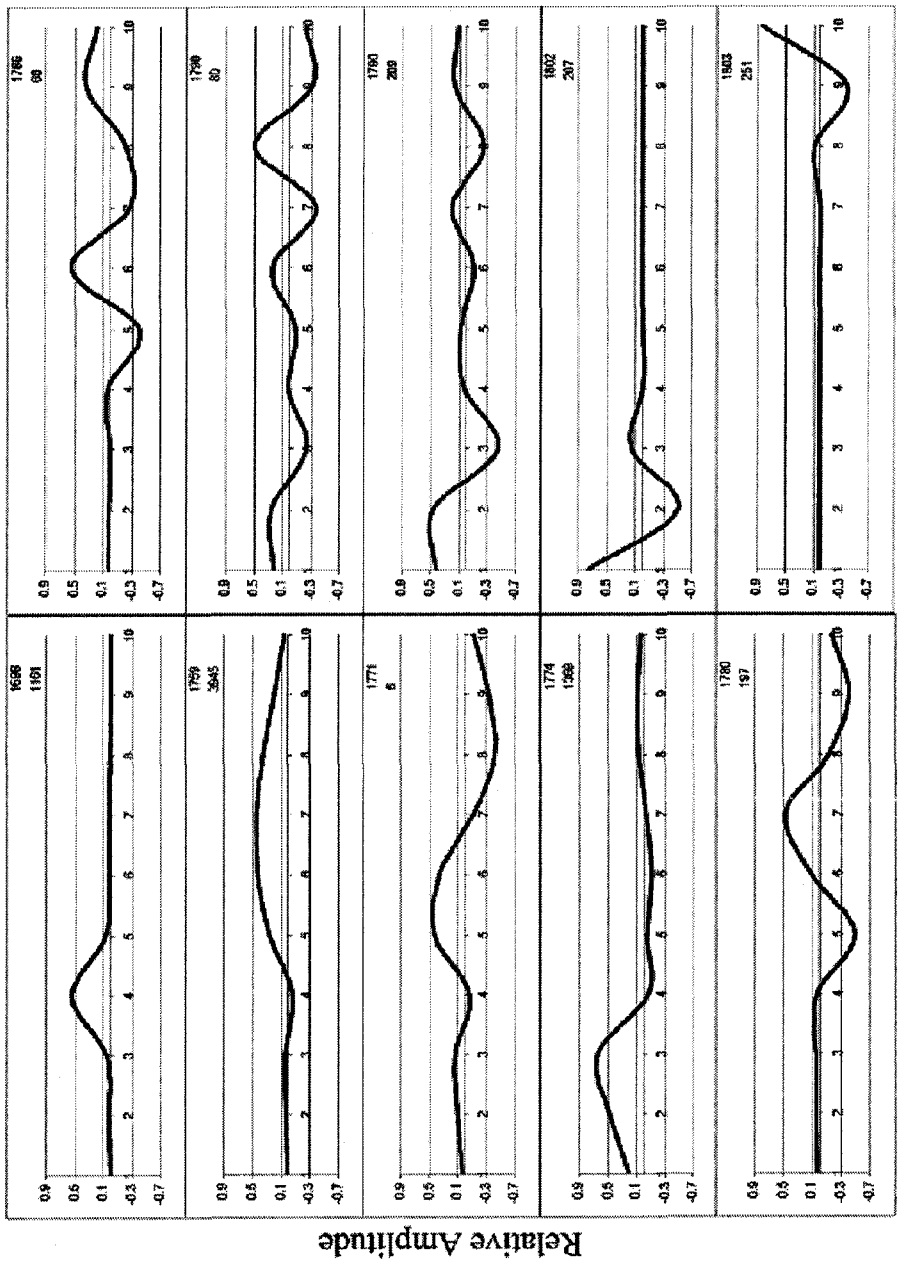


Figure 4.10. Amplitudes for each formamide in the coupled C=O stretches (amide I) with ^{14}C in the fourth formamide. See text for explanation; see Figure 4.1 for numbering.

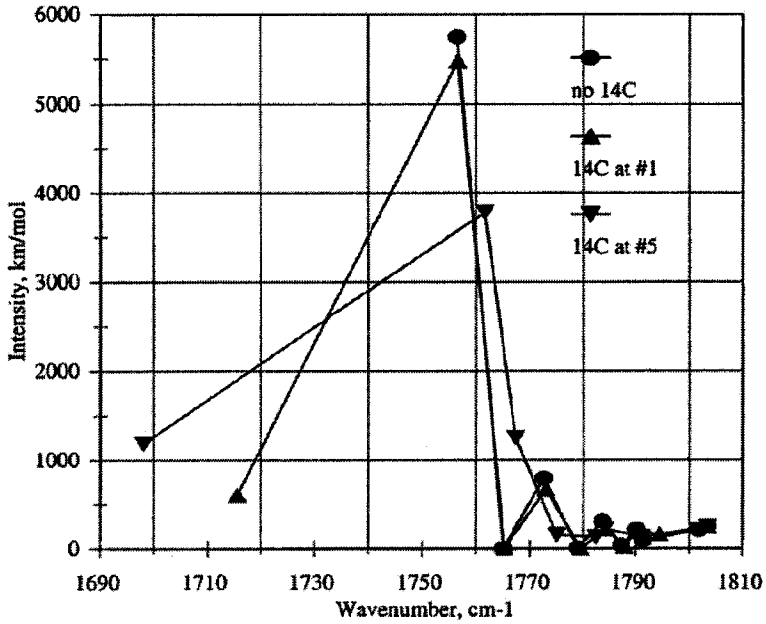


Figure 4.11. Variation of frequencies and intensities for the C=O (amide I) stretches in formamide chains upon ^{14}C substitution.

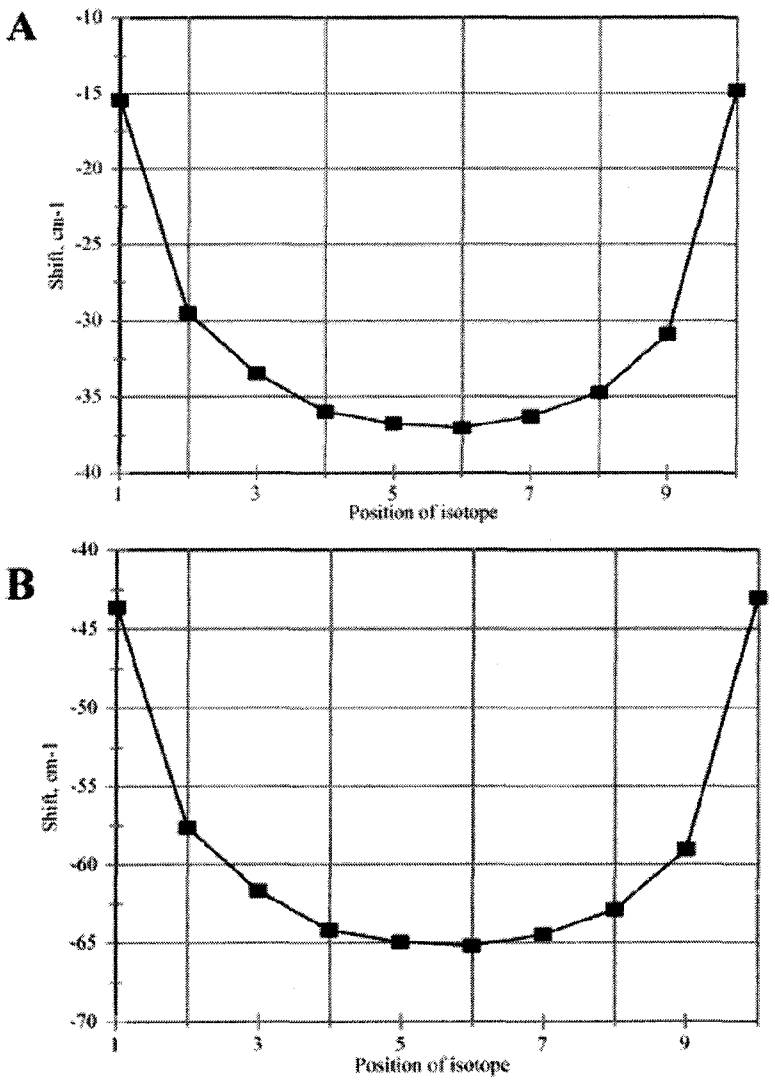


Figure 4.12. Frequency shift for C=^{18}O from isotopically substituted (A) and unsubstituted (B) formamide for monosubstituted decamers with same isotopic substitution at indicated formamide. See Figure 4.1 for numbering.

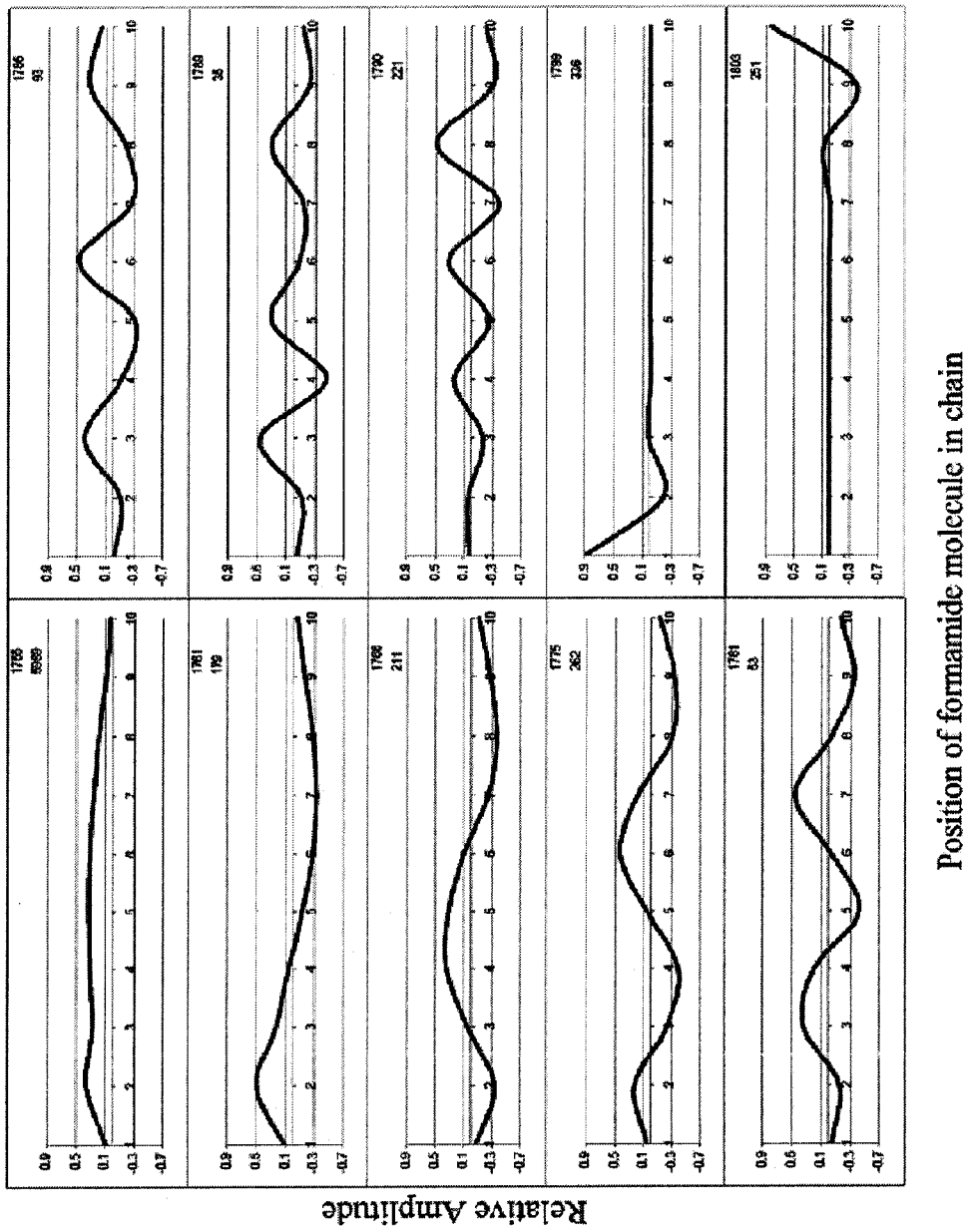


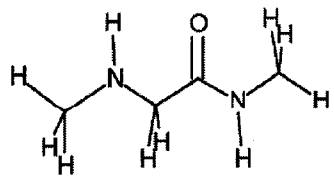
Figure 4.13. Amplitudes for each formamide in the coupled C=O stretches (amide I) with ^{18}O substitution in second formamide. See text for explanation; see Figure 4.1 for numbering.

The frequencies of the other N-Hs (non-H-binding) and the C=Os are not appreciably affected by D-substitution. For example, the lowest C=O stretch varies less than 1 cm^{-1} while the second lowest N-H stretch (lowest after the N-D) varies by only 14 cm^{-1} upon substitution of D for any H-bonding H.

4.3.2 Extended β-Strands

The results for the β-strands are collected in Table 4.5 and Figures 4.14-6.

In contrast to the formamide chains, the carboxyl C=O (amide I) stretches remain little changed in oligoglycines containing from two to ten glycine residues. The amide C=O stretching modes do not extensively couple with the higher frequency C=O stretch of the carboxyl group. The coupling of the other C=O stretches differs *somewhat* from those of the H-bonding chains. Unlike the H-bonding chains (where all the C=Os are aligned in approximately the same direction as the long axis of the chain), the C=Os of the β-strand are roughly perpendicular to the long axis of the strand and oriented in the opposite direction to that of its nearest neighbors (see Figure 4.1). Since there are no cooperative H-bonds in the β-strand, all coupling must come via a through-bond (or through space) mechanism involving the amide linkages. To model a single C=O in β-strand, I used α-methylamino,-methylacetamide, (CH3-NH-CH2-**C**=**O**-NH-CH3), **I**,



I

Table 4.5: Vibrational Frequencies and Absorption Intensities for the Model Compound (**I**), the Glycine (1Gly), the Dipeptide (2Gly), and the Polyglycine Decapeptide (10Gly)

	N-H stretch		C=O stretch (amide I)		C-N stretch /N-H bend (amide II)	
	Frequency cm^{-1}	Intensity km/mol	Frequency cm^{-1}	Intensity km/mol	Frequency cm^{-1}	Intensity Km/mol
I			1769	243		
1Gly			1833	296		
2Gly	3622	77	1764	305	1536	232
			1829	244		
10Gly	3563	511	1741	1185	1515	3970
	3566	2	1743	176	1525	880
	3567	128	1746	185	1536	31
	3570	17	1750	18	1546	73
	3570	234	1755	85	1552	9
	3575	154	1760	1	1557	13
	3576	146	1764	85	1559	4
	3587	120	1767	60	1561	13
	3610	107	1769	686	1562	13
			1831	255		

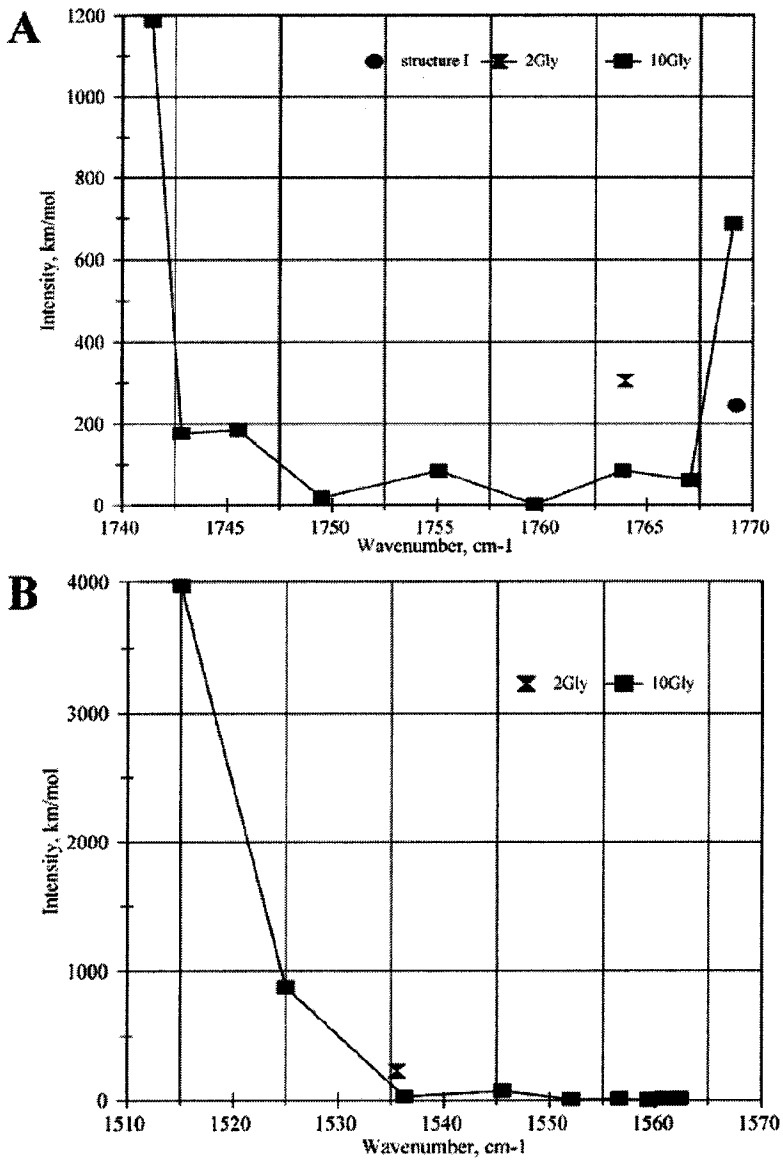


Figure 4.14. Variation of frequencies and intensities for the amide I (C=O) stretch (A) and amide II bend (B) in diglycine and decaglycine.

which avoids the coupling between the amide C=O and the COOH. Using this model, the lowest amide I frequency shifts from 1769 cm^{-1} for single C=O bond to 1741 cm^{-1} in the decapeptide. At the same time the highest frequency stays unchanged at 1769 cm^{-1} compared to **I** (Table 4.5). The lowest, most intense, frequency increases further in intensity as the strand grows in size (Figure 4.14a). However the intensities are much lower than those of the H-bonding chains of comparable sizes and change from 305 km/mol in the dipeptide to 1185 km/mol only in the decapeptide.

The variation of the C=O bond lengths is very small compared to that of the H-bonding chains. All, except the two terminal bonds, vary only from 1.2348 to 1.2352 Å, the terminal bonds are 1.2324 and 1.2322 Å.

The amide II frequencies involve C-N stretching (as well as, N-H deformation) which is more aligned with the length of the β-strand than the C=O stretches. The frequency moves from the dipeptide value of 1536 cm^{-1} to a range of 1515 to 1562 cm^{-1} in the decapeptide. The most intense frequency is always the lowest, which involves the in phase C-N stretches approximately aligned along the length of the strand (in contrast to the amide chains - see below). This transition increases with intensity as the strand becomes larger, becoming more intense than the lowest amide I transition for tripeptides and larger strands. While the range of amide II vibrations extends both above (26 cm^{-1}) and below (21 cm^{-1}) the frequency of the dipeptide, the most intense transition has the largest red shift (21 cm^{-1}). Thus while the centroid of the amide II frequencies exhibits a slight blue shift, the most intense band exhibits a somewhat larger red shift (Figure 4.14b).

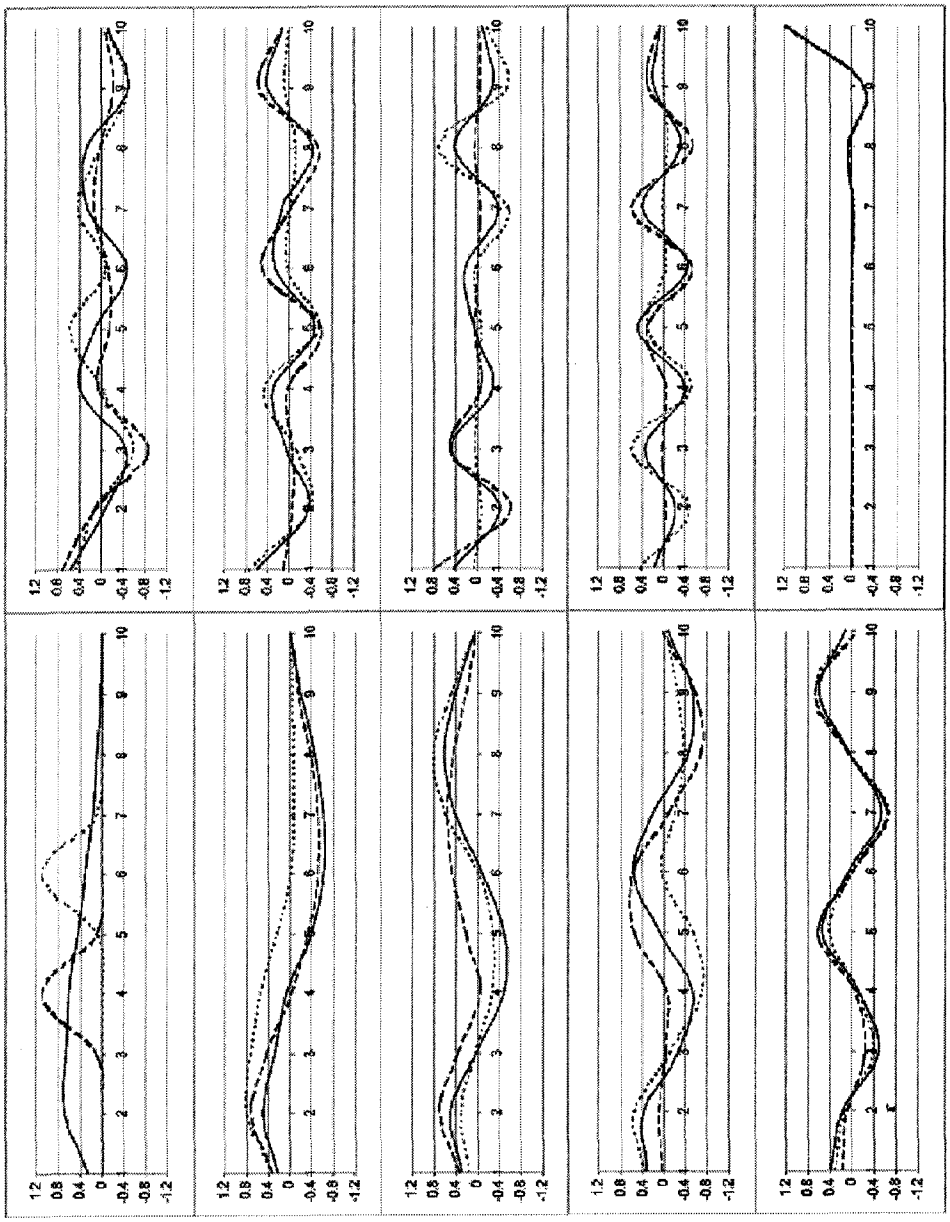


Figure 4.15. Amplitudes for each C=O in the coupled stretches (amide I) with C=O position in decaglycine with no isotopic substitution (solid line), ^{14}C at the fourth C=O (dashed line) and ^{14}C at the sixth C=O (dotted line).

Upon isotopic substitution with ^{14}C at individual positions within the chain containing 10 glycine residues, the carbonyl frequency of the substituted carbonyl decouples with the other unsubstituted carbonyls. The terminal C=Os are different due to the effects of the chain ends. The 'natural' (uncoupled) C=O could be obtained from the frequencies of the ^{14}C substituted decaglycines in a manner analogous to that used for the H-bonding chains. However, unlike the case of the formamide chains, the ^{14}C=O stretching vibrations varied only within a range of 5 cm^{-1} upon ^{14}C-substitution at positions 2-9 in the polyglycine chain. Thus, the natural frequencies of each of these carbonyls are not significantly perturbed by the change in the environment from one position to another within the polyglycine chain. The wavefunctions for decoupled carbonyls are displayed in Figure 4.15, along with the same wavefunctions for the decaglycine chain substituted with ^{14}C in positions four and six. The lowest frequency is delocalized among the all the carbonyls. Unlike the case of the chain of formamides, where the ^{14}C labeled carbonyl acted as a blocker to coupling of the other carbonyls (see Figure 4.10), the substituted carbonyl in the chain of decaglycine does not block coupling of the carbonyls on either side of it. This observation is consistent with a model in which the C=Os couple through the C-C and C-N (but not C=O) bonds.

I have not made any direct comparisons with the through-space dipole-dipole coupling model. However, this model using equivalent dipoles for all C=Os in either the formamide chain or polyglycine should not lead to any shift of the centroid of the amide I frequencies (for example). Clearly, the formamide chain vibrations (which exhibit a large red shift) cannot be adequately described by such a model. I should emphasize that the

dipole coupling theory is simply a model. All the appropriate interactions are implicitly included in the DFT calculations. Furthermore, the observations that the isotopic shifts of the vibrations are quite dependant on the position of the isotope in the formamide chain, and the observations that the C=O bond distances are also dependent on the position in the chain suggest that the C=O dipoles should not be taken as equivalent.

4.4. Comparison of Amide I and Amide II Vibrations in H-Bonding Chains and in β-Strands

4.4.1. Amide I

The coupling of the C=Os in both the H-bonding chains and the β-strands leads to red shifts and increased intensities with increasing size. However, the effects are predicted to be significantly greater for the H-bonding chains than for the β-strands. All frequencies are red shifted in the H-bonding chains. For the β-strands, all but the highest are red shifted. The intensity is concentrated in the lowest (in-phase) frequency for the H-bonding chains. No other frequency reaches 15% of this for the decamer. For the β-strands, the lowest frequency is most intense but generally less than 25% of the corresponding intensities of the H-bonding chains. The highest (slightly blue shifted from the dimer, but unshifted from **I**) frequency is roughly half as intense as the lowest frequency in the decapeptide.

4.4.2. Amide II

The amide II (like) frequencies of the H-bonded chains are all blue shifted while those of the β-strands extend above and below that of the dipeptide with the collective center of the frequencies only slightly blue shifted (see Figure 4.6 and Figure 4.14b). The intensities of these vibrations in the H-bonded chain are all relatively low. They do not change appreciably with chain length. In contrast to the C=O stretches, the lowest frequencies are localized at the ends of the chains, while the highest frequencies are those more delocalized and central.

On the other hand, the intensities of the lowest vibrations of the β-strands are by far the largest in this series. They increase with the size of the β-strand. The second most intense band is the second lowest frequency, which had about 20% of the intensity of the lowest one. Since the most intense vibration is red shifted, the observed spectrum ought to appear red shifted as the β-strands grow.

4.5. Conclusions

Through H-bond coupling of the C=O stretches is clearly stronger than the through (covalent) bond coupling for the models studied. The lowest frequency vibrations are the most red-shifted and most intense. To the extent that these model studies approximate the amide I frequencies in proteins, the implications are clear:

i) Longer H-bonding chains lead to lower frequency, more intense amide I vibrations.

ii) When long H-bonding chains are present, the intensities of these low frequency amide I vibrations might dominate the spectra.

iii) Isotopic substitution at the C=O carbon will lower the red-shifts and the intensities of

the amide I frequencies of the H-bonded chains. Consequently, the amide I vibrations of such isotopically labeled proteins might be dominated by the parts of the protein containing chains that are not labeled. The isotopically labeled regions that have the label near the end of an H-bonding chain will be more prominent than those with the label near the middle of the H-bonding chain. The natural abundance ^{13}C in proteins will have similar effects upon the amide I frequencies.

5.
*COOPERATIVE HYDROGEN-BONDING IN ADENINE-THYMINE AND GUANINE-CYTOSINE BASE PAIRS**

5.1. Introduction

Since the initial report by Clementi,[1] the hydrogen bonds in the purine/pyrimidine base pairs of DNA have been studied many times using molecular orbital methods.[2-11] The aggregate energies of the hydrogen bonds in these base pairs have been carefully evaluated

---

* The results presented in this chapter are published in Kobko, N.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441. This work was supported in part by grants from the National Institutes of Health (S06GM60654), and from PSC-CUNY.

[1] Clementi, E.; Mehl, J.; Von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.

[2] Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

[3] Sponer, J.; Hobza, P.; Leszczynski, J. *Comput.Chem.* **1996**, *1*, 185.

[4] Sponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365.

[5] Brameld, K. A.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1999**, *121*, 985.

[6] Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, *101*, 4851.

[7] Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. -A Europ. J.* **1999**, *5*, 3581.

[8] Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1999**, *103*, 2528.

[9] Dannenberg, J. J.; Tomasz, M. *J. Am. Chem. Soc.* **2000**, *122*, 2062.

[10] Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117.

[11] Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M. *J. Phys. Chem. B* **2000**, *104*, 6286.

and are well documented. The cooperativity in these interactions has been addressed.[7] However, the extent to which hydrogen bond cooperativity contributes to the stability of these base pairs has not previously been quantitatively investigated. In this work, we evaluate the energies of the individual H bonds in the two base pairs of normal DNA (Figure 5.1). We compare these energies with the total H-bonding interaction of each base pair to evaluate the cooperative contributions to the overall stabilization of the pairs. While the interaction energies presented here have been calculated at a reasonably high level, this study aims to characterize the cooperative interactions in the base pairs rather than to perform a state of the art determination of the interaction energies, which would require larger basis sets and optimization on a potential energy surface corrected for BSSE. While such calculations are possible,[12] they would require an enormous amount of computer time for H-bonding systems like G-C whose geometry converges extremely slowly.

Hydrogen-bond cooperativity has been extensively studied for H-bonds that form chains, such as those formed by the amidic functions in peptides and proteins[13-20] (see also

---

[12] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

[13] Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.

[14] (a) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348; (b) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 10389.

[15] Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.

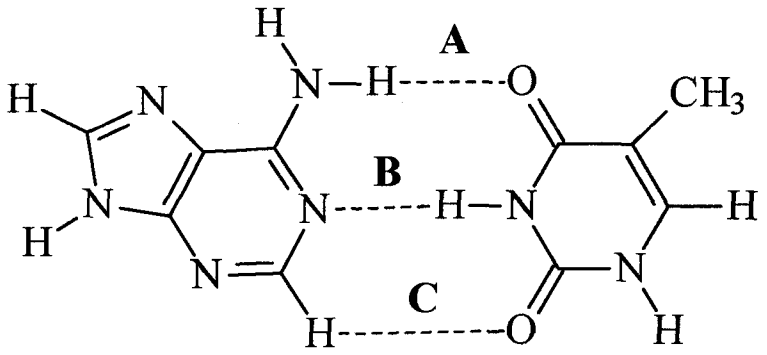
[16] Suhai, S. *Int. J. Quantum Chem.* **1994**, *52*, 395.

[17] Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273.

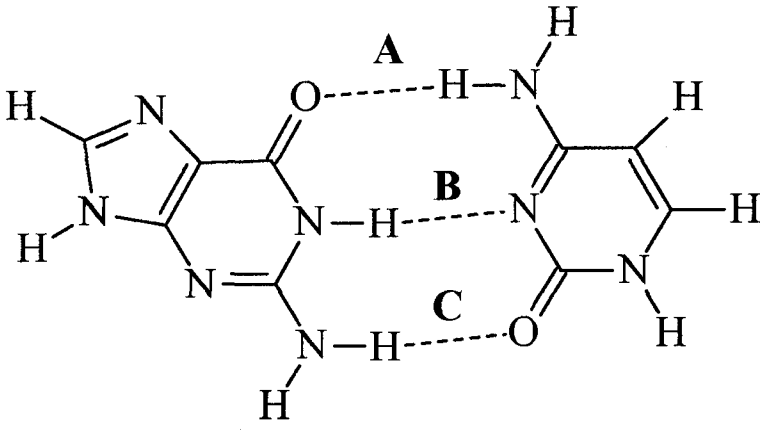
[18] Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451.

[19] Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1748.

[20] (a) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313; (b) Zhao, Y.-L.; Wu, Y.-D. *J. Am. Chem. Soc.* **2002**, *124*, 1570.



AT



GC

Figure 5.1. Base pairs with the H-bonds labeled.

Chapter 2), and in molecular crystals, such as urea, [21-23] acetic acid[24] and the enol of cyclohexane-1,3-dione.[25] These systems generally have only one H-bond between pairs of molecules. Thus, the cooperativity could be studied by determining the interaction energies as molecular aggregates increase in size, or by doing calculations on infinite periodic systems.[23, 26]

The DNA base pairs contain multiple H-bonds between one pair of molecules. In order to evaluate the individual H bonds in the DNA base pairs, one needs to be able to evaluate the energy of a particular hydrogen bond in the absence of the others. To do this, we have initially optimized the geometries of the normal base pairs to obtain the interaction energies of each pair. In order to obtain an estimate of the energies of the individual H-bonds in each pair, we rotated one of the bases with respect to the other about the axis of each hydrogen bond, in turn, so that the planes of the individual bases become perpendicular to each other. We, then, optimized the geometries of these structures with the constraints that (a) the planes of the bases remain perpendicular to each other and (b) that the angle of the hydrogen bond in which the hydrogen atom is central remains fixed. All other geometric parameters were allowed to vary freely. We take the difference between the sum of the individual H-bond energies and the total interaction of the base pairs in their normal coplanar geometries as the cooperative part of the hydrogen bonding interactions.

---

[21] Belosludov, R. V.; Li, Z.-Q.; Kawazoe, Y. *Mol. Eng.* **1999**, *8*, 105.

[22] Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806.

[23] Dovesi, R.; Causa, M.; Orlando, R.; Roetti, C.; Saunders, V. R. *J. Chem. Phys.* **1990**, *92*, 7402.

[24] Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714.

[25] Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1992**, *96*, 5819.

[26] Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3950.

Due to the constraints described above, the structures that contain the single H-bonds are not true minima on a potential energy surface (PES). Thus, the vibrational analyses that are necessary for the evaluation of enthalpies and entropies could not be properly performed. The cooperative contribution to the energies of interactions would be equivalent to the enthalpies if the vibrational energies of the planar and twisted base pairs be the same. Clearly, the model we use for evaluating the cooperativity does not directly address the entropic contribution.

5.2. Computational Details

DFT (B3LYP) and MP2 *ab initio* molecular orbital calculations were performed with the Gaussian 98 suite of computer programs (see Ref. 28, Chapter 1) on our cluster of computers parallelized using LINDA (Ref. 29, Chapter 1). All calculations employed the D95** basis set. The geometries of all species were completely optimized with the following constraints: (1) each of the bases was kept planar, (2) in the perpendicular species, each planar base was kept perpendicular to the other with the X-H...Y angle fixed at its value in the optimized planar base-pair (X and Y refer to the heavy atoms in the H-bond). The number of computed nodes used for each calculation varied with the sizes of the systems studied. The vibrational frequencies were calculated for the planar structures, using the normal harmonic approximations employed in the Gaussian 98 program, to verify the stationary points and calculate the enthalpies of the various species. All frequencies were real except for some very low frequency imaginary vibrations that involved pyramidalization of the NH_2 groups of the individual bases. Vibrational analyses were not

carried out on the perpendicular species, as they are not true minima on the PES. The single-point *a postieri* counterpoise corrections (CP) were calculated using the procedure incorporated in Gaussian 98. Optimization on the CP-corrected PESs (CP-OPT)[12] were not completed due to the excessive CPU time required for the MP2 calculations.

5.3. Results and Discussion

The energetic results are collected in Table 5.1. Table 5.2 shows the bond lengths in the planar and twisted base pairs. The A-T and G-C base pairs with the individual H-bonds labeled to facilitate the following discussion are depicted in Figure 5.1.

We will start by discussing each individual base pair first and then will compare them.

5.3.1. Adenine-Thymine

The A-T base pair has two traditional H-bonds and one weak C-H···O interaction. After BSSE correction, the interaction energies are calculated to be -11.7 kcal/mol and -12.6 kcal/mol by MP2 and DFT respectively. The MP2 calculations have a much greater CP-correction than the DFT calculations. This implies that had the CP-OPT geometry been calculated, it should differ from the normal optimized geometry more for the MP2 than for the DFT optimized geometry. Thus, the MP2 geometry is farther from the minimum on the CP-corrected surface than that for DFT. This explains the weaker interaction calculated

Table 5.1: Energies for H-Bonds in Base Pairs. The Interaction Energy (kcal/mol) Followed by the Energies of Each H-Bond Individually and the Cooperativity (Total Interaction less the Sum of the Individual Bonds)

	MP2/D95**		**B3LYP/D95****	
	no correction	CP-corrected	no correction	CP-corrected
A-T				
Interaction	-17.58	-11.69	-14.70	-12.56
NH..O (A)	-8.78	-4.68	-5.90	-4.59
N..H (B)	-5.98	-3.38	-4.79	-3.85
CH..O (C)	-2.29	-0.06	-1.01	-0.16
Cooperativity	-0.53	-3.57	-3.00	-3.95
G-C				
Interaction	-31.60	-24.50	-29.33	-26.65
O..H (A)	-6.56	-3.55	-5.39	-4.22
N..H (B)	-12.53	-7.92	-8.78	-7.24
NH..O (C)	-13.04	-9.98	-12.45	-10.87
Cooperativity	0.53	-3.05	-2.70	-4.32

Table 5.2: H-Bonds Distances (Å) for the Planar and Twisted Base Pairs

	MP2/D95**			B3LYP/D95**		
	planar	twisted	Difference	planar	twisted	Difference
A-T						
NH..O (A)	1.915	1.997	0.082	1.793	1.923	0.130
N..H (B)	1.774	1.863	0.089	1.887	1.991	0.104
CH..O (C)	2.741	2.336	-0.405	2.834	2.406	-0.429
G-C						
O..H (A)	1.742	1.880	0.138	1.719	1.881	0.162
N..H (B)	1.890	1.882	-0.008	1.879	1.963	0.084
NH..O (C)	1.883	1.843	-0.040	1.881	1.831	-0.050

using MP2. The sums of the interactions for the individual H-bonds are 3.57 and 3.95 kcal/mol less than interaction energy of the planar base pair for MP2 and DFT, respectively. Thus, the cooperative contribution to the H-bonding interaction is 31% of the total interaction for each MO method. Closer examination of the individual H-bonds in the perpendicular base pairs shows that each of the traditional H-bonds is longer than in the planar structure (see Table 5.2). This result agrees with the observation that shorter H-bonds are generally stronger. On the other hand, the C-H···O interaction shortens considerably upon twisting the planar structure about the C-H...O axis. Thus, the C-H···O interaction must be weakened in the planar base pair. This interaction is sacrificed so that the two other stronger interactions can be optimized. The C-H···O interaction is negligible, even in the perpendicular structure. One can safely conclude this last interaction makes no significant contribution to the stability of the planar base pair, where the C-H···O interaction is much longer. Previous studies found no evidence for a C-H···O stabilizing interaction in A-T, despite finding such interactions stabilizing in uracil dimers.[7,11]

5.3.2. Guanine-Cytosine

The energetic analysis for the G-C base pair is more complex. Once again, the CP-corrected DFT interaction of -26.65 kcal/mol is slightly stronger than for MP2 (-24.50 kcal/mol). The cooperative contributions of 3.05 and 4.32 kcal/mol for MP2 and DFT respectively are similar in magnitude to those calculated for A-T (see above). However, this corresponds to only 12-16% of the total interaction of the three H-bonds. Examination of the individual H-bonds in the perpendicular structures shows that only one of them, the

C=O···H interaction (H-bond **A** in Figure 5.1), lengthens upon twisting about its H-bond axis for the MP2 calculations, while both this and the central N-H···N interaction (H-bond **B**) lengthen in the DFT optimization (see Table 5.2). For both calculational methods, the third H-bond shortens. The clear implication is that the molecular geometries do not allow all three H-bonds to simultaneously achieve their optimal interactions. A system with three H-bonds between rigid monomers cannot adjust its intermolecular geometry to simultaneously optimize all three interactions. Formation of an optimal H-bond **A**, for example, requires some geometrical compromise which may lead to less than optimal formation of the other two H-bonds, etc. For a system with only two H-bonds, small adjustments in the angle of one of the H-bonds can lead to a more stable H-bonding distance of the other without appreciably affecting the first H-bonding distance. However, with three H-bonds, such an adjustment of an angle might lead to a more stable interaction of one of the other two H-bonds, but not necessarily both. Thus, the total cooperativity for the three H-bond system (G-C) remains similar in magnitude (but lower in terms of the fraction of the overall interaction) to that for the two H-bond system (A-T).

5.3.3. Comparison of the Adenine-Thymine and Guanine-Cytosine Base Pairs

A more detailed comparison of the H-bonds in the two base pairs indicates that H-bond **A** in G-C has the shortest interaction distance in the planar structure and lengthens considerably more than any of the other H-bonds in either base pair when the bases are held perpendicular to each other. One might be tempted to interpret this as an indication of a large cooperative interaction involving this H-bond. However, the observation that both of

the other H-bonds contract when the bases are perpendicular suggests that H-bond **A** is compressed in the planar base pair. This compression, which allows the other two H-bonds to enhance their interactions, relaxes in the perpendicular structure that preserves H-bond **A**. Variation of the structure of G-C by tilting (while keeping the base pair planar) about the central H-bond (**B** in Figure 5.1) would simultaneously shorten (strengthen) **A** and lengthen (weaken) **C**, or vice versa. Thus, small tilts of this nature would have a much more significant effect upon the structure than upon the energy of the base pair.

Bickelhaupt et al. have noted that the best calculated values for A-T are in better agreement with experiment than those for G-C. They have attributed the poorer agreement of the G-C H-bond lengths to the experimental environment. The inclusion of environmental effects in their calculations improved the agreement with experiment.[10] Another possible explanation might be that BSSE affects the geometry of the G-C pair more than A-T. The combined BSSE of the three H-bonds should exert a larger non-physical attraction between G and C than between A and T, while the ability of the G-C pair to tilt about H-bond **B** would provide it with an additional means of relaxing under this artefactual force. We have previously shown the H-bonding distances on counterpoise-corrected surfaces to be very sensitive to small changes in the corrected interaction energies.[27]

One should note that the observations that both base pairs have substantial cooperative interactions differ from the predictions that might be expected from the secondary electrostatic interactions described originally by Jorgensen.[28] However, we had previously noted that the interaction energy of acetic acid dimer is more than twice that

---

[27] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640.

expected for one H-bond,[29] in disaccord with these predictions. More recently, Leszczynski has also concluded that secondary electrostatic interactions can be deceptive based upon a comprehensive study of 17 multiply H-bonded complexes.[30]

5.4. Conclusions

The cooperative contribution to each of the A-T and G-C pairs is similar in magnitude, but provides a much larger fraction of the overall stability of A-T, which has the weaker overall interaction. The relatively small fraction of cooperativity in the G-C pair appears to be due to the inability of two rigid molecules to form three H-bonds with close to optimal geometries for each individual H-bond. A generalization of this observation suggests that similar problems should be inherent in other systems with more than two H-bonds between inflexible molecules. Thus, the entropic advantage of multiple H-bonds between rigid (rather than flexible) molecules can be partially offset by enthalpic disadvantage of less than optimal individual H-bonding structures.

---

[28] Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.

[29] Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993**, *97*, 12197.

[30] Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 6775.

Part II

EFFECT OF BSSE IN TRANSITION STATES

6.
*EFFECT OF BASIS SET SUPERPOSITION ERROR UPON AB INITIO CALCUATIONS OF ORGANIC TRANSISTION STATES**

6.1. Introduction

Basis set superposition error (BSSE) has long been recognized as a serious problem when calculating the interactions between two (or more) species using *ab initio* or Density Functional Theory (DFT) molecular orbital calculations with basis sets substantially below the Hartree-Fock limit. This error occurs because the energy of each unit within an associated complex will be lowered by the basis functions of the other units. In other words, BSSE introduces nonphysical stabilization in case of supermolecular complex at equilibrium.

The counterpoise (CP) correction proposed by Boys and Bernardi[1] in 1970 continues to be the most prominent tool of correcting for BSSE despite the fact that other

---

* Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2001**, *105*, 1940. Financial support was provided (in part) by PSC-CUNY grant.

[1] (a) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553. (b) Meunier, A.; Levy, B.; Berthier, G. *Theor. Chim. Acta* **1973**, *29*, 49. (c) Jansen, H. B.; Ross, P. *Chem. Phys. Lett.* **1969**, *3*, 40.

methods for correcting this error have been discussed in the literature.[2] The CP method calculates the energy of each unit of the complex with the basis functions of the other units (but without the nuclei or electrons), using so-called "ghost orbitals". The original procedure was conceived for molecules in fixed geometries and atoms. Geometric optimization was not considered.[1] The following equation is a modification of the original procedure, which accounts for the changes in the geometries of the monomeric units upon forming the intermolecular complex

$$E^{CP}_{interaction} = E_{super} - \sum_{i=1}^{n} E_{m^i_{opt}} + \sum_{i=1}^{n} (E_{m^i_f} - E_{m^{i*}_f}) \qquad (6.1)$$

where $E^{CP}_{interaction}$ is the CP-corrected interaction energy, E_{super} is the energy of the complex and E_ms represent the energies of the individual monomers. The subscripts "opt" and "f" denote the individually optimized monomers and those frozen in their supermolecular geometries and the asterisk (*) denotes monomers calculated with ghost orbitals. This method has proven to be somewhat controversial.[3] Nevertheless, van Duijneveldt has shown that CP rigorously corrects for BSSE in certain cases.[3i]

[2] For example, see: (a) Mayer, I.; Surjan, P. R., *Int. J. Quantum Chem.* **1989**, *36*, 225. (b) Almlöf, J.; Taylor, P. R. *J. Chem. Phys.* **1987**, *86*, 553. (c) Saebø, W.; Pulay, P. *J. Chem. Phys.* **1988**, *88*, 1884 and references cited. (d) Mayer, I.; Valiron, P. *J. Chem. Phys.* **1998**, *109*, 3360. (e) Famulari, A.; Specchio, R.; Sironi, M.; Raimondi, M., J. *Chem. Phys.* **1998**, *108*, 3296. (f) Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*, 157. (g) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251. (h) Famulari, A.; Raimondi, M.; Sironi, M.; Gianinetti, E. *Chem. Phys.* **1998**, *232*, 275.

[3] (a) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1984**, *82*, 2418. (b) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer H. F., III *J. Chem. Phys.* **1986**, *84*, 2279. (c) Szalewicz, K.; Cole, S. J.; Kolos, W.; Bartlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3662. (d) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993**, 97, 2488. (e) van Duijneveldt - van Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Ab Initio Methods in Quantum Chemistry*; Lawley, K. P., Ed.; John Wiley & Sons: New York, 1987; Vol. II. (f) Gutowski, M.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Chem. Phys.* **1993**, *98*, 4728. (g) Cook, D. B.; Sordo, J. A.; Sordo, T. L. *Int. J. Quant. Chem.* **1993**, *48*, 375. (h) van Duijneveldt, F. B.; van Duijneveldt - van Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873. (i) Van Duijneveldt, F. B. in *Molecular Interaction: From van der Waals to Strongly Bound Complexes;* Scheiner, S., Ed.; Wiley: Chichester, U.K., 1997, 81-104. (j) Tolosa, L.; Olivares del Valle, F. J. *THEOCHEM* **1994**, *11*, 109. (k) Collins, J. R.; Gallup, G. A. *Chem. Phys. Letters* **1986**, *123*, 56.

Traditionally, one applied CP correction as a single point correction to a previously optimized geometry of the complex.

BSSE introduces a nonphysical attraction between the units of the complex. Thus, the CP correction generally makes intermolecular complexes less stable with longer intermolecular distances than apparent from the normally optimized structure. That means that one should use the CP method to correct the optimized geometry as well as the interaction energy. A few years ago the CP procedure was expanded to gradients and Hessians by S. Simone, M. Duran and J. J. Dannenberg. Their procedure for geometry optimization on the CP-corrected potential energy surfaces (PESs)[4] is now implemented in *Gaussian 03* (Ref. 23, Chapter 1). Others had previously optimized some surfaces using point by point calculations.[5] There are also some reports of geometric optimizations on BSSE-corrected PESs.[6] Usually correction for basis set superposition error (BSSE) is applied to the calculations of intermolecular complexes. It has rarely been applied to the calculation of the energies and never to the geometries of transition states (TSs). Nevertheless, many TSs involve interactions between molecules (or molecular fragments) that should lead to errors similar to those encountered for intermolecular complexes.

In this chapter, we perform the geometric optimization of the TSs of several simple organic reactions on the CP-corrected PESs. For this study we have chosen three different kinds of reactions that exemplify different effects of CP optimization. The first

---

[4] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

[5] (a) Bouteiller, Y.; Behrouz, H. *J. Chem. Phys.* **1992**, *96*, 6033. (b) Del Bene, J. E.; Mettee, H. D. *J. Phys. Chem.* **1991**, *95*, 5387. (c) Leclercq, J. M.; Allavena, M.; Bouteiller, Y. *J. Chem. Phys.* **1983**, *78*, 4606.

[6] (a) Daza, M. C.; Dobado, J. A.; Molina, J. M.; Salvador, P.; Duran, M.; Villaveces, J. L. *J. Chem. Phys.* **1999**, 11806. (b) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640. (c) Van Mourik, T.; Price, S. L.; Clary, D. C. *J. Phys. Chem. A* **1999**, *103*, 1611. (d) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372. (e) Novoa, J. J.; Planas, M. *Chem. Phys. Lett.* **1998**, *285*, 186. (f) Schuetz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997. (g) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575.

reaction that we considered is the concerted Diels-Alder (DA) reaction between ethylene and butadiene, which has a reaction coordinate at the TS that is primarily the approach of the two fragments toward each other. The second is the 1,2-H-atom shift in ethyl radical, which appears to have a transition state with the H atom symmetrically placed with respect to the two carbons (but which may be an artifact of the calculations). And, finally, we examine the two cases of the hydrogen atom transfer between organic radicals: (1) the methyl radical and methane; (2) the methyl radical and propene. The first one has a symmetric TS and a reaction coordinate that primarily involves motion of the H-atom between the two carbons. The second, on the other hand, has an unsymmetric TS in which the reaction coordinate involves both the motion of the H atom relative to the carbons and the approach of the two fragments.

6.2. Computational details

All molecular orbital (MO) calculations were performed using the Gaussian 98 suite of computer programs (see Ref 28, Chapter 1). We obtained the CP-optimized TSs by converging to the geometrically optimized structures using the energy of the CP-corrected first-order saddle points by means of derivatives calculated using equation 6.2 as previously described.[4]

$$\frac{\partial E^{\mathrm{CP}}_{\mathrm{interaction}}}{\partial p_j} = \frac{\partial E^{\mathrm{CP}}_{\mathrm{super}}}{\partial p_j} = \frac{\partial E_{\mathrm{super}}}{\partial p_j} + \sum_{i=1}^{n} \left(\frac{\partial E_{m_i^{\dagger}}}{\partial p_j} - \frac{\partial E_{m_i^{*}}}{\partial p_j} \right) \qquad (6.2)$$

We used the CP-Optimizer program[7] to drive *Gaussian 98* (see Ref. 28, Chapter 1) for the CP-optimization. This program allows the user complete flexibility to specify the individual fragments of the associated complex. The charges and multiplicities of each fragment can be individually defined. Basis sets of varying complexity (from 3-21G to 6-311++G**) were used in this study to illustrate the extent of the CP-optimization effect. We report Hartree-Fock (HF), frozen core second order Moller-Plesset (MP2), and density functional theory (DFT) calculations. For the DFT calculations we used the two hybrid functionals - B3PW91[8] and B3LYP (see Ref. 26 and 27, Chapter 1).

Vibrational frequencies were calculated for all TSs to verify that they are first order saddle-points and to determine the enthalpies. Vibrational frequencies for the CP-optimized structures were obtained from the CP-opt program using second derivatives calculated in a manner analogous to equation 6.1, as previously described.[4]

6.3. Results and Discussion

We first present and discuss the results for the individual reactions, followed by a general discussion.

6.3.1. Concerted Diels-Alder Reaction between Ethylene and Butadiene

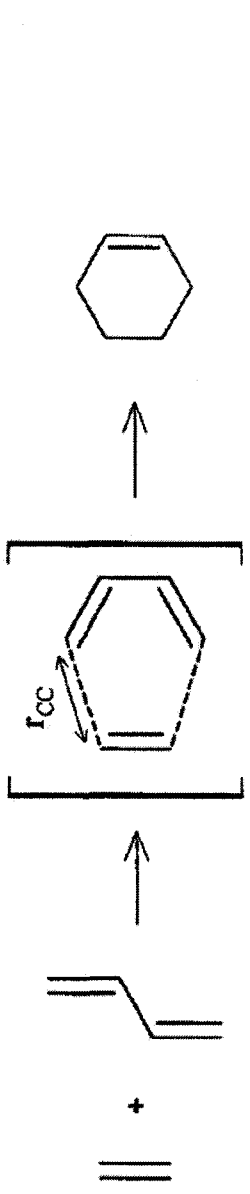
The activation energies, $\Delta H^{\ddagger}$s and distances of the incipient bonds for various TSs optimized with and without CP-correction using different MO methods are summarized in Table 6.1.

---

[7] Duran, M.; Simon, S.; Salvador, P.; Dannenberg, J. J. *Code to Optimize a Molecular System Including Counterpoise Correction* ;Version 4.1 (December 1999).

[8] The B3PW91 method combines Becke's 3-parameter functional (Ref. 3, Chapter 1) with the non-local correlation provided by the Perdew-Wang expression: Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

Table 6.1: Comparison of Normal and CP-Optimized TSs for the Diels-Alder Reaction between Ethylene and Butadiene



method	normal (uncorrected)			CP-corrected			difference		
	ΔE	$\Delta H^{\ddagger}$	r_{CC}, Å	ΔE	$\Delta H^{\ddagger}$	r_{CC}, Å	$\Delta\Delta E$	$\Delta\Delta H^{\ddagger}$	Δr_{CC}, Å
HF/3-21G	33.21	34.58	2.2096	43.27	44.60	2.1949	10.06	10.02	−0.0147
HF/6-31G	40.70	41.86	2.2039	44.55	45.75	2.1962	3.86	3.89	−0.0077
HF/6-31G**	42.26	43.39	2.2002	45.83	46.99	2.1947	3.57	3.60	−0.0055
HF/D95**	43.75	44.80	2.2002	44.73	46.37	2.1983	0.97	1.57	−0.0019
HF/D95++**	43.84	44.89	2.1999	44.82	45.87	2.1980	0.98	0.98	−0.0018
HF/6-311++G**	45.39	46.50	2.1925	46.10	47.22	2.1910	0.71	0.72	−0.0015
B3PW91/3-21G	10.66	11.80	2.3280	18.46	19.66	2.2837	7.79	7.86	−0.0442
B3PW91/6-31G	17.35	18.37	2.2885	20.21	21.29	2.2772	2.86	2.92	−0.0113
B3PW91/6-31G**	16.30	17.29	2.2950	18.95	20.01	2.2863	2.65	2.72	−0.0087
B3PW91/D95**	17.42	18.22	2.2982	18.21	19.06	2.2938	0.78	0.84	−0.0044
B3PW91/D95++**	17.64	18.48	2.2974	18.32	19.22	2.2950	0.69	0.74	−0.0023
B3PW91/6-311++G**	18.43	19.39	2.2736	19.06	20.08	2.2703	0.63	0.69	−0.0033
B3LYP/6-311++G**	21.99	23.00	2.2469	22.54	23.62	2.2464	0.55	0.62	−0.0005
MP2/3-21G	15.70	16.81	2.2677	28.82	29.97	2.2182	13.12	13.16	−0.0495
MP2/6-31G	20.76	21.73	2.2557	28.55	29.58	2.2240	7.80	7.85	−0.0316
MP2/6-31G**	14.45	15.35	2.2832	21.46	22.42	2.2563	7.01	7.07	−0.0269
MP2/D95**	13.74	14.79	2.2997	19.60	20.72	2.2720	5.86	5.93	−0.0277
MP2/D95++**	12.66	13.65	2.2982	18.36	19.39	2.2773	5.70	5.74	−0.0209
MP2/6-311++G**	12.93	13.84	2.2936	16.87	17.80	2.2804	3.94	3.96	−0.0133

We considered two fragments, ethylene and butadiene, in the CP-opt procedure. The TSs optimized with CP-correction all have higher energies and *shorter incipient bond lengths* than the TSs optimized using the traditional procedure. The higher energies are a reasonable expectation based upon previous experience with the stable intermolecular complexes, whose energies also increase (stabilization decreases) upon application of CP. However, intermolecular distances generally increase when these complexes are optimized on a CP-corrected surface. In these cases removal of the nonphysical attraction (attributed to BSSE) between the molecules moves the minimum on the PES to a larger intermolecular separation.

For the TS in the DA reaction, we must consider the effect of the nonphysical attractive force upon the reaction coordinate, which has its maximum value at the TS (see Figure 6.1). At the TS calculated on the normal (uncorrected) PES (point *a* in Figure 6.1), the gradient of the force along the reaction coordinate is zero (as for all the other internal coordinates). The effect upon the reaction coordinate of the uncorrected PES of removing the nonphysical attraction is equivalent to adding a repulsion of equal magnitude. This repulsion will increase along the reaction coordinate from reagents to product as the molecular fragments approach each other. If one imagines adding such a repulsion to the traditional uncorrected PES (point *d* of Figure 6.1), the gradient of the reaction coordinate at the original TS will no longer be zero on the CP-corrected PES. Rather, it will become positive in the direction of product. Upon moving along the reaction coordinate from the traditional TS towards product, this repulsive force increases, whereas the gradient of the normal PES decreases. At some point on the product-side of the original TS, the repulsive force will be exactly canceled by the downward slope of the traditional uncorrected PES.

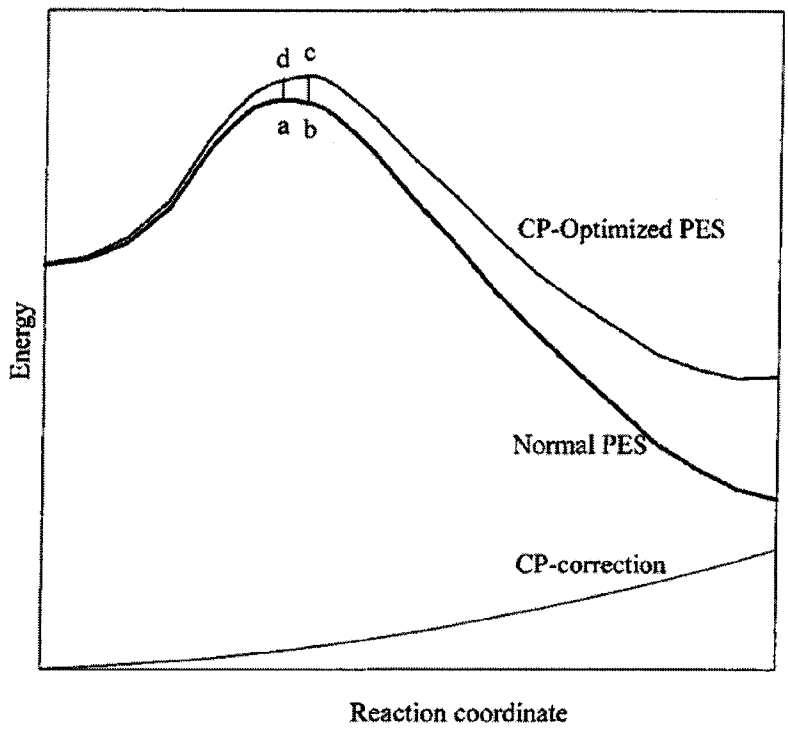


Figure 6.1. Schematic comparison of normal and CP-optimized PESs for the Diels-Alder reaction between ethylene and butadiene. The CP-optimized surface can be understood as the sum of the normal (uncorrected) surface and the CP correction. Points *a* and *c* represent the optimized structures for the transition state on the normal and CP-optimized surfaces, respectively. Point *b* represents the CP-optimized structure on the normal surface. Point *d* represents the normal optimized structure on the CP-corrected surface.

This point becomes the new (CP-corrected) TS (point *c* in Figure 6.1). Because its reaction coordinate is on the product-side of the original TS, its incipient bond-lengths will be shorter than those of the uncorrected PES.

The MP2 calculations all predict activation parameters that are significantly higher than the HF and DFT calculations. The effects of CP-correction increase in the order DFT < HF < MP2 both for increase in activation and decrease in incipient bond-length. We had previously noted that the single-point CP-correction for the TS of the DA reaction between butadiene and cyclopropene is quite large when quadratic configuration interaction (QCI) is used.[9] The effects diminish as the basis sets increase in complexity, as expected.

For large basis sets, only the B3LYP calculations are reasonably consistent with the experimental activation energy. The kinetics of the gas phase reaction between ethylene and butadiene were determined at temperatures between 760 and 921 K[10]. An Arrhenius plot in this temperature range gave and activation energy of 27.5 kcal/mol (extrapolated to 25.1 at 0 K). Because of the high temperatures employed in the study, there is reason to believe that a significant fraction of the reaction follows a step-wise rather than concerted reaction. The $\Delta S^{\ddagger}$ for the former should be significantly less negative, making the step-wise reaction more competitive at high temperatures.[11] B3LYP/6-31G* calculations on the reaction between ethylene and butadiene place the step-wise (biradical) TS only 3.4 kcal/mol higher than the concerted TS.[12] Therefore, the

---

[9] Sodupe, M.; Rios, R.; Branchadell, V.; Nichols, T.; Oliva, A.; Dannenberg J. J. *J. Am. Chem. Soc.* **1997**, *119*, 4232.

[10] Rowley, D.; Steiner, H. *Discuss. Faraday Soc.* **1951**, *10*, 198.

[11] Dannenberg, J. J.; Liotard, D., in preparation.

[12] Goldstein, E.; Beno, B.; Houk, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 6036.

expected activation energy for the concerted reaction might be a few kcal/mol less than the experimental value.

6.3.2. Hydrogen 1,2 Shift in the Ethyl Radical

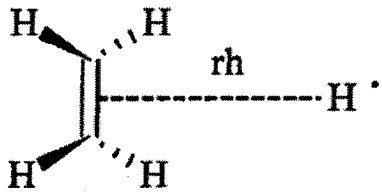
The results presented in this section are collected in Table 6.2.

The 1,2-sigmatropic shift of an H-atom in ethyl radical is a symmetry-forbidden process because the lowest unoccupied molecular orbital (LUMO) of the ethylene has a nodal plane that bisects the C-C bond axis. Thus the hydrogen's 1s orbital has zero overlap with the LUMO as it crosses the nodal plane (at the presumed TS). One may reasonably question whether a TS exists for this reaction or whether the system would preferentially dissociate into ethene and a H-atom. In this case, the PES for ethylene plus an H-atom would have two valleys for approach of an H-atom to either carbon, separated by a ridge that increases in magnitude as the H-atom approaches the ethylene. This situation is qualitatively similar to that previously proposed for the differentiation of the reaction paths for coupling versus disproportionation of two ethyl radicals.[13] Simple MO calculations do predict a well defined TS (only one imaginary frequency which connects the two equivalent ethyl radicals) for this process. There is a slight increase in energy as the H-atom is removed to infinite distance from its position in the TS. The obvious question arises: Does this TS owe its existence to weak bonding of the H-atom to orbitals of the ethylene other than the LUMO, or is the apparent TS simply an artifact on the PES caused by BSSE?

---

[13] Dannenberg, J. J.; Baer, B. *J. Am. Chem. Soc.* **1987**, *109*, 292.

Table 6.2: Hydrogen 1,2 Shift in the Ethyl Radical



	normal (uncorrected)		CP-corrected		difference	
method	ΔE^a	rh, Å	ΔE	*rh*, Å	$\Delta\Delta E$	Δrh, Å
HF/3-21G	−0.04	3.709	0.00	5.083	0.04	1.374
HF/6-31G	−0.03	3.961	0.00	5.376	0.03	1.414
HF/6-31G**	−0.03	3.933	0.00	5.258	0.03	1.325
HF/D95**	−0.01	4.322	0.00	5.626	0.01	1.303
HF/D95++**	−0.01	5.398	0.00	5.475	0.01	0.077
HF/6-311++G**	−0.00	6.131	0.00	6.131	0.00	0.000
B3PW91/3-21G	0.01	4.963	0.00	5.571	0.01	0.608
B3PW91/6-31G	0.00	6.201	0.00	6.201	0.00	0.000
B3PW91/6-31G**	0.00	6.097	0.00	6.097	0.00	0.000
B3PW91/D95**	0.01	4.619	0.00	6.499	−0.01	1.880
B3PW91/D95++**	−0.00	6.318	0.00	7.502	0.00	1.184
B3PW91/6-311++G**	0.01	6.019	0.00	7.813	−0.01	1.794
B3LYP/6-311++G**	0.04	3.576	0.05	3.660	0.01	0.084
MP2/3-21G	−0.06	3.632	0.00	5.072	0.06	1.440
MP2/6-31G	−0.04	3.868	0.00	5.356	0.04	1.488
MP2/6-31G**	−0.05	3.778	−0.00	5.080	0.05	1.302
MP2/D95**	−0.03	4.062	−0.00	5.211	0.03	1.148
MP2/D95++**	−0.03	4.508	−0.00	4.880	0.02	0.372
MP2/6-311++G**	−0.03	4.325	−0.01	4.534	0.02	0.208

In order to answer this question, we optimized the geometry of a system in which an H-atom is constrained to the plane that perpendicularly bisects the C=C bond of ethylene (the nodal plane of the LUMO). A minimum found in this plane should correspond to a TS for the 1,2-shift. In Table 6.2 the ΔE values refer to the stabilization of the TS relative to separated ethylene and an H-atom. Inspection of this table shows that HF and MP2 calculations with small basis sets predict a well-defined TS with a distance of about 3.6-3.7 Å between the H-atom and the midpoint of the C=C bond. The energies for removal of the H-atom are about 0.04 kcal/mol (HF) and 0.06 kcal/mol (MP2). As the basis sets become more complex within a series, the predicted distance between the H-atom and the midpoint of the C=C bond increases to values as large as 6.13 Å (HF/6-311++G**) with an ΔE of about 15×10^{-3} kcal/mol.

Optimization of these structures on the CP-corrected surfaces appeared to give TSs with much longer H atom-carbon separations and very low E_as. However, CP-corrected frequency calculations indicate these structures not to be true TSs. For these CP-opt calculations, the fragments considered were ethylene and a hydrogen-atom. The gradients of the TSs continue to decrease as the H-atom moves away from the ethylene (Figure 6.2); however, the surface had become so flat that the gradients of the energy with respect to this distance had become effectively zero (thus satisfying the program's convergence criterion). Clearly, this is a case where the qualitative appearance of the PES changes when it is calculated with CP-correction.

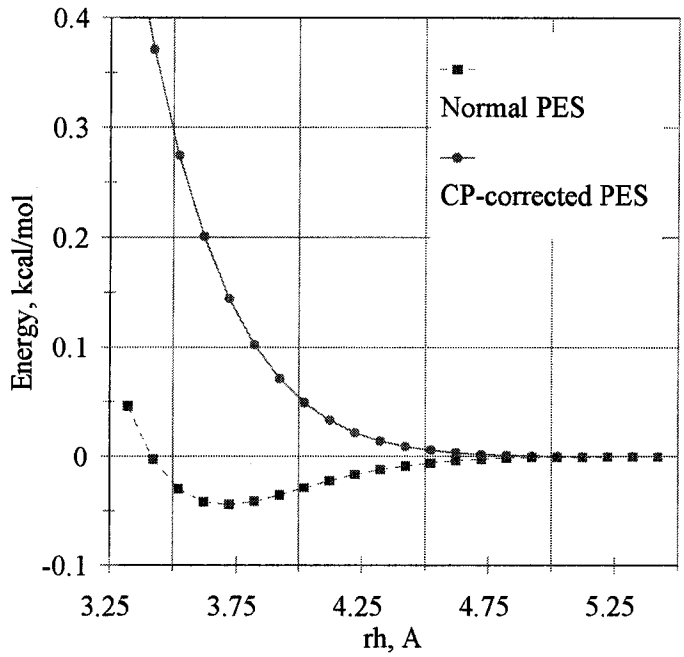


Figure 6.2. Comparison of the normal and CP-corrected PESs for the H migration in ethyl radical at the HF/3-21G level. The figure represents a cross section of the surface in a plane that bisects and is perpendicular to the C=C bond, whereas *rh* is the distance between the midpoint of the C=C bond and the H atom (see Table 6.2).

One should note that another saddle point for H-transfer in ethyl radical exists. This structure, which has the H-atom much more tightly bound to the ethylene, can be thought of as an H-atom forming a stabilizing interaction with the π-bonding orbital of an excited ethylene. Removal of the H-atom via dissociation of the H-atom along a path that preserves the plane of symmetry leads to an excited triplet state of ethylene,[14] clearly an endoergonic process (see Figure 6.3). Because H transfer via this saddle point will surely be of higher free energy, than the dissociation of the H-atom on the ground-state surface, the associated saddle point cannot be properly called a TS for a thermal reaction. Nevertheless, we have calculated the CP-corrected saddle point for this process at two levels: HF/6-31G and B3LYP/6-311++G**. In this case the two fragments used for the CP-optimization procedure are an H-atom and a triplet ethylene. Since the normal coordinate, which separates the H-atom from the triplet ethylene, has a rapidly increasing slope, the CP-correction on the PES is minimal. The optimization of this point on the CP-optimized surface is noteworthy as it is the first example to our knowledge where CP-optimization has been performed using an excited state of one of the fragments in the complex.

The energies of the CP-corrected saddle point are +18.45 and -0.44 kcal/mol compared to those of separated (ground state) ethylene and H atom for the HF and DFT CP-optimized species, respectively. At the lower level of calculation, this saddle point is considerably higher than the dissociated species, implying that it could be reasonably be

---

[14] (a) Sevin, A.; Yu, H. T.; Evleth, E. M. *THEOCHEM* **1983**, *104*, 163. (b) Evleth, E. M.; Cao, C. Z.; Kassab, E.; Sevin, A. *Chem. Phys. Lett.* **1984**, *109*, 45.

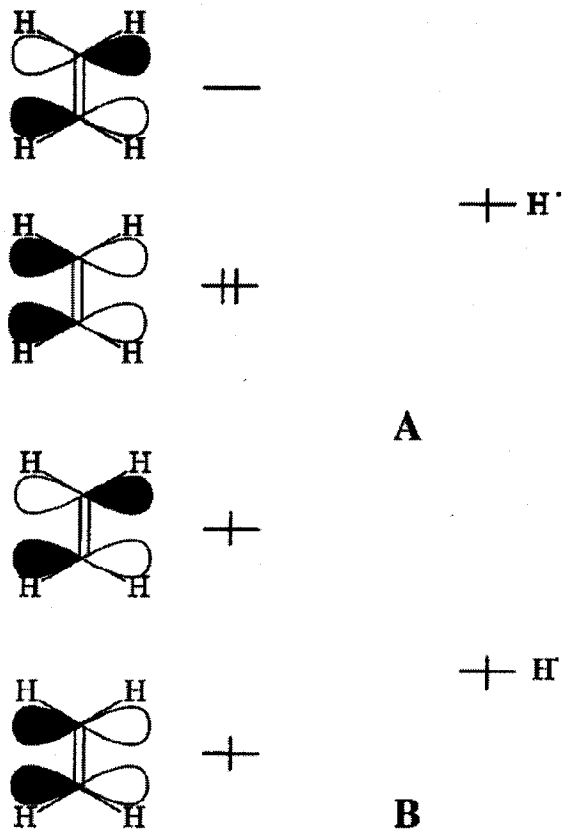


Figure 6.3. Orbital interaction diagrams for ethylene plus a hydrogen atom in the plane perpendicular to and bisection the C=C bond. (A) ground singlet-state ethylene plus H; (B) excited triplet-state ethylene plus H.

neglected. However, the more sophisticated calculation suggests this saddle point is energetically equivalent to the separated species. The necessarily more negative entropy for this saddle point makes it an extremely unlikely candidate for the TS on a free energy surface.

6.3.3. H-atom Transfer from Methane to Methyl Radical

The results discussed in this section are collected in Table 6.3.

For this and the following reaction, three fragments were considered for the CP-opt procedure: an H-atom and the two appropriate radicals (two methyls for the first reaction and a methyl and an allyl for the second one). Using only two fragments (i. e. methyl radical and methane) for CP-optimization would destroy the symmetry of the TS, as noted by Mayer.[15] However, it is entirely appropriate to use the three fragments system as the quasi-thermodynamic state properties of the TS should be independent of whether it be formed from a methyl radical and methane or two methyl radicals and an H-atom.

In this process, the reaction coordinate involves: a) the approach of the reactants, b) the transfer of the H-atom, and c) and the separation of the products. All MO methods that we have tried predict the reaction coordinate at the TS to primarily involve the motion of the migrating H-atom between the carbons. The distance between the carbons remains relatively insensitive to the reaction coordinate at the TS. Nevertheless, the nonphysical force caused by BSSE should affect this distance at the TS. Removing this force would cause the C...C separation to increase at the TS. If the C...C separation were greater, the hydrogen-atom would have to traverse a larger distance during its migration.

---

[15] Lendvay, G.; Mayer, I. *Chem. Phys. Lett.* **1998**, *297*, 365.

Table 6.3: Hydrogen Transfer from Methane to the Methyl Radical[a]

method	normal (uncorrected)			CP-corrected			difference			
	ΔE	$\Delta H^{\ddagger}$	r_1, r_2, Å	ΔE	$\Delta H^{\ddagger}$	r_1, r_2, Å	ΔE(TS)	$\Delta\Delta E$	$\Delta\Delta H^{\ddagger}$	$\Delta r_1, \Delta r_2$, Å
HF/3-21G	27.20	26.22	1.3562	30.97	29.36	1.3676	5.18	3.77	3.15	0.0114
HF/6-31G	29.69	28.34	1.3555	30.84	29.44	1.3582	1.63	1.15	1.10	0.0027
HF/6-31G**	29.73	28.39	1.3560	30.71	29.30	1.3578	1.46	0.98	0.91	0.0018
HF/D95**	30.61	29.10	1.3547	31.12	30.62	1.3558	1.12	0.51	1.51	0.0011
HF/D95++**	30.61	29.11	1.3551	30.86	29.35	1.3563	0.52	0.24	0.25	0.0012
HF/6-311++G**	30.53	29.08	1.3569	30.71	29.25	1.3578	0.34	0.17	0.17	0.0009
B3PW91/3-21G	11.35	10.29	1.3446	15.57	13.88	1.3554	5.67	4.22	3.59	0.0108
B3PW91/6-31G	13.88	12.43	1.3430	15.27	13.22	1.3458	1.91	1.38	0.79	0.0028
B3PW91/6-31G**	13.76	12.21	1.3431	14.96	13.02	1.3455	1.75	1.20	0.81	0.0024
B3PW91/D95**	14.31	12.67	1.3416	15.02	13.00	1.3429	1.52	0.71	0.33	0.0013
B3PW91/D95++**	14.48	12.83	1.3414	14.70	12.66	1.3425	0.46	0.21	−0.17	0.0011
B3PW91/6-311++G**	14.84	13.22	1.3423	15.01	13.39	1.3431	0.27	0.17	0.17	0.0008
B3LYP/6-311++G**	15.34	13.72	1.3468	15.49	13.97	1.3475	0.24	0.15	0.25	0.0007
MP2/3-21G	21.14	20.05	1.3435	25.93	24.25	1.3606	7.15	4.79	4.21	0.0171
MP2/6-31G	22.97	21.42	1.3442	25.13	23.02	1.3530	3.50	2.16	1.60	0.0088
MP2/6-31G**	20.81	19.22	1.3241	22.74	20.60	1.3308	3.30	1.94	1.38	0.0067
MP2/D95**	20.57	18.92	1.3225	22.20	20.50	1.3303	3.17	1.63	1.58	0.0078
MP2/D95++**	20.36	18.61	1.3226	21.78	20.10	1.3303	2.62	1.42	1.49	0.0077
MP2/6-311++G**	19.77	17.90	1.3265	20.69	18.85	1.3320	1.79	0.92	0.95	0.0055

[a] Transition state, kcal/mol. ΔE(TS) indicates ΔE for TS only. $\Delta\Delta E$ and $\Delta\Delta H^{\ddagger}$ are activation parameters where CP optimization is applied to the reactants as well as the TS.

Consequently, the H-atom would have decreased overlap with each C at the TS. Thus, one would expect the calculated activation energy to increase after CP-correction is applied to the PES. The results collected in Table 6.3 and displayed in Figure 6.4 confirm these expectations. The activation parameters in the table refer to the energy differences between the TS and the reagents calculated using the same three fragments. Thus, methane is calculated using the CP-optimization procedure with a methyl radical and an H atom as the two fragments. This breaks the symmetry of methane but provides a continuous potential-energy surface for the H-transfer reaction. Because the changes in activation energies upon CP optimization derive both from the energetic change in the TS and the methane, the change in the energy of the TS alone, $\Delta E(TS)$, is included in Table 6.3 along with the changes in the activation parameters. Once again, the effect of CP-correction upon the TS decreases as the basis set becomes more complex. Also, the effect is greater for MP2 than for HF or DFT calculations.

HF calculations greatly overestimate the activation parameters for this reaction, while the best DFT calculations predict a $\Delta H^{\ddagger}$ very close to the experimental[16] value of 14.9 kcal/mol for abstraction from methane by $CD_3\bullet$. The MP2 calculations also predict $\Delta H^{\ddagger}$s that are higher than the experimental reports. There may be a contribution from H-atom tunneling that would make the calculated $\Delta H^{\ddagger}$ higher than that derived from experimentally measured rates. However, the apparently small deviations from linearity of the Arrhenius plots suggest tunneling to be unimportant. Nevertheless, Truhlar[17] has

[16] Taken from Trotman-Dickenson, A. F.; Milne, G. S. *Tables of Bimolecular Gas Reactions*; NSRDS-NBS 9, U. S. Department of Commerce, 1967.

[17] (a) Steckler, R.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys.* **1986**, *84*, 6712. (b) Hu, W. P.; Liu, Y. P.; Truhlar, D. G. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1715.

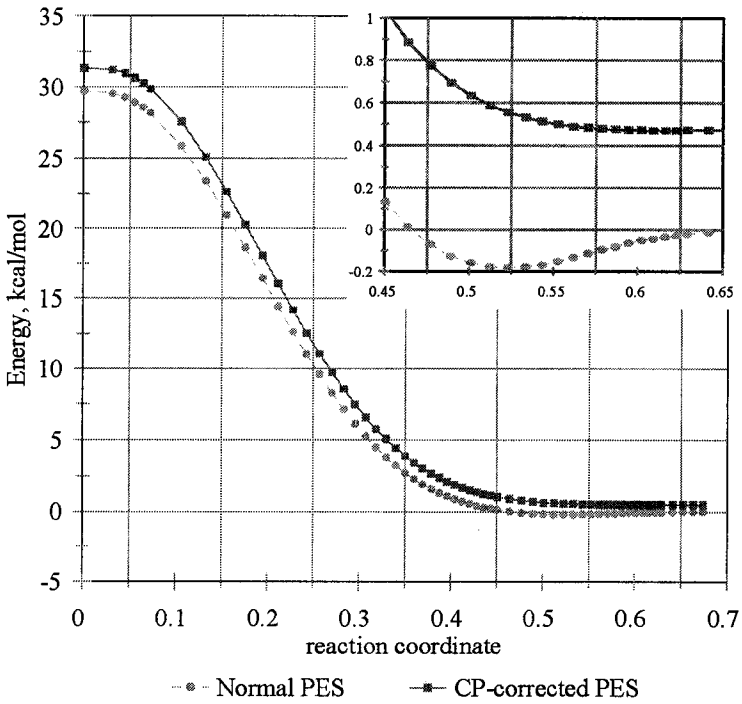


Figure 6.4. Normal and CP-corrected PES for the H3C-H-CH3 system (HF/6-31G). To perform a scan, distance r_1 was gradually increase while the rest of the system was optimized, whereas the single-point CP correction was performed for every optimized geometry. The ratio $(r_1 - r_2)/r_c$ is used ad the reaction coordinate to generate the graph. The inset shows a magnified version of the region for $(r_1 - r_2)/r_c = 4.5 - 6.5$. Note the minimum is on the uncorrected PES only. See Table 6.3 for definitions of parameters.

suggested that there may be increased tunneling as the system approaches the TS. Such a situation could be consistent with the experimental Arrhenius plots.

In addition to the effect upon the TS, BSSE has another noticeable effect upon the PES. The methyl radical and methane molecule are calculated to form a weak complex when small basis sets are used. This phenomenon is illustrated in Figure 6.4 (inset) for the PES calculated at the HF/6-31G level. As noted in the figure, CP-correction removes this (apparently anomalous) minimum on the PES.

6.3.4. H-Atom Transfer from Propane to the Methyl Radical

This reaction is similar to the H-transfer for methane to methyl radical. However, it does not have a symmetrical PES as the products (methane and ally radical) are more stable than the reactants. The Hammond postulate dictates that the TS should be earlier along the reaction path than for the previous reaction. Thus, one might reasonably expect the reaction coordinate at the TS to involve approach of the two reactive carbon centers, as well as, movement of the H-atom between these centers. As in the case of the Diels-Alder reaction, the gradient of the reaction coordinate will be positive at the traditionally optimized (uncorrected) TS (because the two molecular fragments are approaching each other). Thus, the CP-optimized TS will be somewhat closer to product. The progress along the reaction coordinate can be measured by the relative C...H distances for the breaking and forming bonds. Because the TS is early, the ratio of the breaking to forming C...H bond distances should be <1. As one moves towards product along the reaction coordinate, this ratio will increase. The data in Table 6.4 shows that this ratio at the TS increases upon CP-optimization of the TS. We have already seen (from the previous

Table 6.4: Hydrogen Transfer from Propene to the Methyl Radical[a]

method	normal				CP-corrected				difference			
	ΔE	$\Delta H^{\ddagger}$	rc, Å	$r1/r2$, Å	ΔE	$\Delta H^{\ddagger}$	rc, Å	$r1/r2$, Å	ΔE(TS)	$\Delta\Delta E$	$\Delta\Delta H^{\ddagger}$	Δrc, Å	$\Delta(r1/r2)$, Å
HF/3-21G	20.34	18.70	2.738	1.1001	24.55	22.26	2.769	1.0965	5.77	4.22	3.56	0.030	−0.0036
HF/6-31G	22.55	20.48	2.746	1.1031	23.77	21.11	2.753	1.1028	1.76	1.22	0.63	0.008	−0.0003
HF/6-31G**	23.80	21.78	2.742	1.0912	24.86	22.84	2.747	1.0909	1.59	1.06	1.06	0.005	−0.0003
HF/D95**	25.09	22.93	2.736	1.0850	25.64	23.48	2.739	1.0846	1.27	0.55	0.55	0.003	−0.0003
HF/D95++**	25.08	22.93	2.737	1.0849	25.38	23.23	2.740	1.0843	0.58	0.30	0.30	0.002	−0.0005
HF/6-311++G**	25.13	23.07	2.741	1.0845	25.34	23.28	2.742	1.0842	0.38	0.21	0.21	0.002	−0.0003
B3PW91/3-21G	5.44	4.52	2.733	1.1822	9.99	8.41	2.758	1.1656	6.25	4.55	3.89	0.024	−0.0166
B3PW91/6-31G	8.01	6.72	2.739	1.1846	9.44	7.56	2.746	1.1828	2.11	1.43	0.83	0.008	−0.0018
B3PW91/6-31G**	8.06	6.64	2.736	1.1775	9.32	7.90	2.741	1.1761	1.94	1.26	1.26	0.005	−0.0013
B3PW91/D95**	8.79	7.28	2.730	1.1716	9.56	7.46	2.733	1.1704	1.78	0.77	0.18	0.003	−0.0012
B3PW91/D95++**	9.00	7.47	2.730	1.1710	9.29	7.74	2.732	1.1698	0.59	0.29	0.27	0.002	−0.0012
B3PW91/6-311++G**	9.29	7.79	2.732	1.1708	9.51	8.02	2.733	1.1700	0.36	0.22	0.23	0.001	−0.0008
B3LYP/6-311++G**	9.80	8.29	2.736	1.1574	9.99	8.49	2.737	1.1568	0.31	0.19	0.20	0.001	−0.0006
MP2/3-21G	20.72	20.56	2.689	1.0542	26.10	25.33	2.733	1.0515	8.30	5.38	4.76	0.043	−0.0026
MP2/6-31G	23.46	22.89	2.698	1.0621	25.81	24.66	2.721	1.0621	4.16	2.35	1.78	0.023	−0.0000
MP2/6-31G**	19.77	19.23	2.656	1.0539	21.97	20.86	2.673	1.0535	3.93	2.20	1.63	0.017	−0.0004
MP2/D95**	19.62	19.03	2.651	1.0469	21.56	20.39	2.669	1.0461	3.94	1.94	1.37	0.018	−0.0008
MP2/D95++**	19.16	18.56	2.652	1.0468	21.08	19.90	2.669	1.0454	3.38	1.92	1.34	0.017	−0.0014
MP2/6-311++G**	18.24	17.64	2.659	1.0474	19.46	18.28	2.672	1.0467	2.31	1.22	0.65	0.013	−0.0006

[a] Transition state, kcal/mol. ΔE(TS) indicates ΔE for TS only. $\Delta\Delta E$ and $\Delta\Delta H^{\ddagger}$ are activation parameters where CP optimization is applied to the reactants as well as the TS.

discussion) that the H-atom transfer is facilitated by a close approach of the reactive carbon centers. This distance increases upon CP-optimization of the TS causing increases in the calculated activation parameters.

As in the case of the previous reaction, the HF calculations give activation parameters that are much too high. The best DFT calculations give results in accord with the reported experimental[20] value of 7.7 kcal/mol for $\Delta H^{\ddagger}$.

6.4. General Discussion

Clearly, CP-corrected surfaces lead to noticeable improvements in the geometrically optimized TSs obtained with relatively small basis sets. The geometries and associated activation energies become closer to those obtained with larger basis sets. The CP corrections to these parameters tend to diminish as the basis sets are improved. In other words, optimization on the CP-corrected PESs lead to the results that are less dependent on the choice of the basis set. These observations may be contrasted with the earlier suggestion by Lendvay and Mayer that CP-correction is inappropriate for the optimizations of TSs.[15,18] They quite correctly noted that certain choices for the fragments used in the CP correction calculation would be inappropriate. In particular, they criticized two applications of CP to TSs.[19] For example, had we used two fragments (CH_4 and $CH_3^{\bullet}$) for the methane/methyl H-transfer reaction, we would have obtained an inappropriate, unsymmetrical TS. However, a flexible choice of the fragments will usually allow reasonable CP calculations of the TS, at least for the cases where all

[18] See also on the problem of the choice of fragments: Chalasinski, G.; Szczesniak, M. *Chem. Rev.* **2000**, *100*, 4227.

[19] (a) Nanayakkara, A. A.; Balint-Kurti, G. G.; Williams, I. H. *J. Phys. Chem.* **1992**, *96*, 3662. (b) Tooley, D. W.; Anderson, J. G. *J. Phys. Chem.* **1989**, *93*, 1049.

fragments are neutral. Nevertheless, the seemingly arbitrary nature of the fragments choices bear some detailed discussion.

Since the TS is defined by transition-state theory as a quasithermodynamic state, it is approximated to have the associated thermodynamic state functions. Thus, the First Law can be applied to the TS. As a consequence, one can construct the TS from any fragments, not simply from the reagents and products. Clearly, the BSSE will depend upon the choice of fragments, just as it will depend upon the choice of basis set. Thus, the extent of the BSSE and the CP will differ somewhat depending upon how the fragments are chosen. In principle, the TS could be constructed from the individual atoms. This would lead to a different BSSE. While CP is not usually now applied to the construction of molecules from atoms, it is noteworthy that this was the original application by Jansen and Ros.[1c]

One can consider the choice of the basis set in an analogous manner. Typically, nucleus-centered gaussians are used to construct basis sets. These lead to obvious assignments of these gaussians to the atom on whose nucleus it is centered and the molecular fragment to which this atom belongs. However, other choices for basis function exist and have been used. For example, gaussians can be placed along the bond axes between the atoms. If such a bond were broken into two fragments, the arbitrary decision must be made to associate the bond-gaussian with one of the two fragments. If the other functions in the basis set were sufficient to provide HF-limit calculations for the original structure, as well as, each of the two fragments, the energy of bond rupture would be insensitive to the assignment of the bond Gaussian to one or the other fragment. The use of inadequate nucleus centered gaussian basis sets presents an analogous

problem. Because the basis set of the first fragment is inadequate, it borrows density from the bases of the other fragment. Thus, the apparent basis functions of the second fragment are also being used by the first fragment. In fact these basis functions on the second fragment are also basis functions of the first (just as the bond Gaussian was). This problem will disappear as the originally inadequate basis set is improved. Obviously, this is simply another way of describing BSSE phenomenologically. In any case, any reasonable construction of the TS from fragments suitably chosen to preserve the essential properties of the TS (i. e., correct symmetry) should lead to an improvement in the TS calculated with relatively small basis sets. The correction should tend to disappear as the basis set is improved. We see that both the choice of the fragments and basis sets used to construct a molecular complex will influence the calculated energies and geometries of the molecular complex when the basis sets are inadequate, but not when they are complete. Even the aforementioned TS for the $CH_4/CH_3^{\bullet}$ H-atom transfer will converge to the correct TS on the CP-optimized surface as the HF limit is approached.

Appendix 1

Estimation of H-bonds Strength in Optimized α-Helix

In this study I perform the estimate of the H-bonding strengths in the optimized α-helix (polyglycine) containing 16 peptide units. I do it to verify the validity of the Formula 2.13, Chapter 2.

I started with full optimization of the polyglycine constructed of 16 peptide units in the α-helical conformation at the B3LYP/D95** level. I took into account the fact that clockwise and counterclockwise coiled α-helical polyglycines have the same energies. One by one, starting from the CH3-CO-NH- end, I then changed the backbone dihedrals of each residue to their negative values. By doing that in each step I was breaking two consecutive N-H···O hydrogen bonds (one in case of the first and the last residues) and obtaining a system of two helices (coiled clockwise and counterclockwise) which differ from initial helix by the absence of two hydrogen bonds without any other significant changes to the initial geometry. Figure A1.1 shows one of the steps when the part containing 9 peptide unites is coiled in the opposite direction relatively to the rest of the helix (7 peptide unites). At each step I performed single point calculations for these geometries at the B3LYP/D95** level. This gave me a set of equations that allowed me to estimate all H-bonding energies in this helix. I check my results against the values obtained using an equation similar to Equation 2.13. Since I did not perform the complete optimization of the partially recoiled structures the H-bonding enthalpies were not

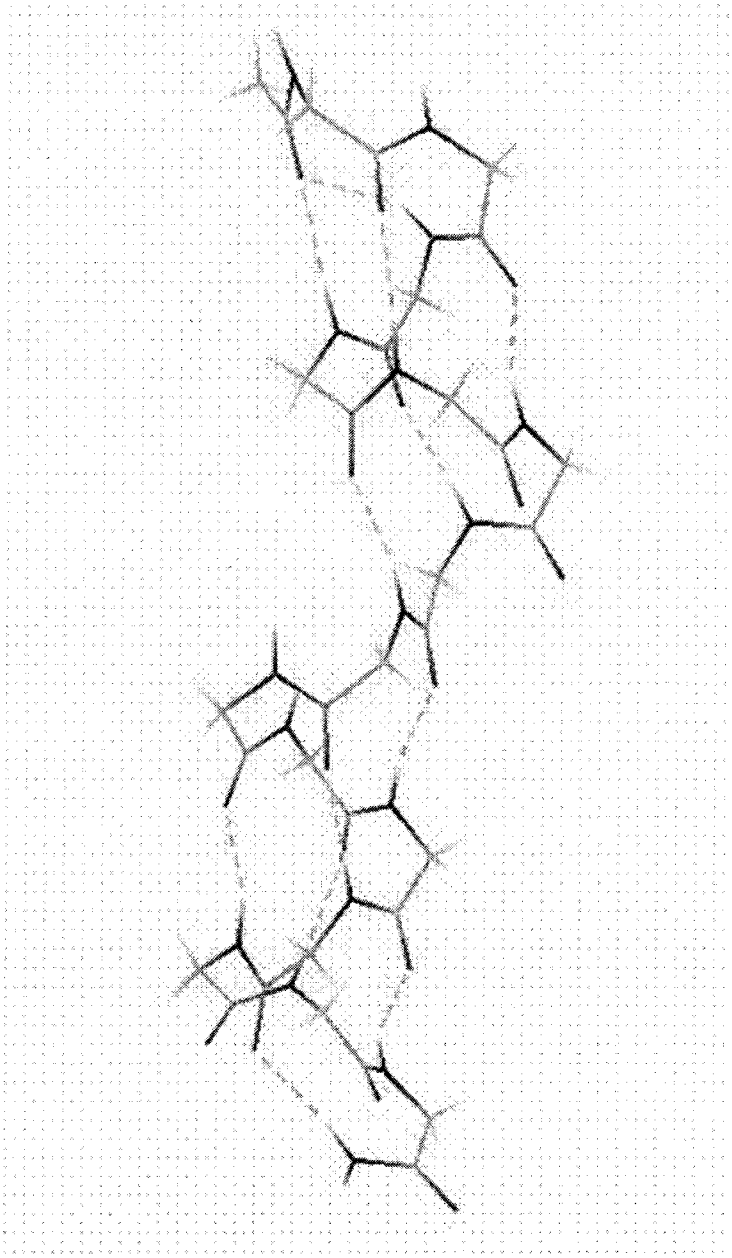


Figure A1.1. Example of the conformation studied. The top part of this α-helix is coiled in one direction and its bottom part in the opposite direction. As one can see the two H-bonds in the middle part of the helix are missing.

obtained in this study. First I thought to modify Equation 2.13 for the H-bonding energies in α-helical polyglycines without zero-point energy correction and CP-correction in order to match the data. But when I compared the average energy values for the H-bond types in this study with the data presented in Table 2.6, Chapter 2 (the CP-corrected enthalpies of the hydrogen bonds in formamide chains calculated using Equation 2.13) they were relatively close to each other. I believe this is mainly due to the steric strain that exists in α-helices making each H-bond weaker than the corresponding bond in formamide chains. So I decided to use Equation 2.13 without any changes since at this point we only need this data for comparison.

Figures A1.2 and A1.3 present the estimated hydrogen bonding energies and distances for the three chains in the α-helical polyglycine. One can clearly see the effect of cooperativity on the strength and the length of the H-bonds in this system. The middle hydrogen bonds in each chain are shorter and stronger than the terminal ones. This once again illustrates that the simple pair-wise potentials are not capable to model the hydrogen bonding in the systems that contain H-bonded chains. Though the hydrogen bonding energies in the optimized helix and the values obtained with the Equation 2.13 are close to each other, the differences are larger at the CH_3-CO- end of the helix. This is probably due to the fact that the first H-bond is bifurcated. I believe that the correctly designed formula should work even better for the long H-bonded chains since the effects of the ends in that case may not be that dramatic. I should once again point out that if one is to calculate the CP-corrected H-bonding enthalpies for the α-helix, all the values

should lie above the ones calculated using Equation 2.13, as one would expect because of the steric strain that exists in the helices.

I should also note that the H-bond lengths in the optimized α-helical polyglycine are very close to those obtained by Robert Wieczorek for the α-helical polyalanine (Figure A1.4).[1] One can also see from this figure that most of the hydrogen bonds in the polyalanine are slightly longer than in the polyglycine, indicating that there is relatively more steric strain in the former.

[1] Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.

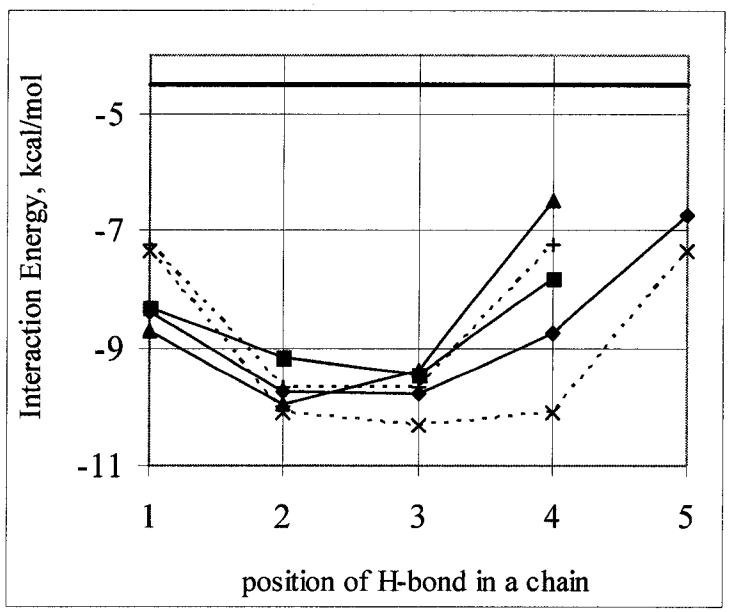


Figure A1.2. H-bonding energies for the α-helical polyglycine containing 16 peptide unites (solid lines) compared to the energies calculated using Equation 2.13 (dashed lines). Thick solid line represents the H-bonding enthalpy in the formamide dimer.

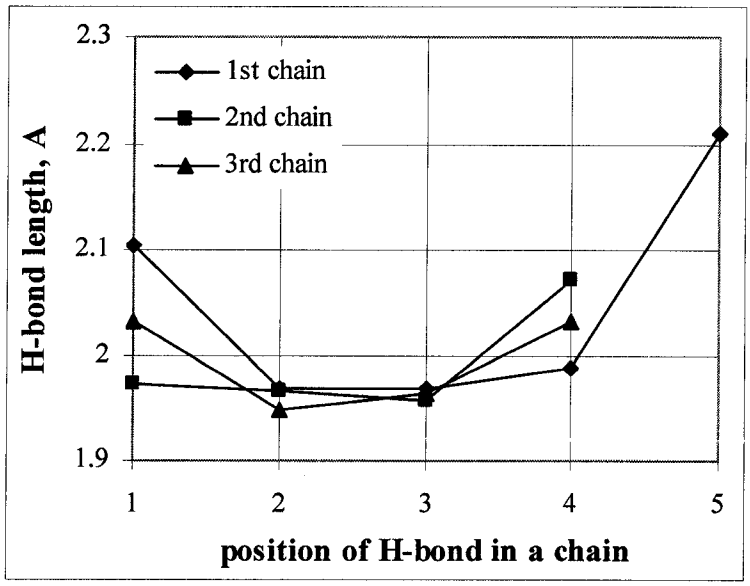


Figure A1.3. H-bonding distances in the three chains of the α-helical polyglycine containing 16 peptide unites.

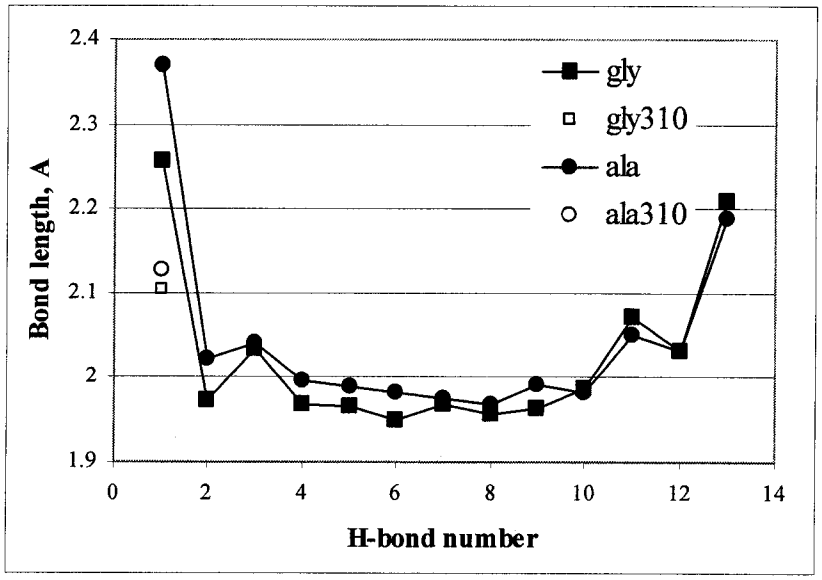


Figure A1.4. Comparison of the H-bonding lengths for the α-helical polyglycines and polyalanines. Note that the first H-bonds from the CH_3-CO- end in both sets are bifurcated.

BIBLIOGRAPHY

Chapter 1

Introduction

(1) *Protein Structure: Determination, Analysis, and Applications for Drag Discovery*; Chasman, D. E., Ed.; Marcel Dekker, Inc.: New York and Basel, 2003.

(2) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868.

(3) Friesner, R. A.; Gunn, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 315.

(4) Weiner, S.; Kollman, P.; Nguyen, D.; Case, D. *J. Comput. Chem.* **1986**, *7*, 230; McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*,7743.

(5) Brooks, C. R.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187; Mackerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

(6) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins* **2000**, *39*, 252; Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474; Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849.

(7) Dykstra, C. E. *Theochem* **1996**, *362*, 1.

(8) Van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M., Jr. *J. Phys. Chem. B* **2000**, *104*, 9554.

(9) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.

(10) Guo, H.; Gresh, N. Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.

(11) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

(12) Scheiner, S. *Hydrogen Bonding*; Oxford University Press: New York, 1997.

(13) Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714.

(14) Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806.

(15) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1996**, *100*, 9638.

(16) Turi, L.; Dannenberg, J. J. *Chem. Mater* **1994**, *6*, 1313.

(17) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1992**, *96*, 5819.

(18) Kang, Y. K. J. *Phys. Chem. B* **2000**, *104*, 8321; for NMA see also Aleman, C. *J. Phys. Chem. A* **2001**, *105*, 6717.

(19) Creighton, T. E. *Proteins: Structure and Molecular Properties*; W. H. Freeman and Co.: New York, 1993.

(20) Boys, S. F.; Bernardi, F *Molecular Physics*, **1970**, *19*, 553.

(21) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

(22) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewsk, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; J. Tomasi; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.;

Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03*, Gaussian, Inc.: Pittsburgh PA, 2003.

(24) Tsuzuky, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906.

(25) (a) Lozynski, M; Rusinska-Roszak, D. *J. Phys. Chem.* **1998**, *102*, 2899. (b) Tsuzuly, S.; Luthai, P. H. *J. Chem. Phys.* **2001**, *114*, 11906. (c) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705.

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

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

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewsk, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; J. Tomasi; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen,

W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople J. A. *Gaussian 98*, Gaussian, Inc.: Pittsburgh PA, 1998-2001.

(29) LINDA. Scientific Computing Associates: New Haven.

Chapter 2.

Cooperativity in Amide Hydrogen Bonding Chains. The Relationships between Energy of H-Bond, its Position and H-Bonding Chain Length in Peptides and Protein Folding

(1) (a) Dill, K. A. *Biochem.* **1990**, *29*, 7133. (b) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868. (c) Schuster, P.; Wolschann, P. *Monatsh. Chem.* **1999**, *130*, 947. (d) Kaya, H.; Chan, H. S. *Phys. Rev. Lett.* **2000**, *85*, 4823. (e) Skolnick, J.; Kolinski, A. *Comput. in Sci. Eng.* **2001**, *3*, 40. (f) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499. (g) Mirny, L.; Shakhnovich, E. *Annu. Rev. of Biophys. Biomol. Struct.* **2001**, *30*, 361. (h) Chasse, G. A.; Rodriguez, A. M.; Mak, M. L.; Deretey, E.; Perczel, A.; Sosa, C. P.; Enriz, R. D.; Csizmadia, I. G. *Theochem* **2001**, *537*, 319. (i) Arteca, G. A.; Reimann, C. T.; Tapia, O. *Mass Spectrom. Rev.* **2002**, *20*, 402.

(2) *Acc. Chem. Res.* **1998**, *31*, entire issue.

(3) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M. J. *J. Phys. Chem. B* **2000**, *104*, 9554.

(4) King, B. F.; Weinhold, F. *J. Chem. Phys.* **1995**, *103*, 333.

(5) Vendruscolo, M.; Domany, E. *J. Chem. Phys.* **1998**, *109*, 11101.

(6) Kennedy, R. J.; Tsang, K.-Y.; Kemp, D. S. . *J. Am. Chem. Soc.* **2002**, *124*, 934.

(7) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, 123, 5313.

(8) Lin, J.-Q.; Juo, S.-W.; Wu, Lin, J.-Q.; Juo, S.-W.; Wu, Y.-D. *J. Comput. Chem.* **2002**, *23*,1551.

(9) Zhao, Y.-L.; Wu, Y.-D. . *J. Am. Chem. Soc.* **2002**, *124*, 1570.

(10) Viswanathan, R.; Asensio, A.; Dannnenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205.

(11) Jaravine, V. A.; Alexandrescu, A. T.; Grzesiek, S. *Protein Sci.* **2001**, *10*, 943.

(12) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.

(13) (a) Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1748. (b) Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451. (c) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273. (d) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104. (e) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J Phys. Chem. B* **2000**, *104*, 9746. (f) Ludwig, R. *J. Mol. Liq.* **2000**, *84*, 65. (g) Cabaleiro-Lago, E. M.; Otero, J. R. *J. Chem. Phys.* **2002**, *117*, 1621.

(14) Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3950.

(15) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024. See also Chapter 6 on BSSE.

(16) Kang, Y. K. *J. Phys. Chem. B* **2000**, *104*, 8321.

(17) Hoeflich, K. P.; Tsukita, S.; Hicks, L.; Kay, C. M.; Tsukita, Sh.; Ikura, M. *Biochemistry* **2003**, *42*, 11634.

(18) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339.

(19) (a) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Acta Crystallogr., Sect. B: Struct. Sci.* **1998**, *B54*, 50. (b) Gilli, G.; Bertolasi, V.; Ferretti, V. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *B49*, 564. (c) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Chem.-Eur. J.* **1996**, *2*, 925.

(20) Dannenberg, J. J.; Haskamp, L.; Masunov, A. *J. Phys. Chem. A* **1999**, *103*, 7083.

(21) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A*, **2003**, *107*, 6688.

(22) Applequist, J.; Mahr, T. G. *J. Am. Chem. Soc* **1966**, *88*, 5419.

(23) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Molec. Biol.* **1987**, *44*, 97.

(24) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.

(25) Salvador, P.; Kobko, N.; Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14190.

Chapter 3

Hydrogen Bonding and Relative Stability of Polyglycine Conformers

(1) (a) Dill, K. A., *Biochem.* **1990**, *29*, 7133. (b) Friesner, R. A.; Gunn, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 315. (c) Dobson, C. M.; Šali, A.; Karplus, M., *Angew. Chem. Int. Ed.* **1998**, *37*, 868. (d) *Accounts of Chemical Research* **1998**, *31 (11)*, entire issue.

(2) See, for example, (a) Tobias, D. J.; Brooks, C. L., III *Biochemistry* **1991**, *30*, 6059. (b) Wang, Y.; Kuczera, K. *J. Phys. Chem. B* **1997**, *101*, 5205. (c) Hudgins,

R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974. (d) Hudgins, R. R.; Jarrold, M. F. *J. Phys. Chem. B* **2000**, *104*, 2154. (e) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 14065.

(3) See, for example, (a) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605. (b) Kim, K.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 12952.

(4) Viguera, A. R.; Serrano, L. *Biochemistry* **1995**, *34*, 8771.

(5) Iprota, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Am.Chem. Soc.* **2001,** *123*, 3311.

(6) Park, C.; Goddard, W. A., III *J. Phys. Chem. B* **2000**, *104*, 7784.

(7) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.

(8) See, for example, (a) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriat, I.; Muir, T.; Gellman, S. H. *PNAS* **2001**, *98 (21)*, 12015. (b) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124. (c) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198.

(10) Schafer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272.

(11) Iprota, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *114 (6)*, 2541.

(12) See for example, Sirois, S.; Proynov, E. I.; Nguyen, D. T.; Salahub, D. R. *J. Chem. Phys.* **1997**, *107 (17)*, 6770.

(13) Creighton, T. E. *Proteins: Structure and Molecular Properties*; W. H. Freeman and Co.: New York, 1993.

(14) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Freisner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.

(15) Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. *J. Am. Chem. Soc.* **1993**, *115*, 11590.

(16) Topol, I. A.; Stanley, K. B.; Tang, T.-H.; Perczel, A.; Rashin, A.; Csizmadia, I. G. *J. Am. Chem. Soc.* **2001**, *123*, 6054.

(17) (a) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873. (b) Hanson, P.; Martinez, G.; Millhauser, G.; Formaggio, F.; Crisma, M.; Toniolo, C.; Vita, C. *J. Am. Chem. Soc.* **1996**, *118*, 271. (c) Bolin, K. A.; Millhauser, G. L. *Acc. Chem. Res.* **1999**, *32*, 1027.

(18) Long, H. W. Tycko, R. *J. Am. Chem. Soc.* **1998**, *120*, 7039.

(19) Sudha, R.; Kohtani, M.; Breaux, G. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2004**, *126*, 2777.

(20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3909.

Chapter 4

Comparison between Vibrational Coupling through Hydrogen Bonds and Covalent Bonds. Implications for Peptide Vibrational Spectra

(1) Yoder, G.; Pancoska, P.; Keiderling, T. A. *Biochemistry* **1997**, *36*, 15123.

(2) Krimm, S.; Reisdorf, W. C., Jr. *Faraday Discuss.* **1995**, *Volume Date 1994, 99*, 181.

(3) Yoder, G.; Keiderling, T. A.; Formaggio, F.; Crisma, M.; Toniolo, C. *Biopolymers* **1995**, *35*, 103.

(4) Manas, E. S.; Getahun, Z.; Wright, W. W.; GeGrado, W. F.; Vanderkooi, F. M. *J. Am. Chem. Soc.* **2000**, *122*, 9883.

(5) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628.

(6) (a) Chirgadze, Y. N.; Nevskaya, N. A. *Dokl. Akad. Nauk SSSR* **1973**, *208*, 447. (b) Chirgadze, Y. N.; Nevskaya, N. A. *Biopolymers* **1976**, *15*, 627.

(7) Cha, S.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 740.

(8) Rubtsov, I. V.; Hochstrasser, R. M. *J. Phys. Chem. B* **2002**, *106*, 9165.

(9) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem.* **2003,** *107*, 10389.

(10) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348.

(11) Fillaux, F.; De Loze, C. *J. Chim. Phys. Physicochim. Biol.* **1972**, *69*, 36.

(12) Mehler, E. L. *J. Am. Chem. Soc.* **1980**, *102*, 4051.

(13) Bour, P.; Keiderling, T. A. *J. Am. Chem. Soc.* **1993**, *115*, 9602.

(14) Aleman, C.; Navas, J. J.; Munoz-Guerra, S. *J. Phys. Chem.* **1995**, *99*, 17653.

(15) Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283.

(16) Torii, H.; Tatsumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309.

(17) Torii, H.; Tasumi, M. *AIP Conf. Proc.* **1998**, *430*, 719.

(18) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142.

(19) Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2000**, *53*, 380.

(20) Evans, J. C. *J. Chem. Phys.* **1954**, *22*, 1228.

(21) Ludwit, R.; Reis, O.; Winter, R. *J. Phys. Chem. B* **1998**, *102*, 9312.

(22) Eberhardt, E. S.; Raines, R. T. *J. Am. Chem. Soc.* **1994**, *116*, 2149.

(23) Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T. *Biopolymers* **2001**, *59*, 396.

(24) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727.

(25) Brauner, J. W.; Dugan, C.; Mendelsohn, R. *J. Am. Chem. Soc.* **2000**, *122*, 677.

(26) Fang, C.; Wang, J.; Kim, Y. S.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B., III; Decatur, S. M.; Hochstrasser, R. M. *J. Phys. Chem. B* **2004**, *108*, 10415.

(27) Barber-Armstrong, W.; Donaldson, T.; Wijesooriya, H.; Silva, R. A. G. D.; Decatur, S. M. *J Am. Chem. Soc.* **2004**, *126*, 2339.

(28) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R. A. G. D.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.

Chapter 5

Cooperative Hydrogen-Bonding in Adenine-Thymine and Guanine-Cytosine Base Pairs

(1) Clementi, E.; Mehl, J.; Von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.

(2) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

(3) Sponer, J.; Hobza, P.; Leszczynski, J. *Comput. Chem.* **1996**, *1*, 185.

(4) Sponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365.

(5) Brameld, K. A.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1999**, *121*, 985.

(6) Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, *101*, 4851.

(7) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.-Eur. J.* **1999**, *5*, 3581.

(8) Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1999**, *103*, 2528.

(9) Dannenberg, J. J.; Tomasz, M. *J. Am. Chem. Soc.* **2000**, *122*, 2062.

(10) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117.

(11) Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M. *J. Phys. Chem. B* **2000**, *104*, 6286.

(12) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

(13) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.

(14) (a) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348. (b) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 10389.

(15) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.

(16) Suhai, S. *Int. J. Quantum Chem.* **1994**, *52*, 395.

(17) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273.

(18) Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451.

(19) Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1748.

(20) (a) Wu, Y.-D.; Zhao, Y.-L. *J. Am. Chem. Soc.* **2001**, *123*, 5313. (b) Zhao, Y.-L.; Wu, Y.-D. *J. Am. Chem. Soc.* **2002**, *124*, 1570.

(21) Belosludov, R. V.; Li, Z.-Q.; Kawazoe, Y. *Mol. Eng.* **1999**, *8*, 105.

(22) Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806.

(23) Dovesi, R.; Causa, M.; Orlando, R.; Roetti, C.; Saunders, V. R. *J. Chem. Phys.* **1990**, *92*, 7402.

(24) Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714.

(25) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1992**, *96*, 5819.

(26) Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3950.

(27) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640.

(28) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.

(29) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993**, *97*, 12197.

(30) Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 6775.

Chapter 6

Effect of Basis Set Superposition Error upon Ab Initio Calculations of Organic Transition States

(1) (a) Boys, S. F.; Bernardi, F. *Mol. Phys.*, **1970**, *19*, 553. (b) Meunier, A.; Levy, B.; Berthier, G. *Theor. Chim. Acta* **1973**, *29*, 49. (c) Jansen, H. B.; Ross, P. *Chem. Phys. Lett.*, **1969**, *3*, 40.

(2) For example, see: (a) Mayer, I.; Surjan, P. R., *Int. J. Quantum Chem.* **1989**, *36*, 225. (b) Almlöf, J.; Taylor, P. R. *J. Chem. Phys.* **1987**, *86*, 553. (c) Saebø, W.; Pulay, P. *J. Chem. Phys.* **1988**, *88*, 1884 and references cited. (d) Mayer, I.; Valiron, P. *J. Chem. Phys.* **1998**, *109*, 3360. (e) Famulari, A.; Specchio, R.; Sironi, M.; Raimondi, M., J. *Chem. Phys.* **1998**, *108*, 3296. (f) Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*, 157. (g) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251. (h) Famulari, A.; Raimondi, M.; Sironi, M.; Gianinetti, E. *Chem. Phys.* **1998**, *232*, 275.

(3) (a) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1984**, *82*, 2418. (b) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer III, H. F. *J. Chem. Phys.* **1986**, *84*, 2279. (c) Szalewicz, K.; Cole, S. J.; Kolos, W.; Bartlett, R. J. *J. Chem. Phys.* **1988,** *89*, 3662. (d) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993,** 97, 2488. (e) van Duijneveldt - van Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Ab Initio Methods in Quantum Chemistry*, Lawley, K. P., Ed.; John Wiley & Sons: New York, 1987; Vol.

II. (f) Gutowski, M.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Chem. Phys.* **1993,** *98*, 4728. (g) Cook, D. B.; Sordo, J. A.; Sordo, T. L. *Int. J. Quant. Chem.* **1993**, *48,* 375. (h) van Duijneveldt, F. B.; van Duijneveldt - van Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94,* 1873. (i) Van Duijneveldt, F. B. in *Molecular Interaction: From van der Waals to Strongly Bound Complexes;* Scheiner, S., Ed.; Wiley: Chichester, U.K., 1997, 81-104. (j) Tolosa, L.; Olivares del Valle, F. J. *THEOCHEM* **1994**, *118,* 109. (k) Collins, J. R.; Gallup, G. A. *Chem. Phys. Letters* **1986**, *123*, 56.

(4) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

(5) (a) Bouteiller, Y.; Behrouz, H. *J. Chem. Phys.* **1992,** *96*, 6033. (b) Del Bene, J. E.; Mettee, H. D. *J. Phys. Chem.* **1991**, *95*, 5387. (c) Leclercq, J. M.; Allavena, M.; Bouteiller, Y. *J. Chem. Phys.*, **1983**, *78*, 4606.

(6) (a) Daza, M. C.; Dobado, J. A.; Molina, J. M.; Salvador, P.; Duran, M.; Villaveces, J. L. *J. Chem. Phys.* **1999**, 11806. (b) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640. (c) Van Mourik, T.; Price, S. L.; Clary, D. C. *J. Phys. Chem. A* **1999**, *103*, 1611. (d) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372. (e) Novoa, J. J.; Planas, M. *Chem. Phys. Lett.* **1998**, *285*, 186. (f) Schuetz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997. (g) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575.

(7) Duran, M.; Simon, S.; Salvador, P.; Dannenberg, J. J. *Code to Optimize a Molecular System Including Counterpoise Correction* Version 4.1 (December 1999).

(8) The B3PW91 method combines Becke's 3-parameter functional (Ref. 3, Chapter 1) with the non-local correlation provided by the Perdew-Wang expression: Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(9) Sodupe, M.; Rios, R.; Branchadell, V.; Nichols, T.; Oliva, A.; Dannenberg J. J. *J. Am. Chem. Soc.* **1997**, *119*, 4232.

(10) Rowley, D.; Steiner, H. *Discuss. Faraday Soc.* **1951**, *10*, 198.

(11) Dannenberg, J. J.; Liotard, D., in preparation.

(12) Goldstein, E.; Beno, B.; Houk, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 6036.

(13) Dannenberg, J. J.; Baer, B. *J. Am. Chem. Soc.* **1987**, *109*, 292.

(14) (a) Sevin, A.; Yu, H. T.; Evleth, E. M. *THEOCHEM* **1983**, *104*, 163. (b) Evleth, E. M.; Cao, C. Z.; Kassab, E.; Sevin, A. *Chem. Phys. Lett.* **1984**, *109*, 45.

(15) Lendvay, G.; Mayer, I. *Chem. Phys. Lett.* **1998**, *297*, 365.

(16) Taken from Trotman-Dickenson, A. F.; Milne, G. S. *Tables of Bimolecular Gas Reactions*; NSRDS-NBS 9, U. S. Department of Commerce, 1967.

(17) (a) Steckler, R.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys.* **1986**, *84*, 6712. (b) Hu, W. P.; Liu, Y. P.; Truhlar, D. G. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1715.

(18) See also on the problem of the choice of fragments: Chalasinski, G.; Szczesniak, M. *Chem. Rev.* **2000**, *100*, 4227.

(19) (a) Nanayakkara, A. A.; Balint-Kurti, G. G.; Williams, I. H. *J. Phys. Chem.* **1992**, *96*, 3662. (b) Tooley, D. W.; Anderson, J. G. *J. Phys. Chem.* **1989**, *93*, 1049.

Appendix 1

(1) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.