Secondary Structure Motifs in Proteins: A QTAIM Study of the α-Helix

by

Shenna LaPointe

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2008

Your file Votre référence
ISBN: 978-0-494-50169-6
Our file Notre référence
ISBN: 978-0-494-50169-6

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages
- Examiners Signature Page (pii)
- Dalhousie Library Copyright Agreement (piii)

Appendices
- Copyright Releases (if applicable)

To my parents, Eugene and Joan LaPointe

Table of Contents

List of Figures

List of Tables

Abstract

This thesis describes the application of high-level ab initio molecular orbital calculations to the study of protein secondary structure. The focus of the study is a polyalanine peptide in the α-helix conformation. Amino acid substitutions are made at the centre of the peptide to determine the effect of different residues on the hydrogen bond network of the α-helix. The hydrogen bond network is characterized according to the quantum theory of atoms in molecules. Characterization of the hydrogen bond network reveals three hydrogen bonding interactions that contribute to the stabilization of α-helices. Also, the results suggest that in some cases, differences in helix propensities may be explained by contributions from the hydrogen bond network.

List of Abbreviations and Symbols Used

Abbreviations

B3	Becke's three-parameter exchange functional
B971	Exchange correlation functional proposed by Handy, Tozer and coworkers
BCP	Bond critical point
CCP	Cage critical point
CI	Configuration interaction
CIS	Configuration interaction using singles
CID	Configuration interaction using doubles
CISD	Configuration interaction using singles and doubles
DFT	Density functional theory
GTO	Gaussian-type orbital
HBCP	Hydrogen bond critical point
HF	Hartree-Fock
LCAO	Linear combination of atomic orbitals
LYP	Lee-Yang-Parr correlation functional
MO	Molecular orbital
MP*n*	MPPT applied to the *n*th order correction
MPPT	Møller-Plesset perturbation theory
NCP	Nuclear critical point
NMR	Nuclear magnetic resonance
PES	Potential energy surface
QCISD	Quadratic configuration interaction using singles and doubles
QTAIM	Quantum theory of atoms in molecules
RCP	Ring critical point
RHF	Restricted HF
SCF	Self-consistent-field
STO	Slater-type orbital
STO-*n*G	Minimal basis set utilizing a contraction of *n* GTOs
UHF	Unrestricted HF

Symbols

Ψ	Molecular wavefunction
$\hat{H}$	Hamiltonian operator
$\hat{T}$	Kinetic energy operator
$\hat{V}$	Potential energy operator
$\mathbf{r}$	Vector describing positions of all electrons
r_{ij}	Distance between electron i and electron j
R_{AB}	Distance between nucleus A and nucleus B
θ, ϕ	Polar angles
Z_A	Atomic number of atom A
N	Number of electrons
$\hat{H}_{elect}$	Electronic Hamiltonian operator
E	Energy
ψ	Spatial orbital
$\hat{F}$	Hartree-Fock operator
ε_i	Energy of the ith orbital
$\hat{H}^{core}$	One electron Hamiltonian
$\hat{J}$	Coulomb operator
$\hat{K}$	Exchange operator
Φ	Trial wave function
P_{ij}	Permutation operator
c	Orbital expansion coefficient
d	Basis set expansion coefficient
φ	Atomic orbital
ϕ	Basis function
$(\mu\upsilon\|\lambda\sigma)$	Two electron integral
$F_{\mu\upsilon}$	Element of the Fock matrix
$S_{\mu\upsilon}$	Element of the overlap matrix

$P_{\lambda\sigma}$	Element of the density matrix
E_{XC}	Exchange-correlation energy functional
λ	Perturbation parameter; electron density curvature
ω	Rank of a critical point in the electron density
σ	Signature of a critical point in the electron density
$\mathbf{\Lambda}$	Diagonalized Hessian matrix of the electron density
Ω	Union of an attractor and its basin in QTAIM
$\mathbf{g}$	Molecular energy gradient with respect to internal coordinates
$\mathbf{q}$	Internal coordinates
S	Surface
ρ	Electron density
$\nabla\rho$	Electron density gradient
$\nabla^2\rho$	Laplacian of the electron density
ε	Ellipticity
$\mathbf{H}$	Hessian matrix
q	Electronic charge

Acknowledgements

First and foremost, I would like to acknowledge my supervisor, Dr. Russell J. Boyd. I say with certainty that I would not have completed this thesis without his encouragement. I am deeply grateful for his kindness and generosity.

I am also thankful to my supervisor, Dr. Donald F. Weaver for putting up with me as I switched through projects and programs.

I would like to thank the members of my supervisory committee including Dr. T. Bruce Grindley, Dr. Laura Turculet and Dr. Peng Zhang for their helpful input.

I am very appreciative of the support from the members of the Boyd and Weaver computational groups. Discussions with them have helped considerably with research and presentations. I would like to give a special thanks to Sarah Whittleton, who has been my labmate for more than six years. Working in a group with her has been incredibly fun.

I also wish to thank the Natural Sciences and Engineering Research Council of Canada and Dalhousie University for financial support.

Finally, I would like to thank my parents, Gene and Joan, my sister, Gina, and two other important family members, Steve and Lyndsay. Your support and encouragement has made me stronger.

Chapter 1. Introduction

1.1 Introduction

Peptides and proteins are organic polymers made up of amino acids that occur in and are vital to every living cell. Keratins, actin, myosin, and collagen are the main components of hair, skin, muscles and tendons, which provide structure, support and protection to multicelled organisms. Hemoglobin, myoglobin and various lipoproteins transport oxygen and other molecules throughout the body and into or out of cells. Enzymes, hormones, antibodies and globulins catalyze and regulate the body chemistry. It is clear that in order to understand the mechanics of life, it is necessary to understand proteins that perform the essential processes and constitute the fundamental physical structure of living things.

The properties of a protein, and therefore its biological function, depend on the conformations that the molecule adopts.[1] Therefore, in order to understand in detail how a protein works it is necessary to obtain its complete three dimensional structure. Unlike most synthetic polymers, which can adopt many different spatial arrangements, proteins usually exist in a single conformation. The most widely used methods of obtaining information about protein structure are X-ray crystallography and nuclear magnetic resonance (NMR). However, crystal structures are available for only a fraction of known proteins. Furthermore, new protein sequences are being determined at a greater rate than protein structures are being determined experimentally. Fortunately, many methods exist that have been shown to be useful in the prediction of protein structure. One tool that is very popular in the prediction of protein structure is computational chemistry.

This thesis uses quantum mechanical calculations in order to obtain information about the structure of proteins. Specifically, the quantum theory of atoms in molecules (QTAIM) is used to investigate the intramolecular hydrogen bond network involved in the stabilization of secondary structural elements of proteins.

1.2 Basic Principles of Protein Structure

1.2.1 Amino Acids

Amino acids are molecules containing both amine and carboxyl functional groups. While there are hundreds of different amino acids, there are only 20 twenty α-amino acids that act as the building blocks of proteins. Each α-amino acid, with the exception of proline, has an amino group and a carboxyl group attached to the same tetrahedral carbon (Cα) as shown in Figure 1.1. Under physiological conditions, the amino group will be protonated and the carboxyl group will be unprotonated. The tetrahedral α-carbon is also attached to hydrogen and an R-group, which distinguishes α-amino acids from one another. Glycine's R-group is a hydrogen atom, making it the only α-amino acid that does not exhibit chirality. The α-carbon of all the other α-amino acids has four distinct substituents and is therefore chiral. All α-amino acids exhibit the same absolute steric conformation as L-glyceraldehyde and thus are L-α-amino acids. D-α-amino acids exist, but are not commonly found in proteins. For the remainder of this thesis, all discussion of amino acids will be in reference to α-amino acids.

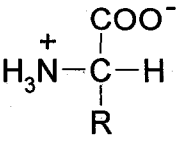

Figure 1.1 General structure of an α-amino acid. The center C is the α-C and R represents the various side chains.

The structure of each amino acid is unique because each has a different R-group or side chain. Amino acids are classified by the properties of their side chains. All twenty amino acids are illustrated in Figure 1.2. It is common to divide the amino acids into polar and non-polar groups. The non-polar amino acids are alanine, cysteine, glycine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, tyrosine and valine. The polar amino acids are arginine, asparagine, aspartic acid, glutamine, glutamic acid, histadine, lysine, serine and threonine. Amino acids with ionizable side chains form another subgroup of amino acids: aspartic acid, glutamic acid, histadine, cysteine, lysine, tyrosine and arginine. Proline is unique because it is the only amino acid whose side chain forms a bond with the amino nitrogen. The different sizes, shapes, hydrogen bonding capabilities and charge distributions of the side chains contribute to the large range of functions carried out by proteins.

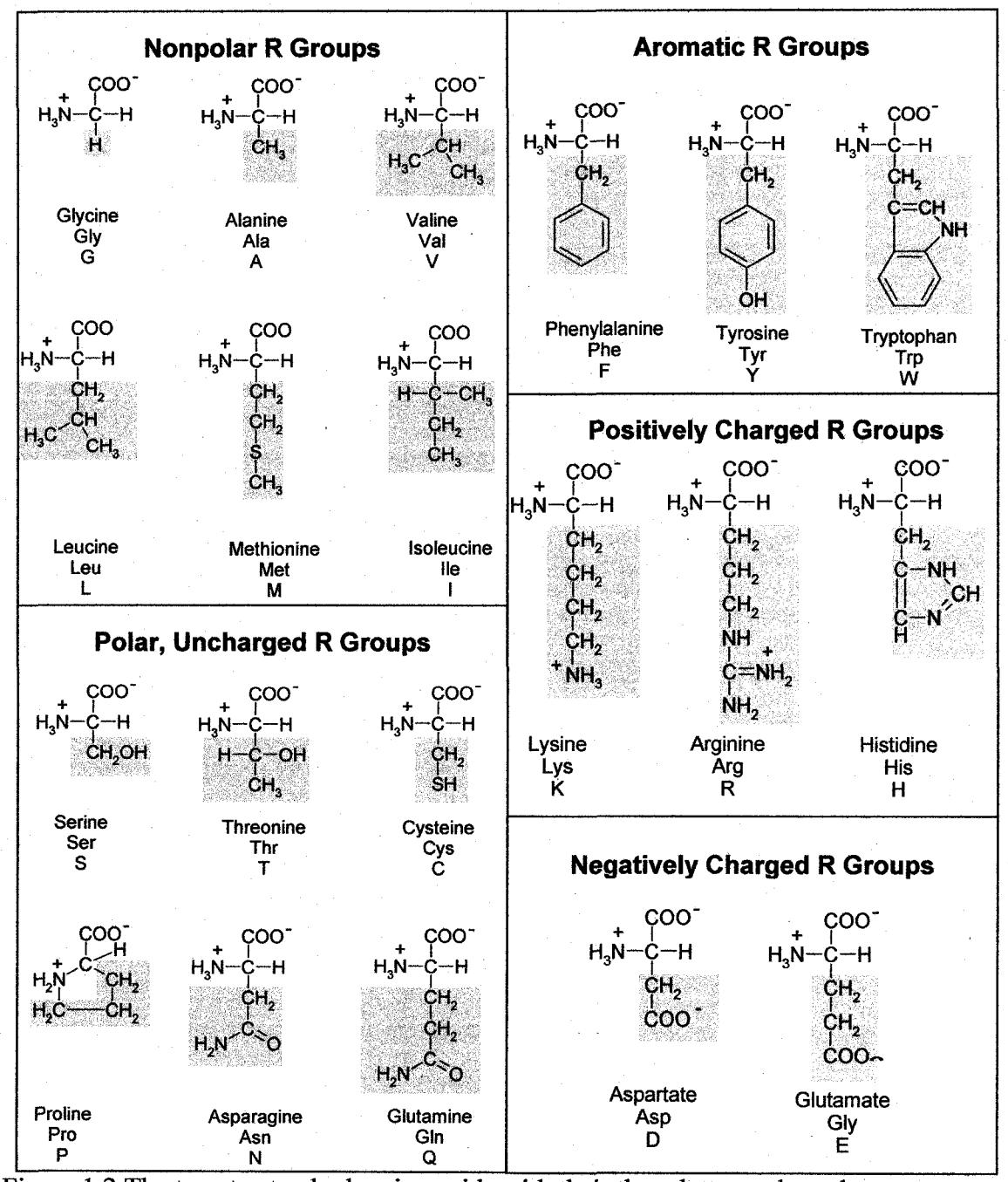

Figure 1.2 The twenty standard amino acids with their three letter and one letter abbreviations. Amino acids are grouped based on characteristics of the side chain and side chains are highlighted in pink.

1.2.2 Peptides

Amino acids are joined together via peptide bonds in order to form peptides. A peptide bond is a covalent bond between the carbon atom of the carboxy group of one

amino acid and the nitrogen atom of the amino group of another amino acid. The dehydration reaction that occurs to form a peptide bond is shown in Figure 1.3. Peptides vary in size with the number of amino acids joined together. Oligopeptides are small peptides consisting of two to twelve amino acids. Peptides with more amino acids are referred to as polypeptides. The regular structure that forms when amino acids are joined together is referred to as the peptide backbone. A peptide backbone is illustrated in Figure 1.4 and consists of a repeated sequence of the amide N, Cα, and the carbonyl C. Each peptide has an N-terminus (amino-terminus) with a free amino group and a C-terminus (carboxy-terminus) with a free carboxyl group. The amino acid sequence of a peptide is always presented in the N to C direction.

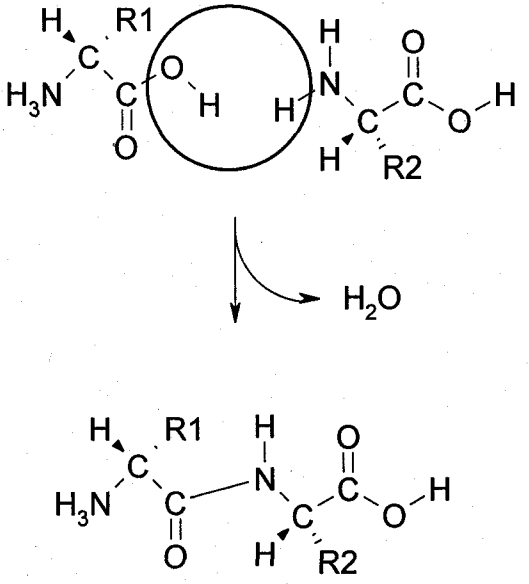

Figure 1.3 Example of a dehydration reaction forming a peptide bond between two amino acids. Water is released.

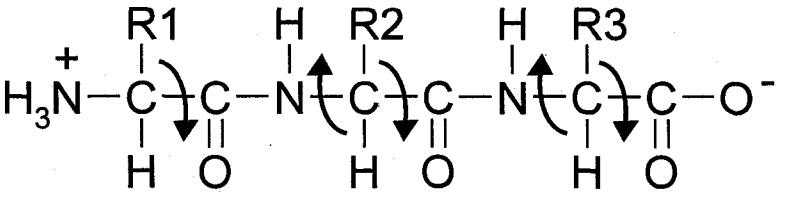

Figure 1.4 Peptide backbone example. Peptide is shown from N-terminus to C-terminus. Arrows represent where rotation may occur to change the conformation of the peptide. Down arrows represent rotation about dihedral angle φ. Up arrows represent rotation about dihedral angle ψ.

1.2.3 Proteins

1.2.3.1 Primary Structure

A protein is a large molecule containing one or more polypeptides. The structure of a protein can be described on four distinct levels. The primary structure of a protein refers to the number and order of amino acids present. The sequence of amino acids in each protein is determined by the gene it is encoded by and determines the unique properties of the protein. The primary structure of a protein is often represented by using the standard three letter abbreviations of the amino acids. For example, the hormone insulin consists of two polypeptide chains and the primary structure representation of both chains is shown in Figure 1.5.

Chain 1 GLY- ILE -VAL- GLU -GLN -CYS -CYS -THR- SER -ILE -CYS- SER -LEU - TYR -GLN -LEU -GLU -ASN -TYR -CYS -ASN

Chain 2 PHE -VAL -ASN-GLN -HIS -LEU -CYS- GLY- ASP -HIS -LEU- VAL- GLU- ALA -LEU- TYR -LEU- VAL- CYS- GLY- GLU- ARG -GLY- PHE -PHE -TYR - THR -PRO -LYS -THR

Figure 1.5 Primary structure of polypeptide chains making up insulin. Amino acids are represented by their three letter abbreviations.

1.2.3.2 Secondary Structure

While the peptide bonds that join amino acids together are rather rigid, the dihedral angles ψ and φ illustrated in Figure 1.4 have a range of possible values, controlling the proteins three dimensional structure. The conformation of the peptide backbone is referred to as its secondary structure.

The most common secondary structure found in proteins is the α-helix, illustrated in Figure 1.6. The backbone of the protein in an α-helix is arranged in a helical structure with the amino acid side chains positioned at the outside of the helix. Each amino acid residue results in a 100° turn in the helix and a translation of approximately 1.5 Å along the helical axis. There are 3.6 amino acid residues per turn, meaning that the helical structure repeats itself every 5.4 Å along the helix axis.

The tightly packed helix is stabilized by hydrogen bonds between the carbonyl group of each peptide bond and the amine group of the peptide bond four amino acids below it in the helix. The result of the arrangement of peptide bonds is that all of the partially negative carbonyl groups point along the helical axis towards the C-terminus and the partially positive amine nitrogens point towards the N-terminus. The aggregate effect is a separation of charge referred to as the helix dipole. The helix dipole destabilizes the helix but is often compensated for by capping the N-terminus end with a negatively charged amino acid such as glutamic acid and neutralizing the dipole.

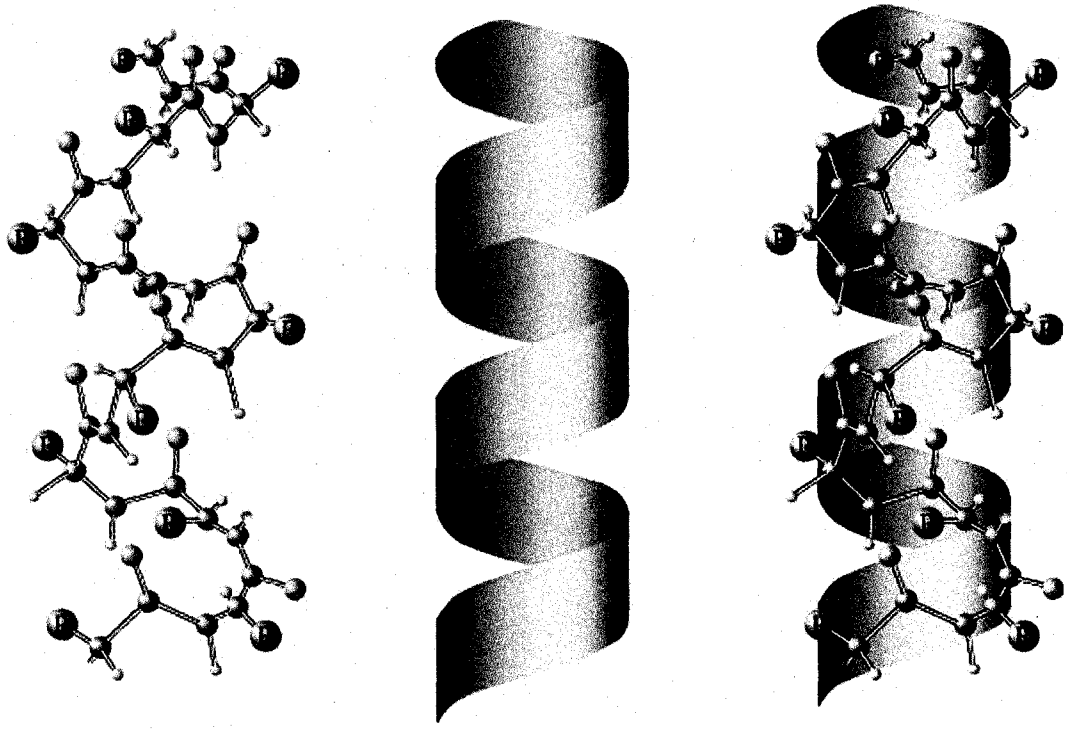

Figure 1.6 Example of an α-helix segment. On the left is a ball-and-stick model, in the centre is a ribbon model and on the right, the two models are superimposed. Figure is modified from Bishop, M. *An Introduction to Chemistry by Mark Bishop*, Chiral Publishing, Accessed October 30, 2008 (http://preparatorychemistry.com/).

The second commonly occurring secondary structure is the β-sheet. The β-sheet is characterized by two or more adjacent amino acid sequences within the same protein arranged in an alternating orientation. The amide groups in one strand can form hydrogen bonds with the carbonyl groups of its neighbor stabilizing the beta sheet. β-sheets can be parallel when the N to C direction is the same for each strand, or anti-parallel when the N to C direction is reversed (Figure 1.7).

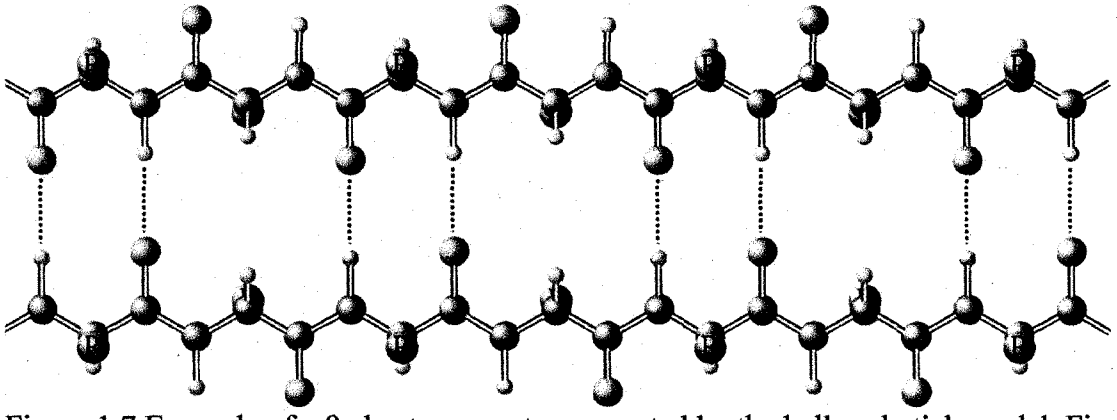

Figure 1.7 Example of a β-sheet segment represented by the ball-and-stick model. Figure is modified from Bishop, M. *An Introduction to Chemistry by Mark Bishop*, Chiral Publishing, Accessed October 30, 2008 (http://preparatorychemistry.com/).

1.2.3.3 Tertiary Structure

α-helices and β-sheets are connected by regions referred to as random coils or loops. Random coils do not adopt regular structures; however, common conformations do exist, such as the β-turn, when the chain makes a sharp, 180° reversal. Such turns or loops allow the secondary structure of the protein to fold back upon itself or twist to give the protein its three dimensional shape. The hydrophobic effect is often a driving force behind folding. The protein chain will fold so that non-polar side chains are hidden within the structure and polar side chains are exposed on the surface. The overall spatial organization and conformation of the protein is referred to as its tertiary structure. An example of how secondary structural elements organize to form the tertiary structure of insulin can be seen in Figure 1.8. Chain 1 is represented in blue and its secondary structure consists of two small α-helices. The α-helices are organized so that they lie side by side, giving chain 1 its tertiary structure. The secondary structure of chain 2, which is represented in green, consists of one larger α-helix. The tertiary structure of chain 2 is folded in a V-shape.

Figure 1.8 Ribbon diagram of insulin. Thick coiled ribbons represent α-helices and thin tubes represent random coils. Reproduced from Brooks, D.W. *Teaching and Research Website*. Accessed October 30, 2008. (http://dwb4.unl.edu/Chem/CHEM869K/CHEM869KLinks/main.chem.ohiou.edu/~wathen/chem302/protein.html)

The tertiary structure of a protein can be stabilized by non-covalent bonding interactions such as van der Waals, ionic and hydrogen bonding between groups from both the peptide backbone and the side chains. Some protein tertiary structures can also be stabilized by the formation of disulphide bonds between cysteine residues. For example, in insulin there is a disulphide bond within chain 1 between cysteine residues in each of the small α-helices.

1.2.3.4 Quaternary Structure

The final level of protein structure is quaternary structure. Often, protein complexes, which are assemblies of protein subunits, carry out important biological functions. Quaternary structure refers to the arrangement into which protein subunits assemble. The organization of the subunits is stabilized mainly by noncovalent

interactions, such as van der Waals interactions and hydrogen and ionic bonding. Stabilization also occurs less commonly by disulphide bonds between cysteine residues in different subunits. Because insulin consists of two polypeptide chains, it also exhibits quaternary structure. Chain 1 and chain 2 are organized as shown in Figure 1.8. There are two disulphide bonds between chain 1 and chain 2 that stabilize this structure.

1.3 Protein Structure Prediction

There is a large gap between the availability of protein sequences and their three dimensional structures. Currently, the Universal Protein Resource Knowledgebase lists more than seven million sequences.[2] However, the Protein Data Bank lists less than 60,000 experimentally determined structures.[3] As previously mentioned, the three dimensional structure of a protein determines its biological function. Many families of proteins exist that have an unknown function. One example of why protein structure prediction is important is that it may provide insight into the functions of proteins that are not understood.

It has been demonstrated by Anfinsen that the amino acid sequence of a protein contains enough information to determine how a protein folds into its three dimensional structure.[4] Consequently, protein structure prediction from the amino acid sequence has been researched for decades and is considered a fundamental scientific problem. There are two broad classes of methods used to predict protein structure: comparative modeling and *ab initio* modeling. Secondary structure prediction is a common step in both approaches[5, 6] and is considered to be very important in the understanding of protein structure. Each method is discussed briefly here.

1.3.1 Comparative Modeling

Comparative modeling or homology modeling makes predictions based on the structures of proteins that are already known. The sequence of an unknown protein is aligned with the sequence of a known protein and if a homology of more than 35% exists, that is, if the sequences are very similar, the three dimensional fold of the protein is considered to be the same.[7] Comparative modeling has proven to be very efficient and is applicable for a majority of proteins.[8] However, there are a number of reasons that *ab initio* modeling is useful. Comparative modeling does not help fundamental understanding of the mechanisms of protein structure formation. Also, there are a large number of proteins that do not show any homology with proteins that are already known. Finally, proteins with a high degree of homology may still adopt a different structure, which implies that comparative modeling is not completely reliable.

1.3.2 Ab Initio Modeling

Ab initio protein modeling attempts to predict protein structure without the use of previously solved structures. Instead, protein structure prediction is based on physical properties. *Ab initio* methods begin with an initial model, define some energy function or force field and then search the energy landscape to locate the lowest energy structure.[9] Locating the lowest energy structure usually involves a conformational scan or a simulation of folding process.

Many *ab initio* modeling studies have used molecular mechanics and molecular dynamics to provide useful information about protein structure. However, in order to more accurately describe the interactions that are important to protein folding and

structure, many studies have investigated smaller fragments of proteins using higher level theory in order to gain more insight.

1.3.3 Secondary Structure Prediction

It has been suggested that predicting a protein's secondary structure is a useful first step in determining its overall three dimensional structure.[10] This stems from the viewpoint that protein folding is initiated by the formation of secondary structural elements, followed by the slower arrangement of the tertiary structure.[11] While there is debate over the validity of this viewpoint,[12] it is generally thought that understanding the formation mechanism of secondary structural elements is important to understanding overall protein folding.[13]

It is a commonly thought that the formation of secondary structural elements cannot be completely understood without a rigorous quantitative description of all the physical forces that stabilize them.[14] Therefore, it is very important to accurately describe the various interactions that act to stabilize secondary structure, such as hydrophobic interactions, hydrogen bonds and electrostatic interactions. Quantum mechanical calculations have been very useful at describing these interactions in small model peptides. However, small models are limited in that they are unable to capture neighboring effects exerted by nearby interactions on the protein backbone, such as hydrogen bonding that is thought to contribute to the stabilization of α-helices and β-sheets.[15]

In the past, quantum mechanical calculations have been limited to small peptide models because of the large computational cost of large molecular systems. Recently, it has become possible to perform high level calculations on polypeptides.[15-17]

Consequently, there is much work to be done on the characterization of hydrogen bonding and hydrogen bond networks of the secondary structural elements of proteins.

This thesis uses quantum mechanical calculations to investigate the hydrogen bond network of secondary structural elements. The results from this work provide insight into the importance of hydrogen bonding in the formation of secondary structures, an important step in understanding why proteins adopt certain structures. Chapter 2 discusses the theoretical background necessary to carry out this investigation, while Chapter 3 outlines the details and results of the study. Finally, Chapter 4 highlights the global conclusions drawn and recommends future work that may further the understanding of protein structure.

Chapter 2. Theoretical Methods

2.1 Introduction

A fundamental concept of chemistry is that matter is composed of atoms.[18] Atoms are, in turn, composed of charged particles [19] whose behavior is governed by the laws of quantum mechanics.[20] Consequently, quantum mechanics can be used to study the behavior of atoms, and therefore the properties of matter.

Quantum chemistry models chemical systems using mathematical descriptions based on the laws of quantum mechanics. Specifically, the energy and many other important properties of a system are determined by the wave function, which is obtained by solving the Schrödinger equation.[21, 22] In practice, it is only possible to solve the Schrödinger equation exactly for the simplest chemical systems. However, a variety of methods are available that generate approximate solutions.

While the discipline of quantum chemistry is based around approximate solutions, some of these solutions are extremely accurate and have been very effective in solving complex chemical and biological problems.[23] Properties that can be determined include energies, reaction pathways, vibrational frequencies, and charge distributions, among many others. Theoretical chemistry does not replace experimental chemistry; however, it is increasingly recognized as a valuable complement. For example, theoretical chemistry may be used to investigate an unstable reaction intermediate that is difficult to study experimentally. Furthermore, computational chemistry has become so reliable that some chemists employ it before beginning an experimental project to determine the feasibility of the project or to guide the project.[24]

This thesis uses theoretical calculations to provide valuable insights into chemical bonding and the electronic interactions that determine geometric configurations. In this chapter, some of the fundamental methods of quantum chemistry, including Hartree-Fock (HF), configuration interaction (CI), Møller-Plesset perturbation theory (MPPT) and density functional theory (DFT), will be discussed. Particular attention will be given to DFT, which is implemented in this thesis. This outline is not comprehensive; more detail on the described methods and others can be found in standard quantum chemistry textbooks.[22-29]

2.2 The Schrödinger Equation

In quantum mechanics, the properties of a system are explicitly defined by the wave function[21, 22], which can be obtained from the Schrödinger equation:

$$\hat{H}\Psi = E\Psi \tag{2.2.1}$$

In this equation, Ψ is the wave function, which exists for any system according to the first postulate of quantum mechanics.[30] $\hat{H}$ is the Hamiltonian operator that acts on Ψ and returns E, the energy of the system. The Hamiltonian is an energy operator and is often expressed as

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{ne} \tag{2.2.2}$$

The five terms in Equation (2.2.2) describe the kinetic energies of the electrons and the nuclei, the interelectronic and internuclear repulsions, and the attraction between the electrons and the nuclei, respectively.

Because wave functions are functions of the coordinates and motions of all the electrons in the system, it is extremely difficult to express accurate wave functions for many particle systems. In fact, there are only a few systems for which the Schrödinger

equation can be solved exactly. For most real systems, only an approximate solution to the Schrödinger equation can be obtained through the application of various assumptions.

2.3 Fundamental Approximations

2.3.1 The Born-Oppenheimer Approximation

The first approximation that is made to simplify the problem is to separate the electronic and nuclear motions. It is reasonable to assume that the nuclei are motionless because they are much more massive and move much more slowly than the electrons. This is referred to as the Born-Oppenheimer approximation[25] and under this approximation, the goal is to solve the electronic Schrödinger equation:

$$\hat{H}_{elect}\Psi = E_{elect}\Psi \tag{2.3.1}$$

where

$$\hat{H}_{elect} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} \tag{2.3.2}$$

There is no longer a contribution from the kinetic energy of the nuclei because they are fixed. The term accounting for the internuclear repulsion has also disappeared because this contribution can be added to the energy as a classical term later.

2.3.2 The Orbital Approximation

The wave function of an N-electron system is expressed in terms of the coordinates of the N electrons ($\Psi(1,2,\ldots,N)$). The motions of the electrons are assumed to be independent, which allows each electron to be assigned a different spatial function. This is known as the orbital approximation.[25] Under this approximation, the wave function can be written as a product of one electron wave functions:

$$\Psi(1,2,\ldots,N) = \psi_1(1)\psi_2(2)\ldots\psi_N(N) = \prod_1^N \psi_i(i) \tag{2.3.3}$$

The form of the wave function given by Equation (2.3.1) describes the spatial orbitals of the electrons but it does not describe their spin. Electrons are characterized by a spin quantum number that is either ½ or -½. In order to describe the electrons more completely, each spatial orbital should be assigned a spin function. These spin functions are denoted by α and β. A spin orbital is the product of a spatial function and a spin function denoted by $\psi_i\alpha(i)$ or $\psi_i\beta(i)$. The wave function, as originally proposed by Hartree, is a Hartree product of spin orbitals:

$$\Psi(1,2,...N) = \psi_1\alpha(1)\psi_2\beta(2)\psi_3\alpha(3)\psi_4\beta(4)... \tag{2.3.4}$$

The form of the wave function expressed in Equation (2.3.4) is deficient. A Hartree product of spin orbitals is not consistent with the indistinguishability of electrons. Because electrons are indistinguishable, interchanging any two electrons should not change the probability density $\Psi^2(1,2,\ldots,N)$. This implies the equality

$$\Psi(1,2,...i,j,...,N) = \pm\Psi(1,2,...j,i,...,N) \tag{2.3.5}$$

From experimental evidence we know that electrons are antisymmetric.[28] This is known as the antisymmetry principle and can be written as

$$\hat{P}_{ij}\Psi(1,2,...,i,j,...,N) = \Psi(1,2,...,j,i,...,N) = -\Psi(1,2,...,i,j,...,N) \tag{2.3.6}$$

where $\hat{P}_{ij}$ is the permutation operator that interchanges electrons i and j. The antisymmetry principle also leads to the Pauli principle that states that no two electrons can be assigned to identical spin orbitals. If two electrons were assigned identical spin orbitals, they would be symmetric with respect to interchange.

As previously mentioned, a Hartree product does not satisfy the antisymmetry principle. Expressing the wave function as a Slater determinant guarantees that the

antisymmetry principle, along with the Pauli principle, are satisfied. A Slater determinant representing a $2N$ electron wave function has the form

$$\Psi(1,2,...,2N) = [(2N)!]^{-1/2} \begin{vmatrix} \psi_1\alpha(1) & \psi_1\beta(1) & \psi_2\alpha(1) & \cdots & \cdots & \psi_N\beta(1) \\ \psi_1\alpha(2) & \psi_1\beta(2) & \psi_2\alpha(2) & \cdots & \cdots & \psi_N\beta(2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \psi_1\alpha(2N) & \psi_1\beta(2N) & \psi_2\alpha(2N) & \cdots & \cdots & \psi_N\beta(2N) \end{vmatrix} \quad (2.3.7)$$

The constant, $[(2N)!]^{-1/2}$ is the normalization factor. This factor is present as a consequence of the assumption that the wave functions are normalized. Normalized wave functions obey the condition

$$\int \Psi_i \Psi_j d\bar{r} = 1 \quad \text{for } i = j \quad (2.3.8)$$

where $d\bar{r}$ is a generalized $3N$-dimensional volume element.

In a Slater determinant, the rows correspond to electrons while the columns correspond to spin orbitals. Interchanging any two rows, which corresponds to interchanging electrons, leads to a change in the sign of the determinant. Also, if any two columns are the same, which would correspond to two electrons being assigned to the same spin orbital, the determinant becomes equal to zero.

2.4 Hartree-Fock Theory

2.4.1 The Variation Theorem

The form of the wave function is given by Equation (2.3.7). The variation theorem[25] provides a method for obtaining the "best" wave function. Given a trial wave function, Φ, the equation for the energy becomes

$$\hat{H}\Phi = E\Phi \quad (2.4.1)$$

The energy can be found by multiplying Equation (2.4.1) on the left by Φ and integrating. After these operations and rearranging, Equation (2.4.2) becomes

$$E = \frac{\int \Phi \hat{H} \Phi d\bar{r}}{\int \Phi \Phi d\bar{r}} \tag{2.4.2}$$

Written using bra-ket notation, Equation (2.4.2) is equivalent to

$$E = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \tag{2.4.3}$$

The variation theorem states that for any trial wave function Φ, the energy of the system obeys the following inequality:

$$E_{\Phi} = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq E_0 \tag{2.4.4}$$

where E_0 is the ground state energy. In order to find the best wave function, parameters of the trial wave function are varied until the energy reaches a minimum. Hartree-Fock is a variational method which gives an upper bound estimate of the true ground state energy.

2.4.2 The Hartree-Fock Equations

The Hartree-Fock equations are used to obtain an approximate wave function. If the form of the wave function is a Slater determinant, the spin orbitals must be determined. The spin orbitals are obtained by solving the Hartree-Fock equations. The goal is to find the ψ_i that result in an energy minimum and that obey the orthonormal condition. Orthonormal wave functions are normalized, as previously discussed, and orthogonal. Orthogonal wave functions obey the condition

$$\int \Psi_i \Psi_j d\bar{r} = 0 \quad \text{for } i \neq j \tag{2.4.5}$$

This is a constrained optimization that can be solved by the method of Lagrange multipliers.[25] The result is the set of Hartree-Fock equations:

$$\hat{F}\psi_i = \varepsilon_i \psi_i \qquad i = 1,2,\ldots,N \tag{2.4.6}$$

$\hat{F}$ is the Hartree-Fock Hamiltonian operator given by

$$\hat{F} = \hat{H}^{core} + \sum_{j=1}^{N} (2\hat{J}_j - \hat{K}_j) \tag{2.4.7}$$

The eigenfunctions of $\hat{F}$ are the best ψ_i. The eigenvalues, ε_i are the orbital energies associated with an electron in orbital ψ_i. $\hat{H}^{core}$ is the core Hamiltonian operator and the operators $\hat{J}_j$ and $\hat{K}_j$ are called the Coulomb operator and the exchange operator, respectively. These operators are defined in Equations (2.4.8), (2.4.9) and (2.4.10).

$$\hat{H}_i^{core} = -\frac{1}{2}\nabla_i^2 - \sum_{A}^{n} \frac{Z_A}{r_{Ai}} \tag{2.4.8}$$

$$\hat{J}_j(1) = \int \psi_j(2) \frac{1}{r_{12}} \psi_j(2) d\bar{r} \tag{2.4.9}$$

$$\hat{K}_j(1)\psi_i(1) = \left[\int \psi_j(2) \frac{1}{r_{12}} \psi_i(2) d\bar{r} \right] \psi_j(1) \tag{2.4.10}$$

$\hat{H}_i^{core}$ represents the motion of a single electron moving in the field of the bare nuclei. Z_A is the nuclear charge of nucleus A while r_{Ai} is the distance between electron i and nucleus A. $\hat{J}_j$ describes the potential experienced by electron 1 in the field of electron 2. $\hat{K}_j$ has no classical interpretation.

The Hartree-Fock equations must be solved iteratively because the operator $\hat{F}$ depends on the spin orbitals, ψ_i and the ψ_i depend on the operator. Normally, an initial guess is made for the ψ_i and this guess is used to find the operator. The new operator is then used to find a new set of ψ_i. This process continues until the orbitals do not change appreciatively. When this happens, they are said to be self-consistent with the field they generate. This is referred to as the self-consistent field (SCF) method or Hartree-Fock (HF) technique. It was first described by Hartree and Fock.

2.4.3 The Roothaan-Hall Equations

2.4.3.1 Closed Shell Systems

While the Hartree-Fock equations have been solve numerically for atoms, it is not practical to solve them for polyatomic systems. This problem can be made easier by expressing molecular orbitals (MOs) as a linear combination of one electron functions called basis functions:

$$\psi_i = \sum_{\mu=1}^{M} c_{\mu i} \varphi_\mu \tag{2.4.11}$$

where φ_μ are the basis functions and $c_{\mu i}$ are the coefficients which indicate the contribution of the basis function to the spin orbital.

The basis functions are used to model atomic orbitals this approximation is therefore called the linear combination of atomic orbitals (LCAO) approximation. Basis functions are used because it is not practical to use hydrogen-like atomic orbitals for molecular calculations.

The goal is now to calculate the coefficients ($c_{\mu i}$) that lead to an energy minimum and make the ψ_i orthonormal. These coefficients are found by solving the Roothaan-Hall equations:

$$\sum_{\upsilon=1}^{M}(F_{\mu\upsilon}-\varepsilon_i S_{\mu\upsilon})c_{\upsilon i}=0 \qquad \mu=1,2,\ldots,M \tag{2.4.12}$$

ε_i is the orbital energy of ψ_i. $S_{\mu\upsilon}$ is the overlap integral given by

$$S_{\mu\upsilon}=\int \varphi_\mu \varphi_\upsilon dr_1 dr_2 \tag{2.4.13}$$

and $F_{\mu\upsilon}$ is a matrix element of the Hartree-Fock operator give by

$$F_{\mu\upsilon}=H_{\mu\upsilon}+\sum_{\lambda=1}^{M}\sum_{\sigma=1}^{M}P_{\lambda\sigma}[(\mu\upsilon\,|\,\lambda\sigma)-\frac{1}{2}(\mu\lambda\,|\,\upsilon\sigma)] \tag{2.4.14}$$

$H_{\mu\upsilon}$ represents the energy of a single electron moving in the field of the bare nuclei and is expressed as

$$\left\langle\varphi_\mu\left|-\frac{1}{2}\nabla^2\right|\varphi_\upsilon\right\rangle-\sum_{A}^{nuclei}Z_A\left\langle\varphi_\mu\left|\frac{1}{r_A}\right|\varphi_\upsilon\right\rangle \tag{2.4.15}$$

$P_{\lambda\sigma}$ is an element of the density matrix given by

$$P_{\mu\upsilon}=2\sum_{i}^{occupied}c_{\mu i}c_{\upsilon i} \tag{2.4.16}$$

Finally, the $(\mu\upsilon|\lambda\sigma)$ are the two electron repulsion integrals that describe the repulsion between the two local product densities $\varphi_\mu\varphi_\upsilon$ and $\varphi_\lambda\varphi_\sigma$. They can be expressed

$$(\mu\upsilon\,|\,\lambda\sigma)=\int \varphi_\mu(1)\varphi_\upsilon(1)\frac{1}{r_{12}}\varphi_\mu(2)\varphi_\upsilon(2)dr_1 dr_2 \tag{2.4.17}$$

Because the Fock matrix depends on the expansion coefficients ($c_{\mu\nu}$), the SCF method must be used to solve these equations.

2.4.3.2 Open Shell Systems

The methods and equations discussed above are referred to as restricted Hartree-Fock (RHF) methods and are only valid when all electrons in a system are spin paired. When unpaired electrons are present in the system, it is more appropriate to use different spatial orbitals to describe electrons with different spin. This method is referred to as unrestricted Hartree-Fock (UHF). When electrons of α and β spin are assigned different spatial orbitals, there are two sets of molecular orbitals:

$$\psi_i^{\alpha} = \sum c_{\mu i}^{\alpha} \varphi_{\mu}$$
$$\psi_i^{\beta} = \sum c_{\mu i}^{\beta} \varphi_{\mu} \qquad (2.4.18)$$

It is also possible to use the restricted open shell Hartree-Fock method for open shell systems. In this method, spatial orbitals with paired electrons are treated with the RHF method while singly occupied orbitals are treated independently in a more complicated manner. This method does not account for the interaction between paired and unpaired electrons and is therefore not as accurate as UHF.

2.5 Basis Sets

When using the LCAO approach described above to construct MOs, the quality of the MOs and therefore the accuracy of the solution depend on the set of atomic orbitals used. The set of atomic orbitals, or basis functions, is referred to as the basis set.

The two most commonly used types of basis functions used in constructing basis sets are Slater-type orbitals[31] (STOs) and Gaussian-type orbitals[32] (GTOs).

2.5.1 Slater-Type Orbitals

Slater-type orbitals have the form:

$$\varphi^{STO} = NY_{l,m}(\theta,\phi)e^{-\xi r} \tag{2.5.1}$$

where N is a normalization constant, $Y_{l,m}(\theta,\phi)$ is an angular function and $e^{-\xi r}$ is a radial function. The variable r is the distance between the electron and the nucleus and the spatial extent of the orbital is controlled by ξ.

STOs are good approximations of atomic orbitals because they obey the electron-nucleus cusp condition and therefore accurately describe behavior near the nucleus. Also, as the distance between the electron and nucleus increases, STOs decay correctly, accurately describing the 'tail' of the wave function.[26] However, STOs are generally not used for theoretical calculations because some of the integrals required for the calculation of approximate solutions must be evaluated numerically and take considerable computational time. Instead, Gaussian-type orbitals (GTOs) are the most widely used form of basis functions.

2.5.2 Gaussian-Type Orbitals

Gaussian-type orbitals have the form:

$$\varphi^{GTO} = NY_{l,m}(\theta,\phi)e^{-\xi r^2} \tag{2.5.2}$$

where all symbols and variables have the same meanings defined for STOs. The advantage of GTOs is that the integrals involved in obtaining approximate solutions are relatively easy to evaluate. However, GTOs do not meet the electron-nucleus cusp condition and they decay too rapidly far from the nucleus. Therefore, they do not describe atomic orbitals as accurately as STOs.

In order to maintain the accuracy of STOs while retaining the computational efficiency of GTOs, it is common to use a linear combination of GTOs to represent an

STO. That is, each basis function (φ_μ) in Equation (2.4.11) can be represented by a series GTOs (g_i)[24]

$$\varphi_\mu = \sum_i d_{i\mu} g_i \tag{2.5.3}$$

where $d_{i\mu}$ are the coefficients. In order to maintain computational efficiency, the coefficients, $d_{i\mu}$, are chosen to best imitate STOs and are held fixed during the calculation. This is referred to as a contracted basis set where φ_μ are the basis functions representing atomic orbitals.

A minimal basis set uses the least number of basis functions possible to describe the atomic orbitals. Each type of core and valence orbital will be described by only one basis function. An example of a minimal basis set is STO-3G. Each basis function is represented by a contraction of three GTOs (g_i) and each atomic orbital is represented by one basis function. For example, hydrogen would be represented by only one basis function ($1s$) and oxygen would be represented by five basis functions ($1s$, $2s$, $2p_x$, $2p_y$, $2p_z$).

For more accurate results, it is necessary to increase the flexibility of the basis set by increasing the number of basis functions. A common way to expand the basis set is to double the number of basis functions to give a double-zeta basis set. Tripling the number of basis functions creates a triple-zeta basis set, and so on.

2.5.3 Split Valence Basis Sets

In addition to adding more basis functions, a common way to increase basis set flexibility is to use a split valence basis set. A split valence basis set describes core orbitals and valence orbitals differently. Core orbitals are treated minimally, with a single

basis function. Valence orbitals are treated with a double- or triple-zeta set. For example, a commonly used basis set is 6-31G. When considering atoms Li to F, the core orbitals (1*s*) will be described by one basis function, which is made up of a linear combination of six GTOs. The valence orbitals (2*s*, $2p_x$, $2p_y$, $2p_z$) will each be described by two basis functions. The first basis function will be a linear combination of three GTOs and the second will consist of a single GTO.

Additional flexibility can be added by describing the valence orbitals with three basis functions or using a triple-zeta split valence (for example, the 6-311G basis set).

2.5.4 Polarization Functions

It is possible to increase the accuracy of the basis set by including more Gaussians. Polarization functions are basis functions with higher angular momentum. For example, *p*-type functions can be included in the description of hydrogen atoms and *d*-type functions can be included in the description of heavier atoms, and so on. Polarization functions allow for the distortion of atomic orbitals in their molecular environments. For example, the electronic distribution around a bonded hydrogen atom is not expected to have spherical symmetry. However, the use of only *s*-type orbitals to describe hydrogen would result in a symmetric electron distribution. Including a *p*-type orbital in the description of hydrogen better represents the unsymmetrical atomic orbital. Including *d* functions to heavy atoms and *p* functions for hydrogen, in addition to the double-zeta basis set described above, results in the 6-31G(d,p) basis set.

2.5.5 Diffuse Functions

Diffuse functions can also be added to heavy atoms or basis sets. Diffuse functions have large radial distributions and therefore allow orbitals to occupy larger

regions in space. They are useful for describing molecules where electrons are loosely bound, such as anions. The inclusion of diffuse functions is denoted by a '+' where 6-31++G is a split valence double-zeta basis set with diffuse functions on both hydrogen and heavy atoms.

2.6 Electron Correlation

In Hartree-Fock theory, under the orbital approximation, the electrons move independently of each other. Only average electron repulsion is included in the calculation. However, it is energetically unfavorable for two electrons to come in close proximity to each other. Thus, the motion of the electrons is correlated. The energy associated with this correlation is called the correlation energy and is expressed as

$$E_{corr} = E_{exact} - E_{HF} \tag{2.6.1}$$

Hartree-Fock methods do not account for electron correlation. Electron correlation accounts for approximately 1% of the total energy for a given basis set. This 1% is often very important for describing chemical phenomena, such as the formation and breaking of bonds. In this section, methods that improve upon HF solutions by incorporating electron correlation effects are described.

2.6.1 Configuration Interaction

Configuration interaction (CI) accounts for electron correlation by including excited states in the description of the electronic state. While HF uses a single determinant to describe the wave function, CI is a post-Hartree-Fock method that expresses the total wave function as a linear combination of Slater determinants (Φ_i)[29]

$$\Psi = \sum_i c_i \Phi \tag{2.6.2}$$

where c_i is the expansion coefficient of the i_{th} determinant. The first determinant (Φ_0) is taken to be the HF determinant. Subsequent determinants are referred to as excited determinants. Excited determinants are generated by exciting electrons from the occupied orbitals of the HF determinant to the unoccupied or virtual orbitals. The CI wave function is represented as

$$\left|\Psi\right\rangle = c_0\left|\Psi_{HF}\right\rangle + \sum_{a,r} c_a^r \left|\Psi_a^r\right\rangle + \sum_{\substack{a<b,\\ r<s}} c_{a,b}^{r,s}\left|\Psi_{a,b}^{r,s}\right\rangle + \ldots \tag{2.6.3}$$

which can also be written as

$$\Psi = c_0\Phi_0 + \sum_{ar} c_a^r \Phi_a^r + \sum_{\substack{a<b\\ r<s}} c_{ab}^{rs}\Phi_{ab}^{rs} + \ldots \tag{2.6.4}$$

In Equation (2.6.4), Φ_a^r represents one electron being excited from an occupied orbital a to unoccupied orbital r and Φ_{ab}^{rs} is a double excitation, where electrons from orbitals a and b have been promoted to orbitals r and s.

A full CI expansion includes all the excitations and is most accurate. In fact, full CI gives the most accurate upper bound to the exact energy for a given basis set. For a complete basis set, the full CI energy is equal to the exact energy. However, full CI is computationally expensive. The number of excitations is very large, even with a minimal basis set, for all but the smallest systems. Usually, CI is truncated to include only certain excitations. For example, CIS included only single excitations while CID includes only doubles excitations and CISD includes both single and double excitations. Singly excited CI (CIS) does not improve the description of the ground state, by Briollouin's theorem, which states that singly excited determinants do not interact with the HF solution.[33] CISD, on the other hand, is a useful expansion and is often used as a compromise

between chemical accuracy and computational efficiency. The double excitations play an important role in calculating the correlation energy and are therefore important to include.

2.6.1.1 Size Consistency

There are some drawbacks to the truncated CI methods. One obvious drawback is that as the number of electrons in the system increases, so does the number of possible excitations that are being neglected. As a result, the accuracy deteriorates. Another problem with truncated CI is that it is not size consistent. That is, the energy of a system is not equal to the sum of its parts. For small systems, calculations including up to quadruple excitations can be effectively size-consistent. However, including triple and quadruple excitations can be computationally expensive. Therefore, for most truncated CI calculations, as the size of the system grows, the proportion of the correlation energy recovered decreases and the error associated with this energy decreases.

2.6.1.2 Quadratic Configuration Interaction

In an attempt to correct for the lack of size consistency in truncated CI expansions, Pople and coworkers introduced the quadratic configuration interaction with singles and doubles (QCISD) technique.[34] The QCISD expansion contains all single and double excitations as well as contributions from some quadruple excitations. Enough quadruple excitations are included to ensure that the method is size consistent. However, QCISD is not variational.

2.6.2 Møller-Plesset Perturbation Theory

While the CI method is variational and can be very accurate, it has the disadvantage that it is not size consistent. Møller-Plesset perturbation theory[35] is a method that systematically finds the correlation energy and is size-consistent. The electron correlation is treated as a perturbation to the HF Hamiltonian operator discussed above.

2.6.2.1 General Principles of Perturbation Theory

Perturbation theory uses the known eigenfunctions and eigenvalues of a simplified operator to approximate a solution to the exact operator. In other words, if in Equation (2.6.5) the eigenfunctions (ψ_n^0) and eigenvalues (E_n^0) for the operator $\hat{H}^0$ have been solved for a system and the system is perturbed slightly, perturbation theory can be used to approximate a solution to Equation (2.6.6) for the new eigenfunctions (ψ_n) and eigenvalues (E_n) for the operator $\hat{H}$.

$$\hat{H}^0\psi_n^0 = E_n^0\psi_n^0 \tag{2.6.5}$$

$$\hat{H}\psi_n = E_n\psi_n \tag{2.6.6}$$

In order to solve this problem, the exact Hamiltonian is written as the sum of two parts

$$\hat{H} = \hat{H}^0 + \lambda\hat{H}' \tag{2.6.7}$$

where $\hat{H}^0$ is the zeroeth part of the Hamiltonian with known eigenfunctions and eigenvalues and $\hat{H}'$ is the perturbation. The parameter λ is a device used to keep track of

terms. It is first taken to be a small number and is later set to 1 when it is no longer necessary.

The exact eigenfunctions and eigenvalues are expanded as a Taylor series in λ, expressed by Equations (2.6.8) and (2.6.9), respectively.

$$\psi_n = \psi_n^0 + \lambda\psi_n^1 + \lambda^2\psi_n^2 + \cdots \tag{2.6.8}$$

$$E_n = E_n^0 + \lambda E_n^1 + \lambda^2 E_n^2 + \cdots \tag{2.6.9}$$

In the above equations, ψ_n^1 is referred to as the first-order correction to the n^{th} eigenfunction, E_n^1 is referred to as the first-order correction to the n^{th} eigenvalue, ψ_n^2 and E_n^2 are the second-order corrections, and so on.

When Equations (2.6.7), (2.6.8) and (2.6.9) are substituted into Equation (2.6.6), and the products are expanded, the coefficients of equal powers of λ can be equated. The result is a set of equations that represent all of the higher orders of perturbation.

2.6.2.2 Møller-Plesset Perturbation Theory

In order to apply perturbation theory to the calculation of correlation energy, Møller and Plesset proposed choices for $\hat{H}^0$ and $\hat{H}'$.[35] The zeroeth-order part of the Hamiltonian and the perturbed part of the Hamiltonian are expressed in Equations (2.6.10) and (2.6.11), respectively. In these equations, $\hat{H}^{core}$, $\hat{J}_i$ and $\hat{K}_i$ are as defined in Equations (2.4.8), (2.4.9) and (2.4.10). The zeroeth order part of the Hamiltonian, the Hartree-Fock Hamiltonian operator, accounts only for the average electron-electron repulsion. However, when the perturbation term in Equation (2.6.11) is added, the result is the correct Hamiltonian, $\hat{H}$ because the perturbation represents the difference between

the Hartree-Fock averaged interelectronic interaction and the exact electron-electron interaction.

$$\hat{H}^0 = \sum_{i=1}^{N} \hat{H}^{core} + \sum_{j=1}^{N} (\hat{J}_i + \hat{K}_i)) \tag{2.6.10}$$

$$\hat{H}' = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} - \sum_{j=1}^{N} (\hat{J}_i + \hat{K}_i) \tag{2.6.11}$$

The application of this method is referred to as MP*n* where *n* is the order at which the perturbation is truncated. For example, MP2 includes first and second order corrections to the wave function and the energy, MP3 includes first, second and third order corrections, and so on. An MP∞ calculation is equivalent to a full CI calculation. When this theory is applied to the first order, it turns out that $E_0^{(0)} + E_0^{(1)} = E_{HF}$.[36] That is, MP1 returns the same energy as Hartree-Fock theory. Therefore, in order to improve upon the Hartree-Fock approximation, one must at least include the second order correction.

2.7 Density Functional Theory

Traditional methods discussed up until now, including Hartree-Fock (HF), configuration interaction (CI) and Møller-Plesset perturbation theory (MP*n*) can be computationally expensive for large molecules. This is due to the complexity of the *N*-electron wave function and the Schrödinger equation. Density functional theory (DFT) is a method that replaces the use of the wave function and the Schrödinger equation with the much simpler electron density. As a result, DFT is capable of calculating the electronic structure of larger molecules with significantly less computational effort.

2.7.1 The Hohenberg-Kohn Theory

The basic notion of DFT is that the ground state energy and all other molecular properties are uniquely determined by the electronic density, $\rho(x, y, z)$. This idea was present in the work of Thomas[37] and Fermi[38-40] in the 1920s. However, Hohenberg and Kohn were the first to formalize a proof in 1964.[41] The result of their proof, that the ground state energy E_0 is a functional of the density, is expressed in Equation (2.7.1). A functional, which is denoted by square brackets, maps a function to a value. A value for E_0 is associated with each density function ρ.

$$E_0 = E_0[\rho] \tag{2.7.1}$$

According to the Hohenberg-Kohn theorem, given the electron density function, it is possible to calculate the ground state energy along with all molecular properties. However, the theorem does not give the electron density or the form of the energy functional; the theorem merely states that such a functional exists. The Kohn-Sham theorem supplies both an expression for the ground state energy and a method for obtaining the electron density.

2.7.2 The Kohn-Sham Theorem

Kohn and Sham showed that the ground state energy of an N-electron system can be calculated by[42]

$$E_0[\rho] = -\frac{1}{2}\sum_{i=1}^{N}\langle\psi_i(1)|\nabla_1^2|\psi_i(1)\rangle - \sum_{\alpha}\int\frac{Z_\alpha\rho(1)}{r_{1\alpha}}d\upsilon_1 + \frac{1}{2}\int\int\frac{\rho(1)\rho(2)}{r_{12}} + E_{XC}[\rho] \tag{2.7.2}$$

where $\psi_i(1)$ are the Kohn-Sham orbitals, which are functions of the spatial coordinates of electron 1, and $E_{XC}[\rho]$ is the exchange correlation energy. The exchange correlation energy is known to be a function of the electron density because of the Hohenberg-Kohn

theorem. The first term in Equation (2.7.2) accounts for the kinetic energy of the system. The second term accounts for the interaction of the N electrons with each nucleus α of distance $r_{1\alpha}$ with charge Z_α. The third term accounts for the Coulomb interaction between the total charge distribution. Finally, the last term is the one that differs from the Hartree-Fock energy. Equation (2.7.2) does not include an exchange energy term, however, the exchange-correlation energy, $E_{XC}[\rho]$ includes both the exchange energy and the correlation energy.

Kohn and Sham also provided a method for finding the ground state electron probability density. The electron density is found from the Kohn-Sham orbitals ψ_i by

$$\rho = \sum_{i=1}^{N} |\psi_i|^2 \tag{2.7.3}$$

There is a constraint on the electron density because the integration over all space must be equal to the number of electrons in the system:

$$\int \rho(x,y,z)dxdydz = N \tag{2.7.4}$$

Therefore, the Kohn-Sham orbitals must give an energy minimum and obey the constraint in Equation (2.7.4). When the method of Lagrangian multipliers is applied to this constrained variation problem, the Kohn-Sham equations are derived:[36]

$$\hat{F}_{KS}(1)\psi_i(1) = \varepsilon_{i,KS}\psi_i(1) \tag{2.7.5}$$

where ψ_i are the Kohn-Sham orbitals, ε_i are the orbital energies and

$$\hat{F}_{KS} = -\frac{1}{2}\nabla_1^2 - \sum_\alpha \frac{Z_\alpha}{r_{1\alpha}} + \sum_{j=1}^{N} \hat{J}_j(1) + V_{XC}(1) \tag{2.7.6}$$

In the above equation, $\hat{J}_j$ is the Coulomb operator defined in Equation (2.4.9) and V_{XC} is the exchange correlation potential. The exchange correlation potential is the functional derivative of the exchange correlation functional:

$$V_{XC} = \frac{\delta E_{XC}[\rho]}{\delta \rho} \tag{2.7.7}$$

Therefore, if the exchange correlation functional is known, the exchange correlation potential can be easily found in order to solve the Kohn-Sham equations. However, the correct form of the functional for molecules is not known and so various approximations are used. These approximations can be tested by using them to run calculations and comparing the results to experimental values. However, there is no systematic way of improving the exchange-correlation functional and this is the main drawback of DFT.[43]

2.7.3 Solving the Kohn-Sham Equations

Most DFT programs solve the Kohn-Sham equations by expressing the Kohn-Sham orbitals in terms of a set of basis functions[44]

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu \tag{2.7.8}$$

where ϕ_μ are the basis functions and $c_{\mu i}$ are the coefficients in the expansion. Gaussian basis functions have been used for DFT calculations, however, Slater-type orbitals, which are discussed above, and numerical orbitals have also been employed. Numerical basis functions are generated by iteratively solving the Kohn-Sham equations for isolated atoms. This is done by an iterative procedure similar to that for solving the HF equations outlined previously.

Solving the Kohn-Sham equations using orbitals expanded in terms of basis functions amounts to finding the coefficients in the basis set expansion. If Equation

(2.7.8) is substituted into the Kohn-Sham equations, the result is a set of equations similar to the Roothaan-Hall equations:

$$\sum_{\upsilon=1}^{K}(\hat{K}_{\mu\upsilon}-\varepsilon_i S_{\mu\upsilon})c_{\mu i}=0 \qquad \mu=1,2,...,K \tag{2.7.9}$$

where ε_i are the orbital energies, $c_{\mu i}$ are the desired coefficients, $S_{\mu\upsilon}$ is the overlap integral defined in Equation (2.4.13) and

$$\hat{K}_{\mu\upsilon}=H_{\mu\upsilon}+\iint\frac{\phi_\mu(1)\rho(2)\phi_\upsilon(1)}{r_{12}}d\nu_1 d\nu_2+\int\phi_\mu(1)V_{XC}\phi_\upsilon(1)d\nu_1 \tag{2.7.10}$$

$H_{\mu\upsilon}$ represents the kinetic energy of an electron moving in the field of the bare nuclei. This is the same $H_{\mu\upsilon}$ defined by Equation (2.4.15). The second term represents the Coulomb repulsion and the third represents the exchange correlation effect.

Once an approximate $E_{XC}[\rho]$ is obtained, the set of Equations (2.7.9), like the Roothaan-Hall equations, are solved iteratively. An initial guess for the electron density ρ is used to calculate the approximate $E_{XC}[\rho]$, which is then used to calculate V_{XC}. The initial electron density is usually taken to be a superposition of atomic densities. V_{XC} is then used to solve the Kohn-Sham equations for an initial set of coefficients $c_{\mu i}$. These coefficients are used in the basis set expansion of the ψ_i, which are used in turn to get a better electron density. The process is repeated until the density and exchange correlation are converged. Once convergence is achieved, the electron density can be used to calculate the ground state energy.

It is important to note that the Kohn-Sham orbitals have no physical significance other than that they allow the electron density ρ to be calculated. Therefore, the Kohn-Sham orbital energies are not molecular orbital energies.

2.7.4 Key Similarities and Differences between DFT and HF

DFT and HF methods share similarities beyond their iterative procedures. The operator $\hat{K}_{\mu\upsilon}$ is very similar to the operator used in the Roothaan-Hall equations. The kinetic energy and the electron-nuclei attraction components are exactly the same. Also, if the density ρ that appears in the Coulombic repulsion operator (the second term in Equation (2.7.10)) is expressed in terms of some basis functions, then both the HF and DFT methods treat the electron repulsion in the same way.[44] The difference between the two operators is that while the operator expressed in Equation (2.4.10) contains a term which accounts for exchange, $\hat{K}_{\mu\upsilon}$ contains a term involving V_{XC}, which accounts for the effect of exchange and electron correlation.

Another key difference between DFT and HF is that DFT does not calculate the full wave function while HF does. DFT only calculates the total electronic energy and the overall electronic distribution. Because DFT optimizes the electron density and not the wave function, it is necessary to know how properties of interest depend on the density. Also, if the true E_{XC} was known, DFT would be an exact method while the HF method is only an approximation. Exact DFT is variational, like HF, however, the approximate DFT methods that are actually used due to the fact that E_{XC} is unknown are not variational.

2.7.5 Performance of DFT

DFT can be implemented in a number of different ways depending on the selection of the exchange correlation functional E_{XC} and the basis set. As previously mentioned, the main drawback of DFT is that the correct form of E_{XC} is not known and there is no systematic method for improving approximations of E_{XC}. The use of an

approximate E_{XC} also means that DFT is not variational. However, there are many advantages to DFT that make it a very useful method for electronic structure calculations.

The main advantage of DFT is its computational efficiency. In fact, DFT is the most cost effective method to achieve a given level of accuracy.[36] DFT is faster that both CI and MP2 and for large molecules and it is also faster than HF. Therefore, DFT makes it possible to study large molecules with moderate basis sets and methods that include electron correlation at a reasonable computational cost. This is not always possible with CI and MP2. Furthermore, depending on the exchange correlation function and basis set used, DFT can give competitive and sometimes superior results to *ab initio* calculations that include electron correlation such as MP2. Comparative studies can be performed by keeping the basis set constant and using the different methods for calculations. Examples of such studies have found that for energetics[45-48] and geometries[45, 47, 49], DFT can give mean absolute errors that are almost equal to or better in quality than more expensive correlation methods such as MP2 and CI, depending on the form of the exchange correlation functional. DFT especially outperforms HF and MP2 for transition metals.[50] Similar results have also been found when comparing DFT to HF and MP2 for calculating dipole moments.[51]

2.8 Potential Energy Surfaces

The theoretical methods described thus far have all applied the Born-Oppenheimer approximation, which separates electronic and nuclear motion. Consequently, the electronic energy of a system can be expressed as a function of its nuclear coordinates. A potential energy surface (PES) describes the energies associated with all possible configurations of the nuclei. Because chemical reactions usually involve

a reorganization of nuclei, information provided by a PES would be very useful in understanding chemical systems. Unfortunately, calculating an entire PES, for any of the theoretical methods discussed above, is only computationally feasible for very small systems.

However, the stationary points of a PES are often of primary interest. By definition, a stationary point is a point where the first derivative of the energy, with respect to the nuclear coordinates, is zero.[24] Generally, there are two types of stationary points of interest: minima and first order saddle points. A minimum on a PES corresponds to an equilibrium structure. A first order saddle point corresponds to a transition structure. Many problems of chemical interest involve characterization of equilibrium structures and/or transition states.

Locating stationary points on a PES is referred to as a geometry optimization. A geometry optimization is accomplished by changing a given geometry so that it approaches a stationary point. The commonly used quasi-Newton method[52, 53] makes changes to the nuclear coordinates (**q**) according to the following:

$$\Delta \mathbf{q} = -\mathbf{H}^{-1}\mathbf{g} \qquad (2.8.1)$$

where **H** is the Hessian matrix and **g** is the gradient of slope of the PES at a given point. Once certain criteria are met, the geometry is considered to be a stationary point.

Located stationary points can be characterized as minima, first order saddle points, or higher order saddle points by calculating vibrational frequencies. A vibrational frequency calculation involves calculating the second derivatives of the energy with respect to the nuclear coordinates. The second derivatives are expressed in the form of a Hessian matrix. If all the eigenvalues are positive, the stationary point is a minimum. If

one eigenvalue of the Hessian is negative, the stationary point is a first order saddle point. Two negative eigenvalues indicate a second order saddle point, and so on.

2.9 The Quantum Theory of Atoms in Molecules

For many decades, the transferability of atomic and functional group properties has been experimentally observed.[54] Furthermore, the fact that atoms and functional groups behave similarly from one molecule to another has been instrumental in the development of chemistry.[55] Consequently, it is desirable to understand or define an atom, within a molecule, in the context of the Schrödinger equation or the quantum theories described thus far.

The quantum theory of atoms in molecules (QTAIM), developed by Richard F.W. Bader and his co-workers[56], provides a solution, which partitions a molecular system based on a quantum observable, the electron density. From the topology of the electron density, a molecule can be uniquely divided into a set of atoms. Atomic properties, such as energy, charge and dipole moment, can then be derived by integrating their corresponding operators over the atomic volume. The resulting atomic properties can be summed to yield the value of that property for the entire system. The implication is that it is possible to partition an electronic property into individual atomic contributions. The mathematical basis for QTAIM can be found in Bader's book.[56] This section will outline some of the main concepts of QTAIM and will highlight the features of the theory that will be used in this thesis.

2.9.1 The Topology of the Electron Density

The attractive force of the nuclei on the electron density results in a substantial local maximum in the electron density, $\rho(\mathbf{r})$, at each nuclear position. However, the

topology of the electron density is also characterized by other critical points, specifically, minima and saddle points. The critical points in the electron density reveal valuable information about the chemistry of the system.

A critical point in the electron density is a point at which the first derivative, or gradient, of the electron density, $\nabla\rho(\mathbf{r})$, is equal to zero:

$$\nabla\rho = i\frac{\partial\rho}{\partial x} + j\frac{\partial\rho}{\partial y} + k\frac{\partial\rho}{\partial z} = \vec{0} \tag{2.9.1}$$

where $\vec{0}$ indicates that each partial derivative is equal to zero and not just their sum.

It is possible to characterize critical points as local maxima, minima and saddle points by calculating the second derivative of $\rho(\mathbf{r})$ at the point of interest. There are nine possible second derivatives, which can be represented by a diagonalized Hessian matrix. The diagonal form of the Hessian, $\mathbf{\Lambda}$, can be expressed as[55]:

$$\Lambda = \begin{pmatrix} \frac{\partial^2\rho}{\partial x^2} & 0 & 0 \\ 0 & \frac{\partial^2\rho}{\partial y^2} & 0 \\ 0 & 0 & \frac{\partial^2\rho}{\partial z^2} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \tag{2.9.2}$$

where λ_1, λ_2, and λ_3 are the curvatures of the electron density with respect to the axis system.

The curvatures are used to classify the critical points. Critical points are classified by assigning each a rank (ω) and a signature (σ). The rank is defined as the number of non-zero curvatures at the critical point. Any critical point with a rank lower than 3 is mathematically unstable and will disappear with any perturbation of the electron density caused by nuclear motion.[55] Therefore, only critical points with a rank of 3 are physically relevant. The signature is defined as the sum of the signs of the curvatures. That is, each

curvature contributes ±1 to the signature, depending on whether it is a positive or negative curvature. Therefore, there are four different mathematically stable critical points, which are symbolized as (ω, σ). Each critical point corresponds to an element of chemical structure. The four critical points are described below.

A (3, -3) critical point has three negative curvatures and represents a point at which $\rho(\mathbf{r})$ is local maximum. A (3, -3) critical point is referred to as a nuclear critical points (NCP) and corresponds to the position of a nucleus.

A (3, -1) critical point has two negative curvatures and is a point at which $\rho(\mathbf{r})$ is a maximum in the plane defined by the corresponding negative curvatures, and a minimum in the third axis, which is perpendicular to that plane. A (3, -1) critical point is a saddle points and is referred to as a bond critical point (BCP) and corresponds to the point where electron density between two nuclei is at a minimum.

A (3, +1) critical point has two positive curvatures and is a point at which $\rho(\mathbf{r})$ is a maximum in the plane defined by the corresponding positive curvatures, and a minimum in the third axis, which is perpendicular to that plane. A (3, +1) critical point is referred to as a ring critical point (RCP) and corresponds to the point where electron density at the center of a ring is a minimum.

A (3, +3) critical point has three positive curvatures and represents a point at which $\rho(\mathbf{r})$ is a local minimum. A (3, +3) critical point is referred to as a cage critical points (CCP) and corresponds to areas of low electron density in enclosed spaces.

2.9.2 Definition of an Atom in a Molecule

The first derivative of $\rho(\mathbf{r})$ or the gradient, $\nabla\rho(\mathbf{r})$, is a vector, which points in the direction of the greatest increase in $\rho(\mathbf{r})$ and has a magnitude equal to the rate of increase

in that direction. Accordingly, all gradient vectors near a nucleus will point to and terminate at the nucleus. In fact, all trajectories, which are obtained by tracing gradient vectors, terminate at the nuclei. The nuclei are said to be attractors in the gradient vector field and the trajectories create basins. The existence of attractors and basins in topology of the electron density gives rise to a natural partitioning of the molecular system. Atoms are defined as the union of an attractor and its basin and are represented by the symbol Ω. The surface bounding an atom S(Ω) is defined by the zero-flux condition:

$$\nabla \rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0, \quad \text{for all } \mathbf{r} \text{ belonging to the surface } S(\Omega) \tag{2.9.3}$$

where $\mathbf{n}(\mathbf{r})$ is the unit vector normal to S(Ω). The zero-flux condition implies that the surface is not crossed by any gradient vectors at any point. Figure 2.1 illustrates electron density and gradient vector field of BF_3 as well as the partitioning of the molecule.

The partitioning of molecular space into atomic basins allows for the partitioning of electronic properties into atomic contributions. Atomic contributions are calculated by integrating the appropriate operator over the atomic volume. For example, the total electron population of an atom (N(Ω)) can be obtained by integrating the electron density over the atomic basin.

$$N(\Omega) = \int_{\Omega} \rho(r) dr \tag{2.9.4}$$

The atomic charge can then be determined by subtracting N(Ω) from the nuclear charge (Z_Ω):

$$q(\Omega) = Z_\Omega - N(\Omega) \tag{2.9.5}$$

Other properties that are often discussed include polarizations, volumes and energies. The integrals required to calculate these properties can be found in other texts.[55]

One important result of Bader's work is that the atomic energies are additive.[56] That is,

$$E_{total} = \sum_{\Omega} E(\Omega) \tag{2.9.6}$$

where E_{total} is the total energy of the molecular system and $E(\Omega)$ are the individual atomic energies. Equation (2.9.6) is very significant because it provides an understanding of atomic contributions to important chemical phenomena. For example, QTAIM has been used to understand the atomic origins of the relative energies of isomers[57-59] as well as the atomic origins of potential energy barriers[60, 61] and atomic contributions to dissociation energies[62].

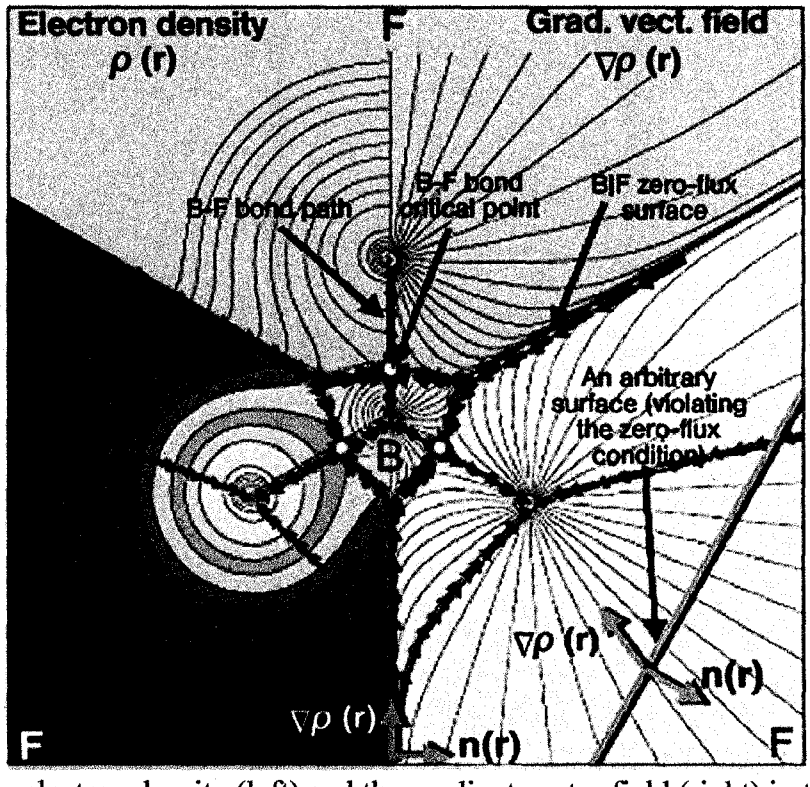

Figure 2.1 The electron density (left) and the gradient vector field (right) in the molecular plane of BF_3. The nuclei are connected by bond paths, which are illustrated by the blue arrows. The purple lines represent the separation of the atomic basis and represent the intersection of the plane with the zero-flux surfaces. The small circles drawn on the bond paths are the bond critical points. Reproduced from Matta, C. F.; Boyd, R. J.; Editors *The Quantum Theory of Atoms in Molecules* Wiley-VCH: Weinheim, 2007.

2.9.3 Definition of a Bond

There is a unique set of gradient vectors on the zero-flux surface that terminate at BCPs. This set of gradient vectors only exists between bonded atoms and corresponds to an interatomic surface. The interatomic surface is always accompanied by a pair of trajectories that originate at the BCP and terminate at the nuclei. These trajectories are referred to as bond paths and are paths of maximum electron density between the BCP and the nucleus. Bond paths provide a completely unambiguous definition of a bond in a molecular system as only atoms that share an intermolecular surface are linked by a bond path and are therefore bonded.[63, 64] Bond paths indicate bonding of all kinds, regardless of the nature or strength of the interaction.

The properties of the energy and electron densities at the BCP provide useful information about chemical bonding interactions. For example, the magnitude of the electron density at the BCP is a reflection of the strength of the bond or the bond order. The electron density at the BCP is generally more than 0.20 au for covalent bonds and less than 0.10 for closed shell interactions including ionic, van der Waals and hydrogen bonding. It has been shown that the electron density at the bond critical point is strongly correlated with binding energy for several types of bonds. The correlation between binding energy and hydrogen bond strength is of particular relevance to this thesis.[63-68]

The Laplacian ($\nabla^2\rho$) of the electron density also provides useful information about bonding. The Laplacian at the BCP is defined as the sum of the three curvatures at that point. As previously mentioned, at a BCP, two curvatures are negative and one is positive. The negative curvatures indicate the concentration of the density along the bond path while the positive curvature measures how the electron density is depleted along the

interatomic surface and concentrated on the individual atomic basins. Generally, for covalent bonds, the negative curvatures are dominant and $\nabla^2\rho < 0$. For ionic, hydrogen bonding and van der Waals interactions, there is a depletion of density along the interatomic surface and the positive curvatures dominates ($\nabla^2\rho > 0$).

The ellipticity (ε) at a BCP is defined as

$$\varepsilon = \frac{\lambda_1}{\lambda_2} - 1 \quad \text{when } |\lambda_1| > |\lambda_2| \text{ and}$$
$$\varepsilon = 0 \quad \text{when } \lambda_1 = \lambda_2 \tag{2.9.7}$$

The ellipticity is a measure of how much the electron density is accumulated in a given plane that contains the bond path. In other words, it is a measure of the π-character of the bonding.

It has been suggested that these local properties of the electron density evaluated at the bond critical points can be used to characterize chemical bonds.[69] This thesis uses the quantum theory of atoms in molecules to extract chemical information from the electron density in order to characterize hydrogen bonding, specifically.

Chapter 3. Analysis of Hydrogen Bonding in the α-Helix

3.1 Introduction

The α-helix is a common conformational motif in peptides and proteins.[70] Furthermore, the formation of α-helices is thought to have an important role in the early stages of protein folding[70-73] and therefore, the prediction of secondary structure is an important first step in understanding overall protein structure.[10] Accordingly, the formation of α-helices has been the focus of numerous experimental[74-79] and theoretical studies.[14, 70, 80-85] Molecular mechanics and molecular dynamics are widely used theoretical methods in the modeling of large peptides.[86-89] Due to the computational speed of these methods, they are sometimes the only feasible means of studying large biological systems. However, molecular mechanics neglects explicit treatment of electrons and therefore cannot account for important electronic effects, such as hydrogen bonding. Only quantum mechanics based methods can be used to account for the important electronic effects that influence protein structure. Many studies on the α-helix have employed quantum mechanical calculations with the view that a fundamental understanding of protein structure requires a more accurate quantitative description of all its governing forces.[14]

One important force stabilizing the structure of α-helices is intramolecular hydrogen bonding in the peptide backbone.[90, 91] Because the energy of a hydrogen bond, which ranges from 5 to 10 kcal/mol, is comparable to the free energy of folding of proteins, an accurate characterization of hydrogen bonding is essential for understanding the forces that stabilize proteins.[85] As quantum mechanical calculations are restricted by the size of the system, many previous studies on hydrogen bonding have used dipeptides

and tripeptides in order to model hydrogen bonding in the peptide backbone.[84, 85] While these small models have some benefits: they allow the calculation of hydrogen bonding energies, as well as the inclusion of solvation effects, they do not describe the neighboring effects of nearby hydrogen bonds in the protein backbone. Neighboring effects are an important factor in the stabilization of proteins due to the cooperativity of hydrogen bonds.[92, 93] That is, hydrogen bonds may interact with each other. As a result, both experimental[94, 95] and theoretical studies[96, 97] have recently begun to focus on hydrogen bond networks in secondary structures in order to account for the cooperativity of hydrogen bonding. However, due to the large size of systems required to study hydrogen bond networks, quantum mechanical calculations have been difficult and a detailed characterization of hydrogen bonding in secondary structure has not been performed.

It has been suggested that QTAIM would be very useful for analyzing peptides.[98-100] Analysis of the electron density can be used to confirm hydrogen bonding.[101] Furthermore, the electron density at bond critical points has been found to be correlated with bond strength[63-68, 102] and with bond length in hydrogen bonds.[103] Consequently, QTAIM has been widely applied to characterizing hydrogen bonding interactions.[56, 66, 81, 84, 104] In fact, the sum of electron density at the hydrogen bond critical points has been found to be correlated with helix stability.[81] Motivated by the potential for QTAIM to provide valuable information about hydrogen bonding in peptides, an electron density analysis has been performed on an α-helix model, in order to characterize the hydrogen bond network.

In order to investigate the hydrogen bond network of α-helices, full quantum mechanics optimizations have been performed on 19 different peptide sequences with formyl (COH-) and amino (-NH_2) end groups. The sequences are 13 amino acids in length and contain 12 alanines and one guest residue: for-AAAAAAXAAAAAA-NH_2, where X designates one of 19 possible amino acids.*† The model was chosen because alanine based peptides form stable α-helices[105] and have therefore been recognized as important models for the study of helix formation.[70] The length of the peptide was chosen because the average length of an α-helix found in a protein is 12 amino acid residues long[106] and it has been found that at least seven residues are required to form stable α-helices in the gas phase.[81] Substitutions were made at the centre of the sequence to ensure that end effects, such as helix capping, are minimized. This ensures that results reflect the intrinsic α-helix forming properties of the amino acids rather than their interactions with the helix dipole.

The effects of the substitutions on the hydrogen bond network of the α-helix model are compared and interactions that cause different amino acids to be stabilizers or destabilizers of α-helices are identified through analysis of bond critical points. No previous reports discuss the electron density analysis of hydrogen bond networks in fully optimized peptides of this length. The full optimization ensures that accuracy is maintained while the large model naturally includes longer range forces that affect helix stability.

---

* The original intent of the study was to substitute each of the 20 standard amino acids at the central position of the peptide sequence. However, after months of calculations, no optimized structure could be located for sequences containing arginine and lysine at the central position

† Two structures have been calculated for histidine. Depending on the environment, histidine may be protonated and positively charged or unprotonated and neutral.

3.2 Computational Methods

Initial geometries were constructed using Molden[107] and Gaussview[108] by Sarah Farrag, who carried out some preliminary work.[109] Geometry optimizations were performed using the hybrid functional B971,[110] which has been found to be superior to five other density functionals, including the popular B3[111]LYP[112], for predicting hydrogen bonding energies.[113] The 6-31G(d,p) basis set was used in all calculations, as it was suggested by the same article to be suitable for problems involving large molecules where larger basis sets are not computationally feasible.[113] Frequency calculations were performed with the B971/6-31G(d,p) method on all optimized structures to confirm that each is a local minima.

All calculations were performed in the gas phase with the Gaussian 03[114] suite of programs. While the effect of solvation is very important for understanding protein behavior, it is also necessary to understand the unsolvated model and the intrinsic helix forming properties before the effects of solvation can be determined.

The electron density analysis was performed with the AIM2000[115] software package using the density obtained from the B971/6-31G(d,p) calculations described above.

3.3 Results and Discussion

This thesis characterizes hydrogen bonding on the basis of properties of the hydrogen bonding critical points (HBCPs) in the electron distribution. Figures 3.1-3.19 each show two representations of the model peptides, which have different amino acids at the central residue. Part a) is a simple ball and stick representation while part b) provides a view of the bond paths as well as the locations of the BCPs. All peptide models studied

share some common properties. For example, with the exceptions of aspartate and glutamate, the substitutions of all amino acids studied at the central position produce a similar helical structure (Figures 3.1-3.17). When aspartate and glutamate are substituted at the central position, the N-terminus of the peptide loses any helical resemblance as it twists to interact with the anionic side chain (Figures 3.18 and 3.19) The common properties of the model helices are discussed first, followed by a discussion of each model separately.

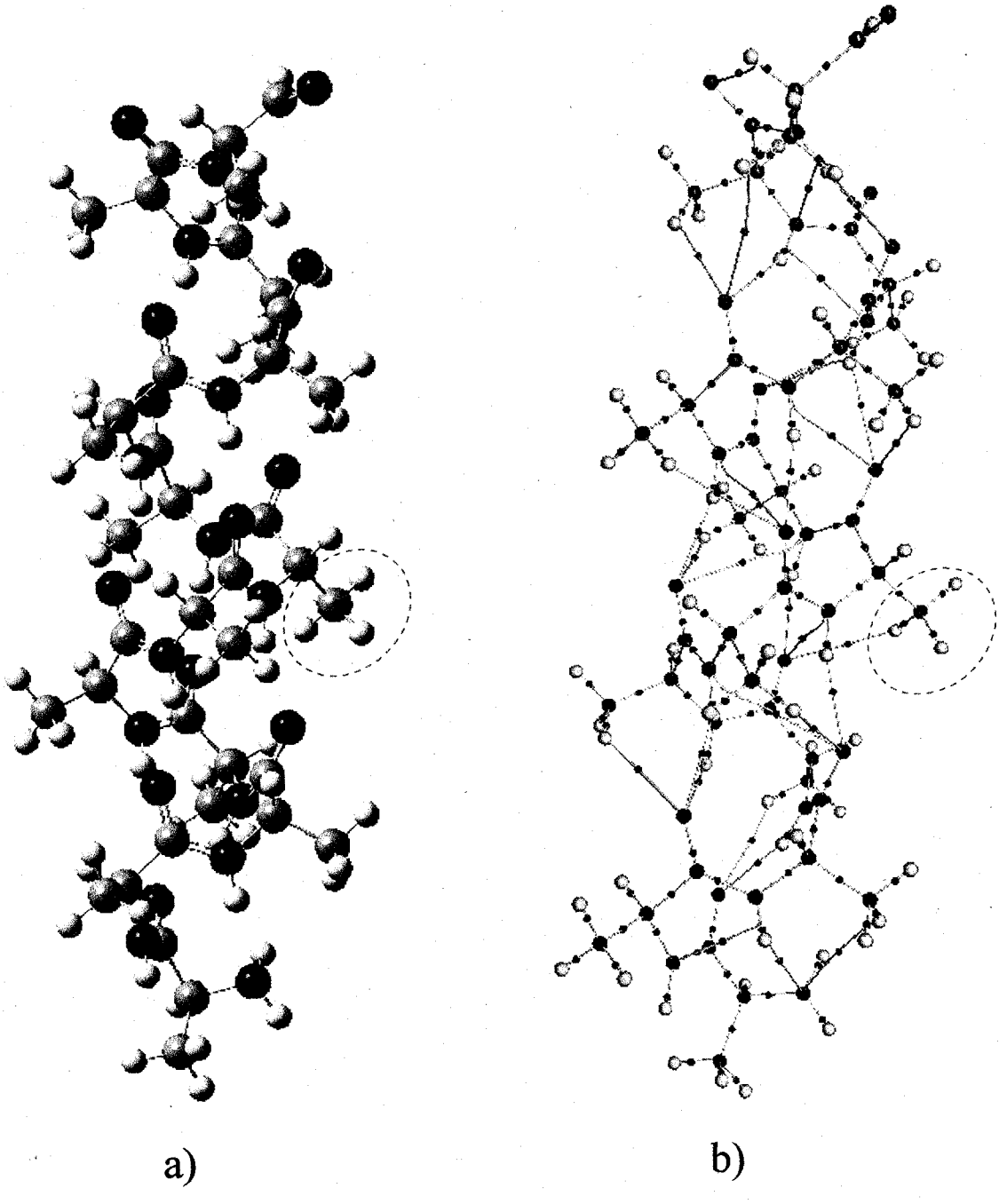

Figure 3.1 a) Ball and stick representation of For-AAAAAAAAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAAAAAAAA-NH_2. The side chain is encircled by a dotted line.

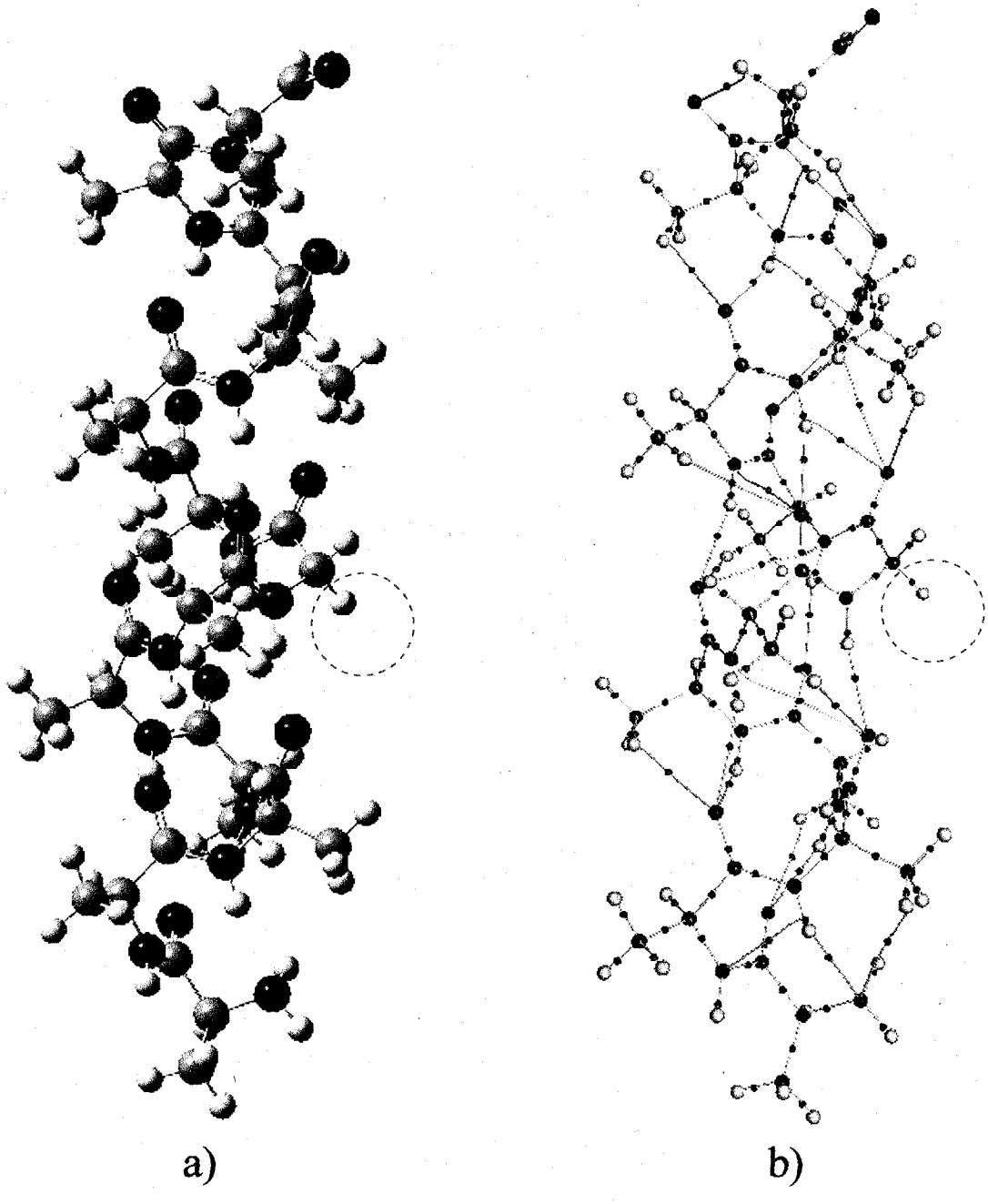

Figure 3.2 a) Ball and stick representation of For-AAAAAAGAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAGAAAAAA-NH_2. The side chain is encircled by a dotted line.

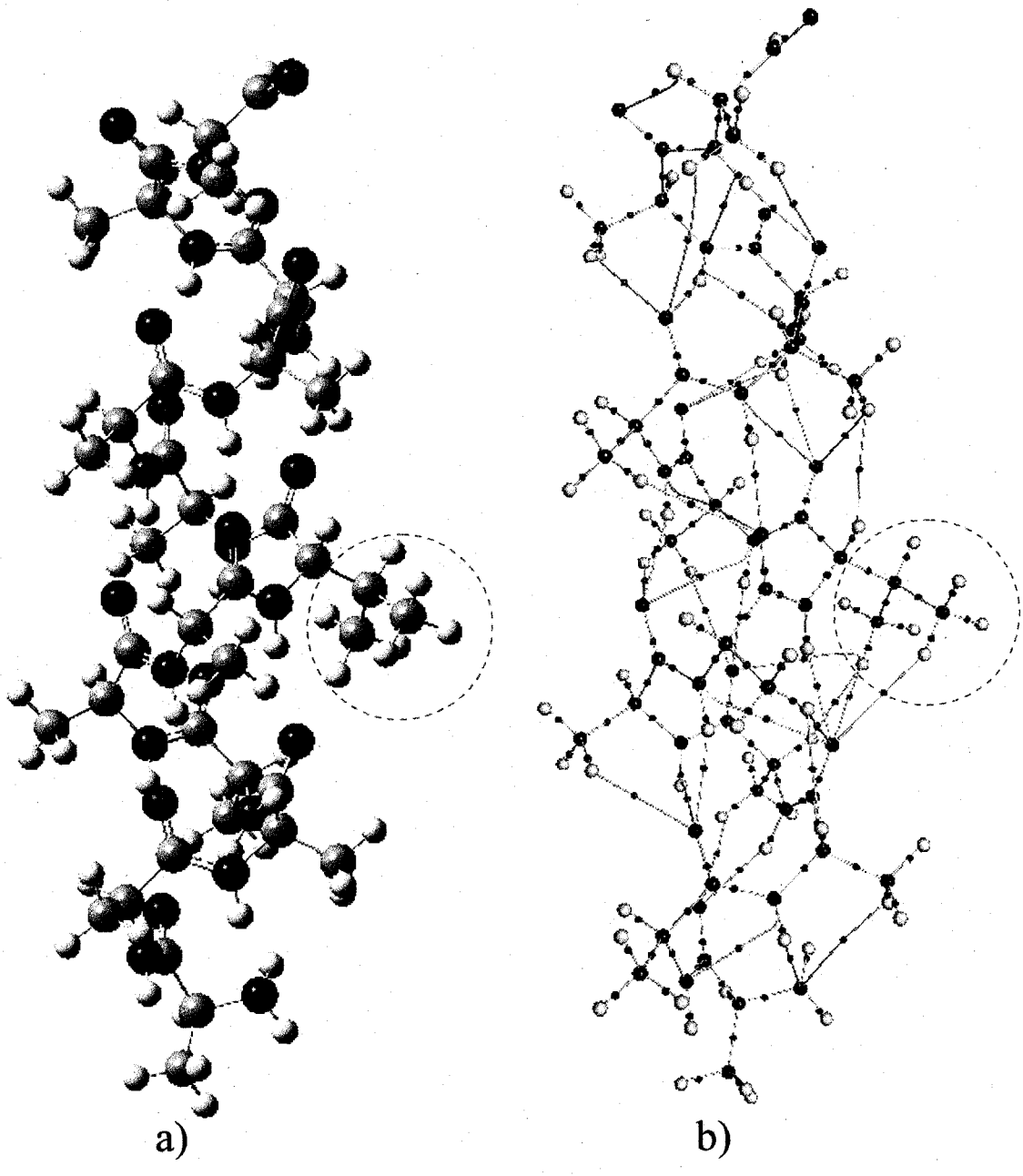

Figure 3.3 a) Ball and stick representation of For-AAAAAAVAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAVAAAAAA-NH_2. The side chain is encircled by a dotted line.

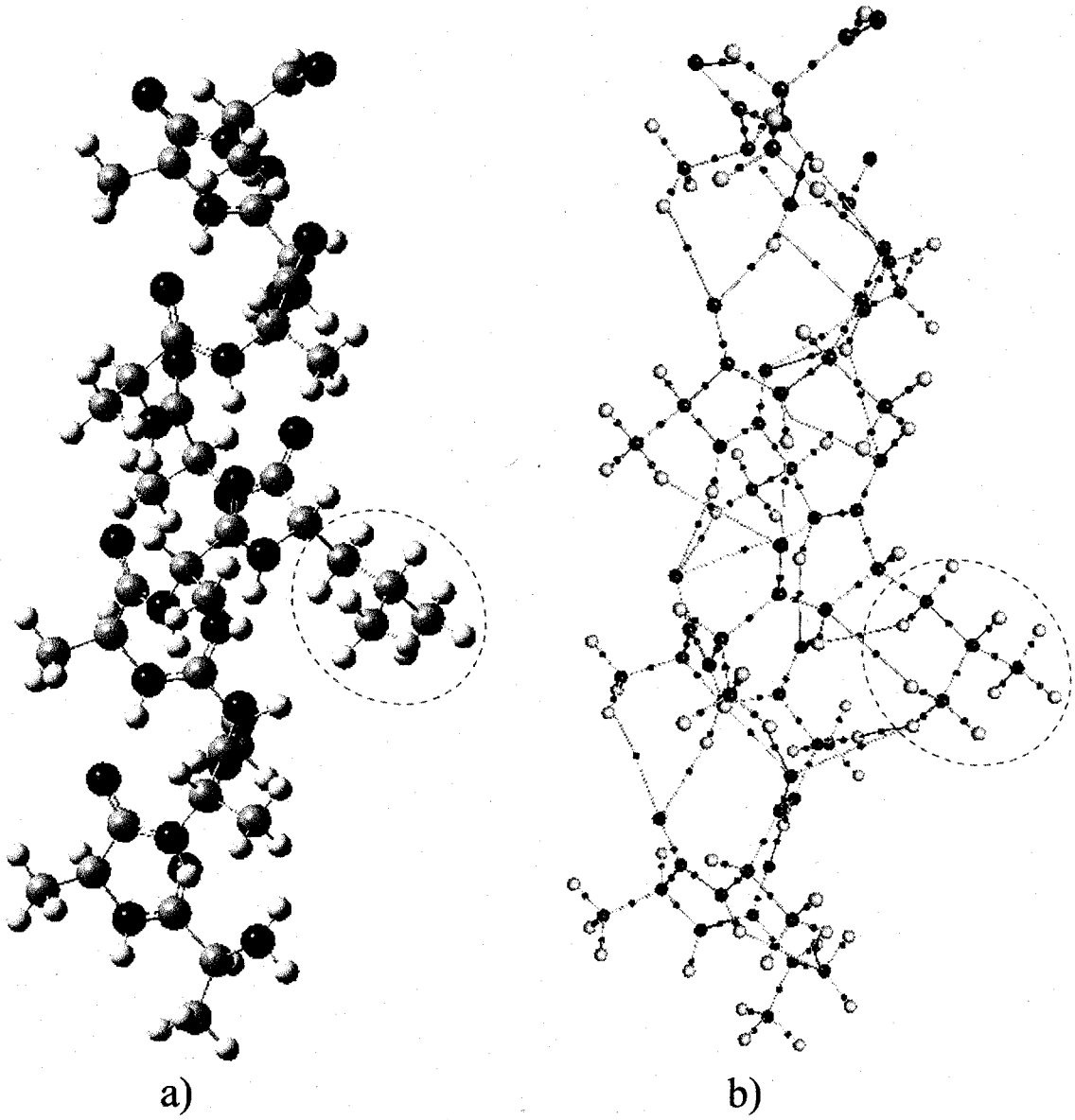

Figure 3.4 a) Ball and stick representation of For-AAAAAALAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAALAAAAAA-NH_2. The side chain is encircled by a dotted line.

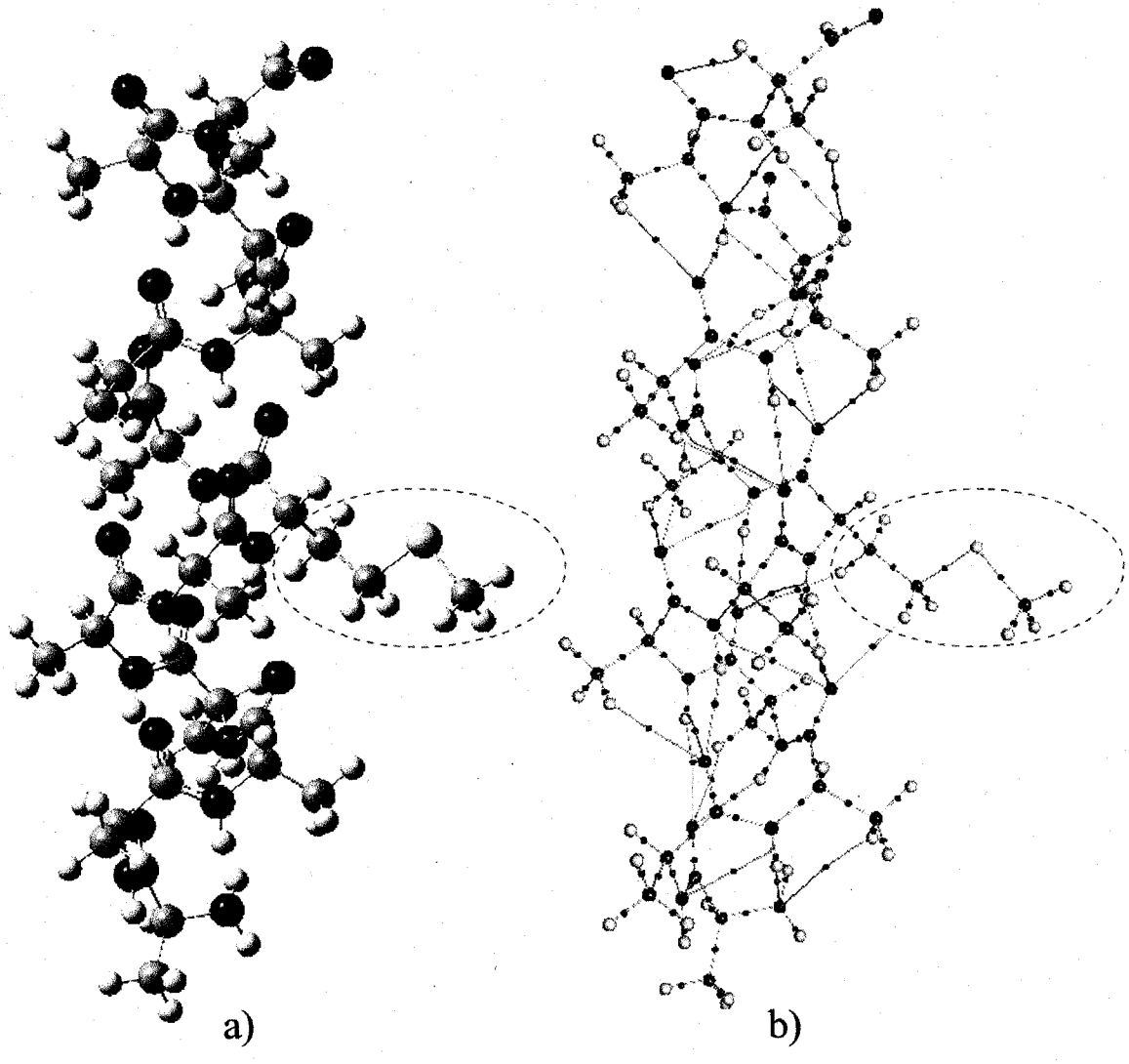

Figure 3.5 a) Ball and stick representation of For-AAAAAAMAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAMAAAAAA-NH_2. The side chain is encircled by a dotted line.

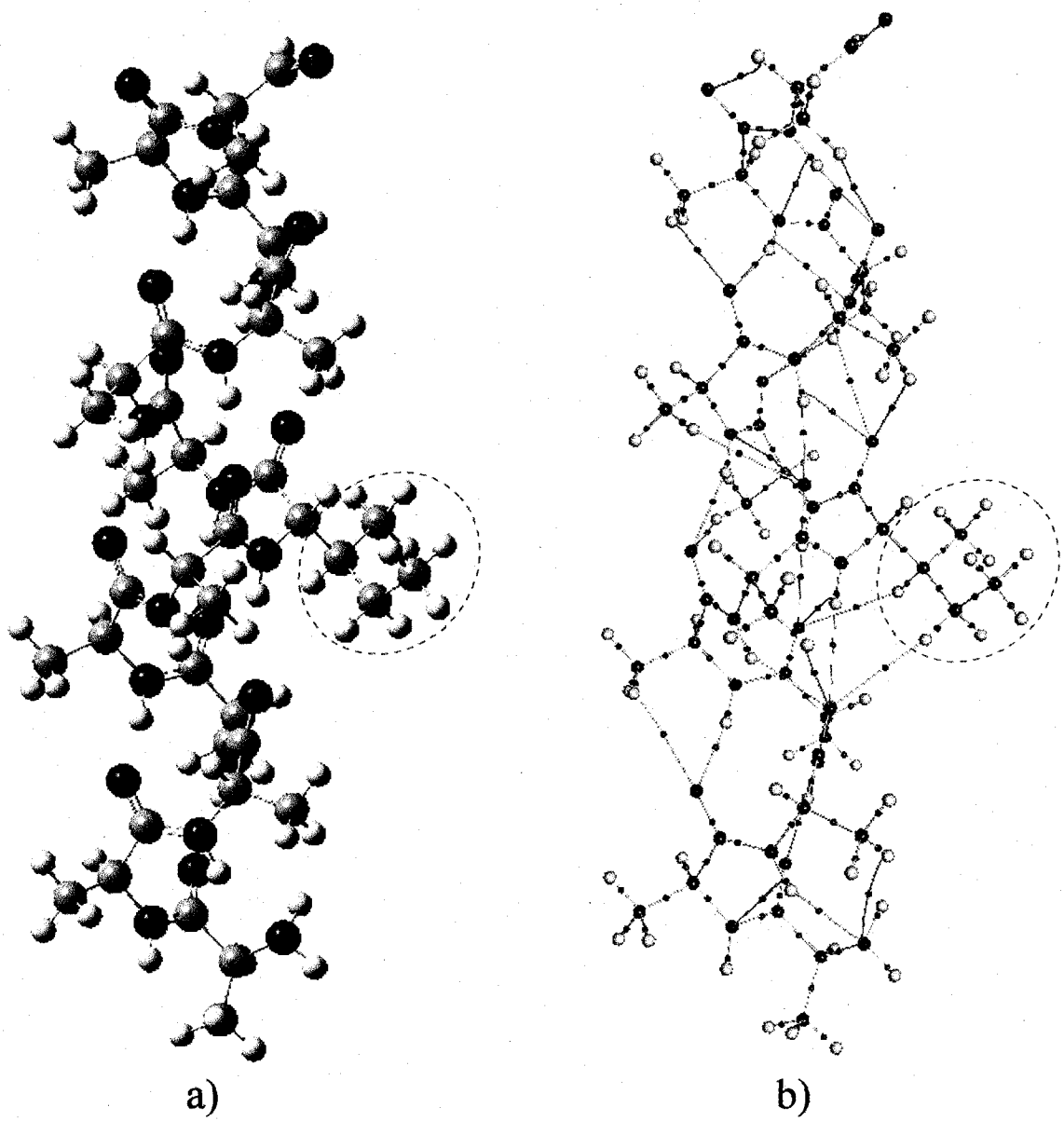

Figure 3.6 a) Ball and stick representation of For-AAAAAAIAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAIAAAAAA-NH_2. The side chain is encircled by a dotted line.

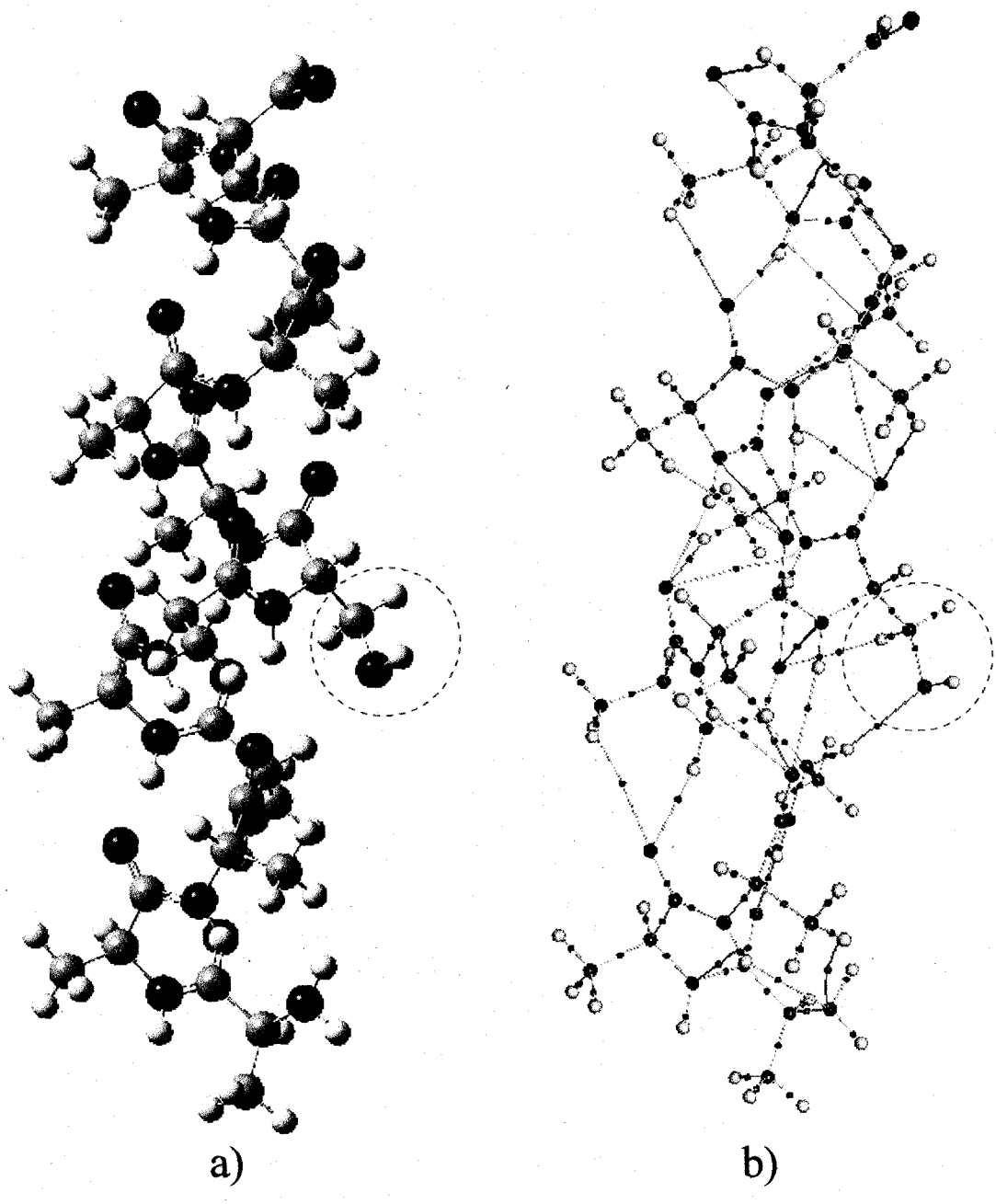

Figure 3.7 a) Ball and stick representation of For-AAAAAASAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAASAAAAAA-NH_2. The side chain is encircled by a dotted line.

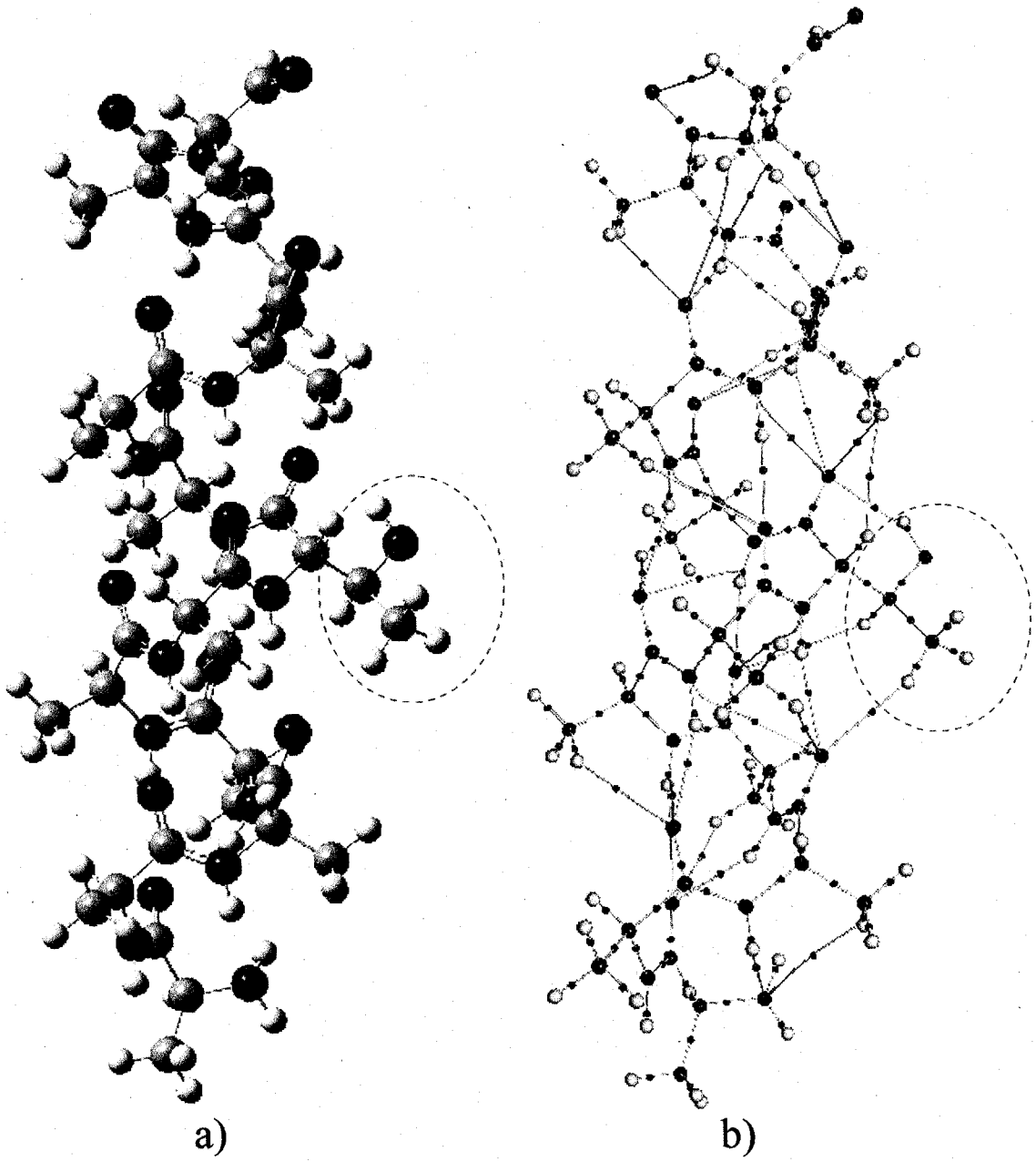

Figure 3.8 a) Ball and stick representation of For-AAAAAATAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAATAAAAAA-NH_2. The side chain is encircled by a dotted line.

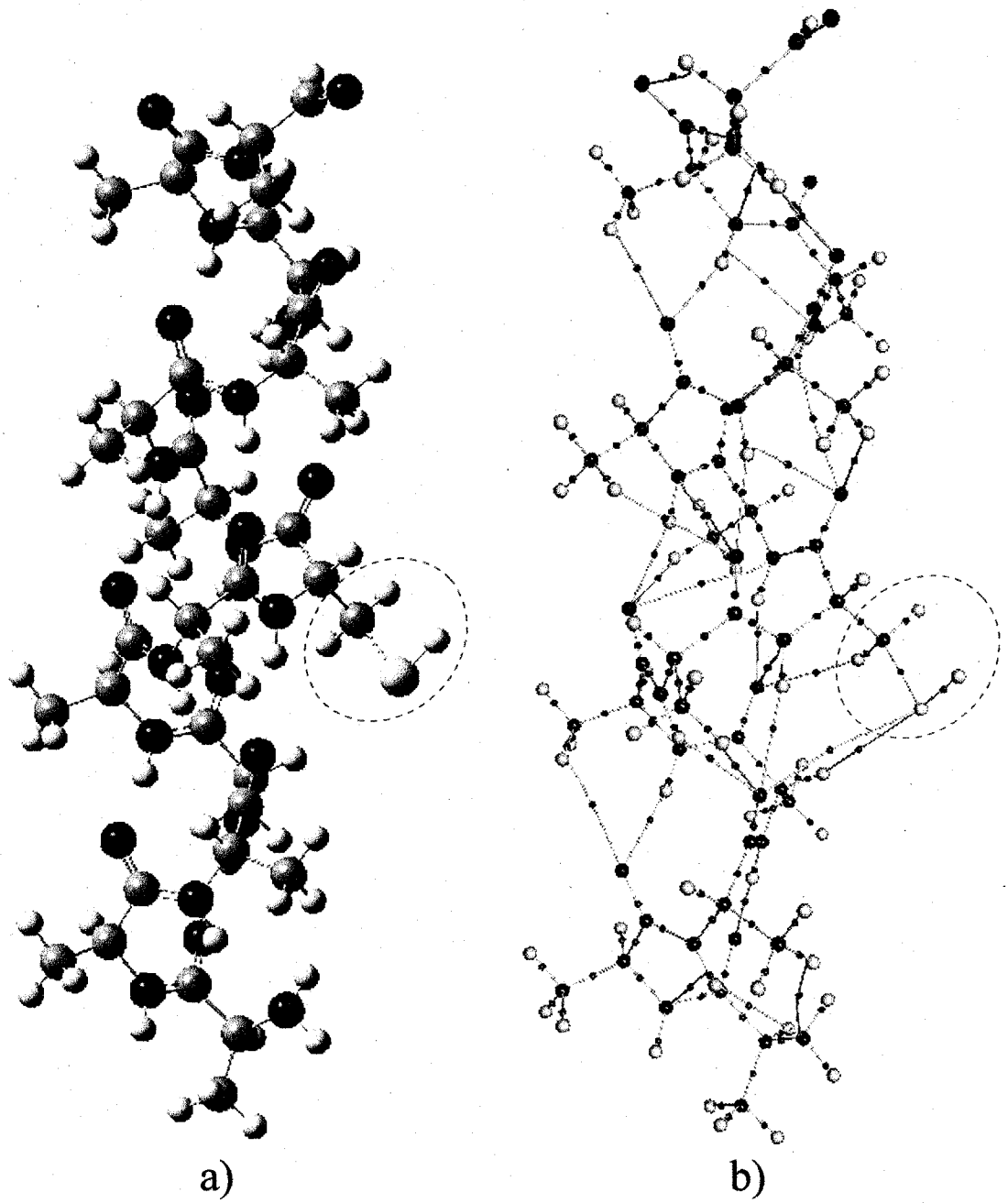

Figure 3.9 a) Ball and stick representation of For-AAAAAACAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAACAAAAAA-NH_2. The side chain is encircled by a dotted line.

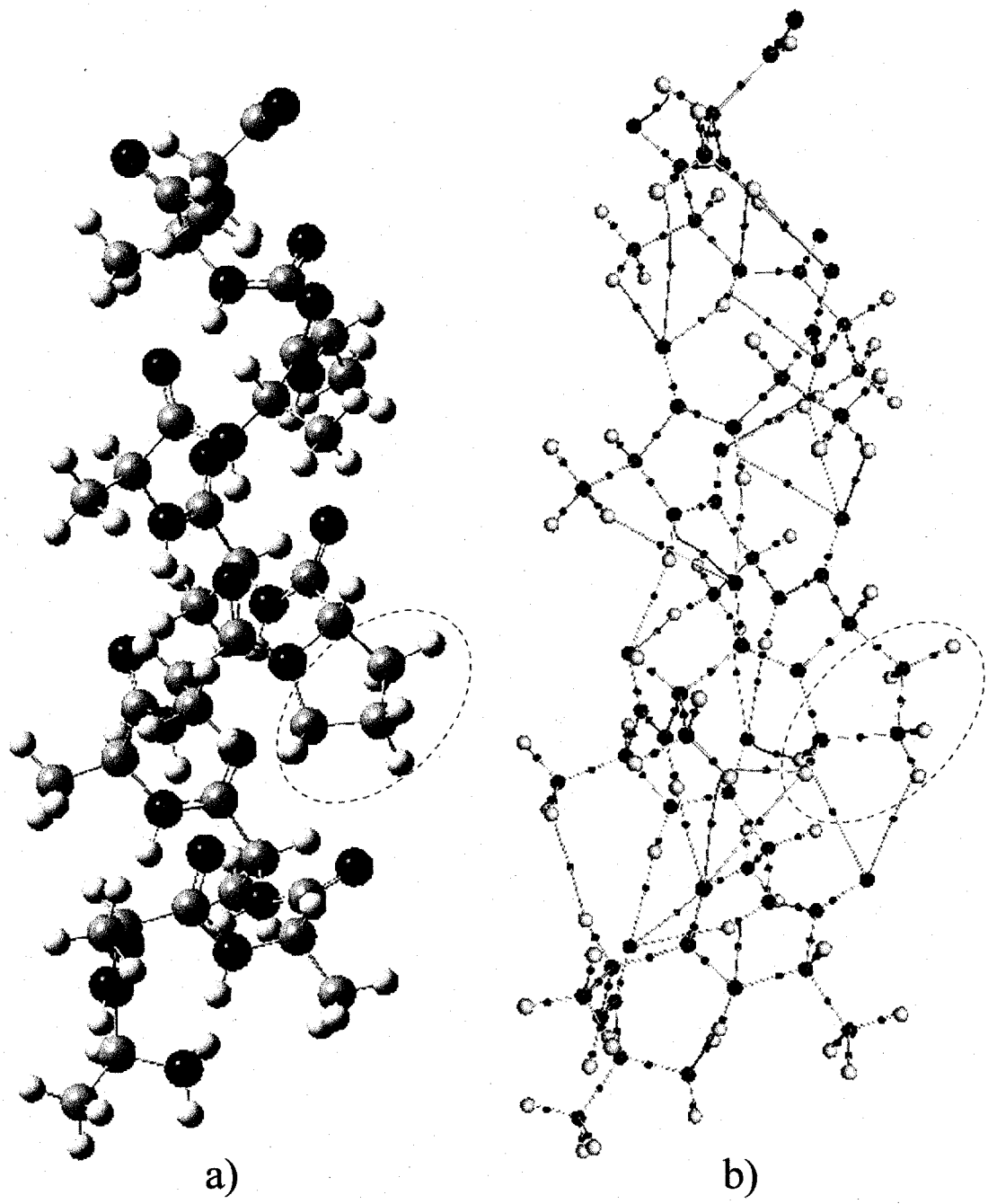

Figure 3.10 a) Ball and stick representation of For-AAAAAAPAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAPAAAAAA-NH_2. The side chain is encircled by a dotted line.

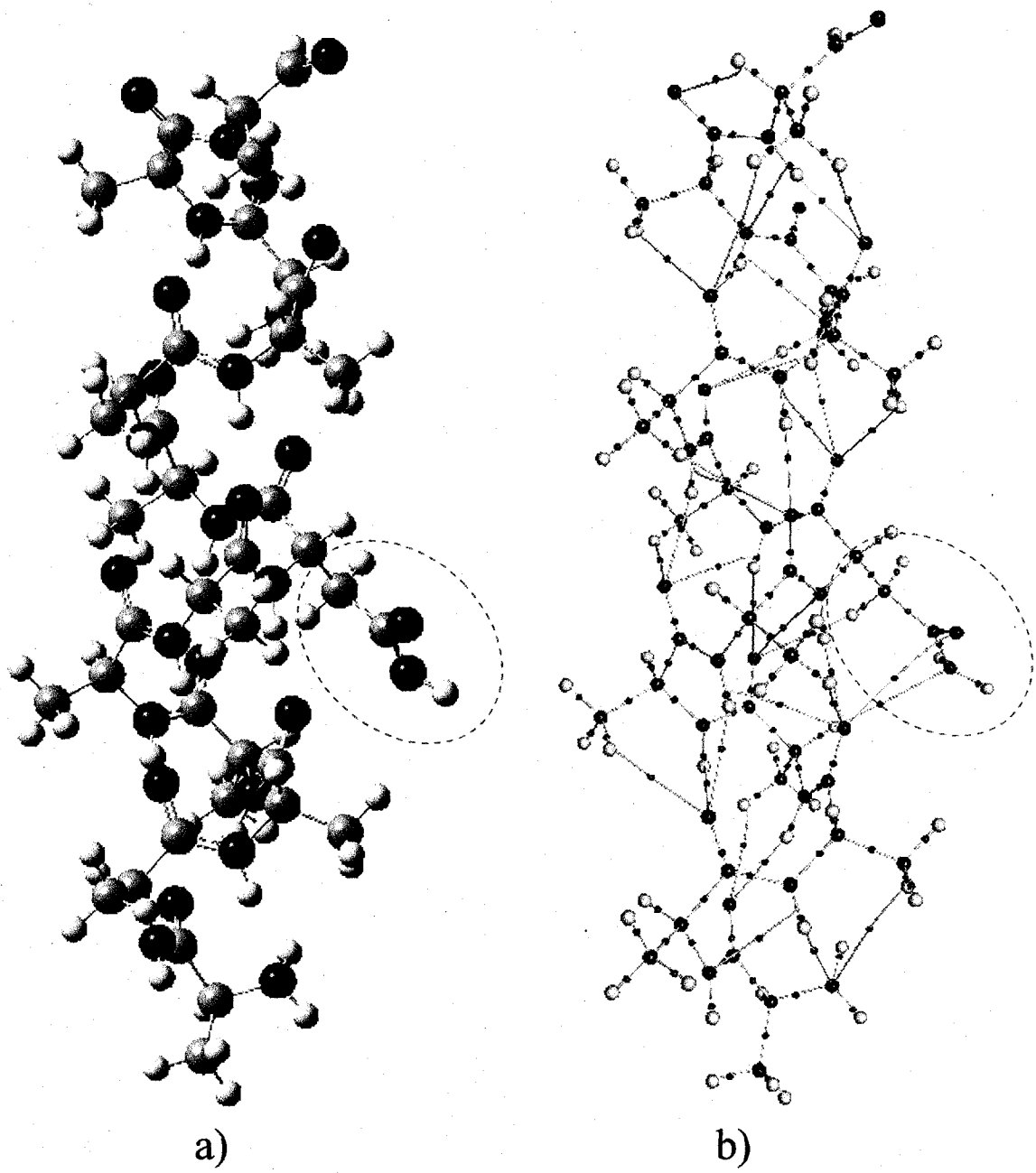

Figure 3.11 a) Ball and stick representation of For-AAAAAANAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAANAAAAAA-NH_2. The side chain is encircled by a dotted line.

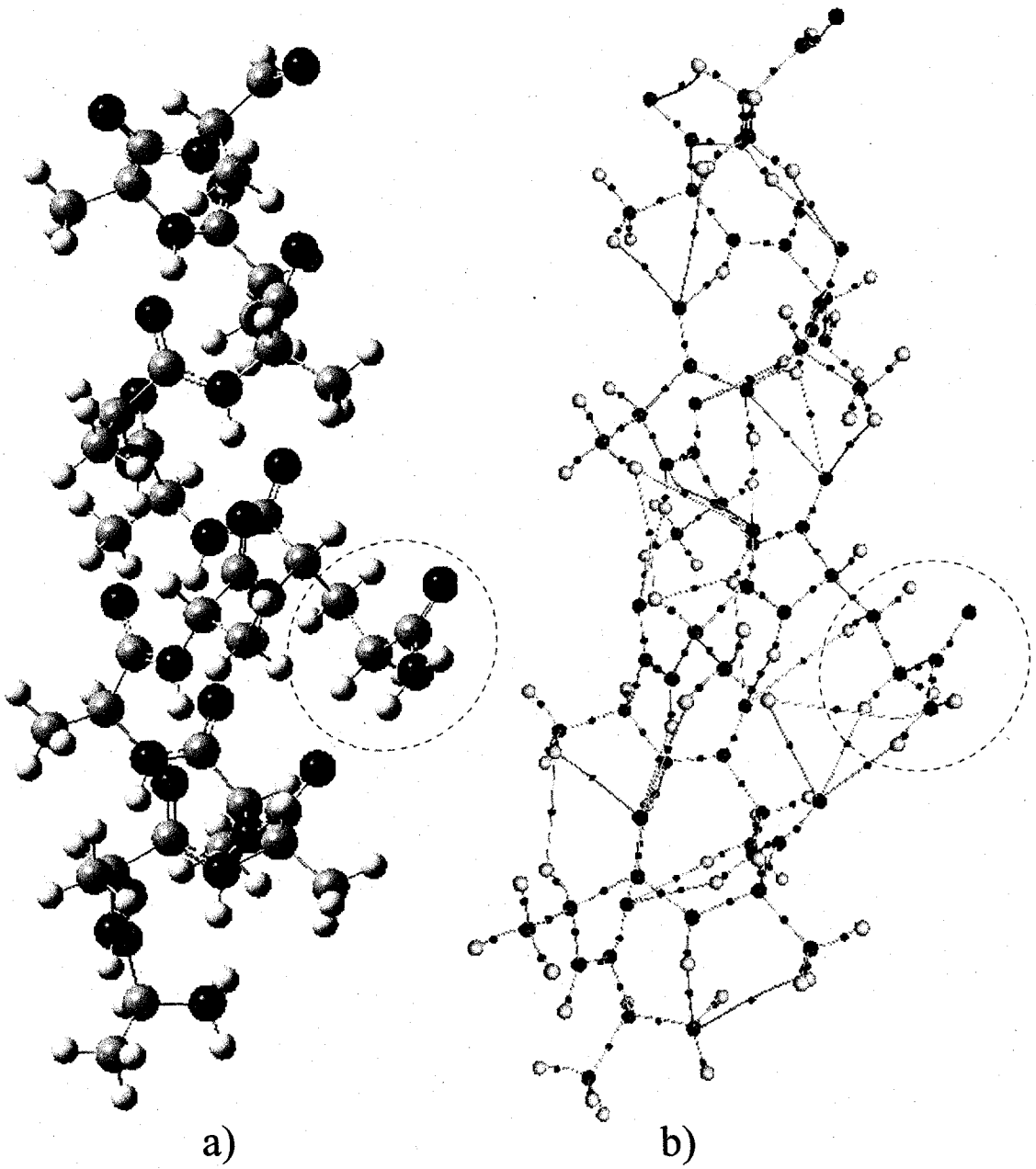

Figure 3.12 a) Ball and stick representation of For-AAAAAAQAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAQAAAAAA-NH_2. The side chain is encircled by a dotted line.

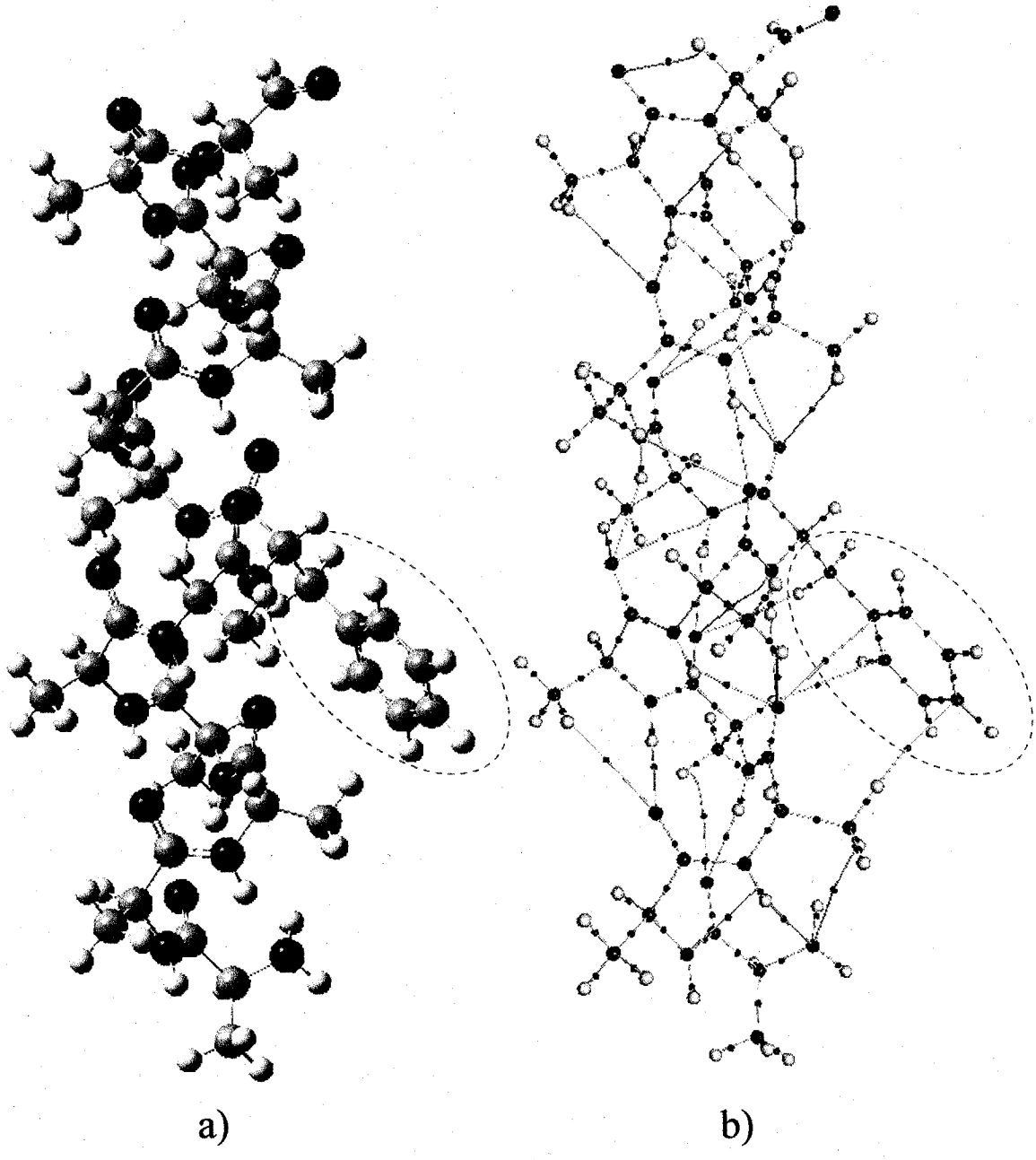

Figure 3.13 a) Ball and stick representation of For-AAAAAAFAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAFAAAAAA-NH_2. The side chain is encircled by a dotted line.

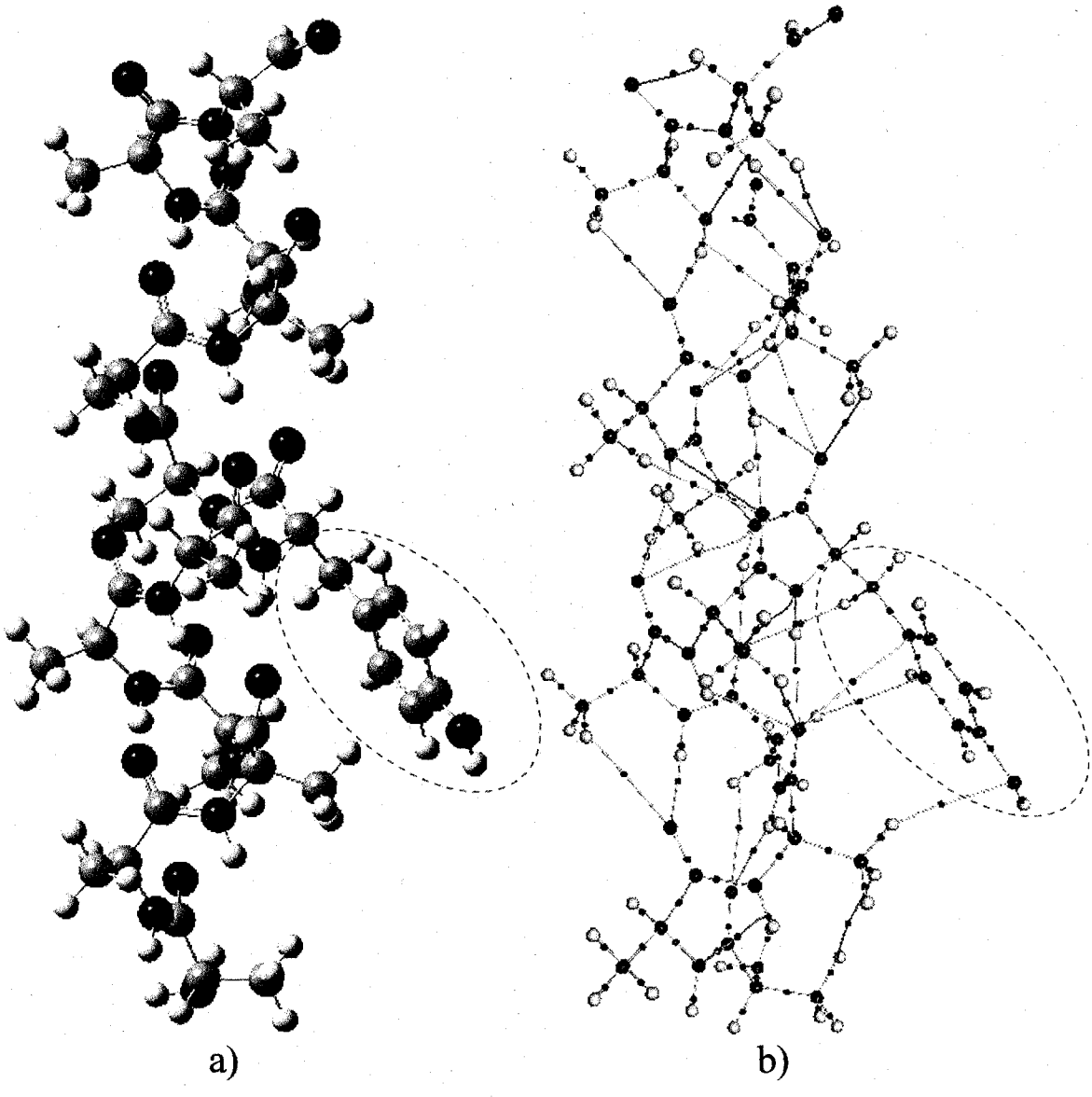

Figure 3.14 a) Ball and stick representation of For-AAAAAAYAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAYAAAAAA-NH_2. The side chain is encircled by a dotted line.

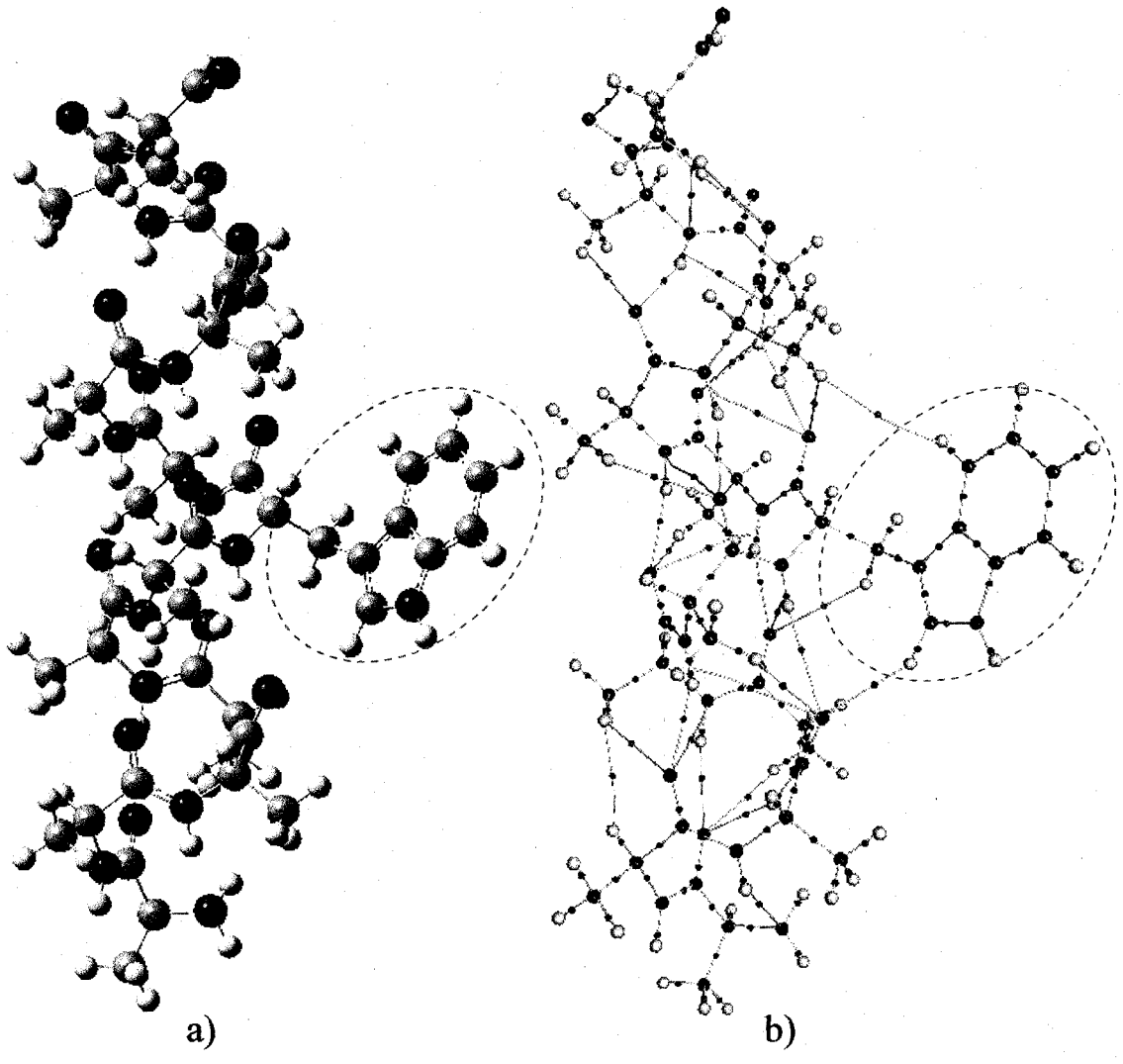

Figure 3.15 a) Ball and stick representation of For-AAAAAAWAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAWAAAAAA-NH_2. The side chain is encircled by a dotted line.

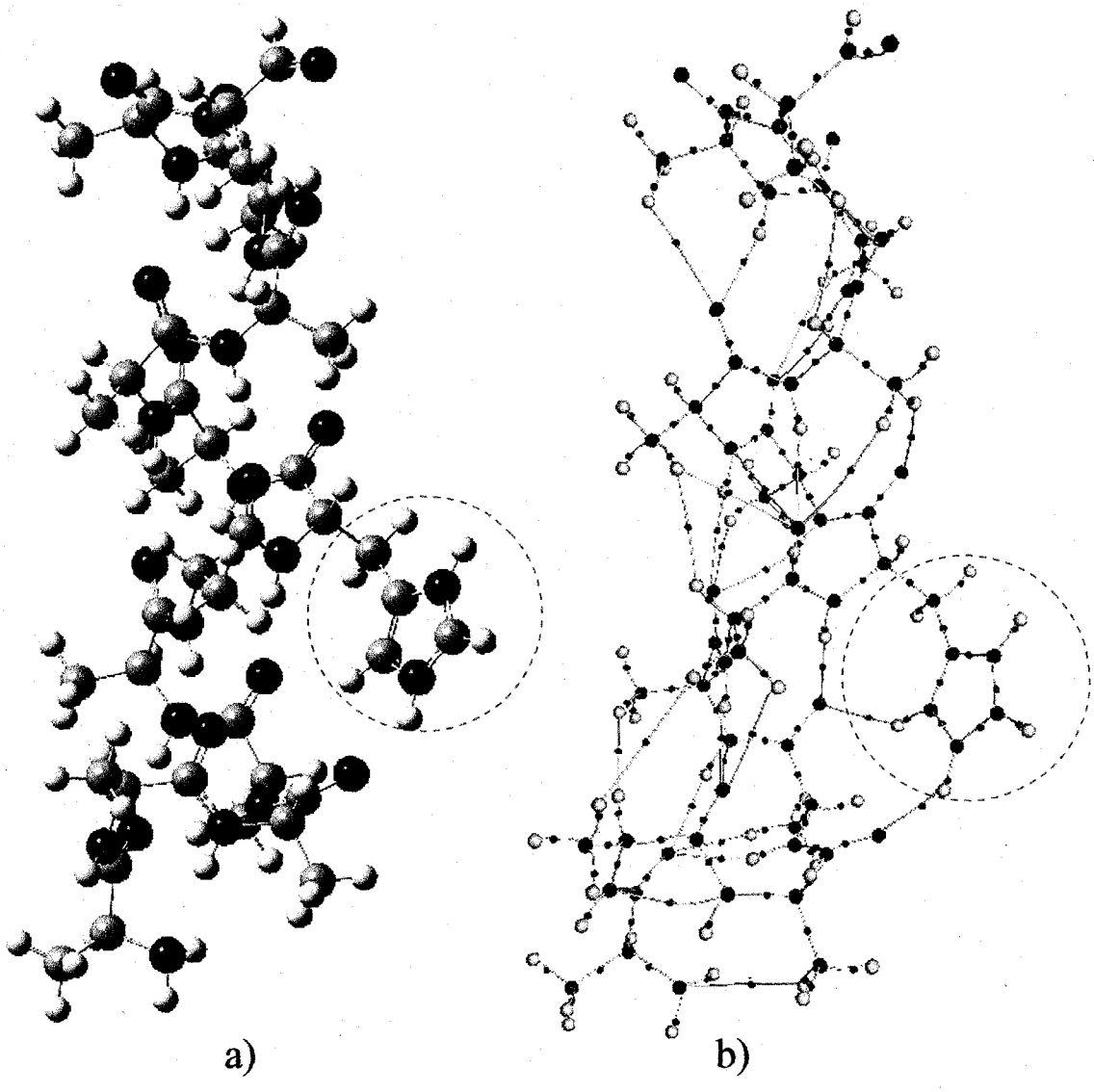

Figure 3.16 a) Ball and stick representation of For-AAAAAAHAAAAAA-NH_2. Histidine is protonated and positively charged. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAHAAAAAA-NH_2. The side chain is encircled by a dotted line.

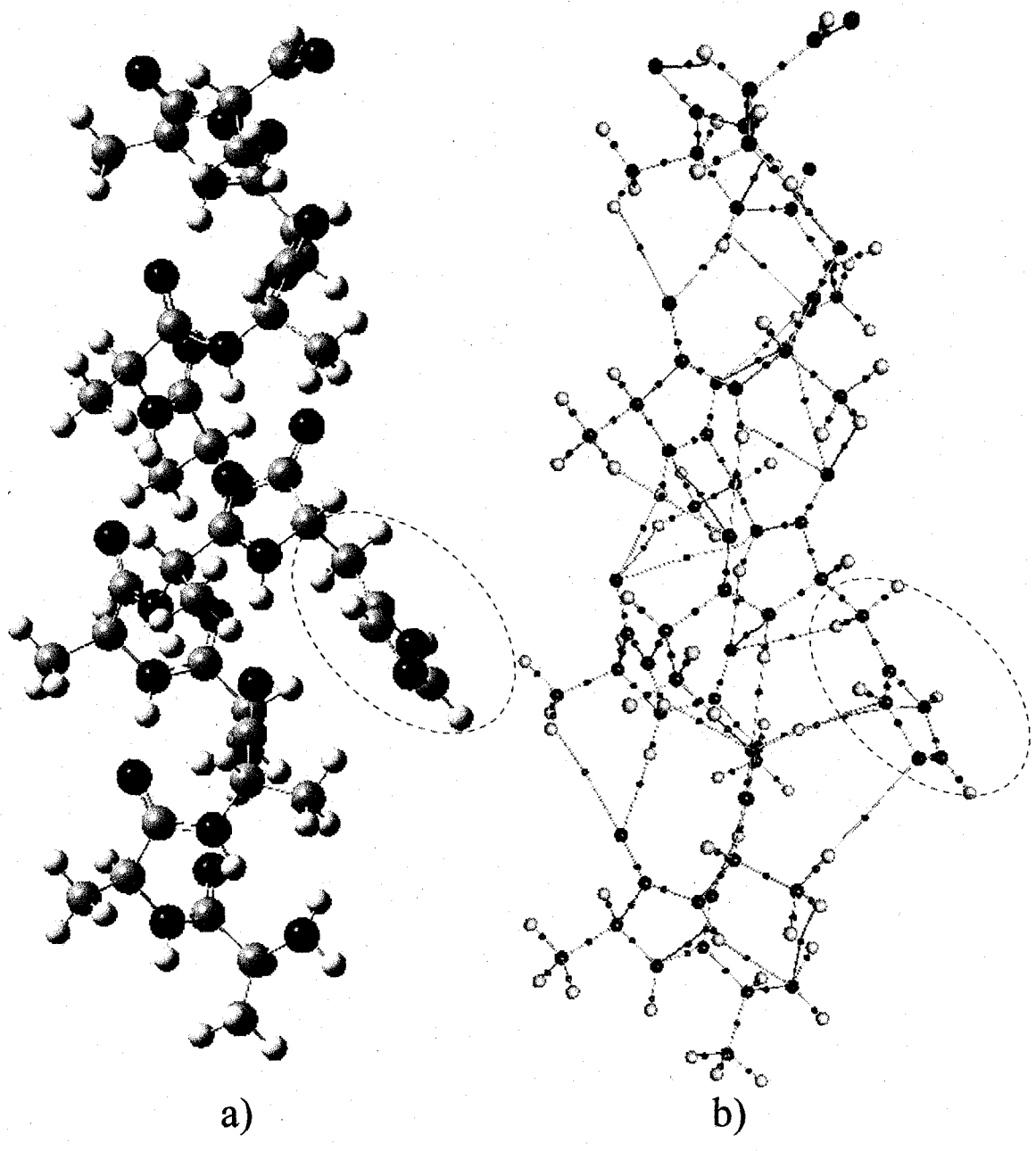

Figure 3.17 a) Ball and stick representation of For-AAAAAAHAAAAAA-NH_2. Histidine is not protonated and is neutral. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAHAAAAAA-NH_2. The side chain is encircled by a dotted line.

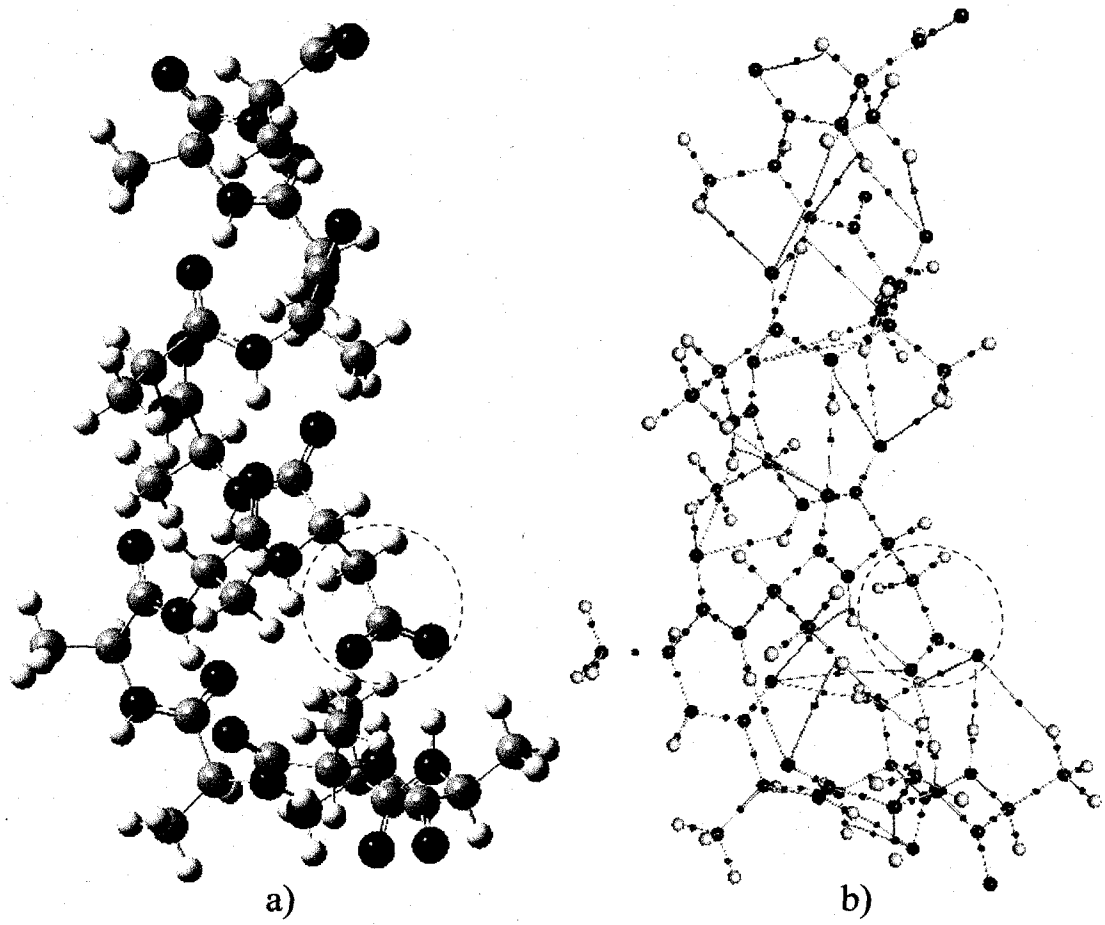

Figure 3.18 a) Ball and stick representation of For-AAAAAADAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAADAAAAAA-NH_2. The side chain is encircled by a dotted line.

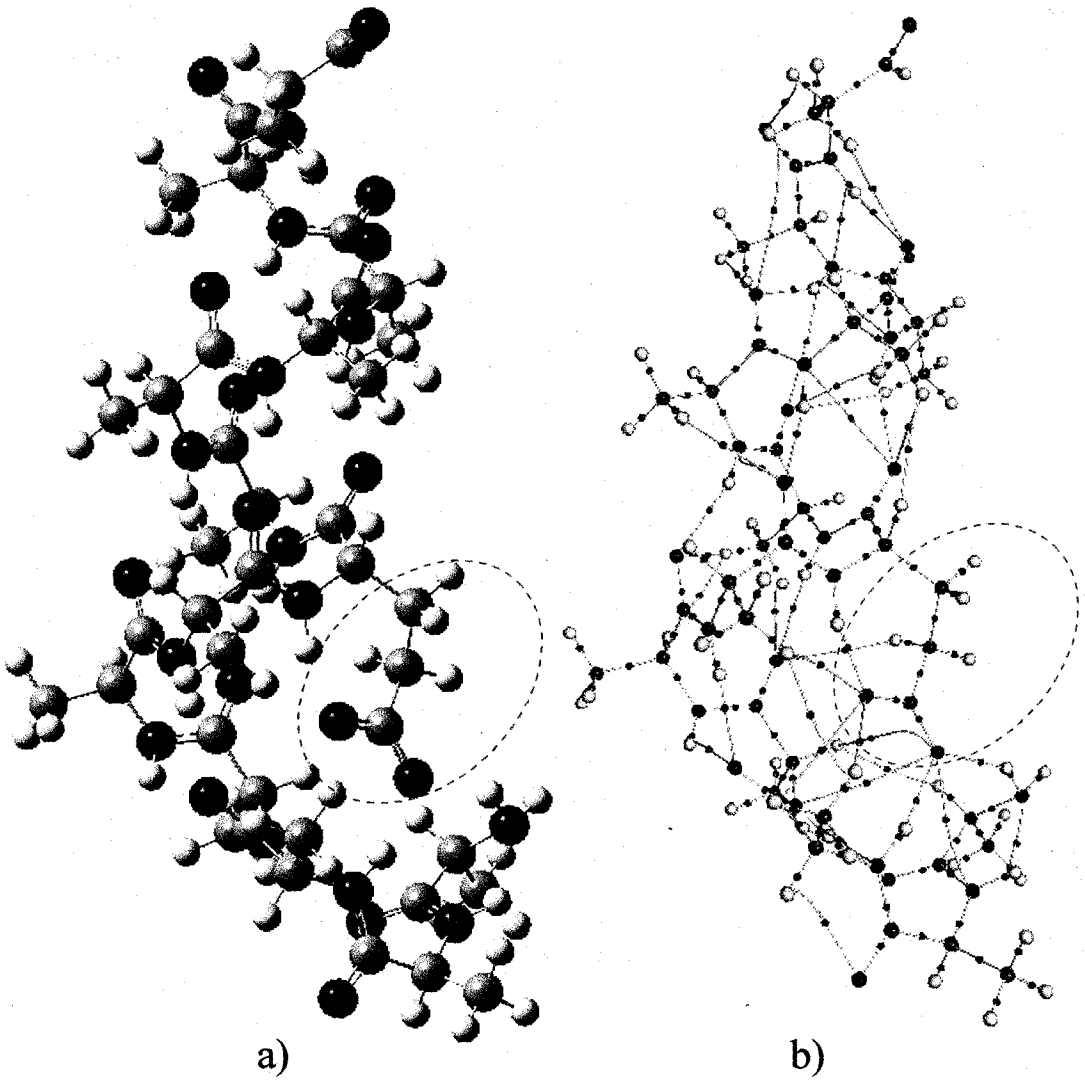

Figure 3.19 a) Ball and stick representation of For-AAAAAAEAAAAAA-NH_2. The N-terminus is located at the bottom of the figure. b) Molecular graph of For-AAAAAAEAAAAAA-NH_2. The side chain is encircled by a dotted line.

3.3.1 Common Characteristics of the Model Peptides

3.3.1.1 Hydrogen Bonding

There are three types of hydrogen bonds that are present within the backbone of each model helix studied. The values of the electron density at all HBCPs, $\rho(r_c)$, all fall in the range 0.002-0.022, which is within the range of common hydrogen bonds.[101, 116] The Laplacian, $\nabla^2\rho(r_c)$, is positive for all bonds, which is expected for hydrogen bonds. Also, no bonds exhibit a notable ellipticity. The average properties of each type of bond are described in Table 3.1.

One type of hydrogen bonding interaction found is between the N-H group at position *i* and the C=O group at position *i* + 4. This bonding interaction will be referred to as N-H•••O, *i* + 4 and is the characteristic hydrogen bond of the α-helix structure. The average bond length of a N-H•••O, *i* + 4 bond is 2.281 Å and the average $\rho(r_c)$ found at the HBCP is 0.014 au. Depending on the substitution at the centre of the helix, there are between four and seven N-H•••O, *i* + 4 bonds present, which are normally located between the center of the helix and the N-terminus. This is consistent with the experimental finding that the greatest α-helix content in alanine-based peptides is between the middle of the peptide and the N terminus.

A second type of hydrogen bond is found between the side chain C-H group at position *i* and the C=O group at position *i* + 3. The presence of a C-H•••O, *i* + 3 interaction is consistent with other studies that find that this type of interaction is involved in the stabilization of the α-helix.[14, 84, 117] The average bond length (2.732 Å) is longer than that of the N-H•••O, *i* + 4 interaction and the average $\rho(r_c)$ (0.006 au) found at the HBCP is smaller, indicating that the C-H•••O, *i* + 3 is a weaker interaction.

However, these interactions are significant stabilizers of the α-helix: there are between six and ten interactions found throughout the helix, at the N-terminus and at the interior of the helix. Near the C-terminus, there are no C=O groups for the side chain C-H groups to interact with.

The third type of hydrogen bond is between the N-H group at position *i* and the C=O group at position *i* + 3. The average bond length (2.257 Å) and $\rho(r_c)$ (0.014 au) indicate that the N-H•••O, *i* + 3 interaction is stronger than the C-H•••O, *i* + 3 interaction and comparable to the N-H•••O, *i* + 4 interaction. There are between three and eight N-H•••O, *i* + 3 interactions, depending on the residue at the centre of the helix. These interactions are usually divided between the outer ends of the helix and are not usually present in the centre. This type of bonding is characteristic of the 3_{10}-helix. The 3_{10}-helix has 3 residues per turn rather than the 3.6 residues per turn of an α-helix. Formation of 3_{10}-helices has been recognized as a widely occurring, yet minor contributor to overall protein secondary structure.[70, 118, 119] The presence of interactions characteristic of 3_{10}-helices at the termini of the helices in this study is consistent with the experimental finding that 3_{10}-helices are probable at the termini of alanine-based peptides.[120]

Table 3.1 Average properties of the hydrogen bond critial points of the three types of hydrogen bonds found in the backbone of the model helices. All units are in atomic units (au) except bond lengths, which are in angstroms (Å).

Bond type	Average bond length	Average $\rho(r_c)$	Average $\nabla^2\rho(r_c)$	Average ε
N-H···O *i* + 3	2.257	0.014	0.0445	0.1326
C-H···O *i* + 3	2.732	0.006	0.0209	0.1787
N-H···O *i* + 4	2.281	0.013	0.0425	0.0854

The three types of hydrogen bonds are distributed differently throughout the different helices; also, they have different interaction strengths. This implies that each hydrogen bond contributes differently to the stabilization of the helix. Table 3.2 lists the number of each type of bond for each helix model, as well as the total electron density at the HBCPs. Table 3.3 lists the total number of HBCPs for each model helix, as well as the total electron density and Laplacian of the electron density. In general, C-H•••O, $i + 3$ bonds are most abundant, however, due to their weak interaction strength, they usually contribute least to electron density at the HBCPs, which has been shown to be correlated with helix stabilization.[81] Both N-H•••O, $i + 3$ and N-H•••O, $i + 4$ bonds tend to contribute more electron density to the network of HBCPs than C-H•••O, $i + 3$ bonds do. However, which of these stronger types of interactions dominates depends on the amino acids residue that is substituted at the central position. For approximately half of the helix models, N-H•••O, $i + 3$ bonds contribute most to electron density at HBCPs and for the other half, N-H•••O, $i + 4$ bonds do.

3.3.1.2 Fourth Backbone Interaction

It should be noted that there is a fourth type of interaction that is common to all model helices. This interaction is not a hydrogen bond but an interaction between the backbone N at position i and the C=O group $i + 4$ positions away. Table 3.4 lists the number of N···O interactions in each helix as well as the properties at the BCPs. The N···O interactions are usually found at the centre of the helix and while they have significantly less electron density at the BCPs than the hydrogen bonds discussed above, the total contribution is on the order of a single hydrogen bond. As previously mentioned,

this contribution can be significant because the free energy of folding of a protein is on the order of magnitude of a single hydrogen bond.[85]

Table 3.2 Number of hydrogen-bonded critical points (nN-H•••O, i + 3,nC-H•••O, i + 3 and nN-H•••O, i + 3) for the three types of interactions found in helix backbones and total electron density ($\sum\rho(r_c)$) found at the bond critical points for each type of bond.

Central residue	nN-H···O, $i+3$	$\sum\rho(r_c)$ at N-H···O, $i+3$	nC-H···O, $i+3$	$\sum\rho(r_c)$ at C-H···O, $i+3$	nN-H···O, $i+4$	$\sum\rho(r_c)$ at N-H···O, $i+4$
Ala	5	0.0717	10	0.0568	6	0.0856
Gly	6	0.0876	9	0.0523	6	0.0791
Val	5	0.0678	10	0.0559	6	0.0851
Leu	7	0.1091	9	0.0526	4	0.0523
Met	5	0.0702	10	0.0587	7	0.0843
Ile	8	0.1115	9	0.0528	5	0.0601
Ser	7	0.1067	10	0.0613	5	0.0600
Thr	7	0.0787	10	0.0562	7	0.0862
Cys	7	0.1087	9	0.0557	5	0.0599
Pro	4	0.0597	8	0.0472	6	0.0867
Asn	6	0.0784	10	0.0621	6	0.0841
Gln	6	0.0781	10	0.0472	6	0.0725
Phe	7	0.1040	10	0.0566	5	0.0652
Tyr	7	0.1025	10	0.0571	5	0.0633
Trp	4	0.0623	10	0.0578	7	0.0745
His_c	6	0.0647	8	0.0410	4	0.0387
His_n	7	0.1036	9	0.0523	5	0.0661
Asp	4	0.0313	6	0.0386	4	0.0507
Glu	3	0.0413	7	0.0394	5	0.0679

Table 3.3 Number of hydrogen-bonded critical points (nHBCP), total electron density ($\sum\rho(r_c)$) and total Laplacian of electron density ($\sum\nabla^2\rho(r_c)$) found at the bond critical points.

Central residue	nHBCP	$\sum\rho(r_c)$	$\sum\nabla^2\rho(r_c)$
Ala	21	0.2141	0.7076
Gly	21	0.2190	0.7189
Val	21	0.2088	0.6938
Leu	20	0.2140	0.6933
Met	22	0.2132	0.7130
Ile	21	0.2244	0.7383
Ser	22	0.2280	0.7424
Thr	24	0.2211	0.7476
Cys	22	0.2243	0.7274
Pro	18	0.1936	0.6329
Asn	22	0.2246	0.7461
Gln	22	0.1978	0.7011
Phe	22	0.2258	0.7382
Tyr	22	0.2229	0.7326
Trp	21	0.1946	0.6534
His_c	18	0.1444	0.6349
His_n	21	0.2220	0.7231
Asp	14	0.1206	0.4612
Glu	15	0.1486	0.4894

Table 3.4 Number of N•••O interactions (nN•••O),), total electron density ($\sum\rho(r_c)$), total Laplacian of electron density ($\sum\nabla^2\rho(r_c)$) found at the bond critical points and average ellipticity (ε) of N•••O interaction.

Central residue	nN···O	$\sum\rho(r_c)$ at N···O	$\sum\nabla^2\rho(r_c)$ at N···O	$\sum\varepsilon$ at N···O	Average ε at N···O
Ala	5	0.0235	0.0929	7.6395	1.5279
Gly	3	0.0140	0.0548	5.0299	1.6766
Val	4	0.0186	0.0735	5.2683	1.3171
Leu	2	0.0092	0.0368	3.2149	1.6075
Met	5	0.0244	0.0981	6.2019	1.2404
Ile	1	0.0044	0.0170	1.9837	1.9837
Ser	3	0.0152	0.0572	3.5662	1.1887
Thr	3	0.0137	0.0549	4.0222	1.3407
Cys	3	0.0148	0.0563	4.6993	1.5664
Pro	2	0.0116	0.0464	2.0513	1.0257
Asn	4	0.0214	0.0840	4.0653	1.0163
Gln	3	0.0120	0.0484	4.0289	1.3430
Phe	3	0.0136	0.0526	7.5092	2.5031
Tyr	3	0.0138	0.0533	5.2401	1.7467
Trp	6	0.0284	0.1142	6.9667	1.1611
His_c	2	0.0069	0.0306	1.3497	0.6749
His_n	3	0.0138	0.0535	5.3509	1.7836
Asp	3	0.0141	0.0587	2.8804	0.9601
Glu	2	0.0099	0.0381	3.2016	1.6008

3.3.2 Effect of Substitution at the Central Position of the α-Helix Model

The total electron density at the hydrogen bond critical points ($\sum\rho(r_c)$), which has been found to be correlated with helix stabilization in the gas phase,[81] ranges from 0.1206 to 0.2280 au depending on the amino acid residue in the central position (Table 3.3). This large range (0.102 au) arises because amino acid residues with charged side groups tend to destabilize the helix significantly relative to the reference polyalanine model (For-AAAAAAAAAAAAA-NH_2). The range of electron density at HBCPs is much smaller for model helices with neutral amino acid side chains in the central position (0.1936-0.2280 au). This implies that for helices containing neutral amino acids, the difference in intrinsic stabilization energy due to the hydrogen bond network is on the order of a hydrogen bond (0.034 au). As previously mentioned, stabilization of this magnitude can be significant to protein folding. Therefore, some differences in the electron density of the hydrogen bond network found here may help to explain why some amino acids have been found to be helix stabilizers while others are helix destabilizers.

3.3.2.1 Comparison to Theory

Until recently, it has been impossible to study large systems, such as α-helices using quantum mechanical methods, due to the computational cost. Therefore, there are few theoretical studies that are suitable for comparison. However, Wieczorek and Dannenberg[121] examined the substitution of seven different amino acids near the central position of a polyalanine peptide using DFT for the backbone and semi-empirical methods for the side chains. They found that in the gas phase, substitutions of Gly, Leu and Ser for Ala in polyalanine caused stabilization of the helix relative to the component amino acids. They reported that the Leu and Ser substitutions resulted in very small

stabilizations. This is relatively consistent with the finding of this study that peptides containing Gly and Ser have stronger hydrogen bond networks, measured by total electron density, than the peptide with Ala at the central position. Also, the hydrogen bond network of the peptide containing Leu is very similar to that of the polyalanine peptide, which may explain why the stabilization found by Wieczorek and Dannenberg is small. That is, the intrinsic hydrogen bonding properties of helices containing Gly, Ser and Leu may partially explain why they have been found to be more stable than polyalanine in the gas phase.

In the same study, Wieczorek and Dannenberg found that substitutions of Val, Phe and Pro destabilize polyalanine. This is consistent with the QTAIM result that substitutions of Val, and Pro have weaker hydrogen bond networks than polyalanine. That is, the strength of the hydrogen bond networks of helices containing these residues may contribute to the fact that they destabilize polyalanine. The only major discrepancy between this study and the Wieczorek and Dannenberg study is the substitution of Phe in polyalanine. While they find that Phe destabilizes polyalanine, this study reports a stronger hydrogen bond network for the helix containing Phe. In this case, the strength of the hydrogen bond network cannot explain the previous theoretical finding. The reason may be the difference of methods used to describe the side chains (quantum mechanical versus semi-emperical), or else, factors other than the hydrogen bond network are leading to the destabilization of Phe-containing helices.

While there is only one theoretical study available for comparison, the present method of using the strength of the hydrogen bond network as a measure of stabilization is relatively consistent with the method of comparing the helix to the amino acid

components. This may indicate that the hydrogen bond network is an important intrinsic contributor to protein secondary structure.

3.3.2.2 Comparison to Experiment

Comparison of the hydrogen bond network of α-helices containing different amino acids would be most useful if it could help to explain the α-helix propensities of the amino acids found experimentally. For example, the substitution of Pro into an α-helical peptide is known to destabilize the helix because of the strain it introduces.[122-125] Results from this QTAIM study may complement this knowledge by suggesting an additional reason for the destabilization of Pro containing helices. The total electron density of the hydrogen bond network of the model containing Pro is the lowest of the neutral amino acids (0.1936 au); this is accompanied by the loss of some stabilizing hydrogen bonds near the N terminus and towards the centre of the strand. The disruption of the hydrogen bond network may contribute to Pro's low α-helix propensity.

However, gas phase calculations of peptides do not usually compare so favorably to experiment. In fact, there are two problems that impede the comparison of gas phase calculations with experimental reports: (1) the populations of non-helical peptides in experiment are unknown and (2) the effect of solvation has not been treated theoretically.[121]

The reason that (1) is problem stems from the fact that most experiments studying α-helix propensity determine the effect of amino acid mutation on α-stability with respect to a random coil. That is, while the mutation of an amino acid may stabilize the α-helix, it may stabilize a different conformation, such as a β-strand or a random coil, more. For example, Gly is known to stabilize the α-helix structure.[121, 123] This is consistent with the

finding that the model helix containing Gly has more electron density at the HBCPs than the polyalanine model helix (0.2190 au compared to 0.2140). However, the addition of Gly to a peptide stabilizes a random coil more because of the conformational flexibility in that state.[122, 123] Therefore, with the exception of Pro, Gly actually has the lowest propensity to form α-helices due to the entropy cost associated with the loss of conformational flexibility in the α-helix formation. In fact, the entropy costs associated with helix formation are thought to be a significant determinant of α-helix propensity.[126] This behavior cannot be accounted for ine present study. Potential methods for accounting for this effect are discussed in Chapter 4.

The second problem (2) arises from the importance of solvation in experimental studies. Depending on the environment, α-helices can be stabilized by solvation relative to random coil, or vice versa. Furthermore, it is thought that hydrophobicity is an important contributor to differences in α-helix propensities of amino acids.[122] That is, the side chains of residues in an α-helix have contact with the peptide backbone and are therefore removed from solvent. This creates a hydrophobic stabilization that contributes to the formation of α-helices. Again, this important behavior is not accounted for in this study. For example, substitution of Ser in the helix model results in a stronger hydrogen bonding network than for the polyalanine model (0.2280 compared to 0.2141 au). However, Ala has a larger hydrophobicity index (1.8) than Ser (-0.8).[127] Therefore, Ala experiences some additional stabilization in an α-helix due to desolvation of its side chain. This may partially account for the fact that Ala has a higher α-helix propensity than Ser.[122, 124, 125] The inclusion of solvation is discussed further in Chapter 4.

While the gas phase calculations presented in this thesis do not account for important entropic and hydrophobic effects, insights can still be provided if the results are examined with an understanding of these limitations. It has been found that entropic and hydrophobic effects cannot completely account for the α-helix propensities of amino acids.[128, 129] There are instances where amino acid residues with similar side chain flexibility or similar hydrophobicities have different α-helix propensities. In these cases, the intrinsic energy of the helix itself may be the dominating factor deciding α-helix propensity, which provides an opportunity for gas phase studies of the hydrogen bond network to provide insight.

3.3.2.2.1 Amino Acids with Non-Polar Side Chains

Hydrophobicity is not thought to contribute significantly to the α-helix propensity of amino acids with non-polar side chains.[129] In some cases, entropic factors can explain helix propensity, but in others it is not clear why some non-polar amino acids are favored over others. Table 3.5 presents the strength of the hydrogen bond network of five non-polar amino acids compared to experimental results for α-helix propensity.[‡] In general, a weaker hydrogen bond network corresponds to more destabilization (and smaller α-helix propensity) of model helices. The one exception is the substitution of Ile, which is found to stabilize the helix in gas phase calculations but is experimentally found to destabilize the helix. The destabilizing effect of Ile compared to Ala in α-helices has been attributed to the entropy loss that occurs when Ile is substituted into the helix.[125] However, experiment has not been able to explain why Ile is favored over Val in α-helices, as both have similar hydrophobic and entropic contributions. This QTAIM study suggests that Ile

[‡] Glycine is not included in the discussion because it is a special case due to its structure, as described in 3.3.2.2.

is favored in α-helices over Val because it results in a hydrogen bonding network that is stronger (by 0.0152 au) than the Val hydrogen bond network by approximately one hydrogen bond. Similarly, Leu is more stable in α-helices than Met because its hydrogen bond network is more stable.

Table 3.5 Strength of hydrogen bond network (au) and α-helix propensity (kcal/mol) of model helices substituted with Ile, Leu, Met and Val relative to model helix containing Ala in the central position. Positive result for strength result indicates hydrogen bond network is stronger than for Ala and negative result indicates a weaker hydrogen bond network. Positive number for propensity indicates the magnitude of destabilization.

Central Residue	Strength of Hydrogen Bond Network Relative to Ala	α-Helix Propensity Relative to Ala[125]
Ile	0.0099	0.41
Ala	0	0
Leu	-0.0001	0.21
Met	-0.0009	0.24
Val	-0.0053	0.61

3.3.2.2.2 Amino Acids with Comparable Polar and Charged Side Chains

Polar and charged side groups have a wide range of hydrophobic indexes and side chain flexibilities. Therefore, it is not realistic to simultaneously compare all polar and charged side groups on the basis of their effect on the hydrogen bond network. Within this group, hydrophobic and entropic properties, which are not accounted for in this study, will have a large role in α-helix propensity. However, there are certain subgroups with similar hydrophobicities and side chains and within these subgroups, the hydrogen bond network can a useful tool for explaining α-helix propensities.

The strength of the hydrogen bond network, as well as the experimental propensity to form α-helices, for Ser and Thr helices are presented in Table 3.6. Serine and threonine have similar hydrophobicities and similar side chains, yet serine has a higher propensity to form α-helices.[125] Results from the present QTAIM study suggest

that the substitution of Ser at the central position of polyalanine results in a stronger hydrogen bond network than Thr (0.0139 au compared to 0.0070 au), which results in a higher propensity for Ser to form α-helices.

Table 3.6 Strength of hydrogen bond network (au) and α-helix propensity (kcal/mol) of model helices substituted with Ser and Thr relative to model helix containing Ala in the central position. Positive result for strength result indicates hydrogen bond network is stronger than for Ala. Positive number for propensity indicates the magnitude of destabilization.

Central Residue	Strength of Hydrogen Bond Network Relative to Ala	α-Helix Propensity Relative to Ala[125]
Ala	0	0
Ser	0.0139	0.50
Thr	0.0070	0.66

Two more subgroups of amino acid residues that can be compared on the basis of the hydrogen bonding network are presented in Table 3.7 and Table 3.8. It has been suggested that the longer chained Glu and Gln are found to be more stable than the short chained Asp and Asn in α-helices because of entropic factors affecting the random coil.[125] However, this argument cannot explain why Asn is more stable than Asp and Gln is more stable than Glu, when substituted in the centre of α-helices. Because the side chains of these residues have similar hydrophobicities, other factors must contribute to the differing helix propensities in these cases. The results of the QTAIM study indicate that substituting Asn into the model helix results in a stronger hydrogen bond network than Asp (by 0.1040 au), contributing to Asn's higher α-helix propensity. Similarly, Gln, which has a higher helix propensity than Glu, has a stronger hydrogen bond network (by 0.0492 au) when substituted into an α-helix model.

Table 3.7 Strength of hydrogen bond network (au) and α-helix propensity (kcal/mol) of model helices substituted with Asn and Asp relative to model helix containing Ala in the central position. Positive result for strength result indicates hydrogen bond network is stronger than for Ala and negative result indicates a weaker hydrogen bond network. Positive number for propensity indicates the magnitude of destabilization.

Central Residue	Strength of Hydrogen Bond Network Relative to Ala	α-Helix Propensity Relative to Ala[125]
Ala	0	0
Asn	0.0105	0.65
Asp	-0.0935	0.69

Table 3.8 Strength of hydrogen bond network (au) and α-helix propensity (kcal/mol) of model helices substituted with Gln and Glu relative to model helix containing Ala in the central position. Negative result for strength result indicates hydrogen bond network is weaker than for Ala. Positive number for propensity indicates the magnitude of destabilization.

Central Residue	Strength of Hydrogen Bond Network Relative to Ala	α-Helix Propensity Relative to Ala[125]
Ala	0	0
Gln	-0.0163	0.39
Glu	-0.0655	0.40

Due to differing hydrophobicities and side chain flexibilities, it is difficult to compare other groups of amino acid side chains with experiment. However, based on the hydrogen bond network, the intrinsic preferences of amino acids to form α-helices in the absence of solvent can be presented as in Table 3.9. All hydrogen bond strengths are relative to Ala, as polyalanine is usually used as a reference model. Because Ala, the amino acid with the highest propensity to form helices, is in the middle of the group, it is clear that other forces also contribute to the stabilization of helices. The information in this table will be useful to future studies that attempt to model these other forces, for example, the effects of solvation.

Table 3.9 Strength of the hydrogen bond network (au) of model helices with different amino acids placed at the central position, relative to the polyalanine model helix. Residues are listed in order of network strength and a positive value indicates that the network is stronger than the polyalanine model.

Central Residue	Strength of Hydrogen Bond Network Relative to Ala
Ser	0.0139
Phe	0.0117
Asn	0.0105
Cys	0.0102
Ile	0.0099
Tyr	0.0088
His_n[§]	0.0079
Thr	0.0070
Gly	0.0049
Ala	0
Leu	-0.0001
Met	-0.0009
Val	-0.0053
Gln	-0.0163
Trp	-0.0195
Pro	-0.0205
Glu	-0.0655
His_c[**]	-0.0697
Asp	-0.0935

[§] Subscipt "n" refers to the unprotonated, neutral histidine.
[**] Subscript "c" refers to the protonated, positively charged histidine.

3.4 Conclusions

Model α-helical peptides containing all the natural amino acids, except lysine and arginine, have been constructed, optimized and characterized using QTAIM. Three types of hydrogen bonds have been identified that contribute to the stability of all model peptides. The values of $\rho(r_c)$ at the HBCPs distinguish the relative strengths of these bonds. The C-H•••O bonds found between side chains and backbone carbonyl groups three positions away are weaker than both types of N-H•••O bonds found within the backbone. However, the C-H•••O bonds are expected to collectively contribute to the overall stabilization of the α-helix. Both types of N-H•••O bonds identified also contribute significantly to the helix stabilization.

A weak N•••O interaction was also identified within the helix backbone. While only a small number of N•••O interactions were found in the centre of the helix, they were present in all helices and may contribute to helix stability.

In general, helix stability is a result of a balance of many contributing factors. Two of these factors, solvation and entropy, are not accounted for in the present study. As both factors are thought to be significant contributors to helix stability and differences in helix propensity, it is not expected that results from this study will completely agree with experimental results. Rather, this study accounts for a third contribution to helix stability, the hydrogen bond network, which is also thought to be very important.

In order to determine if the hydrogen bond network could be used to explain differences in helix propensities, groups of acid side chains with similar hydrophobicity and similar side chain flexibility were compared. Within these groups, hydrophobic and entropic effects are not as significant as other contributions, which must explain the

different helix propensities. In each case where amino acids had similar hydrophobic and entropic properties, the strength of the hydrogen bond network corresponded to relative α-helix propensity. That is, residues with higher helix propensities had stronger hydrogen bond networks (more electron density at the hydrogen bond critical points) than those with lower helix propensities. Therefore, hydrogen bonding could be used to explain helix propensities in cases where hydrophobicity and entropy arguments could not. This result provides promising evidence that further examination of the hydrogen bond network, which is discussed in Chapter 4, could provide even greater insight into helix stabilization.

Chapter 4. Conclusion

The computational study in this thesis represents one of the few available examples of quantum mechanical calculations applied to polypeptides. This chapter summarizes the conclusions and implications of the study and discusses avenues for future work.

4.1 Global Conclusions

The focus of this thesis has been the characterization of the hydrogen bond network of the α-helix, as well as the effect of amino acid substitution at the centre of the helix, with the aim to understand why amino acids favor or disfavor the α-helix formation. The structures of 19 model helices were optimized with density functional theory and the topology of the electron density derived from these calculations was analyzed according to the quantum theory of atoms in molecules. The analysis identified three types of hydrogen bonds present in α-helical peptides that contribute to the stabilization of the structure: 1) the well known bond between the backbone N-H group and the C=O group four residues away, 2) bonds between the backbone N-H and the C=O group three residues away and 3) bonds between side chain C-H groups and C=O groups three residues away have also been detected. This result is consistent with other theoretical and experimental studies that also described helix stabilization by way of these bonds. Furthermore, the results agree with theory and experiment on the relative strength of these three interactions, with the C-H•••O bonds being weaker than both types of N-H•••O bonds.

In addition to the three hydrogen bonds, an N•••O interaction was found between atoms three residues apart in the helix backbone. While this interaction is much weaker

than the hydrogen bonds described, it is common to all model helices in this structure and may partially contribute to helix stabilization.

In the literature described in this thesis, the stabilization of the α-helix has been attributed to a balance of different contributing factors. These factors include the effects of solvent as well as entropic effects associated with different conformations of the peptide. In some cases, these effects are predominant factors in the α-helix propensities of different amino acids. As the study of the hydrogen bonding network in the gas phase does not account for either of these factors, results from this study are not expected to, and do not completely agree with experimental results that report α-helix propensities. However, in subgroups of amino acids where solvation and entropic effects are not expected to dictate relative α-helix propensity, contributions from the hydrogen bond network were compared to experiment. In such cases, amino acids found to contribute to the strongest hydrogen bond networks also have the highest α-helix propensities. Therefore, contributions from the hydrogen bond network may provide an explanation of the relative helix propensities of these amino acids that experiment has not been able to provide.

The results of this study demonstrate the importance of hydrogen bonding in the stabilization of α-helices. They also demonstrate the potential for quantum mechanical calculations to provide insight into the forces that govern protein structure.

4.2 Future Work

In some ways, this thesis has laid the foundation for more in depth studies of the α-helix. Certainly, the effects of solvation are important for understanding how a helix behaves in its environment. However, the effects of solvation cannot be known without

first understanding the unsolvated system. Also, the effect of entropy plays a role in α-helix propensity; however, the entropy cost associated with α-helix formation cannot be accounted for without a structure for the α-helix. Potential research directions for incorporating these effects are discussed below. In addition to these suggestions, the research completed here, and the research suggested below can also be completed on β-sheet structures to provide insight into the stability of the other common secondary structural elements.

4.2.1 Inclusion of Solvation

Many experiments on α-helical peptides are performed in aqueous solution. However, in nature, α-helices can be shielded from the solvent. In addition, the in vivo environment contains ingredients other than water. Therefore, in order to truly understand the implications of experiments, it is necessary to better understand how solvation affects the stability of α-helices. Once the unsolvated system has been characterized, a solvation model can be included and both results can be used to better understand the effects of the environment on peptide stability. It will also be possible to determine if the effects of changing individual amino acids in the helix are due to solvation or due to intrinsic properties of the helix itself.

While a solvation study would provide significant insight into the forces stabilizing secondary structural elements, it is expected that the computational cost of including a solvation model to a system with this large number of atoms may be too high to be feasible at this time. However, with improving technology it is likely that such a study will be possible in the future.

4.2.2 Accounting for Entropic Effects

It has been suggested that the entropic costs associated with helix formation are high for some amino acids and that these amino acids would therefore favor another conformation. That is, even if a particular amino acid stabilizes an α-helix, it may stabilize another conformation more. Experiments often measure the stabilization of the α-helices relative to random coils. Therefore, in order to account for entropic effects, and to compare reasonably with experiment, theoretical studies should measure the stabilization of a peptide in α-helix formation relative to another conformation of the peptide that represents a mixture of random coils. The choice of one conformation to represent a mixture of random coils can be difficult because the substitution of different amino acids often affects the random coil mixture. The few quantum mechanical studies reported use an extended β-strand as a reference structure. A comparison of the β-strands of the peptides investigated in this study to the helices described here would be useful in quantifying the role of hydrogen bonding while accounting for some entropic effects.

The study of protein structure is a very active area of research. Reviews can be found from recent months that describe the importance of defining protein structure[130131, 132] and the work being carried out to accomplish this goal.[131, 132] News articles can be found from recent weeks that describe how knowledge of protein structure has provided important insights in many areas. For example, data on the structure of inclusion bodies may guide research efforts aimed at preventing these protein aggregates from trapping proteins desired for experiment or therapeutic use.[133, 133] Understanding the structure of spider silk proteins may aid in the design of new synthetic materials with similar strength and elasticity.[134] A structural study of the interaction of cholesterol and proteins in the

brain suggests a direct role for cholesterol in brain function.[135] Also, a protein found in the bird flu virus strain has been found to form tiny tubules that "hide" RNA during the infection, which would otherwise prompt an immune response; it may be possible to design drugs based on this knowledge to block this action.[136] It is certain that there are many areas where understanding protein structure is important and there is much work to be done towards this goal. It is hoped that this thesis provides some insight into the stabilizing forces governing protein secondary structure while stimulating new research directions in this area.

References

(1) Buxbaum, E. In *Fundamentals of Protein Structure and Function;* Springer: New York ; London, 2007, pp 367.

(2) UniProt Consortium *Nucleic Acids Res.* **2008**, *36*, D190-5.

(3) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.

(4) Anfinsen, C. B. *Science* **1973**, *181*, 223-230.

(5) Chen, C. C.; Singh, J. P.; Altman, R. B. *Bioinformatics* **1999**, *15*, 53-65.

(6) Baldi, P.; Pollastri, G.; Andersen, C. A.; Brunak, S. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2000**, *8*, 25-36.

(7) Edwards, Y. J.; Cottage, A. *Mol. Biotechnol.* **2003**, *23*, 139-166.

(8) Zhang, Y.; Skolnick, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1029-1034.

(9) Osguthorpe, D. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 146-152.

(10) Chandonia, J. M.; Karplus, M. *Proteins* **1999**, *35*, 293-306.

(11) Floudas, C. A. *Biotechnol. Bioeng.* **2007**, *97*, 207-213.

(12) Dobson, C. M.; Scaronali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868-893.

(13) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363-1376.

(14) Shi, Z.; Olson, C. A.; Bell, A. J., Jr; Kallenbach, N. R. *Biopolymers* **2001**, *60*, 366-380.

(15) Scheiner, S. *J. Phys. Chem. B* **2006**, *110*, 18670-18679.

(16) Wu, Y.; Zhao, Y. *J. Am. Chem. Soc.* **2001**, *123*, 5313-5319.

(17) Bour, P.; Keiderling, T. A. *J. Mol. Struct. (THEOCHEM)* **2004**, *675*, 95-105.

(18) Zumdahl, S. S.; Zumdahl, S. A. In *Chemistry;* Houghton Mifflin: Boston, 2007; , pp 1056.

(19) Dunlap, R. A. In *An Introduction to the Physics of Nuclei and Particles;* Thomson Brooks/Cole: Belmont, CA, 2004, pp 284.

(20) Griffiths, D. J. In *Introduction to Quantum Mechanics;* Pearson Prentice Hall: Upper Saddle River, N.J., 2005, pp 468.

(21) McQuarrie, D. A. In *Quantum Chemistry;* University Science Books: Mill Valley, Calif., 1983, pp 517.

(22) Levine, I. N. In *Quantum chemistry;* Prentice Hall; Prentice-Hall Canada: Englewood Cliffs, N.J.; Toronto, 1991, pp 629.

(23) Ramachandran, K. I.; Deepna, G.; Namboori, K. In *Computational Chemistry and Molecular Modeling Principles and Applications;* Springer-Verlag: Berlin, 2008, pp 419.

(24) Lewars, E. In *Computational Chemistry : Introduction to the Theory and Applications of Molecular and Quantum Mechanics;* Kluwer Academic: Boston, 2003, pp 471.

(25) Cramer, C. J. In *Essentials of Computational Chemistry : Theories and Models;* Wiley: Chichester, West Sussex, England ; Hoboken, N.J., 2004, pp 596.

(26) Jensen, F. In *Introduction to Computational Chemistry;* Wiley: Chichester ; New York, 1999, pp 429.

(27) Koch, W.; Holthausen, M. C. In *A Chemist's Guide to Density Functional Theory;* Wiley-VCH: Weinheim; New York, 2001, pp 300.

(28) McWeeny, R. In *Methods of Molecular Quantum Mechanics;* Theoretical chemistry; Academic Press: London ; New York, 1989, pp 573.

(29) Szabo, A.; Ostlund, N. S. In *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory;* Free Press; Collier Macmillan: New York; London, 1982, pp 446.

(30) Ratner, M. A.; Schatz, G. C. In *Introduction to Quantum Mechanics in Chemistry;* Prentice Hall: Upper Saddle River, NJ, 2001, pp 305.

(31) Slater, J. C. *Phys. Rev.* **1930**, *36*, 57-64.

(32) Boys, S. F. *Proc. R. Soc. Lond., A* **1950**, *200*, 542-554.

(33) Brillouin, L. *Actualities Sci. Ind.* **1934**, *71*, 159.

(34) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968-5975.

(35) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.

(36) Cramer, C. J. In *Essentials of Computational Chemistry : Theories and Models;* J. Wiley: West Sussex, England ; Rexdale, Ont., 2002, pp 542.

(37) Thomas, L. H. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542-548.

(38) Fermi, E. *Z. Phys.* **1928**, *48*, 73-79.

(39) Fermi, E. *Rend. Accad. Naz. Lincei* **1928**, *7*, 342-346.

(40) Fermi, E. *Rend. Accad. Naz. Lincei* **1927**, *6*, 602-607.

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

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

(43) Atkins, P. W.; Friedman, R. S. In *Molecular Quantum Mechanics;* Oxford University Press: New York, 1997, pp 545.

(44) Leach, A. R. In *Molecular Modelling : Principles and Applications;* Prentice Hall: Harlow, England ; Toronto, 2001, pp 744.

(45) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612-5626.

(46) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221-7230.

(47) Andrew C. Scheiner, Jon Baker, Jan W.Andzelm, *J. Comput. Chem.* **1997**, *18*, 775-795.

(48) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42-55.

(49) Redfern, P. C.; Blaudeau, J. P.; Curtiss, L. A. *J. Phys. Chem. A* **1997**, *101*, 8701-8705.

(50) Frenking, G.; Antes, I.; Bohme, M.; Dapprich, S.; Ehlers, A. W.; Jonas, V.; Neuhaus, A.; Otto, M.; Stegmann, R.; Veldkamp, A.; Vyboishchikov, S. F. In Lipkowitz, K. B., Boyd, D. B., Eds.; Reviews in Computational Chemistry; VCH: New York, 1996; Vol. 8, pp 63.

(51) Cohen, A. J.; Tantirungrotechai, Y. *Chem. Phys. Lett.* **1999**, *299*, 465-472.

(52) Dennis, J. E.; Schnabel, R. B. In *Numerical Methods for Unconstrained Optimization and Nonlinear Equations;* Prentice-Hall series in computational mathematics; Prentice-Hall: Englewood Cliffs, N.J., 1983, pp 378.

(53) Fletcher, R. In *Practical Methods of Optimization;* Wiley: Chichester ; New York, 1987, pp 436.

(54) Parr, R. G.; Yang, W. In *Density-Functional Theory of Atoms and Molecules;* International series of monographs on chemistry; Oxford University Press; Clarendon Press: New York; Oxford England, 1989; Vol. 16, pp 333.

(55) Matta, C. F.; Boyd, R. J. In *The Quantum Theory of Atoms in Molecules : From Solid State to DNA and Drug Design;* Wiley-VCH: Weinheim, 2007, pp 527.

(56) Bader, R. F. W. In *Atoms in Molecules : a Quantum Theory;* International series of monographs on chemistry; Clarendon Press: Oxford ; New York, 1994; Vol. 22, pp 438.

(57) Castillo, N.; Matta, C. F.; Boyd, R. J. *Chem. Phys. Lett.* **2005**, *409*, 265-269.

(58) Matta, C. F.; Hernández-Trujillo, H; Tang, T. H.; Bader, R.F.W. *Chem. --Eur. J.* **2003**, *9*, 1940-1951.

(59) Matta, C. F.; Hernandez-Trujillo, J. *J. Phys. Chem. A* **2003**, *107*, 7496-7504.

(60) Bader, R. F. W.; Cheeseman, J. R.; Laidig, K. E.; Wiberg, K. B.; Breneman, C. *J. Am. Chem. Soc.* **1990**, *112*, 6530-6536.

(61) Pearson, J. K.; Boyd, R. J. *J. Phys. Chem. A* **2007**, *111*, 3152-3160.

(62) Matta, C. F.; Castillo, N.; Boyd, R. J. *J. Chem. Phys.* **2006**, *125*, 204103.

(63) Domagala, M.; Grabowski, S.; Urbaniak, K.; Mloston, G. *J. Phys. Chem. A* **2003**, *107*, 2730-2736.

(64) Grabowski, S. J. *J. Phys. Chem. A* **2001**, *105*, 10739-10746.

(65) Domagala, M.; Grabowski, S. J. *J. Phys. Chem. A* **2005**, *109*, 5683-5688.

(66) Grabowski, S. J.; Sokalski, W. A.; Leszczynski, J. *J. Phys. Chem. A* **2005**, *109*, 4331-4341.

(67) Espinosa, E.; Molins, E.; Lecomte, C. *Chem. Phys. Lett.* **1998**, *285*, 170-173.

(68) Boyd, R. J.; Choi, S. C. *Chem. Phys. Lett.* **1986**, *129*, 62-65.

(69) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883-2890.

(70) Petr Bour, Jan Kubelka,Timothy A.Keiderling, *Biopolymers* **2002**, *65*, 45-59.

(71) Baldwin, R. L. *Trends Biochem. Sci.* **1989**, *14*, 291-294.

(72) Levitt, M.; Gerstein, M.; Huang, E.; Subbiah, S.; Tsai, J. *Annu. Rev. Biochem.* **1997**, *66*, 549-579.

(73) Aurora, R.; Creamer, T. P.; Srinivasan, R.; Rose, G. D. *J. Biol. Chem.* **1997**, *272*, 1413-1416.

(74) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519-538.

(75) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95-118.

(76) Miick, S. M.; Casteel, K. M.; Millhauser, G. L. *Biochemistry* **1993**, *32*, 8014-8021.

(77) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691-697.

(78) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200-9210.

(79) Toniolo, C.; Benedetti, E. *Trends Biochem. Sci.* **1991**, *16*, 350-353.

(80) Park, C.; Goddard, W. A. *J. Phys. Chem. B* **2000**, *104*, 7784-7789.

(81) Parthasarathi, R.; Raman, S. S.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2007**, *111*, 7141-7148.

(82) Park, C.; Carlson, M. J.; Goddard, W. A. *J. Phys. Chem. A* **2000**, *104*, 2498-2503.

(83) Ireta, J.; Neugebauer, J.; Scheffler, M.; Rojo, A.; Galvan, M. *J. Phys. Chem. B* **2003**, *107*, 1432-1437.

(84) Vener, M. V.; Egorova, A. N.; Fomin, D. P.; Tsirelson, V. G. *Chem. Phys. Lett.* **2007**, *440*, 279-284.

(85) Wang, Z.; Wu, C.; Lei, H.; Duan, Y. *J.Chem.Theory Comput.* **2007**, *3*, 1527-1537.

(86) Prigogine, I.; Rice, S. A. In Friesner, R. A., Ed.; Computational Methods for Protein Folding. Advances in Chemical Physics; J. Wiley & Sons: New York, 2001

(87) Stocker, U.; van Gunsteren, W. F. *Proteins* **2000**, *40*, 145-153.

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

(89) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139-145.

(90) Branden, C.; Tooze, J. In *Introduction to Protein Structure;* Garland Pub.: New York, 1991, pp 302.

(91) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. In *Lehninger Principles of Biochemistry;* W.H. Freeman: New York, 2005

(92) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198-14205.

(93) Horvath, V.; Varga, Z.; Kovacs, A. *J. Phys. Chem. A* **2004**, *108*, 6869-6873.

(94) Koch, O.; Bocola, M.; Klebe, G. *Proteins* **2005**, *61*, 310-317.

(95) Lario, P. I.; Vrielink, A. *J. Am. Chem. Soc.* **2003**, *125*, 12787-12794.

(96) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 14065-14071.

(97) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205-9212.

(98) Popelier, P. L. A.; Bader, R. F. W. *J. Phys. Chem.* **1994**, *98*, 4473-4481.

(99) Chang, C.; Bader, R. F. W. *J. Phys. Chem.* **1992**, *96*, 1654-1662.

(100) Parthasarathi, R.; Subramanian, V. In Grabowski, S. J., Ed.; Hydrogen Bonding: New Insights, Challenges and Advances in Computational Chemistry and Physics; Springer: Dordrecht, 2006; Vol. 3, pp 1.

(101) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747-9754.

(102) Carroll, M. T.; Bader, R. F. W. *Mol. Phys.* **1988**, *65*, 695-722.

(103) Boyd, R. J.; Choi, S. C. *Chem. Phys. Lett.* **1985**, *120*, 80-85.

(104) Parthasarathi, R.; Amutha, R.; Subramanian, V.; Nair, B. U.; Ramasami, T. *J. Phys. Chem. A* **2004**, *108*, 3817-3828.

(105) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 5286-5290.

(106) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632-1641.

(107) Schaftenaar, G.; Noordik, J. H. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123-134.

(108) Dennington II, R.; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W. L.; Gilliland, R. *Gaussview, Version 3.09* Semichem, Inc. Shawnee Mission, KS, **2003**.

(109) Farrag, S. Stability of Secondary Structure Motifs in Proteins, Dalhousie University, 2008. BSc Honours Thesis

(110) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264-6271.

(111) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

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

(113) Johnson, E. R.; DiLabio, G. A. *Chem. Phys. Lett.* **2006**, *419*, 333-339.

(114) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. **2003**, *Gaussian 03, Revision, B.05*, Gaussian, Inc., Wallingford CT, **2004**.

(115) Biegler-Konig, F.; Schonbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545-559.

(116) Tang, T.; Hu, W.; Yan, D.; Cui, Y. -. *J. Mol. Struct. (THEOCHEM)* **1990**, *207*, 319-326.

(117) Jiménez, A. I.; Ballano, G.; Cativiela, C. *Angew. Chem. Int. Ed.* **2005**, *44*, 396-399.

(118) Bolin, K. A.; Millhauser, G. L. *Acc. Chem. Res.* **1999**, *32*, 1027-1033.

(119) Toniolo, C.; Benedetti, E. *Trends Biochem. Sci.* **1991**, *16*, 350-353.

(120) Millhauser, G. L.; Stenland, C. J.; Hanson, P.; Bolin, K. A.; van de Ven, F. J. *J. Mol. Biol.* **1997**, *267*, 963-974.

(121) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2005**, *127*, 17216-17223.

(122) Blaber, M.; Zhang, X. J.; Matthews, B. W. *Science* **1993**, *260*, 1637-1640.

(123) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669-673.

(124) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648-1652.

(125) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422-427.

(126) Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5937-5941.

(127) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105-132.

(128) Avbelj, F.; Moult, J. *Biochemistry (N. Y.)* **1995**, *34*, 755-764.

(129) Wang, J.; Purisima, E. O. *J. Am. Chem. Soc.* **1996**, *118*, 995-1001.

(130) Munoz, I. G.; Blanco, F. J.; Montoya, G. *Clin. Transl. Oncol.* **2008**, *10*, 204-212.

(131) Zhang, Y. *Curr. Opin. Struct. Biol.* **2008**, *18*, 342-348.

(132) Fersht, A. R. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 650-654.

(133) Kemsley, J. *Chem. Eng. News* **2008**, *86*, 31.

(134) Townsend, M. "INADEQUATE Analysis of Protein Structure" Research News Volume 12, October 21, 2008, *Chemical Biology*, December 8, 2008, http://www.rsc.org/Publishing/Journals/cb/Volume/2008/12/protein_structure.asp

(135) "Biophysicists create new model for protein-cholesterol interactions in brain tissue" Sciencecentric.com, October 27, 2008, *Science Centric,* December 8, 2008, http://www.sciencecentric.com/news/article.php?q=08102726.

(136) Gutierrez, G. "Protein 'tubules' free avian flu virus from immune recognition" Science News, November 6, 2008, *Baylor College of Medicine*, December 8, 2008, http://www.bcm.edu/news/item.cfm?newsID=1260