Quantum Mechanical Studies of N-H···N Hydrogen Bonding in Acetamide Derivatives and Amino Acids

Sandra J. Lundell
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Chemistry Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/7309
QUANTUM MECHANICAL STUDIES OF N-H···N HYDROGEN BONDING
IN ACETAMIDE DERIVATIVES AND AMINO ACIDS

by

Sandra J. Lundell

A thesis submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
Chemistry

Approved:

Steve Scheiner
Major Professor

David Farrelly
Committee Member

Alexander I. Boldyrev
Committee Member

Richard S. Inouye, Ph.D.
School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah
2018
Copyright © Sandra J. Lundell 2018

All Rights Reserved
ABSTRACT

Quantum Mechanical Studies of N-H···N Hydrogen Bonding in Acetamide Derivatives and Amino Acids

by

Sandra Lundell, Master of Science

Utah State University, 2018

Major Professor: Dr. Steve Scheiner
Department: Chemistry and Biochemistry

The stability and structure of proteins is due, in part, to extensive intramolecular hydrogen bonding networks. The most common of these, which has been known for decades, is the N-H···O bond. Large numbers of these form between amide groups along the peptide backbone and are necessary for the structures of α-helices and β-pleated sheets. Recently, the complete characterization of other types of hydrogen bonds that occur in proteins have gained interest and among these is the N-H···N hydrogen bond. A small number of amino acids have been reported to form N-H···N hydrogen bonds in recent years, yet a full investigation of the essential amino acids has not been done. This thesis is focused on expanding the investigation of N-H···N hydrogen bonds to a wider group of amino acids, including both polar and nonpolar residues. Better understanding of the electronic properties of these bonds will have applications in curing diseases, pharmaceutical development, and other areas.

There were two types of computational studies designed to identify N-H···N
hydrogen bonds in this thesis. The first used three simple acetamide derivatives to mimic portions of protein backbone and complexed them together. In five of the six complexes, stable N-H---N hydrogen bonds formed in structures representing local minima. Next, ten amino acids were complexed with N-methylacetamide and six of these formed N-H···N hydrogen bonds. The amino acids were a mix of polar and nonpolar residues and the N-H···N bonds were commonly stabilizing. Researchers can use this work to aid in further understanding the noncovalent interactions that provide the structure and stability of proteins. The computational studies also provide a knowledge base that should help guide future work in this area of research.

(128 pages)
PUBLIC ABSTRACT

Quantum Mechanical Studies of N-H···N Hydrogen Bonding

in Acetamide Derivatives and Amino Acids

Sandra Lundell

Proteins are made of vast chains of amino acids that twist and fold into intricate designs. These structures are held in place by networks of noncovalent interactions. One of these, the hydrogen bond, forms bridges between adjacent pieces of the protein chain and is one of the most important contributors to the shape and stability of proteins. Hydrogen bonds come in all shapes and sizes and a full understanding of these not only aids in our understanding of proteins in general but can bridge the gap to finding cures to many protein-related diseases, such as sickle-cell anemia. The primary aim of this thesis is to discover if a specific type of hydrogen bond, the N-H···N bond, occurs within proteins and if so, if it contributes to the structure and stability of proteins.
ACKNOWLEDGMENTS

I would like to extend a very sincere thank you to everyone in the Chemistry Department at Utah State. I am grateful for all of the help and guidance from the professors. A special thank you to Dr. Steve Scheiner for the direction, teaching, and explanations throughout my time working with him.

To my friends and former co-workers, Dr. Vincent de Paul N. Nziko and Dr. Binod Nepal, I want to thank you for all your help. You both were instrumental in my success and I wish you the best in the future. Also, to Buck Banham, Dr. Tapas Kar, and Scott Nielson who always fixed whatever I broke, and to Geri, Cara, and Maury, who knit the department together and without whom, I never would have navigated the paperwork.

My journey in chemistry never would have endured if not for my beloved roommates Dr. Alison Webb and Nat Freestone. Our memories are unforgettable, cherished, and completely inappropriate to be recalled here. I love you both so much and I thank you for all the support you’ve given me.

To my lifelong friend, Deja Fife, you have taught me what courage and bravery are. I’ve learned that at times it is screaming hysterically and trying not to cry on the log ride you insisted we go on, and sometimes it is the quiet decision to face another day, though you’re not confident you can make it through. You are an angel on earth. And for the last time, it’s not my fault the mattress ate me. Blame Romper.

It’s a bird, it’s a plane, no, it’s my Superman. My endless gratitude goes to Curtis Brown. You’ve been there at my very worst and still you love me the same. Everyone
should have a best friend like you. Thank you for always believing I could do absolutely anything. It turns out, you were right.

I am indebted to my graduate committee, Professors D. Farrelly, A. Boldyrev, S. Bialkowski, and F. Edwards. Farrell, thank you for letting us take naps in your class, and even more so for unravelling the mysteries of the quantum universe. David, your support these past six months has been invaluable to me, and you’re not rid of me yet. I look forward to my doctoral research with you.

Acknowledging everyone I wish to would be impossible, but to those I have not yet mentioned and who believed in me, I give my most humble appreciation: E.K., J.B., B.L., R.R., K.d.J., G.L., M.L., M.S., R.B.; and to J.M., thank you.

Finally, this piece is not complete without the most wonderful woman in my life. I am eternally grateful to my angel, goddess, and wife, Katie Lundell. You are my shining star and dearest friend, and writing about who you are to me would be a thesis on its own. Thank you for your endless patience and support through the long nights, set-backs, and uncertainty. This thesis would not exist without your love and so I humbly dedicate it to you.

Sandra Lundell
## CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT .................................................................................................................. iii</td>
</tr>
<tr>
<td>PUBLIC ABSTRACT ....................................................................................................... v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS .................................................................................................... vi</td>
</tr>
<tr>
<td>LIST OF TABLES ......................................................................................................... x</td>
</tr>
<tr>
<td>LIST OF FIGURES ....................................................................................................... xii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS ............................................................................................ xv</td>
</tr>
<tr>
<td>1. INTRODUCTION AND LITERATURE REVIEW ............................................................ 1</td>
</tr>
<tr>
<td>1.1. Hydrogen Bonds .................................................................................................. 1</td>
</tr>
<tr>
<td>1.2. Hydrogen Bonds in Proteins .............................................................................. 4</td>
</tr>
<tr>
<td>1.3. N-H···N Bonds in Proteins ............................................................................... 6</td>
</tr>
<tr>
<td>1.4. Conclusion .......................................................................................................... 21</td>
</tr>
<tr>
<td>References .................................................................................................................. 24</td>
</tr>
<tr>
<td>2. IDENTIFICATION AND CHARACTERIZATION OF N-H···N HYDROGEN BONDS BETWEEN PROTEIN BACKBONE MIMICS .............................................................. 29</td>
</tr>
<tr>
<td>2.1. Abstract .............................................................................................................. 29</td>
</tr>
<tr>
<td>2.2. Introduction ....................................................................................................... 29</td>
</tr>
<tr>
<td>2.3. Computational Methods ..................................................................................... 33</td>
</tr>
<tr>
<td>2.4. Results and Discussion ..................................................................................... 34</td>
</tr>
<tr>
<td>2.5. Conclusion .......................................................................................................... 55</td>
</tr>
<tr>
<td>References .................................................................................................................. 57</td>
</tr>
<tr>
<td>3. IDENTIFICATION AND CHARACTERIZATION OF SIDECHAIN-BACKBONE AND BACKBONE-BACKBONE N-H···N HYDROGEN BONDS BETWEEN AMINO ACID MIMICS .............................................................. 60</td>
</tr>
<tr>
<td>3.1. Abstract .............................................................................................................. 60</td>
</tr>
<tr>
<td>3.2. Introduction ....................................................................................................... 60</td>
</tr>
<tr>
<td>3.3. Computational Methods ..................................................................................... 62</td>
</tr>
<tr>
<td>3.4. Results and Discussion ..................................................................................... 62</td>
</tr>
<tr>
<td>3.5. Conclusion .......................................................................................................... 79</td>
</tr>
<tr>
<td>References .................................................................................................................. 80</td>
</tr>
</tbody>
</table>
APPENDIX A: COMPUTATIONAL METHODS ......................................................... 82

A.1. Gaussian Methods, and Basis Sets ............................................................. 83
A.2. Counterpoise Correction and Basis Set Superposition Error ...................... 87
A.3. Natural Bond Orbital Analysis .................................................................... 89
A.4. Symmetry-Adapted Perturbation Theory .................................................... 91
A.5. Electron Density Shift Map ........................................................................ 94
A.7. Atoms in Molecules Analysis ................................................................... 95
References ........................................................................................................ 97

APPENDIX B: COPYRIGHT PERMISSIONS ....................................................... 101
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1.</td>
<td>The geometry criteria set by Deepak et al. to determine the presence of hydrogen bonds in Data Set I and II. AA – the acceptor antecedent atom</td>
</tr>
<tr>
<td>2-1.</td>
<td>Properties of the N-H bonds in NHH and NHC. The NHH (A) and NHH (B) labels are depicted in Fig. 2-2.</td>
</tr>
<tr>
<td>2-2.</td>
<td>Geometry properties of five dimers (donor/acceptor): the change in the N-H bond length (dimer-monomer) in angstroms; the change in the vibrational frequency of N-H (dimer-monomer) in cm$^{-1}$; the N···H bond length in angstroms; the N-H···N bond angle in degrees. Data is reported with both the MP2/6-31+g* and MP2/aug-cc-pvdz levels of theory where possible.</td>
</tr>
<tr>
<td>2-3.</td>
<td>The binding energy ($E_b$) and interaction energy ($E_i$) of the five dimers reported in kcal/mol. The change in the isotropic shielding of the bridging proton reported in parts per million (dimer-monomer).</td>
</tr>
<tr>
<td>2-4.</td>
<td>The electron density ($\rho$) and the Laplacian of the electron density ($\nabla^2 \rho$) in atomic units at the bond critical point of the N-H···N bond for each dimer. Values calculated with MP2/aug-cc-pvdz level of theory except NHC/NHC (MP2/6-31+g*).</td>
</tr>
<tr>
<td>2-5.</td>
<td>Second-order perturbation E(2) energies (kcal/mol) of the intermolecular interactions of the five dimers, calculated with NBO analysis. Results were calculated with MP2/aug-cc-pvdz except where indicated *<em>, which is MP2/6-31+g</em>.</td>
</tr>
<tr>
<td>2-6.</td>
<td>The dissection of the interaction energy, $E_i$, by SAPT analysis into electrostatic, induction, dispersion, and exchange forces.</td>
</tr>
<tr>
<td>2-7.</td>
<td>Geometry properties of the trimer complexes (donor/acceptor): the change in the N-H bond length (trimer-monomer) in angstroms; the change in the vibrational frequency of N-H (trimer-monomer) in cm$^{-1}$; the N···H bond length in angstroms; and the N-H···N bond angle in degrees. The binding energy ($E_b$) and interaction energy ($E_i$) of the five dimers reported in kcal/mol. Results were calculated with</td>
</tr>
</tbody>
</table>
2-8. Second-order perturbation E(2) energies of the intermolecular interactions of the trimers, calculated with NBO analysis. Results calculated with MP2/aug-cc-pvdz and reported in kcal/mol.

3-1. Properties of the N-H bonds in the amino acid mimics and NHC: the N-H bond length in angstroms; the N-H vibrational frequency in cm\(^{-1}\); and the isotropic shielding of the hydrogen of N-H in parts per millions. The A-C labels correlate with those in Fig. 3-3.

3-2. The binding energy (E\(_b\)) and interaction energy (E\(_i\)) of the five dimers in kcal/mol. The change in the isotropic shielding of the bridging proton reported in parts per million (dimer-monomer). Dimer labeling is by the amino acid mimic. Each was paired with NHC.

3-3. Geometry properties of nine dimers: the change in the N-H bond length (dimer-monomer) in angstroms; the change in the vibrational frequency of N-H (dimer-monomer) in cm\(^{-1}\); the N···H bond length in angstroms; and the N-H···N bond angle in degrees. Data is reported with MP2/aug-cc-pvdz.

3-4. Second-order perturbation E(2) energies of the intermolecular interactions between each amino acid and NHC calculated with NBO analysis. Results calculated with MP2/aug-cc-pvdz and reported in kcal/mol.
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>In a hydrogen bond, the D-H σ* orbital gains electron density from a lone pair or π-orbital.</td>
<td>1</td>
</tr>
<tr>
<td>1-3</td>
<td>The secondary structure of an α-helix and β-pleated sheet holds its shape with hydrogen bonds.</td>
<td>5</td>
</tr>
<tr>
<td>1-4</td>
<td>The computationally determined structure of D,L,-Histidyl-L,D-histidine pentahydrate has a N$<em>2$D-H···N$</em>{1D}$ H-bond.</td>
<td>7</td>
</tr>
<tr>
<td>1-5</td>
<td>The structure of His24 and His119 in the sperm whale apomyoglobin protein. The geometry of the imidazole groups suggests the formation of an N-H···N hydrogen bond.</td>
<td>8</td>
</tr>
<tr>
<td>1-6</td>
<td>Conformations of the four proline residues in nSH3 based on the crystal structure (PDB ID 1CKA), with φ and ψ angles indicated. Green: carbon, blue: nitrogen, white: hydrogen, red: oxygen, hydrocarbon hydrogen atoms not shown.</td>
<td>11</td>
</tr>
<tr>
<td>1-7</td>
<td>The N-terminal Src homology 3 domain from the murine Crk-II adaptor protein with Pro152, Pro183, Pro185, and Pro165 shown. Pro152 and Pro183 are incapable of forming N-H···N bonds due to their position at the perimeter of the protein.</td>
<td>11</td>
</tr>
<tr>
<td>1-8</td>
<td>(A) The chemical structure of Ace-Pro-NMe, used for quantum chemical calculations. The dihedral angles φ and ψ were varied in a step-wise fashion and the energy profile determine. (B) The resulting structure with the most stable configuration.</td>
<td>15</td>
</tr>
<tr>
<td>1-9</td>
<td>The model compounds N-methylformamide and N-acetyl pyrrolidine are involved in the N-H···N H-bond. The distance $d$ between the N-H group and the N was varied from 2.1 to 4.1Å. (B) The interaction energy profile between the two molecules as a function of the distance $d$.</td>
<td>15</td>
</tr>
<tr>
<td>1-10</td>
<td>Molecular plots of Ace-Pro-NMe, representing different regions of the potential energy profile and corresponding to the points A-E in</td>
<td></td>
</tr>
</tbody>
</table>
1-11. The potential energy profile generated from the single-point energy calculation of the model compound Ace-Pro-Nme as a function of $\psi$. (See Fig. 1-7).

1-12. Histograms showing the distribution of BSSE-corrected interaction energies for imidazole pairs participating in N-H···N hydrogen bonds: (A) both imidazole rings are neutral and (B) donor imidazole is protonated and the acceptor imidazole is neutral.

2-1. Conformations of the four proline residues in nSH3 based on the crystal structure (PDB ID 1CKA), with $\phi$ and $\psi$ angles indicated. Green: carbon, blue: nitrogen, white: hydrogen, red: oxygen, hydrocarbon hydrogen atoms not shown.

2-2. The optimized structure of (I) NHH, (II) NHC, and (III) NCC; and the corresponding molecular electrostatic maps. (red) electron-rich (blue) electron-poor.

2-3. The structures of (I) NHC/NHC, (II) NHC/NCC, (III) NHH/NCC, and (IV) NHH/NHH. Structure I optimized at the MP2/6-31+g* level of theory. Structures II-IV optimized at the MP2/aug-cc-pvdz level of theory.

2-4. The gain (red) and loss (blue) of electron density upon complexation of the monomers. The pattern along the N-H···N bond is indicative of a H-bond.

2-5. The gain (red) and loss (blue) of electron density upon complexation of (I) NHC/NCC, (II) NHH/NHC, (III) NHH/NCC. The pattern along the N-H···N bond is indicative of a H-bond.

2-6. The gain (red) and loss (blue) of electron density upon complexation of the NHH/NHH dimer.

2-7. The optimized structure of NHC/NHC + water with $N_{\text{donor}}$-H···$N_{\text{acceptor}}$, $O$-H···$N_{\text{donor}}$, and $N_{\text{acceptor}}$-H···$O$ shown. Structure optimized with MP2/aug-cc-pvdz.

2-8. The optimized structures of the NHC/NCC + water complexes labeled I, II, and III.
2-9. The optimized structures of the NHH/NHH + water complexes labeled I and II ................................................................. 55

3-1. The general structure of the amino acid mimics used in this study. The R group represents the side chain of each amino acid. ....................... 63

3-2. The optimized structures of (I) valine, (II) cysteine, and (III) leucine mimics. Optimizations were calculated with MP2/aug-cc-pvdz........... 65

3-3. The optimized structures of (I) serine, (II) asparagine, and (III) glutamine mimics. Labels A-C indicate hydrogen atoms in (II) and (III). Optimizations were calculated with MP2/aug-cc-pvdz......................... 66

3-4. The optimized structure of N-methylacetamide (NHC), the molecule paired with amino acids residues to mimic protein intramolecular bonding ........................................................................................................ 67

3-5. The optimized structures of the dimers labeled: (I) valine (II) leucine (III) cysteine A, and (III) cysteine B................................................................. 68

### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUPAC</td>
<td>International Union of Pure and Applied Chemistry</td>
</tr>
<tr>
<td>D</td>
<td>Hydrogen-Bond Proton Donor</td>
</tr>
<tr>
<td>A</td>
<td>Hydrogen-Bond Proton Acceptor</td>
</tr>
<tr>
<td>σ*</td>
<td>Antibonding Sigma Orbital</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>$^{2}\text{J}_{\text{NN}}$</td>
<td>The triplet coupling constant between two $^{15}\text{N}$ atoms</td>
</tr>
<tr>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>B3LYP</td>
<td>A DFT quantum mechanical method with two types of exchange correlation.</td>
</tr>
<tr>
<td>3-11+g(d,p)</td>
<td>A Pople basis set with polarization and diffuse functions on heavy atoms.</td>
</tr>
<tr>
<td>NBO</td>
<td>Natural Bond Orbital Analysis</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlation Spectroscopy</td>
</tr>
<tr>
<td>pKa</td>
<td>A measure of acid strength.</td>
</tr>
<tr>
<td>Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>BP86</td>
<td>A DFT quantum mechanical method.</td>
</tr>
<tr>
<td>Triple-ξ def2-TZVP</td>
<td>A Karlsruhe basis set with valence triple-zeta polarization.</td>
</tr>
<tr>
<td>M06-2X</td>
<td>A DFT quantum mechanical method.</td>
</tr>
<tr>
<td>Aug-cc-pvqz</td>
<td>Dunning’s augmented correlation consistent polarized</td>
</tr>
</tbody>
</table>
valence quadruple zeta basis set

\( E_{\text{DH-AH}} \) The single point energy of the structure.

\( E_{\text{DH}} \) The single point energy of the donor imidazole.

\( E_{\text{AH}} \) The single point energy of the acceptor imidazole.

SASA Solvent Accessible Surface Area

NHH Acetamide

NHC \( N\)-methyacetamide

NCC \( N,N\)-dimethylacetamide

MEP Molecular Electrostatic Potential

BSSE Basis Set Superposition Error

SAPT Symmetry-Adapted Perturbation Theory

GIAO Gauge-Independent Atomic Orbital Theory

AIM Atoms in Molecules Theory

\( E(2) \) Second Order Perturbation Energy

6-31+g* A Pople basis set with diffusion and polarization functions.

MP2 Second-order Møller-Plesset perturbation theory

aug-cc-pVDZ Dunning’s augmented correlation consistent polarized valence double zeta basis set

\( E_b \) Binding Energy

\( E_i \) Interaction Energy

HF Hartree-Fock Theory

RS-PT Rayleigh-Schrödinger Perturbation Theory
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STO</td>
<td>Slater Type Orbital</td>
</tr>
<tr>
<td>GTO</td>
<td>Gaussian Type Orbital</td>
</tr>
<tr>
<td>AO</td>
<td>Atomic Orbital</td>
</tr>
<tr>
<td>CP</td>
<td>Counterpoise Correction</td>
</tr>
<tr>
<td>NAO</td>
<td>Natural Atomic Orbital</td>
</tr>
<tr>
<td>NHO</td>
<td>Natural Hybrid Orbital</td>
</tr>
<tr>
<td>NLS</td>
<td>Natural Lewis Structure</td>
</tr>
<tr>
<td>MO</td>
<td>Molecular Orbital</td>
</tr>
<tr>
<td>NMB</td>
<td>Natural Minimal Basis</td>
</tr>
<tr>
<td>RS</td>
<td>Rayleigh-Schrödinger</td>
</tr>
<tr>
<td>QTAIM</td>
<td>Quantum Theory of Atoms in Molecules</td>
</tr>
</tbody>
</table>
1.1. Hydrogen Bonds

One of the most crucial chemical interactions in biological systems is the hydrogen bond. The textbook definition of a hydrogen bond (H-bond) is an electrostatic attraction between a H atom covalently bonded to O, N, or F (highly electronegative atoms) and the lone pair of another such atom nearby.\textsuperscript{1-3}

![Figure 1-1. In a hydrogen bond, the D-H $\sigma^*$ orbital gains electron density from a lone pair or $\pi$-orbital. Reprinted by permission of Pearson Education, Inc., New York, New York.\textsuperscript{4}}]

This definition, while not incorrect, has proven to be too narrow and a new definition was proposed in 2011 by the International Union of Pure and Applied Chemistry (IUPAC).\textsuperscript{5} Evidence had shown that the hydrogen donor (D) does not always need be one of the electronegativity giants, N, O, or F. An atom with a higher electronegativity than hydrogen will suffice. Furthermore, the role of the hydrogen acceptor (A) is to provide electron density. This can be done with the traditional lone pair, but also with a $\pi$-bond of a double or triple bond (C=O for example).\textsuperscript{6-12} This new
definition expands the field to include many new interactions that have been proposed as H-bonds that include less electronegative atoms such as Cl, S, P, C, or even metals to act as the proton donor or double and triple bonds as the acceptor.\textsuperscript{13-14}

It is now known that contrary to what was formerly believed, the forces involved in H-bond formation are not solely electrostatic. Many studies show that electrostatic and dispersion forces play nearly an equal role, followed by a smaller but still significant induction force.\textsuperscript{13-14} A H-bond forms when the antibonding orbital of D-H (\(\sigma^*\)) accepts electron density from the lone pair or \(\pi\)-orbital of A. Because of this, H-bonds are strongest when the D-H···A angle is linear, and the orbitals align accordingly, though deviations of more than 80\(^\circ\) have been reported.\textsuperscript{5,12,15}

Along with bond angle, the bond lengths of D-H and H···A are also important, largely in how they change upon H-bond formation. The attraction between the hydrogen and acceptor draws the H atom away from the donor, elongating and subsequently weakening the D-H bond while the H···A bond becomes shorter and stronger. The degree of these shifts is proportional to the strength of the H-bond.\textsuperscript{5,12,15}

H-bonds can often be detected using spectroscopic methods. For example, the lengthening of the D-H bond can cause a red shift (decrease) in the infrared D-H stretching frequency that is proportional to the H-bond strength along with a broadening of the signal.\textsuperscript{12,16-19} It is worth noting, however, that increases (blue shifts) in the frequencies sometimes occur.\textsuperscript{16,20-28} As well as a chemical shift, new vibrational modes associated with the formation of the H···A bond are generated.

The source of the blue-shift is very much debated in the current literature though
it’s commonly agreed that whether red- or blue-shift occurs is determined by a pair of competing forces. Li et al. proposed that the competition between the short-range Pauli repulsion of D-H and the long-range electrostatic attraction between D and A is the origin of both the red- and blue-shifting bonds.\textsuperscript{29} Alternatively, Joseph and Jemmis proposed that the heightened electrostatic attraction between D and H caused by the presence of A, and the electrostatic attraction between H and A shortens the H···A distance and elongates the D-H bond.\textsuperscript{30}

Still other explanations exist, but most agree that other than the spectroscopic differences, red- and blue-shifted H-bonds are similar.\textsuperscript{31-38} Also, blue-shifts most commonly occur when the donor is a carbon atom.\textsuperscript{9,22-24,26,39-45} As this thesis only considers nitrogen as a donor, it should be reasonable to only consider red-shifts to identify H-bonds.

Nuclear Magnetic Resonance (NMR) can also be used to detect the formation of H-bonds. The electron shielding around the H decreases during H-bond formation as electron density shifts away from H and closer to both D and A. This so-called deshielding is detectable as the proton experiences a higher external magnetic field and subsequently, a higher frequency is required to achieve resonance.\textsuperscript{46}

There are numerous examples demonstrating that H-bonding is necessary for life. For example, intermolecular O-H···O bonds in water provide the source of most of the properties of water as a solvent including the use as a temperature buffer. Water has a high specific heat capacity as well as large enthalpies of vaporization and fusion. These properties provide a buffer against fluctuations in temperature for numerous reactions.
where water is the solvent. It is the cause of the relatively mild climate on Earth and countless temperature-dependent reactions taking place in living organisms.

H-bonds also play a vital role in enzyme function. Most enzymes are proteins that serve as catalysts in living organisms. Without them, most metabolic pathways would proceed too slowly to support life. When a substrate enters the active site of an enzyme, the enzyme binds the substrate to form a temporary complex. The enzyme-substrate bonds need to be strong enough to hold the proper structure of the complex but weak enough to release the substrate when required. H-bonds are strong non-covalent interactions, but not as permanent as covalent or ionic bonds, which provides this functionality.47

1.2. Hydrogen Bonds in Proteins

H-bonding is also crucial for the structure and stability of many biological macromolecules, including proteins.48-52 This has been a field of extensive study for nearly ninety years.48-56 Today we know that many types of interactions contribute to protein properties, including hydrophobic interactions, H-bonds, disulfide bonds, charge-charge interactions, salt bridges, n→σ* interactions, and possibly others.48 Of these, H-bonds and hydrophobic interactions have indisputably the largest impact on structure and stability.48,50,56

Surprisingly, H-bonds, which are polar by nature, are important both in the polar exterior regions of proteins but also in the nonpolar interior. Numerous studies in recent years have concluded that polar residues buried in the protein interior occupy smaller volumes, exhibit tighter packing density, and undergo stronger van der Waal forces when
their H-bonding potential is satisfied. This results in higher stabilization energies than if the H-bonds were absent. In fact, H-bonded residues are the most conserved throughout proteins.

**Figure 1-2.** The secondary structure of an α-helix and β-pleated sheet holds its shape with hydrogen bonds. Reprinted by permission of Pearson Education, Inc., New York, New York.

Buried intramolecular H-bonds are also stronger and therefore more stabilizing than those on the surface of the protein. This is due to the electrostatic nature of H-bonds. In the nonpolar environment of the protein interior where the dielectric constant is lower, the electrostatic interactions are more forcefully felt, resulting in a stronger H-bond.

The strongest H-bond in proteins is the N-H···O=C bond between the carbonyl of one residue and the amino N-H of another (Fig. 1-2). A study published in 2016 reported that 60-76% of H-bonds in proteins are this type. The other ~25% is a mixture of N-H···O, O-H···O, C-H···O, O-H···N, C-H···N, and N-H···N bonds. Little interest has been
shown toward these lesser-known bonds until the past five years.

A full understanding of H-bonds in proteins has become paramount in properly developing molecular modeling software. While great strides have been made in the past two decades, results of many programs still disagree with experimental results.\textsuperscript{58} To overcome this, a better understanding of protein folding is required. This includes a better understanding of the “lesser-known” H-bonds.

1.3. N-H····N Bonds in Proteins

Over the past 25 years, a scattering of papers has been published on the identification of N-H····N H-bonds between amino acids within proteins. A small number of scientists\textsuperscript{59-60} recognized the N-H····N bond in the 90s, but the timing was not quite right for their work to be fully realized and appreciated.

It was not until 2014 that the importance of these bonds became more apparent when Adhikary et al. reported evidence of N-H····N H-bonds.\textsuperscript{15} After the publication, several research groups set out to better explore this phenomenon. A discussion of the studies from the past 25 years is discussed next.

Krause 1991

In 1991, Krause et al.\textsuperscript{59} determined and reported the molecular structure of the racemic dipeptide, D,L-Histidyl-L,D-histidine pentahydrate in their article in *Acta Crystallographica*. Upon acquiring crystals of the species, they performed X-ray crystallography and solved the structure using the software SHELXS86,\textsuperscript{61} the difference Fourier technique, and full-matrix least squares. The final structure is shown in Fig. 1-3.

They claimed that an intramolecular N\textsubscript{2D}-H····N\textsubscript{1D} hydrogen bond exists between
the two imidazole rings with a bond length of 2.724Å and angle 168°. This bond stabilizes the otherwise unfavorable gauche conformation of the rings to the carbonyl group and amino group of the neighboring ring. Based on the short H···N length, they predicted that the strength of the bond was on the same order of magnitude as the more well-known N-H···O or O-H···O bonds.

**Figure 1-3.** The computationally determined structure of D,L-Histidyl-L,D-histidine pentahydrate has a N\textsubscript{2D}-H···N\textsubscript{1D} H-bond.\textsuperscript{59} (Reproduced with permission of the International Union of Crystallography. See Appendix B.)

**Hennig and Geierstanger 1999**

In 1998, Dingley et al. were able to detect N-H···N H-bonds in both RNA and DNA directly using a heteronuclear HNN-COSY nuclear magnetic resonance technique.\textsuperscript{62} The following year, Hennig and Geierstanger used the same technique on the sperm whale apomyoglobin protein.\textsuperscript{63} X-ray crystallography had previously suggested an
N-H···N interaction between His24 and His119 in the protein based on the locations of the potential H-bond donor and acceptor groups (Fig 1-4). To determine if the perceived interaction was significant, they measured the $^2J_{NN}$ scalar couplings associated with the N-H···N bonds. A J-coupling is an indirect interaction between two nuclear spins which arises from hyperfine interactions between the nuclei and local electrons. The presence of such couplings between the nitrogen atoms would indicate an interaction since scalar couplings are generally only detected between chemically bonded nuclei. The coupling constants they gained were comparable to other confirmed N-H···N H-bonds. They believed the bond to significantly impact the stability and folding of the native state of the protein.

![Figure 1-4. The structure of His24 and His119 in the sperm whale apomyoglobin protein. The geometry of the imidazole groups suggests the formation of an N-H···N hydrogen bond.](image)

**Adhikary 2014**

After Krauses’ and Hennig’s publications, the N-H···N bond went uninvestigated for fifteen years as in the 1990s there was as yet no recognizable need; but that has now changed. Protein binding is significant when designing molecular modeling software and
over the past two decades, the programs have become more advanced and accurate when predicting protein structures. However, there is still a frustrating amount of error when compared to experimental data. To overcome these flaws, the software creators must have a full understanding of how proteins fold and what bonds are involved. Including the most prominent N-H···O=C H-bond alone does not appear to be enough.

Consequently, renewed interest has occurred in the lesser-known bonds within proteins, including the N-H···N H-bond.

In 2014, Adhikary et al. published “Evidence of an Unusual N-H···N Hydrogen Bond in Proteins” and brought interest back to the field. They simulated a novel approach to detecting N-H···N hydrogen bonds within proteins using IR spectroscopy. Traditionally, N-H···N hydrogen bond formation is observed by a decrease (red shift) in the stretching vibration of the N_donor-H bond. However, detecting a single N-H stretching frequency within a protein is nearly impossible due to spectral congestion. To overcome this, they studied the more accessible stretching frequency of the C-H bond adjacent to the protein backbone amino group. Their reasoning was that due to the hyperconjugation between the lone pair on the nitrogen and the σ* orbital of the adjacent C-H, the C-H stretching frequency would be affected by N-H···N hydrogen bond formation. They predicted that upon N-H···N formation, the hyperconjugation would be reduced, causing an increase (blue shift) of the C-H stretching frequency.

To test their hypothesis, they first examined the free amino acid, proline. Proline was ideal due to its C-H group adjacent to the amino group that would be capable of hyperconjugation. There are also other C-H groups that are not adjacent and could serve
as controls. To mimic the conditions of N-H···N formation, they obtained IR spectra for free proline throughout the pH range 7.3 to 13.0. Above proline’s pk_{a2} of 10.60, the amino acid is in its anion state and the amino group is deprotonated leaving the π-orbital free to hyper-conjugate with the adjacent C-H group. This mimicked the non-hydrogen bonded structure. The C-H absorption at pH 13.0 was roughly 2120 cm\(^{-1}\).

Then by decreasing the pH below 10.60, the amino acid is in the zwitterionic state and the amino group is protonated creating a similar configuration as an H-bond. The π-orbital is no longer available and hyperconjugation decreases. By pH 7.3, the C-H stretching frequency had increased to roughly 2140 cm\(^{-1}\) creating the predicted blue shift.

To confirm their results and test their hypothesis in a more accurate environment, Adhikary et al. synthesized the N-terminal Src homology 3 domain from the murine Crk-II adaptor protein. The protein contained four proline residues (Pro152, Pro165, Pro183, and Pro185) and from the crystal structure they determined Pro165 and Pro185 had the correct geometry to form N-H···N interactions while Pro152 and Pro183 didn’t (Fig. 1-5). By comparing the IR absorptions for the two groups, they once again concluded that the formation of N-H···N hydrogen bonds occurs based on the blue shift in the adjacent C-H.
Figure 1-5. Conformations of the four proline residues in nSH3 based on the crystal structure (PDB ID 1CKA), with $\phi$ and $\psi$ angles indicated. Green: carbon, blue: nitrogen, white: hydrogen, red: oxygen, hydrocarbon hydrogen atoms not shown.\textsuperscript{15}

Figure 1-6. The N-terminal Src homology 3 domain from the murine Crk-II adaptor protein with Pro152, Pro183, Pro185, and Pro165 shown. Pro152 and Pro183 are incapable of forming N-H···N bonds due to their position at the perimeter of the protein. (b) The computationally determined structures of Pro165 and Pro185 participating in N-H···bonds. Green: carbon, blue: nitrogen, white: hydrogen, red: oxygen, hydrocarbon hydrogen atoms not shown.\textsuperscript{15}
As a final test, they performed density functional theory (DFT) calculations of methyl-terminated proline dipeptide mimics whose configurations resembled the four proline residues in the previous protein. A molecular dynamic simulation of the protein was performed in a water solvent using the Amber ff99SB protein force field to view the protein in solution rather than as a crystal. The proline structures were then exported and optimized using the B3LYP/3-11+g(d,p) level of theory with the ψ and φ angles constrained to the average values obtained in the simulation (Fig. 1-6). Natural bond orbital (NBO) analysis, which converts the full electron density from the DFT calculation into a set of localized natural atomic and bonding orbitals, was performed using the NBO 5.9 package built into Gaussian09. Adhikary et al. reported the calculated stabilization energies due to the n→σ* charge transfer were 0.6 and 0.1 kcal/mol for Pro165 and Pro185, respectively. From this evidence and the previous blue shifts of the C-H stretching frequencies, they concluded that the N-H···N interactions, were, in fact, hydrogen bonds.

Preimesberger 2015

Preimesberger et al. conducted a series of studies investigating N-H···N interactions between histidine caps on the N-terminus of α- and 310- helices of proteins. The studies began with ankyrin repeating proteins, first with three repeating units, then four, then expanded to include a host of histidine N-caps in heme proteins (truncated hemoglobins and cytochrome b5). They used the same method as Hennig in 1999, that is heteronuclear HNN-COSY nuclear magnetic resonance.

Preimesberger et al. found that the helix-capping N-H···N H-bonds can be
routinely detected in $^{15}$N-labeled proteins using H-bond scalar coupling experiments. Direct assignment of H-bonding nuclei was achieved by tailoring HNN-COSY and CTSE difference experiments for protein amide $^{15}$N-$^1$H and histidine $^{15}$N nuclei. Comparing the $^{2}$H$_{NN}$ coupling constants with those for histidine pKa allows for a convenient comparison for the length and relative strength of N-H···N H-bonds. They found that compared to results by Dingley of the same nature for RNA and DNA, the N-H···N H-bonds in proteins were slightly weaker.$^{62}$

**Deepak 2016**

Deepak et al. studied the crystal structures of a large number of proteins to identify six different types of H-bonds by the distances and angles of the acceptors and donors in the bonds.$^{57}$ Their interest was specifically in N-H···N bonds and they used quantum chemical calculations to ascertain if these interactions were stabilizing. They found that proline commonly participates in N-H···N interactions and provides additional stability to loops and capping regions of many secondary structures.

To begin, Deepak et al. compiled two data sets of crystal structures from the Protein Data Bank. Data Set I comprised structures determined by neutron diffraction or ultra-high-resolution x-ray diffraction with the hydrogen atoms determined experimentally. This data set had a resolution of 0.9 Å or better and was made up of 68 polypeptide chains. Data Set II had a lower resolution (1.8 Å or better) but a greater number of polypeptide chains (5542). These had been imaged by x-ray diffraction and the hydrogens were determined using the software REDUCE.$^{73}$

Deepak et al. studied six types of H-bonds of the D-H···A type: N-H···O, O-H···O,
C-H···O, N-H···N, O-H···N, and C-H···N. To do this, they scanned each polypeptide chain for the above atoms that were within the criteria in Tbl. 1-1. For the weaker C-H···N and C-H···O bonds, only the distance criteria were used and $\theta(C-H\cdots A) \geq 120^\circ$.

**Table 1-1.** The geometry criteria set by Deepak et al. to determine the presence of hydrogen bonds in Data Set I and II. AA – the acceptor antecedent atom.\(^{57}\)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d(D\cdots A)$</td>
<td>$\leq 3.5\text{Å}$</td>
</tr>
<tr>
<td>$d(H\cdots A)$</td>
<td>$\leq 2.5\text{Å}$</td>
</tr>
<tr>
<td>$\Theta(D-H\cdots A)$</td>
<td>$\geq 90^\circ$</td>
</tr>
<tr>
<td>$\Theta(H\cdots A-AA)$</td>
<td>$\geq 90^\circ$</td>
</tr>
</tbody>
</table>

While Deepak et al. identified as many H-bonds as possible, their main interest was in the existence and properties of N-H···N interactions. They found that out of all the interactions identified, roughly 1.1% were of the N-H···N type. Surprisingly, almost 90% of these were between a donor proline and the N-H amino group of the succeeding residue ($N_{i+1}-H_{i+1}\cdots N_i$ where proline is the $i^{th}$ residue). This agrees with the previous study of proline by Adhikary et al.\(^{15}\)

To determine if the N-H···N interactions were stabilizing, Deepak et al. performed quantum chemical calculations on a model of Pro94 from the cytochrome c peroxidase protein. This residue had met the previous criteria for N-H···N interactions and was a part of Data Set I which had higher resolution crystal structures than Data Set II. The coordinates of each atom of Pro94 and Ile95 were taken directly from the crystal structure and the ends were methylated to form the model compound, N-acetyl L-proline N-methylamide. The hydrogen atoms were optimized computational using the BP86/triple-ζ def2-TZVP level of theory.\(^{74-75}\) (BP86 is a density functional theory model.) The
electronic structure program package ORCA v3.0.2 was used for the optimizations. The resulting structure is shown in Fig. 1-7.

![Figure 1-7](image)

**Figure 1-7.** (A) The chemical structure of Ace-Pro-NMe, used for quantum chemical calculations. The dihedral angles $\phi$ and $\psi$ were varied in a step-wise fashion and the energy profile determined. (B) The resulting structure with the most stable configuration.

![Figure 1-8](image)

**Figure 1-8.** The model compounds N-methylformamide and N-acetyl pyrrolidine are involved in the N-H···N H-bond. The distance $d$ between the N-H group and the N was varied from 2.1 to 4.1 Å. (B) The interaction energy profile between the two molecules as a function of the distance $d$.\(^{57}\)

Single-point energy calculations were gathered at M06-2X/aug-cc-pvqz\(^{77-78}\) with Gaussian09\(^\text{72}\) by varying $\phi$ and $\psi$ in 5° steps while keeping all other bond angles and lengths constant. Seventy-two structures were generated, and the average potential energy
profile of these is shown in Fig. 1-8. NBO analysis was also performed on each of the 72 structures.

Three minima and two maxima stood out in the structures found by Deepak et al. (Fig. 1-9 and 1-10). Structure A had the lowest energy minimum making it the most stable and, furthermore, this structure had the closest resemblance to the crystal structure. NBO analysis revealed that it was stabilized by both N-H⋯N (n→σ* ) and n→π* interactions with stabilizing energies of 0.75 kcal/mol and 0.56 kcal/mol, respectively. Structures B and C also had minimal zero-point energies but these were not due to N-H⋯N interactions. Instead, N-H⋯O and n→π* interactions stabilize these configurations. The two maxima D and E exhibit large steric hindrances that prevent a stable structure from forming. This information showed that for proline and isoleucine, the most stable configuration is that with an N-H⋯N interaction.

![Molecular plots of Ace-Pro-NMe](image)

**Figure 1-9.** Molecular plots of Ace-Pro-NMe, representing different regions of the potential energy profile and corresponding to the points A-E in Fig. 1-10.57
Deepak et al. also performed NBO analysis on other structures from Data Set I. The coordinates of the atoms were exported from 18 crystal structures then the hydrogen atoms optimized as in the previous experiment. It is worth noting that after optimization, the N-H···N geometries no longer met their bond length and angle criteria to classify them as H-bonds. However, they still performed NBO analysis on each structure. Of the 18 structures, 9 had stabilizing energies for the N-H···N (n→σ*) values greater than 0.5 kcal/mol. In four structures, even though visually the crystal structures had suggested N-H···N stabilizing interactions, the NBO data did not detect anything.

**Iyer and Deepak 2017**

Subsequently, Deepak published an article with Iyer which further analyzed the data gained from the H-bond search of Data Set II. Rather than focusing on proline,
however, they turned to histidine. After proline, histidine was the most common participant in N-H···N interactions. 285 examples were found among the 5542 polypeptide chains. Interestingly, unlike proline, which generally bonded with the residue adjacent to it, histidine residues largely bonded to residues at least eight locations away. Less than 7% of interactions occurred between neighboring residues.

Histidine residues have three nitrogen atoms available to H-bond, one on the main chain, and the other two in the imidazole ring. In this study, their interest was in the properties, such as strength and environment, of the N-H···N interactions namely between the imidazole rings of two histidine residues. For this, Deepak and Iyer used quantum computational methods.

There are two nitrogen atoms in the imidazole group of the histidine residue. This creates different N-H···N H-bonds with their own properties. In one type, both the donor and acceptor histidine residues are neutral. In the second, the donor is protonated, and the acceptor is neutral. Out of the 285 N-H···N histidine interactions found, 223 were from the neutral category.

For simplification in their calculations, Deepak and Iyer used imidazole groups to represent the histidine residues. These were optimized independently using the BP86/triple-ζ def2-TZVP level of theory\textsuperscript{74-75} and the ORCA v3.0.2 software.\textsuperscript{76} The structures were then superimposed on the side chains of the histidine crystal structures they wished to study. Single-point energies were computed at the M06-2X/aug-cc-pvqz level with Gaussian09 for the 285 structures where the donor imidazole was protonated or neutral accordingly. Interaction energies were calculated using the following equation
\[ E_i = E_{DH-AH} - E_{DH} - E_{AH} \]

\( E_{DH-AH} \) is the single point energy of the structure. \( E_{DH} \) and \( E_{AH} \) are the single point energies of the donor and acceptor imidazoles, respectively. Boys and Bernardi’s standard counterpoise correction method\(^\text{80}\) was employed to account for the basis set superposition error (BSSE).

A tally of \( E_i \) for the neutral and protonated complexes are shown in Fig. 1-11. 282 of the 285 cases had favorable \( E_i \). The majority of \( E_i \) for neutral complexes were between -6 and -8 kcal/mol. When the donor was protonated, \( E_i \) was much more favorable in the range of -20 and -25 kcal/mol. These values indicate the N-H···N interactions between histidines are stabilizing and in the protonated case, very stabilizing.

Figure 1-11. Histograms showing the distribution of BSSE-corrected interaction energies for imidazole pairs participating in N-H···N hydrogen bonds: (A) both imidazole rings are neutral and (B) donor imidazole is protonated and the acceptor imidazole is neutral.\(^\text{79}\)

They did not elaborate on why they believed the protonated donor created such a strong bond. They continued to classify it as a H-bond. It could be argued, however, that the interaction is less likely to be an H-bond and more likely to be a dipole-ion interaction. Imidazole rings are aromatic, which means when the ring becomes protonated, the donor-nitrogen becomes more positive due to the shared electron density
decreasing which would destabilize a H-bond. However, a new type of interaction would be possible, an ion-dipole interaction. In this type of bond, a charged species attracts the oppositely partially-charged dipole of another species. It is not a true ionic bond because both species are not charged, but it is still, in general, a stronger interaction than a H-bond. For the imidazole complex, the now more positively charged nitrogen would attract the negative electron density of the polar nitrogen on a different imidazole ring, creating the ion-dipole interaction. This would account for the significantly stronger interaction energies that Deepak et al. reported.

Next, the imidazole pairs were further characterized using NBO analysis. Four structures were considered, the least favorable and most favorable of both the neutral and protonated groups. For the protonated-neutral pair, the second order perturbation energies were 40.1 and 25.16 kcal/mol and for neutral-neutral, 10.9 and 0.1 kcal/mol. This indicates that the charge transfer due to the interactions was favorable for all cases.

Deepak and Iyer wanted to further understand the environment of the bonded histidines. They used Solvent Accessible Surface Area (SASA) Analysis to investigate if the N-H···N bonds show a preference to be on the surface of a protein, in contact with the solvent, or buried in the hydrophobic interior. To compare, they calculated SASAs for all the histidines in Data Set II, both those with N-H···N interactions and those that did not.

When the histidine side chain reacted with another histidine sidechain through N-H···N interactions, the average SASA of the residue was 253 Å². For comparison, a histidine without this interaction had an average SASA of 287 Å². This difference is significant and indicates histidines with N-H···N interactions are more likely to be buried.
As a final experiment, Deepak and Iyer obtained the B-factors of Data Set II provided with the x-ray crystallography data. (B-factors are also known as temperature factors or atomic displacement parameters.) These values, when compared, give an indication of the degree of mobility of each atom. Comparing the normalized average values of the histidines that N-H···N bond and those that do not found that the bonded histidines had significantly less mobility (-0.279 vs. +0.259). This agrees with the theory that atoms involved in stronger bonds would have less mobility than those that had weaker or no interactions associated with them. The difference in values is extremely significant and suggests a possible way to identify N-H···N H-bonds from the x-ray crystallography data alone.

1.4. Conclusion

The work done in the past few years has significantly improved our understanding of N-H···N interactions in proteins. However, a full understanding is yet to be realized. Adhikary et al. did groundbreaking work investigating N-H···N bonds using proline. Their use of blue-shifted neighboring C-D bonds was an ingenious way of overcoming the usual IR congestion problems with proteins. Their use of NBO analysis to further confirm the H-bonds, however, was flawed. They reported a second-order perturbation energy of an N-H···N interaction as 0.1 kcal/mol and then claimed it was a H-bond. Not only is that very low for H-bonds in general, but the makers of the NBO software strongly suggest in the result files themselves that a threshold of 0.5 kcal/mol should be used to ensure reliable data. Anything below that is heavily subject to error. The other low energy Adhikary et al. reported was 0.6 kcal/mol which is also low for a H-bond but
considering the bond angle (130°) the energy is reasonable. The true critique is not so much against their work, however, but that it is just the beginning. There are 19 other essential amino acids to explore. There is plenty more work to be done.

Deepak and Iyer used a significant amount of x-ray crystallography and neutron diffraction data from a protein data bank to identify H-bonds based on geometries and angles. Not only is there still speculation as to whether crystal structures are truly accurate depictions of native proteins (a hazard of the field, but not to be ignored), but a H-bond is a very specific type of interaction that cannot be identified based on geometry alone. Knowledge of the electronic structure of the interactions is necessary to confirm the presence of a H-bond.

Deepak et al. tried to do this computationally, but once again, there were possibilities of errors because of particular assumptions. At the beginning of the computational calculations, when forming the structures they would use, they optimized only the hydrogen atoms rather than the entire structures. Also, during many of the calculations, most bond lengths and angles were held fixed, rather than being allowed to adjust and optimize. While this saves a great deal of computational time and can simplify a calculation, it also opens the possibility of a large degree of error, most prevalently shown in the interaction energies. Their simplifications were understandably chosen, but more sophisticated calculations and experiments are needed to confirm their work.

The research reported to date in the literature has confirmed that N-H···N interactions occur within proteins. Histidine and proline have specifically been investigated and their participation as H-bond donors has been shown. There remains a
gap in our understanding of the roles of the other eighteen essential amino acids in these interactions and more substantial evidence that the interactions are electronically true H-bonds is needed to understand their importance or otherwise.

The remainder of this thesis will detail studies to start to fill this gap in knowledge. Quantum chemical calculations were performed to identify N-H···N H-bonds in protein-like systems and elucidate detailed properties of these bonds. Chapter 2 focuses on three simple peptide backbone mimics. Chapter 3 expands on this work to consider amino acid residues. The computational methods used in both chapters are discussed in detail in the Appendix.
REFERENCES


(68) Pople, J. A.; Binkley, J. S.; Seeger, R. J. Quant. Chem. 1976, 10, 1-19.

(69) Pople, J. A.; Seeger, R.; Krishnan, R. J. Quant. Chem. 1977, 12, 149-63.


(71) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. NBO Version 5.9.


CHAPTER 2
IDENTIFICATION AND CHARACTERIZATION OF N-H···N HYDROGEN BONDS BETWEEN PROTEIN BACKBONE MIMICS

2.1. Abstract

The stability and structure of proteins is due, in part, to extensive intramolecular hydrogen bonding networks. Recently, the complete characterization of the different types of these hydrogen bonds has gained interest, including the N-H···N hydrogen bond. This work is an \textit{ab initio} quantum chemical study of N-H···N hydrogen bonding between the peptide backbones of amino acid residues. Geometry optimizations were performed on dimers constructed from combinations of three protein backbone mimics, acetamide, \(N\)-methylacetamide, and \(N,N\)-dimethylacetamide; and the intermolecular interactions are described with Natural Bond Orbital analysis, Symmetry-Adapted Perturbation Theory, Nuclear Magnetic Resonance, electron density maps, and \textit{Atoms in Molecules} methodology. The experiments were performed with and without an aqueous environment. It was concluded that intramolecular N-H···N H-bonds along the backbone of proteins can be a stabilizing force both in the hydrophobic protein interior and the aqueous protein exterior.

2.2. Introduction

Being able to understand and predict the pathway of a protein folding is of paramount importance to discovering cures for protein-related diseases. Many of these diseases, such as sickle-cell anemia, are caused by a mis-folding of the protein which
then impedes its function.\textsuperscript{1} Current understanding of these processes has led to the
creation of protein modeling software that attempts to predict the folding dynamics and
final structure of proteins.\textsuperscript{2}

Our understanding of protein folding is still in its adolescence. Evidence of this is
apparent in the inaccuracy of simulated proteins when compared to their
spectroscopically-determined structures. For example, Stranges and Kuhlman tested the
molecular modeling software, Rosetta, by comparing over 150 Rosetta-predicted protein
structures with their experimentally-determined structures.\textsuperscript{3} Only five of the proteins met
the criteria they had set for successful predictions and these were small, largely
hydrophobic molecules. Interactions that included electrostatic forces, which includes
most noncovalent interactions, had large errors in the calculated energetics. This resulted
in serious flaws in the predicted folding of these proteins and very inaccurate structures.
The inaccuracies in the software are due, in part, to our not having a complete
understanding of the intramolecular bonding that occurs within the proteins. Until we
have this, we won’t be able to advance the software nor push through the current
roadblocks that keep us from pursuing cures for diseases such as sickle-cell anemia.

An important piece to predicting protein folding is a complete knowledge of the
intramolecular bonding within proteins. For many years, we have known that hydrogen
bonding plays a key role in this, especially the N-H···O bond between carbonyl and amino
groups. Other types of hydrogen bonds have been suggested in the past, but it wasn’t
until recently that detailed studies have begun to be noticed.\textsuperscript{4–9}

One such study was previously introduced in Chapter 1; Adhikary et al. identified
N-H···N H-bonds between proline and adjacent residues in a Src homology 3 domain protein in 2014. Upon finding N-H···N interactions by IR analysis (see Chap. 1 for more details), they turned to density functional (DFT) computational methods to confirm their characterizations as H-bonds. There were four proline residues in the original protein, so they made four molecules that closely mimicked these, each with proline and its adjacent residue within the protein (Fig. 2-1). To terminate the molecules without drastically changing their electronic environments, they methylated the ends.

**Figure 2-1.** Conformations of the four proline residues in nSH3 based on the crystal structure (PDB ID 1CKA), with φ and ψ angles indicated. Green: carbon, blue: nitrogen, white: hydrogen, red: oxygen, hydrocarbon hydrogen atoms not shown.

Using these mimics with Natural Bond Orbital (NBO) analysis, they were able to show a charge transfer from the lone pair of the H-bond acceptor N into the sigma antibonding orbital of N-H, indicating a small interaction and possibly an N-H···N
H-bond.

Their conclusion, as well as the mimic molecules they used in the DFT calculations, inspired the idea for this thesis. Specifically, to confirm and expand Adhikary et al.’s work on N-H···N H-bonding in proteins, in the present work, three monomers that mimic protein backbone have been dimerized computationally both in and out of an aqueous environment and the resulting intermolecular interactions investigated for N-H···N H-bonds. The three molecules used were acetamide, N-methylacetamide, and N,N-dimethylacetamide, shown in Fig. 2-2. These molecules are amides with a methyl group adjacent to the carbonyl carbon. The C-N covalent bond represents the peptide bond along the backbone of proteins. The R groups on the nitrogen vary from two hydrogens in acetamide (NHH), H and CH₃ in N-methylacetamide (NHC), and two methyl groups in N,N-dimethylacetamide (NCC). The abbreviations NHH, NHC, and NCC represent whether the R groups are CH₃ (C) or hydrogen (H) and will be used throughout this thesis. The changing R groups change the electronic environment of the nitrogen so as to mimic different protein residues.

To ascertain the types of bonding possible, combinations of the monomers, two at a time, were brought in close proximity and the natural bonding that would result, simulated using *ab initio* computational methods. The resulting geometries of the dimers were examined, and the electronic structures determined using multiple approaches to identify the types and strengths of the intermolecular interactions. Five dimers formed intermolecular N-H···N H-bonds and these were further tested by simulating an aqueous environment to determine if the complexes were stable, and therefore possible, *in vivo.*
N-H···N H-bonds in three of the five dimers proved to be further stabilized upon the addition of water. It was concluded that intramolecular N-H···N H-bonds along the backbone of proteins can be a stabilizing force. This is especially so in the hydrophobic protein interior, but not uncommon in the aqueous protein exterior as well.

The rest of this chapter includes a brief description of the computational methods used (for a more detailed explanation see the Appendix); the results and a discussion interpreting the finding of the calculations; and the conclusions drawn from those results.

2.3. Computational Methods

All calculations were carried out with the MP2/6-31+g* and MP2/aug-cc-pVDZ levels of theory unless otherwise indicated. Wherever possible, results will be presented with the larger aug-cc-pvdz basis set. Geometries and vibrational frequencies were optimized using the Gaussian09 suite of programs. Minima were confirmed by the absence of imaginary frequencies. The molecular electrostatic potentials (MEPs) were evaluated for the monomers in their optimized geometries at the MP2/aug-cc-pVDZ level. Electron density shifts caused by complexation were calculated as the difference between the electron density of the complex and the sum of those of the monomers, then diagrams created. The binding energy, $E_b$, was calculated as the difference between the total energy of the complex and the sum of the isolated optimized monomers. Interaction energy, $E_i$, was defined relative to the monomers in their geometries within the context of the complex. The energies were corrected for basis set superposition error (BSSE) via the counterpoise technique. The interaction energy was partitioned into separate contributions by Symmetry Adapted Perturbation Theory (SAPT) at HF/aug-cc-pVDZ
as implemented in the MOLPRO software.\textsuperscript{22} NBO analysis was used to evaluate the charge transfer effects using the NBO-3 program\textsuperscript{23} incorporated in the Gaussian09 software. The $^1$H isotropic shielding was calculated with the Gauge-Independent Atomic Orbital (GIAO) method\textsuperscript{24} using the optimized parameters. The electron density was analyzed via the \textit{Atoms in Molecules} (AIM) procedure to determine the positions of bond critical points and the electron density at these points calculated using the Laplacian operator, via the Multiwfn v3.4.1 software.\textsuperscript{29}

\section*{2.4. Results and Discussion}

\subsection*{Monomers}

For this study, three monomers were used due to their resemblance to the repeating unit in the backbone of proteins. The molecular electrostatic maps of these are shown in Fig. 2-2. Electron rich regions (red) are seen surrounding the oxygen atoms, as expected, along with smaller regions near the nitrogen atoms perpendicular to the plane of the molecules. This region is the largest in NHH, followed by NHC and NCC. Electron-poor regions (blue) are most concentrated on the amide hydrogens and methyl hydrogens adjacent to the carbonyl group.

Simulations of $^1$H NMR spectra calculated the isotropic shielding of the bridging proton in the N-H bond and the results are shown in Tbl. 2-1. This will be compared to the same data in the dimer complexes as an indicator of H-bonds. NCC was not included because it does not have an amino hydrogen and it can only act as an N-H···N H-bond acceptor. The vibrational frequencies and bond lengths of the N-H bonds in NHH and NHC are also included in Tbl. 2-1 and will be used to compare with the dimer complexes.
Table 2-1. Properties of the N-H bonds in NHH and NHC. The NHH (A) and NHH (B) labels are depicted in Fig. 2-2.

<table>
<thead>
<tr>
<th></th>
<th>N-H (Å)</th>
<th>v (cm⁻¹)</th>
<th>NMR (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHH (A)</td>
<td>1.01189</td>
<td>3677.33</td>
<td>27.0361</td>
</tr>
<tr>
<td>NHH (B)</td>
<td>1.00930</td>
<td>3706.63</td>
<td>27.1118</td>
</tr>
<tr>
<td>NHC</td>
<td>1.01054</td>
<td>3677.71</td>
<td>27.1359</td>
</tr>
</tbody>
</table>

Figure 2-2. The optimized structure of (I) NHH, (II) NHC, and (III) NCC; and the corresponding molecular electrostatic maps. (red) electron-rich (blue) electron-poor.

Dimers

Next, pairs of the test molecules were positioned in such a way to promote N-H···N interactions. In each pair, NHH or NHC acted as the H-bond donor and was paired with NHH, NHC, or NCC as the acceptor then complexes were optimized computationally without constraints. Six pairs were investigated and five formed structures with N-H···N interactions. These are shown in Fig. 2-3. The complexes are written with the H-bond hydrogen donor listed first, and the hydrogen acceptor second (Donor/Acceptor). The only complex that did not form an N-H···N interaction was
It is important that each structure represents a minimum on its potential energy surface, otherwise it may only be a temporary (unstable) state, represented by a maximum or saddle point (e.g., in the case of a transition state). To do this, the local vibrational frequencies were calculated for each dimer at all suspected potential minima (equilibrium points). A local expansion about an equilibrium point will yield imaginary frequencies unless it is a minimum. There were no imaginary frequencies present in the five complexes, so they all correspond to bound (stationary) states.

The rest of this section will discuss the five dimers in detail beginning with NHC/NHC; then NHH/NHC, NHC/NCC, and NHH/NCC which are grouped together as they have similar properties; and finally, NHH/NHH which has unique properties different from the others.

Table 2-2. Geometry properties of five dimers (donor/acceptor): the change in the N-H bond length (dimer-monomer) in angstroms; the change in the vibrational frequency of N-H (dimer-monomer) in cm$^{-1}$; the N···H bond length in angstroms; and the N-H···N bond angle in degrees. Data is reported with both the MP2/6-31+g* and MP2/aug-cc-pvdz levels of theory where possible.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta r_{\text{N-H}}$ (Å)</th>
<th>$\Delta \nu$ (cm$^{-1}$)</th>
<th>$R_{\text{N···H}}$ (Å)</th>
<th>$\Theta_{\text{NH···N}}$ (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHC/NHC</td>
<td>0.00419</td>
<td>-56.44</td>
<td>2.204</td>
<td>167.1</td>
</tr>
<tr>
<td>NHC/NCC</td>
<td>0.00592 0.00649</td>
<td>-99.00 -118.36</td>
<td>2.193 2.166</td>
<td>162.2 156.4</td>
</tr>
<tr>
<td>NHH/NHC</td>
<td>0.00426 0.00526</td>
<td>-51.00 -97.12</td>
<td>2.225 2.183</td>
<td>175.0 176.5</td>
</tr>
<tr>
<td>NHH/NCC</td>
<td>0.00723 0.00682</td>
<td>-128.54 -132.90</td>
<td>2.115 2.082</td>
<td>163.7 177.1</td>
</tr>
<tr>
<td>NHH/NHH</td>
<td>0.00408 0.00536</td>
<td>-146.68 -88.80</td>
<td>2.294 2.298</td>
<td>140.9 140.0</td>
</tr>
</tbody>
</table>
Table 2-3. The binding energy ($E_b$) and interaction energy ($E_i$) of the five dimers reported in kcal/mol. The change in the isotropic shielding of the bridging proton reported in parts per million (dimer-monomer).

<table>
<thead>
<tr>
<th></th>
<th>$E_b$ (kcal/mol)</th>
<th>$E_i$ (kcal/mol)</th>
<th>$\Delta$NMR (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6-31+g* pvdz</td>
<td>6-31+g* pvdz</td>
<td>6-31+g* pvdz</td>
</tr>
<tr>
<td>NHC/NHC</td>
<td>-2.71</td>
<td>-3.14</td>
<td>-1.12</td>
</tr>
<tr>
<td>NHC/NCC</td>
<td>-3.29</td>
<td>-5.56</td>
<td>-0.31</td>
</tr>
<tr>
<td>NHH/NHC</td>
<td>-2.58</td>
<td>-6.40</td>
<td>-0.72</td>
</tr>
<tr>
<td>NHH/NCC</td>
<td>-5.51</td>
<td>-4.34</td>
<td>-3.56</td>
</tr>
<tr>
<td>NHH/NHH</td>
<td>-7.32</td>
<td>-8.06</td>
<td>-1.59</td>
</tr>
</tbody>
</table>

Table 2-4. The electron density ($\rho$) and the Laplacian of the electron density ($\nabla^2 \rho$) in atomic units at the bond critical point of the N-H···N bond for each dimer. Values calculated with MP2/aug-cc-pvdz level of theory except NHC/NHC (MP2/6-31+g*).

<table>
<thead>
<tr>
<th></th>
<th>$\rho$ (a.u.)</th>
<th>$\nabla^2 \rho$ (a.u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHC/NHC</td>
<td>0.0187*</td>
<td>0.0541*</td>
</tr>
<tr>
<td>NHC/NCC</td>
<td>0.0207</td>
<td>0.0522</td>
</tr>
<tr>
<td>NHH/NHC</td>
<td>0.0187</td>
<td>0.0490</td>
</tr>
<tr>
<td>NHH/NCC</td>
<td>0.0234</td>
<td>0.0632</td>
</tr>
<tr>
<td>NHH/NHH</td>
<td>0.0247</td>
<td>0.0747</td>
</tr>
</tbody>
</table>

Table 2-5. Second-order perturbation $E(2)$ energies (kcal/mol) of the intermolecular interactions of the five dimers, calculated with NBO analysis. Results were calculated with MP2/aug-cc-pvdz except where indicated **, which is MP2/6-31+g*.

<table>
<thead>
<tr>
<th></th>
<th>$N_{lp} \rightarrow NH\sigma^*$</th>
<th>$CO\sigma \rightarrow CH\sigma^*$</th>
<th>$O_{lp} \rightarrow NH\sigma^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHC/NHC</td>
<td>7.40**</td>
<td>0.93**</td>
<td></td>
</tr>
<tr>
<td>NHC/NCC</td>
<td>5.98</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>NHH/NHC</td>
<td>6.30</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>NHH/NCC</td>
<td>7.04</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>NHH/NHH</td>
<td>4.80</td>
<td>0.81</td>
<td>12.06</td>
</tr>
</tbody>
</table>
Table 2-6. The dissection of the interaction energy, $E_i$, by SAPT analysis into electrostatic, induction, dispersion, and exchange forces.
Figure 2-3. The structures of (I) NHC/NHC, (II) NHC/NCC, (III) NHH/NCC, and (IV) NHH/NHH. Structure I optimized at the MP2/6-31+g* level of theory. Structures II-IV optimized at the MP2/aug-cc-pvdz level of theory.
NHC/NHC

A stable dimer of NHC/NHC with the N-H···N H-bond was found at the smaller level of theory, MP2/6-31+g*, but was not a minimum at the larger MP2/aug-cc-pVDZ level. However, when water was added to form a trimer, which is discussed following this section, the structure was a minimum at both levels of theory. For this reason, the NHC/NHC dimer will not be dismissed and the following results will be based on the smaller MP2/6-31+g* level. Note that the data for this dimer cannot be directly compared to the other dimers due to the difference of basis sets; nonetheless, the data still has relevance to this study.

A study of the geometry of the optimized complex suggests the presence of an N-H···N H-bond (Fig. 2-3 and Tbl. 2-2). The near linear bond angle, 167°, and the H···N distance, 2.204Å, are both near-ideal conditions for an H-bond which is the strongest at 180° where the orbital overlap is the greatest. To support this, upon complexation, the N\textsubscript{donor}-H covalent bond lengthened by 0.00419Å and its vibrational stretching frequency exhibited a red shift of 56\text{cm}^{-1}, both of which indicate a weakening of the bond as occurs with H-bonds.

Natural Bond Orbital analysis calculates second-order perturbation energies due to the transfer of electron density from one orbital to another. Included in the analysis results is the identification of the orbitals involved. The analysis of NHC/NHC reveals two intermolecular interactions with second-order perturbation energies, E(2), greater than the 0.50 kcal/mol threshold (Tbl. 2-5). The strongest is the N\textsubscript{donor}-H···N\textsubscript{acceptor} interaction wherein a portion of the N\textsubscript{acceptor} lone pair is transferred into the N-H
antibonding sigma orbital. The other interaction is much smaller and is a transfer of electrons from the C=O sigma bond to a C-H antibonding sigma orbital.

The N-H···N interaction is nearly eight times stronger than the C-H···O=C. This can be explained if we make the, somewhat justified, assumption that both interactions are H-bonds. Comparatively, nitrogen makes a much stronger H-bond donor than carbon due to its higher electronegativity. It is this electronegativity that causes the hydrogen to be electropositive. When carbon is the donor, this decreases the electrostatic attraction between H and the acceptor and results in a weaker bond.

Figure 2-4. The gain (red) and loss of (blue) electron density upon complexation of the monomers. The pattern along the N-H···N bond is indicative of a H-bond.

The changes in electron density due to complexation are seen in Fig. 2-4 and are indicative of H-bond formation. An increase in electron density (red) can be seen along both the N-H and H···N bonds, in each case closer to the more electronegative N than the H. A small amount of density is lost (blue) directly on the H-bond acceptor and a larger loss on the N donor. The electron density does not directly indicate where the electron
density originates from, so a direct correlation with the NBO data is not possible; however, basic considerations about the nature of the H-bond can explain the changes. The bridging proton becomes more positive as both the donor and acceptor nitrogen atoms electrostatically pull the electron density away from the H and toward each N. The gain in electron density along the N···H is due to the polarization of the N\textsubscript{acceptor} atom as a result of the electropositive hydrogen. It is important to point out that the entire electron density corresponding to the lone pair of N\textsubscript{acceptor} atom is not transferred to the N\textsubscript{donor}–H bond upon H-bond formation. Rather, the majority of the density remains with the N\textsubscript{acceptor} which is shown by the density along the N···H bond. Similar trends can be seen along the C-H···O=C bond, though to a smaller extent.

Because the N-H···N interaction is much stronger than that of C-H···O=C, it contributes the most to the interaction energy between the monomers. Therefore, the dissection of E\textsubscript{i} by SAPT analysis is an accurate depiction of the forces involved in the interaction (Fig. 2-6). The largest force in the N-H···N interaction is electrostatic (-5.60 kcal/mol) with a smaller dispersion force contribution (-3.78 kcal/mol) and an even smaller induction (-2.98 kcal/mol). Different hydrogen bonds exhibit a variety of ratios of these energies, but in general, the electrostatic and dispersion contributions are the two largest with a smaller induction force.\textsuperscript{30} This information supports the classification of the N-H···N interaction as an H-bond.

Simulations of \textsuperscript{1}H NMR spectra calculated the isotropic shielding of the bridging proton in the N-H···N bond. The values are reported as the actual shielding and not the difference compared to TMS; therefore, less shielding causes a decrease in the signal. For
this complex, the shielding decreased by 1.12 ppm (Tbl. 2-3), agreeing with the electron density loss in Fig. 2-4. A decrease in electron density would decrease the amount of shielding the nucleus is experiencing from the electrons.

As a final confirmation that an N-H···N interaction occurs and that it is a H-bond, the complex was analyzed using AIM theory (Tbl. 2-4). A bond critical point was detected along the bond path of H···N with an electron density of 0.0187 au. A publication by Parthasarathi et al. details the experimentation of comparing electron density calculated by AIM theory against experimental stabilization energies and by their results, this would constitute a weak to moderate H-bond.

In summary, the NHC/NHC dimer, at the MP2/6-31+g* level of theory, contains an N-H···N H-bond, as verified by numerous methods. A second, smaller possible H-bond also exists, that is, C-H···O=C, with a much smaller stabilization energy.

**NHC/NCC, NHH/NHC, and NHH/NCC**

Examination of the geometries of NHC/NCC, NHH/NHC, and NHH/NCC show that they have properties similar to the NHC/NHC dimer with an N-H···N interaction (Fig. 2-3 and Tbl. 2-2). The N-H···N bond angle and H···N bond distance for each is well within typical H-bond ranges. To support this, upon complexation, the N_{donor}=H covalent bonds lengthened between 0.00526Å and 0.00682Å and their vibrational stretching frequencies exhibited red shifts of roughly 100cm⁻¹, both of which indicate bond weakening as occurs with the formation of H-bonds.

The binding and interaction energies of the complexes vary slightly between the dimers, but all are comparable to typical H-bonds when using this level of theory in
organic compounds. A large difference in the interaction and binding energies occurred in NHH/NHC amounting to 2.5 kcal/mol (Tbl. 2-3). This difference is due to the energy used in moving the monomers into the proper internal structure to form the complex from their “starting positions.” In the context of this study, the movement of the monomers is not relevant.

NBO analysis reveals that for each complex, there are two intermolecular interactions with second-order perturbation energies, E(2), greater than the 0.50 kcal/mol threshold (Tbl. 2-5). The strongest is the N\textsubscript{donor}-H···N\textsubscript{acceptor} interaction where a portion of the N\textsubscript{acceptor} lone pair is transferred into the N-H antibonding sigma orbital. The E(2) energy of the complexes range from 5.98 – 7.04 kcal/mol, all of which are possible for H-bonds. The other interaction is much smaller and is a transfer of electrons from the C=O sigma bond to a C-H antibonding orbital. In each dimer, the two interactions appear to be proportional; the N-H···N interaction is roughly 5.5 times stronger than the C-H···O=C. This indicates that the changes in N-H···N bond strengths between the dimers are not due to competition with the second interaction; rather, the cause affects the strength of both interactions.

As was also apparent in the NHC/NHC dimer, the N-H···N interaction is much stronger than the C-H···O=C, and contributes the most to the interaction energy. This enables the dissection of E\textsubscript{i} by SAPT analysis to give an accurate depiction of the forces involved in the N-H···N interaction. (Tbl. 2-6). In all three dimers, the electrostatic and dispersion forces are the strongest with a weaker induction force. In NHH/NHC and NHH/NCC, the electrostatic and dispersion forces are nearly equal; but in NHC/NCC, the
dispersion force is greater than electrostatic (-8.55 kcal/mol and -6.80 kcal/mol).

Isotropic shielding for each of the three N-H⋯N bridging protons decreased upon complexation by greater than 1.50 ppm. Bond critical points along the N-H⋯N interactions were detected by AIM analysis with slightly greater electron density and may be classified as moderately strong H-bonds. Finally, the electron density shifts (Fig. 2-5) all bear the same pattern as the NHC/NHC dimer of electron density loss at the bridging hydrogen, and to a lesser extent at the nitrogen atoms, and electron density gain along the N-H and N⋯H bonds. This is indicative of H-bond formation.

**Figure 2-5.** The gain (red) and loss of (blue) electron density upon complexation of (I) NHC/NCC, (II) NHH/NHC, (III) NHH/NCC. The pattern along the N-H⋯N bond is indicative of a H-bond.
In conclusion, the data confirms that the N-H···N interactions in the three dimers, NHC/NCC, NHH/NHC, and NHH/NCC, are H-bonds and the strongest intermolecular interactions in these systems. The NHC/NCC dimer has a weaker N-H···N H-bond as shown by the non-linear bond angle and lower E(2) energy. The dissected interaction energy differs from the other two dimers; dispersion is the dominant force. These differences may be explained by steric hindrance. The monomers in the NHC/NCC complex are bulkier and lie closer to being parallel to each other than in the other two structures. This creates more repulsion and less-ideal bond angles for H-bonds between the two, causing a weaker N-H···N bond (and C-H···O=C). This also accounts for the higher dispersion force; greater contact creates more instantaneous dipoles for attraction. Be that as it may, the NHC/NCC complex is not less stable than the other two. The additional dispersion forces compensate for the reduced energy resulting from the weaker H-bonds.

**NHH/NHH**

The NHH/NHH geometry appears to contain the well-known N-H···O and desired N-H···N H-bonds (Tbl. 2-2). The two interactions make a bifurcated system where the N_{acceptor} in N_{donor}-H···N_{acceptor} is the donor for the N-H···O H-bond. Because of this, the N-H···N bond angle is far from linear at 140°. The N-H···O interaction has a slightly larger, more favorable, angle of 154° which may indicate that it is the stronger bond. The N-H···N H-bond length (H···N) is 2.298Å which is greater than the other dimers by at least 0.115Å, which may indicate that the N-H···N interaction in this dimer is weaker than in the others. Upon complexation, the N-H···N covalent bond lengthened by 0.00536Å
and the vibrational frequency red shifted by 89 cm\(^{-1}\), in good agreement with the other dimers.

The binding and interaction energies were greater than previously calculated (-8.06 and -8.46 kcal/mol) which is expected from the presence of the additional intermolecular interaction (Tbl. 2-3). NBO analysis confirms this with the detection of three intermolecular interactions (Tbl. 2-5). The strongest is an N-H...O interaction wherein a portion of the oxygen lone pair density is transferred into the N-H antibonding sigma orbital. The N-H...O interaction largely dominates the dimer and is nearly three times stronger than the N-H...N interaction. Once again, the C-H...O=C interaction is present but with a significantly smaller energy than the others.

Of the five dimers examined, this complex has the weakest N-H...N interaction by over 1.0 kcal/mol. It is also the only dimer with an N-H...O interaction. The N-H...O H-bond is stronger than the N-H...N H-bond. Delocalization of the electron density over the amide O=C-N in each NHH occurs due to pi orbitals extending above and below the plane at each atom. (This occurs in NHH, NHC and NCC.) The reduction of concentrated electron density on the nitrogen atom decreases its basicity, causing a weaker N-H...N bond than if compared to an N-H...N bond between two ammonia molecules without the delocalization. The carbonyl oxygen has two electron pairs, so even with it also being affected by the delocalization, it remains more basic than the nitrogen atom, resulting in a stronger N-H...O bond.

It is significant that this dimer has both N-H...O and N-H...N H-bonds. It has been long known that N-H...O bonds are prevalent throughout proteins. That both H-bonds are
present here confirms that the presence of one does not imply the absence of the other. Although some competition between them is apparent from the distortion of the angles of each bond.

Due to the larger N-H⋯O H-bond, SAPT analysis of the interaction energy cannot be attributed mostly to the N-H⋯N H-bond (Tbl. 2-6). However, since all three interactions appear to be H-bonds, we may still expect typical H-bond trends. A strong electrostatic force (-15.4 kcal/mol) is more than twice that of dispersion and induction (-7.0 and -6.9 kcal/mol, respectively). The dispersion force is a much smaller percentage than in the previous dimers, but consideration of the monomer positions may account for this. The two monomers are perpendicular to each other with very little surface areas coming in close contact. This would reduce the influence of the dispersion forces as commonly seen in the other dimers.

NMR isotropic shielding exhibited the greatest decrease in the N-H⋯N bridging proton compared to the other dimers at 2.0ppm (Tbl. 2-6). The bifurcating H-bonds at the N-H⋯N acceptor nitrogen atom causes this. Additional induction forces from the electronegative oxygen impacts the nitrogen of NHH2, which in turn pulls electron density from the N-H⋯N hydrogen more so than in previous dimers. This causes less electron-shielding of the hydrogen nucleus and a decreased isotropic shift.

The electron density at the bond critical point, as analyzed using AIM theory, was 0.0247 a.u., indicating a moderately strong hydrogen bond. The electron density shifts (Fig. 2-5) show clearly both the N-H⋯N and N-H⋯O bonds. Rather than a gain of electron density in a sphere along the N⋯H bond of N-H⋯N, the gain merges with that
along the N-H···O bond. The smaller areas of color along the N-H···N compared to the N-H···O is another confirmation that the N-H···O bond is the stronger of the two.

![Image](image.jpg)

**Figure 2-6.** The gain (red) and loss of (blue) electron density upon complexation of the NHH/NHH dimer.

In summary, the NHH/NHH complex is unique to the set of dimers studied. It has a bifurcated H-bond arrangement with a N-H···N H-bond and a stronger N-H···O H-bond. This competition of bonds prevents either from achieving linearity. Analysis reveals the N-H···O H-bond is three times as strong as in the N-H···N bond, which was expected.

**Trimers**

As a final test, a water molecule was added to each of the five dimers and computations were then re-optimized. This was done to ascertain if water aided or abetted N-H···N H-bond formation. Adding more than one water molecule was too taxing on the computational power and wasn’t feasible. An alternative approach is to model solvent interactions in an average sense by treating the solvent as a dielectric field; however, this
does not allow for a detailed study of solvent-monomer bonding. Therefore, the simplified approach of adding one water molecule to the complex was used. Again, as with the dimers, after optimization, several methods were used to characterize any N-H···N interactions. Due to the additional computational demand of the larger complexes, the calculations were limited to counterpoise correction of the binding energies and NBO analysis. Ideally these would not be used alone to identify H-bonds but when used in conjunction with the dimer results, confirmation is possible.

Many of the trimers tested resulted in the loss of the N-H···N interactions, however, five complexes retained it. These five bore many similarities to the dimers in their geometries and NBO analysis and since the N-H···N interactions in the dimers were confirmed as H-bonds through SAPT, electron density shifts, etc., we can conclude that the trimer interactions are also H-bonds.

**Table 2-7.** Geometry properties of the trimer complexes (donor/acceptor): the change in the N-H bond length (trimer-monomer) in angstroms; the change in the vibrational frequency of N-H (trimer-monomer) in cm\(^{-1}\); the N···H bond length in angstroms; and the N-H···N bond angle in degrees. The binding energy (\(E_b\)) and interaction energy (\(E_i\)) of the five dimers reported in kcal/mol. Results were calculated with MP2/aug-cc-pvdz.

<table>
<thead>
<tr>
<th></th>
<th>(E_b) (kcal/mol)</th>
<th>(E_i) (kcal/mol)</th>
<th>(\Delta r_{N-H}) (Å)</th>
<th>(\Delta \nu) (cm(^{-1}))</th>
<th>(R_{N..H}) (Å)</th>
<th>(\Theta_{N..HN}) (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHC/NHC/H(_2)O</td>
<td>-11.07</td>
<td>-12.18</td>
<td>0.00905</td>
<td>-142.44</td>
<td>2.170</td>
<td>154.7</td>
</tr>
<tr>
<td>NHC/NCC/H(_2)O/1</td>
<td>-10.67</td>
<td>-11.17</td>
<td>0.00663</td>
<td>-121.40</td>
<td>2.152</td>
<td>158.1</td>
</tr>
<tr>
<td>NHC/NCC/H(_2)O/2</td>
<td>-11.66</td>
<td>-12.15</td>
<td>0.00813</td>
<td>-159.37</td>
<td>2.048</td>
<td>173.2</td>
</tr>
<tr>
<td>NHC/NCC/H(_2)O/3</td>
<td>-11.04</td>
<td>-12.14</td>
<td>0.01163</td>
<td>-222.07</td>
<td>2.049</td>
<td>167.2</td>
</tr>
<tr>
<td>NHH/NHH/H(_2)O/1</td>
<td>-17.16</td>
<td>-18.04</td>
<td>0.00445</td>
<td>-69.80</td>
<td>2.347</td>
<td>138.5</td>
</tr>
<tr>
<td>NHH/NHH/H(_2)O/2</td>
<td>-14.27</td>
<td>-15.04</td>
<td>0.00636</td>
<td>-106.22</td>
<td>2.260</td>
<td>141.8</td>
</tr>
</tbody>
</table>
Table 2-8. Second-order perturbation E(2) energies of the intermolecular interactions of the trimers, calculated with NBO analysis. Results calculated with MP2/aug-cc-pvdz and reported in kcal/mol.

<table>
<thead>
<tr>
<th></th>
<th>$N_{lp}$ $\rightarrow$ $N_{H\sigma^*}$</th>
<th>$N_{lp}$ $\rightarrow$ $O_{H\sigma^*}$</th>
<th>$C_{O\sigma}$ $\rightarrow$ $C_{H\sigma^*}$</th>
<th>$O_{lp}$ $\rightarrow$ $C_{H\sigma^*}$</th>
<th>$O_{lp}$ $\rightarrow$ $N_{H\sigma^*}$</th>
<th>$O_{lp}$ $\rightarrow$ $O_{H\sigma^*}$</th>
<th>$O_{H\sigma}$ $\rightarrow$ $N_{H\sigma^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHC/NHC/H$_2$O</td>
<td>5.22</td>
<td>9.24</td>
<td>0.56</td>
<td>0.92</td>
<td>6.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHC/NCC/H$_2$O/1</td>
<td>6.23</td>
<td>3.97</td>
<td>1.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHC/NCC/H$_2$O/2</td>
<td>8.47</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHC/NCC/H$_2$O/3</td>
<td>8.10</td>
<td>10.85</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHH/NHH/H$_2$O/1</td>
<td>3.76</td>
<td>0.82</td>
<td>19.92</td>
<td>16.67</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHH/NHH/H$_2$O/2</td>
<td>5.64</td>
<td>0.60</td>
<td>9.50</td>
<td>11.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NHC/NHC + water

The NHC/NHC dimer previously discussed was a temporary (unstable) state when optimized with the aug-cc-pvdz basis set. However, upon the addition of the water molecule, the complex stabilized and became a bound (stationary) state. From the geometry (Fig. 2-7), there appear to be three H-bonds in close proximity to each other in the complex. In companion to the $N_{donor}$-$H$-$N_{acceptor}$ $H$-bond, there are also $O$-$H$-$N_{donor}$ and $N_{acceptor}$-$H$-$O$ bonds with the water molecule. Of the three H-bonds, the $N$-$H$-$N$ bond has the most linear angle (155.7°), suggesting that it is the strongest. However, the $O$-$H$-$N_{donor}$ $H$-bond length is slightly shorter than $N$-$H$-$N$, which would suggest the opposite is true (Tbl. 2-7). Further information is required to better gauge the bond strengths. A fourth H-bond is also possible based on the calculated geometry, that is a weaker $C$-$H$-$O$ bond. This is not physically close the $N$-$H$-$N$ interaction.
Figure 2-7. The optimized structure of NHC/NHC + water with $N_{\text{donor}}$-$H$···$N_{\text{acceptor}}$, $O$-$H$···$N_{\text{donor}}$, and $N_{\text{acceptor}}$-$H$···$O$ shown. Structure optimized with MP2/aug-cc-pvdz.

The binding and interaction energies for the trimer cannot be directly compared to the NHC/NHC energies because they were calculated with different basis sets. However, the trimer energies ($E_b = -11.07$ kcal/mol and $E_i = -12.18$ kcal/mol) are very similar to the other trimer complexes and it can be said, within reason, that the energies for the trimers are twice as great as the dimers (Tbl. 2-7). This is expected with the addition of the water molecule. An increase in intermolecular interactions causes higher binding and interaction energies.

These energies are large enough to be due to hydrogen bonds, but energy considerations alone cannot confirm this definitively. It is reasonable to conclude, however, that the interactions are noncovalent.

Natural Bond Orbital analysis detected five intermolecular interactions with $E(2)$ above a 0.50 kcal/mol threshold (Tbl. 2-8); the three involving the water molecule are the strongest. Both the $O$-$H$···$N_{\text{donor}}$ and $N_{\text{acceptor}}$-$H$···$O$ interactions (9.24 and 6.21 kcal/mol)
are stronger than the \( N_{\text{donor}} - H \cdots N_{\text{acceptor}} \) interaction (5.22 kcal/mol). These \( E(2) \) values are within typical values for H-bonds calculated via NBO using the MP2/aug-cc-pvdz level of theory. The two stronger bonds can be attributed to the participation of oxygen in \( \text{H}_2\text{O} \). Oxygen is more electronegative than nitrogen and therefore is a stronger H-bond donor. The delocalization of electron density over the amide groups in each NHC weakens the nitrogen atoms’ role in the H-bonds. The other two interactions are significantly weaker.

Again, the energies from NBO analysis cannot be directly compared to the previous NHC/NHC dimer. However, the energy of the N-H\cdots N interaction in this trimer is in the same range as the other four dimers (4.80 - 7.04 kcal/mol). The other dimers have similar environments to this trimer and since the dimers have confirmed N-H\cdots N H-bonds, it is reasonable to conclude that the same interaction in the NHC/NHC trimer is also an H-bond.

**NHC/NCC + water**

From the NHC/NCC dimer, three different trimers formed with the N-H\cdots N interaction. The differences in these three are attributable to the location of the water molecule as shown in Fig. 2-8. The binding and interaction energies for the three trimers are similar, roughly -11 kcal/mol, which is greater than that of the dimer at -5 kcal/mol (Tbl. 2-7). This is to be expected because the addition of the water molecule increases the intermolecular interactions. These energies suggest that noncovalent, rather than covalent, interactions are involved.

In every case, the N-H\cdots N H-bond is stronger in the trimers than the dimer; the H-bond length is shorter, and the N-H\cdots N angle is closer to linear in each trimer as
compared to the dimer. Also, the lengthening of the covalent bond and the concomitant red shift in vibrational frequency are each greater in the trimers (Tbl. 2-7).

Figure 2-8. The optimized structures the NHC/NCC + water complexes labeled I, II, and III.

The NBO data provides further confirmation that the N-H···N interaction is a H-bond (Tbl. 2-8). Each trimer has different accompanying interactions apart from the N-H···N bond as would be expected with the water molecule varying in position. Only in the first trimer is the N-H···N interaction the strongest interaction. But in every case, the E(2) energy corresponding to the H-bond is stronger than was the case for the dimer.

In the case of the NHC/NCC complex, the addition of a water molecule, in three different locations, further stabilized the N-H···N H-bond as is shown not only in the geometries, but also in the bond energies of the complexes.
Figure 2-9. The optimized structures the NHH/NHH + water complexes labeled I and II.

NHH/NHH + water

As was the case with the NHH/NHH dimer, the NHH/NHH trimers are unique. Two trimers retained the N-H···N interaction and are shown in Fig. 2-9. In the first trimer, the N\textsubscript{donor} atom in the N\textsubscript{donor}-H···N\textsubscript{acceptor} bond is involved in three separate noncovalent interactions. These result in a less ideal geometry for the N-H···N bond which is weaker than that in the dimer (see Tbl. 2-7). In the second trimer, on the other hand, the N-H···N bond is further stabilized by the presence of the water molecule. In this case, the N-H···O interaction weakened due to the second interaction at the amide oxygen atom with the water molecule. This brought about a more ideal N-H···N angle and the interaction was strengthened.

2.5. Conclusion

To determine if N-H···N H-bonds occur between portions of protein backbones, the intermolecular bonding between pairs of three amide test molecules were simulated using \textit{ab initio} computational methods. The resulting geometries of the dimers were examined, and the electronic structures determined using multiple approaches in order to
identify the types and strengths of the intermolecular interactions. Five dimers formed intermolecular N-H···N H-bonds and these were further tested by simulating an aqueous environment to determine if the complexes were stable or not, and therefore potentially physically possible, in vivo. N-H···N H-bonds in three of the five dimers proved to be further stabilized upon the addition of water. It is concluded that intramolecular N-H···N H-bonds along the backbone of proteins can be a stabilizing force both in the hydrophobic protein interior and the aqueous protein exterior.

The next step in this work is to consider larger test molecules that not only mimic the backbone but also the side chain of the different amino acids. These calculations have shown that the smaller molecules can be calculated with reasonable computational cost and the confirmation of the existence of the N-H···N H-bonds provides the rationale to investigate larger systems.

Only a few amino acids have, so far, been studied for their ability to form N-H···N H-bonds. Therefore, a more comprehensive and systematic analysis will be valuable to understanding the role of these interactions in protein dynamics. In addition, understanding these interactions will help in the developments of more sophisticated protein modeling software.
REFERENCES


(3) Stranges, P. B.; Kuhlman, B. Protein Sci. 2013. 1, 74-82.


CHAPTER 3
IDENTIFICATION AND CHARACTERIZATION OF SIDECHAIN-BACKBONE
AND BACKBONE-BACKBONE N-H---N HYDROGEN BONDS
BETWEEN AMINO ACID MIMICS

3.1. Abstract

This work is a study of N-H···N hydrogen bonds (H-bonds) in proteins and the role they play in the stability of the structures. Combinations of amino acid mimics and N-methylacetamide were made into pairs and the bonding simulated using ab initio computational methods. This was done to ascertain the types of bonding possible, with a specific interest in N-H···N H-bonds. The resulting geometries of the dimers were examined, and the electronic structures were determined using multiple approaches to identify intermolecular bonding patterns. Six amino acids of the ten studied were found to form N-H···N H-bonds when complexed with N-methylacetamide, three nonpolar and three polar amino acids.

3.2. Introduction

As noted in Chap. 2, being able to understand and predict the pathway of a protein folding is of paramount importance to discovering cures for protein related diseases. Chapter 2 detailed the results of a series of computational experiments in which three peptide backbone mimics were studied for their ability to form N-H···N hydrogen bonds (H-bonds). This study expands on that to include the possibility of backbone-backbone and backbone-sidechain N-H···N H-bonds in methyl-terminated amino acids. In Chap. 2
it was concluded that N-H···N interactions could exist in protein environments and it justified expanding that study to consider complete amino acid residues to incorporate not just the backbone, but the sidechain of the residues as well. The previous project was performed first to determine if it was promising to do the more computationally intensive experiments.

In this chapter, both polar and nonpolar amino acid residues are considered. Of the ten residues studied, six formed minima with N-H···N H-bonds when paired with N-methylacetamide (NHC) and two of these formed multiple structures. The bonding involved both the peptide backbone and the side chains.

Each amino acid was combined with NHC and the bonding simulated using ab initio computational methods. This was done to ascertain the types of bonding possible, with a specific interest in N-H···N H-bonds. The resulting geometries of the dimers were examined, and the electronic structures were determined using multiple approaches to identify intermolecular bonding patterns. Pairing the amino acids with NHC rather than another amino acid was done due to the computational demand of ab initio calculations on increasingly larger systems. NHC mimics the amino acid backbone without the sidechain which decreases the number of atoms in the calculations. When complexes were tested with two amino acids paired, the time-requirement was unfeasible.

The rest of this chapter includes a brief description of the computational methods used (for a more detailed explanation see the Appendix); the results and a discussion interpreting the finding of the calculations, including the amino acids that did and did not form N-H···N hydrogen bonds with NHC; and the conclusion drawn from these results.
3.3. Computational Methods

All calculations were carried using Second-Order Møller-Plesset perturbation theory\(^2\) (MP2) to include the effects of electron correlation with Dunning’s\(^3-4\) augmented correlation consistent polarized valence double zeta basis set (aug-cc-pVDZ). This level of theory is widely used in the literature and provides accurate data for systems of this sort.\(^5-7\) Geometries were optimized, and non-imaginary vibrational frequencies confirmed using the Gaussian09 suite of programs.\(^8\) The binding energy, \(E_b\), was calculated as the difference between the total energy of the complex and the sum of the isolated, optimized monomers. The interaction energy, \(E_i\), was defined relative to the monomers in their geometries within the context of the complex. Basis set superposition error (BSSE) was corrected via the counterpoise technique.\(^9\) Natural Bond Orbital (NBO) analysis was used to evaluate the charge transfer effects using the NBO-3 program\(^10\) incorporated in the Gaussian09 software. The \(^1\)H isotropic shielding was calculated with the Gauge-Independent Atomic Orbital (GIAO) method\(^11-15\) using the optimized parameters.

3.4. Results and Discussion

Amino Acids Mimics

The basic structure used to build the amino acid residues for this study is shown in Fig. 3-1. To mimic the environment of N-H···N bonds as accurately as possible, this structure has the amide nitrogen in the center with the N-C\(_\alpha\) peptide bond adjacent. The amino acid-specific side chains are attached to the C\(_\alpha\); bonded to both this and the carbonyl carbon are methyl groups to terminate the structure with a similar electronic environment that a continuing protein chain would experience. The structures of the side
chains are identical to the actual amino acids.

![Amino acid structure](image)

**Figure 3-1.** The general structure of the amino acid mimics used in this study. The R group represents the side chain of each amino acid.

Six amino acids successfully formed an N-H···N interaction, and these will be the main focus of this chapter; three of these are nonpolar (valine, cysteine, and leucine) and three are polar (serine, asparagine, and glutamine). The structures of these are shown in Figs. 3-2 and 3-3. Each residue has an amino group on the main chain that may participate in N-H···N bonds. Two of the polar residues, asparagine and glutamine, also have amino groups in the side chains; their ability to form N-H···N bonds was determined as well.

*Ab initio* calculations performed on molecules of this size are highly computationally demanding, especially when a second structure is added to form N-H···N bonds. To minimize this effort, rather than pairing two of the amino acids in Figs. 3-2 and 3-3 together, a single amino acid was paired with N-methylacetamide (NHC). This smaller molecule (see Fig. 3-4) was used in Chap. 3 to mimic a segment of protein backbone and is nearly identical to the amino acids discussed above minus the side chain. Due to its smaller size, this lessens the computational demands and can shorten a single simulation by up to a week.
Properties of the amino N-H bonds are listed in Tbl. 3-1 and will be used for comparison with the dimers. In asparagine and glutamine, where three N-H bonds exist, they are labeled A-C and correspond to the structures in Fig. 3-3. The isotropic shielding of the amino hydrogens was calculated and these shieldings are also included in the table. The change in these properties upon dimerization can all be used as indicators of the presence of H-bonds.
Figure 3-2. The optimized structures of (I) valine, (II) cysteine, and (III) leucine mimics. Optimizations were calculated with MP2/aug-cc-pvdz.
Figure 3-3. The optimized structures of (I) serine, (II) asparagine, and (III) glutamine mimics. Labels A-C indicate hydrogen atoms in (II) and (III). Optimizations were calculated with MP2/aug-cc-pvdz.
Figure 3-4. The optimized structure of N-methylacetamide (NHC), the molecule paired with amino acids residues to mimic protein intramolecular bonding.

Table 3-1. Properties of the N-H bonds in the amino acid mimics and NHC: the N-H bond length in angstroms; the N-H vibrational frequency in cm\(^{-1}\); and the isotropic shielding of the hydrogen of N-H in parts per millions. The A-C labels correlate with those in Fig. 3-3.

<table>
<thead>
<tr>
<th></th>
<th>N-H (Å)</th>
<th>(\nu) cm(^{-1})</th>
<th>NMR (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valine</td>
<td>1.01322</td>
<td>3637.54</td>
<td>27.52</td>
</tr>
<tr>
<td>Cysteine</td>
<td>1.01573</td>
<td>3604.37</td>
<td>25.90</td>
</tr>
<tr>
<td>Leucine</td>
<td>1.01426</td>
<td>3629.53</td>
<td>26.49</td>
</tr>
<tr>
<td>Serine</td>
<td>1.01301</td>
<td>3643.18</td>
<td>27.41</td>
</tr>
<tr>
<td>Asparagine A</td>
<td>1.01300</td>
<td>3642.45</td>
<td>27.33</td>
</tr>
<tr>
<td>Asparagine B</td>
<td>1.01345</td>
<td>3617.22</td>
<td>27.33</td>
</tr>
<tr>
<td>Asparagine C</td>
<td>1.01146</td>
<td>3617.22</td>
<td>27.12</td>
</tr>
<tr>
<td>Glutamine A</td>
<td>1.01359</td>
<td>3635.13</td>
<td>27.14</td>
</tr>
<tr>
<td>Glutamine B</td>
<td>1.00979</td>
<td>3703.15</td>
<td>27.05</td>
</tr>
<tr>
<td>Glutamine C</td>
<td>1.01248</td>
<td>2675.10</td>
<td>26.97</td>
</tr>
<tr>
<td>N-methylacetamide</td>
<td>1.01054</td>
<td>3677.71</td>
<td>27.14</td>
</tr>
</tbody>
</table>

Dimers

Next, each amino acid was paired with NHC and positioned to promote N-H···N interactions. Ten to twenty arrangements were made for each pair with variations where the amino acid was both the H-bond donor and acceptor. For glutamine and asparagine, both amino groups were used. Each complex was optimized computationally to find the structure with the lowest energy on the local potential energy map. Nine of the optimized
structures exhibited N-H···N interactions.

The rest of this section will discuss the nine dimers in detail beginning with a general discussion, then a focus on the nonpolar amino acids, followed by the polar structures. Four amino acids did not form N-H···N bonds when tested without restrictions about the N-H···N. These will be briefly discussed at the end of the section.

As before, the local vibrational frequencies were calculated for each dimer at all suspected potential minima (equilibrium points). There were no imaginary frequencies present in the nine complexes, so they all correspond to bound (stationary) states.

Several properties can be used to identify an H-bond. Due to the attraction between the bridging hydrogen and H-bond acceptor, the covalent bond weakens which can be identified by a lengthening of the bond and a red shift in the vibrational frequency when compared with the monomer. Another indicator is a decrease in the electron shielding of the hydrogen nucleus as electron density is pulled towards the H-bond acceptor. This, again, is apparent when compared to the original H-bond donor monomer. The changes of these properties in the complexes are listed in Tables 3-2 and 3-3.

The strength of an H-bond can be interpreted by the degree of the changes in the above parameters and in the length and angle of the H-bond (shorter and more linear is stronger); the binding and interaction energies; and second-order perturbation energies calculated by NBO analysis, see Tbls. 3-2, 3-3, and 3-4.
Table 3-2. The binding energy ($E_b$) and interaction energy ($E_i$) of the five dimers in kcal/mol. The change in the isotropic shielding of the bridging proton reported in parts per million (dimer-monomer). Dimer labeling is by the amino acid mimic. Each was paired with NHC.

<table>
<thead>
<tr>
<th></th>
<th>$E_b$ (kcal/mol)</th>
<th>$E_i$ (kcal/mol)</th>
<th>NMR (Δppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valine</td>
<td>-10.56</td>
<td>-11.13</td>
<td>-0.9529</td>
</tr>
<tr>
<td>Cysteine A</td>
<td>-4.76</td>
<td>-5.32</td>
<td>-1.7278</td>
</tr>
<tr>
<td>Cysteine B</td>
<td>-5.64</td>
<td>-6.60</td>
<td>-1.1463</td>
</tr>
<tr>
<td>Leucine</td>
<td>-4.78</td>
<td>-5.20</td>
<td>-1.7278</td>
</tr>
<tr>
<td>Serine</td>
<td>-4.20</td>
<td>-5.11</td>
<td>-1.0285</td>
</tr>
<tr>
<td>Asparagine A</td>
<td>-7.93</td>
<td>-8.32</td>
<td>-1.3409</td>
</tr>
<tr>
<td>Asparagine B</td>
<td>-5.59</td>
<td>-6.36</td>
<td>-2.1846</td>
</tr>
<tr>
<td>Asparagine C</td>
<td>-6.28</td>
<td>-9.77</td>
<td>-0.5834</td>
</tr>
<tr>
<td>Glutamine</td>
<td>-4.69</td>
<td>-5.44</td>
<td>-1.00229</td>
</tr>
</tbody>
</table>

Table 3-3. Geometry properties of nine dimers: the change in the N-H bond length (dimer-monomer) in angstroms; the change in the vibrational frequency of N-H (dimer-monomer) in cm$^{-1}$; the N···H bond length in angstroms; and the N-H···N bond angle in degrees. Data is reported with MP2/aug-cc-pvdz.

<table>
<thead>
<tr>
<th></th>
<th>$R_{N\cdots H}$ (Å)</th>
<th>$\Theta_{N\cdots HN}$ (deg)</th>
<th>$\Delta r_{N-H}$ (Å)</th>
<th>$\Delta \nu$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valine</td>
<td>2.19265</td>
<td>168.094</td>
<td>0.00446</td>
<td>-82.13</td>
</tr>
<tr>
<td>Cysteine A</td>
<td>2.09982</td>
<td>171.576</td>
<td>0.00620</td>
<td>-127.50</td>
</tr>
<tr>
<td>Cysteine B</td>
<td>2.15238</td>
<td>175.452</td>
<td>0.00325</td>
<td>-74.90</td>
</tr>
<tr>
<td>Leucine</td>
<td>2.30939</td>
<td>153.747</td>
<td>0.00194</td>
<td>-30.82</td>
</tr>
<tr>
<td>Serine</td>
<td>2.25482</td>
<td>157.518</td>
<td>0.00406</td>
<td>-71.08</td>
</tr>
<tr>
<td>Asparagine A</td>
<td>2.27801</td>
<td>140.684</td>
<td>0.00671</td>
<td>-80.03</td>
</tr>
<tr>
<td>Asparagine B</td>
<td>2.37594</td>
<td>140.013</td>
<td>0.00343</td>
<td>-90.93</td>
</tr>
<tr>
<td>Asparagine C</td>
<td>2.42298</td>
<td>153.848</td>
<td>0.00366</td>
<td>-62.44</td>
</tr>
<tr>
<td>Glutamine</td>
<td>2.30755</td>
<td>166.646</td>
<td>0.00438</td>
<td>-75.21</td>
</tr>
</tbody>
</table>
**Table 3-4.** Second-order perturbation $E(2)$ energies of the intermolecular interactions between each amino acid and NHC calculated with NBO analysis. Results calculated with MP2/aug-cc-pvdz and reported in kcal/mol.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>$N_{lp} \rightarrow NH\sigma^*$</th>
<th>$O_{lp} \rightarrow NH\sigma^*$</th>
<th>$S_{lp} \rightarrow NH\sigma^*$</th>
<th>$O_{lp} \rightarrow CH\sigma^*$</th>
<th>CO$\sigma$ $\rightarrow$ CH$\sigma^*$</th>
<th>NH$\sigma$ $\rightarrow$ NH$\sigma^*$</th>
<th>$N_{lp} \rightarrow$ CO$\sigma^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valine</td>
<td>6.17</td>
<td>0.7</td>
<td>1.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cysteine A</td>
<td>9.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cysteine B</td>
<td>7.72</td>
<td>12.87</td>
<td>0.94</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>3.72</td>
<td>0.64</td>
<td>0.71</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>4.75</td>
<td>0.88</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine A</td>
<td>4.59</td>
<td>13.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine B</td>
<td>4.03</td>
<td></td>
<td>1.03</td>
<td>1.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine C</td>
<td>2.96</td>
<td>4.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamine</td>
<td>4.24</td>
<td>0.75</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3-5.** The optimized structures of the dimers labeled: (I) valine (II) leucine (III) cysteine A (III) cysteine B.

**Nonpolar amino acids**

Three nonpolar amino acids formed N-H···N bonds when complexed with NHC; these are valine, leucine, and cysteine. The electronic environments of valine and leucine...
are like those of the test molecules used in Chapter 2 with the addition of short, bulky hydrocarbon sidechains. The side chains are not very reactive when compared to the amide group, but they do cause steric hindrance which may restrict intermolecular bonding of the amide nitrogen. Both formed a single complex with NHC that had the N-H⋯N interaction and in both cases, the amino acid was the H-bond donor.

In both complexes, the N-H bond lengthened and the vibrational frequency decreased upon complexation which indicates that the N-H⋯N bond is a hydrogen bond. The NMR and NBO results further confirm this with the decrease in isotropic shielding of the bridging proton (NMR) and the transfer of electron density from the NHC nitrogen lone pair to the σ* orbital of N-H on the amino acid (NBO).

The strengths of the N-H⋯N bonds do differ in valine and leucine. The N-H⋯N bond in valine is stronger with a second-order perturbation energy of 6.17 kcal/mol while leucine is lower at 3.12 kcal/mol. This is supported by the more linear bond angle for the valine complex and the greater weakening of the N-H covalent bond.

Both complexes have additional interactions to the N-H⋯N bond which include C-H⋯O and C-H⋯O=O. Leucine also has one additional bond between the lone pair of nitrogen to the C-O σ* orbital which, with the N-H⋯N bond, creates a bifurcated H-bond system at the amine on leucine. The N⋯C=O bond is much weaker than the N-H⋯N bond but would still cause a small shift of the electron density away from the nitrogen and as a result, the nitrogen becomes a poorer H-bond donor, hence weakening the N-H⋯N bond. This also explains the disproportionately large electron deshielding of the N-H⋯N hydrogen when compared with the strength of the bond itself. At the formation of the N⋯C=O bond, the nitrogen loses a portion of its electron density, and so, by
induction, draws more from the hydrogen. This causes greater deshielding but it is not due to the N-H···N bond itself, giving a false indication of it being a strong bond.

The binding and interaction energies for the valine/NHC complex are double those of any other complex studied, but with no extraordinary properties to explain this. NBO analysis shows the N-H···N bond is the strongest intermolecular bond in the complex, with two weaker interaction as well. But the strength of the N-H···N bond (E(2) = 6.17 kcal/mol), while a strong H-bond, is not greater than in the some of the other complexes. So why the high binding and interaction energies?

One explanation is the position of the two molecules relative to each other. As shown in Fig. 3-5, valine and NHC are nearly parallel and held in place by three interactions perpendicular to the molecules. (The COσ → COσ* interaction is not shown but occurs at the same location as the COσ → CHσ* bond.) This creates a large area for London dispersion forces to occur which can contribute significantly to E_b and E_i.

Next, cysteine formed two different complexes with NHC which formed the N-H···N bond. These are labeled cysteine A and cysteine B. In complex A, NHC is the H-bond donor, and in complex B, cysteine is the donor. As with the valine and leucine complexes, all the parameters shown in Tbls. 3-2, 3-3, and 3-4 indicate that the N-H···N interactions are H-bonds. In fact, cysteine A has the strongest N-H···N H-bond out of all the complexes studied which may be due to a lack of competition by other bonds, as NBO analysis shows, it is the only intermolecular interaction between the two molecules. This means that the binding and interactions energies of roughly 5.0 kcal/mol are a good estimate of the energy of the bond. The H-bond length of cysteine A is also the shortest of all the complexes studied and has an almost linear bond angle of 172°. The red shift in
the vibrational frequency was the greatest at 127 cm\(^{-1}\) and its deshielding is also quite large. This complex represents a strong N-H\(\cdot\cdot\cdot\)N bond that holds the complex together and would have a large impact within a protein environment between two amino acids.

Cysteine B is unique in that the sulfur atom in the side chain plays an important role. The N-H\(\cdot\cdot\cdot\)N H-bond acceptor is the donor in the N-H\(\cdot\cdot\cdot\)S bond, creating a bifurcated H-bond system. The N-H\(\cdot\cdot\cdot\)S bond is stronger than the N-H\(\cdot\cdot\cdot\)N, as shown by the greater E(2) energy via NBO. However, this strongly contradicts the geometries of the two bonds. The N-H\(\cdot\cdot\cdot\)N bond has a bond angle that is more linear by 20 degrees and a shorter H-bond length by 0.035 Å. The greater strength of N-H\(\cdot\cdot\cdot\)S may be explained with the basicity of each electron donor. The delocalization of the amide group of NHC causes the amide to be more stable than it would be with localized electron density. This stability reduces the basicity of the nitrogen atom, making it a poorer hydrogen acceptor in the N-H\(\cdot\cdot\cdot\)N bond than the sulfur atom in N-H\(\cdot\cdot\cdot\)S, which would otherwise be the poorer acceptor. Also, sulfur has two electron pairs which provides more density to donate into the bond.

In addition to these intermolecular interactions, cysteine B also has a C-H\(\cdot\cdot\cdot\)C=O bond, shown in Fig. 3-5, and an N-H\(\cdot\cdot\cdot\)H-N bond. In this interaction, the N-H sigma bond of cysteine donates electron density into the N-H σ* orbital of NHC. This is a new finding and may be due to the shifted electron density into the NHC N-H bond due to the N-H\(\cdot\cdot\cdot\)S bond. Both the C-H\(\cdot\cdot\cdot\)O=C and N-H\(\cdot\cdot\cdot\)H-N interactions are very small compared to the N-H\(\cdot\cdot\cdot\)N and N-H\(\cdot\cdot\cdot\)S bonds.

In summary, three nonpolar amino acids were computationally tested for their ability to form N-H\(\cdot\cdot\cdot\)N H-bonds when complexed with NHC. Leucine and valine each
formed a complex with the N-H⋯N bond being stronger in valine. This is likely due to the bifurcated H-bond system at the amino group in leucine which divides the nitrogen electron density between the two bonds and lessens its electronegativity. Cysteine formed two different structures with the N-H⋯N bond; In cysteine A, cysteine is the H-bond donor while in cysteine B, it is the acceptor. Both complexes have stronger N-H⋯N bonds than either valine or leucine, and cysteine A had the strongest bond of all the complexes studied. Like leucine, cysteine B also had a bifurcated H-bond about the N-H⋯N acceptor which may have weakened the bond somewhat.

**Polar amino acids**

Three polar amino acids formed N-H⋯N bonds when complexed with NHC; these are serine, asparagine, and glutamine (Fig. 3-6). Both serine and glutamine formed one complex with NHC that has an N-H⋯N bond with the amino acid as the H-bond donor. Asparagine formed three complexes, labeled A-C. In asparagine A, the side chain amine is the H-bond donor, in asparagine B, NHC is the donor and bonds to the side chain, and in asparagine C, the main chain amide on asparagine is the H-bond donor. The properties of these complexes are listed in Tbls. 3-2, 3-3, and 3-4.

The serine and glutamine complexes have similar properties though the locations of the N-H⋯N bonds are different. Serine only has one amino group, so the N-H⋯N bond takes place along the main chain. Glutamine has two nitrogen atom, one in the main chain and the other in the side chain. Only one formed N-H⋯N bonds, however, and it is on the side chain. Both the serine and glutamine complexes have the expected lengthening of the N-H covalent bond and the red shift in the vibrational frequency, almost of the same amounts, which are indicative of H-bonds. Each has three
Figure 3-6. The structures of the polar amino acid – NHC complexes labeled by the amino acid: (I) Glutamine, (II) Asparagine A, (III) Asparagine B, (IV) Asparagine C, (V) Serine.
intermolecular interactions, the largest being the N-H⋯N bond followed by C-H⋯O and C-H⋯O=C which are significantly weaker.

The asparagine/NHC complexes have been labeled A-C and can be seen in Fig. 3-6. Going one by one, asparagine A has the strongest N-H⋯N bond as shown by its larger E(2) energy. However, it has a non-linear bond angle of 141°. This is explained by the presence of a second H-bond, N-H⋯O, forming along the N-H⋯N acceptor. The E(2) energy of N-H⋯O is over twice that of N-H⋯N and also has a more linear bond angle of 156° and a significantly shorter H-bond length. As was seen in leucine and cysteine B, this bifurcated system creates a competition over the nitrogen and results in a weaker N-H⋯N bond.

In asparagine B, asparagine is the N-H⋯N H-bond acceptor, rather than the donor. The N-H⋯N bond is the strongest intermolecular interaction but is still slightly weaker than the bond in asparagine A. Once again, the acceptor N is part of a bifurcated H-bond system, this time with a N⋯C=O bond which is half the strength of N-H⋯N. It still weakens the N-H⋯N bond, however, as can be seen by the non-ideal bond angle of 140°. Without this second bond, there would have been less restriction on the movement of NHC and the N-H⋯N angle would have been free to approach closer to 180°.

Finally, asparagine C has the lowest N-H⋯N E(2) energy of any of the complexes studied, indicating it is the weakest bond. This is reflected in many of the other properties including the N⋯H bond length of 2.42Å, the longest of any of the complexes, and only a small red shift of 62.44cm⁻¹ of the vibrational frequency. Also, the deshielding of the bridging proton is small compared to the other complexes by more than a factor of two. This may be attributed to strain or steric hindrance caused by two
other H-bonds occurring along the side chain of asparagine. Two bifurcated H-bonds occur at the NHC oxygen spanning out to two different atoms along the asparagine side chain. These bonds, together, are stronger than the N-H···N bond and as such have more control over the bending of asparagine. The result is that the N-H···N bond cannot form a linear interaction without putting strain on the complex.

In summary, three polar amino acids were computationally tested for their ability to form N-H···N H-bonds when complexed with NHC. Serine and glutamine each formed one complex, while asparagine formed three. In each of the five complexes, the properties verify that these are H-bonds. In some instances, the N-H···N bonds were the strongest intermolecular interactions. In asparagine A and C, however, other interactions were stronger. When a bifurcated H-bond system forms at the N-H···N acceptor, a slight weakening of the N-H···N bond occurs. Both the main chain and side chain (where applicable) can participate in the interaction. The most reasonable conclusion is that these three polar amino acids can form N-H···N H-bonds when complexed with a protein backbone.

Amino acids that did not form N-H···N bonds

Four amino acids that were studied did not form N-H···N bonds when complexed with NHC; these were glycine, alanine, proline, and isoleucine. In each experiment, an N-H···O bond formed instead. Nevertheless, the lack of an N-H···N bond can provide useful information. Leucine formed an N-H···N bond, as discussed above, but isoleucine did not. The electronic environments do not differ significantly between the two amino acids, so it must be concluded that it is most likely a physical difference which prevents the N-H···N bond from forming. The sidechain of isoleucine is bulky near the main chain
and may inhibit an approaching molecule from binding with the amino group.

Proline was reported by Adhikary et al.\textsuperscript{16} to form N-H\cdots N bonds but when tested in the present work, would not do so without constraints. In the specific protein Adhikary et al. studied, two proline residues formed N-H\cdots N bonds with the adjacent amino acids, tryptophan or another proline. In both these cases, the local structure was held in place by a strong N-H\cdots O bond and the N-H\cdots N bond was able to form due to the resulting geometry (see Fig. 1-6). When they performed computational simulations of the two complexes, the $E(2)$ energies of the N-H\cdots N bonds were extremely low (0.1 and 0.6 kcal/mol) compared to the complexes in the present study. These calculations were performed at a level of theory proven to be accurate for hydrogen bonding (B3LYP/6-311+G(d,p)) but the energy values are so low that it is questionable if a claim can be made that N-H\cdots N H-bonds really formed. The creators of the NBO program state that energies below 0.5 kcal/mol may not be reliable. This, combined with the failure to form N-H\cdots N bonds in the present work, leads to the conclusion that while proline might form quite weak N-H\cdots N interactions in the correct environment, the interaction has little control over the local structure and stability of the protein.

Lastly, glycine and alanine did not form N-H\cdots N bonds when combined with NHC. These complexes are very similar to the NHC/NHC dimer studied in Chap. 2 which also did not form the N-H\cdots N unless water was added. With this, it is reasonable to conclude that in aqueous solution, glycine and alanine will likely form N-H\cdots N bonds. But in the hydrophobic interior of the protein, they will not.

\textit{Future Work}

In addition to the amino acid complexes discussed in this thesis, the preliminary
calculations of several more have been performed and suggest the presence of N-H···N bonds. These include tryptophan, methionine, threonine, and tyrosine. The characterization of these complexations as well as investigating the six remaining amino acids, will complete the study of whether N-H···N hydrogen bonds can occur in proteins.

3.5. Conclusion

Six amino acids were found to form N-H···N hydrogen bonds when complexed with NHC; three nonpolar and three polar amino acids. These occurred along both the main chains, and the side chains when amino groups were present. The bond strengths varied among nonpolar and polar residues but, on average, were stronger for the nonpolar residues. In some cases, there were other intermolecular bonds that were stronger. The strengths of those bonds do not directly correlate with the strengths of the N-H···N bond. However, when bifurcated H-bond systems formed at the N-H···N acceptor nitrogen, it weakened the N-H···N bond.

Four amino acids did not exhibit N-H···N bonds when allowed to optimize freely, though studies have shown that interactions can occur if the N-H···N parameters are held constant, for example if a protein folded into a position where the N-H···N is held in place by other interactions, then the bond may form. Further studies can be done holding the N-H···N interaction in place, but must be done with caution so as not to call into question the validity of the structure, unless this is done to directly mimic a structure known to exist, as in the case of Adhikary et al.

Future work will continue this study of the essential amino acids and their ability to form N-H···N bonds in proteins.
REFERENCES


A.1. Gaussian, Quantum Methods, and Basis Sets

Gaussian is an electronic structure software package capable of predicting many properties of atoms, molecules, and reactive systems, such as: molecular energies, structures, vibrational frequencies, and electron densities, utilizing ab initio, density functional theory, semi-empirical, molecular mechanics, and various hybrid methods. The calculations in this thesis utilized the Gaussian-09 suite of programs.

The Gaussian software can utilize many different quantum-chemical methods that attempt to solve the molecular Schrödinger equation associated with the molecular Hamiltonian. Methods that do not include any empirical or semi-empirical parameters in their equations, being derived directly from theoretical principles, with no inclusion of experimental data are called ab initio methods. This does not mean that the solution is an exact one, they are still approximations, but the approximations are based on rigorously defined quantum principles.

An ab initio method was used in this thesis, specifically, Møller-Plesset second-order perturbation (MP2) theory, which is based on the subdiscipline of mathematical physics known as many-body perturbation theory. This method adds corrections to the simplest ab initio theory, Hartree-Fock (HF), to achieve much better accuracy for systems by incorporating electron correlation effects by means of Rayleigh-Schrödinger perturbation theory (RS-PT). The principle of Møller-Plesset theory is that an assumption can be made that there exists a HF wavefunction $\Psi_0$ and energy $E_0$ that lie near the exact wavefunction $\Psi$ and energy E. This assumption then allows the
Hamiltonian operator to be written as an expression of the HF Hamiltonian operator, plus a small perturbation \(V\) that can be tuned by some dimensionless parameter \(\lambda\).

\[
H = H_0 + \lambda V
\]

Expanding the exact wavefunction with the incorporated perturbations yields a HF wavefunction \(\Psi\) and energy \(E\) of:

\[
\Psi = \Psi_0 + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \ldots
\]

and

\[
E = E_0 + \lambda E^{(1)} + \lambda^2 E^{(2)} + \ldots
\]

The result is that the HF energy is simply the sum of the zeroth- and first- order energies and the remaining \(n^{th}\)-order energies correspond to the electron correlation energies, of which the 2\(^{nd}\) order energy is the associated MP2 energy.\(^6,7\) Methods that incorporate further perturbation energies can be used and are written MP3, MP4, etc. The increase in accuracy, however, is not great and rarely worth the increased computational cost.\(^8\) MP2 theory is used extensively in the literature and has proven to have high accuracy for organic systems and intermolecular interactions which are the topic of this thesis.\(^9-13\)

Together with a quantum chemical method, such as MP2, a basis set must also be chosen. A **basis set** is a set of (basis) functions that are used to represent the electronic wavefunction in the chosen theory (MP2) in order to turn the partial differential equations that describe the system into a problem in linear algebra.\(^4\) A further explanation of basis sets is detailed here followed by the reasons for the two different basis sets used in this thesis.
There are two typical flavors of basis function: the Slater Type Orbital (STO) and the Gaussian Type Orbital (GTO). The STOs are expressed in terms of the radius, \( r \), and a parameter, \( \zeta \), which describes the nature of the orbital, and are introduced in the function:

\[
STO(r) = \left( \frac{\zeta^3}{\pi} \right)^{1/2} e^{-\zeta r}.
\]

The behaviors of the STO functions are well-known in their ability to excellently describe the near and far regions of the atomic nucleus. They satisfy the nuclear cusp condition because of their exponential relationship with the nuclear distance, both as \( r \to 0 \) and \( r \to \infty \). This relationship allows the STO basis set (a collection of numbers inserted into the basis function) to reproduce the regions near the nucleus, without increasing the angular quantum numbers, correctly. The problem that arises with STOs is, despite their excellent ability to describe energy of an orbital as a function of \( r \), they are only practical with one-electron systems. Thus, an approximation has to be made to apply STO calculations to multi-electron systems, which leads to the Gaussian Type Orbitals (GTOs). GTOs have the form:

\[
GTO(r) = \left( \frac{2\alpha}{\pi} \right)^{0.75} e^{-\alpha r^2},
\]

where \( \alpha \) is a numerical value that describes the orbital and \( r \) is the radius of orbital. Mathematically, GTOs are a simpler and less computationally costly alternative than STOs due in part to the squaring of \( r \). Although GTOs are simpler to employ they cannot accurately represent STOs by themselves. To overcome this difficulty, several researchers in quantum chemistry have curve fitted Slater orbitals to sums of Gaussian functions, the fit improving with \( N \), the number of Gaussian functions.
The STO-6G basis function offers a simple example for the combination of GTOs to represent a STO. In the STO-6G basis set, all atomic orbitals (AOs) are described by the sum of six Gaussian functions, an example of the 1s AO is shown below

$$\phi_{1s}^{STO}(r) = \sum_{i=1}^{N=6} c_{1si} \phi_{1s}^{GF}(r, \alpha_{1si}),$$

where $\phi_{1s}^{STO}$ is the approximate Slater orbital, $\phi_{1s}^{GF}$ is the GTO and $c_{1si}$ is the contraction coefficient, a value chosen to optimize the shape of the basis function sum and ensure normalization.4,17

One of the major limitations of STO-6G, and any of the other STO-NG basis sets is that they use fixed exponents, $\alpha_{ki}$, that is, all orbitals of the same type are identical in size, which will generally not provide an accurate picture of the electron density of a particular atom within a molecule. A solution to this particular problem is to express each atomic orbital by a sum of two or more STOs, referred to as split-valence, that differ only in the value of the exponent $\zeta$. For example, a 2s orbital written as the sum of three STOs, called a triple-$\zeta$, is written as

$$\phi_{2a}(r) = d_1 \phi_{2s}^{STO}(r, \zeta_1) + d_2 \phi_{2s}^{STO}(r, \zeta_2) + d_3 \phi_{2s}^{STO}(r, \zeta_3),$$

where $d$ is a contribution constant indicating how much each STO contributes to AO.

Addition of diffuse functions to the description of each AO allows the orbitals to occupy larger regions of space using large-size versions of $s$- and $p$-type functions, which is vital for systems where electrons are relatively far from the nucleus. Split-valence basis sets and diffuse functions allow orbitals to change size but not shape. To remove this limitation, incorporation of polarization functions can add orbitals with angular momentum beyond what is required for the ground state to the description of each atom.1
Two basis sets were used in this thesis; a smaller set for preliminary geometry optimizations, followed by a larger set for more accurate results. This is a common method that results in shorter calculation times. The smaller basis set 6-31+g* function by Pople is a split-valence double zeta basis set and incorporates 6 \( d \)-type polarizations on Li through Ca and 10 \( f \)-type polarization functions on Sc through Zn. Diffuse functions are also incorporated on atoms other than hydrogen and are important for accurate modeling of intermolecular bonding.\(^{18-20}\)

The second basis set used is Dunning’s\(^2\)\(^\text{a}\)\textsuperscript{2}\text{augmented, electron correlation-constant polarized (valence-only) double zeta (aug-cc-pvdz)}.\(^{23}\) This is one of the most widely used basis sets due to the highly accurate results achieved when paired with post-HF methods such as MP2.\(^9,10,24\)

**A.2. Counterpoise Correct and Basis Set Super Position Error**

Basis set superposition error (BSSE) is the phenomenon referring to the artificial shortening of intermolecular distance and artificial strengthening of the intermolecular interactions between two or more species.\(^{25}\) It should be noted that this error is not a result of the basis set itself but instead arises from an inconsistent treatment of the basis set for each of the monomers. Duijneveldt\(^2\)\textsuperscript{6} explains this inconsistency in terms of two hypothetical monomers A and B as follows: both monomers A and B possess their own set of basis functions when infinitely separated. As monomer A approaches B, however, the dimer can be artificially stabilized as monomer A utilizes the extra basis functions on B to describe its own electron density, and vice versa. The inconsistent treatment of each monomer as the intermolecular distance is varied is the source of the BSSE.
The Boys and Bernardi counterpoise correction (CP) offers a solution for removing the BSSE. In principle the interaction energy of the two molecules A and B can be computed exactly as:

\[ \Delta E_{\text{int}}(AB) = E_{AB}^{AB}(AB) - E_A^A(A) - E_B^B(B), \]

where the superscripts denote the basis used, the subscripts denote the geometry, and the symbol in parentheses denotes the chemical system considered. Thus \( E_{AB}^{AB}(AB) \) represents the energy of the bimolecular complex AB evaluated in the dimer basis, computed at the geometry of the dimer. This designation also holds true for the monomers.

The current issue with the above equation is that it does not account for the stabilization that may occur for either the A or B monomers when they use extra basis functions from one another as described previously. A correction that can be added in to attempt to adjust for the BSSE can be expressed as:

\[ E_{\text{BSSE}}^A(A) = E_{AB}^{AB}(A) - E_A^A(A) \]

\[ E_{\text{BSSE}}^B(B) = E_{AB}^{AB}(B) - E_B^B(B) \]

where the energy of monomer A, in its monomer basis, is subtracted from the energy of monomer A in the dimer basis. Equation 3 is the same approach for monomer B. It should be noted that at this point the assumption is that there will be no deformation of the geometries of A and B as they form the bimolecular complex. This is often a very good approximation and simplifies the procedure. However, for the sake of completeness, it is important to also address the treatment for when geometries are deformed during complexation. In this treatment the energy of monomer A in the dimer basis is taken to be
lower (more stable) than the energy of monomer A in the monomer basis, so $E_A^{AB}(A) < E_A^A(A)$, which makes $E_A^{BSSE} < 0$ as defined above resulting in an error that is stabilizing. Subtracting this error from Equation 1, the terms $E_A^A(A)$ and $E_B^B(B)$ cancel yielding the corrected counterpoise interaction energy:

$$\Delta E_{int}^{CP}(AB) = E_A^{AB}(AB) - E_A^A(A) - E_B^B(B).$$

A.3. Natural Bond Orbital Analysis

Natural Bond Orbital (NBO) analysis calculates the electronic density distribution of atoms and bonds by analyzing a many-electron molecular wavefunction in terms of localized electron-pair bonding units. The input atomic orbital basis set is transformed via natural atomic orbitals (NAOs) and natural hybrid orbitals (NHOs) into natural bond orbitals (NBOs). The NBOs obtained in this fashion correspond to the widely used Lewis picture, in which two-center bonds and lone pairs are localized with maximum electron density. This section provides a brief introduction to NBO algorithms and nomenclature.

NBO analysis is based on a method for optimally transforming a given wavefunction into localized form, corresponding to the one-center “lone pair” and two-center “bond” elements of the Lewis structure picture. The NBOs are obtained as local block eigenfunctions of the density matrix and are hence “natural” in the sense of Löwdin, having optimal convergence properties for describing the electron density. The set of high-occupancy NBOs, each taken doubly occupied, is said to represent the “natural Lewis structure” (NLS) of the molecule. Delocalization effects appear as weak departures from this idealized localized picture. The various natural localized sets can be considered to result from a sequence of transformations of the input atomic orbital basis
set,

\[ \text{input basis } \text{AOs } \rightarrow \text{NAOs } \rightarrow \text{NHOs } \rightarrow \text{NBOs } \rightarrow \text{NLMOs} \]

Each natural localized set forms a complete orthonormal set of one-electron functions for expanding the delocalized molecular orbitals (MOs) or forming matrix representations of one-electron operators. The overlap of associated “pre-orthogonal” NAOs, lacking only the interatomic orthogonalization step of the NAO procedure, can be used to estimate the strength of orbital interactions in the usual way, based on Mulliken-type approximations.

The optimal condensation of occupancy in the natural localized orbitals leads to partitioning into high- and low-occupancy orbital types (reduction in dimensionality of the orbitals having significant occupancy), as reflected in the orbital labelling. The small set of most highly-occupied NAOs, having a close correspondence with the effective minimal basis set of semi-empirical quantum chemistry, is referred to as the “natural minimal basis” (NMB) set. The NMB (core + valence) functions are distinguished from the weakly occupied “Rydberg” (extra-valence-shell) functions that complete the span of the NAO space, but typically make little contribution to molecular properties. Similarly, in the NBO space, the highly occupied NBOs of the natural Lewis structure (NLS) can be distinguished from the “non-Lewis” antibonding and Rydberg orbitals that complete the span of the NBO space. Each pair of valence hybrids \( h_A \), \( h_B \) in the NHO basis give rise to a bond (\( \sigma_{AB} \)) and antibond (\( \sigma^*_{AB} \)) in the NBO basis,

\[
\sigma_{AB} = c_A h_A + c_B h_B \\
\sigma^*_{AB} = c_B h_A - c_A h_B
\]
the former a Lewis (occupied) and the latter a non-Lewis (unoccupied) orbital. The antibonding (valence shell non-Lewis orbitals) typically play the primary role in departures (delocalization) from the idealized Lewis structure.

The NBO program also makes extensive provision for energetic analysis of NBO interactions. This analysis is carried out by examining all possible interactions between "filled" (donor) Lewis-type NBOs and "empty" (acceptor) non-Lewis NBOs and estimating their energetic importance by 2nd-order perturbation theory. Since these interactions lead to donation of occupancy from the localized NBOs of the idealized Lewis structure into the empty non-Lewis orbitals (and thus, to departures from the idealized Lewis structure description), they are referred to as "delocalization" corrections to the zeroth-order natural Lewis structure. For each donor NBO \(i\) and acceptor NBO \(j\), the stabilization energy \(E(2)\) associated with delocalization ("2e-stabilization") \(i \rightarrow j\) is estimated as

\[ E(2) = \Delta E_{ij} = q_i \frac{F(i, j)^2}{\varepsilon_j - \varepsilon_i} \]

where \(q_i\) is the donor orbital occupancy, \(\varepsilon_i, \varepsilon_j\) are diagonal elements (orbital energies) and \(F(i, j)\) is the off-diagonal NBO Fock matrix element.\(^{28}\)

**A.4. Symmetry-Adapted Perturbation Theory**

Symmetry-adapted perturbation theory (SAPT) is a method designed to calculate intermolecular interaction energies by starting with isolated monomers and treating the interactions as small perturbations of the system.\(^{29}\) Intermolecular interactions are generally quite small energetically when compared with intramolecular interactions.
within the monomers and so the distortion of the molecules due to their mutual interactions is relatively minor. This is important for any perturbation theory. SAPT starts from unperturbed molecules (isolated monomers) and treats the interaction energy and wave function as small quantities resulting from the mutual perturbation of monomers by coulombic intermonomer interactions.

First, the Schrödinger equation is solved for the isolated monomers A and B:

\[ H_X \Phi_X = E_X \Phi_X; \quad X = A \text{ or } B, \]

where \( H_X, \Phi_X, \) and \( E_X \) are, respectively the Hamiltonian, wave function, and energy of monomer X. Next, the monomers are placed in the dimer configuration and all electrons and nuclei of monomer A interact with those of monomer B according to Coulomb’s Law. The sum of the Coulomb interaction terms is the intermolecular interaction operator, V, and the Hamiltonian of the dimer is

\[ H = H_A + H_B + V. \]

The effect of V can be accounted for using Rayleigh-Schrödinger (RS) perturbation theory. (For a detailed explanation of RS perturbation theory see Szalewicz’s book.)

The interaction energy is expressed as the sum of the perturbation corrections

\[ E_i = E_{RS}^{(1)} + E_{RS}^{(2)} + ... \]

This approach fails to predict the existence of van der Waals minima on potential energy surfaces. This is due to the unfulfilled Pauli Exclusion Principle by both unperturbed and RS wave functions. (The permutation [exchange] of electrons between monomers does not give the same wavefunction to within a sign.) The Pauli-correct wavefunction is \( \mathcal{A} \Psi \) where the operator \( \mathcal{A} \) is called the antisymmetrizer, which is the
sum of intermonomer perturbations operators with the appropriate signs.

\( A \Psi \) is not an eigenfunction of \( H_0 \), and so RS theory can no longer be used, but there have been symmetry-adapted perturbation theories that are able to incorporate it. The most commonly used theory is the symmetrized RS (SRS) method. Wave functions are computed by RS equations, then are antisymmetrized before computing the interaction energy contributions. The difference between the original RS energies and the antisymmetrized contributions is termed the exchange energy, \( E_{Ex}^{(1)} \), and the interaction energy can now be written as

\[ E_i = E_{RS}^{(1)} + E_{Ex}^{(1)} + E_{RS}^{(2)} + E_{Ex}^{(2)} \ldots \]

**Interpreting the Results**

Perhaps the most important feature of SAPT is that it allows one to predict qualitatively how strong the intermolecular interactions will be for different dimers. This is possible because SAPT interaction energies are directly related to monomer properties and are naturally decomposed into physically interpretable components. The main components are the electrostatic, induction, dispersion, and exchange contributions.

The *electrostatic energy* is simply \( E_{RS}^{(1)} \) and describes the Coulomb interaction of charge distributions of monomer A with that of monomer B. At large separation, this can be reduced to the interactions between multipole moment of the monomers.

The *induction (polarization) and dispersion* energies are the two parts of \( E_{RS}^{(2)} \). The wave function component giving \( E_{RS}^{(2)} \) is a sum of the products of the monomer wavefunctions. Large separation shows a term that can be interpreted as the response of the density of this monomer to the electric field of the permanent charge distribution of
the other monomer. It is the *induction (polarization) energy*. At large separations, the induction energy can be reduced to products of polarizabilities and permanent multipole moments.

The remaining term, where both monomers are excited, is the *dispersion energy*. This term does not have any classical interpretation; it results from the quantum correlation of electronic motion between the monomers.

Finally, the *exchange energies* (the difference between the SAPT and TS values of perturbation corrections) result from the action of the antisymmetrizer and can be physically interpreted as the effects of electron tunneling through the potential barrier.29

### A.5. Electron Density Shift Maps

The electron density maps reported in this thesis represent the gain or loss of electron density during complexation as shown by red and blue areas on the structure. These are made by first calculating the total electron density for the complex, and each monomer. Then, it is a simple matter to subtract the electron density of the monomers from the complex and graphing the results.

### A.6. Nuclear Magnetic Resonance

The basic principle of Nuclear Magnetic Resonance (NMR) is to apply an external magnetic field, $B_0$, and measure the frequency at which the nucleus achieves resonance. Electrons orbiting around the nucleus generate a small magnetic field that opposes $B_0$; the electrons are shielding the nucleus from $B_0$.

The greater the electron density around the nucleus, the higher the opposing magnetic field to $B_0$ from the electrons, and the greater the shielding of the nucleus.
Because the proton experiences a lower external magnetic field, it requires a lower frequency to achieve resonance.

If the electron density around the nucleus decreases, as happens at the formation of a hydrogen bond, the opposing magnetic field becomes smaller and therefore, the nucleus experiences more of $B_0$, and a higher frequency is required to achieve resonance. This phenomenon is called deshielding and can be used, in conjunction with other evidence, to show that a hydrogen bond has formed.\textsuperscript{32}

\textbf{A.7. Atoms in Molecules (AIM) Analysis}

The quantum theory of atoms in molecules (QTAIM) is a model of molecular and condensed matter electronic systems in which the principle objects of molecular structure – atoms and bonds – are natural expressions of a system’s observable electron density distribution function, $\rho(\mathbf{r})$.\textsuperscript{33} The Hohenberg-Kohn theorem\textsuperscript{34} confirms that $\rho(\mathbf{r})$ is the fundamental property that characterizes the ground state of a system – once $\rho(\mathbf{r})$ is known, the energy of the system is uniquely defined, and from there a diverse range of molecular properties can, in principle, be deduced. An analysis of the topology of $\rho(\mathbf{r})$ leads directly to the chemical concepts of atoms, molecules, structures, and bonds.

The greatest amount of $\rho(\mathbf{r})$ directly surrounds the nuclei and fans out spherically. For a molecule in the gas phase, the trajectories will mostly terminate at infinity. In special cases, however, they will terminate at another nucleus; these special trajectories are known as \textbf{bond paths}. At the point of termination for both atoms is the minimum of electron density along the bond path and is called the \textbf{bond critical point}. Two other trajectories leave the critical point perpendicular to the bond path making it a saddle
The Laplacian of the electron density, $\nabla^2(\rho)$, traces the effects of chemical bonding in the total charge density. It is used to identify both the bond path and the bond critical point. Along the bond path, $\nabla^2(\rho)$ is negative and at the bond critical point, $\nabla^2(\rho)$ is still negative perpendicular to the bond path, but is positive along it, identifying the saddle point.

The electron density at the critical point is thought to be very closely related to the strength of the bond. The greater the density at that point, the greater the electron density, in general, along the bond path, and the stronger the bond. This is true for both covalent and noncovalent bonds. Popelier identified ranges of density at the bond critical point and associated the strength of bonds by it. His work was used in this thesis to categorize H-bond strengths.
REFERENCES


(28) Weinhold, F.; Glendening, E. D. *NBO 6.0 Program Manual: Natural Bond Orbital Analysis Programs*. Board of Regents of the University of Wisconsin: Madison,
2013.


APPENDIX B
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.
Order Completed

Thank you for your order.

This Agreement between Sandra Lundell ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number 4373210931499
License date Jun 20, 2018
Licensed Content Elsevier
Publisher
Licensed Content Publication
License Content Title Unconventional N-H...N Hydrogen Bonds Involving Proline Backbone Nitrogen in Protein Structures
License Content Author R. N. V. Krishna Deepak, Ramasubbu Sankaramakrishnan
License Content Date May 10, 2016
License Content Volume 110
License Content Issue 9
License Content Pages 13
Type of Use reuse in a thesis/dissertation
Portion figures/tables/illustrations
Number of figures/tables/illustrations 4
Format both print and electronic
Are you the author of this Elsevier article? No
Will you be translating? No
Original figure numbers Figures 1, 2, 3a, 4
Title of your thesis/dissertation Quantum Mechanical Studies of N-H...N Hydrogen Bonding in Acetamide Derivatives and Amino Acids
Expected completion date Jun 2018
Estimated size (number of pages) 116
Requestor Location Sandra Lundell
1620 N 200 E
325
LOGAN, UT 84341

https://s100.copyright.com/AppDispatchServlet
Direct Detection of a Histidine-Histidine Side Chain Hydrogen Bond Important for Folding of Apomyoglobin

Author: Mirko Hennig, Bernhard H. Geierstanger
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Jun 1, 1999

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions. Comments? We would like to hear from you. E-mail us at customerservice@copyright.com
Imidazole Nitrogens of Two Histidine Residues Participating in N-H···N Hydrogen Bonds in Protein Structures: Structural Bioinformatics Approach Combined with Quantum Chemical Calculations

Author: Abhishek Hariharan Iyer, R. N. V. Krishna Deepak, Ramasubbu Sankararamakrishnan

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Jan 1, 2018

Quick Price Estimate

I would like to... reuse in a Thesis/Dissertation

Requestor Type Non-profit

Portion Table/Figure/Micrograph

Number of Table/Figure/Micrographs 1

Format Print

Select your currency USD - $

Quick Price Click Quick Price

To request permission for a type of use not listed, please contact the publisher directly.

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement, Terms and Conditions, Comments? We would like to hear from you. E-mail us at customercare@copyright.com

https://s100.copyright.com/AppDispatchServlet#termTop

1/1
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.
Jul 16, 2018

Sandra Lundell
UTAH STATE UNIVERSITY
1620 N 200 E, #325
LOGAN, UT 84341

Dear Sandra Lundell,

You have our permission to include content from our text, *FUNDAMENTALS OF GENERAL, ORGANIC, AND BIOLOGICAL CHEMISTRY, 7th Ed.* by McMurry, John E.; Hoeger, Carl A.; Peterson, Virginia E.; Ballantine, David S., in your dissertation or masters thesis at Utah State University.

Content to be included is:
218 Figure 8.6

Please credit our material as follows:

Sincerely,
Michael Prince,
Permissions Granting Analyst
Jun 29, 2018

Sandra Lundell
UTAH STATE UNIVERSITY
1620 N 200 E, #325
LOGAN, UT 84341

Dear Sandra Lundell,

You have our permission to include content from our text, CAMPBELL BIOLOGY, 10th Ed. by REECE, JANE B.; URRY, LISA A.; CAIN, MICHAEL L.; WASSERMAN, STEVEN A.; MINORSKY, PETER V.; JACKSON, ROBERT B., in your dissertation or masters thesis at Utah State University.

Content to be included is:

80 Figure 5.18, secondary structure

Please credit our material as follows:
REECE, JANE B.; URRY, LISA A.; CAIN, MICHAEL L.; WASSERMAN, STEVEN A.; MINORSKY, PETER V.; JACKSON, ROBERT B., CAMPBELL BIOLOGY, 10th, ©2014.

Sincerely,
Michael Prince,
Permissions Granting Analyst