Probing Secondary Structures of Small Peptides in Gas Phase as Well as Condensed Phase

A thesis

submitted in partial fulfilment of the requirements

of the degree of

Doctor of Philosophy

By

Satish Kumar

Registration No. 20153422

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

(2022)

In Memory of my Grandfather

Late Shri Visheshwar Mahto

Dedicated to

Late Mr. Kshetrimayum Borish

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pashan Pune – 411008

Declaration

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus notbeen properly cited or from whom proper permission has not been taken when needed.

The work reported in this thesis is the original work done by me under the guidance of Prof. Aloke Das.

Date :17/06/2022

Mr. Satish Kumar

20153422

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pashan Pune – 411008

Certificate by Supervisor

I certify that the thesis entitled "Probing Secondary Structures of Small Peptides in Gas Phase as well as condensed phase" presented by Mr. Satish Kumar represents his original work which was carried out by him at IISER, Pune under my guidance and supervision during the period from 2015 to 2022.

The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution. I further certify that the above statements made by him regarding his thesis are correct to the best of my knowledge.

Date: 17/06/2022

Prof. Aloke Das

(Research Supervisor)

Acknowledgment

I want to acknowledge my Ph.D. thesis supervisor, Prof. Aloke Das, for a very constant role in guidance, motivation, discussions, and experiment design throughout my Ph.D. tenure. His invaluable ideas and comments on my research work enabled me to carry out the experiments efficiently and smoothly. The constant input towards scientific writing and thinking helped me to better my skills in writing and presenting the research work.

I sincerely thank my research advisory committee members, Prof. H. N. Gopi and Dr. Sayan Bagchi for their invaluable scientific discussions and suggestions throughout my research tenure at IISER PUNE.

I would like to acknowledge my research collaborators Dr. Biplab Sarkar, Dr. Jayashree Nagesh, Mr. Sanjit Dey, and Mr. Manjeet Singh for their support and help in various research work.

I would like to thank the former director for establishing a world-class research facility at IISER PUNE. I would like to sincerely acknowledge all the instruments technicians of IISER Pune for their support during my Ph.D. tenure. I further like to acknowledge all the staff members of the chemistry department at IISER PUNE.

I would like to thank my former lab member Dr. Santosh Kumar Singh and Dr. Kamal Kumar Mishra for their experimental teachings and scientific discussions. A special thanks go to the late Kshetrimayum Borish for making the laser desorption set-up very efficient. His hard-working ability and innovative thinking will always be remembered. I would like to thank my present lab members Mr. Prakash Panwaria, Mr. Surajit Metya, Sourav Mandal, and Shubham for their constant support and help. I would like to thank all members of Prof. H.N. Gopi lab for their help.

I would like to thank my friends at IISER PUNE and home for their support in tough times including Mohit, Dheeraj, Rishabh, Virender, Sachin, Manu, Aman, Rakesh, Rohit, Jitender, Prakash, Divya, Jyoti, Yashwant, and many more. A lot of thanks to all the cricket friends at IISER PUNE.

I would like to thank my parents Shri Chandrika Prasad and Shrimati Urmila Devi for their unconditional support during the challenging period at IISER PUNE. I want to thank my elder brothers Mr. Shambhu Kumar, Mr. Santosh Kumar, and Mr. Sudhir Kumar for their guidance and support. My eldest brother Mr. Shambhu Kumar is the main reason behind the motivation behind joining the research field. I want to thank my sisters-in-law Mrs. Sweta, Mrs. Sony, and Mrs. Neha for their faith and support. I would like to thank my younger brother Mr. Sachin Kumar for the encouragement and support. I want to thank my elder sisters Meena Didi and Veena didi for their care and support. I would thank Neha, Nikki, Ravina, Shubham, Aryan, Sanvi, Deveshi and Setu for creating a positive environment whenever I took a break from research and went home.

Table of content

List of Abbreviations

NMR	Nuclear Magnetic Resonance
1C-R2PI	One color resonantly enhanced two-photon ionization
2C-R2PI	Two-color resonantly enhanced two-photon ionization
Z	Benzyloxycarbonyl
Aib	Aminobutyric acid
Ala	Alanine
Gly	Glycine
Pro	Proline
FTIR	Fourier transform infrared
CD	Circular dichroism
SMD	Solvation method based on Density
PDB	Protein data bank
CSD	Cambridge structural database
Phe	Phenylalanine
Glu	Glutamic acid
Asp	Aspartic acid
PCM	Polarizable continuum model
DFT	Density functional theory
BBO	Beta barium borate
KDP	Potassium dihydrogen phosphate
Boc	Tert-Butyloxycarbonyl

List of Figures

List of Tables

SYNOPSIS

Proteins, which are one of the most important biomolecules for living beings, consist of amino acids joined by peptide bonds.[1, 2] The functions and properties of proteins are not only dependent on the sequence of amino acids but also on the spatial conformations adopted by proteins due to the hydrogen bonding network along the polypeptide chain.[3] Other factors such as hydrophobic interactions etc. are also very important for specific folded structures of proteins.[4] The study of overall conformation adopted by proteins is a challenging task and it can be studied by focusing on selective parts of the protein. Peptides are defined as smaller segments of proteins having two or more amino acid residues joined by peptide bonds.

Condensed phase study of peptides is performed through various techniques such as Nuclear Magnetic Resonance, FTIR, X-ray crystallography, Cryogenic electron microscopy, circular dichroism spectroscopy, etc.[5, 6] Condensed phase study generally provides very important information about the global minimum structure of the peptide or average information on multiple low energy conformers from their broad spectroscopic features.[7] However, an advantage of the gas phase spectroscopy over condensed phase spectroscopy is that one can obtain detailed information on the structures of several low-energy conformations of small peptides which are important for the secondary structures of various neighbouring states including the global minimum of polypeptides or proteins.[8] Essentially, gas-phase spectroscopy combined with quantum chemical calculations can be used to reveal the intrinsic conformational preferences of peptides.[9]

The protein structure can be broadly classified into four different types namely primary, secondary, tertiary, and quaternary structures. One of the most important and useful classes of

protein structures for understanding the properties and functions of protein is the secondary structure.[10] The secondary structure of the proteins or peptides can be classified into mainly turns, sheets, and helices based on their hydrogen-bonding pattern and conformation.[8, 11] Although sheets and helices are more commonly found in large proteins, turns can be regularly observed in smaller peptides also.

The turns in proteins are the region where the overall direction of the polypeptide chain is reversed.[12] Many significant and important biological phenomena such as hydroxylation and enzymatic activity mostly occur through the loop region of the turns.[13, 14] Based on the hydrogen bonding between different amino acid residues of the polypeptide chains, turns can be classified into δ, γ, β, α, and π.[15, 16] In general, the turn structures arise due to inter-residue hydrogen bonding interactions. However, an intra-residue hydrogen bond termed as C5 hydrogen bond has been recently found to be quite important for the stabilization of the secondary structures of the peptides.[17] A non-covalent interaction similar to the C5 hydrogen bond was observed by Nyquist in 1963.[18] Since then, the C5 hydrogen bond has been observed in the solution, crystal structure, and gas-phase studies of peptides through various spectroscopic techniques.[1, 17, 19, 20]

The secondary structures of the peptides and proteins are primarily based on the hydrogen bonding interactions in the backbone as well as the sequence of the amino acid residues present there. It has been found that the Pro-Gly sequence prefers folded β-turn conformation whereas the Gly-Pro sequence has more propensity to form extended polyproline II type structure.[21] The β-turn conformation present at the Pro-Gly sites in procollagen is a prerequisite for the necessary enzymatic hydroxylation which provides the triple-helical structure of the collagen.[13] Studies by Brahmachari and co-workers further highlighted that the neighbouring residue X plays a vital role in the β-turn stabilization in a Pro-Gly-X sequence.[14] Very selective amino acids such as glycine, alanine, leucine, isoleucine, and phenylalanine as an X residue were

found to have a high propensity to stabilize the β-turn formation in a Pro-Gly-X (X=Leu > Ala > Gly, Ile > Phe) sequence.[22] The studies on the conformational preference of peptides in the condensed phase have a long history whereas the studies in the gas phase have been popular in the last 2-3 decades.[8]

Isolated gas-phase studies can reveal the intrinsic properties of the peptides void of external interactions such as solvent interaction and intermolecular interactions.[23] Various gas-phase spectroscopic groups have revealed the conformational preferences of several small peptides with the use of gas-phase spectroscopic techniques and quantum chemical calculations.[24-27] This thesis aims to study the conformational details of small peptides containing glycine, proline, and alanine amino acids in gas as well as condensed phases. Through this work, we have tried to address the following aspects of the peptide conformations in the crystal structure as well as the gas and solution phase.

1. The intra-residue C5 hydrogen bond has shown to exist alongside the conventional turns in peptide sequences in gas-phase and condensed phases. We tried to show that a relatively weaker C5 hydrogen bond is solely able to stabilize the most stable conformation of a dipeptide.
2. Sequence-dependent studies on model capped gly-pro and pro-gly sequences in condensed and gas phase were performed to find out the conformational preferences of the selected sequences.
3. The role of residue X in stabilizing the β-turn conformation adopted by the Pro-Gly-X residue was also addressed in this thesis work.

Chapter 1 provides a brief introduction to the structures of proteins, protein folding, various secondary structures in proteins, and studies on sequence-dependent conformational preferences of small peptides in solution and gas phases. A special emphasis has been put on the discussion of the peptides having Gly-Pro and Pro-Gly sequences. A brief discussion of the

previous work on isolated gas-phase studies of small peptides carried out by various research groups has been presented here. Chapter 1 also contains the aim of the present thesis.

Chapter 2 describes different experimental methods along with computational methods used for the thesis work. Home-built jet-cooled laser desorption Time-of-Flight mass spectrometer set-up for performing gas-phase spectroscopy experiments has been discussed in detail. We have also discussed basic principles of supersonic jet cooling, laser desorption, Time-of-Flight mass spectrometry, and various gas-phase spectroscopic techniques to measure conformation-specific electronic and IR spectra of the peptides. Various spectroscopic techniques in the condensed phase such as 1D/2D NMR and FTIR spectroscopy including single-crystal X-ray crystallography have been discussed briefly. This chapter also provides a very brief description of the computational methods employed to carry out the quantum chemical calculations of the peptide molecules.

Chapter 3 includes mostly the gas-phase studies of Z-Gly-Pro-OH revealing the role of the C5 hydrogen bond in stabilizing the most stable conformation of the peptide. Gas-phase UV/IR spectroscopy experiments combined with DFT calculations reveal that the most stable conformer of Z-Gly-Pro-OH has extended conformation solely stabilized by an intra-residue C5 hydrogen bond. The other experimentally observed higher energy conformer in the gas phase is stabilized by an OH-π hydrogen bond.

In **Chapter 4**, we have studied sequence-dependent conformational preferences of capped Pro-Gly and Gly-Pro dipeptides in the gas phase as well as the condensed phase. In the case of the Boc-^{D}Pro-Gly-NHBn-OMe peptide, it has been found that both the observed conformers in the gas phase have C7-C7 hydrogen-bonded conformation. The solution phase 2D NMR and FTIR spectroscopy reveal that the Pro-Gly sequence adopts the β-turn conformation in the solution. All of the three observed conformers of Boc-Gly-^{D}Pro -NHBn-OMe peptide in the gas phase

show extended β-strand or polyproline-II (PP-II) structure, having C5-C7 hydrogen bonding interactions, which correlates well with the structure obtained from solution phase spectroscopy and XRD. Overall results of these two sequences demonstrate a fine interplay of the intrinsic conformational properties, solvation, and crystal packing of the peptides.

Chapter 5 includes the studies on the role of the neighbouring residue X (X=Ala) on the stabilization of the β-turn in the Pro-Gly-X sequence through solution-phase 2D NMR, FTIR spectroscopy, and X-ray crystal structure combined with solution-phase and gas-phase quantum chemistry calculations. The crystal structure of the Boc-^{D}Pro-Gly-Ala sequence has a helical C10-C10 conformation. A similar conformation of the Pro-Gly-Ala sequence has been observed in the solution phase through 2D NMR and FTIR spectroscopy. The gas-phase DFT calculations further show that the most stable conformer PGA1-F-C10-C10 has very similar conformations as observed in solution and crystal structure.

In **Chapter 6**, a summary of the work described in the thesis and future perspectives have been presented.

List of Publications

1. **Satish kumar**, Kshetrimayum Borish, Sanjit Dey, Jayashree Nagesh* and Aloke Das*: Sequence dependent folding motifs of the secondary structures of Gly-Pro and Pro-Gly containing oligopeptides, *Phys. Chem. Chem. Phys.*, **2022**, *24*, 18408-18418.
2. ***Satish Kumar***, Kamal K. Mishra, Santosh K. Singh, Kshetrimayum Borish, Sanjit Dey, Biplab Sarkar* and Aloke Das*: Observation of a weak intra-residue C5 hydrogen-bond in a dipeptide containing Gly-Pro sequence *J. Chem. Phys.*, **2019**, *151*, 104309-104317.
3. Kamal K. Mishra, Santosh K. Singh, ***Satish Kumar***, Gulzar Singh, Biplab Sarkar*, M. S. Madhusudhan* and Aloke Das*: Water-Mediated Selenium Hydrogen-Bonding in Proteins: PDB Analysis and Gas Phase Spectroscopy of Model Complexes: *J. Phys. Chem. A.*, **2019**, *123*, 5995-6002.
4. Santosh K. Singh, Shahaji More, ***Satish Kumar***, Kamal K. Mishra, Krishna. N. Ganesh* and Aloke Das*: Conformation-specific IR Spectroscopic Signature for Weak C=O...C=O n→π* Interaction in Capped 4R-Hydroxyproline *Phys. Chem. Chem. Phys.*, **2019**, *21*, 4755-4762.

Chapter 1

Introduction

Introduction:

1.1 Structures of protein

Proteins consisting of a large number of amino acids were the most essential materials required for the creation and function of life in the universe.[15, 28] The peptides can be termed as the small segment of a protein or a short amino acid chain. The structures and functionalities of peptides/proteins depend not only on the sequence of the amino acids and side-chains but also on the spatial arrangements of the amino acids which result from the non-covalent interactions present along the main chain and side-chain residues of the peptides/protein.[2, 3, 11, 29] The first complete structure of a protein was determined in 1952 using an X-ray beam on the crystallized protein.[30, 31] Nucleic Magnetic Resonance (NMR), X-ray crystallography, and Cryogenic electron microscopy among other tools have been used then extensively to determine the structures of proteins in the condensed phase.[5, 6, 32, 33] Molecular dynamics simulation can now provide the conformational analysis of larger proteins in a reasonable time using modern computation facility.[34, 35] Proteins are synthesized by ribosomes in the cell by assembling the amino acids in a particular order linked by the peptide bonds.[36, 37] The synthesized proteins are then folded into their native functional states through various non-covalent interactions along the amino acid chain. Water molecules surrounding proteins also play a very essential role in providing the specific folded structures through hydrophobic interactions.[38, 39] The native conformation of a protein is the state of the protein in which it is active and functional.

1.1.1 Role of protein folding

The folding processes of protein molecules include various high-energy intermediary conformations along the protein folding funnel. The protein folding funnel diagram was introduced by Wolynes, Onuchic, and Thirumalai in 1971 to give a pictorial representation of protein folding.[40] The protein folding funnel is based on the famous Levinthal's paradox which

tells that it is impossible to find the native state of protein folding among the numerous possible conformations of the protein.[41-43] The folding of the protein in its native state is the central part as the misfolding of the protein into an unwanted confirmation leads to various diseases in the living

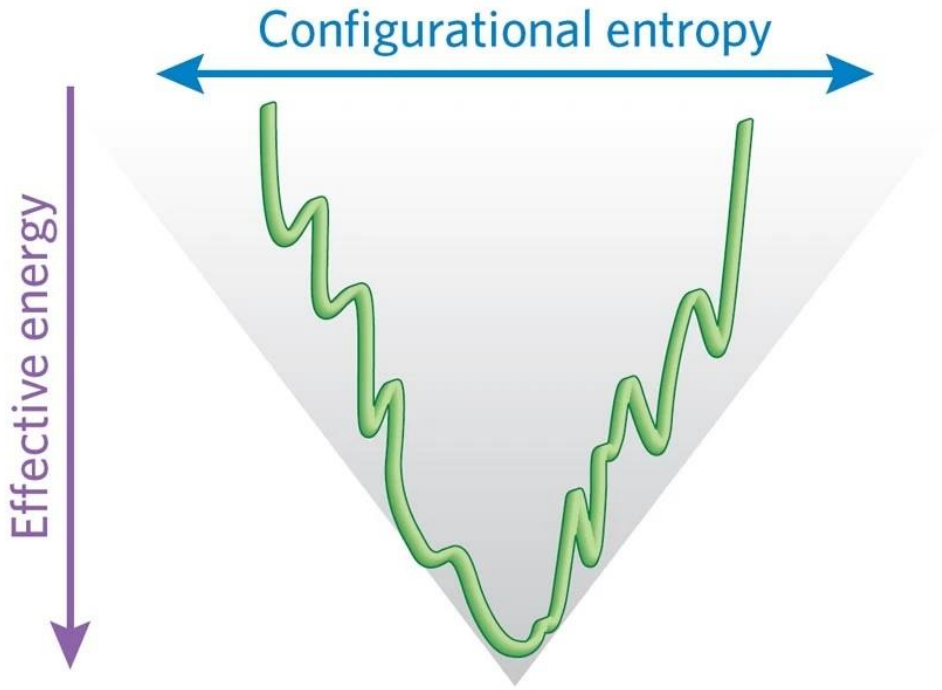

Figure 1.1. The protein folding funnel diagram based on Levinthal diagram. The x-axis represents the configurational entropy whereas y-axis represents the energy. The figure has been adapted in parts with permission from reference no. 43. Copyright [2011] Springer Nature.

organism. In terms of energetics, the native state of protein can be defined as the lowest energy conformation among all the other possible conformations in the protein folding funnel. The quantum chemical calculation of peptides/proteins can reveal the global minima or native state of the protein and the different conformations of the protein folding funnel. The complexity of

the protein folding problem can be addressed by simplifying it to smaller to medium-sized peptides and studying it through different spectroscopic or kinetics techniques.[44, 45]

The work reported in this thesis combines quantum chemical calculations with isolated gas-phase spectroscopy to reveal several low-energy conformers of small peptides. The structures of peptides/proteins can be broadly classified into four different types namely primary, secondary, tertiary, and quaternary structures. A brief overview of all the different types of protein structures has been provided below.

1.1.2 Primary structure of the protein

The primary structures of proteins represent the sequence of the amino acids along the whole polypeptide chain. It was illustrated by Chris Anfinsen in 1973 that the primary structure (sequence of amino acids) of a protein determines its higher-order structures (i.e., secondary, tertiary, and quaternary structures).[46] The primary structures of proteins allow us to compare the proteins in a very simple step and know the percentage occurrence of a particular amino acid in the very long polypeptide chain. The unique primary structure or amino acid sequence has some limitations and cannot dictate the properties of protein as these depend also on the local conformation and environment. The primary structure of insulin was fully obtained by Tuppy and co-workers for the first time in 1951.[47] The primary structure of human insulin has been shown in figure 1.2.[48]

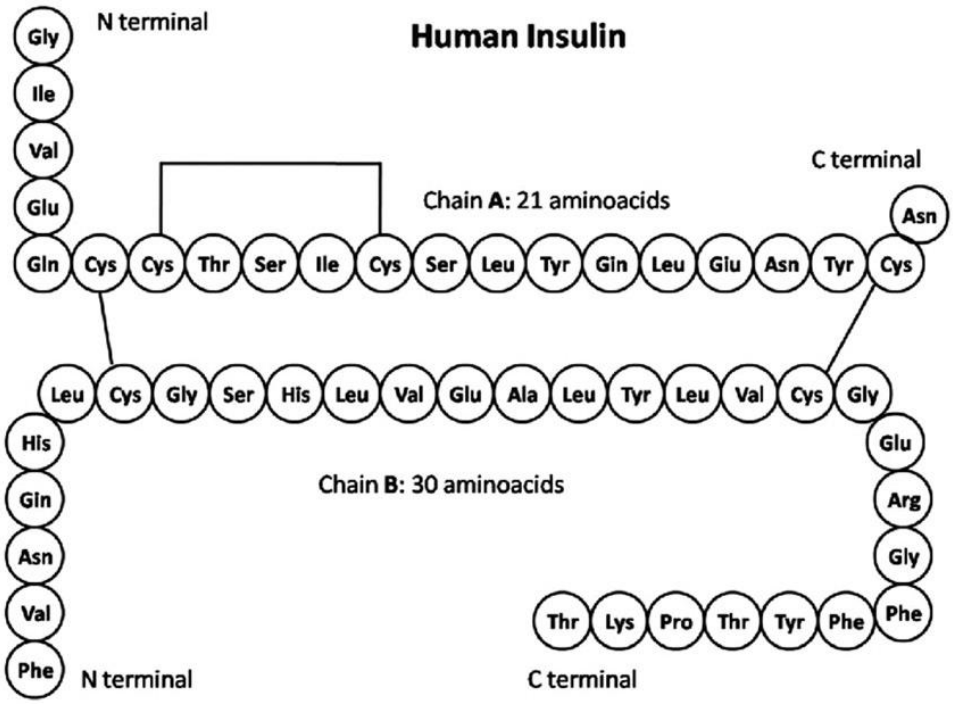

Figure 1.2. The primary structure of human insulin. The figure has been taken with permission from reference no. 48. Copyright [2012] Taylor & Francis.

1.1.3 Secondary structure of the protein

The secondary structures of peptides/proteins are defined as the local conformations governed mainly by the hydrogen bonding interactions in the backbone as well as the sequence of the amino acid residues.[10, 49-51] The secondary structures can be defined by the Ramachandran dihedral angles (Φ and Ψ) along with the hydrogen bonding patterns around the main chain of the polypeptide backbone. The secondary structures of the proteins/peptides are generally stabilized by inter-residue hydrogen bonds between the N-H and C=O groups of the peptide units. However, intra-residue hydrogen bonds named C5 involving the same residue also contribute to the stabilization of the secondary structures, specifically, the β-sheet type of structures. [17] The secondary structures of the proteins/peptides can be further categorized into

mainly three types based on the hydrogen bonding patterns along the main polypeptide chain and these are helices, sheets, and turns. [52]

1.1.4 Tertiary structure of the protein

The Tertiary structure of a protein refers to the three-dimensional structure of a particular single strand or polypeptide chain. The tertiary structure can either have only a single type of secondary structure like turn, sheet or helices or have multiple types of secondary structures.[52] In this configuration, the protein becomes operational as the functional groups of the amino acid residues are aligned outwards and hence allowing the protein strand to interact with other molecules or biological systems. The tertiary structure is further categorized into two parts named globular and fibrous proteins. Apart from the non-covalent interactions between the constituting amino acid residues, the tertiary structure is stabilized by the polar hydrophilic hydrogen and ionic bond interactions. Detailed study of the tertiary structures of proteins has helped the researchers to understand various important biological phenomena such as hydroxylation, and enzymatic activity among others.[53, 54] The tertiary structures of proteins also play a key role in the hormone-regulated receptor activation.[55] The complexity of the tertiary structures of proteins is hard to be analyzed and hence chemists usually study the secondary structures of the active sites of the proteins. The tertiary structure of a protein showing different types of secondary structures has been shown in figure 1.3.[56]

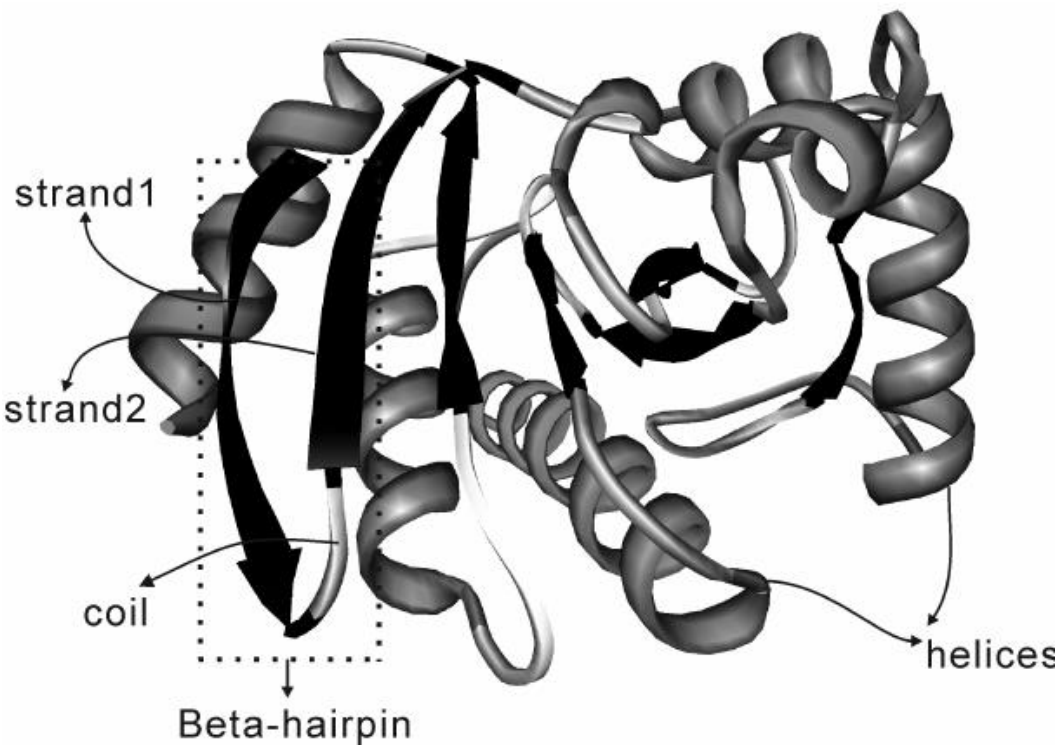

Figure 1.3. Schematic representation of the tertiary structure of the protein showing all the secondary structures such as sheet, helices and loops. The figure has been taken with permission from reference no. 56. Copyright [2012] Springer Nature.

1.1.5 Quaternary structure of the protein

In 1958, Bernal first introduced the term quaternary structure as a supplement to the already known primary, secondary and tertiary structures of proteins.[57] The quaternary structure of proteins is a complete three-dimensional structure containing two or more polypeptide chains stabilized by various types of non-covalent interactions and ionic interactions among many other interactions.[58] Each of the polypeptide chains of the protein has well-defined secondary and tertiary structures.[53, 59, 60] The smallest subunit of a protein that can be identified independently from other subunits is called a monomer. The specific function of a particular

protein depends on the arrangement of the quaternary structure consisting of more than one subunit of the protein. The quaternary structure is one of the most complex structures of proteins and can be understood with the help of smaller secondary and tertiary structures of small units of proteins. The idea of the quaternary structure was first discovered by Svedberg in 1927.[61] He found out the molecular weight of hemoglobin by use of sedimentation in the ultracentrifuge. The value of 68000 Da for the molecular weight of hemoglobin implied that there are four subunits in the protein molecule.

Figure 1.4. Quaternary structure of protein showing all the secondary structures like sheet, helices and loops. The figure has been taken with permission from reference no. 60. Copyright [2007] Elsevier.

1.2 Importance of Secondary structures of protein

Pauling and co-workers first demonstrated that proteins can have their secondary structures as α-helix and the parallel and antiparallel pleated sheets.[30] The secondary structure of the peptide/protein is the most important structure among all the three different types of structures. As we know, the secondary structure of a protein is the three-dimensional structure of a local segment of the protein. Although it is local, the secondary structure has the most important role in the folding process, functioning, and characterization of protein. The secondary structure of a protein is stabilized mainly by hydrogen bonding interactions.[10, 49, 62] The hydrogen bond in the protein can be between the main chain residues as well as the side chain residues.[8, 63, 64] The

hydrogen bond network can also run between the two different strands/subunits of protein. The hydrogen bond network or pattern in a protein induces the different types of secondary structures in protein.

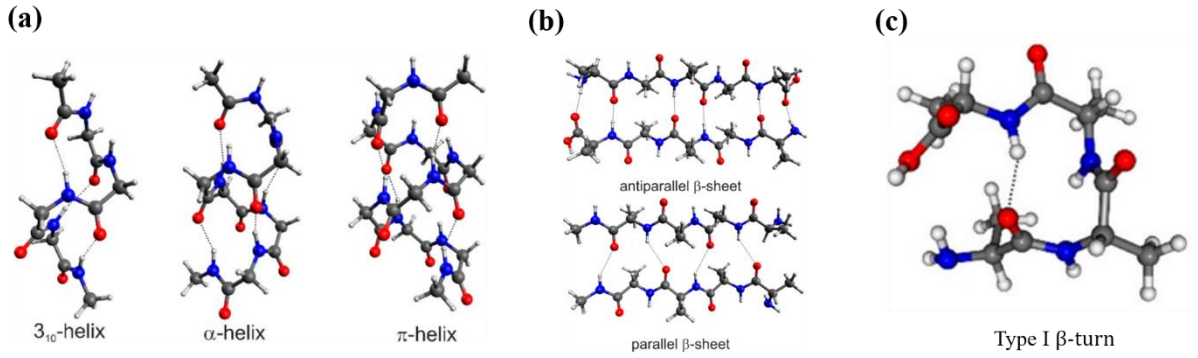

Figure 1.5. A schematic representation of helices (a), sheets (b), and turn (c). The figure has been adapted with permission from reference no. 8. Copyright [2020] American Chemical Society.

The most common and important secondary structures of proteins are helices, sheets, and turns.[65-69] The names for different secondary structures have been given based on the shape of that particular structure which is due to their distinctive hydrogen-bonding pattern. Helices are one of the important characteristics of the secondary structures of proteins where amino acid residues are arranged in a spiral shape. The contributing factor to the stabilization of the helices' secondary structure is the hydrogen bond between the amide hydrogen and carbonyl oxygen. The most common types of helices are α-helix, π-helix, and 3_{10} helices, which are described below. [70]

1.3 Helix secondary structures in protein

Helix is one of the most common secondary structures present in globular proteins.[71, 72] It can be found mostly in the membrane proteins and membrane-associated peptides. Helices can be further categorized into different types based on the hydrogen bond interactions between the main-chain carbonyl oxygen of residue i and the main chain amide hydrogen of residue n+1.

The α-helices are formed when there is a hydrogen bond formation between the i and i+4 amino acid residues.[72, 73] The α-helix structure repeats itself after a distance of 5.4 Å and has 3.6 amino acid residues per turn. Apart from the most common α-helix, the other helices such as 3_{10} and π-helices are also observed in the protein secondary structures.[74] 3_{10} and π-helices are formed when there is a hydrogen bond between the i→i+3 and i→i+5 residues, respectively. Most of the helices found in nature are right-hand helices. The right-handed α-helices are more favourably formed by the L-amino acids whereas the D-amino acids are preferable for the formation of the left-handed α-helices.[75] The Ramachandran dihedral angles Φ and Ψ for a right-handed α-helix and 3_{10} helices are −57°, −47°, and −49°, −26°, respectively.[76] Helical secondary structures are found mostly in relatively larger peptide molecules as it is difficult to attain the dihedral angle requirement in the smaller peptides.

1.4 Sheets secondary structure in protein

Sheets are one of the unique secondary structures where a segment or the whole set of residues of the polypeptide chain has an extended conformation and has hydrogen bonding interaction with other strands of protein having similar conformation.[77, 78] The sheets are formed when at least two extended β-strands are having the extended conformation. The β-strands are stabilized by the hydrogen bonds running along the two or more strands of the protein unit chain. The β-strands can be defined as the super secondary structure whose antiparallel strands are connected via a loop consisting mostly of the β-turn. The model of sheets was derived by Pauling and Corey for the fibrous protein known as beta-keratins.[30] Pauling and co-workers showed that the newly discovered configuration of the proteins can have a planar peptide group in the plane of a sheet.

The propensity of the β-sheets formation in proteins depends largely on certain amino acids present in the polypeptide backbone. The formation of the β-strands is most likely to occur

when hydrophilic and hydrophobic peptide side chains of amino acids are positioned on opposite sides of a main-chain peptide backbone. The β-sheet secondary structure of proteins gives the extended conformation of proteins in the condensed phase. The extended conformation of sheets in amyloid protein has a key role in Alzheimer's and Parkinson's diseases.[79, 80] The β-sheet structure has the elements of secondary as well as tertiary structures whereas helices have mostly the local conformation. The occurrence of the sheets in peptides was further probed through NMR spectroscopy by Raines and co-workers. [17]

1.5 Turns in protein

Turns in the proteins or peptides are important secondary structures as they allow the reversal of the direction of the polypeptide chains to provide folded structures of the proteins.[12, 14, 70] Thus, the turns, which have unique chain reversal properties, can be observed in a large number of proteins.[81] Turns or loops in proteins are generally found at the junction of different secondary structures like sheets and helices. Turns in proteins are a very important local region where many biological processes such as hydroxylation, enzymatic activity, etc. occur.[14] It is estimated that at least one-third of the globular protein has turns at one of the local sites.[82] Turns are mostly located at the protein surface containing polar and charged residues. Turns in proteins mostly occur when there is a hydrogen bonding between two different amino acid residues of the polypeptide backbone although there is a weak possibility of intra-residue hydrogen bond formation. Turns can be further categorized into δ, γ, β, α, and π resulting in the formation of C6, C7, C10, C13, and C15 membered hydrogen-bonded rings when hydrogen bond formation takes place between the N−H and C=O groups of i → i + 1, i → i+ 2, i→ i+ 3, i → i + 4, and i → i+ 5 amino acid residues, respectively.[12, 16] The relatively unexplored and weaker turn is formed when there is hydrogen bonding in a single amino acid residue resulting in the formation of a five-membered ring.[17] This type of hydrogen bonding has been named the intra-residue C5 hydrogen bond.

Turns were first identified by Venkatachalam in 1968 when it was found that there is hydrogen bonding between the carbonyl oxygen of residue i and amide hydrogen of the i+3 residue.[83] β-turn is the most important and abundant one out of all the possible types of tight turns found in peptides and proteins. It has been shown by Raghava and co-workers that almost 9% of all the amino acids in the BT6376 datasheet exist in β-turn conformation.[84] Upon inspection of 1397857 amino acids in the BT6376 database, 126016 β-turns were found. It was further concluded from their study that glycine, proline, asparagine, and aspartic acid are favored for the formation of the β-turn in the peptides and proteins and hence, those can be called the β-turn formers. Glycine with a flexible backbone and proline with the ability to break ordered structures make them suitable for β-turn formation in peptides and proteins. Asparagine and Aspartic acid have a side chain that can form hydrogen bonds with the main chain of the polypeptide chain. Apart from the β-turn inducer amino acids such as glycine, proline, asparagine, and aspartic acid, there are some other amino acids that act as β-turn breakers.[14] The β-turn breakers include the amino acids namely isoleucine, leucine, valine, and methionine that have hydrophobic nature in them. It was further revealed that β-turn favors the glycine and proline residues at the first and fourth positions.[14]

1.6 Standard types of β-turn

The β-turn induces the formation of loop segments and hairpin strands in the secondary structures of peptides and proteins. It has been well established that β-turn has a key role in mediating interactions between peptide ligands and their receptors.[85]

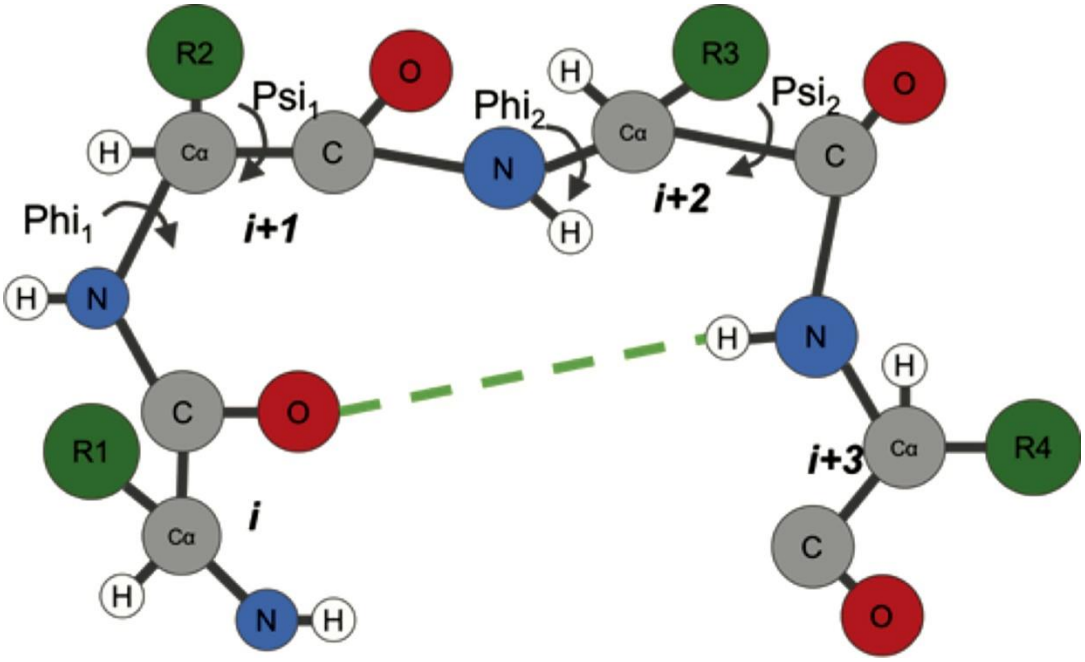

Figure 1.6. A schematic representation of β-turn in peptides. The figure has been adapted form reference no. 86. Copyright [2014] John Wiley and Sons. The different types of β-turn can be categorized by the dihedral angles of the polypeptide backbone.

There are nine types of β-turn observed in protein namely, Type I, Type II, Type I' and Type II', IV, VIII, VIb, VIa1, VIa2.[86] All these different types of β-turn are categorized based on the Ramachandran dihedral angles (Φ, Ψ). The most common type of β-turn is Type I, Type II, Type I' and Type II'. It has been understood now that the consecutive existence of type II β-turns will lead to the formation of 3_{10} helices in peptides. The Ramachandran angle parameters for the different types of β-turn have been shown in figure 8. It has been suggested that the introduction of proline residue increases the chances of formation of β-turn in cyclic peptides.

1.7 Intra-residue C5 hydrogen bond

Although most of the hydrogen bonds found in the proteins and peptides are inter-residue in nature, one of the little-known but equally important intra-residue named the "C5 hydrogen bond" has also been found to stabilize the secondary structure of the protein and peptides. The C5 hydrogen bond is most commonly observed in the extended β-sheet type of secondary

structures of proteins.[17, 29] The C5 hydrogen bond is intra-residue in nature as the lone pair from the carbonyl group and amine hydrogen of the same residue/amino acid are involved in the hydrogen bonding. Although relatively weak compared to other well-studied hydrogen bonds, C5 has been shown to stabilize almost 5% of all the residues in proteins. One of the early findings on the existence of this type of hydrogen bond was reported in the solution phase by R. A. Nyquist.[18] In that work, it was observed that the trans conformer of N-alkyl α-halo/alkoxy acetamides is more stable compared to its cis counterpart in the solution phase. With the help of the IR spectroscopy measurements, Nyquist concluded that the trans conformation of α-substituted secondary acetamides was more stabilized in dilute CCl_4 solution due to the formation of the C5 hydrogen bond.

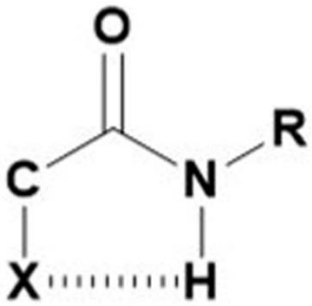

Figure 1.7. Trans conformer of N-alkyl α-halo/alkoxy acetamides as studied by R. A. Nyquist through solution phase IR spectroscopy. The figure has been adapted with permission from reference no. 18. Copyright [2014] Elsevier.

In conventional inter-residue hydrogen bonds in the peptide backbone such as π, α, β, δ, and γ turns, the s-type lone pair electrons of carbonyl group oxygen interact with the σ* orbital of the amide NH group. However, the C5 hydrogen bond deals with the interaction of the p-type lone pair of the C=O group corresponding to a particular amino acid residue with the σ* orbital of the amide NH group of the same residue.[17] As the C5 hydrogen bond is formed in the same residue of amino acid, it is highly constrained in nature.

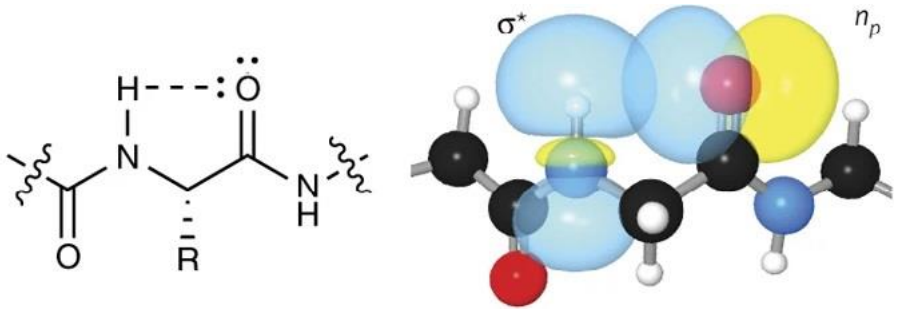

Figure 1.8. The Natural bond orbital (NBO) analysis of model peptide chain for the demonstration of intra-residue C5 hydrogen bond. It suggests that n_p lone pair orbital can form a hydrogen bond with the carbonyl group of the same residue. The Figure has been adapted in parts with permission from reference no. 17. Copyright [2016] Springer Nature.

1.8 Intra-residue C5 hydrogen bond in solution

The existence of the C5 hydrogen bond was further probed by studying model dipeptides using solution-phase FTIR spectroscopy. Neel and co-workers studied conformational preferences of dipeptides in CCl_4 solvent using IR spectroscopy.[87] They observed that the origin of the vibrational band around 3420 cm^{-1} is not due to the free amide frequency but it is due to the formation of an intra-residue C5 hydrogen bond by the amide N-H group. These extended conformations adopted by model dipeptides in the CCl_4 solvent were further explored by Scheraga and co-workers.[21] Maxfield and co-workers studied the IR vibrational frequencies of N-Acety1-N'-methylamide derivatives of glycine, L-alanine, and L-leucine in dilute solutions of chloroform and carbon tetrachloride. They confirmed the presence of C5 hydrogen bonds by monitoring the frequency of the N-H band in chloroform solvent and concluded the existence of the extended conformation in the solution.[88] Scheraga and co-workers also used NMR spectroscopy to study the conformational preferences of blocked gly-pro sequence in CD_2Cl_2 and DMSO-d_6 solution. They concluded from their study that the peptide with gly-pro sequence might exist in an extended C5 conformation. Balaram et. al. further studied Z-Aib-Ala-OMe and Z-Aib-Aib-OMe peptides in chloroform solvent.[89] They reported that both of

these dipeptides have extended conformation having an intra-residue C5 hydrogen bond.[90] The C5 interactions were further observed in model β-hairpin systems TrpZip2 through NMR and CD spectroscopy. The presence of the C5 interaction was also demonstrated by studying small peptides using Density functional theory and NMR spectroscopy by Steve Scheiner.[91]

1.8.1 Intra-residue C5 hydrogen bond in Protein Database

Raines and co-workers have done an extensive search to find the presence of the C5 hydrogen bond in proteins through Protein Data Bank (PDB) analysis.[17] They found that the C5 hydrogen bond exists in almost 94% of the proteins in the PDB. It was found that the C5 hydrogen bond was mostly present in the β-sheet structures of proteins. To verify the inherent existence of the C5 hydrogen bond, Raines and co-workers studied a few model peptides using IR and NMR spectroscopy combined with quantum chemical calculations. It has been found that the intra-residue C5 hydrogen bond is indeed very important for the overall stabilization of these peptides. They further found through natural bond orbital (NBO) analysis that the C5 hydrogen bond is very weak i.e. the interaction energy is ~0.25 kcal/mol.

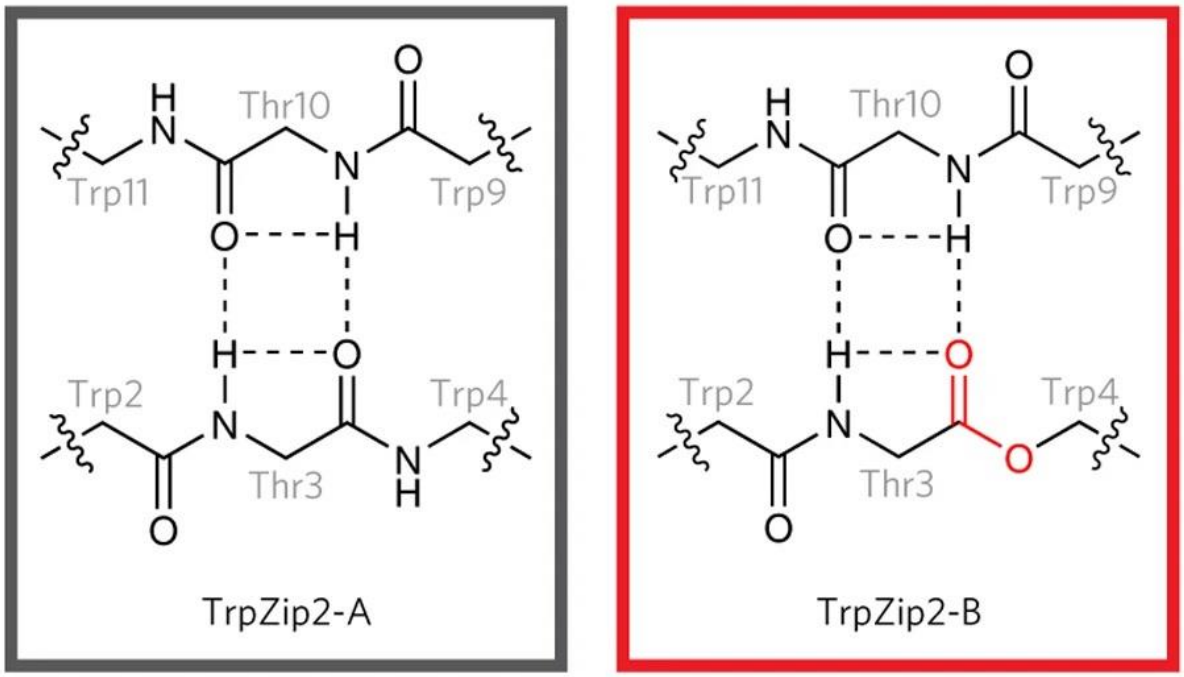

Figure 1.9. The intra-residue C5 hydrogen bond in proteins. The Figure has been adapted in parts with permission from reference no. 17. Copyright [2016] Springer Nature.

1.8.2 Intra-residue C5 hydrogen bond in the gas phase

As we know that the C5 hydrogen bond is very weak and rarely studied in the literature, a gas-phase study in isolated conditions can reveal the intrinsic nature of this non-covalent interaction compared to what has been reported in the condensed phase studies. One of the early reports of the C5 hydrogen bond in the gas phase was reported in tryptophan-containing capped peptides by Zwier and co-workers. [26] Very recently, new insights about the C5 hydrogen bond have been obtained from isolated gas-phase spectroscopic studies of small peptides along with solution-phase studies.[1, 8, 20, 92] It has been observed that the most commonly occurring amino acid residue associated with the C5 interaction is glycine.[1] The C5 interactions were indeed detected in glycine-containing peptides such as Z-Gly-OH, Gly-Gly, and Ac-Gly-Phe-NH_2 using several gas-phase laser spectroscopic techniques combined with the quantum chemical calculations.

1.9 sequence-dependent folding of peptides/proteins

The functionalities and properties of a protein/peptide are not only determined by the secondary and tertiary structure of proteins but also depend on the amino acid sequence of the strand.[93] It has been observed that selective residues of amino acids are very often involved in a particular biological phenomenon like enzymatic and hydroxylation processes.[14] The selective amino acid residues required for a specific property of proteins are determined by the Ramachandran angles (Φ, Ψ) adopted by the polypeptide backbone.[94] The Ramachandran plot uses the main chain's dihedral angles to predict the secondary structure of long residue peptides and proteins.

The Ramachandran plot can predict the feasibility of a particular stereochemistry in the polypeptide chain.

1.10 Sequence-dependent conformational studies of peptides in the solution phase

Scheraga and co-workers used the Empirical Conformational Energy Program for Peptides (ECEPP) to study the conformational preferences of X-Pro and Pro-X (where X = Ala, Asn, Asp, Gly, Leu, Phe, Ser, and Val, and of X-Pro, where X = Ala, Asn, Gly, and Pro) sequences. [95] They found from their extensive studies that the Pro-X sequences had more propensity to form a bend in polypeptide chains compared to that in the X-Pro sequences.[96] Further, they explored the conformational preferences of peptide sequences containing glycine and proline residues using NMR, IR, and CD spectroscopic techniques.[97] They concluded from their overall experimental results that the major conformation observed for the gly-pro sequence was β-strand or extended polyproline II type where the glycine residue forms an intra-residue C5 hydrogen bond along with a C7 hydrogen bond. On the other hand, the dominant conformation observed for the pro-gly sequence was β-turn. However, their results also indicated that other minor conformations such as C7 or γ-turn were also present for the pro-gly sequence.

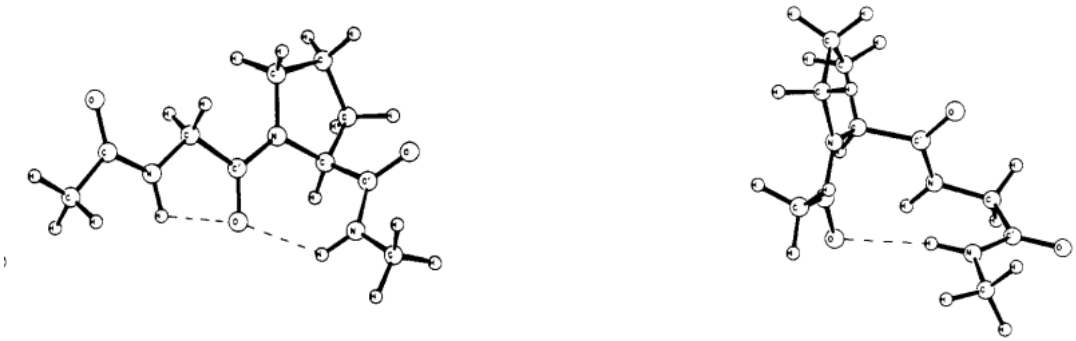

Figure 1.10. The figure represents the conformation adopted by the gly-pro (left) and pro-gly (right) sequence in solution phase conditions. The gly-pro sequence shows C5-C7 conformation whereas the pro-gly shows the C10 conformation in solution. The figure has been adapted with permission from reference 21. Copyright [1977] American Chemical Society.

Figure 1.10 shows a comparison of the major conformation adopted by the gly-pro and pro-gly sequence in the solution phase. The propensity of the Pro-Gly-X sequence to form a β-turn has been also observed in many crystal structures reported by Balaram and co-workers. [67, 98] The crystal structure of the peptide-containing D-Pro-Gly sequence shows a hairpin type of structure where the bend or β-turn occurs at the pro-gly position. The turn-inducing property of the Pro-Gly sequence was also reported from solution phase calculations by Park and Co-workers.[99] They used the SMD solvation method to study the conformational landscape of Ac-Pro-Gly-NHMe in water and CH_2Cl_2 solvents and found the β-turn conformation of the peptide for both the solvents.[99]

1.10.1 β-turn requirement for Hydroxylation in pro-collagen

It has been shown that β-turn conformation of the Pro-Gly sequence in procollagen is required for hydroxylation which is the driving force for the formation of the triple helical structure of collagen.[14] Prolyl hydroxylase is the main enzyme that helps in the hydroxylation of selected proline moiety of the procollagen. It was observed that selectively the proline residue in position three of the general collagen subchain or building block -Gly-R_2-R_3 was susceptible to hydroxylation by the enzyme.

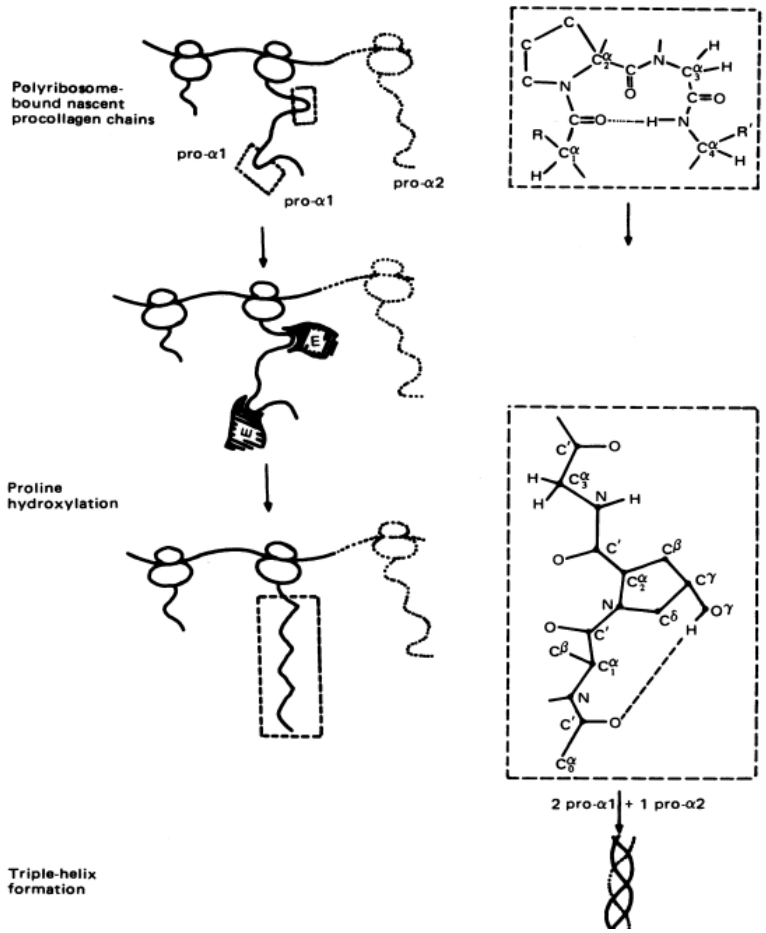

Figure 1.11. The figure represents the important steps in the synthesis of collagen from procollagen. The figure has been adapted with permission from reference 14. Copyright [1979] National Academy of Sciences (NAS).

1.11 Importance of D-Amino acids

The amino acids except glycine has at least one chiral carbon resulting in the formation of two different configuration L (Levo, left-handed) and D (Dextro, right-handed). D-amino acids, the enantiomers of L- amino counterparts show some very specific properties that are opposite to their counterparts. Although the D-amino acids show similar chemical and physical properties compared to their L-amino acid counterparts, they rotate the plane of polarised light in the opposite direction compared to their counterparts. As we know all the amino acids used for the

living organism's protein have L-configuration whereas D-amino acids are most prominent in the posttranslational modification of peptidoglycan cell walls of bacteria.[100] One of the most common applications of D-amino acids nowadays includes increasing the biostability of proteins for the making of hydrogels, protein inhibitors, and uses in drug delivery systems.[101] The driving force behind the existence of unusual left-handed α-helices has been attributed to the D-amino acids.

1.12 Stabilization of Pro-Gly-X sequence on the β-turn formation

Studies on the conformational preferences of Pro-Gly-X residues reveal that X residue has an important role in the stabilization of the β-turn formed by the sequence. It has been further identified that the order of stabilization of the β-turn is Leu>Ala>Gly, Ile, and Phe.[14] It was revealed through spectroscopic techniques such as X-ray, circular dichroism, Infrared spectroscopy, and NMR that Ac-Pro-Gly-X-OH (X=Gly, Ala, Leu, Ile, and Phe) indeed show β-turn type of conformation.[22] The statistical analysis of crystal structures by Chou and Fasman showed that in a four residue amino acid chain proline in the second position and glycine in the third position are preferred. It can be further concluded that X residue affects the β-turn stabilization in Pro-Gly-X residue.

1.13 Importance of Isolated Gas-Phase studies of Small peptides.

Gas-phase studies of small peptides are of great importance as they reveal intrinsic conformational preferences of specific peptide sequences in absence of any solvent molecules or any other interactions.[9, 23] Moreover, it is possible to probe multiple low-energy conformations of the peptides apart from the global minimum in isolated gas phase laser spectroscopy study using supersonic expansion technique. Very often, it is important to unravel the structures of several low-energy conformations of the peptides which are slightly higher in energy than the global minimum as very little change in the structure of the global minimum

can lead to those higher energy structures. On the other hand, solution-phase studies can provide either only the global minimum structure or an average structure of several low-energy conformations of a peptide.[20] It has been possible to study the peptides in the gas phase without any fragmentation due to the combination of gas-phase jet-cooled laser spectroscopy study with laser desorption technique which has been described in detail in chapter 2. Brief detail about the work done by various research groups to understand the conformational preferences of biomolecules such as amino acids, peptides, and DNA base-pair has been presented here.

The generation of the supersonic molecular beam of tryptophan by Levy et. al paved the way to study the conformation of small bio-molecules in isolated gas-phase conditions.[24, 102, 103] They used the thermospray jet and supersonically expanded helium molecules to generate the molecular beam of tryptophan. The mass spectrum of tryptophan was recorded by the nonresonant two-photon ionization technique with the use of the second-harmonic output of an Nd: YAG pumped pulsed laser. The electronic spectrum of tryptophan was obtained by the Resonantly enhanced two-photon ionization spectroscopy technique (R2PI) in the argon carrier gas.[24] They were further able to identify different conformers contributing toward the electronic spectrum of tryptophan in the gas phase by selectively exciting the different ground state conformers and recording their dispersed fluorescence.[103] The techniques such as resonantly enhanced two-photon ionization spectroscopy (R2PI) and laser-induced fluorescence were then used to record the electronic spectra of tryptophan analogs in a supersonic molecular beam. The dispersed fluorescence spectra of tryptophan analogs such as 3-indole acetic acid, 3-indole propionic acid, tryptamine, and N-acetyltryptophan ethyl ester revealed the existence of different conformation in isolated gas-phase conditions.[104] The incorporation of conformer selective infrared-ultraviolet double resonance spectroscopy by Lee and co-workers revealed the local modes in benzene and benzene dimer.[105, 106] The conformational selective UV-UV hole burning or UV-UV double resonance spectroscopy was further incorporated in gas-phase

experiments to identify the different conformation of molecules.[107] The infra-red spectra of benzene-water clusters in the gas phase were recorded by using the IR-UV double resonance technique.[108]

One of the most common methods to bring the molecules into the gas phase was thermal heating of sample molecules before the supersonic expansion. The molecules such as peptides that have low vapour pressure and thermal stability require a stable and efficient method named laser desorption for the generation of vapours of peptides.[109, 110] In this method, the peptide was pressed into a solid pellet system for absorption of peptide and then an IR laser was directed onto the pellet to bring the peptide molecule into the gaseous phase. A detailed understanding of the laser desorption method has been discussed in chapter two of the present thesis.

Initially, gas-phase UV and IR spectroscopy were recorded for the uncapped amino acid having a chromophore group such as tryptophan phenylalanine and tyrosine.[24, 104, 111]The conformational landscape of such uncapped amino acids was heavily determined by their $-COOH$ and $-NH_2$ end group and thus had a very limited implication in large biological molecules such as proteins and peptides.[8] The gas-phase study of capped tryptophan residues N-acetyl tryptophan amide and N-acetyl tryptophan methyl amide by Zwier and co-workers revealed the existence of γ-turn in model peptides.[26] The gas-phase studies of these model peptides were compared with the DFT calculations performed at the B3LYP/6-31+G(d) level of theory GAUSSIAN 98 suite of programs. The infra-red spectra showed that N-acetyl tryptophan amide (NATA) existed in two conformations in the gas phase, one having C5 conformation while the other having $C7_{eq}$ conformation. The infra-red spectra of N-acetyl tryptophan methyl amide in the gas phase indicate the existence of two conformers having C5 conformation and one having C7eq conformation. Zwier and co-workers extended the study on the conformational landscape of larger capped peptide systems in the gas phase with the help of DFT calculations.[112, 113] The Gerhards and co-workers explored the conformational

landscape of the peptide Ac-Val-Phe-OMe and other phenylalanine derivatives in the gas phase revealing the preference of an extended β-sheet conformation.[114]

The conformational details of DNA bases, base pairs, and nucleosides in the isolated gas phase were brought to light by M. de Vries and co-workers. [25, 115, 116] The study revealed the structural details of rotamers of a base such as Adenine, guanine, and cytosine as well as the base pair including guanine-cytosine, guanine-guanine, and cytosine-cytosine in the gas phase. They further revealed the different folding motifs in different components of gramicidin Peptides in isolated gas-phase conditions.[117] It was recently highlighted by Rijs and co-workers through IR multiple-photon dissociation vacuum–ultraviolet (IRMPD-VUV) that the gly-gly sequence exists in a nearly planar structure.[118] Biologically important molecules such as synephrine and other peptide sequences were studied in gas-phase by Fujii and co-workers.[119-121]

Mons and co-workers revealed the conformational preferences of model tripeptides N-acetyl-Phe-Pro-NH_2 and N-acetyl-Pro-Phe-NH_2 in the gas phase for the first time with the help of gas-phase laser spectroscopy and DFT calculations.[122] Studies revealed that the Phe-Pro sequence preferentially adopts β-turn conformation whereas the Pro-Phe sequence prefers double γ-turn structures. They further showed that the capped tripeptide system Ac-Phe-Gly-Gly-NH_2 exists in a double β-turn conformation in the gas phase. [123]

1.14 Aim of the thesis

Secondary structures of proteins and peptides play an important role in the formation of their unique tertiary and quaternary structures and functions. The secondary structures are governed mostly by the hydrogen bonding interactions present along the polypeptide backbone while the sequence of the amino acids also contributes to the overall structures. Conformations of the secondary structures of proteins and peptides are generally elucidated by, XRD, and Cryogenic electron microscopy in the condensed phase.[67, 89, 124, 125] However, the secondary structures,

being local, can be studied in small peptides in absence of other neighbouring residues as well as surrounding water molecules. Isolated gas-phase spectroscopy techniques combined with quantum chemical calculations enable one to study the inherent folding motifs of multiple low-energy conformations of small peptides with various amino acid residues.

In this thesis, we have tried to explore the folding motifs of the low energy conformations of a few selected peptides in the isolated gas phase as well as the condensed phase in combination with quantum chemistry calculations. In this thesis, we have chosen the Z-Gly-Pro-OH molecule as a model peptide to study the C5 hydrogen bond in an isolated gas-phase condition. The protecting group Z (benzyloxycarbonyl) also acts as a chromophore for the UV laser to do the gas phase spectroscopic studies. We have shown that the C5 hydrogen bond indeed stabilizes the most stable conformer of Z-Gly-Pro-OH in isolated gas-phase conditions. The non-covalent interactions present in the observed conformers for Z-Gly-Pro-OH have been validated by Natural Bond Orbital (NBO) analysis and Non-Covalent Interaction (NCI) analysis. The peptides studied in this thesis include Z-Gly-Pro-OH, Boc-Gly-D-Pro-NHBn-OMe, Boc-D-Pro-Gly-NHBn-OMe, and Boc-D-Pro-Gly-Ala-NHBn-OMe. In the first working chapter, we have shown the importance of the intra-residue C5 hydrogen bond in the stabilization of small dipeptide molecules. The second working chapter of the thesis deals with the study of the sequence dependence folding motifs of the secondary structures of small peptides. In particular, the conformational preferences of the gly-pro and pro-gly sequences have been revealed in the isolated gas phase and condensed phase. From the overall study of these two peptides, it has been found that the gly-pro sequence has the propensity to form an extended conformation while the pro-gly sequence forms C10 (β-turn) and C7 conformations in the solution and gas phase, respectively. In the last working chapter, we studied the effect of the neighbouring residue (X) on the stabilization of the β-turn in the Pro-Gly-X sequence of a

peptide chain from the solution phase study. The brief objectives of the study of the secondary structures of these small peptides are as follows.

I. Importance of a relatively weak C5 hydrogen bond in the stability of peptides.

It is known that the stability of the secondary structures of proteins and peptides is mostly due to the hydrogen bonding interactions between the N-H and C=O groups of two neighbouring residues in the polypeptide backbone. However, intra-residue hydrogen bonds between the carbonyl oxygen and amide hydrogen the of same amino acid residue have been underestimated for a long time and have not been explored very much in the literature. The intra-residue C5 hydrogen bond has been shown to exist in extended β-sheet conformations of proteins and peptides.[17] We have studied the conformational landscape preferred by Z-Gly-Pro-OH peptide in an isolated gas phase with the help of IR/UV double resonance spectroscopy and quantum chemical calculations. The results of our experiments on this dipeptide reveal that the most stable conformer of Z-Gly-Pro-OH in the isolated gas-phase condition is stabilized exclusively by the intra-residue C5 hydrogen bond. The C5 hydrogen bond in Z-Gly-Pro-OH enables this peptide to have an extended type of conformation. The other conformer adopted by the Z-Gly-Pro-OH molecule in isolated gas-phase conditions has a folded conformation where there is a hydrogen bond between the hydroxyl group of proline residue with the benzene ring of the protecting group carboxybenzyl (Z). This work demonstrates that a single weak C5 hydrogen bond can direct a dipeptide towards the global conformation despite the presence of other conformations which are stabilized due to a relatively stronger hydrogen bond.

II. Sequence-dependent folding of peptides.

It is known that particular secondary structures of a polypeptide chain not only depend on the hydrogen bonding interactions in the polypeptide backbone but also on the specific amino acid residues and their sequence. There have been reports of a completely different propensity of

the secondary structures of the polypeptide backbone based on a particular sequence of specific amino acid residues and its reverse sequence. The conformation adopted by the Pro-Gly-X sequence has been shown to have a preference to form a β-turn or bend type of structure in polypeptides and proteins.[14, 22, 83] On the other hand, the Gly-Pro-X sequence in polypeptides and proteins has a propensity for extended conformation.[97] Numerous reports on solution-phase spectroscopic and quantum chemical calculations of the Gly-Pro and Pro-Gly peptides demonstrate the folding motifs observed in proteins and polypeptides.[22, 97, 99] Requirement of the β-turn conformation at the Pro-Gly position of the procollagen for selective hydroxylation of the Pro residue for the formation of collagen. However, there was no report of gas-phase spectroscopic studies of peptides containing the gly-pro and pro-gly sequences.

In this thesis, we have explored the conformation preferences of capped dipeptides containing the gly-pro and pro-gly sequences using isolated gas phase as well as condensed phase spectroscopic techniques. In our work, we have used gas-phase double resonance spectroscopy combined with quantum chemical calculations to determine the structures of a few low-energy conformations of both the dipeptides. The solution phase conformational preferences of peptides were predicted with the help of solution-phase NMR spectroscopy and FT-IR spectroscopy in $CDCl_3$. The solid-state structure of the gly-pro peptide was determined by the X-ray crystallography technique. The gas-phase studies reveal that the pro-gly sequence has a C7-C7 conformation whereas the gly-pro sequence has an extended conformation with a C5 hydrogen bond. The solution phase studies suggest that the gly-pro peptide adopts an extended β-strand conformation whereas the pro-gly sequence prefers the folded β-turn conformation in the $CDCl_3$ solution. The extended C5 structure is also observed in the crystal structure of the gly-pro peptide.

III. Effect of the amino acid residue X on the stabilization of the β-turn in the Pro-Gly-X sequence of a polypeptide backbone

The importance of the β-turn conformation has been discussed above in the hydroxylation of the procollagen protein and reversal of the direction of the polypeptide chain towards the folding of proteins. It is reported that some specific amino acid residues have a very high propensity to form a β-turn or β-bend conformation in the polypeptide backbone of proteins and peptides.[14] The propensity to form β-turn or bend type of conformation has been frequently discussed in the case of the Pro-Gly-X sequence in proteins and peptides. The β-turn or bend conformation stability also depends on the X amino acid residue attached at the end of the sequence. The common trend for the β-turn or bend stabilization by the X amino acid residue can be given as follows, Leu>Ala>Gly, Ile, and Phe. We have studied here the effect of the Ala residue on the strength of the β-turn structure of the end-protected Pro-Gly-Ala tripeptide by FT-IR, 2D-NMR spectroscopy, and X-ray crystallography.

Chapter 2

Experimental and Computational Methods

In this chapter, we have provided a detailed description of our gas-phase experimental setup, laser-based various spectroscopic techniques, and various computational methods used for performing the research. A brief description of various spectroscopic methods used for solution-phase spectroscopy is also provided here.

2.1 Experimental Methods

2.1.1 Gas-phase Laser Spectroscopy

The gas-phase laser spectroscopy experiments for the determination of the structures of various low-energy conformers of several small peptides have been performed using home-built jet-cooled laser desorption resonantly enhanced multiphoton ionization (REMPI) Time of Flight mass spectrometer.[102]

2.1.1.1 Supersonic expansion technique

The use of the supersonic expansion technique for experiments was pioneered by Levy and co-workers.[24, 102, 126-128] The supersonic expansion technique allows the gaseous molecules to undergo cooling of internal degrees of freedom i.e. translation, rotational, and vibrational. Vibrational and rotational cooling is the key to obtaining very high-resolution electronic spectra of large polyatomic molecules. The supersonic expansion of gaseous molecules creates a collision-free isolated low-density molecular beam. A detailed description of this technique has been discussed in several reviews, books, and articles.[8, 9, 23, 129] A brief description of the supersonic expansion technique has been provided here.

The supersonic molecular beam of samples is produced when sample vapor seeded in inert carrier gases (i.e. He, Ar, Ne, etc.) is expanded from high-pressure to a vacuum chamber through a small orifice. The only condition required to generate the supersonic molecular beam is to have the orifice diameter (D) larger than the mean free path of the gaseous molecules. However, if the mean free path (λ_0) becomes larger than the diameter of the orifice (D), the beam becomes effusive. As the gaseous molecules pass through the pulse valve orifice, they

undergo enormous collisions within a few mm distance from the orifice. Numerous collisions near the orifice result in the conversion of the enthalpy associated with the random motion of the molecules into a directed mass flow. The supersonic expansion cools down the translational degrees of freedom and then the cold translational bath acts as a refrigerant for the other degrees of freedom i.e., rotational and vibrational. This process allows the gas mixture to have a very low translational, rotational and vibrational temperature. The translational temperature of the gas molecules can be expressed in terms of the speed of sound (a) according to the following equation $a=(\gamma k_B T/m)^{1/2}$ where γ is the heat capacity ratio defined as Cp/Cv. The other parameters m, k_B, and T are the mass, Boltzmann constant, and temperature, respectively. The equation suggests that the speed of sound decreases as the temperature of the molecules decreases. The supersonic or effusive nature of the molecular beam is expressed by a quantity called Mach number (M), which is defined as the ratio of the velocity of the molecules in the beam (u) and the speed of sound (a) i.e, M=u/a. The Mach number is greater than one for the supersonic molecular beam and this implies that the velocity of the molecules is greater than the speed of sound. A comparison between the velocity distribution of molecules in the high-pressure reservoir and the supersonic beam has been shown in figure 2.1.

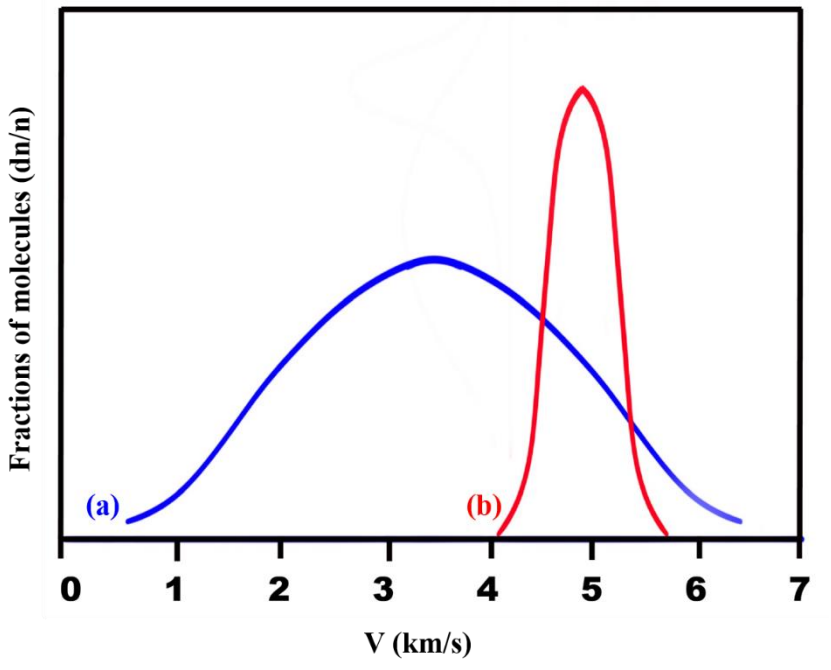

Figure 2.1. A schematic representation of normalized velocity distribution curves for the molecules in (a) high-pressure reservoir and (b) supersonic beam.

The supersonic beam has a significantly narrow velocity distribution along with a shift of peak maxima towards a higher value as the flow velocity of the molecules increases. The narrow velocity distribution of supersonically expanded molecules implies that the translational temperature of the molecules is significantly lower (~0.5-5 K). The rotational cooling is higher than the vibrational cooling as equilibration between the rotational and translational degrees of freedom is substantially faster than that between the vibrational and translational degrees of freedom. After a certain distance from the pulse valve nozzle, a non-equilibrium state is achieved where there are no more collisions. The supersonic expansion cools the rotational and vibrational degrees of freedom up to 5-10 K and 20-50 K, respectively.

The supersonic expansion is an isentropic process under adiabatic reversible flow conditions. The correlation between the temperature, pressure and density of a supersonic molecular beam can be defined as follows on the assumption that the expanding gas is ideal.

$$\frac{\boldsymbol{T}}{\boldsymbol{T_0}} = \left(\frac{\boldsymbol{P}}{\boldsymbol{P_0}}\right)^{(\boldsymbol{\gamma-1})/\boldsymbol{\gamma}} = \left(\frac{\boldsymbol{\rho}}{\boldsymbol{\rho_0}}\right)^{\boldsymbol{\gamma-1}} = \frac{\mathbf{1}}{\mathbf{1}+\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\gamma-1})\boldsymbol{M^2}} \qquad \text{…….2.1}$$

Where, T_0, P_0, and ρ_0 are the temperature, pressure, and density of molecules in the high-pressure reservoir, and T, P, and ρ are the temperature, pressure, and density of molecules in the supersonic molecular beam. γ is the heat capacity ratio (C_p/C_v) and M is the Mach number. Considering the expanding molecular beam as a continuous medium, the Mach number at a distance (X) from the nozzle can be described by equation 2.2.

$M=A(X/D)^{\gamma-1}$……(2.2)

Here, X is the distance from the nozzle and D is the diameter of the nozzle while A is a constant depending on γ (γ=3.26 for monoatomic gases). This equation suggests that the Mach number increases with the distance (X) from the nozzle. However, Fenn and Anderson have shown that

the Mach number can increase up to a certain finite value as the gas flow is not continuous and gas molecules are discrete particles in reality. As the distance from the orifice increases, there is less number of collisions due to the reduction of molecular density. The terminal Mach number of molecules can be described by the following equation as derived by Fenn and Anderson.[130]

$M_T = 2.05\epsilon^{-(1-\gamma)/\gamma}\ (\lambda_0/D)^{(1-\gamma)/\gamma}$ …….(2.3)

$=133(P_0D)^{0.4}$ for Argon

Here, M_T represents the terminal Mach number, ϵ is the collisional effectiveness constant, *D* is the orifice diameter, λ_0 is the mean free path of molecules in the reservoir and γ is the ratio of heat capacity (C_p/C_v). The equation implies that the terminal Mach number depends on the diameter of the orifice and the pressure of the reservoir. P_0D can be defined as the total number of possible collisions of the molecule before reaching a given Mach number. It implies that the Mach number depends on the total number of collisions a molecule experiences during the supersonic expansion.

2.1.1.2 Time of Flight mass spectrometry

Time of Flight mass spectrometry (TOF-MS) is coupled with supersonic expansion to obtain mass-selected high-resolution electronic and vibrational spectra of molecules and weakly bound complexes. The cold molecular beam produced from supersonic expansion is ionized by a tunable UV laser through the resonantly enhanced multiphoton ionized (REMPI) technique. The TOF-MS setup consists of a field-free TOF tube, three-electrode plates, and a microchannel plate (MCP) detector. Ions are directed towards the MCP ion detector by applying potential on three-electrode designs pioneered by Wiley and McLaren and then allowed to move in a field-free region.[131] The ions with different mass-to-charge ratios travel at different speeds in the TOF tube and hence reach the MCP detector at different times. Ions having a lower mass-to-charge ratio reach the detector earlier than those with a larger mass-to-

charge ratio. The first, middle, and third electrodes located at the beginning part of the TOF-MS setup (Figure 2.2) are called repeller plate, extraction grid, and accelerating grid, respectively. The repeller plate has a higher positive electric field than the extraction grid. The ions are generated at the center of the repeller plate and extraction grid through REMPI spectroscopy. The constant electric field (E_s) between the repeller plate and the extraction grid directs the ions towards the extraction grid. The higher electric field (E_d) between the extraction grid and accelerating grid accelerates the ions molecule into the TOF tube towards the MCP detector.

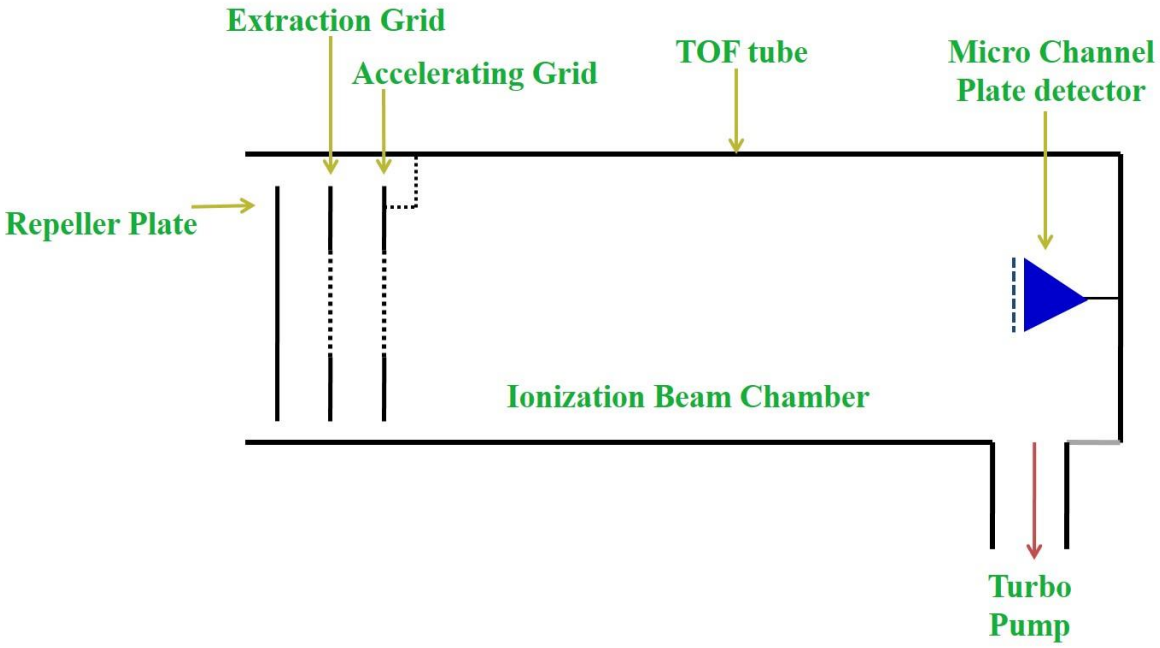

Figure 2.2. A schematic diagram of Time of Flight mass spectrometer assembly which is coupled with supersonic expansion setup.

A schematic representation of TOF mass spectrometer assembly has been shown in figure 2.2. As the ions are generated in the region between the repeller plate and extraction grid, they experience different velocities according to their position. The ions closer to the repeller plate experiences higher acceleration compared to the ions closer to the extraction grid. This can

impact the resolution of the TOF mass spectrometer as the ions with the same charge-to-mass ratio will reach the detector at different times. The adjustment of the two field sources E_d/E_s between the plates allows the ions to move from the same position into the field-free region of the TOF tube. If U_0 is the initial energy of the ions, then it is increased to U when it experiences an electric field Es between the repeller plate and the extraction grid and an electric field E_d between the extraction grid and the accelerating grid.

$U = U_0 + qsE_s + qdE_d$, ……2.1

Total time of flight (T) for an ion with mass and charge ratio as m/q can be given as

$T = T_s + T_d + T_D$, ……2.2

where T_s can be defined as the time recquired by the ion source to pass the ionization region,

$T_s = 1.02\frac{(2m)^{1/2}}{qE_d}[(U_0+qsE_s)^{1/2} \pm (U_0)^{1/2}]$ ……2.3

T_d is the time required by the ion source to pass the acceleration region,[131]

$T_d = 1.02\frac{(2m)^{1/2}}{qE_d}[U^{1/2}-(U_0+qsE_s)^{1/2}]$ ……..2.4

and T_D is the time reacquired by ion beams to travel field-free drift tubes.

$T_D = 1.02(2m)^{\frac{1}{2}}(\frac{D}{2U})^{\frac{1}{2}}$ … … ….2.5

In the gas-phase experimental set-up, the TOF mass spectrometer module is placed at a mutually perpendicular position to the molecular beam axis, and the direction of the UV/IR laser beam. The molecular beam intersects the UV laser beam at the center position between the repeller plate and extraction grid. The voltages applied on the repeller plate and extraction grid are +3060 V and +2780 volt, respectively, while the accelerating grid is grounded. The potential difference between the repeller plate and extraction grid directs the ions into the 1m long TOF tube. The positively charged ions arrive at the negatively charged MCP detector (-3060V) placed at the end of the TOF tube. The MCP detector consists of millions of micro-

channels that act as an electron multiplier. As any ion comes into contact with the microchannel plate, the generation of new electrons takes place from the channel wall. These electrons in turn generate secondary electrons upon hitting the walls of the micro-channel. This process generates millions of electrons at the rear end of the plate which are converted into electronic signals. The electronic signal from the MCP detector is visualized and recorded in the digital oscilloscope.

2.1.1.3 Experimental set-up

A schematic diagram of a home-built jet-cooled REMPI time of flight mass spectrometer coupled with a laser desorption assembly has been shown in figure 2.3. The gas-phase spectroscopy experiment is carried out in this setup consisting of two differentially pumped vacuum chambers called molecular beam chamber and ionization beam chamber. Both chambers are connected by a skimmer of 2 mm diameter. The molecular beam chamber contains a sample holder assembly connected with a pulse valve (General valve, series 9, 0.5 mm diameter and 10 Hz repetition rate). A pulse driver (IOTA ONE, Parker Instrumentation, USA) controls the opening time and pulse duration of the valve. A diffusion pump (OD 250, Hind Hivac) with a pumping speed of 2000 L/s pumps down the molecular beam chamber. The diffusion pump (OD 250, Hind Hivac) is backed up by a roughing pump (FD 60, Hind Hivac) with a pumping speed of 17 L/s. On the other hand, the ionization chamber is pumped separately by another diffusion pump (OD 114, Hind Hivac) with a pumping speed of 280 L/s and the diffusion pump is backed up by another roughing pump (ED-21, Hind Hivac) having a pumping speed of 6 L/s. The diffusion pumps are cooled down through a water chiller (Refricon Hivac system). The liquid nitrogen traps situated between the vacuum chambers and diffusion pumps prevent the contamination of the diffusion pump oil as well as assist to get a better vacuum. The pressure of the molecular and ionization beam chambers is measured by

cold cathode ion gauges attached to both chambers. The base pressure achieved in the molecular and ionization beam chambers is ~$1x10^{-7}$ and $3x10^{-7}$torr, respectively.

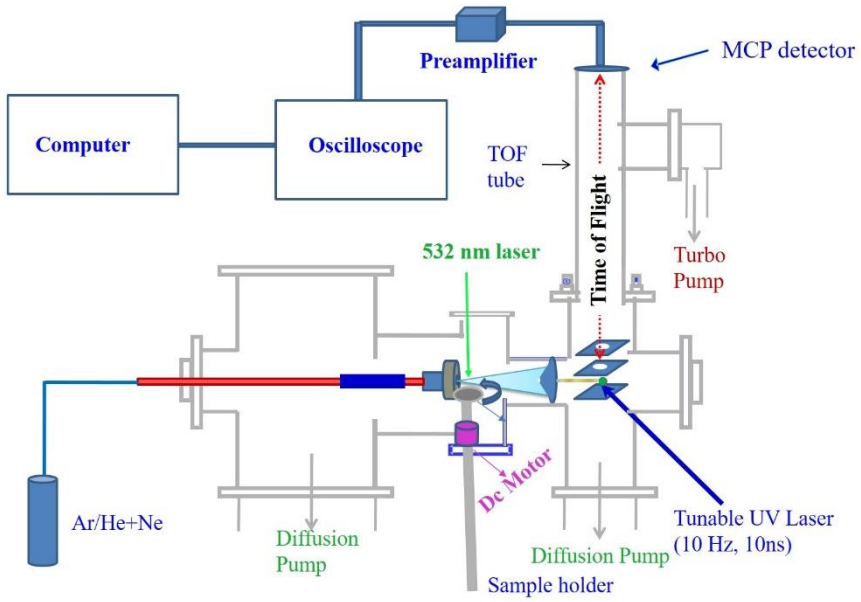

Figure 2.3. A schematic diagram of home built experimental set-up for carrying out the gas-phase experiments.

The time of the flight tube is connected with the ionization beam chamber in a perpendicular orientation to the molecular beam axis. The MCP detector situated at the end of the TOF tube is housed in a small vacuum chamber pumped by a small turbo molecular pump (V70, Varian) with a pumping speed of 70 L/s. The turbo-molecular pump is backed by a dry scroll pump (SH-110, Varian) with a pumping speed of 1.5 L/s. The detector chamber is separated from the TOF tube and ionization chamber by a gate valve. The voltages on the electrodes in the TOF tube and detector are controlled by a TOF power supply (Jordan TOF). The ion signal obtained from the detector is further amplified by a preamplifier (SRS, Model SR445A) and then sent to a digital oscilloscope (Tektronix, 350 MHz, DPO 4034). The signal from the oscilloscope is acquired to a computer via a USB port using LabView-based programs. A delay generator

(BNC, Model 575) is used to control the time delay adjustment between the pulsed valve, UV laser, and IR laser. As the molecular systems studied in this work are peptides, which are non-volatile, and are vaporized using the laser desorption technique.

2.1.1.4 Laser desorption technique

The laser desorption method is used for molecules having low vapor pressure and less thermal stability. The solid peptide sample is mixed with graphite powder and the mixture is pressed with 2-3 tons of pressure using a hydraulic press to make a pellet of 6 mm diameter and 2 mm thickness. The pellet is kept just after the pulse valve assembly in the molecular beam chamber. A schematic of the laser desorption assembly attached to the pulse valve is shown in figure 2.4. The pellet is mounted on the shaft of a DC motor fixed with a sample holder. The pellet is placed near the pulse valve in an orientation perpendicular to the face plate of the pulse valve. The DC motor is connected with a variable power source placed outside the vacuum chamber.

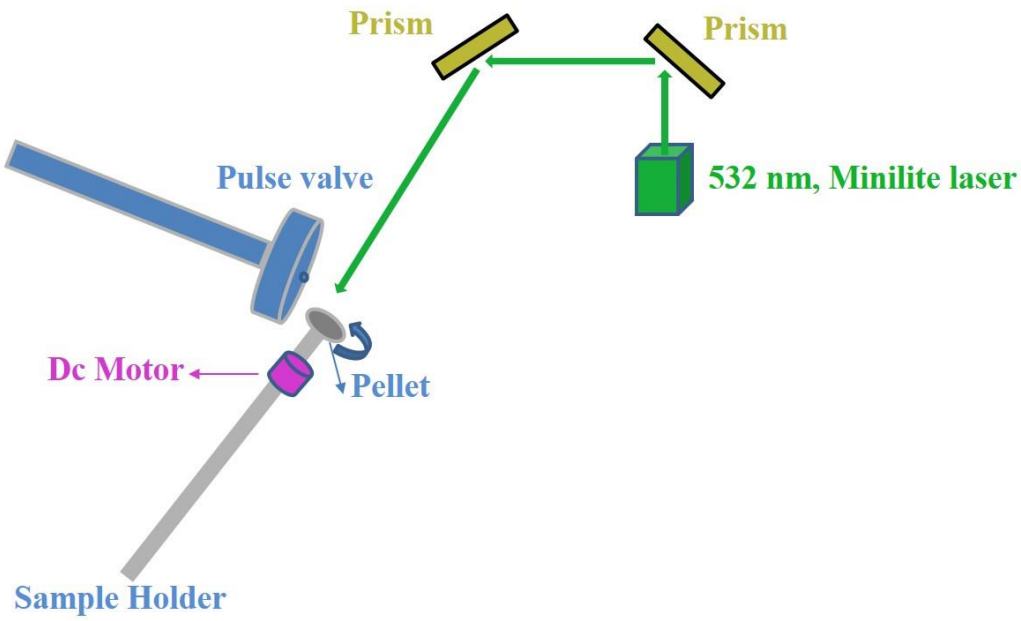

Figure 2.4. Schematic representation laser desorption technique used to bring the solid peptide molecules into the gas-phase.

The DC motor rotates the sample pellet so that a fresh part of the pellet is exposed to the desorption laser to prevent damage to the sample and reduce signal fluctuation. A 532 nm laser beam (0.7-0.8 mJ, Minilite-I, 10 ns, 10 Hz) is shone on the rotating pellet and the desorbed sample is entrained in the supersonic expansion of the Ar carrier gas (~ 70-80 psi) towards the ionization region. The desorption laser beam having a spot size of ~2 mm is focused onto the pellet. The distance between the pulse valve orifice and the edge of the sample pellet exposed to the laser beam is maintained at around 2 mm. The surface of the pellet is kept ~1.5-2 mm away from the molecular beam axis to optimize the amount of the desorption signal as well as supersonic cooling.

2.1.1.5 Laser systems used for gas-phase experiments

2.1.1.5.1 Dye Laser

Tunable UV output from the dye lasers is used to measure the electronic spectra of molecules and complexes in the gas phase. The dye laser uses organic dyes as active gain media for the generation of the tuneable UV-Vis laser output. The dye molecules should have good photochemical stability, strong fluorescence ability, and limited inter-system crossing for stable and tunable laser output. A schematic diagram for the dye laser used for the gas-phase experiments has been provided in figure 2.5.

Second harmonic output (532 nm) from an ND: YAG laser (Continuum, Surelite, 10 Hz rep. rate, 10 ns pulse width) is used to pump the dye laser. The dye laser contains an oscillator, preamplifier, and amplifier cells. The second harmonic output of 532 nm laser light is incident on a beam splitter (BS1). The beam splitter (BS1) reflects 5% of the 532 nm beam towards the oscillator cavity and thus 95% of the 532 nm beam gets transmitted through. The transmitted portion of the 532 nm laser light is directed toward another beam splitter (BS2). The beam splitter (BS2) reflects 10% of the beam towards the pre-amplifier while the transmitted beam is directed towards the final amplifier. The dye cell resonator/cavity contains two gratings

having 2400 lines/grooving, a beam expander, an output coupler called Moya mirror, and the oscillator dye cell. The 5% of reflected 532 nm light from BS1 passes through the beam expander before falling onto the grating inside the oscillator cavity. The grating inside the oscillator cavity is used to select the output wavelength from a range of the wavelength available from the lasing emission of the dye solution. The laser output from the oscillator cavity is then amplified at the preamplifier and final amplifier. The fundamental output beam from the final amplifier then enters into the UV tracker (UVT) and generates frequency-doubled UV output.

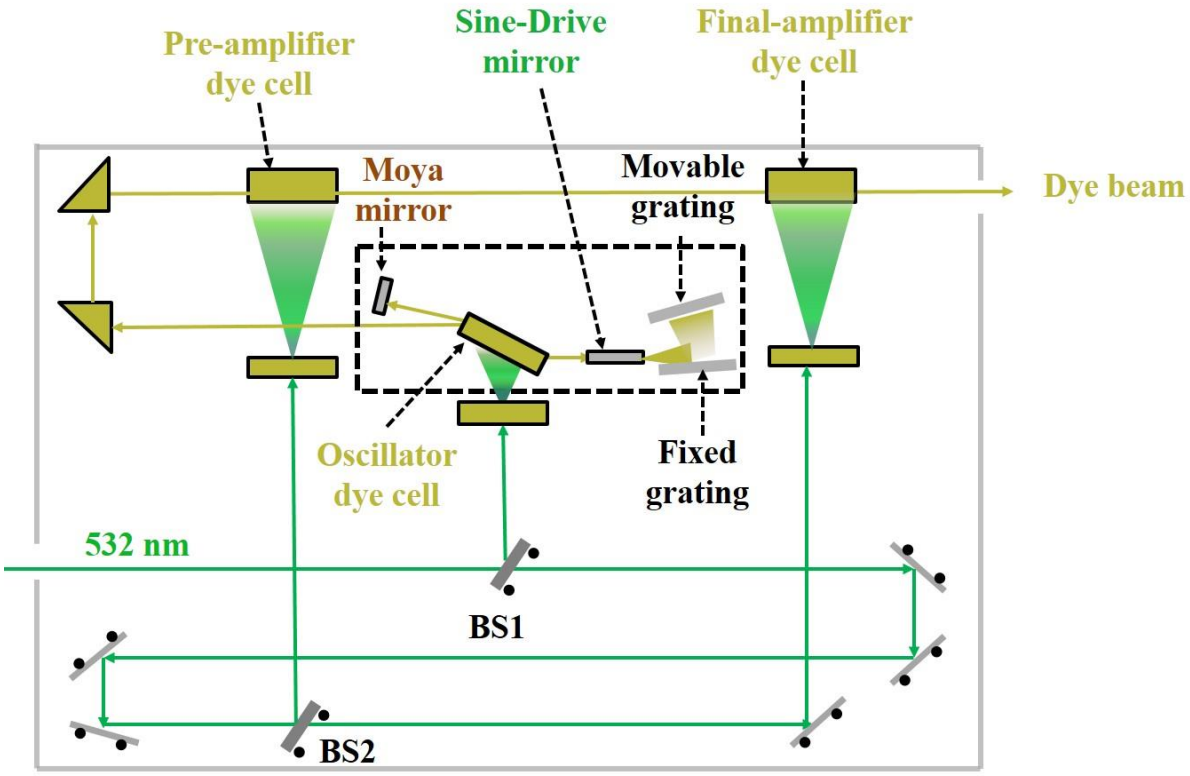

Figure 2.5. Schematic representation of Dye-laser system used for the generation of tunable UV laser output.

2.1.1.5.2 UV tracker

The UV tracker or UVT is used to generate the tuneable UV beam from a visible fundamental beam output from the dye laser with the use of non-linear KDP or BBO crystal. The tunable UV laser output can be generated from the fundamental output by two common methods

namely doubling and mixing after doubling. In the doubling method, the fundamental output of the dye laser is frequency-doubled by the KDP or BBO crystal. In the mixing after doubling method, the frequency-doubled output of the dye fundamental is mixed with the residual 1064 nm beam from the Nd: YAG laser in a BBO crystal.

2.1.1.5.3 IR OPA/OPA system

A combination of optical parametric oscillation (OPO) and optical parametric amplification (OPA) is used to generate the laser beam in the near-IR region (710 to 885 nm) and the mid-IR region (1.35 to 5 μm). The OPO system uses a non-linear KTP crystal pumped by 532 nm light whereas the OPA system uses four non-linear KTA crystals. The KTP and KTA crystals are mounted on stepper motors. Stepper motors enable the crystals to rotate in such a way that proper phase matching can be achieved for the generation of the desired IR radiation. A schematic for the generation of the tunable IR frequency has been provided in figure 2.7. The IR OPA/OPO system (Laser Vision) is pumped by the 1064 nm laser light generated from an Nd:YAG laser (Surelite II-10, 10 ns, 10 Hz) having a pulse energy of around 530 mJ. The 1064 nm laser output from the Nd:YAG laser is directed into the IR laser system by two steerings 1064 mirrors. The 1064 nm pump beam is collimated onto a beamsplitter by a telescope. One-third of the pump beam is directed towards the OPO system whereas the transmitted beam is directed towards the OPA crystals. A BBO crystal doubles the 1064 nm pump beam. The doubled 532 nm beam is directed towards the KTP crystal by the HR1 and HR2 mirrors and two new photons are generated through the optical parametric oscillation process. The sum of the frequencies of the two newly generated photons i.e. signal (ω_S) and idler (ω_I) are equal to the frequency of the pump beam. The rotation of the KTP crystal at different phase-matching angles provides tunable OPO output (signal and idler) in the 710-885 nm range. The OPO system has a linear cavity containing a KTP crystal, an output coupler, and a rear mirror. The signal and idler output from the OPO passes through a dove prism and MgF_2 waveplate. The

dove prism is used for the collimation of the beam whereas the MgF_2 waveplate changes the polarization of the idler beam from vertical to horizontal as KTA crystals are placed horizontally. A silicon filter is placed between the MgF2 and KTA crystals to reject the near-infrared signal(s) so that only the idler beam(i) reaches the KTA crystals of the OPA stage.

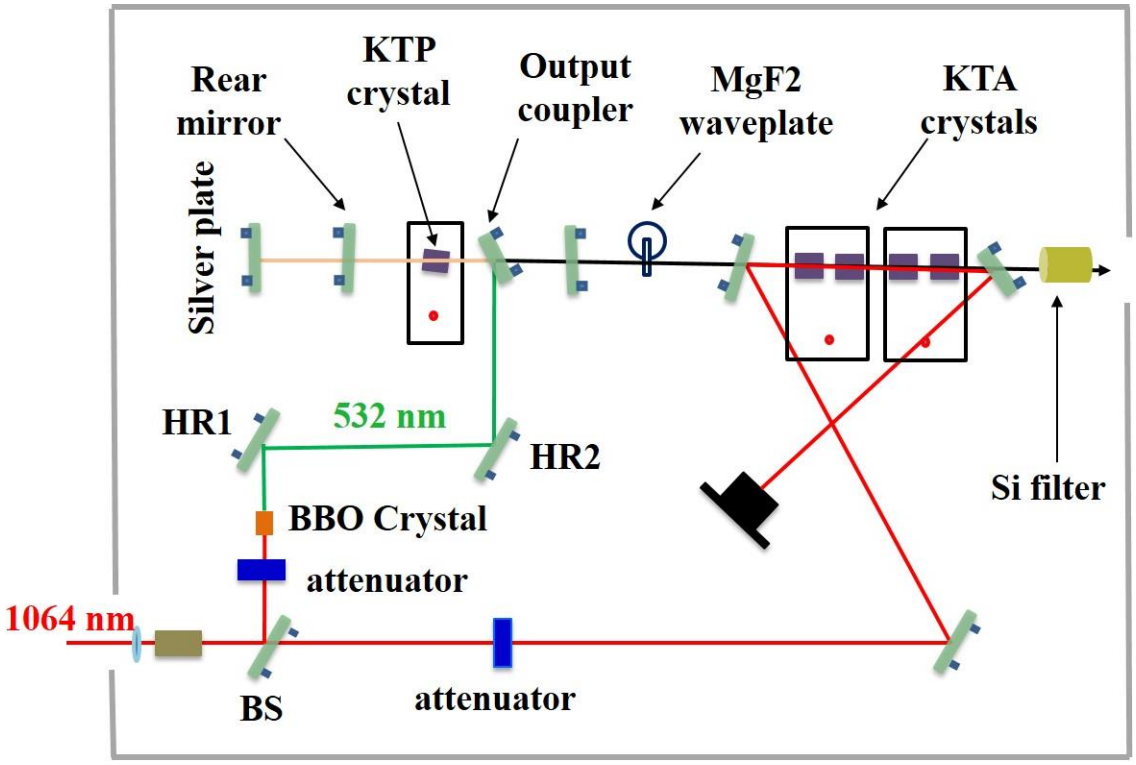

Figure 2.6. Schematic representation of IR OPA/OPO system used for the generation of laser wavelength in infra-red region.

The transmitted 1064 beam from the beam splitter moves through the half-wave plate before being directed towards the OPA setup. The half-wave plate provides desired polarization of the pump beam. The four KTA crystals at the OPA stage alternately rotate in the opposite direction so that path of the beam doesn't change. At the OPA stage, there is a mixing of the 1064 nm pump beam and idler beam and new signal and idler beams are generated through a difference frequency mixing process. The newly generated beams after the OPA stage are mid-IR idler and intermediate-IR signals, which are polarized vertically and horizontally, respectively. The

residual 1064 nm beam after the OPA stage is directed towards the beam dump by the 1064 nm separator mirror whereas the mid-IR idler and intermediate-IR signal are allowed to pass through. A "Stack of plates" Si filter is placed just before the exit of the IR laser setup to select either idler or signal IR beam. Mostly the idler IR beam is used for the experiment described in the thesis.

2.1.1.6 Gas-phase laser spectroscopy techniques

Various gas-phase laser spectroscopic techniques are used to measure mass-selected conformation-specific electronic and IR spectra of different peptide molecules in the gas phase. These techniques have been explained here briefly.

2.1.1.6.1 Resonantly Enhanced multi-photon Ionization Spectroscopy (REMPI)

One-color resonant 2-photon ionization (1C-R2PI) technique has been used to measure mass-selected electronic spectra of the peptides in the isolated gas phase. In this technique, two photons of the same frequency (color) or wavelength of the laser resonantly ionize the gaseous molecules from the neutral ground state. A schematic for the 1C-R2PI and 2C-R2PI techniques has been provided in figure 2.6. The first photon of a tunable UV laser excites the molecules from the ground electronic state to the first excited electronic state (S_1) and subsequently, the second photon excites the molecules from the first excited state to the ionized state. Apart from the 1C-R2PI technique, 2C-R2PI can be also used for the ionization of molecules. In the case of 2C-R2PI, two photons of different frequencies are used to ionize the molecules.

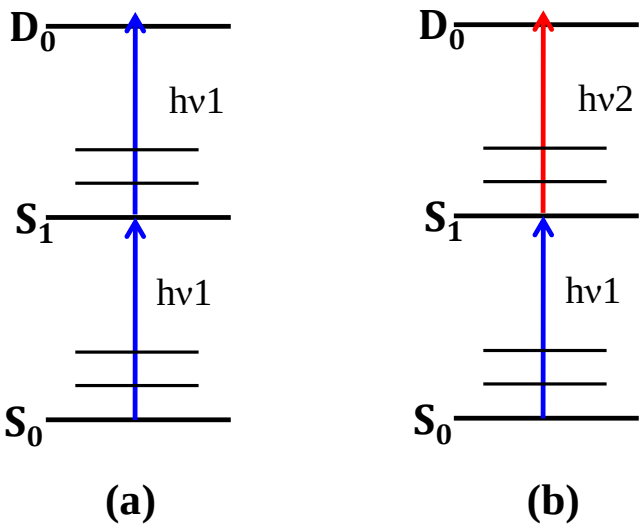

Figure 2.7. Schematic representation of the 1C-R2PI and2C-R2PI techniques used to record the electronic spectra of the molecules in the gas-phase.

2.1.1.6.2 Resonant Ion-dip Infrared spectroscopy (RIDIRS)

This technique is used to record conformation-specific infrared spectra of the molecules in the gas phase. The IR and UV laser beams are spatially overlapped in the ionization beam chamber. The IR laser precedes the UV laser by 150-200 μs. The UV laser wavelength is fixed at one of the electronic peaks whereas the IR laser is scanned throughout the region of the concerned IR frequency. The ion signal generated by the UV laser is depleted as the IR frequency matches the vibrational transition of the molecule. A schematic for the RIDIR spectroscopy has been shown in figure 2.8.

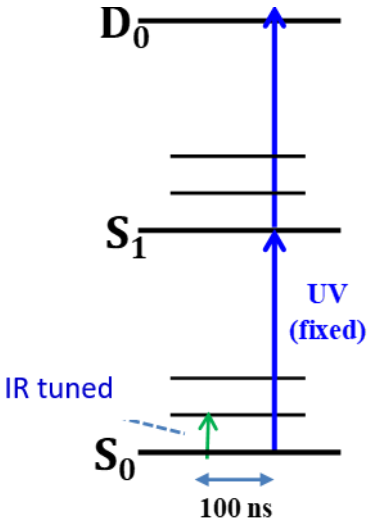

Figure 2.8. Schematic representation of RIDIR technique used for recording the IR spectrum in gas-phase.

2.1.1.6.3 IR-UV hole-burning spectroscopy

IR-UV double resonance spectroscopy has been used to determine the number of conformers contributing to the electronic spectrum of the molecule. In this technique, the IR laser is fired 150-200 ns before the UV laser. The IR laser is fixed at one of the vibrational frequencies of one of the conformers, and the UV laser is scanned through the whole region of the electronic spectrum. The electronic bands corresponding to the vibrational frequency of a particular conformer are depleted in intensity while the electronic bands which don't correspond to the vibrational frequency of the same conformer are unchanged in intensity. IR-UV double resonance Spectroscopy is also called IR-UV hole burn spectroscopy. The schematic for the IR-UV double resonance Spectroscopy has been provided in figure 2.9.

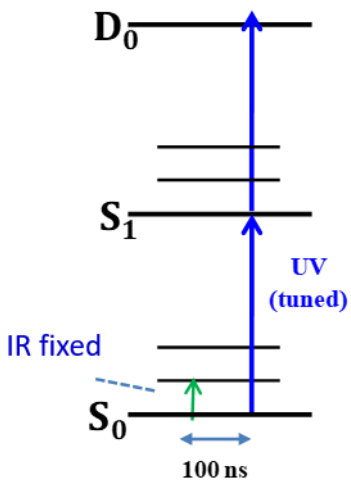

Figure 2.9. Schematic representation of IR-UV double resonance spectroscopy used for determining the presence of different conformers in the gas-phase experiment.

2.1.1.6.4 UV-UV hole-burning spectroscopy

The presence of different conformers in the gas phase experiment can also be discriminated by the UV-UV double resonance spectroscopy. In this technique, two UV laser beams are spatially

overlapped with the cold molecular beam. while being temporally delayed. One of the UV lasers (probe laser, 10 Hz rep. rate, 10 ns pulse width, pulse energy 0.2-0.4 mJ) fixed at one of the electronic bands is delayed by 100-200 ns from the other UV laser (pump laser, 10 Hz rep. rate, 10 ns pulse width, pulse energy 0.5-0.7 mJ), which is scanned through the whole region of the electronic spectrum. The electronic bands which belong to the same conformer are depleted in intensity. A schematic of the UV-UV hole-burning spectroscopy has been provided in figure 2.10.

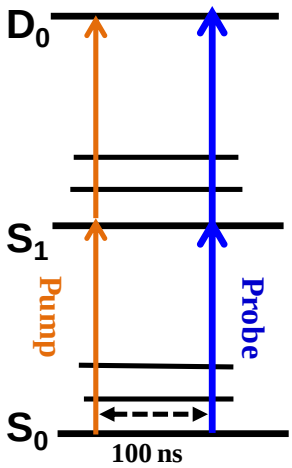

Figure 2.10. Schematic representation of the UV-UV hole-burning spectroscopy used for measuring the conformation-specific electronic spectra in gas-phase.

2.1.2 Solution phase spectroscopy

2.1.2.1 FT-IR spectroscopy

FTIR spectroscopy is used quite extensively to measure the molecular vibrations that absorb the radiation in the infrared region. The functional groups present in the molecule absorb infrared radiation giving a unique fingerprint characteristic. The FTIR spectrum of molecules is most commonly recorded in the solution phase apart from the gaseous and solid phases. The most common solvents for FTIR experiments are Carbon Tetrachloride (CCl_4), Carbon Disulfide (CS_2), and $CDCl_3$. FT-IR spectra of all the peptides studied in this thesis have been recorded using a Fourier-Transform IR spectrometer (Bruker Vertex 70).

2.1.2.2 NMR spectroscopy

The NMR (Nuclear magnetic resonance) spectroscopy is one of the most common spectroscopic techniques to identify and characterize molecules. The two of the most useful NMR techniques used to study the various peptides in this thesis have been described briefly, here.

2.1.2.2.1 1D NMR spectroscopy

The 1D NMR spectroscopy in solution is used to analyze the local chemical environment of nuclei which are present in the molecules. The most used 1D NMR technique for chemists is the ^{1}H NMR technique or proton NMR technique. All the ^{1}H NMR spectra of the peptides studied in this thesis have been recorded in $CDCl_3$ solvent using a 400 MHz NMR spectrometer (Bruker-400).

2.1.2.2.2 2D NMR spectroscopy

The two-dimensional NMR detects the signal (S) from the different nuclei as a function of two-time variables namely t_1 and t_2.[132, 133] In this method t_1 is time evolved under the first pulse whereas signal-induced due to spin change upon introduction of an external magnetic field is directly recorded as an accusation time t_2. This results in the formation of an outgoing signal f (t_1, t_2) which results in transformation into a two-dimensional spectrum of F (ω_1, ω_2) upon Fourier transformation. The most common and useful 2D NMR used for the characterization of molecules in the solution phase are ^{1}H-^{1}H COSY (COrrelated SpectroscopY), TOCSY (Total Correlation Spectroscopy), ROESY (Rotating Frame Overhauser Enhancement Spectroscopy) and Nuclear Overhauser Effect Spectroscopy (NOESY). All the 2D NMR experiments performed for the thesis have been recorded in $CDCl_3$ solvent using a 400 MHz NMR spectrometer (Bruker-400).

2.1.3 Single-crystal X-ray diffraction (XRD)

Single-crystal X-Ray diffraction technique has been used to obtain the crystal structure of various peptides studied in this thesis.[134] X-ray diffraction of all the peptides crystal were recorded using the APEX(II) DUO CCD diffractometer. This technique allows chemists to study the molecular structure at an atomic level.

2.2 Computational Methods

2.2.1 Gas-phase quantum chemistry calculation

The ground state geometry optimization and harmonic vibrational frequency calculations of several peptides are performed using Density Functional Theory (DFT) implemented in Gaussian 09 program package.[135] Various DFT functional used in the calculations are M05-2X, M06-2X, B97-D, ωB97X-D, B97-D3, and B3LYP-D3. The basis sets include Pople type i.e., 6-31+G(d), 6-311++G(d,p), 6-311++G(2d, 2p)) along with Dunning's correlation consistent one i.e., def2-TZVPP. DFT calculations are done using "opt=tight" convergence criteria and an "ultrafine" numerical integration grid.

2.2.1.1 Natural Bond Orbital (NBO) calculation

Various non-covalent interactions, especially hydrogen bonding interactions, present in the backbone of the peptides are visualized through NBO calculations implemented in NBO 6.0 program developed by Weinhold and co-workers.[136] The localized natural bond orbitals used for this analysis are constructed from the delocalized molecular orbitals. The natural bond orbitals are localized on one-center, two-center, or three-center systems. The NBO program uses the mathematical algorithm to generate the natural atomic orbitals (NAOs) from the input atomic orbital functions. The linear combination of the natural atomic orbitals (NAOs) is used to construct the natural hybrid orbitals (NHOs). The linear combination of the natural hybrid orbitals (NHOs) further produces the natural bond orbitals (NBOs). In the case of a linear

diatomic molecule i.e., “XY”, two types of natural bond orbitals (NBOs) are generated namely bonding NBO orbital (Lewis type) and anti-bonding NBO orbital (non-Lewis type).

The bonding NBO orbital (donor orbital) has more electron occupancy compared to the anti-bonding (acceptor orbital) one. In the case of hydrogen bonding, there is delocalization of the electron density from the donor NBO orbital (lone pair) to the acceptor NBO orbital. The second-order perturbative energy, $E^{(2)}_{i\rightarrow j*}$ (i and j* stand for the donor and acceptor orbitals) is used to quantify the strength of hydrogen bonding.

2.2.1.2 Non-covalent Interaction (NCI) index calculations

Non-covalent interaction (NCI) index calculations are used to further validate various non-covalent interactions observed in the peptide.[137, 138] NCI plot provides an insight into non-covalent interactions using electron density and its derivatives. The NCI index is constructed as a two-dimensional plot of electron density (ρ) and reduced density gradient (*s*).

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}}$$

The correlation between electron density (ρ) and reduced density gradient (s) has been provided above. In the case of weak inter or intramolecular interactions, there is a significant change in the reduced gradient which in turn brings out density critical points between interacting atoms. The strength of non-covalent interactions is determined based on the sign of Laplacian of density (ρ) i.e., $\nabla^2\rho$.

2.2.1.3 Solution phase quantum chemistry calculation

The solution phase quantum chemistry calculations for peptides have been performed using the polarizable continuum model (PCM) method. In this method, the solvent molecules are treated as a polarizable continuum. This method replaces the usual microscopic specification of a solvent system with a polarizable continuum medium with a suitable dielectric constant and

thermal expansion coefficient. This method creates an efficient way to optimize the structure of molecules in the solution.

Chapter 3

Observation of a Weak Intra-residue C5 Hydrogen-Bond in Z-Gly-Pro-OH

3.1 Introduction

Hydrogen-bonding between the amino hydrogen and carbonyl oxygen atoms of different amino acid residues of the peptide backbone is the key to the stabilization of various secondary structures in proteins.[12, 51, 139-144] The most common secondary structures in proteins are α-helices, β–sheets, and turns.[12, 30, 143-145] In general, the turns are classified into δ, γ, β, α , and π corresponding to the C6, C7, C10, C13, and C15 hydrogen-bonded rings involving the backbone N-H and C=O groups of $i \rightarrow i+1$, $i \rightarrow i+2$, $i \rightarrow i+3$, $i \rightarrow i+4$, and $i \rightarrow i+5$ residues, respectively.[12, 16] However, a hydrogen bond between the C=O and N-H groups of the same amino acid residue of proteins forming a C5 ring is unprecedented and not explored very much in the literature.[17] The C5 interaction being an intra-residue hydrogen bond is highly constrained and relatively weaker than other hydrogen-bonded rings present in the secondary structures of proteins.[146]

A five-membered hydrogen-bonded ring, albeit not termed as a C5 hydrogen-bond, to stabilize the trans conformer of N-alkyl α-halo/alkoxy acetamides over the cis conformation was reported by Nyquist using IR spectroscopy in the solution phase.[18] Later, Burgess et al. and several other researchers showed the existence of the C5 hydrogen-bond in several dipeptides using solution phase as well as matrix-isolation FT-IR spectroscopy.[87, 146, 147] The presence of the C5 hydrogen-bond in dipeptides has also been demonstrated through NMR spectroscopy and quantum chemistry calculations.[17, 91, 148, 149]

C5 hydrogen-bond is mostly observed in the β-sheet structures of proteins having an extended configuration.[17] In the β-sheet structures, s-type lone pair on carbonyl oxygen of the amide backbone of one strand overlaps with the σ* orbital of the amide N-H group of another strand to form a regular strong N-H…O hydrogen-bond. However, the p-type lone pair on the backbone carbonyl oxygen atom can be involved in a constrained C5 hydrogen-bond

interaction with the amide N-H group of the same residue. Raines and co-workers have found from an extensive PDB analysis that C5 hydrogen-bond is present in 94% of proteins.[17] They have verified the existence of the C5 hydrogen bond by studying monomeric peptide models as well as tryptophan zipper peptides through NMR and temperature-dependent CD spectroscopy. They have further reported that a fine interplay between the C5 hydrogen bonds and $n\rightarrow\pi^*$ interactions of the disulfide bonds of the cysteine residues contribute significantly to the stability of proteins.[150] $n\rightarrow\pi^*$ interaction is a weak non-covalent interaction that originates due to the delocalization of lone pair electrons on oxygen, nitrogen, and sulfur atoms into π^* orbitals of C=O, C=S, aromatic moieties, etc.[151-158]

The intrinsic nature and strength of the C5 hydrogen-bond have also been investigated in amino acids, monomeric peptide models, and dipeptides using gas-phase laser spectroscopy.[1, 26, 27, 112, 159-169] The propensity of the C5 hydrogen-bond interaction has been observed generally in glycine (Gly) containing small peptides i.e. Z-Gly-OH, Gly-Gly, Ac-Gly-Phe-NH_2.[1, 161, 167] However, most of the dipeptides, containing two different amino acid residues having C5 interaction, reported in the literature are stabilized by additional hydrogen-bond interactions present there.[112, 164, 167] The current investigation aims to study a dipeptide of two different residues stabilized by only a C5 hydrogen-bond interaction in order to demonstrate the significance of this weak interaction in the stability of the peptides.

In this work, we have studied the inherent conformational preferences of Z-Gly-Pro-OH through isolated gas-phase spectroscopy and quantum chemical calculations. Here, the capping group "Z" on the N-terminal of the peptide stands for benzyloxycarbonyl which is used as a UV chromophore for electronic absorption. Figure 3.1 shows the chemical structure of Z-Gly-Pro-OH describing the Ramachandran angles (φ, ψ). Two conformers of the peptide have been observed in the experiment. One of the conformers of Z-Gly-Pro-OH has an extended β-strand like structure solely stabilized by a C5 hydrogen bond while the other one

has a folded conformation stabilized by an OH…π interaction. Natural Bond Orbital (NBO)[170] and Non-Covalent Interaction (NCI)[137] calculations are performed to validate the non-covalent interactions present in the two conformers of the peptide studied here. The most intriguing finding of the present study is that the extended conformer of the peptide is much more stable than the folded conformer at the experimental temperature. Investigation of the conformational preferences of a peptide containing –Gly-Pro- sequence is also significant as this is a part of the repeating tripeptide (-Gly-Pro-Hyp-) sequence of collagen[66] which has a long extended triple helical structure.

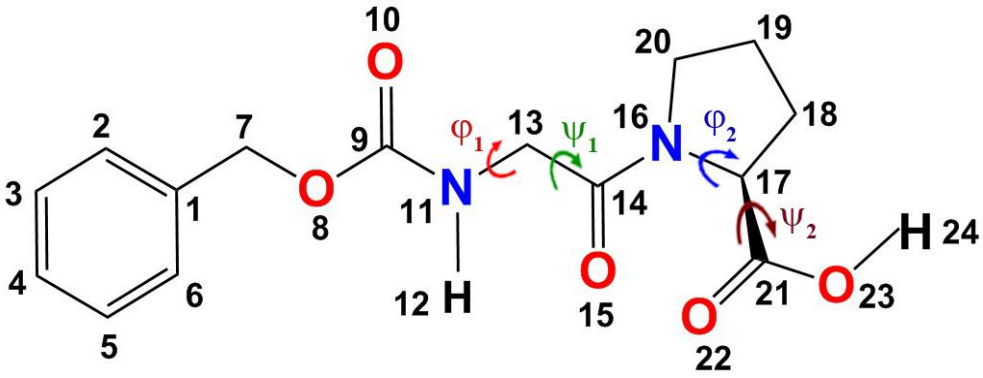

Figure 3.1. Skeletal structure of Z-Gly-Pro-OH showing the atom numbering scheme and definition of symbols representing the geometrical parameters. Dihedral angles φ_1 (C9-N11-C13-C14), ψ_1 (N11-C13-C14-N16), φ_2 (C14-N16-C17-C21) and ψ_2 (N16-C17-C21-O23) are the Ramachandran angles.

3.2. Methods and Characterization

3.2.1 Synthesis and characterization of Z-Gly-Pro-OH

Materials and Methods: All the amino acids, Cbz-Cl, N-Hydroxysuccinimide (NHS), N, N'-Dicyclohexylcarbodiimide (DCC), Na_2CO_3, $NaHCO_3$, and NMR solvents, THF, EtOAc, were obtained from commercial sources and used without further purification. Column chromatography was performed on silica gel. ^{1}H and ^{13}C NMR spectra were recorded on a 400 MHz (or 101 MHz for ^{13}C) spectrometer using residual solvent signals as an internal reference (DMSO-d_6 ^{1}H-2.50 ppm, ^{13}C-39.52 ppm). The chemical shifts (δ) are reported in ppm and coupling constants (J) in Hz.

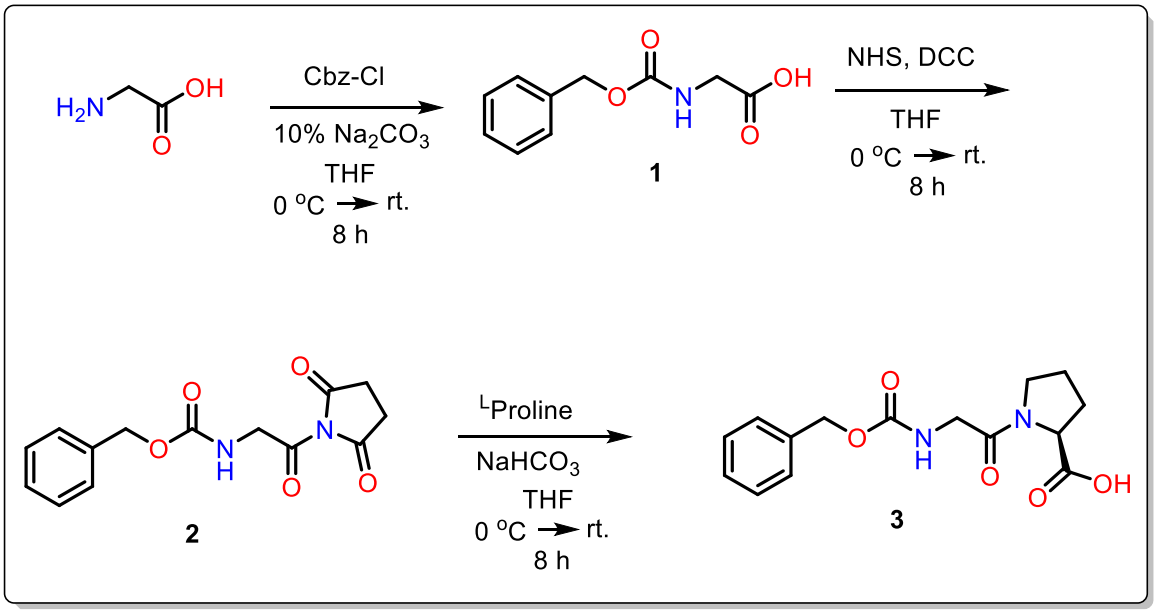

Figure 3.2. A schematic for the synthesis of Z-Gly-Pro-OH.

Z-Gly-Pro-OH (compound 3) was synthesized from the previously reported method. Glycine (15 mmol, 1.13 g) was added to 20 mL of THF (12.7 mmol, 2.67 g). Subsequently, Cbz-Cl in THF was added dropwise to the reaction mixture in ice condition and stirred for 8 h. Upon completion of the reaction, THF was removed by rota evaporation and worked up by acidification with 10 % HCl and the compound was extracted with EtOAc (150 ml X 1).

Finally, the organic layer was washed with brine solution (100 mL X 1). The organic layer was dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get Compound 1 (Yield- 2.67 g, 85%).

Cbz-Glycine i.e compound 1 (12.7 mmol, 2.67 g) was dissolved in 20 ml THF and this solution was then cooled to 0 ^oC using an ice bath in N_2 atmosphere. NHS (25 mmol, 2.9 g) was added to Cbz-Glycine and the reaction mixture was stirred for 10 mins. After that, DCC (15.5 mmol, 3.2 g) was added to the reaction mixture and stirred for 8 h. After completion of the reaction, it was filtered off and the filtrate was collected to get compound 2 (Yield- 2.96 g, 80%).

The active ester i.e compound 2 (10.2 mmol, 2.96 g) was dissolved in 20 ml THF and this solution was then cooled to 0 ^oC using an ice bath in N_2 atmosphere. ^{L}Proline (11.2 mmol, 1.3 g) in $NaHCO_3$ was added to the reaction mixture and stirred for 8 h. Upon completion of the reaction, THF was evaporated and the compound was extracted with EtOAc (150 mL X 1) and then the organic layer was washed with brine (100 mL X 1). The organic layer was dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to give a crude peptide which was then purified by precipitation method to get pure compound 3 as a white solid (Yield- 3.0 g, 70%).

Characterization of Z-Gly-Pro-OH:

^{1}H NMR (400 MHz, DMSO-d_6): δ 7.41 – 7.26 (m, 5H), 5.03 (d, *J* = 2.2 Hz, 2H), 4.23 (dd, *J* = 8.8, 3.5 Hz, 1H), 3.92 – 3.68 (m, 2H), 3.50 (q, *J* = 7.0, 6.5 Hz, 4H), 2.22 – 2.02 (m, 1H), 1.93 – 1.80 (m, 2H).

^{13}C NMR (101 MHz, DMSO): δ 173.36, 167.17, 156.50, 137.14, 128.38, 127.70, 65.42, 58.64, 46.30, 45.48, 42.64, 39.52, 28.64, 24.39.

HRMS (ESI-MS) (m/z): [M+H$^+$] calculated for ($C_{15}H_{18}N_2O_5$ + H$^+$) **307.1294** and found **307.1297**

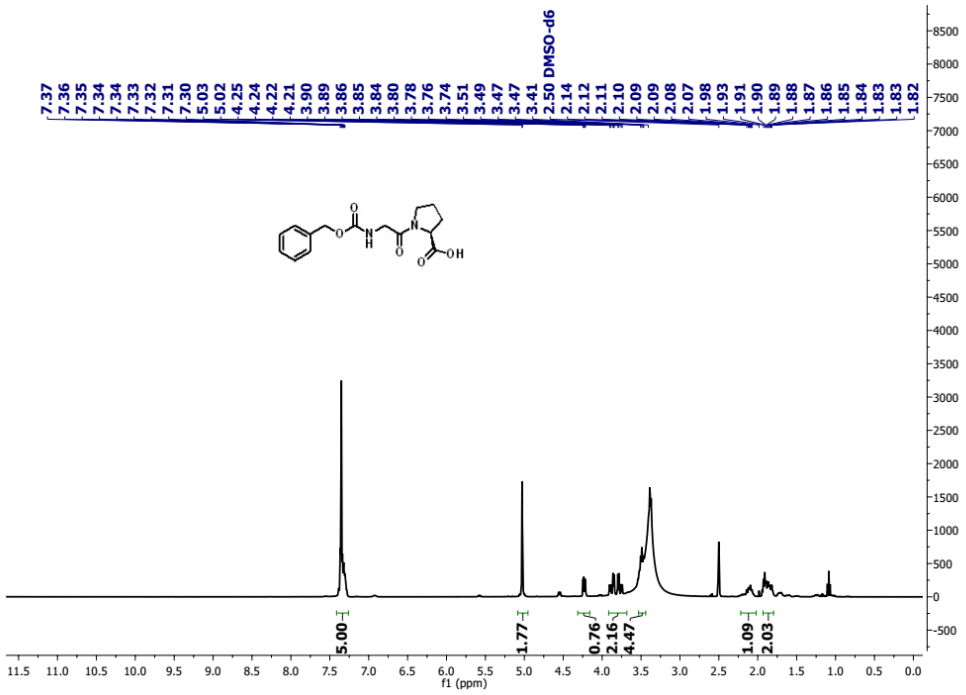

Figure 3.3. [1]H NMR spectrum of Z-Gly-Pro-OH recorded using 400 MHz NMR spectrometer.

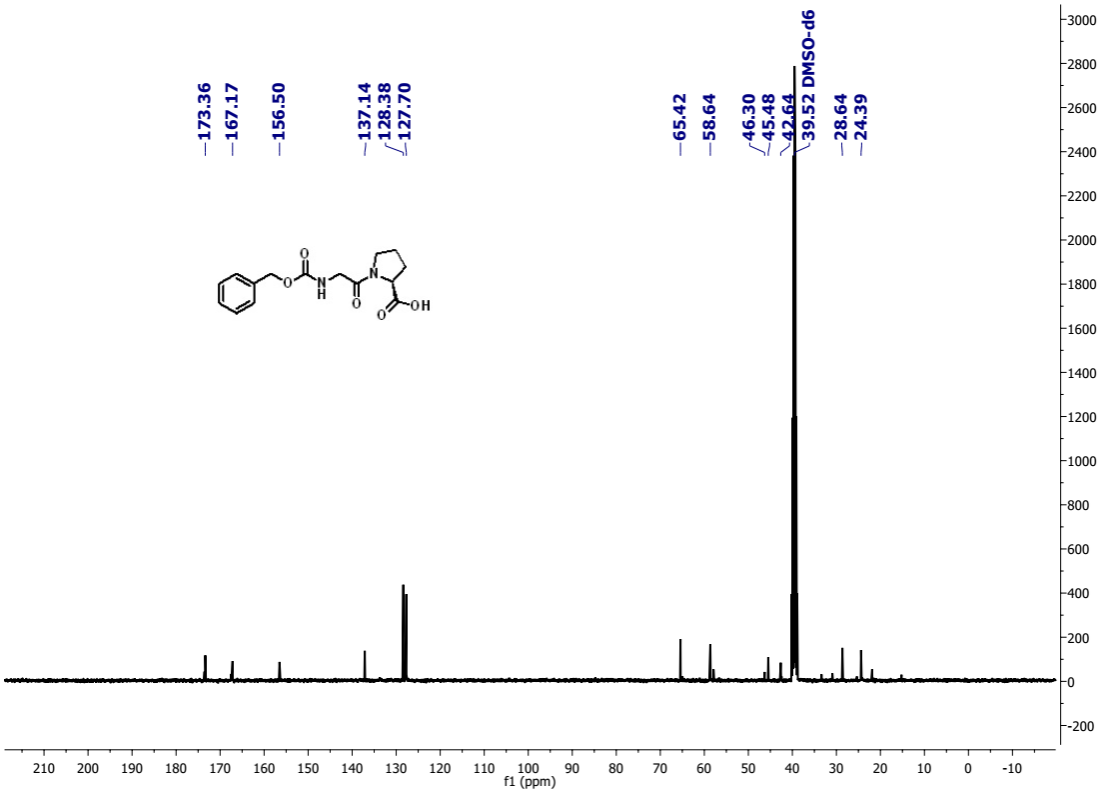

Figure 3.4. ^{13}C NMR spectrum of Z-Gly-Pro-OH recorded using 400 MHz spectrometer.

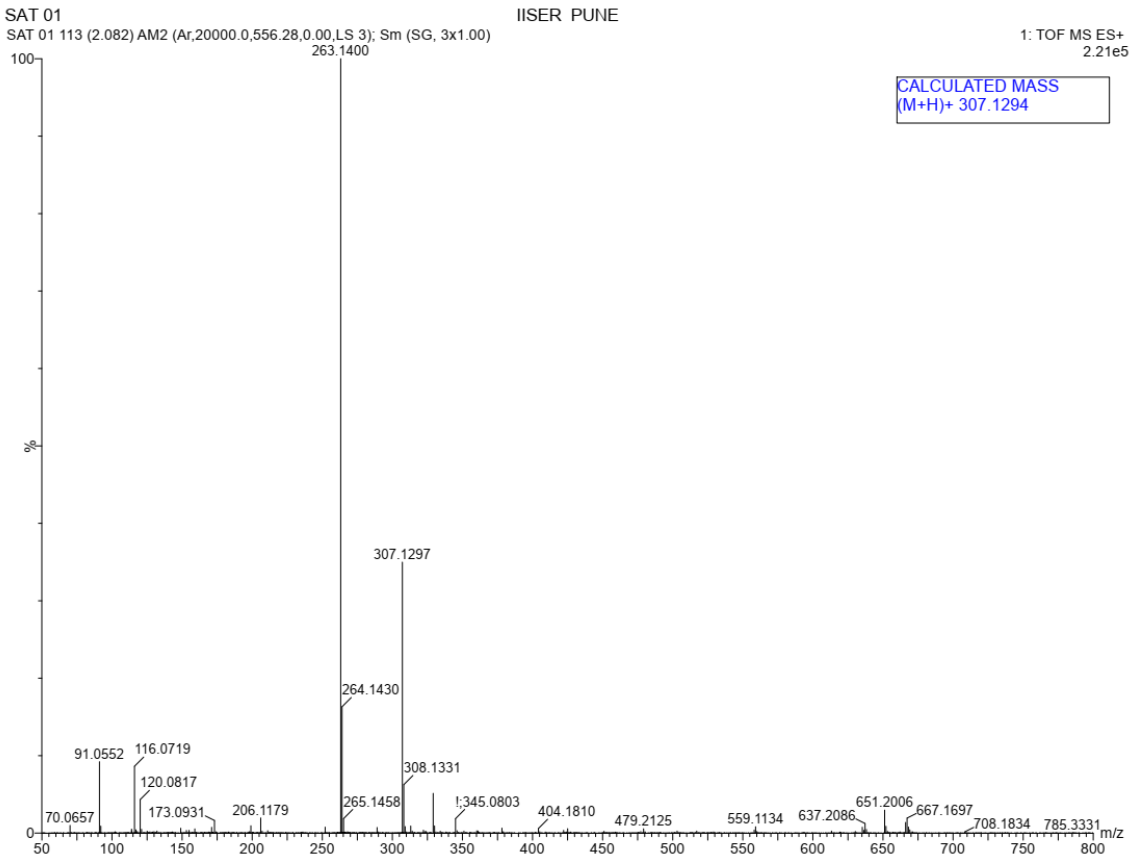

Figure 3.5. High resolution mass spectrum (HRMS) of Z-Gly-Pro-OH measured using electrospray ionization mass spectrometer (ESI-MS).

3.2.2 FTIR spectrum of Z-Gly-Pro-OH

FTIR spectrum of Z-Gly-Pro-OH peptide has been measured in $CDCl_3$ solution using Bruker ALPHA II (2 cm^{-1} resolution) FT-IR spectrometer. The IR spectrum displayed in Figure 3.6 shows the vibrational transitions only in the 2800-3700 cm^{-1} range covering the C-H, N-H, and O-H stretching regions. The N-H and O-H bands are the most characteristic ones for assigning the structure and conformation of the peptide. However, the observed spectral band in the probable position of the N-H and O-H stretching vibration is extremely broad i.e. 3200-3600 cm^{-1}. Hence, it is not straightforward to assign the structure of the probable conformations of

the peptide from the solution phase IR spectrum. It is thus important to measure conformation-specific IR spectra of the peptide in the isolated gas phase.

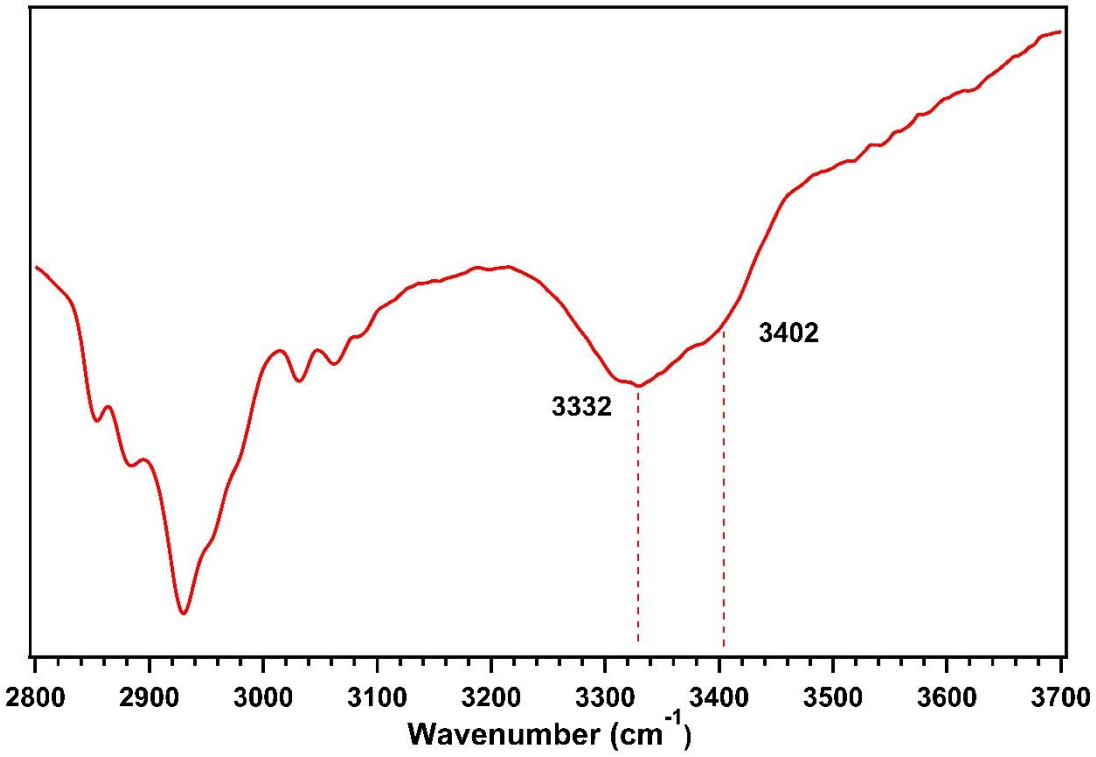

Figure 3.6. FTIR spectrum of Z-Gly-Pro-OH peptide measured in $CDCl_3$ solvent.

3.2.3 Experimental details

A detailed description of the jet-cooled laser desorption Time of Flight (TOF) mass spectrometer coupled with resonantly enhanced multiphoton ionization (REMPI) technique to measure the electronic and IR spectra of Z-Gly-Pro-OH in an isolated gas phase was provided elsewhere.[171-174] Z-Gly-Pro-OH was synthesized according to a method reported previously in the literature.[175] The synthesis method of the peptide and its characterization schemes are described in the supporting information.

A mixture of Z-Gly-Pro-OH sample and graphite powder (Sigma Aldrich, size ~ 20 μm) with a ratio of 70:30 was pressed in a hydraulic press applying ~ 3 tons of pressure to make a pellet of 2 mm thickness and 12 mm diameter. The pellet was then cut into two semi-circular halves

and one of the semi-circular pellets was put into a sample holder fixed with an XYZ manipulator having motorized translation along the Z-axis. 532 nm laser beam (500 μJ/pulse) from an Nd: YAG laser (Continuum, Minilite-I, 10 Hz, 10 nanoseconds) was focused on the pellet through an optical fiber (400 μm core diameter, 4 m length) for desorption of the sample from the surface.[176, 177] The pellet was translated back and forth to avoid any damage to the sample surface by the laser beam. The desorbed peptide molecules were seeded into the molecular beam of Ar gas (5 bar) supersonically expanded through a pulsed nozzle (General valve, series 9, 500 μm diameter orifice, 10 Hz).

The molecular beam of the peptide was ionized by the mixing after doubling output (0.2–0.3 mJ) of a tunable dye laser (ND6000, Continuum) pumped by a second harmonic output of an Nd: YAG laser (10 nanoseconds, 10 Hz, Surelite II-10, Continuum) and analyzed in the TOF mass spectrometer (Jordan TOF, USA). The electronic spectrum of the peptide is measured using the one-color resonant 2-photon ionization (1C-R2PI) technique in which the first photon of the tunable dye laser excites the peptide molecules into the S_1 state while the second photon ionizes the molecules.

The IR spectra of the peptide were recorded using resonant ion dip infrared (RIDIR) spectroscopy.[105] In this technique, the UV laser wavelength was fixed at one of the bands of the electronic spectrum while the IR laser, preceding the UV laser by 150 ns, was scanned in the O-H and N-H stretching frequency of the region of the peptide. The IR spectra were obtained as depletion in the UV ion signal whenever the IR laser frequency was in resonance with any one of the vibrational frequencies of the molecule. The fundamental output from an unseeded Nd:YAG laser (Continuum, Surelite II-10, 10 nanoseconds, 10 Hz) was used to pump the IR laser (Laser Vision, pulse energy ~ 4-5 mJ, resolution ~ 2.5 cm^{-1}).

IR-UV hole-burning spectroscopy was used to determine the presence of different conformers of the peptide in the experiment.[105] In this technique, the IR laser was fixed at a particular

vibrational frequency of the molecule while the UV laser was scanned in the region of the electronic spectrum. The temporal and spatial overlaps in the IR-UV hole-burning and RIDIR spectroscopy are similar.

3.2.4 Computational details

Theoretical calculations were carried out to determine the structures of the conformers of Z-Gly-Pro-OH observed in the experiment and explore the intramolecular interactions present there. Probable initial structures of the peptides were obtained using Dreiding force field calculation[178] implemented in the MarvinSketch software.[179] The conformers with energies within 40 kJ/mol from the global minimum obtained from the force field calculation were subjected to quantum chemistry calculations at different levels of density functional theory (DFT) using various basis sets employing Gaussian 09 software package.[135, 180-184] Various DFT levels used for the calculations were M05-2X/6-31+G(d), ωB97X-D/6-31++G(d,p), ωB97X-D/6-311++G(d,p), and M06-2X/6-311++G(d,p). Electronic energies of different conformers of Z-Gly-Pro-OH were corrected for zero-point vibrational energies. The electronic energies of the conformers were further corrected for their entropy contribution by calculating relative Gibbs free energies at different temperatures.[185, 186] Vibrational frequencies of the two most important functional groups NH and OH of Z-Gly-Pro-OH calculated with harmonic approximations are scaled with respect to those reported experimentally for Z-Gly-OH in the literature.[1] The presence of various non-covalent interactions in the experimentally observed conformers of Z-Gly-Pro-OH was validated through natural bond orbital (NBO)[170] and non-covalent interaction (NCI) calculations.[138, 187, 188] The NBO calculations were done with NBO 6.0 software[136] while the NCI calculations were performed using NCIPLOT program.[137]

3.3 Results and Discussions

3.3.1 Electronic Spectra of Z-Gly-Pro-OH

Figure 3.7a shows the electronic spectrum of Z-Gly-Pro-OH recorded in the S_0-S_1 region of the Z-cap in the 37450-37620 cm^{-1} region using 1C-R2PI spectroscopy. The spectrum shows a few sharp bands along with a very weak broad feature. To check whether various bands present in the electronic spectrum are originating from a single conformer or multiple conformers of Z-Gly-Pro-OH, IR-UV hole-burning experiment has been performed. Figure 3.7b-c shows the IR-UV hole-burning spectra obtained by fixing the IR laser at the N-H stretching frequencies of 3457 and 3450 cm^{-1} of Z-Gly-Pro-OH, respectively, while the UV laser was scanned in the

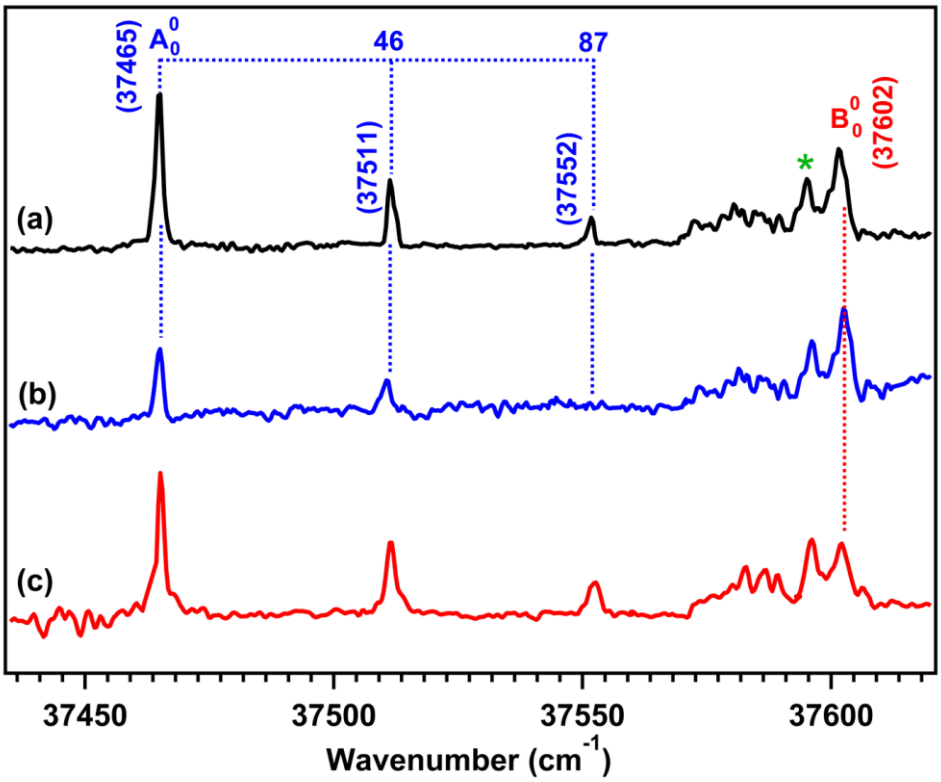

Figure 3.7. (a) Electronic spectrum of Z-Gly-Pro-OH measured using 1C-R2PI spectroscopic technique. (b)-(c) IR-UV hole-burning spectra of Z-Gly-Pro-OH measured by fixing the IR laser at 3457 and 3450 cm^{-1} respectively. The band marked by an asterisk (37596 cm^{-1}) and the weak broad feature in the red-side of this band neither provided any IR spectra nor any hole-burning spectra.

R2PI spectrum region. The N-H stretching frequencies at 3457 and 3450 cm^{-1} are obtained by fixing the UV laser at the 37465 and 37602 cm^{-1} bands, respectively, in the electronic spectrum (Figure 3.10a) by employing RIDIR spectroscopy discussed in section 3.3.3

We can see in the hole-burning spectrum provided in Figure 3.7b that the most intense band at 37465 cm^{-1} and two successive bands at 37511 and 37552 cm^{-1} show the depletion in the ion signal. Thus, these three bands belong to the same conformer of Z-Gly-Pro-OH, labeled as A. The 37465 cm^{-1} band is assigned to the origin band (S_0-S_1) of the conformer A while the A_0^0 + 46 and A_0^0 + 87 cm^{-1} are the vibronic bands of the same conformer. The hole-burning spectrum displayed in Figure 3.7c shows the depletion of only a single band at 37602 cm^{-1}, which is assigned to the origin band (S_0-S_1) of conformer B of Z-Gly-Pro-OH. The sharp band at 37596 cm^{-1} marked with an asterisk and the weak broad feature in the red-part of that band is not affected in both of the hole-burning spectra. However, we were not successful in obtaining any IR or hole-burning spectra by probing these electronic bands as these are weak in intensity. These bands could belong to some other conformers which have very low abundance in the experiment. Difficulties in measuring hole-burning spectra by probing the weak bands in the electronic spectra are reported in the literature.[121, 168] The electronic spectral region of Z-Gly-Pro-OH studied here match well with that of the Z-cap peptides (i.e., Z-Gly-OH, Z-$(Gly)_3$-OH, Z-Aib-OH, Z-Gln-OH) reported in the literature.[1, 159, 160]

3.3.2 Conformational landscape of Z-Gly-Pro-OH

A large number (~200) of conformers of Z-Gly-Pro-OH was generated using a force field based MarvinSketch software.[179] The resulting conformers were arranged in terms of relative energies with respect to the global minimum obtained from the conformational search. The conformers having an energy difference of 0.5 kJ/mol and similar structures were placed in the same group. A total of 65 groups were formed and the conformer with the lowest energy in each group was optimized at the M05-2X/6-31+G(d) level of theory. Finally, the M05-2X/6-31+g(d) optimized structures with energies within 24 kcal/mol were further considered for

geometry optimizations at the ωB97X-D/6-31++G(d,p) level of theory. Thermal correction to the zero-point energy corrected electronic energies of the conformers was done by calculating their relative Gibbs free energies (ΔG_{rel}) at different temperatures ranging from 0-1000 K with an interval of 50 K. It has been emphasized recently that the population of the conformers of the laser-desorbed molecules

in the jet-cooled experiments is best described by their ΔG_{rel} calculated at 300-500 K rather than the default electronic energies calculated at 0 K.[161, 166, 189, 190]

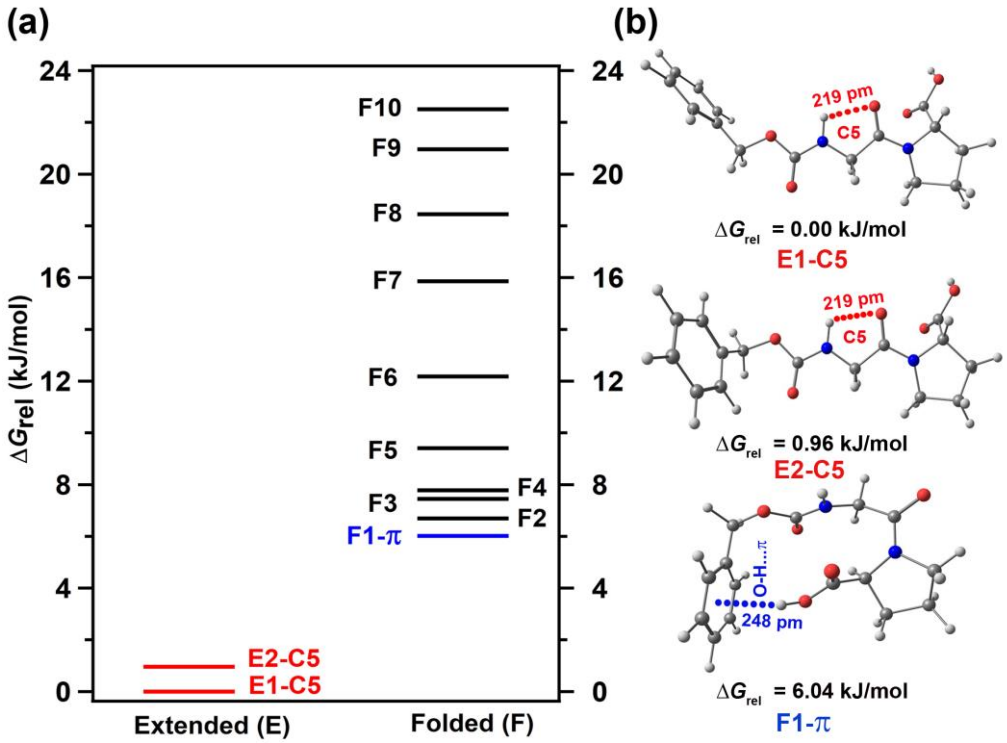

Figure 3.8. (a) Energy landscape of the low energy conformers of Z-Gly-Pro-OH with the relative Gibbs free energy (ΔG_{rel}) values within 24 kcal/mol calculated at 300 K using ωB97X-D/6-31++G(d,p) level of theory. (b) The optimized structures of the three lowest energy conformers of Z-Gly-Pro-OH observed in the experiment.

Figure 3.8a presents the energy landscape of the 12 low energy conformers of Z-Gly-Pro-OH with ΔG_{rel} values within 24 kcal/mol from the global minimum calculated at 300 K at the ωB97X-D/6-31++G(d,p) level of theory. Based on the geometries of the optimized structures

(Figure 3.8b), the conformers are generally named as either extended (E) or folded (F). For the brevity, the optimized structures of only the three lowest energy conformers of Z-Gly-Pro-OH observed in the experiment are provided in Figure 3.8b while the optimized structures of all the 12 conformers are shown in Figure 3.14.

The conformers E1 and E2 shown in Figure 3.8b have an extended β-strand type structure stabilized by an intra-residue C5 hydrogen-bond interaction present between the N-H and C=O groups of the Gly residue and hence these two conformers are named E1-C5 and E2-C5, respectively. It should be noted that the E1-C5 and E2-C5, which are energetically very close to each other, differ structurally only in terms of the orientation of the benzyl group. The remaining conformers shown in Figure 3.14 have folded structure designated as F. Interestingly, the F1 conformer is unique in the orientation of the carboxylic OH group which is involved in the O-H…π interaction with the phenyl group of the Z-cap, and thus, this conformer is named as F1-π. The carboxylic O-H group in the remaining folded conformers is free only.

A few selected important geometrical parameters of the observed conformers of Z-Gly-Pro-OH calculated at the ωB97X-D/6-31++G(d,p) level of theory are listed in Table 3.1 It is noteworthy that the Ramachandran angles (ϕ_1, ψ_1, and ϕ_2, ψ_2) as well as C5 hydrogen-bond distance and angle in the E1-C5 and E2-C5 conformers are quite similar. The Ramachandran angles ϕ_1, ψ_1= -175°, -178° of the Gly residue in the E1-C5 and E2-C5 conformers are the signatures of the presence of the extended β-strand like C5 structure. Similar ϕ_1, ψ_1 angle for the C5 hydrogen-bonded structures of Z-Gly-OH, Z-Aib-OH, Z-Gln-OH, Ac-Phe-NHMe model peptides or Z-$(Aib)_2$-OMe, Gly-Gly dipeptides studied by gas phase spectroscopy have been reported in the literature.[1, 159-161, 165] Raines and co-workers found from the PDB analysis of high-resolution protein crystal structures that approximately 5% of all the residues having

backbone dihedral angles >140° are involved in C5 hydrogen-bonding interactions.[17] The C5 hydrogen-bond distance of 2.19 Å observed in the E1-C5 and E2-C5 conformers of Z-Gly-Pro-OH also match well with that reported for several peptides in the literature.[27, 91, 159, 160, 163]

Table 3.1. Important geometrical parameters of the observed conformers of Z-Gly-Pro-OH calculated at the ωB97X-D/6-31++G(d,p) level of theory

Geometrical parameters	E1-C5	E2-C5	F1-π
φ_1	-175	-175	-140
ψ_1	-178	-178	58
φ_2	-59	-60	-89
ψ_2	150	150	169
C5 bond distance (Å)	2.19	2.19	-
OH...π bond distance (Å)	-	-	2.47
C5 bond angle (^{0})	105.23	105.35	-
OH...π bond angle (^{0})	-	-	121.82

Table 3.2. Comparison of the zero-point corrected energies (kJ/mol) of the first six low energy conformers of Z-Gly-Pro-OH calculated at 0 and 300 K at the ωB97X-D and M06-2X levels of theory using 6-31++G(d,p) and 6-311++G(d,p) basis sets.

	ωB97X-D				M06-2X	
	6-31++G(d,p)		6-311++G(d,p)		6-311++G(d,p)	
Conformers	ΔE_{rel} (0 K)	ΔG_{rel} (300 K)	ΔE_{rel} (0 K)	ΔG_{rel} (300 K)	ΔE_{rel} (0 K)	ΔG_{rel} (300 K)
E1-C5	7.90	0	10.00	0.67	10.21	0.42
E2-C5	5.21	0.97	6.51	0.00	7.98	0.00
F1-π	2.56	6.05	2.39	2.69	0.00	4.33
F2	0.00	6.72	0.00	6.38	5.96	9.45
F3	1.81	7.48	2.69	8.27	7.22	10.88
F4	10.71	7.81	12.05	7.43	14.91	7.10

It is quite intriguing that the two extended structures (E1-C5 and E2-C5) of Z-Gly-Pro-OH are more stable than all the folded structures (Figure 3.8a). Table 3.2 shows a comparison of the zero-point corrected energies of the first six low energy conformers of Z-Gly-Pro-OH calculated at 0 and 300 K at the ωB97X-D and M06-2X levels of theory using 6-31++G(d,p) and 6-311++G(d,p) basis sets. ΔG_{rel} (0 K) and ΔG_{rel} (300 K) values of all the 12 conformers of Z-Gly-Pro-OH calculated at the ωB97X-D/6-31++G(d,p) level of theory are also provided in Table 3.

It is obvious that the folded conformer (either F1-π or F2) is the global minimum at 0 K while the extended conformer (E1-C5 or E2-C5) is the most stable one at 300 K. Thus, it can be concluded that larger entropic effects of the extended conformer compared to the folded conformer at 300 K contribute to the enhanced stability of the former one. Similar observation on the stabilization of the extended structures over the folded structures of several peptides and β-amino acids has been reported in the literature.[161, 162, 189-191]

Table 3.3. Relative Gibbs free energy (ΔG_{rel}) values of the low energy conformers of Z-Gly-Pro-OH calculated at 0 K and 300 K at the ωB97X-D/6-31++G(d,p) level of theory.

Conformers	ΔG_{rel} (kJ/mol) at 0 K	ΔG_{rel} (kJ/mol) at 300 K
E1-C5	7.90	0.00
E2-C5	5.21	0.97
F1-π	2.56	6.05
F2	0.00	6.72
F3	1.81	7.48
F4	10.71	7.81
F5	7.60	9.45
F6	6.05	12.22
F7	19.28	15.92
F8	22.6	18.52
F9	27.55	21.04
F10	24.82	22.60

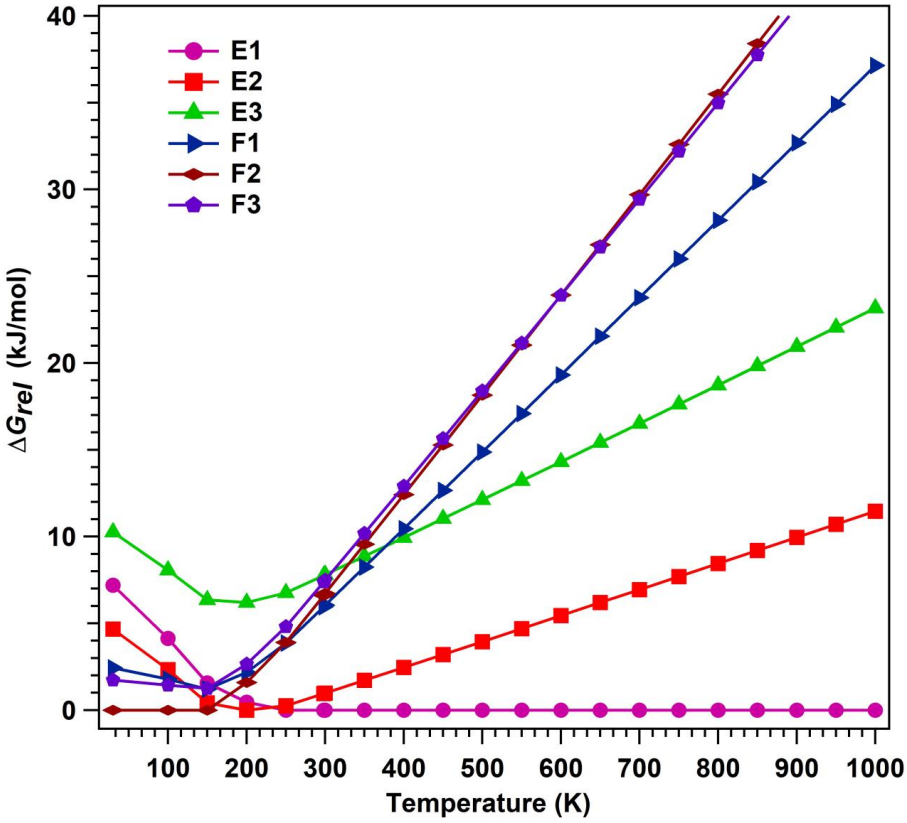

Figure 3.9. A plot of relative Gibbs free energy (ΔG_{rel}) of six low energy conformers of Z-Gly-Pro-OH as a function of temperature (0-1000 K) calculated at the ωB97X-D/6-31++G(d,p) level of theory.

We have also shown the plot of ΔG_{rel} of the six low energy conformers of Z-Gly-Pro-OH as a function of temperature (0-1000 K) in Figure 3.9. It is important to note that the energies of both the extended conformers are lower than those of the folded conformers at a temperature 200 K and higher.

3.3.3 IR spectroscopy of Z-Gly-Pro-OH

Figure 3.10a and 3.10c shows IR spectra of Z-Gly-Pro-OH measured in the N-H and O-H stretching region by probing the A_0^0 band (37465 cm^{-1}) of the conformer A and B_0^0 band (37602 cm^{-1}) of the conformer B, respectively, using RIDIR spectroscopy. The structures of the two observed conformers of the peptide are determined by comparing the IR spectra measured from the experiment with those obtained from the calculation of various probable low-energy

conformers. A comparison of the experimental IR spectra of the two observed conformers with the ωB97X-D/6-31++G(d,p) and M05-2X/6-31+G(d) level calculated, scaled, harmonic IR spectra of the 12 low energy conformers of Z-Gly-Pro-OH with energy within ΔG_{rel} ~ 24 kJ/mol of the global minimum is provided in Figure 3.11. The scaling factors used for the harmonic NH and OH stretching frequencies at the ωB97X-D/6-31++G(d,p) level are 0.944 and 0.933, respectively, while those at the M05-2X/6-31+G(d) level are 0.944 and 0.952, respectively. The scaling factors are obtained by scaling the ωB97X-D/6-31++G(d,p) and M05-2X/6-31+G(d) level calculated harmonic frequencies of the most stable conformer of Z-Gly-OH to the experimental frequency reported in the literature.[1]

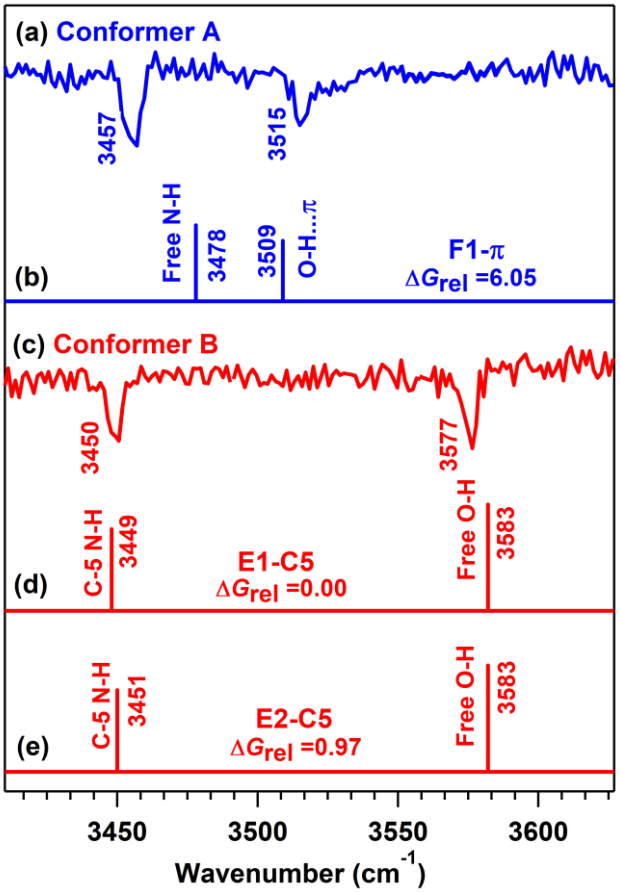

Figure 3.10. (a) and (c) IR spectra of different conformers of Z-Gly-Pro-OH measured using RIDIR spectroscopy by fixing the UV laser at 37465 and 37602 cm^{-1} respectively. (b), (d)-(e) are scaled theoretical IR spectra of different conformers of Z-Gly-Pro-OH with their relative Gibbs free energy (ΔG_{rel}, in kJ/mol) values calculated at the ωB97X-D/6-31++G(d,p) level of theory. The N-H and O-H stretching frequency of all the conformers are scaled with scaling factors of 0.944 and 0.933 respectively (See the text).

A close inspection of Figure 3.11 reveals that the theoretical IR spectra of the global minimum conformer E1-C5 and the next lowest energy conformer E2-C5 of a similar structure match very well with the experimental IR spectrum of the conformer B provided in Figure 3.10c. Although the theoretical IR spectrum of another conformer F3 of ΔG_{rel} (300 K) ~1.5 kcal/mol is similar to the experimental IR spectrum of the conformer B, the possibility of the observation of the high energy F3 conformer in the experiment, compared to the global minimum E1-C5 and its near-isoenergic conformer E2-C5, can be ruled out. Thus, the theoretical IR spectra of only the E1-C5 and E2-C5 are provided in Figures 3.10d and 3.10e, respectively, to assign the experimental IR spectrum of the conformer B provided in Figure 3.10c. On the other hand, the theoretical IR spectrum of solely the F1-π conformer provided in Figure 3.10b corroborates with the experimental IR spectrum of the conformer A presented in Figure 3.10a. Hence, conformer A is assigned to the F1-π structure while conformer B can be assigned to either the E1-C5 or E2-C5 structure of Z-Gly-Pro-OH.

It is worth mentioning here that the intensity of the electronic band of conformer A is higher than that of the conformer B as observed in the electronic spectrum of Z-Gly-Pro-OH (Figure 3.7a) although the latter one is energetically lower than the former one after the thermal correction. The anomaly in the observed intensity of the electronic bands of the conformers in terms of their population can arise as the intensity of the electronic bands does not solely depend on the population of the molecules in the ground electronic state. Thus it could be

speculated that the absorption/photoionization cross section, S1 state lifetime, and Franck-Condon factor of the conformer A might be more favorable towards higher band intensity than that of the conformer B.[159]

The 3515 cm^{-1} band in Figure 3.10a is now assigned to the carboxylic OH group involved in the π-hydrogen-bonding interaction with the phenyl group while the 3457 cm^{-1} band arises due to the free NH group of the F1-π conformer. However, the carboxylic OH group in the E1-C5/E2-C5 conformer is free and appears at 3577 cm^{-1} (Figure 3.10c). Thus, the carboxylic OH group in the F1-π conformer is red-shifted by 62 cm^{-1} compared to that in the E1-C5/E2-C5 conformer. The C5 hydrogen-bonded NH group in the extended conformer, which appears at 3450 cm^{-1} (Figure 3.10c), is red-shifted by 7 cm^{-1} in comparison to the free NH group in the F1-π conformer. Similar stretching frequency value of the amide NH group involving in C5 hydrogen-bond has been reported in the gas phase spectroscopic studies of several dipeptides and model peptides.[1, 159, 160, 165, 167] The red-shift observed in the C5 hydrogen-bonded NH group of Z-Gly-Pro-OH studied here is also consistent with that reported in the gas phase studies of similar molecular systems.[1, 159] It is interesting to point out that the solution phase FT-IR study of the C5 interaction in several peptides reports the N-H stretching frequency in the 3400-3420 cm^{-1} region.[87, 146, 192]

The C5 hydrogen-bond and O-H…π interaction observed in the E1-C5/E2-C5 and F1-π conformers, respectively, of Z-Gly-Pro-OH, are further validated through NBO and NCI analysis discussed later.

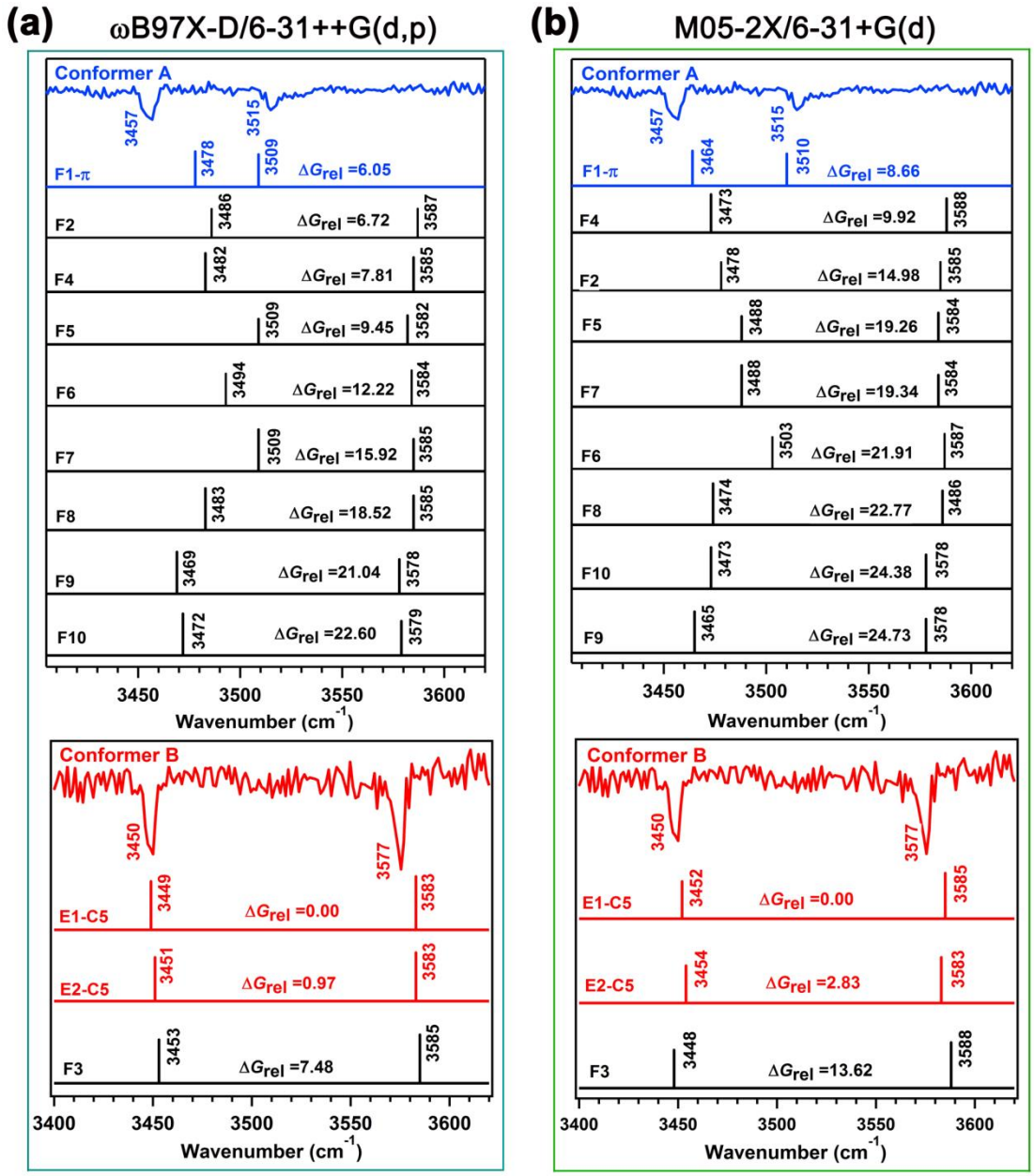

Figure 3.11. A comparison of the experimental IR spectra of the two observed conformers with the (a) ωB97X-D/6-31++G(d,p) and (b) M05-2X/6-31+G(d) level calculated, scaled, harmonic IR spectra of the 12 low energy conformers of Z-Gly-Pro-OH with energy within ΔG_{rel} ~ 24 kJ/mol of the global minimum. The scaling factors used for the harmonic NH and OH stretching frequencies at the ωB97X-D/6-31++G(d,p) level are 0.944 and 0.933, respectively, while those at the M05-2X/6-31+G(d) level are 0.944 and 0.952, respectively. The scaling factors are obtained by scaling the ωB97X-D/6-31++G(d,p) and M05-2X/6-31+G(d) level calculated harmonic frequencies of the most stable conformer of Z-Gly-OH to the experimental frequency reported in the literature.[1]

3.3.4 NBO analysis

According to the NBO formalism, the strength of the hydrogen-bond interaction can be explained in terms of 2[nd]-order perturbative energy ($E^{(2)}_{i\rightarrow j^*}$) for the delocalization of the lone pair electrons in the hydrogen-bond acceptor orbital (i) to the antibonding orbital (j^*) of the hydrogen-bond donor.[170, 193]

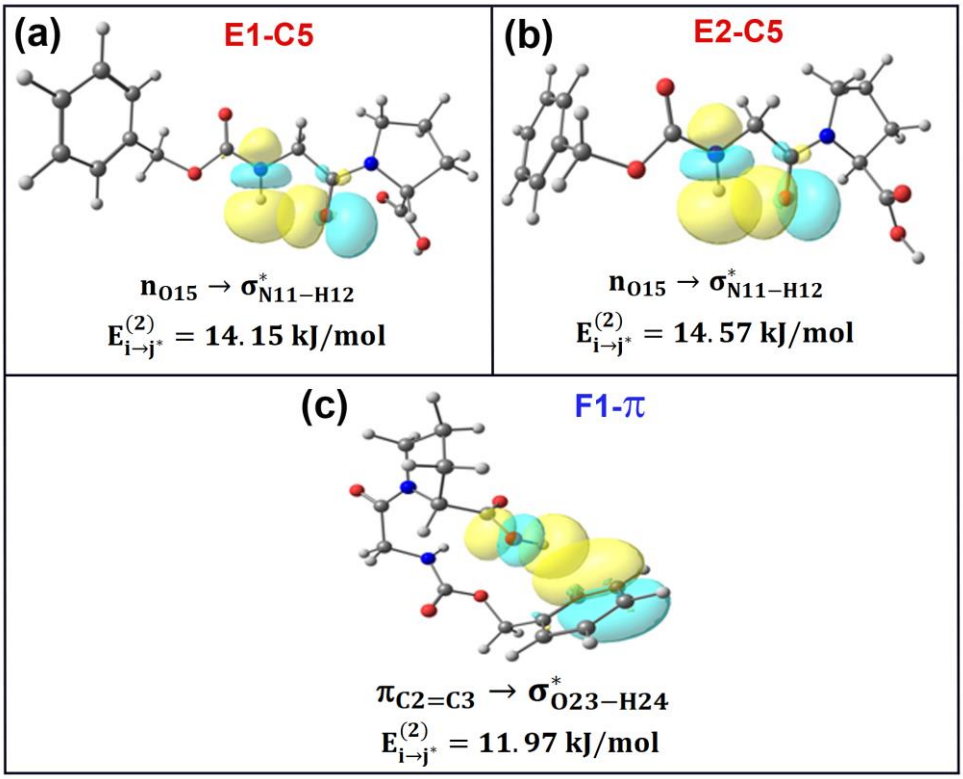

Figure 3.12. (a)-(c) Natural Bond Orbital (NBO) views of the experimentally observed conformers of Z-Gly-Pro-OH showing the interactions between various orbitals as well as their respective second order perturbation energy ($E^{(2)}_{i\rightarrow j^*}$) values. See Figure 1 for atom numbering scheme.

Figure 3.12a-b shows the NBO overlap as well as $E^{(2)}_{i\rightarrow j^*}$ values for the delocalization of the p-type lone pair electrons (n_p) on the oxygen atom of the C=O group into the σ^* orbital of the N-H group i.e. $n_p(O)\rightarrow\sigma^*_{N-H}$ interaction in the C5 hydrogen-bond present in the E1-C5 and E2-C5 conformers of Z-Gly-Pro-OH, respectively. Interestingly, the $E^{(2)}_{i\rightarrow j^*}$ values for the C5 hydrogen-bond in the E1-C5 (14.15 kJ/mol) and E2-C5 (14.57 kJ/mol) conformers are similar.

Raines and co-workers have shown that the C5 hydrogen-bond in the β-sheet structures of proteins indeed involves delocalization of the p-type lone pair on the carbonyl oxygen into the neighbouring N-H group of the same residue of the peptide backbone. Figure 3.12c shows the NBO overlap and $E^{(2)}_{i\rightarrow j^*}$ values for the delocalization of the electrons in the $\pi_{C=C}$ orbital of the phenyl group into the σ^* orbital of the O-H group, i.e. $\pi_{C=C}\rightarrow\sigma^*_{O-H}$ interaction in the O-H…π hydrogen-bond in the F1-π conformer of Z-Gly-Pro-OH. It should be noted that the $E^{(2)}_{i\rightarrow j^*}$ value for the C5 hydrogen-bond in the extended conformer is slightly larger than that for the O-H…π hydrogen-bond in the folded conformer (F1-π).

3.3.5 Non-Covalent Interaction (NCI) calculations

Recently, NCI calculations have been found to be very useful for visualization of weak non-covalent interactions present in molecules as well as weakly bound molecular complexes.[137, 138] NCI calculations show the co-relation between reduced density gradient (RDG) and electron density (ρ) for identification and characterization of favorable or unfavorable interactions of different strength in a semi-quantitative and visual manner.[137] The RDG is represented as follows:

$$RDG = \frac{1}{2(3\pi^2)^{1/3}}\frac{|\nabla\rho|}{\rho^{4/3}}$$

The RDG between the interacting atoms changes when a weak inter- or intra-molecular interaction present in the molecular systems produces density critical points creating iso-surface. The sign of the Laplacian of the density (ρ), $\nabla^2\rho$, is a widely used tool to distinguish the strength of the interactions. The Laplacian is decomposed into three components along the principal axis of maximum variations. These components are the eigenvalues of the electron density Hessian matrix, such that $\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3$. At the nuclei, the density reaches maxima and all the three eigen values are negative. The weaker non-covalent interactions, whether

bonding or We have generated the RDG isosurfaces using the NCIPLOT program[137] and visualized with VMD.[194] A density cut-off of 0.05 a.u. has been applied and the NCI figures have been generated with an isosurface value of 0.042 and colored in the [-4.00, 4.00] a.u. $\boldsymbol{\rho}$ range. Here, the NCI calculations are performed only for the experimentally observed conformers of Z-Gly-Pro-OH, i.e. E1-C5, E2-C5, and F1-π to confirm the presence of the non-covalent interactions revealed by the NBO calculations.

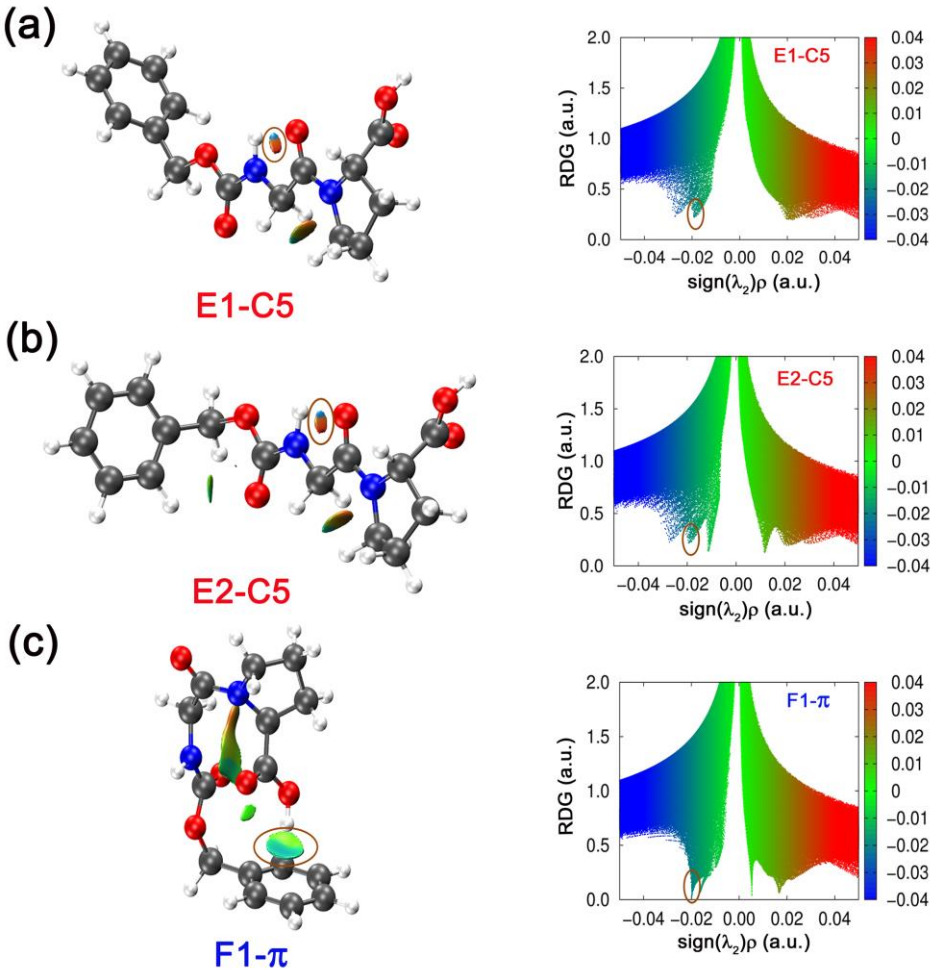

Figure 3.13. (a)-(c) The NCI isosurface and plots of RDG as a function of sign(λ_2)ρ for the lowest energy conformers (E1-C5, E2-C5 and F1-π) of Z-Gly-Pro-OH. The surfaces are colored on a blue-green-red scale according to the values of sign(λ_2)ρ ranging from -0.05 to 0.05 a.u. Blue color indicates strong attractive interactions and red indicates strong repulsive interactions. The geometries are obtained from ωB97XD/631++g(d,p) level.

Figure 3.13a-b shows the NCI isosurface and RDG vs sign (λ_2).ρ plot for the E1-C5 and E2-C5 extended conformers, respectively, while the same for the F1-π conformer has been provided in Figure 3.13c. The RDG minima at a similar value (~ -0.02) of sign (λ_2).ρ for both the extended and folded conformers indicate that the strength of the C5 hydrogen-bond and OH…π interaction present there, respectively, is quite similar.

3.3.6 Structure, energetic and Frequency comparison of high energy conformers.

The structure of 12 low energy conformers lying within the energy limit of 6 kcal/mol with respect to the global minimum conformer has been shown in figure 3.14 after optimization at ωB97X-D/6-31++G(d,p) level. The conformers having extended conformation having an intra-residue C5 hydrogen bond in glycine residue are the most stable conformers.

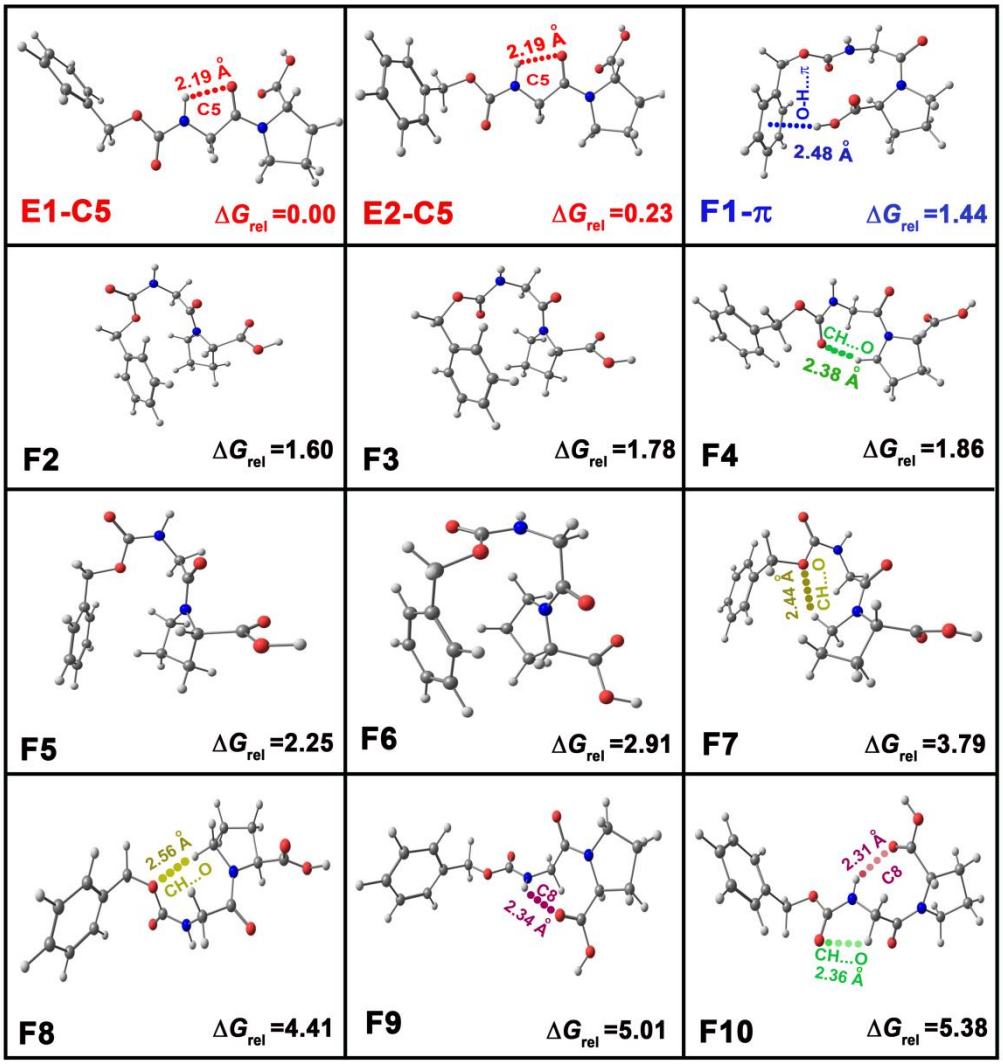

Figure 3.14. Optimized geometries of the low energy conformers of Z-Gly-Pro-OH with relative Gibbs free energy (ΔG_{rel}) values within 6 kcal/mol from the global minimum calculated at the ωB97X-D/6-31++G(d,p) level of theory. ΔG_{rel} values are calculated at 300 K.

3.3.7 Crystal structure analysis of Z-Gly-Pro-OH

The crystal structure of the Z-Gly-Pro-OH molecule was grown in a 2 ml glass vial by keeping the Z-Gly-Pro-OH solution in ethyl acetate and n-hexane for slow evaporation for 5 days. The X-ray diffraction of the crystal of Z-Gly-Pro-OH was performed using APEX(II) DUO CCD

diffractometer. The X-ray diffraction data were collected at 100 K. The structure obtained after diffraction of the Z-Gly-Pro-OH molecule has been presented in figure 3.1

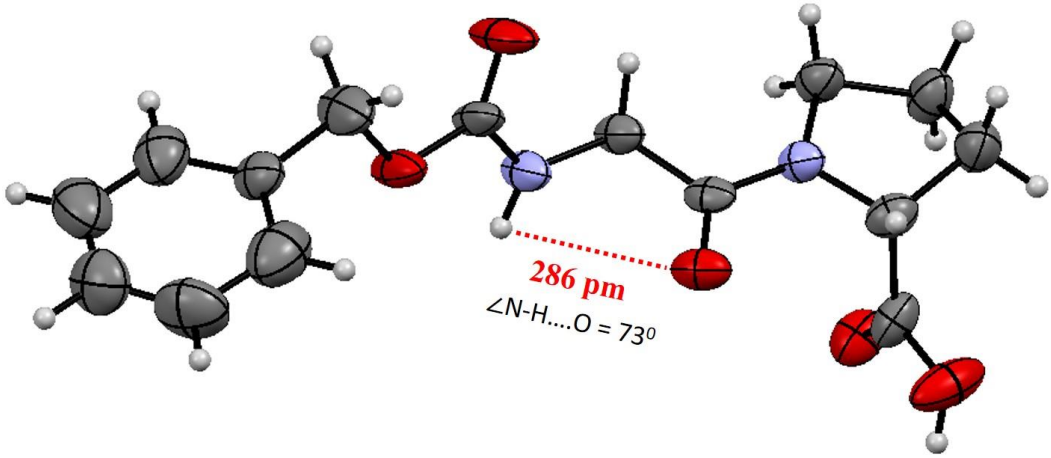

Figure 3.15. ORTEP representative structure of Z-Gly-Pro-OH crystal. Thermal ellipsoid has been kept at 50% level.

The crystal structure of the Z-Gly-Pro-OH molecule shows an extended structure like one observed experimentally in gas-phase spectroscopy although C5 interaction is missing in the case of crystal. This could be due to the tight packing of peptide molecules in the crystal packing. The hydrogen bond angle (73^0) and bond distance (286pm) in the crystal don't allow the formation of intra-residue C5 hydrogen bond in the case of the Z-Gly-Pro-OH molecule.

3.4 Conclusion

Conformational preferences of Z-Gly-Pro-OH have been investigated in an isolated gas-phase experiment using resonant 2-photon ionization and IR-UV double resonance spectroscopy techniques combined with quantum chemistry calculations. Two conformers of the peptide observed in the experiment are assigned to E1-C5/E2-C5 and F1-π structures based on their Gibbs free energies at 300 K or higher as well as vibrational frequencies. It has been found that populations of the conformers of the molecules in the laser desorption experiment are best described in the 300-500 K temperature range. The E1-C5 and E2-C5, which have an extended

β-strand type structure stabilized by an intra-residue C5 hydrogen-bonding interaction (N-H…O), are nearly isoenergetic while E1-C5 is the global minimum. The F1-π conformer, which is about 1.5 kcal/mol higher in energy from the global minimum, has a folded structure stabilized by an O-H…π interaction. Interestingly, both the extended and folded structures of the peptide are observed in the experiment. The higher stability of the extended structure over the folded one obtained from the Gibbs free energy calculation at 300 K or higher temperature indicates a significant contribution of larger entropy in the stabilization of the former structure. The NBO calculations indeed show that the C5 hydrogen-bond interaction in the amide backbone involves delocalization of the electrons in the p-type lone-pair orbital of the carbonyl oxygen atom to the σ^* orbital of the N-H group. The presence of weak C5 hydrogen-bond and O-H…π interaction in the two observed structures of the peptide is further validated by the NCI calculations. The solution phase FTIR suggests that the Z-Gly-Pro-OH molecule can either exist in a folded or extended conformation. The crystal structure of the peptide molecule shows extended conformation devoid of an intra-residue C5 hydrogen bond.

Chapter 4

Sequence-dependent Folding Motifs of the Secondary Structures of Gly-Pro and Pro-Gly Containing Oligopeptides

4.1 Introduction

Primary structures of proteins are based on the sequence of the amino acid residues while the secondary structures are formed through hydrogen bonding interactions of the backbone amide hydrogens and carbonyl oxygens in the polypeptide chains.[146, 195, 196] Although protein folding is a global phenomenon, the secondary structures, which are local in nature, play an important role in the stability and shape of the proteins.[139, 142, 196-199] The secondary structures of proteins and peptides are generally categorized as helices, sheets, and turns depending on their hydrogen-bonding network along the backbone peptide chain.[2, 15, 30, 52, 144, 200]

Turns are one of the most important secondary structures in proteins as they change the overall direction of the polypeptide chains and also provide their compact folded structures.[200] Based on the hydrogen bonding patterns along the polypeptide chains, commonly observed turn structures are γ, β, and α, which are classified in terms of the hydrogen-bonded rings C7, C10, and C13, respectively, formed through the backbone N-H and C=O groups of the $i \rightarrow i+2$, $i \rightarrow i+3$, and $i \rightarrow i+4$ residues, respectively.[15, 16, 64, 201, 202] β-turns are one of the most common secondary structures in proteins and peptides while the γ and α turns are less prevalent.[69, 203] β-turns are very often present in the loop region of the β-hairpin structure made of two anti-parallel β-strands, which are again the building blocks of the β-sheet structures.[204]

Propensities of distinct turn structures in proteins and small peptides are closely related to the specific amino acid residues present there as well as their particular sequences. The specific sequence of the residues permits the polypeptide chain to acquire the requisite Ramachandran (φ, ψ) angles to form the particular turn.[83, 205] It has been reported in the literature that the most common amino acid residues which favor the β-turn formation in proteins and polypeptides are Gly, Pro, Asn, and Asp while the propensity of these residues in helices and β-sheet is negligible.[140] Moreover, Gly and Pro are unique among all the amino acid residues as there is

no side chain in the former one while the side chain of the latter one forms a five-membered ring with the nitrogen atom in the peptide bond. Thus, the combination of the most flexible amino acid Gly and the most rigid amino acid Pro is found to be favorable for the formation of the β-turn and β-bend structures in proteins.[70, 140] However, it has been reported that the Pro-Gly sequence has a higher propensity for the formation of the β-turn conformation in proteins but it is very low for the Gly-Pro sequence, which prefers to form an extended polyproline II type conformation.[21, 83, 97] It is also demonstrated that the β-turn is the conformational requirement of the enzymatic hydroxylation of the specific proline residue of the Pro-Gly segment in procollagen for formation of the stable triple helical structure of collagen.[14]

Balaram and co-workers have illustrated the efficient formation of β-hairpin conformations in the crystals of Hexa- to decapeptides exploiting the ^{D}Pro-Gly segment at the central position to adopt type II′ β-turn conformations.[67, 98, 206, 207] Gellman and co-workers have also validated that peptides having ^{D}Pro-Gly segment have a propensity of β-hairpin formation through β-turn at the loop.[208] Scheraga and co-workers have studied the conformational properties of end-protected Gly-Pro as well as –Pro-Gly- dipeptides using FTIR, CD, and temperature-dependent NMR spectroscopy. They have concluded that the dominant conformation observed for the Gly-Pro peptide is β-strand or extended polyproline II type while the same for the Pro-Gly is β-turn.[21] They reported from their data that other minor conformations of Gly-Pro and Pro-Gly dipeptides were also present in their experiments. Later, it has been shown from the X-ray crystallography study that both t-Boc-Gly-^{L}Pro-OH and t-Boc-Gly-^{L}Pro-OBz peptides indeed exhibit extended conformations.[209] Recently, Kang et al. have demonstrated from DFT and implicit solvation model calculation that ^{D}Pro-Gly containing peptides have a strong preference for β-hairpin and β-turn structures.[99]

Intrinsic folding motifs of small peptides have also been explored using isolated gas phase laser spectroscopy techniques to probe different low-energy conformations as well as various non-covalent interactions present there.[8, 9, 23, 25, 81, 92, 103, 122, 210-214] However, gas-phase conformational studies on peptides consisting of Gly and Pro residues are explored to some extent in the literature but a detailed study combining gas-phase and condensed phase studies is required.[163, 215-217] An advantage of gas-phase spectroscopy over the solution or solid-state spectroscopy is that one can obtain quantitative information on the structures of several low-energy conformations of small peptides which are important for the secondary structures of various neighbouring states including the global minimum of polypeptides or proteins.[8, 23, 218,1, 27, 159, 219] More specifically, gas phase study of the structures of small peptides will resemble the folding motifs of the secondary structures of polypeptides or proteins, which are local and embedded in a hydrophobic environment.

In this work, we have studied sequence-dependent folding motifs of capped dipeptides, Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe employing gas-phase UV/IR laser spectroscopy, quantum chemistry calculations, solution-phase IR, NMR spectroscopy and X-ray crystallography. The aim is to connect the data on the conformations of the peptides revealed from the gas phase with those from the condensed phase.

4.2 Methods

4.2.1 Experimental methods

The end-protected dipeptides Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe were synthesized and characterized following standard methods provided in the supporting information.[175] Solution phase IR and 1D/2D NMR spectra of the peptides were measured in a dilute $CDCl_3$ solution using standard FTIR-Bruker Vertex 70 spectrometer and 400 MHz NMR (Bruker-400) spectrometer, respectively, to determine the structures of the

conformations. Single crystal X-ray diffraction (XRD) of the peptides is performed using APEX(II) DUO CCD diffractometer. A detailed description of the solution phase spectroscopy and single-crystal XRD analysis has been provided in the supporting information.

Conformation-specific electronic and IR spectra of the peptides were measured using various gas-phase laser spectroscopic techniques in a home-built jet-cooled laser desorption Time of Flight (TOF) mass spectrometer, which has been described in detail, previously and also in supporting information.[171, 173, 174, 220-222] In brief, a pellet consisting of the peptide sample and graphite powder was mounted in a rotating sample holder placed in the vacuum chamber. A laser beam of 532 nm (Nd:YAG laser, Continuum, Minilite-I, 10 Hz, 10 nanoseconds) was mildly focused on the rotating pellet to desorb the peptide molecules, which were seeded into the supersonic expansion of Ar (~5 bar) carrier gas. The molecular beam of the peptide was ionized by the second harmonic output (0.2–0.3 mJ) of a tunable dye laser (ND6000, Continuum) pumped by the frequency-doubled output of an Nd:YAG laser (10 nanoseconds, 10 Hz, Surelite II-10, Continuum) and the ions were analyzed in the TOF mass spectrometer (Jordan TOF, USA).

Mass-selected electronic spectra of the peptides were recorded using the one-color resonant 2-photon ionization (1C-R2PI) method. Resonant ion dip infrared (RIDIR) spectroscopy was used to measure the IR spectra of the peptide molecules. A tunable IR laser (Laser Vision, pulse energy ~ 4-5 mJ, resolution ~ 2.5 cm^{-1}) based on an optical parametric oscillator (OPO)/ optical parametric amplifier (OPA) pumped by an unseeded Nd: YAG laser (Continuum, Surelite II-10, 10 nanoseconds, 10 Hz) was used for the IR spectroscopy experiment. IR-UV hole-burning and UV-UV hole-burning spectroscopy were performed to discriminate the presence of different conformers present in the gas-phase experiments.

4.2.2 Computational methods

Quantum chemical calculations were performed to predict the structures of the conformers of the peptides observed in the gas phase spectroscopy experiments. Conformation-specific vibrationally resolved IR spectra obtained from the experiment were compared with the theoretical IR spectra of low-energy conformations. Various types of non-covalent interactions present in the conformers were also visualized through the theoretical calculations. Initially, a large number of conformers (~140) of both Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe peptides were generated using a conformational search software called CONFLEX based on the MMFF94 force field. [179, 223] The conformers within the energy range of 40 kJ/mol with respect to the lowest energy one were grouped by applying a cut-off of 0.4 kJ/mol. About 60 groups with each containing a few conformers were formed for both the peptides. The conformer with the lowest energy in each group was then subjected to preliminary optimization using the HF/6-31G(d) level of theory. A few more conformers within a group showing different conformation were also considered for optimization at HF/6-31G(d). Overall 89 and 83 conformers were optimized at HF/6-31G(d) level of theory for Gly-Pro and Pro-Gly sequence, respectively. Afterward, 52 and 68 non-redundant conformers were re-optimized at the M05-2X/6-31+G(d), B3LYP-D3/def2TZVPP, B97-D3/def2TZVPP, and ωB97XD/def2TZVPP level of theory for Gly-Pro and Pro-Gly sequence, respectively. The conformers within 12 kJ/mol relative energy compared to the global minimum structure after optimization at the B3LYP-D3/def2TZVPP level of theory were further optimized at the M06-2X/6-311++G(2d,2p) level of theory. All calculate ions were performed using Gaussian 09 package[135] and except at M06-2X/6-311++G(2d,2p) level of theory which was carried out using Gaussian 16 package.[180]

4.3 Results and Discussions

4.3.1 Synthesis procedures and characterization of the peptides

Synthesis method

Both the peptides Boc-^{D}Pro-Gly-NHBn-OMe and Boc-Gly-^{D}Pro-NHBn-OMe were synthesized by a standard synthetic procedure reported in the literature.[175]

Synthetic procedure of Boc-D-Pro-Gly-NHBn-OMe: Commercially available Boc-D-Pro-COOH was coupled with N-Hydroxysuccinimide (NHS) in the presence of N,N'-Dicyclohexylcarbodiimide (DCC), tetrahydrofuran (THF) in a round bottom flask kept at ice condition and left at room temperature for 12 h. The product was then coupled with Glycine in the presence of $NaHCO_3$, and THF and then acidified with 10% HCl and this resulted in the formation of Boc-^{D}Pro-Gly-OH. Boc-^{D}Pro-Gly-OH was then mixed with 4-Methoxybenzylamine in the presence of EDC.HCl, HOBt, DIPEA, and DMF at 0^0 C. The final product was received after 12 hours of reaction at room temperature. The final compound was purified by column chromatography using ethyl acetate and hexane as a solvent in a 65% yield. The synthetic procedure has been shown also in Figure S1 provided below.

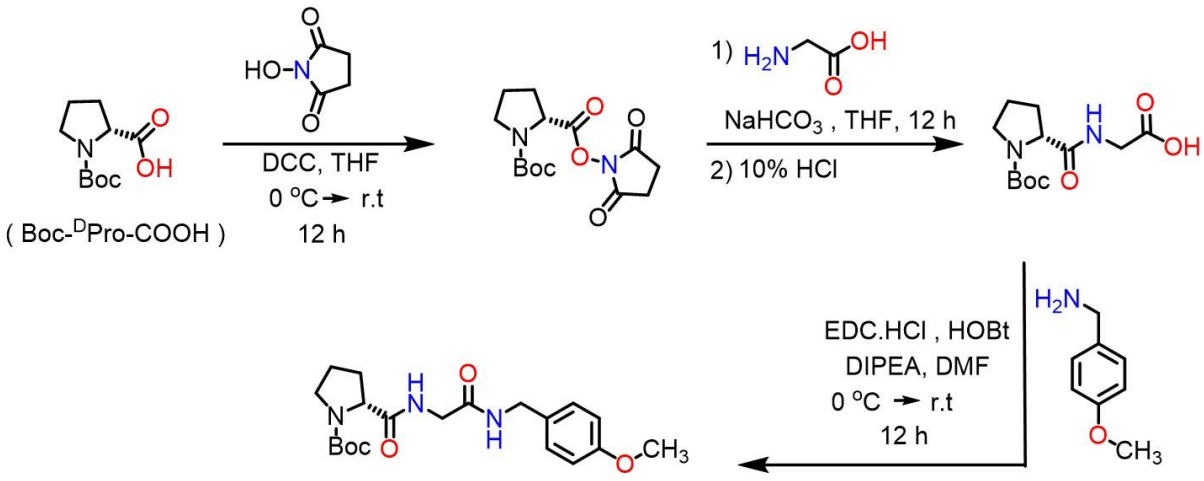

Figure 4.1. Synthetic Scheme of Boc-^{D}Pro-Gly-NHBn-OMe

Synthetic procedure of Boc-Gly-^{D}Pro-NHBn-OMe: Commercially available Boc-Gly-COOH was coupled with N-Hydroxysuccinimide (NHS) in the presence of N, N'-Dicyclohexylcarbodiimide (DCC), tetrahydrofuran (THF) in a round bottom flask kept at ice condition and then left at room temperature for 12 h. The product was then coupled with the Boc-^{D}Pro-OH in the presence of $NaHCO_3$, and THF and then acidified with 10% HCl and this resulted in the formation of Boc-Gly-^{D}Pro-OH. The product Boc-Gly-^{D}Pro-OH was then coupled with 4-Methoxybenzylamine in the presence of EDC.HCl, HOBt, DIPEA, DMF at 0^0 C. Afterwards, the reaction was maintained at room temperature and kept for 12 hours with the continuation of the stirring. The final compound was purified by column chromatography using ethyl acetate and hexane as a solvent with a 72% yield. The synthetic procedure has also been depicted in Figure S2 provided below.

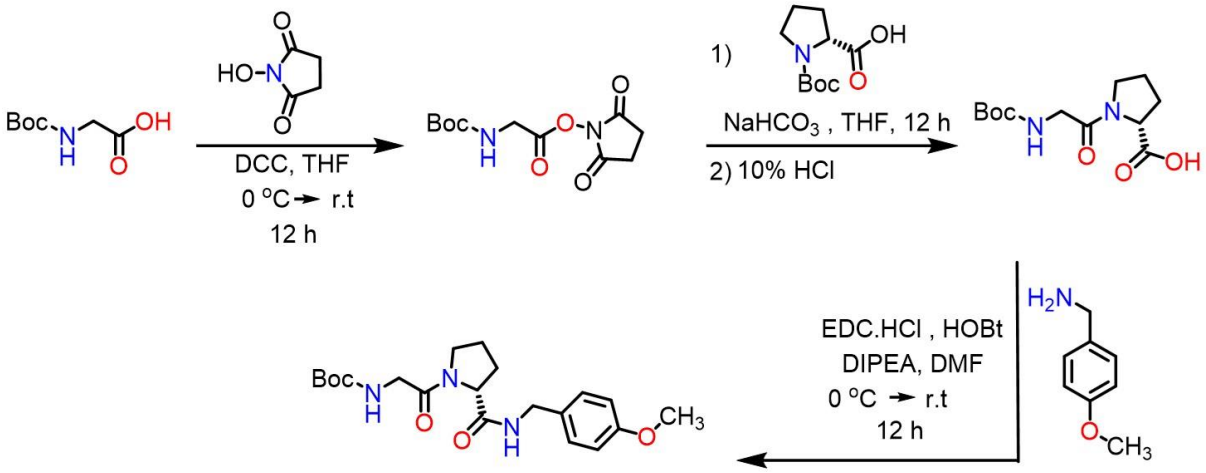

Figure 4.2. Synthetic Scheme of Boc-Gly-^{D}Pro-NHBn-OMe

4.3.2 ^{1}H NMR Characterization

^{1}H NMR of Boc-^{D}Pro-Gly-NHBn-OMe (400 MHz, $CDCl_3$, 298.15 K): δ 6.80-7.50 (m,6H, NH_{Gly}, NH_{NHBn}, CH^{Ar}) 3.4-4.5 (m, 10H, CH_{Gly}, OCH_3, CH_{NHBn}, CH_{Pro}) 1.8-2.2 (m,4H, CH_{Pro}) 1.37 (s, 9H, CH_{Boc})

^{1}H NMR of Boc-Gly-^{D}Pro-NHBn-OMe (400 MHz, $CDCl_3$, 298.15 K): δ 6.80-7.50 (m,5H, NH_{NHBn}, CH^{Ar}), 5.37(s, 1H, NH_{Gly}) 3.4-4.5 (m, 10H, CH_{Gly}, OCH_3, CH_{NHBn}, CH_{Pro}) 1.8-2.2 (m,4H, CH_{Pro}) 1.43 (s, 9H, CH_{Boc})

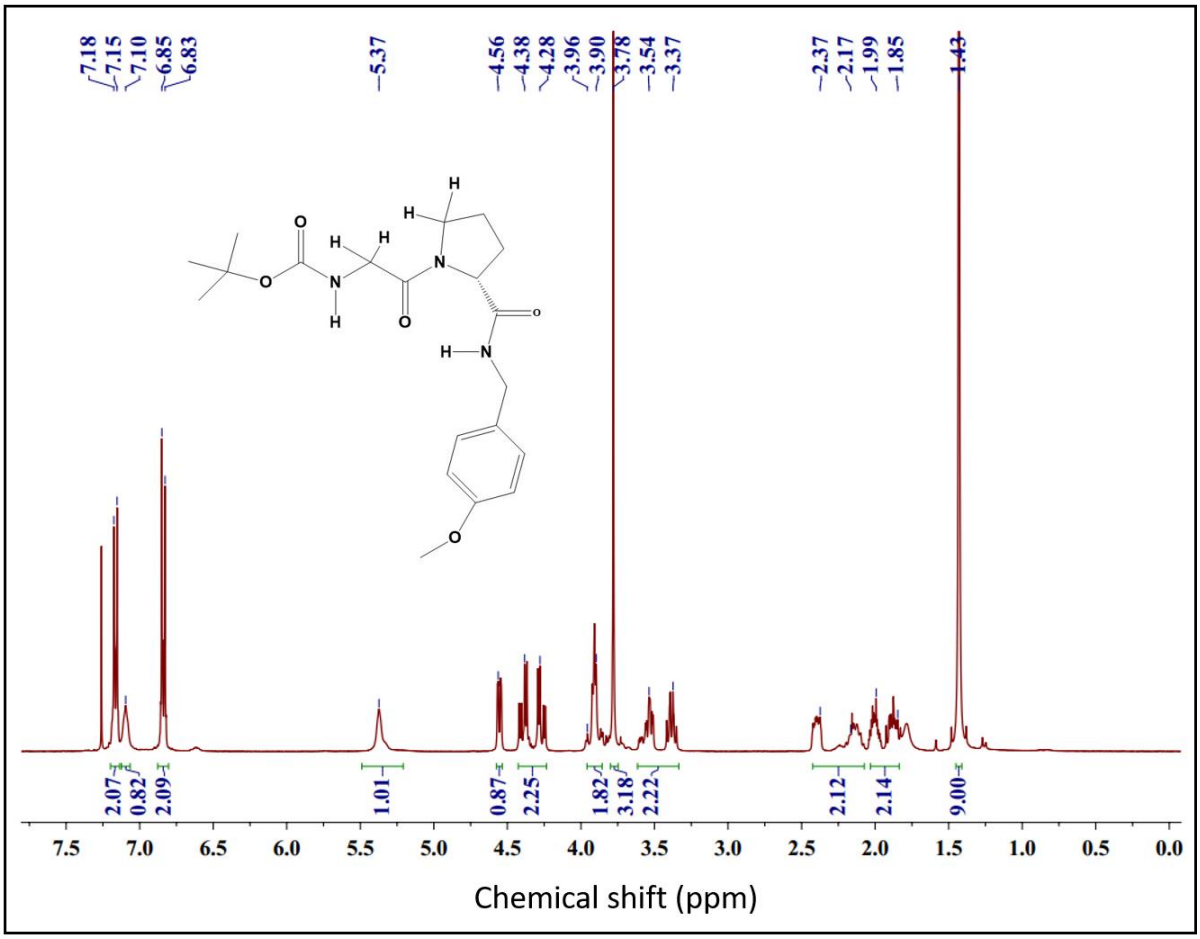

Figure 4.3. ^{1}H NMR spectrum of Boc-Gly-^{D}Pro-NHBn-OMe recorded in $CDCl_3$ solvent.

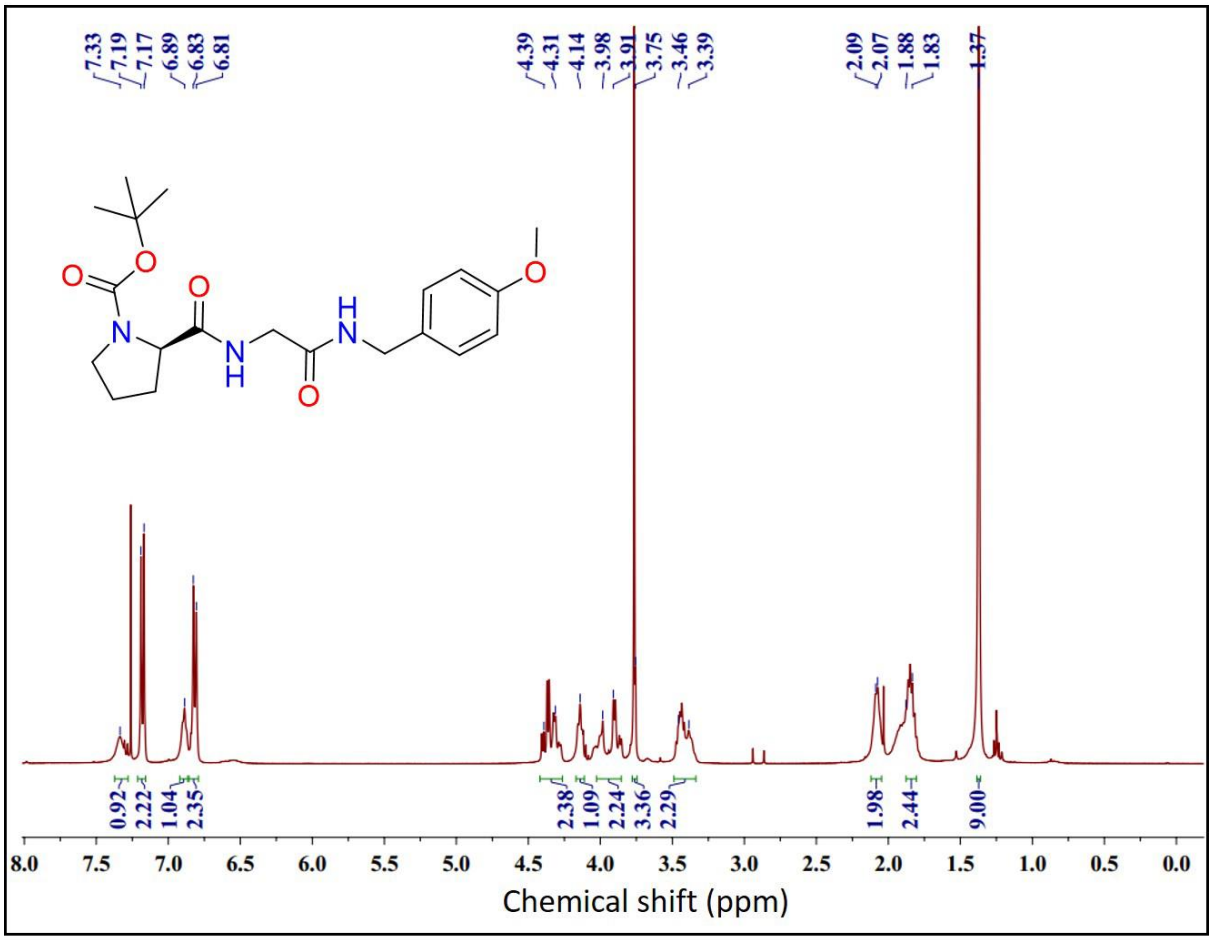

Figure 4.4. ^{1}H NMR spectrum of Boc-^{D}Pro-Gly-NHBn-OMe recorded in $CDCl_3$ solvent.

4.3.3 HRMS spectra of peptides

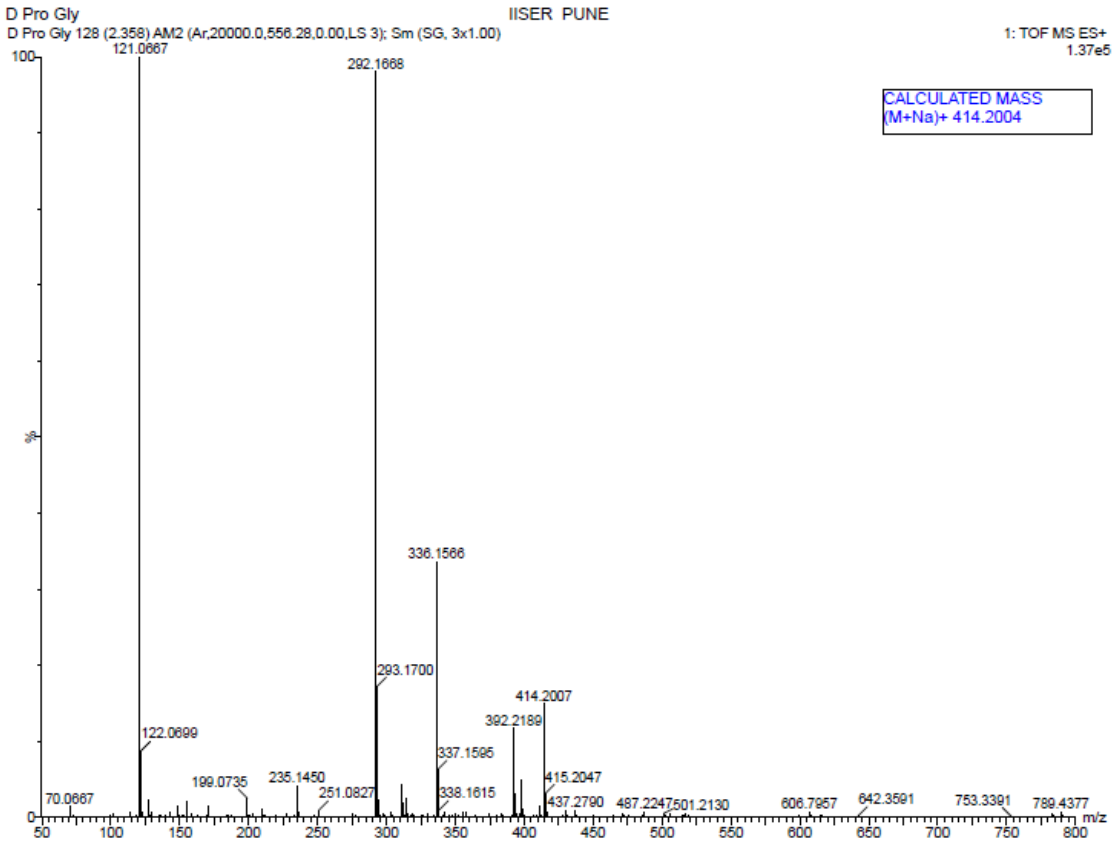

Figure 4.5. HRMS mass spectrum of Boc-^{D}Pro-Gly-NHBn-OMe.

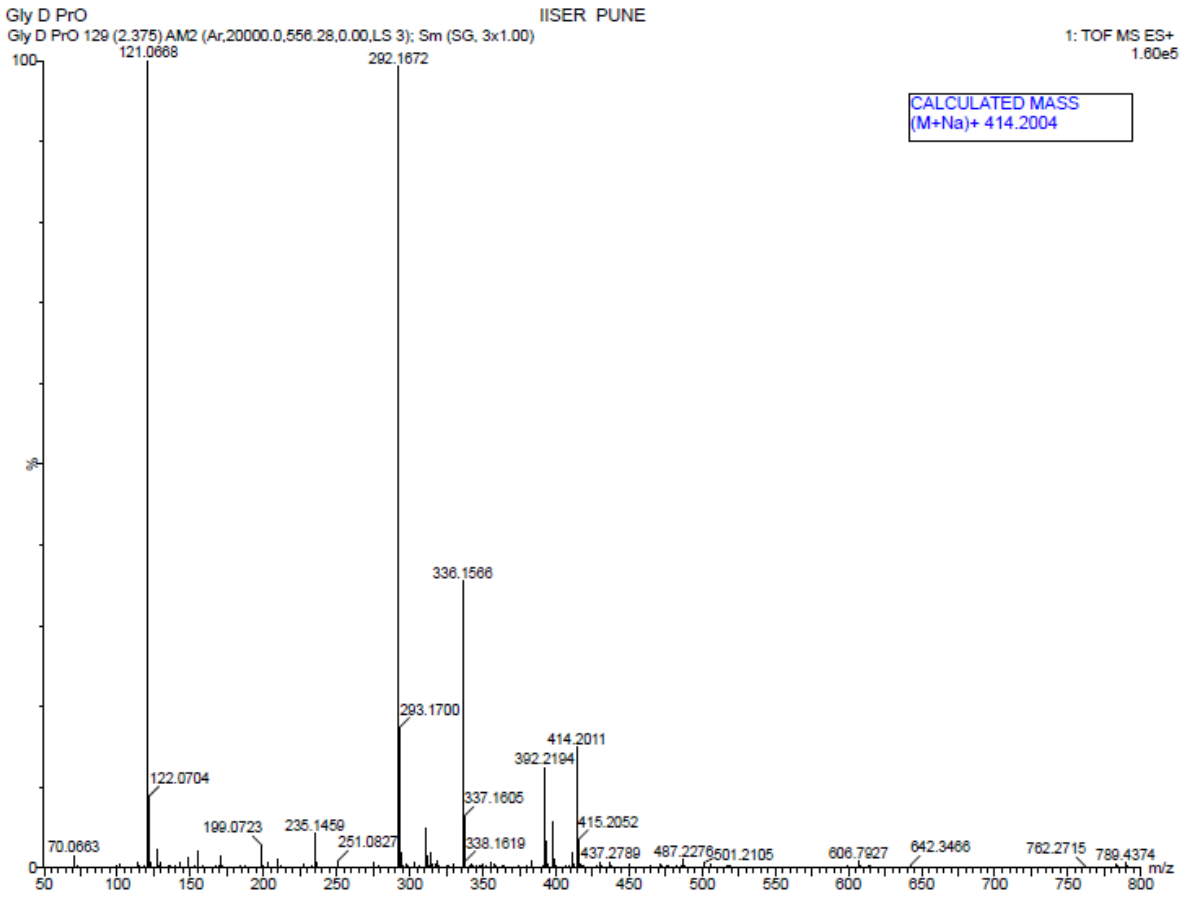

Figure 4.6. HRMS mass spectrum of Boc-Gly-^{D}Pro-NHBn-OMe.

4.3.4. Electronic spectroscopy of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe

Figure 4.7 depicts the chemical structures of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe peptides with marking of the Ramachandran angles (φ,ψ). The NHBn-OMe group at the C- terminal not only acts as a protecting group but also functions as a chromophore for the electronic excitation.[92]

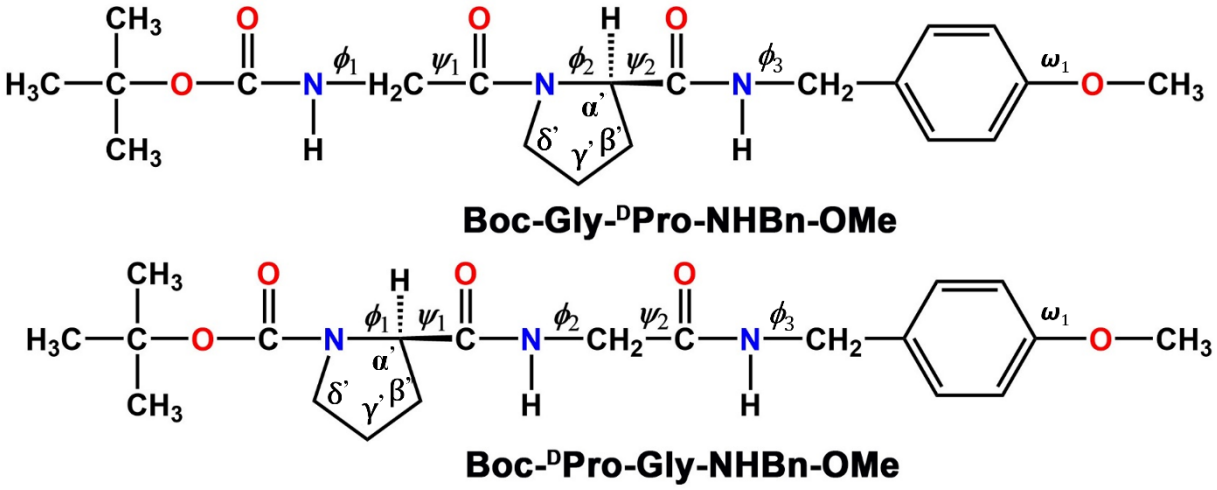

Figure 4.7. Chemical structures of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe marking the Ramachandran angles.

The mass-selected electronic spectrum of Boc-Gly-^{D}Pro-NHBn-OMe measured using the 1C-R2PI method is provided in Figure 4.8a, which shows a large number of sharp well-resolved bands in the 35250-35600 cm^{-1} region. Figure 4.8b-d presents UV-UV hole-burning spectra by probing the electronic bands in Figure 4.8a marked with A_0^0, B_0^0, and C_0^0, respectively. Three conformers of the peptide are observed in the experiment and the origin bands, A_0^0, B_0^0, and C_0^0, show up at 35416 cm^{-1}, 35276 cm$^{-1,}$ and 35267 cm^{-1}, respectively. The conformer A is

Figure 4.9a depicts the electronic spectrum of Boc-^{D}Pro-Gly-NHBn-OMe recorded using the 1C-R2PI technique. The spectrum shows many sharp electronic bands in the 35400-35550 cm^{-1} region. The discrimination of the bands due to the presence of different conformers of this peptide is performed using IR-UV hole-burning spectroscopy. The IR-UV hole-burning spectra shown in Figure 4.9b-3c are obtained by probing the N-H vibrational bands of conformers A rich in vibronic structures while the conformers B and C have very limited and weak vibronic features. and B, respectively. The hole-burning spectroscopy reveals that the electronic transitions of the two observed conformers A and B of the Pro-Gly peptide are quite

overlapping unlike those of the Gly-Pro peptide. The origin bands of the two conformers of the Pro-Gly peptide, $A_0^{0,}$ and B_0^0, appear at 35416 cm^{-1} and 35440 cm^{-1}, respectively.

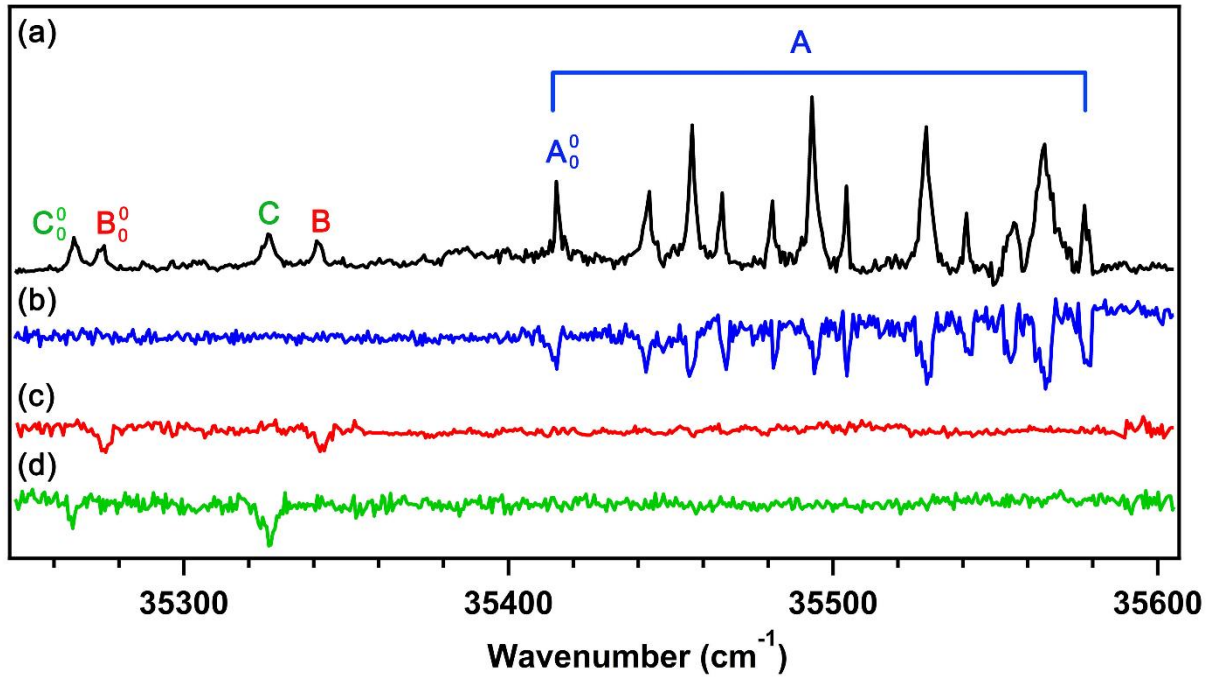

Figure 4.8. (a) Electronic spectrum of Boc-Gly-^{D}Pro-NHBn-OMe measured using 1C-R2PI spectroscopic technique. (b)-(d) UV-UV hole-burning spectra of Boc-Gly-^{D}Pro-NHBn-OMe measured by fixing the pump UV laser at 35416, 35276 and 35267 cm^{-1}, respectively.

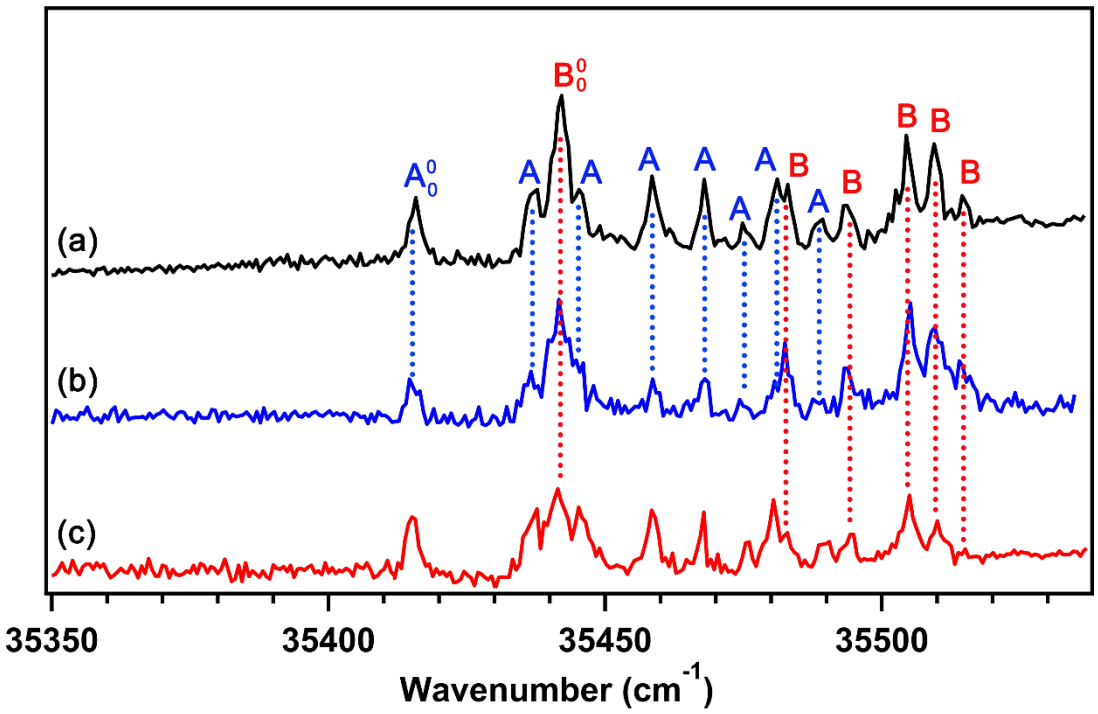

Figure 4.9. (a) Electronic spectrum of Boc-^{D}Pro-Gly-NHBn-OMe measured using 1C-R2PI spectroscopic technique. (b)-(c) IR-UV hole-burning spectra of Boc-^{D}Pro-Gly-NHBn-OMe measured by fixing the pump IR laser at 3346 and 3365 cm^{-1}, respectively.

4.3.5 Conformational landscape of Gly-Pro and Pro-Gly peptides

The energy landscapes of the lowest energy conformers of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe obtained from the M06-2X/6-311++G(2d,2p) level of calculation are provided in Figure 4.10a and b, respectively. The conformers are sorted into different groups in terms of the specific backbone hydrogen bonding interactions present there.

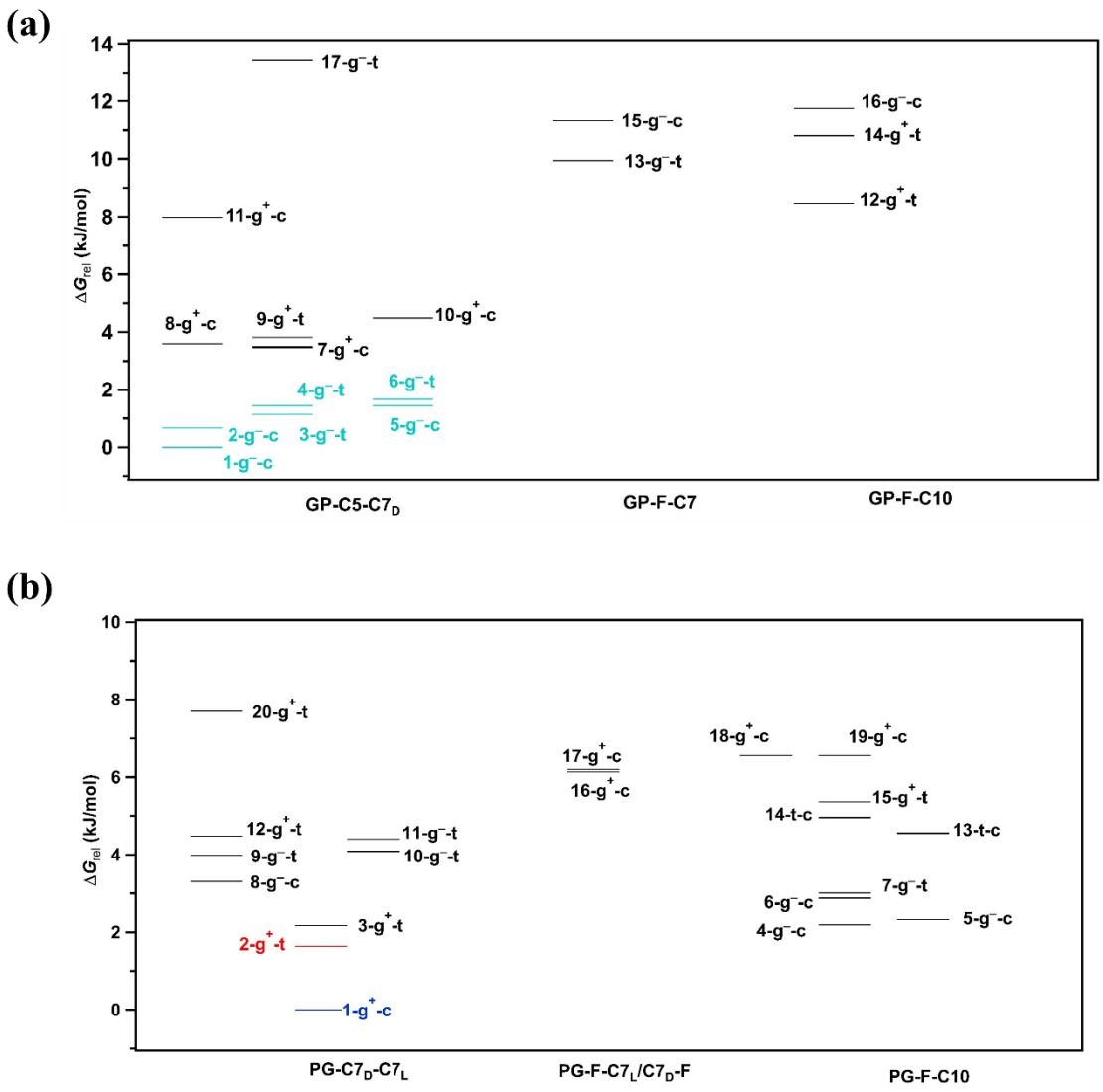

Figure 4.10. Energy landscape of a few low energy conformers of (a) Boc-Gly-^{D}Pro-NHBn-OMe and (b) Boc-^{D}Pro-Gly-NHBn-OMe at 300 K calculated at the M06-2X/6-311++G(2d,2p) level of theory based on zero-point energy corrected Gibbs free energies (ΔG_{rel}) relative to the most stable conformer. The conformers, which are color coded, are observed in the experiment. See the text for the assignment.

For the Gly-Pro peptide, the conformers are arranged into three main categories namely C5-C7, F-C7, and F-C10 while the Pro-Gly peptide has C7-F/F-C7, C7-C7, and F-C10 structures. Apart from that, the benzyl group orientation gauche + (g+), gauche – (g-) and trans (t) have been used to differentiate the conformers.[167] The OMe group orientation with respect to the benzyl group has been either denoted as cis (c) or trans (t). The nomenclature of the conformers is provided by the hydrogen bonding status of the N-H groups of Gly and Bn i.e. whether they are hydrogen-bonded (C5, C7, etc.) or free (F). The C7 hydrogen bond can have two distinct orientations named $C7_D$ and $C7_L$ based on the Ramachandran dihedral angles.[20, 224] The first N-H group considered in the nomenclature is Gly. In the case of Boc-Gly-^{D}Pro-NHBn-OMe, the global minimum conformer is GP1-C5-$C7_D$-g^--c, where the Gly N-H is C5 hydrogen-bonded and the Bn N-H group is $C7_D$ hydrogen bonded. The global minimum conformer of Boc-^{D}Pro-Gly-NHBn-OMe is PG1-$C7_D$-$C7_L$-g^+-c while the low energy C10 structure (PG4-F-C10-g^--c) is about 2.19 kJ/mol higher in energy from the global minimum. A statistical survey of the crystal structures of Gly-Pro and Pro-Gly containing peptides (Figure 4.19) deposited in the Cambridge Structural Database (CSD) demonstrates that there is a significant number of C5 structures apart from the C7 and C10 structures for the Gly-Pro peptides.[225, 226] On the other hand, the Pro-Gly peptides have greater propensities for the C10 and C7 structures compared to forming a C5 structure.

4.3.6 Gas-phase and solution phase IR spectroscopy of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe

4.3.6.1 Boc-Gly-^{D}Pro-NHBn-OMe

The structures of the observed conformers of the two peptides are determined from the comparison of the experimental gas-phase IR spectra with those obtained from the theoretical calculations. Figure 4.11 (a), (b), and (c) shows the experimental IR spectra of Boc-Gly-^{D}Pro-NHBn-OMe measured in the N-H stretching region by probing the A_0^0, B_0^0, and C_0^0 bands, respectively, in the electronic spectra provided in Figure 4.8a using RIDIR spectroscopy. The theoretical IR spectra [M06-2X/6-311++G(2d,2p)] of the five conformers GP1-C5-C7$_D$-g$^-$-c, GP2-C5-C7$_D$-g$^-$-c, GP3-C5-C7$_D$-g$^-$-t, GP4-C5-C7$_D$-g$^-$-t, GP5-C5-C7$_D$-g$^-$-c and GP6-C5-C7$_D$-g$^-$-t among the low energy conformers which best fit with the experimental IR spectra are provided in Figure 5(d)-(i) respectively along with the corresponding optimized structures. A detailed comparison of the experimental IR spectra with the theoretical IR spectra of the low energy conformers calculated at the M06-2X/6-311++G(2d,2p) level of theory is provided in Figure 4.12. The harmonic N-H stretching frequencies are scaled using a factor obtained from the ratio of the experimental to the theoretical N-H frequency of Z-Gly-OH.[1]

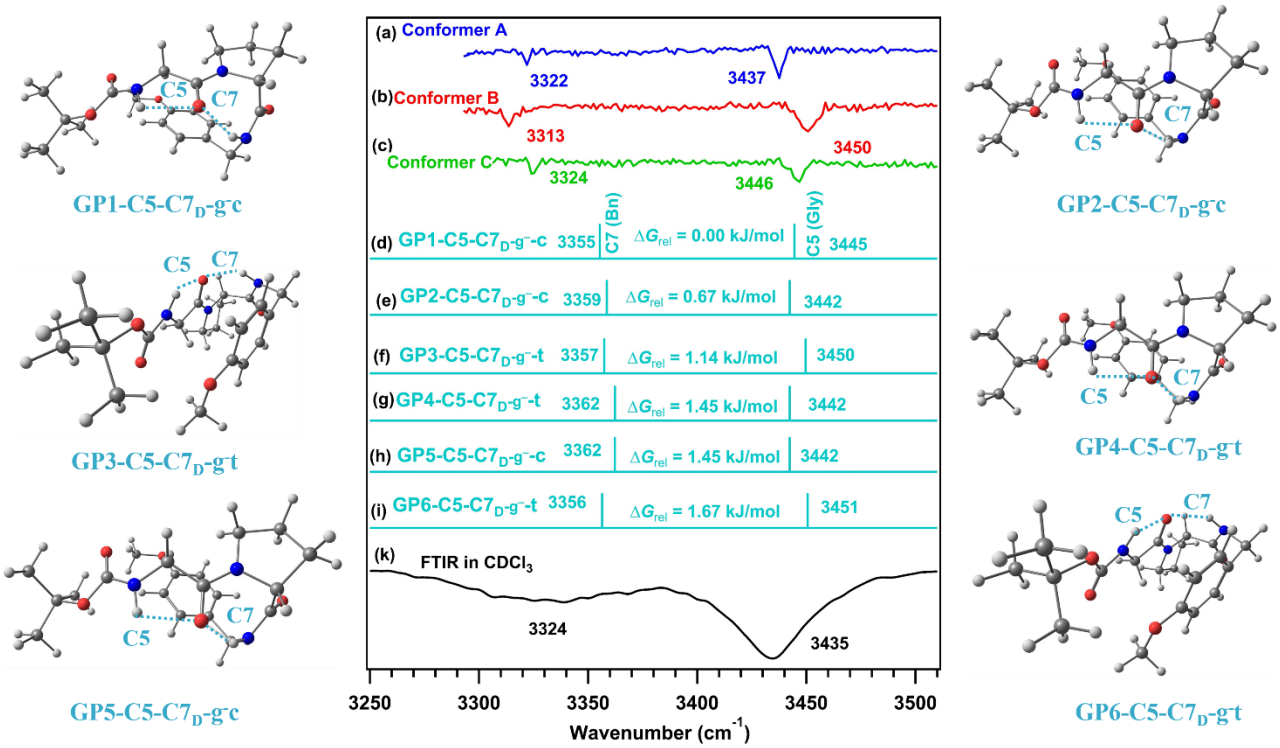

Figure 4.11. (a), (b) and (c) IR spectra of conformers A, B, and C of Boc-Gly-^{D}Pro-NHBn-OMe, respectively, in the N-H stretching region measured in the gas phase using RIDIR spectroscopy. (d)-(i) Scaled theoretical IR spectra of the conformers GP1-C5-$C7_D$-g^--c, GP2-C5-$C7_D$-g^--c, GP3-C5-$C7_D$-g^--t, GP4-C5-$C7_D$-g^--t, GP5-C5-$C7_D$-g^--c and respectively, calculated at the M06-2X/6-311++G(2d,2p) level of theory. The theoretical harmonic N-H stretching frequencies of all the conformers are scaled with a scaling factor of 0.948. (g) Solution phase IR spectrum of the Boc-Gly-^{D}Pro-NHBn-OMe peptide in the N-H region measured in $CDCl_3$ solution.

All the experimental conformers conformer A, Conformer B, and conformer C has their first set of IR bands of around 3437-3450 $cm^{-1,}$ and the second set of IR bands lie in the region of around 3313-3324 cm^-. Three experimentally observed conformers could not be assigned to any specific theoretically calculated structure. Any of the six theoretically calculated low-energy conformers cab be assigned to the experimental conformers. A comparison of experimental IR spectra with theoretically scaled conformers suggests that all three conformers have an extended C5-$C7_D$ conformation. The experimental IR set of bands around 3313-3324 could be assigned to the $C7_D$ hydrogen-bonded N-H group of the NHBn group whereas an experimental set of IR bands around 3437-3450 cm^{-1} could be assigned to the C5 hydrogen-bonded glycine N-H groups. C5 hydrogen-bonded N-H group has been found to have a similar N-H frequency value as we have observed in our case.[1, 19, 112, 159, 160, 167]

Further, we have measured the FTIR spectrum of Boc-Gly-^{D}Pro-NHBn-OMe in $CDCl_3$, which has been provided in Figure 11(k) to compare it with the gas phase IR spectra of the three observed conformers of the peptide. The IR spectral bands in the solution phase are quite broad. However, the positions of the Gly and Bn N-H bands centered at 3324 and 3434 cm^{-1} in the solution phase IR spectrum of the peptide are quite similar to those observed in the gas phase IR spectra. The closely spaced IR bands of the three conformers observed in the gas phase are probably merged under the broad envelope of the solution phase IR band. Thus, the IR spectrum obtained in the solution phase most likely represents the average spectrum due to the

three conformers of the peptide. Scheraga and co-workers have studied the Gly-Pro dipeptide with slightly different protecting groups i.e. Ac-Gly-Pro-NHMe using NMR, IR, and CD spectroscopic techniques.[21] Interestingly, the IR bands in the N-H range of the capped Gly-Pro peptide reported by them in the CD_2Cl_2 solution appear at 3340 and 3440 cm^{-1}. They have concluded from their combined study that the dominant conformation of this dipeptide in the solution phase has extended the C5-C7 structure while other conformations are also present there to a small extent. Recently, Aitken and co-workers have demonstrated a nice correlation between the IR spectra of different conformers of several peptides measured in the solution and gas phase.[7, 227] A comparison between other theoretically low energy conformers and experimentally observed conformers IR has been shown in figure 4.12

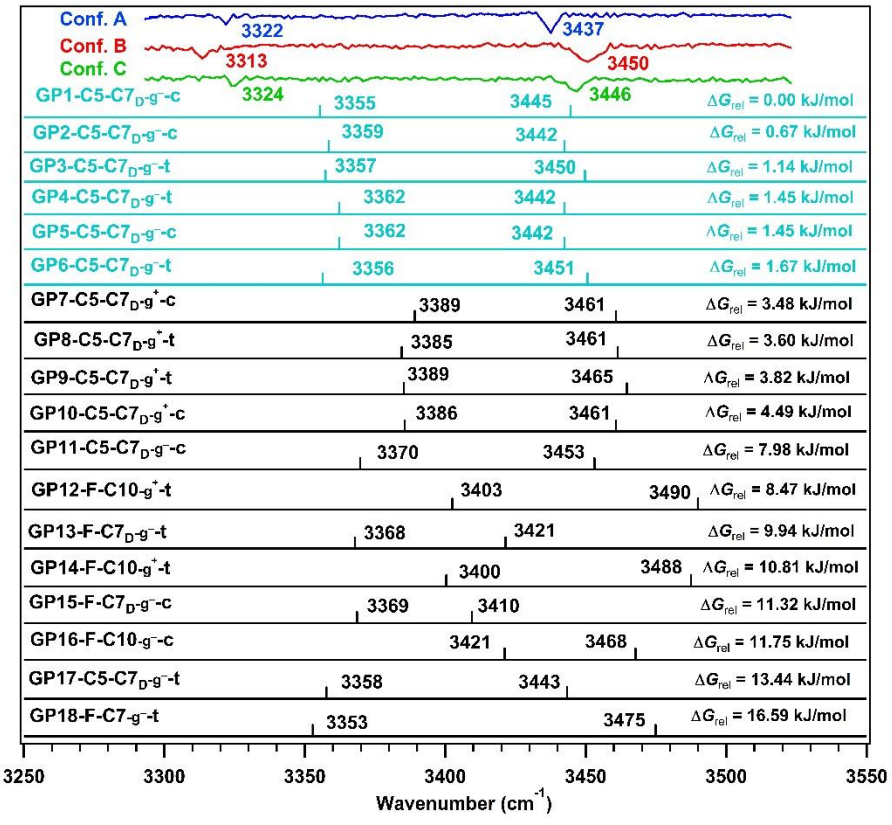

Figure 4.12. Comparison of the experimental IR spectra with the theoretical IR spectra of twelve low energy conformers of Boc-Gly-^{D}Pro-NHBn-OMe calculated at the M06-2X/6-311++G(2d,2p) level of theory. Scaling factor is 0.948 at this particular level of theory.

As we can see from figure 4.11 the conformers apart from GP1-GP6 are either relatively higher in energy or their theoretical frequency is not correlating with the experimental conformer's N-H frequencies.

4.3.6.2 Boc-^{D}Pro-Gly-NHBn-OMe

The experimental gas-phase IR spectra of Boc-^{D}Pro-Gly-NHBn-OMe were measured in the N-H stretching region by probing the A_0^0 and B_0^0 bands of the conformers A and B in the electronic spectra (Figure 4.9a) are provided in Figure 4.13a and c, respectively. The theoretical IR spectra of the two conformers PG1-$C7_D$-$C7_D$-g^+-c and PG2-$C7_D$-$C7_D$-g^+-t, which match closely with the experimental IR spectra, calculated at the M06-2X/6-311++G(2d,2p) level of theory are presented in Figure 4.13b and d, respectively. The IR spectra presented in Figures 4.13a, c, b, and d demonstrate that both the conformers A and B of Boc-^{D}Pro-Gly-NHBn-OMe have a C7-C7 structure. The structures of these two assigned conformers are provided in the figure.

The transitions for the Bn and Gly N-H groups of conformer A appear at 3391 and 3346 cm^{-1}, respectively, while the same transitions for the conformer B are observed at 3411 and 3365 cm^{-1}, respectively. A thorough comparison of the theoretical IR spectra of other low-energy conformers with the experimental IR spectra is provided in Figure 4.14. It is noteworthy from this comparison that the C10 structure is not a probable candidate for the observed conformers of the Pro-Gly peptide in the gas phase. The same scaling factor obtained for the N-H frequencies of the Gly-Pro peptide has been used for the Pro-Gly peptide.

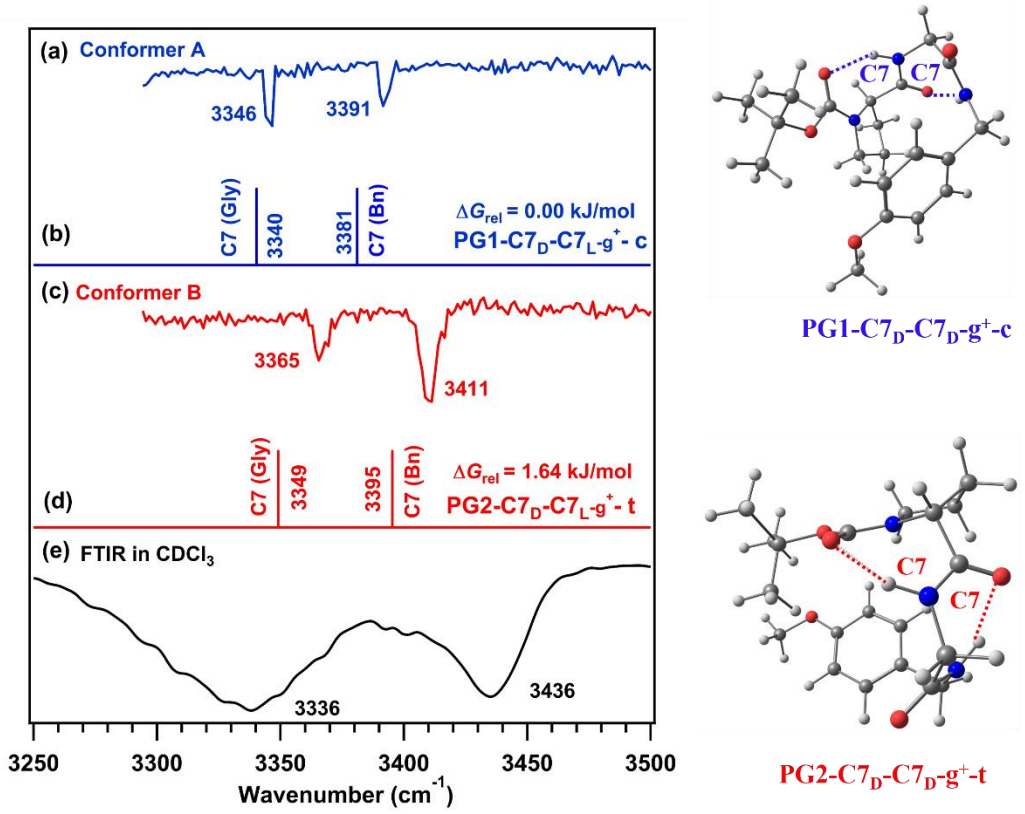

Figure 4.13. (a) and (c) IR spectra of conformers A, and B of Boc-^{D}Pro-Gly-NHBn-OMe, respectively, in the N-H stretching region measured in the gas phase using RIDIR spectroscopy. (b) and (d) Scaled theoretical IR spectra of the conformers PG1-C7-C7 and PG8-C7-C7, respectively, calculated at the M06-2X/6-311++G(2d,2p) level of theory. The theoretical harmonic N-H stretching frequencies of all the conformers are scaled with a scaling factor of 0.948. (e) Solution phase IR spectrum of the Boc-^{D}Pro-Gly-NHBn-OMe peptide in the N-H region measured in $CDCl_3$ solution.

The solution phase IR spectrum of the Boc-^{D}Pro-Gly-NHBn-OMe peptide in the N-H region measured in $CDCl_3$ solution is presented in Figure 4.13e. Both the N-H bands (centered positions at 3336 and 3436 cm^{-1}) in the solution phase IR spectra are extremely broad but appear qualitatively in the similar spectral region of the gas phase IR spectra. The signature for the existence of different low-energy conformers of the Pro-Gly peptide in the solution phase is smeared under the broad envelope of the FTIR spectra. The IR spectrum of the Ac-Pro-Gly-NHMe peptide measured by Scheraga and coworkers in CD_2Cl_2 solution shows the NH bands at 3340 and 3440 cm^{-1}.[21] Thus, it is intriguing to note that the different protecting groups of the

peptides i.e. Boc and NHBn-OMe used in our study do not affect the IR transition frequencies of the NH groups. From combined NMR, CD, and FTIR spectroscopic studies, Scheraga and coworkers surmised that the Pro-Gly peptide in the solution phase exists as various conformations including C10 and C7 structures.[21]

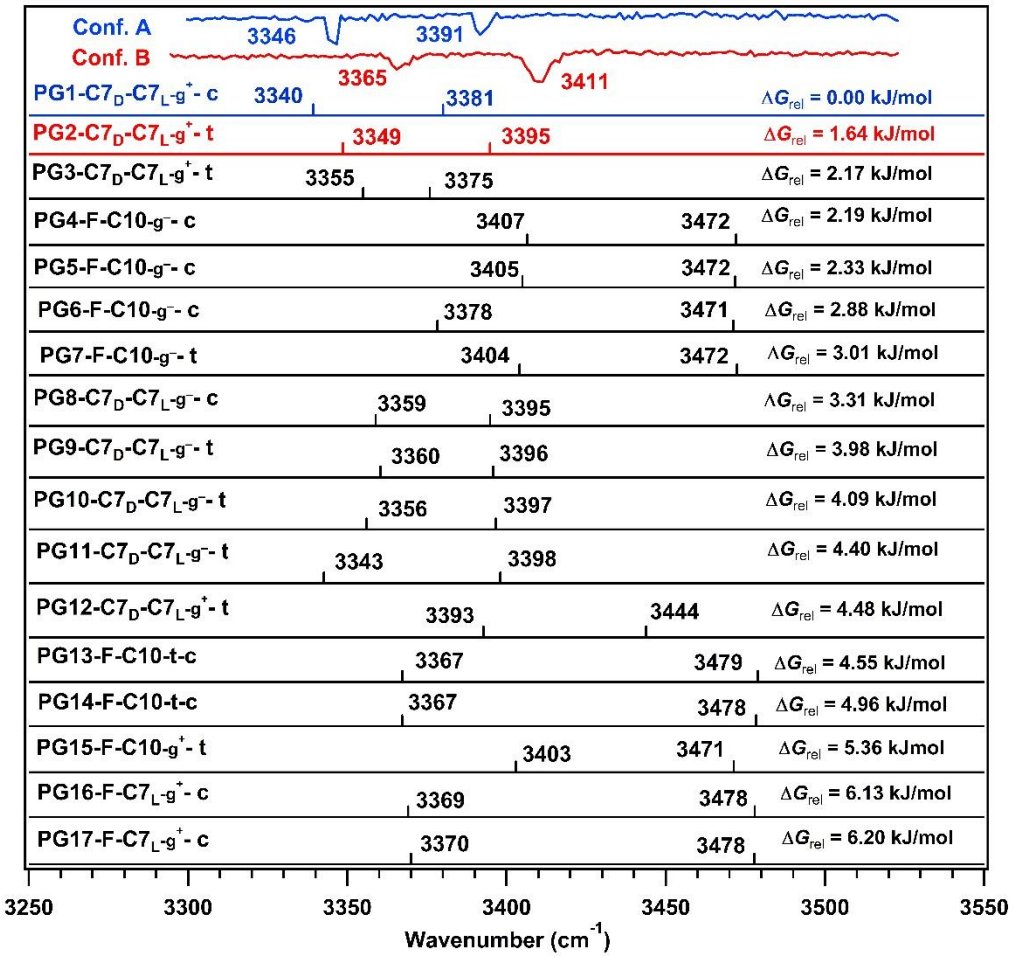

Figure 4.14. Comparison of the experimental IR spectra with the theoretical IR spectra of the twelve low energy conformers of Boc-^{D}Pro-Gly-NHBn-OMe calculated at the M06-2X/6-311++G(2d,2p) level of theory. Scaling has been done by taking Z-Gly-OH molecule as reference. Scaling factor is 0.948 at this particular level of theory.

4.3.7 NMR spectroscopy

4.3.7.1 2D NMR spectroscopy

To get further insight into the structures of the conformations observed in the solution phase, we have measured both 1D and 2D NMR spectra of the Gly-Pro and Pro-Gly peptides in the $CDCl_3$ solution. 1D NMR spectra of Boc-Gly-^{D}Pro-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe are provided in Figures 4.3 and 4.4 respectively. Figure 4.15a shows the partial ROESY (Rotating frame Overhause Effect SpectroscopY) spectrum of Boc-Gly-^{D}Pro-NHBn-OMe measured in $CDCl_3$ solution. The ROESY spectrum provides correlations between the proximal protons in the molecule, which help to determine the 3-D spatial structures of the molecules as well as the possible non-covalent interactions present there.[228-230] The ROESY spectrum in Figure 4.15a shows cross-correlations between glycine CH_2 (3.92 ppm) and proline δCH_2 (3.39 ppm) protons. These correlations indicate that the glycine NH is involved in the intra-residue C5 hydrogen bond and the N-H of the NHBn protecting group is involved in the C7 hydrogen bond.

Additionally, NMR titration of Boc-Gly-^{D}Pro-NHBn-OMe in $CDCl_3$ solution with stepwise addition of DMSO-D6 is performed to monitor the chemical shift positions of Gly N-H and Bn N-H, and the corresponding spectra are provided in Figure 4.16. The chemical shift (δ) values for the Gly N-H and Bn N-H in $CDCl_3$ are 5.36 and 7.07 ppm, respectively, before the addition of DMSO-D6. Upon addition of 30 μL DMSO-D6 in a step of 5 μL, there is a very small downfield shift of both the N-H protons. The downfield shift ($\Delta\delta$) values for the Gly N-H and Bn N-H protons are 0.04 and 0.11 ppm, respectively. Thus, the titration data indicate that both NH_{Gly} and NH_{NHBn} protons are involved in intramolecular hydrogen bonding and are not affected much by DMSO-D6. The overall NMR spectral data substantiate the finding of the most stable conformer of the Gly-Pro peptide i.e. GP1-C5-$C7_D$-g^--c, which is observed from the solution as well as gas phase IR spectroscopy. Figure 4.15b depicts the partial ROESY spectrum of Boc-^{D}Pro-Gly-NHBn-OMe measured in $CDCl_3$ solution. The spectrum shows cross-correlations between Gly_{NH} (6.93 ppm) and Gly_{CH2} (3.90, 4.00 ppm) protons as well as

Gly_{NH} (6.93 ppm) and $Pro\alpha_{CH}$ (4.20 ppm) protons. A C10 structure of Boc-^{D}Pro-Gly-NHBn-OMe, which is compatible with the correlations observed in the ROESY spectrum, is provided in Figure 4.15b. The correlations between the protons in the structure are marked with double-headed arrows. DMSO-d6 induced NMR titration of the Pro-Gly peptide by probing the chemical shift positions of the NH_{Gly} and NH_{NHBn} protons presented in Figure S8 corroborate with the 2D NMR results.[50] The NMR titration data show that the downfield shift of the NH_{Gly} proton ($\Delta\delta$ = 0.68 ppm) is much larger than that of the NH_{NHBn} proton ($\Delta\delta$ = 0.37 ppm). This result indicates that the NH_{Gly} is free and involved in strong intermolecular hydrogen bonding with DMSO while the DMSO-induced intermolecular hydrogen bonding of NH_{NHBn} is weak as it is engaged in intramolecular hydrogen bonding.

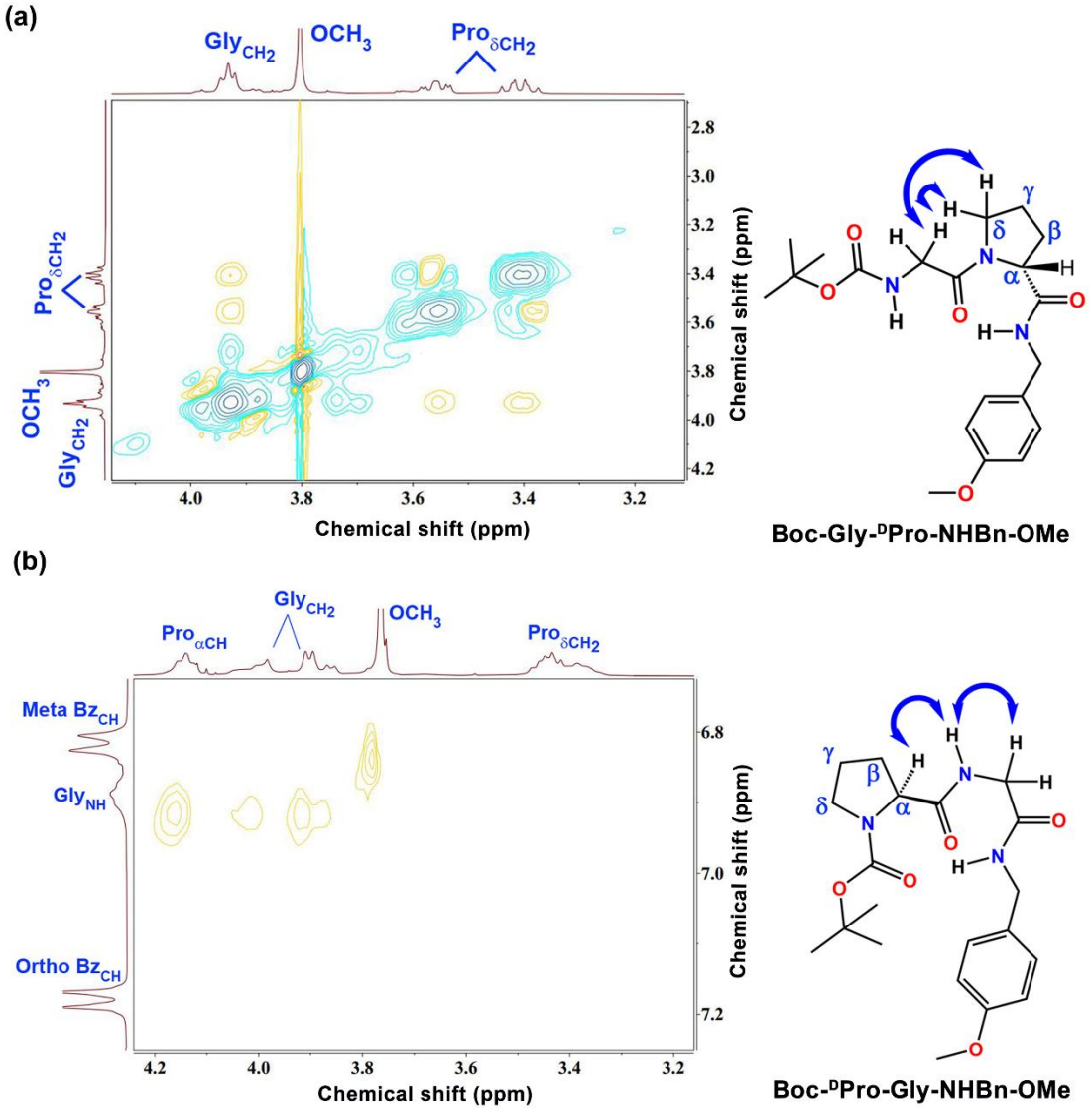

Figure 4.15. Partial ROESY spectra of (a) Boc-Gly-^{D}Pro-NHBn-OMe and (b) Boc-^{D}Pro-Gly-NHBn-OMe measured in $CDCl_3$ solution.

The NMR data bespeak the presence of a C10 structure of Boc-^{D}Pro-Gly-NHBn-OMe in the solution phase. However, the two conformers of the Pro-Gly observed in the gas phase experiment are found to have a C7-C7 structure. The reason for not observing the C10 structure in the gas phase could be the absence of the solvent which may play a role in the preferential stabilization of the C10 structure over the C7-C7 in the solution phase. The C10 structure (PG4-F-C10-g^{-}-c) obtained from the M06-2X/6-311++G(d,p) level of calculation is 2.19 kJ/mol higher in energy than the global minimum structure (PG1-C7$_D$-C7$_L$-g^{+}-c). It is worth

mentioning here that the statistics from the CSD search of the Pro-Gly peptide presented in Figure 4.19 shows a significant number of C10 as well as C7 structures.

4.3.7.2 DMSO-d_6 titration

NMR titration with DMSO-d_6 by probing the chemical shift positions of the N-H groups in the peptides is performed to determine whether the N-H group is free or intramolecular hydrogen-bonded. The chemical shift positions of the N-H groups were monitored by successive addition of 5 µL DMSO-d_6 in the $CDCl_3$ solution of the peptides. If there is a significant downfield chemical shift or deshielding of the N-H proton by the addition of DMSO-d_6, the N-H group in the peptide is free and involved in intermolecular hydrogen bonding with DMSO.

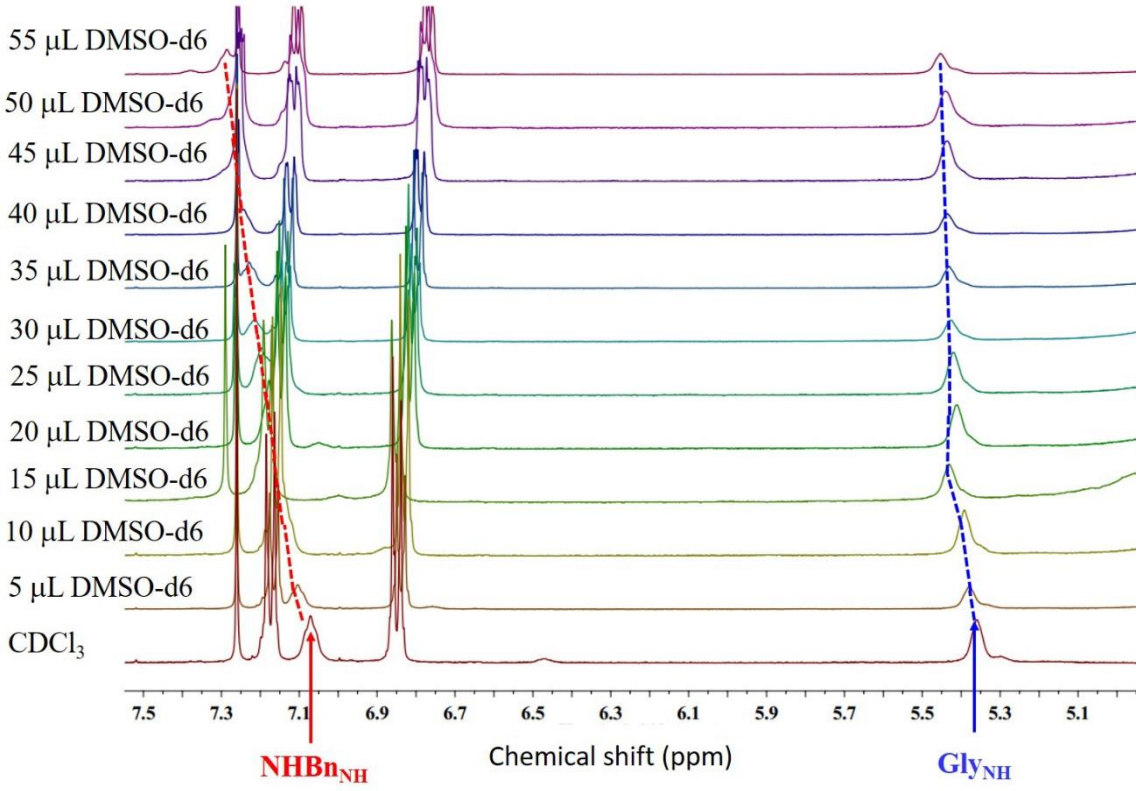

Figure 4.16. NMR titration of Boc-Gly-^{D}Pro-NHBn-OMe in $CDCl_3$ with stepwise addition of DMSO-d_6 by monitoring the chemical shift positions of the Gly and Bn N-H protons.

On the other hand, the change in the chemical shift of the N-H proton will be minimal or negligible on the addition of DMSO-d_6 if the N-H group is already involved in intramolecular hydrogen bonding in the peptide. DMSO-d6 titration of Boc-Gly-^{D}Pro-NHBN-OMe sequence shows both NH are involved in intra-molecular hydrogen bond as the chemical shift of both the NH upon addition of DMSO-d6 is very minimal.

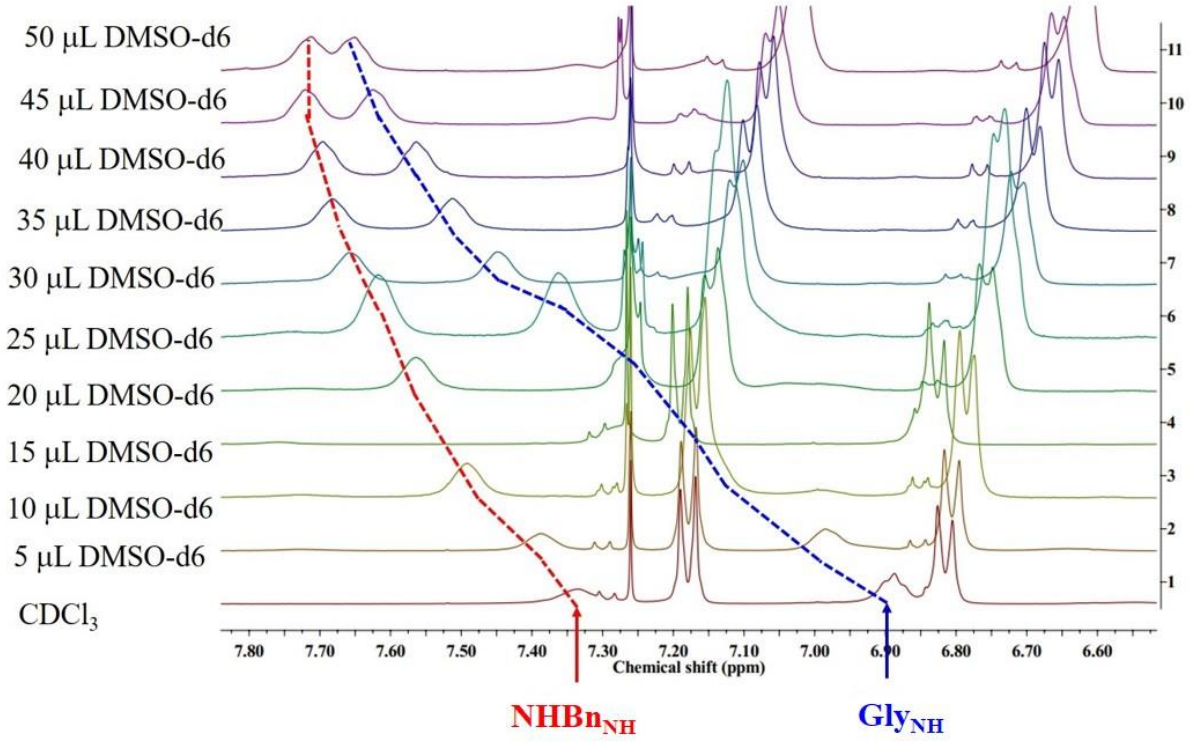

Figure 4.17. NMR titration of Boc-^{D}Pro-Gly-NHBn-OMe in $CDCl_3$ with stepwise addition of DMSO-d_6 by monitoring the chemical shift positions of the Gly and Bn N-H protons.

In the case of Boc-^{D}Pro-Gly-NHBn-OMe NHBn NH shows a very less chemical shift compared to glycine NH upon the addition of DMSO-d6. It shows that in solution glycine NH is free whereas NHBn NH is involved in hydrogen bonding.

4.3.8 X-ray single-crystal structure

A good quality crystal was obtained for the Boc-Gly-^{D}Pro-NHBn-OMe peptide while the quality of the crystal obtained for the Boc-^{D}Pro-Gly-NHBn-OMe peptide was not suitable for the X-ray diffraction. The crystal of Boc-Gly-^{D}Pro-NHBn-OMe was grown in a mixed solvent of ethyl acetate and n-hexane. The crystals were grown through slow evaporation of ethyl acetate and n-hexane solvent mixture. X-ray diffraction of the crystal of Boc-Gly-^{D}Pro-NHBn-OMe was performed using APEX(II) DUO CCD diffractometer. The X-ray data were collected at 100 K temperature. The crystal structure of Boc-Gly-^{D}Pro-NHBn-OMe has been provided in Figure 4.18 while the details of the structure refinement and crystallographic data are provided in Table 4.1. The details of the Ramachandran angles and hydrogen bond parameters of the crystal structure of the Boc-Gly-^{D}Pro-NHBn-OMe peptide have been listed in Table 4.2.

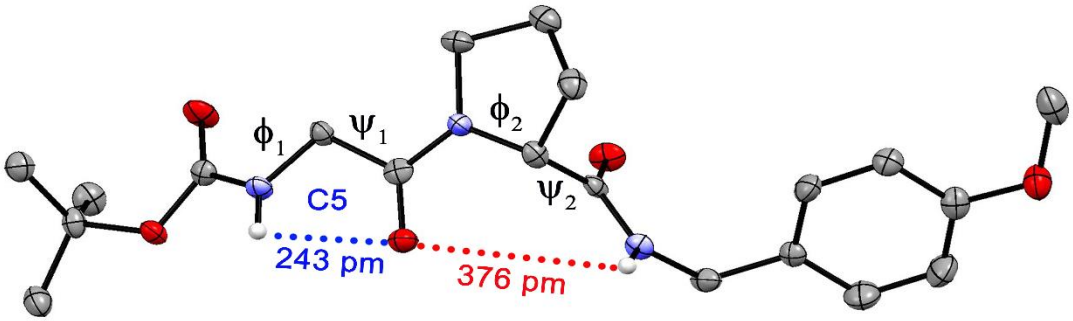

Figure 4.18. ORTEP drawing of the crystal structure of Boc-Gly-^{D}Pro-NHBn-OMe. Thermal ellipsoids have been shown at the 50% level. Only the N-H hydrogen atoms have been shown for the clarity.

Table 4.1 The details of the crystal structure refinement and crystallographic data for Boc-Gly-^{D}Pro-NHBn-OMe

CCDC	2157775

Empirical formula	$C_{20}H_{29}N_3O_5$
M_r	391.46
Crystal size, mm^3	0.12x0.16x0.20
Crystal system	Monoclinic
Space group	P 21
a, Å	11.7946(9)
b, Å	9.3562(8)
c, Å	18.9259(16)
α, °	90
β, °	105.551(2)
γ, °	90
Cell volume, $Å^3$	2012.1(3)
Z ; Z'	4
T, K	100
Radiation type; wavelength Å	MoKα, 0.71073
F_{000}	840
μ, mm^{-1}	0.093
θ range, °	2.234 - 30.170
Reflections collected	11923
Reflections unique	5555
R_{int}	8.35
Parameters	514
wR_2 (all data)	0.1413

The crystal structure was obtained by direct methods using SHELXS-97.[231]

4.3.9 Structural parameters of the assigned structures of the conformers observed in the experiment

Hydrogen bond parameters and Ramachandran angles of the calculated structures assigned for the observed conformers of the Gly-Pro and Pro-Gly peptides are listed in Table 4.2. In the case of the Boc-Gly-^{D}Pro-NHBn-OMe peptide, the C5 hydrogen bond distance for all the

theoretical structures is around 215 pm whereas the C7 hydrogen bond distance is around 200 pm.

Table 4.2. Important geometrical parameters of the observed conformers of Boc-Gly-^{D}Pro-NHBn-OMe calculated at the M06-2X/6-311++G(2d,2p) level of theory

Conformer	Crystal-structure	GP1-C5-$C7_D$-g^--c	GP2-C5-$C7_D$-g^--c	GP3-C5-$C7_D$-g^--t	GP4-C5-$C7_D$-g^--t	GP5-C5-$C7_D$-g^--c	GP6-C5-$C7_D$-g^--t
ϕ_1 (^{0})	150	-155	-155	-167	-154	-154	-167
ψ_1 (^{0})	-177	174	173	175	173	173	176
ϕ_2 (^{0})	-79	84	84	82	84	84	82
ψ_2 (^{0})	143	-72	-72	-80	-72	-72	-80
ϕ_{NHBn}(^{0})	-99	-68	-67	-83	-68	-68	-83
ϕ_{OMe}(^{0})	9	-2	-2	176	178	-2	175
C5 H-bond distance (pm)	243	215	215	220	215	215	220
C5 H-bond angle (^{0})	99	106	106	105	106	106	105
C7 H-bond distance (pm)	376	198	198	203	198	198	203
C7 H-bond angle (^{0})	85	148	147	142	147	147	142

The Ramachandran angles φ_1 (~155°) and ψ_1 (~173°) of the Gly residue of six theoretical structures are commensurate with those which are the characteristics of the C5 hydrogen bond reported in the literature.[1, 23, 27, 118, 159] The dihedral angles φ_2 and ψ_2 of the Pro residue of all the three conformers having C7 hydrogen bond are comparable (~ 80° and ~ -70°, respectively). Similar Ramachandran angles of an amino acid residue involved in C7 hydrogen bonding are reported for various small peptides studied in the literature.[1, 23, 27, 159]

The structural parameters of the crystal of Boc-Gly-^{D}Pro-NHBn-OMe are also provided in Table 4.2. It is worthwhile to note that the Ramachandran angles φ_1 and ψ_1 of the Gly residue in the crystal, which represents the extended C5 structure, corroborate well with those in the most stable conformers of the peptide observed in the solution/gas phase experiment. However,

the ψ_2 angle of the Pro residue in the crystal deviates from that in the observed GP1-C5-$C7_D$-g^--c conformer. Thus, the distance and angle parameters for the C7 hydrogen bond involving the Pro residue in the crystal are out of the range of any regular hydrogen bond. This could be due to crystal packing forces which drive the other unit of the peptide to interact through intermolecular hydrogen bonding. Nevertheless, the overall extended linear structural motif of the Gly-Pro peptide is present in both the crystal and gas/solution phase structures.

Table 4.3. structural parameters of Boc-^{D}Pro-Gly-NHBn-OMe

Conformers	PG1-$C7_D$-$C7_D$-g^+-c	PG1-$C7_D$-$C7_D$-g^+-t
ϕ_1 (^{0})	86	87
ψ_1 (^{0})	-64	-71
ϕ_2 (^{0})	-80	-82
ψ_2 (^{0})	72	66
ϕ_{NHBn}(^{0})	86	72
ϕ_{OMe}(^{0})	-1	-174
C7 H-bond distance Gly (pm)	196	198
C7 H-bond angle Gly (^{0})	149	147
C7 H-bond distance NHBn (pm)	203	203
C7 H-bond NHBn angle (^{0})	145	146

For the Boc-^{D}Pro-Gly-NHBn-OMe peptide, the Ramachandran angles involving the Pro and Gly residues of both the observed conformers PG1-C7-C7 and PG8-C7-C7 in the gas phase corroborate well with those in any peptide residue having C7 hydrogen bond reported in the literature. [1, 23, 27, 159]

4.3.10 A Cambridge Structural Database (CSD) study of peptide sequences

A statistic of various non-covalent interactions (C5, C7, and C10) present in the Gly-Pro and Pro-Gly containing peptides was obtained using CCDC conquest 2020 2.0 software.[232] The Cambridge structural database contains numerous structures of organic and inorganic

materials.[233] CSD studies show for the Gly-Pro-X and Pro-Gly-X sequence have been shown in figure 4.19. We can see that the Gly-Pro-X sequence has a significant number of C5 hydrogen bond interactions compared to the Pro-Gly-X sequence. However, the C10 (β-turn) is significantly higher in the case of the Pro-Gly-X sequence. This further concludes that Gly-Pro-X sequence has a higher propensity for extended conformation whereas the Pro-Gly-X sequence has higher propensity for the folded β-turn conformation in crystals as well.

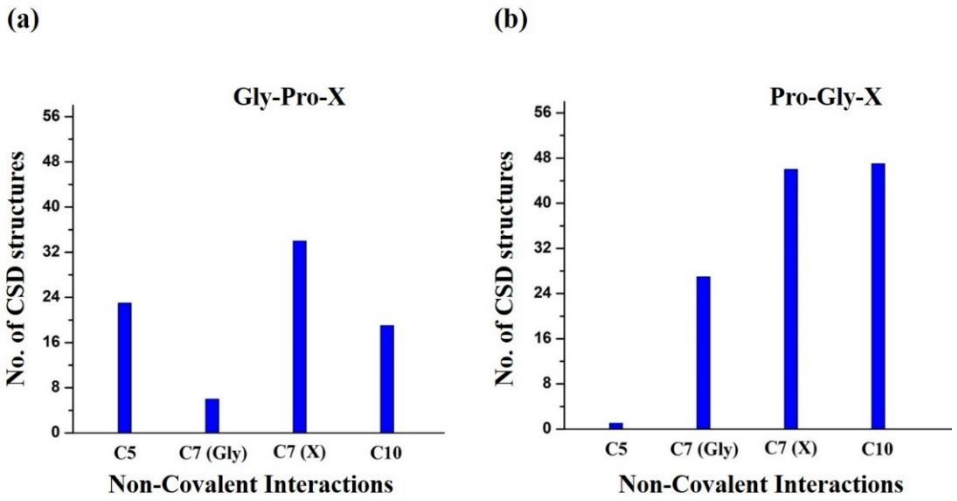

Figure 4.19. Statistics of the number of CSD structures having different non-covalent interactions in (a) Gly-Pro-X and (b) Pro-Gly-X containing peptides.

4.4 Conclusion

Sequence-dependent folding motifs, as well as the conformational landscape of Gly-Pro and Pro-Gly containing dipeptides, have been investigated by employing gas phase laser spectroscopy, quantum chemistry calculations, solution-phase IR, and NMR spectroscopy, and XRD technique. In the gas phase experiment, three low-energy conformers of Boc-Gly-^{D}Pro-NHBn-OMe are observed where all the conformer has extended β-strand or PP-II type structure with C5 and C7 hydrogen-bonded rings. Interestingly, the C5-C7 extended structures of the Gly-Pro peptide in the gas phase has a one-to-one correspondence with the structure observed in the solution phase FTIR, NMR spectroscopy as well as solid-state XRD. In the case of the

Pro-Gly peptide, both the conformers observed in the gas phase have C7-C7 i.e. 2_7-ribbon structure. However, the solution phase NMR and FTIR spectroscopy indicates the presence of the C10 or β-turn structure of the Pro-Gly peptide, which is not observed in the gas phase probably due to the absence of the solvent. Overall, the present study demonstrates that the extended β-strand structure of the Gly-Pro peptide changes to the C10 (β-turn) structure when the sequence is changed to the Pro-Gly peptide. The C7-C7 conformation of the Pro-Gly peptide observed in the gas phase is one of the structural motifs of the Pro-Gly-containing peptides present in the CSD. The gas-phase study has the advantage of exploring multiple low-energy conformations of small peptides and the combination of this investigation with that in the solution phase and solid-state will render a comprehensive picture of the understanding of the folding motifs of peptides containing a variety of residues.

Chapter 5

Effect of Neighbouring Residue on the β-turn in Pro-Gly-Ala sequence

5.1 Introduction

The specific sequences as well as the secondary structures formed by the hydrogen bonding interactions in the amide backbone plays an important role in specific folding patterns of polypeptides and proteins.[3, 11, 12] Among several types of secondary structures, turns are very important as they provide reversal of the polypeptide chain direction and consequently the compactness in the folded structures. β-turns are the most commonly occurring secondary structures in polypeptides and proteins.[14, 22, 69, 84, 200] Different types of β-turn structures are classified according to the Ramachandran angles in the amide backbone.[15, 86] β-turn conformations adopted by the peptide sequences are found to be very important sites for various enzymatic reactions as these sites are mostly located in the exterior regions of the proteins.[14] In the previous chapter, it has been mentioned that there is a requirement for the β-turn conformation of the Pro-Gly segment in procollagen for selective hydroxylation of the Pro residue by prolyl hydroxylase enzyme to form the triple helical structure of collagen.[14, 234-236] β-turns are also the sites for phosphorylation and attachment of the sugar residues to glycoproteins.[237-239]

The formation of the β-turns in peptides or proteins is very sensitive to the presence of a particular sequence of amino acids. It has been found that the Pro-Gly sequence compared to Gly-Pro has more propensity for the formation of the β-turns in peptides and proteins.[21, 97] Brahmachari and co-workers demonstrated the formation of type II β-turn conformations in small peptides i.e., Ac-Pro-Gly-X-OH (X=Ala, Leu, Gly, Ile, and Phe) using NMR, IR, and CD spectroscopy as well as X-ray crystallography technique.[22, 236] Further, they showed that the stability of the β-turn in the Pro-Gly structure increases due to the presence of a specific neighbouring residue X in the order of X=Leu > Ala > Gly, Ile > Phe. It was also concluded from their study that the extent of the enzymatic hydroxylation of the Pro residue in the Pro-Gly-X sequence would follow a similar trend. They analyzed the structures of 20 tripeptides of

Z-Pro-Gly-X sequences for studying the effect of residue X on the stabilization of the β-turn. Studies by Prockop and co-workers revealed that the extent of the hydroxylation of the Pro residue in the peptide segment Pro-Pro-Gly-$NHCH_3$ was the same as that in the ideal substrate protocollagen.[240] Stimson and co-workers have reported from their solution-phase spectroscopic studies that the Ac-Pro-Gly-$NHCH_3$ exists in β-turn conformation.[21] Bhatnagar and Rapaka have shown that the hydroxylation of Pro in the tripeptide unit of $(Pro\text{-}Gly\text{-}X)_n$ is quite considerable when the residue X is Leu, Pro, and Ala but is negligible when the residue X is Val.[13] The amino acids such as Val and others do not stabilize the β-turn due to their relatively large side chain.

In the previous chapter, we performed a comprehensive study on the sequence-dependent folding motifs of Boc-^{D}Pro-Gly-NHBn-OMe and Boc-Gly-^{D}Pro-NHBn-OMe peptides using solution-phase FTIR, NMR spectroscopy, X-ray crystallography as well as gas-phase UV/IR laser spectroscopy and quantum chemistry calculations. Combined gas phase and solution phase spectroscopy along with the X-ray crystallography data reveal that the most stable conformer of the Gly-Pro peptide has an extended β-strand or polyproline-II (PP-II) structure having C5-C7 hydrogen bonding interactions. On the other hand, the Pro-Gly peptide is found to have a C10/β-turn structure in the solution phase although the observed structure in the gas phase is C7-C7 i.e. 2_7-ribbon. In this work, we have studied the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide to probe the effect of the neighbouring residue Ala on the strength of the β-turn in the peptide using solution-phase IR, 2D-NMR spectroscopy, and X-ray crystallography in combination with quantum chemistry calculations. Figure 5.1 shows the chemical structure of the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide.

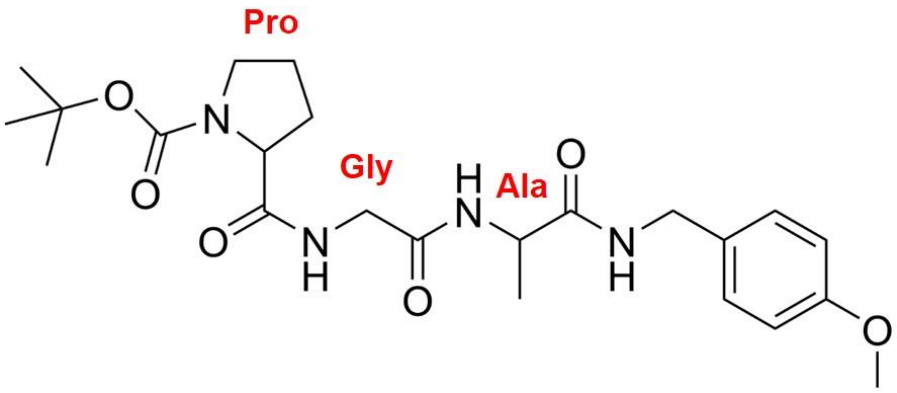

Figure 5.1. Chemical structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide.

5.2 Results and Discussions

5.2.1 Synthesis and characterization of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

Boc-Ala-OH (1.0 g, 5.3 mmol) was converted to Boc-Ala-NHBn-4-OMe by treating it with 1 equiv. of HBTU, 1 equiv. of HOBt, 1.1 equiv. of H2NBn-4-OMe and 5 equiv. of DIPEA in DMF at 0 °C to room temperature for 12 hrs. Progress of the reaction was monitored using TLC. After completion of the reaction, the reaction mixture was acidified using 10% HCl. Using EtOAc (3×50 mL), the product was extracted which was then washed with 10% Na2CO3 (3×50 mL) and with brine solution (3×50 mL) and finally dried over anhydrous Na2SO4. The white color compound was collected after concentrating the solution in a vacuum. The reaction yield was 90%. Further, it was purified using column chromatography by EtOAc and hexanes as mobile phase.

Boc-Ala-NHBn-4-OMe (1.23g, 4 mmol) was converted to H-Ala-NHBn-4-OMe by treating it with 50% TFA-DCM (1:1) for 20 minutes at room temperature. Reaction progress was monitored using TLC. After completion of the reaction, DCM and the remaining TFA were evaporated in a vacuum to get the product. The reaction yield was 98%.

Boc-Pro-Gly-OH (1.0 g, 3.6 mmol) was initially activated with 1 equiv. of HBTU, 1 equiv. of HOBt, 5 equiv. of DIPEA in DMF at ice-cold conditions. After 10 minutes of activation, 1.1 equiv. of H-Ala-NHBn-4-OMe was added to the reaction mixture, and the reaction was allowed to stir for 12 hrs. Progress of the reaction was monitored using TLC. After completion of the reaction, the reaction mixture was diluted, and using EtOAc (3×80 mL) the product was extracted, which was then washed with 10% HCl (3×60 mL), 10% Na2CO3 (3×60 mL), and brine solution (3×60 mL) finally dried over anhydrous Na2SO4. The colorless liquid compound was collected after concentrating the solution in a vacuum. The reaction yield was 86%. Further, it was purified using column chromatography by EtOAc and hexanes as mobile phase.

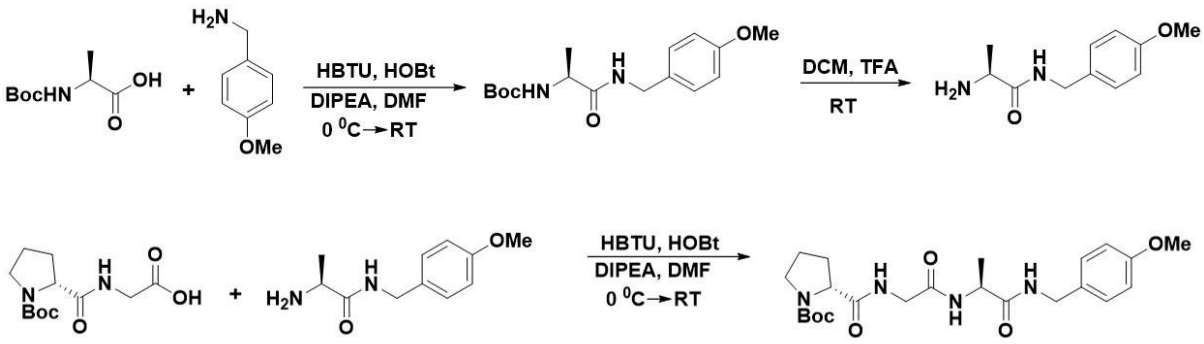

Figure 5.2. Synthetic Scheme of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

5.2.1.2 ^{1}H NMR spectrum of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

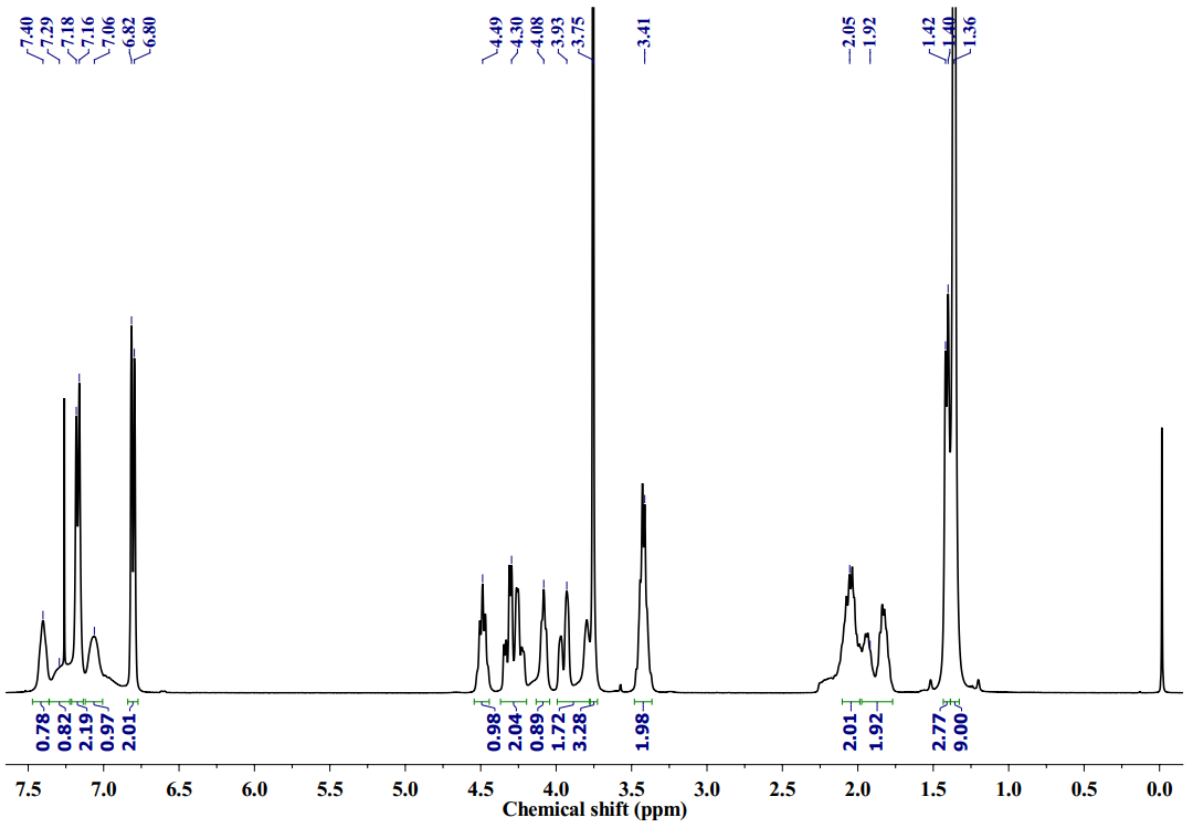

Figure 5.3. 1H NMR spectrum of Boc-D-Pro-Gly-Ala-NHBn-OMe in $CDCl_3$ in a 400 MHz NMR spectrometer (Bruker-400).

The ^{1}H NMR spectrum of Boc-^{D}Pro-Gly-Ala-NHBn-OMe recorded in $CDCl_3$ solvent has been presented in figure 5.3. The assigned 1H NMR peaks have been provided below.

^{1}H NMR of Boc-^{D}Pro-Gly-Ala-NHBn-OMe (400 MHz, $CDCl_3$, 298.15 K): δ 6.80-6.82 (m,2H, CH^{Ar}) 7.16-7.18 (m,2H, CH^{Ar}) 3.75 (s, OCH_3) 1.37 (s, 9H, CH_{Boc})

5.2.1.3 ESI-MS spectrum of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

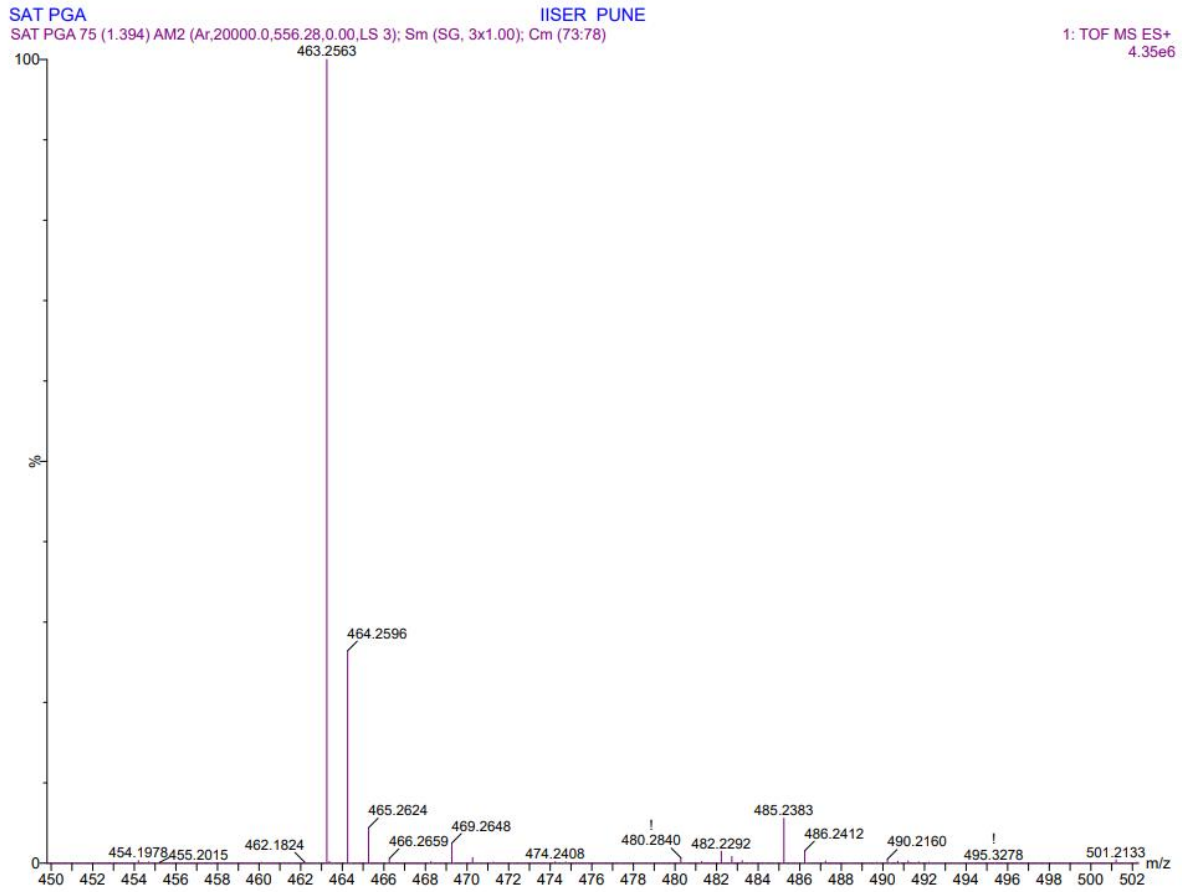

Figure 5.4. ESI-MS mass spectrum of Boc-^{D}Pro-Gly-Ala-NHBn-OMe.

The high-resolution mass spectrum (HRMS) of Boc-^{D}Pro-Gly-Ala-NHBn-OMe has been recorded using the ESI-MS technique. The sharp intense peak at 463.2563 m/z in figure 5.4 corresponds to the peptide+Na^{+} combined molecular weight.

5.2.2 X-ray Crystal Structure

Boc-^{D}Pro-Gly-Ala-NHBn-OMe was dissolved in ethyl acetate taken in a 1ml glass vial and kept for slow evaporation of solvent molecules for ~ 5 days. A good quality needle-like crystal was observed after five days. APEX(II) DUO CCD diffractometer was used to record the X-ray diffraction of the crystal at a temperature of 100 K. The crystal structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe has been provided in Figure 5.5.

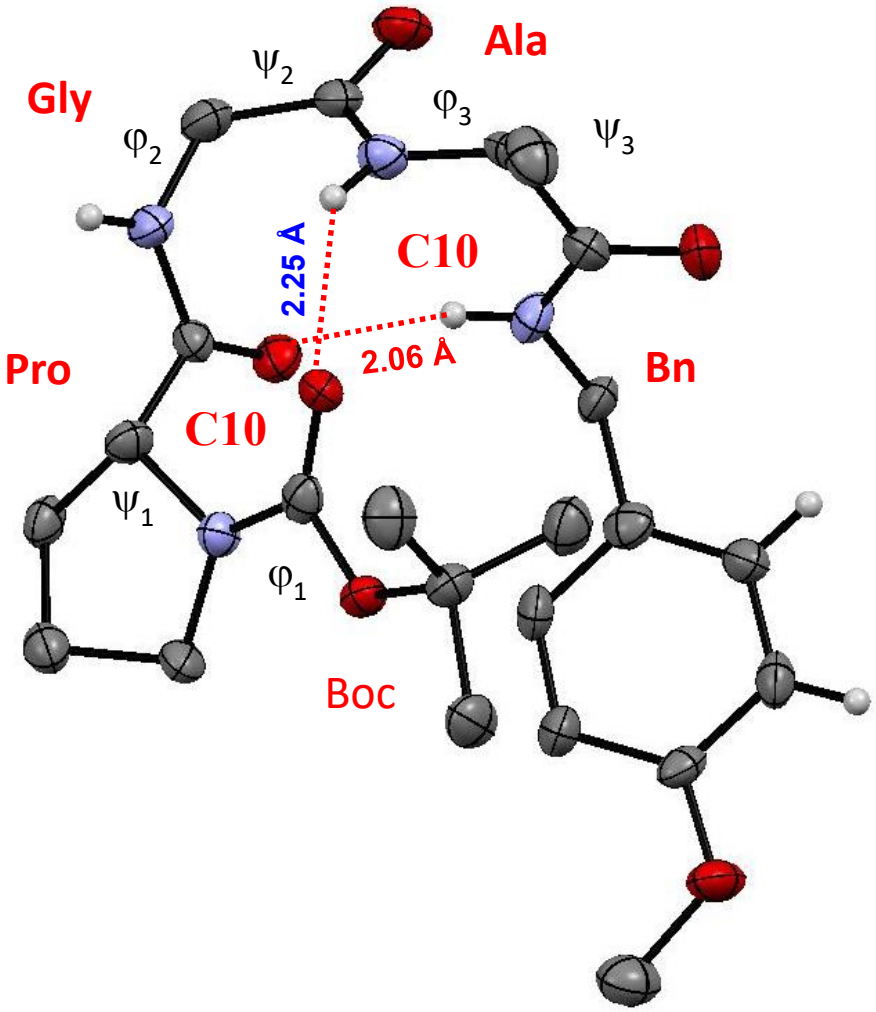

Figure 5.5. ORTEP diagram of the X-ray crystal structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe with 50 % ellipsoid probability.

Interestingly, the crystal structure reveals that the tripeptide exists in a C10-C10 helical fashion in the crystalline form.[241] The amide N-H group of the Ala residue forms a C10 hydrogen bond (β-turn) with the carbonyl group of the "Boc" protecting group while the amide N-H group of the NHBn moiety forms a C10 hydrogen bond (β-turn) with the carbonyl group of the Pro residue. The amide N-H group of the Gly residue remains free from any intramolecular hydrogen bond. The Ramachandran angles of the peptide backbone and selected hydrogen bond parameters of the crystal structure are marked listed in Table 5.1 The φ, ψ angles of the backbone involving the three residues forming the C10 hydrogen bond represent typical type-

II' and type I β-turn structure as reported in the literature.[86] The hydrogen bond parameters listed in the table point out that the C10 hydrogen bond (β-turn) involving the C=O group of Pro and N-H group of NHBn moiety is much stronger than that involving the N-H group of Ala and C=O group of the Boc moiety.

Table 5.1. Selected geometrical parameters of the crystal structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

Parameter	Value
φ_1	65^0
ψ_1	-138^0
φ_2	-63^0
ψ_2	-17^0
φ_3	-84^0
ψ_3	3^0
NH(Ala)…O=C(Boc)	225 pm
NH(Bn)…O=C(Pro)	206 pm
∠NH(Ala)…O=C(Boc)	137^0
∠NH(Bn)…O=C(Pro)	160^0

5.2.3 Solution-phase NMR spectroscopy studies of Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide

Various NMR spectroscopy techniques in the solution phase were used to get the structural information of the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide. All the NMR spectroscopy experiments were carried out in $CDCl_3$ solvent. 1D ^{1}H NMR spectrum of this peptide has been

provided in Figure 5.3 of section 5.2.1.2 of the current chapter. Apart from the 1D ^{1}H NMR spectroscopy, 2-D NMR spectroscopy and DMSO-d_6 NMR titration were performed to obtain detailed structural information about the peptide in the solution phase.

5.2.3.1 2D NMR Spectroscopy

Various 2D NMR spectroscopic techniques are extensively used to predict the structures of the compounds in the solution phase. The partial ROESY (Rotating frame Overhause Effect SpectroscopY) spectrum of Boc-^{D}Pro-Gly-Ala-NHBn-OMe measured in $CDCl_3$ solution has been shown in figure 5.6.

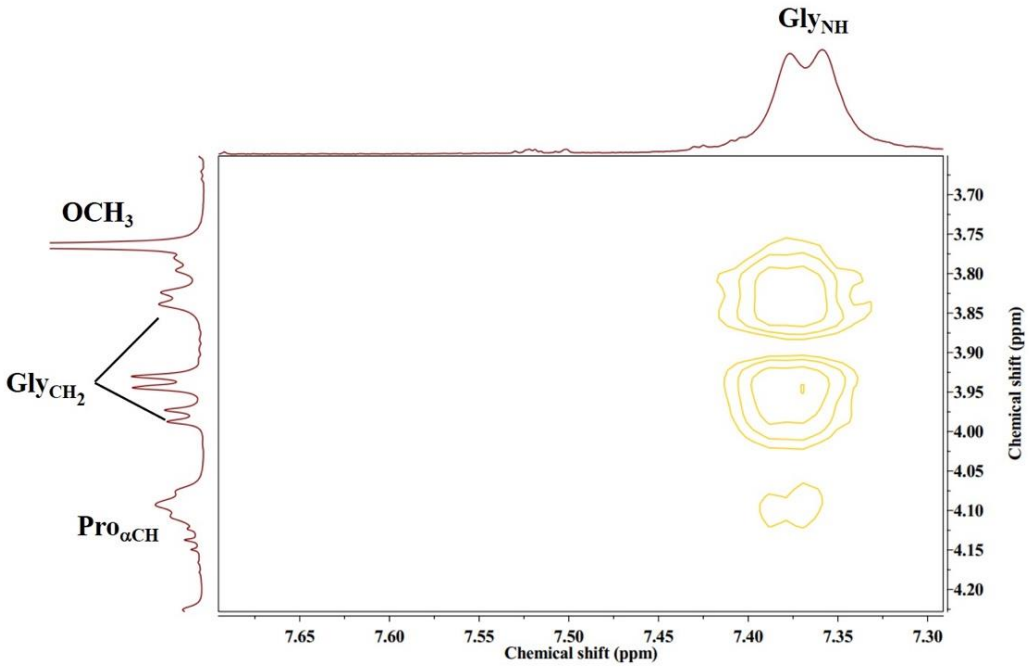

Figure 5.6. ROESY spectra of Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CDCl_3$ showing correlation between various protons using 400 MHz NMR spectrometer.

The partial ROESY spectra shown in figure 5.6a includes three cross-diagonal contour peaks. The cross-diagonal contours in the ROESY spectrum usually arise due to the spatial interaction of two protons even when they are separated by two or more bonds. The top two contours correspond to the correlation between glycine CH_2 proton and Gly_{NH} proton as they are only separated by a bond. The third contour arises as Gly_{NH} comes spatially close to the Proline α_{CH} proton. This correlation suggests that the peptide molecule in the solution has the C10-C10 confirmation as shown in figure 5.7. The solution phase NMR experiments suggest that the peptide molecule exists in the C10-C10 conformation where Alanine and NHBn NH are involved in the C10 hydrogen bond and the Glycine NH is free from any intramolecular hydrogen bonding.

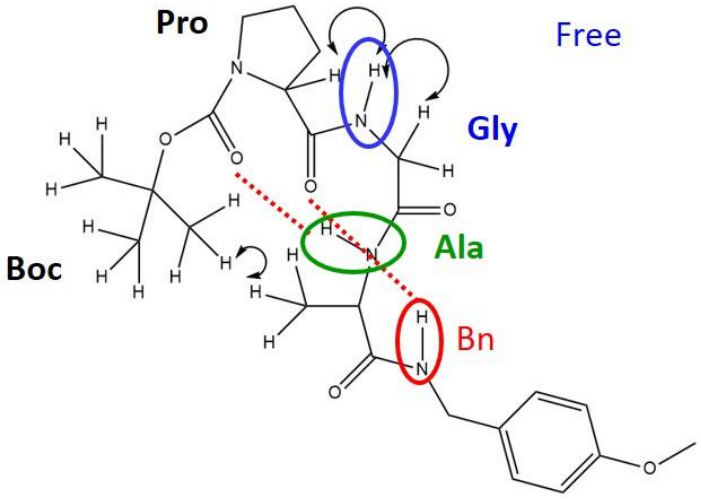

Figure 5.7. Structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide deduced from ROESY performed in $CDCl_3$ solvent.

5.2.3.2 DMSO-d_6 NMR titration of Boc-^{D}Pro-Gly-Ala-NHBn-OMe

DMSO-d_6 titration of Boc-^{D}Pro-Gly-Ala-NHBn-OMe is performed to determine the nature of the NH group i.e., whether it is free or involved in intramolecular hydrogen bonding. In this technique, the relative chemical shift of the concerned proton is monitored upon sequential

addition of DMSO-d_6 in the deuterated solvent $CDCl_3$. If a proton is involved in an intramolecular hydrogen bonding, the change in the chemical shift upon the addition of DMSO-d_6 is quite less compared to a proton that is free from any intramolecular hydrogen bonding or weakly hydrogen-bonded.

Figure 5.8 shows the results obtained by performing DMSO-d_6 titration on Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CDCl_3$. In this experiment, initially, 1H NMR of the peptide molecule was recorded in $CDCl_3$ solvent and then 5 µL DMSO-D_6 was added at each step. After each addition of DMSO-D_6, the 1H NMR spectrum of the peptide molecule was recorded for comparison. All these ^{1}H NMR spectra were stacked vertically in the MNOVA program to have a clear comparison of chemical shifts of concerned protons. We can see from figure 5.8 that the change in the chemical shift of Gly N-H is quite significant compared to that of the other two N-H groups. The glycine N-H chemical shift changes its value f 7.35 ppm to around 7.90 ppm upon the addition of 50 µL of DMSO-d_6. The change in the chemical shift for NHBn NH and Ala NH is from 7.1 to 7.26 ppm and 7.41 to 7.45 ppm, respectively. The plot of the change in the chemical shift vs. the volume of DMSO-d_6 added has been shown in figure 5.9.

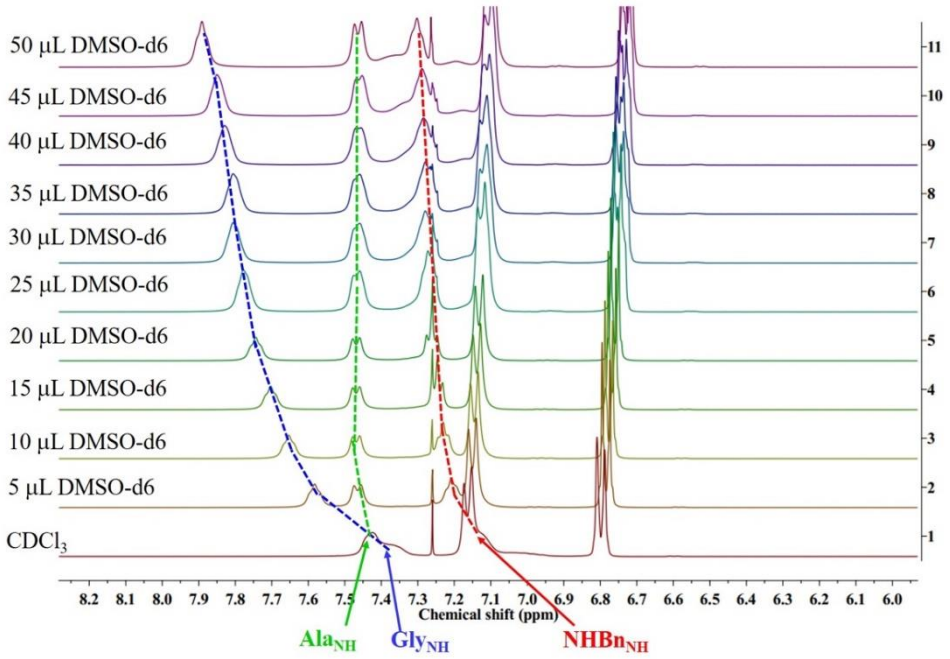

Figure 5.8 DMSO-d_6 ^{1}H NMR titration of Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CDCl_3$ using a 400 MHz NMR spectrometer.

The relative chemical shift of amide proton upon addition of DMSO-d_6 in figure 5.10 suggests that Gly_{NH} is free while the Ala_{NH} and $NHBn_{NH}$ are involved in an intramolecular hydrogen bond. Hence, the ROESY spectra as well as DMSO-d_6 NMR titration in the solution phase corroborate the C10-C10 structure of the peptide obtained from X-ray crystallography.

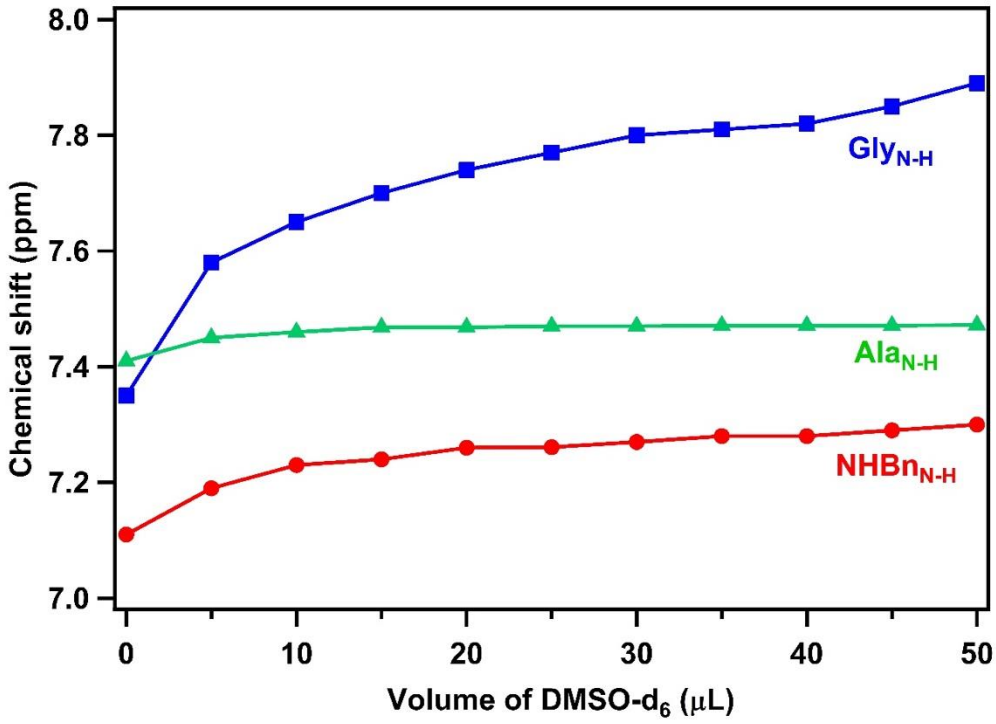

Figure 5.9. A plot of change in the chemical shift of different amide protons vs. volume of DMSO-d_6 added in the original solution of Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CDCl_3$. 5 µL of DMSO-d_6 was added in each step and ^{1}H NMR spectra were recorded thereafter.

5.2.4 Solution Phase FTIR studies of peptides

Solution phase IR spectra of the peptides were measured in $CDCl_3$ solvent at 293 K using a Fourier-Transform IR spectrometer. The FTIR spectra of Boc-^{D}Pro-Gly-Ala-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe measured in the N-H stretching frequency region are shown in

figure 5.10. The observed IR bands in both spectra are quite broad. The C10-C10 structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe predicted from single-crystal XRD and NMR spectroscopy analysis is shown next to its IR spectra. In the case of Boc-^{D}Pro-Gly-NHBn-OMe, we didn't obtain any good quality crystal which could be diffracted by X-ray. Thus the C10 structure of Boc-^{D}Pro-Gly-NHBn-OMe provided next to its IR spectrum is derived from the 2D NMR spectrum and DMSO-d_6 NMR titration in $CDCl_3$. The peak, which appears at almost the same position i.e., 3436/3439 cm^{-1} of both the IR spectra, corresponds to the free N-H group of the Gly residue of the peptides. The band at 3336 cm^{-1} in the IR spectrum of the Pro-Gly peptide could be assigned as the hydrogen-bonded N-H group of the NHBn moiety which forms C10 or β-turn. The corresponding IR band for the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide is much broader than that for the Boc-^{D}Pro-Gly-NHBn-OMe peptide and this additional broadening in the former one could be due to overlap of the unresolved two N-H peaks for the NHBn and Ala moieties. Based on the X-ray crystal structure of Boc-^{D}Pro-Gly-Ala-NHBn-OMe (Figure 5.13), the hydrogen bond formed by the Ala N-H group is weaker than that by the NHBn N-H group. Thus, the most red-shifted peak at 3323 cm^{-1} is assigned as NHBn N-H while the Ala N-H group appears at ~3400 cm^{-1}. A comparison of the IR spectra of the two peptides indicates that the frequency of the NHBn N-H forming the β-turn in Boc-^{D}Pro-Gly-Ala-NHBn-OMe is red-shifted by 13 cm^{-1} compared to that in Boc-^{D}Pro-Gly-NHBn-OMe. Thus it is demonstrated directly from the solution phase FTIR spectra that the strength of the β-turn of the Pro-Gly sequence increases due to the presence of the neighbouring Ala residue.

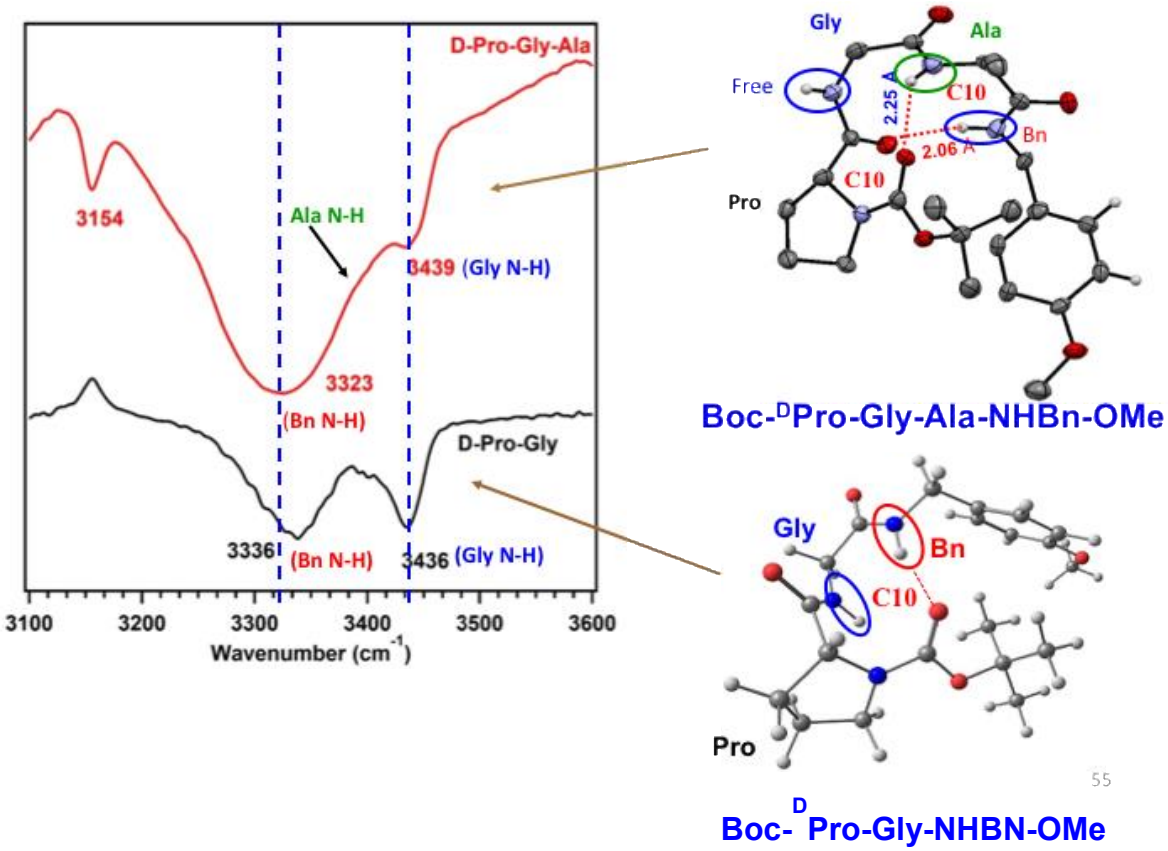

Figure 5.10. A comparison of the FTIR spectra of the peptides Boc-^{D}Pro-Gly-Ala-NHBn-OMe (top) and Boc-^{D}Pro-Gly-NHBn-OMe recorded in $CDCl_3$ solvent. The concentration of the peptide molecules was 8mM.

5.2.5 Theoretical calculations

Both gas and solution phase calculations of the peptides have been performed to further probe their detailed structures.

5.2.5.1 Gas phase calculations

Conformation search for the Boc-^{D}Pro-Gly-Ala-NHNn-OMe was executed by force field method (MMFF94) incorporated in MarvinSketch software. The MarvinSketch software generated around 200 conformers and those conformers were arranged in terms of the relative energies to the global minimum structure. The conformers having similar structures and within the energy difference of 0.5 kJ/mol were kept in a group and the lowest energy conformer in

the group was optimized at the HF/6-31G level of theory. After optimization at the HF/6-31G level of theory, conformers were arranged relative to the global minimum conformer. The structures obtained from the HF/6-31G level of calculation were further optimized at the B97-D/6-31+G(d) level of theory. Thermal corrections were done by calculating the Gibbs free energy at temperatures ranging from 0 to 1000 K with an interval of 50 K.

The energy landscape of the low energy conformers after optimization at the B97-D/6-31+G(d) level of theory has relative Gibbs free energy of 30 kJ/mol from the global minimum conformer of Boc-^{D}Pro-Gly-Ala-NHBn-OMe has been provided in figure 5.11. The most stable conformer of Boc-^{D}Pro-Gly-Ala-NHBn-OMe is named PGA1-F-C10-C10.[242] The conformers have been named based on the hydrogen bonding pattern of the three N-H groups (starting from N-terminus of peptide) present in the peptide molecule. For example, the conformer named PGA1-F-C10-C10 indicates that the glycine NH is free while the alanine NH and NHBn NH groups are involved in type II$^{'}$ β-turn and type I β-turn, respectively. One of the most important findings from the calculations is that most of the low-energy conformers for this peptide sequence are folded in nature. The extended conformation of the backbone is observed in conformer PGA19-C5-C5-F having a C5-C5-F type of hydrogen bonding network which is quite high in energy in comparison to the global minimum. In the case of conformer PGA19-C5-C5-F, the glycine and alanine N-H are involved in the intra-residue C5 hydrogen bond whereas the NHBn N-H is free from any hydrogen bond. This conformer has an extended conformation similar to the conformation adopted by Z-Gly-Pro-OH and Boc-Gly-^{D}Pro-NHBn-OMe in isolated gas-phase conditions.

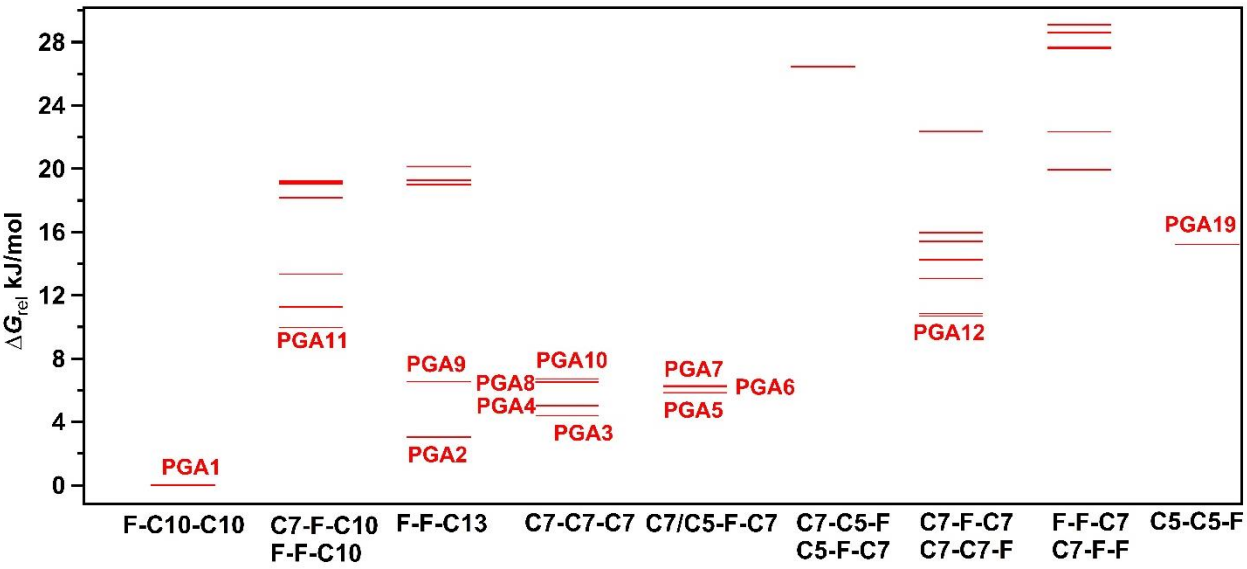

Figure 5.11. Energy landscape of a few low energy conformers of Boc-^{D}Pro-Gly-Ala-NHBn-OMe, having energies within 30 kJ/mol compared to the global minimum conformer, calculated at 300 K at the B97-D/6-31+G(d) level of theory. The conformers are classified into different categories based on their hydrogen bonding patterns such as C13, C10, C7, C5 etc.

The most stable conformer PGA1-F-C10-C10 shows that it has a helical nature with C10-C10 conformation. Thus, the calculation further validates the argument that the propensity to form a β-turn in Pro-Gly-X residue further increases if the X residue is alanine. Apart from the conformers having β-turn, some conformers are solely stabilized by C7 or C7-C7-C7 (PGA3-C7-C7-C7, PGA8-C7-C7-C7, and PGA10-C7-C7-C7) hydrogen-bonded network. The Conformers PGA2-F-F-C13 and PGA9-F-F-C13 even show the existence of α-turn along the main peptide chain. The structures of twelve low-energy conformers optimized at the B97-D/6-31+G(d) level of theory have been shown in figure 5.12.

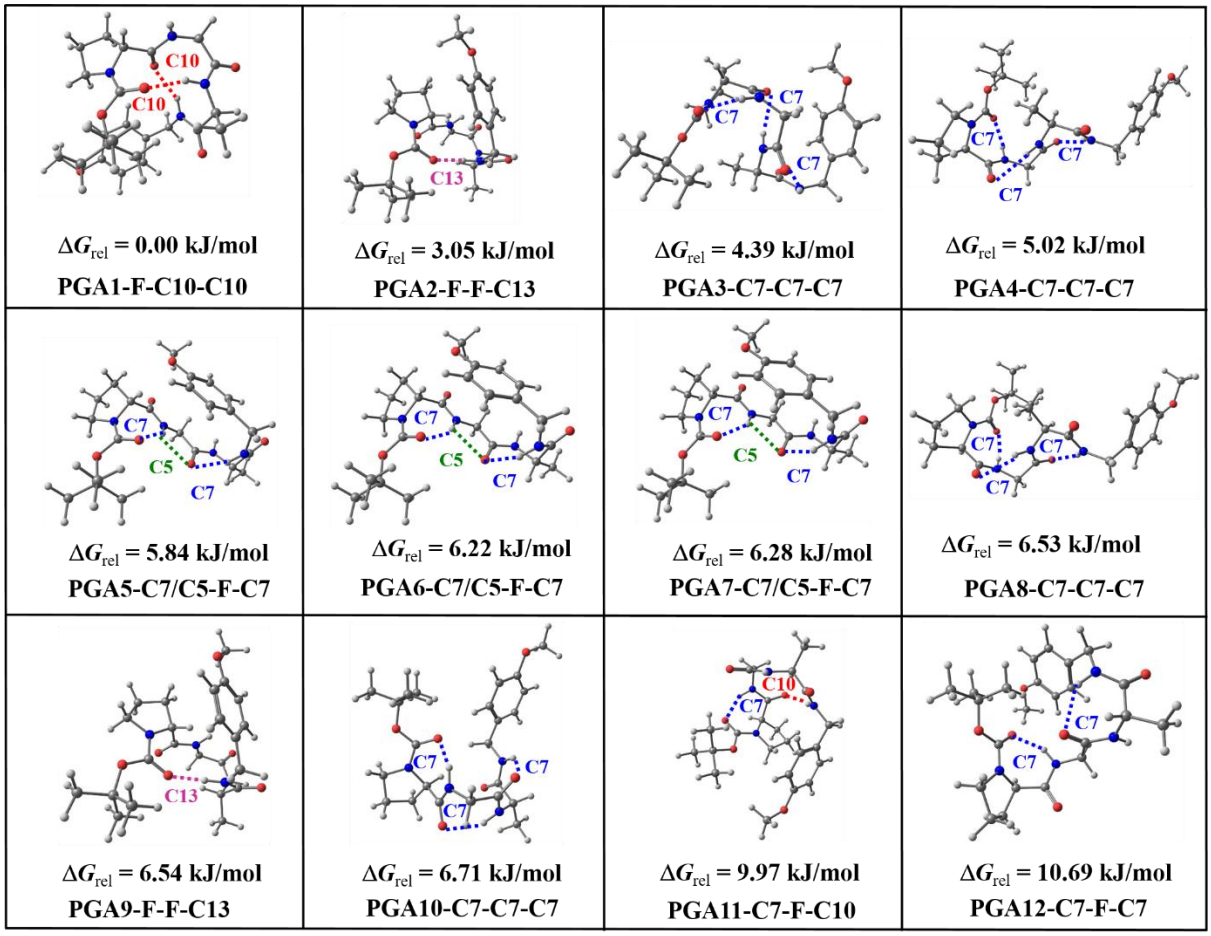

Figure 5.12. Structures of low-energy conformers of Boc-^{D}Pro-Gly-Ala-NHBn-OMe calculated at 300 K at the B97-D/6-31+G(d) level of theory.

Most of the conformers of Boc-^{D}Pro-Gly-Ala-NHBn-OMe exist in folded conformation due to the presence of different types of turns such as α, β, and γ. In conformer PGA2-F-F-C13 and PGA9-F-F-C13, there is an α-turn (C13) formation between the carbonyl group of Boc residue and the amide group of NHBn residue. Conformers such as PGA3-C7-C7-C7, PGA4-C7-C7-C7, and PGA8-C7-C7-C7 have structures where all the N-H groups are involved in γ-turn. A comparison of the energetics of the low energy conformers of Boc-^{D}Pro-Gly-NHBn-OMe and Boc-Gly-^{D}Pro-NHBn-OMe calculated at 300 K at the B97-D/6-31+G(d) level of theory has been provided in Table 5.2.

Table 5.2. Comparison of relative Gibbs free energy of low energy conformers of Boc-^{D}Pro-Gly-Ala-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe at 300 K after optimization at B97-D/6-31+G(d) level of theory

Pro-Gly-Ala		**Pro-Gly**	
Conformers	**ΔG_{rel} (300 K)**	**Conformers**	**ΔG_{rel} (300 K)**
PGA1-F-C10-C10	0	**PG1-C7-C7**	0.00
PGA2-F-F-C13	3.05	**PG4-C7-F**	4.69
PGA3-C7-C7-C7	4.39	**PG3-F-C10**	3.95
PGA4-C7-C7-C7	5.02	**PG2-F-C7**	3.58
PGA5-C7/C5-F-C7	5.84	**PG5-F-C10**	5.88
PGA6-C7/C5-F-C7	6.22	**PG7-C7-F**	8.95
PGA7-C7/C5-F-C7	6.28	**PG8-C7-C7**	9.11
PGA8-C7-C7-C7	6.53	**PG9-F-C10**	9.24
PGA9-F-F-C13	6.54	**PG6-F-C10**	8.14
PGA10-C7-C7-C7	6.71	**PG10-C7-F**	10.93

It can be noticed from Table 5.2 that the most stable conformer in the case of Boc-^{D}Pro-Gly-NHBn-OMe is the PG1-C7-C7 conformer while PG1-F-C10-C10 is the most stable conformer in the case of Boc-^{D}Pro-Gly-Ala-NHBn-OMe. Although β-turn is not the most preferred structure in the case of the Pro-Gly dipeptide in the gas phase, this structure becomes the

prominent one as the Ala residue is added after the Pro-Gly sequence in the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide. This finding demonstrates that the Ala residue increases the strength of the β-turn in the Pro-Gly peptide and makes the β-turn structure most favorable without the presence of any solvent molecule. A comparison of the hydrogen bond parameters in the β-turn structures of Boc-^{D}Pro-Gly-Ala-NHBn-OMe (PGA1-F-C10-C10) and Boc-^{D}Pro-Gly-NHBn-OMe (PG3-F-C10) optimized at the B97-D/6-31+G(d) level of theory is shown in Figure 5.13. It is safe to say from the comparison that the hydrogen bond in the β-turn structure of Boc-D-Pro-Gly-Ala-NHBn-OMe is slightly stronger than that of Boc-D-Pro-Gly-NHBn-OMe.

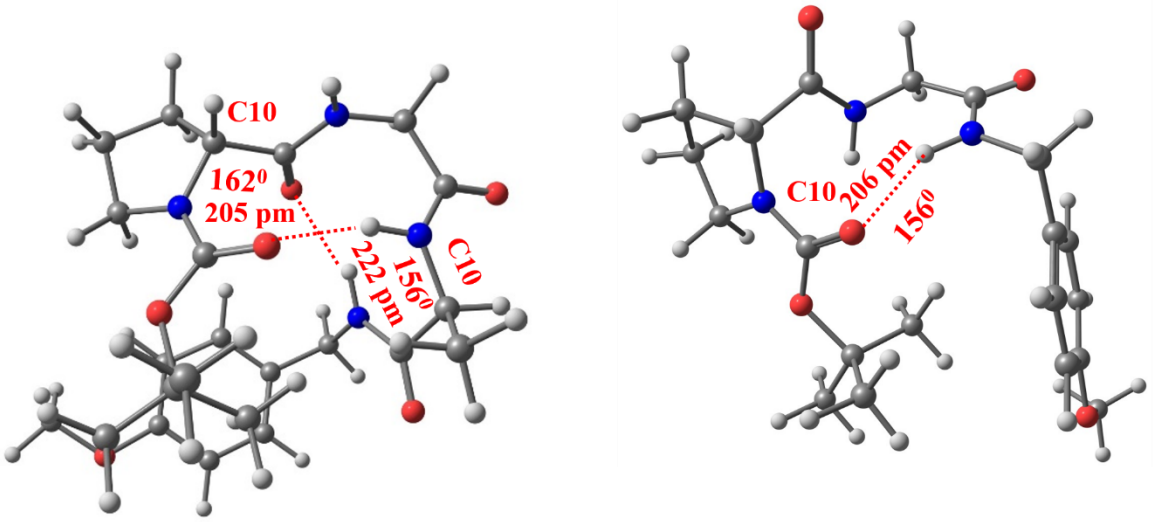

Figure 5.13. A comparison between the β-turn forming conformers of Pro-Gly-Ala and Pro-Gly sequence calculated at 300 K at the B97-D/6-31+G(d) level of theory.

The strength of the C10 hydrogen bond in Boc-D-Pro-Gly-Ala-NHBn-OMe (PGA1-C7-F-C10) and Boc-D-Pro-Gly-NHBn-OMe (PG3-F-C10) has been probed further through NBO calculations. The NBO view for the overlap between the orbitals involved in the C10 hydrogen bond in the two peptides is shown in figure 5.14. The NBO 2nd order perturbation energies

again point out that the C10 hydrogen bond in Boc-D-Pro-Gly-Ala-NHBn-OMe is stronger than that in Boc-D-Pro-Gly-NHBn-OMe.

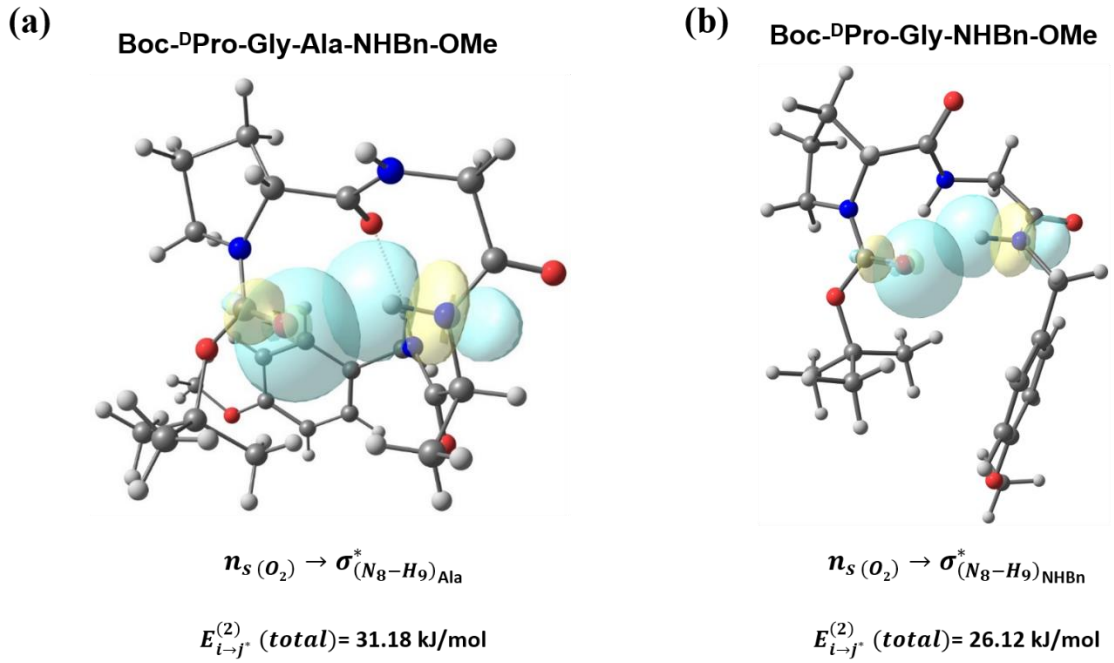

Figure 5.14. Figure shows the NBOs overlap in case of β-turn formation for (a) Boc-^{D}Pro-Gly-Ala-NHBn-OMe and (b) Boc-^{D}Pro-Gly-NHBn-OMe with their respective 2nd order perturbation energies. Calculations were performed at B97-D/6-31+G(d) level of theory.

The NBO interaction energy between oxygen lone pair and N-H antibonding in the case of the pro-gly-ala sequence is significantly higher (31.18 kJ/mol) compared to the Pro-Gly sequence (26.12 kJ/mol). This further highlights the importance of residue X in a Pro-Gly-Ala sequence in the β-turn stabilization.

5.2.5.2 Gas phase conformational preferences of tri-peptides

There have been significant studies regarding the conformational preferences of tri-peptides in gas phase by various spectroscopic groups.[1, 10, 121, 123, 242, 243] A summary of the relevant reported structures of various tri-peptides consisting of α-amino acids has been presented in the table 5.3.

Table 5.3. A summary of the conformation adopted by most stable conformer of various tri-peptides studied in the gas phase

Tri-peptide sequence	Hydrogen bond network of most stable structure
Ac-Phe-Gly-Gly-NH_2[a]	C10-C10
Z-$(Gly)_3$-OH[b]	C7-C7-C7
Ac-Phe-Ala-Ala-NH_2[c]	C5-C7-C7
Ac-Ala-Phe-Ala-NH_2[c]	C10-C10
Ac-Ala-Ala-Phe-NH_2[c]	C7-C10
Ac-Aib-Phe-Aib-NH_2[d]	C10-C10
Z-Pro-Leu-Gly-NH_2[e]	C10-C7

[a] Mons and co-workers, 2005[123]. b Zwier and co-workers, 2012[1]. **[c]** Mons and co-workers, 2005[243]. **[d]** Mons and co-workers, 2007[242], **[e]** Fujii and co-workers, 2013[120]

Literature reports of the study of a few protected tripeptides in the gas phase summarized in Table 5.3 show that the conformation adopted by the most stable conformer can be categorized in mainly three types. The most common conformation preferred by the tripeptides is C10-C10 (β-β turn) while another favorable conformation is the mixed C10-C7. C7-C7-C7 (γ-γ-γ turn) or 2_7 ribbon structure is also observed for some peptides. The ^{D}Pro-Gly-Ala peptide studied in the current chapter of the thesis through FTIR, solution phase NMR spectroscopy, X-ray crystallography and DFT calculations reveals that the most preferred conformation of the tripeptide is C10-C10. Thus, a comparison of the observed conformation of the ^{D}Pro-Gly-Ala peptide studied here in the condensed phase with the preferred conformation of a few tripeptides reported in the literature (Table 5.3) suggests that the most stable conformer of the ^{D}Pro-Gly-Ala sequence might adopt a C10-C10 conformation in the gas phase as well.

5.2.5.3 Solution phase calculations

The Polarizable Continuum Model (PCM) in Gaussian 09 has been used to calculate the conformers energetics of Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CHCl_3$ solvent.[135, 244] The PCM method uses a countless number of solutes having a cavity and apparent surface charges (ASCs).[244-246] The solvent molecules are represented as the polarizable continuum having a dielectric constant ε. The PCM calculation results for Boc-^{D}Pro-Gly-Ala-NHBn-OMe in $CHCl_3$ solvent have been shown in figure 5.15. Results from the PCM calculation show that the conformer PGA1-F-C10-C10 is the global minimum conformer. Energetics from PCM calculations in $CHCl_3$ solvent follows a similar trend compared to the gas phase DFT calculations at the same level of theory. It is interesting to note that the conformation of the Boc-DPro-Gly-Ala-NHBn-OMe peptide observed in the solution phase through the NMR and IR spectroscopy analysis as well as in the crystal form through the XRD analysis is similar. The extended conformer PGA19-C5-C5-F is significantly higher in energy (around 12 kJ/mol) compared to the global minimum conformer PGA1-F-C10-C10. This further strengthens the trend that Pro-Gly-Ala residue has a very high propensity to form β-turn in solution.

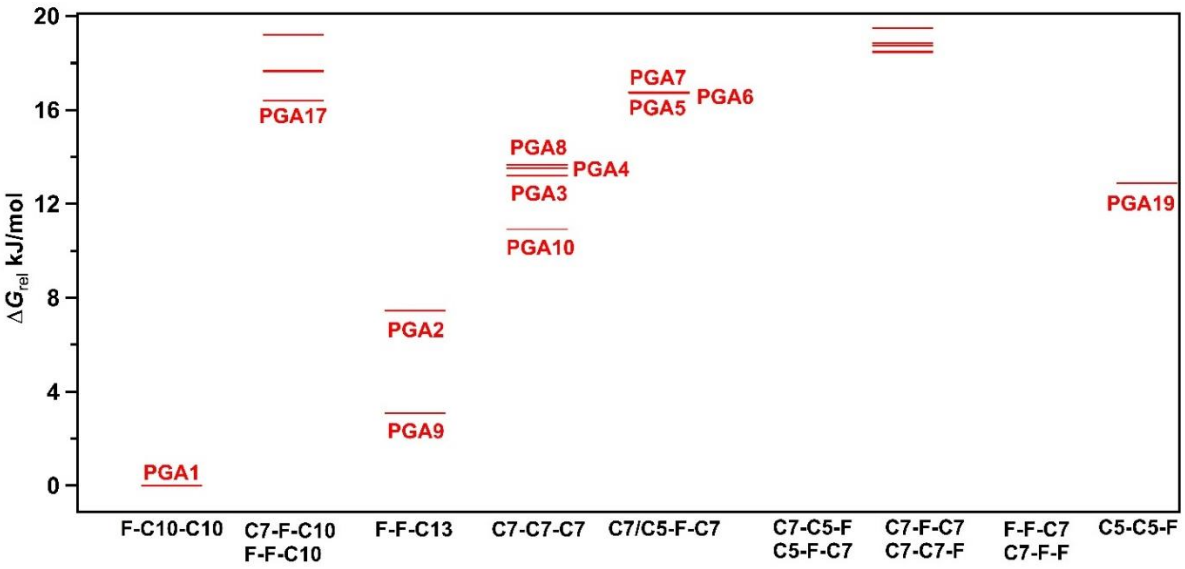

Figure 5.15. Structures of low-energy conformers of Boc-^{D}Pro-Gly-Ala-NHBn-OMe calculated at 300 K at the B97-D/6-31+G(d) level of theory.

5.3 Conclusion

Conformational preferences of Boc-^{D}Pro-Gly-Ala-NHBn-OMe have been studied through solution-phase NMR spectroscopy, X-ray crystallography, solution-phase FTIR spectroscopy, and quantum chemistry calculations. The observed conformer of the peptide from various solution-phase spectroscopy and XRD results has been found to have a C10-C10 structure. It is found that the Gly N-H group is free while the N-H groups of Ala and NHBn moieties are C10 hydrogen bonded. The C10 hydrogen bond between the N-H group NHBn and C=O of Pro is stronger than that between the N-H group of Ala and the C=O group of Boc. A comparison of the solution phase FTIR spectra of the Boc-^{D}Pro-Gly-Ala-NHBn-OMe and Boc-^{D}Pro-Gly-NHBn-OMe peptides in the N-H stretching frequency region indicates that the C10 hydrogen bond (β-turn) involving the Pro-Gly sequence is stronger in the Pro-Gly-Ala tripeptide than the Pro-Gly dipeptide. Thus, the FTIR spectra demonstrate that the presence of the Ala residue increases the strength of the β-turn of the Pro-Gly peptides. The results obtained from the quantum chemistry calculations of the Boc-^{D}Pro-Gly-Ala-NHBn-OMe and Boc-^{D}Pro-

Gly-NHBn-OMe peptides in solution, as well as gas-phase, corroborate the finding from the FTIR, NMR, and XRD data. Interestingly, the gas phase calculations show that β-turn is the most preferred structure for the Pro-Gly-Ala peptide but the same is not true for the Pro-Gly dipeptide. However, the β-turn structure of the Pro-Gly dipeptide is observed in the solution phase IR and NMR experiments. In the future, it will be interesting to study the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide using various gas-phase laser spectroscopic techniques and compare the results with the condensed phase data obtained by NMR, XRD, and FTIR.

Chapter 6

Summary and Future Direction

6.1. Conclusion

We have studied the conformational preferences of small peptides containing glycine, proline, and alanine in the gas as well as condensed phase. Gas-phase conformational preferences of selected peptides were analyzed with the help of gas-phase spectroscopic techniques such as R2PI, RIDIRS, IR-UV hole-burn, and UV-UV hole-burn spectroscopy combined with quantum chemistry calculations at various levels. Condensed phase conformational preferences of the peptides were studied with the help of spectroscopic techniques such as 2D-NMR, FTIR, and XRD spectroscopy. The peptides studied in this chapter are Z-Gly-Pro-OH, Boc-^{D}Pro-Gly-NHBn-OMe, Boc-Gly-^{D}Pro-NHBnOMe, and Boc-^{D}Pro-Gly-Ala-NHBn-OMe. Only the Z-Gly-Pro-OH is unprotected at one terminal while the rest of the peptides are protected at both terminals.

Stabilization of the conformers of a dipeptide exclusively by a weak intra-residue C5 hydrogen bond has been studied by taking the Z-Gly-Pro-OH molecule as a model peptide in isolated gas-phase conditions. The role of the C5 hydrogen bond in the stabilization of extended conformations of proteins and polypeptides is emphasized in the literature only recently. Unlike other hydrogen bonds present in proteins such as C13 (α-turn), C10 (β-turn), C7(γ-turn), and C6 (δ-turn), which are inter-residue, the C5 hydrogen bond exists within a single amino acid residue of proteins and peptides. Our study on Z-Gly-Pro-OH in gas phase conditions reveals that the most stable structure of the peptide exists in an extended conformation which is solely stabilized by the intra-residue C5 hydrogen bond. The other observed conformation of Z-Gly-Pro-OH in the gas phase has a folded conformation having an O–H ⋯π hydrogen bond interaction. The gas-phase studies have been performed by spectroscopic techniques such as resonant 2-photon ionization and IR-UV double resonance spectroscopy along with DFT

calculations using Gaussian 09. However, the study of the peptides protected at both terminals will be mimicking much better the structural prediction of the peptide segment containing Gly and Pro residues in polypeptides and proteins.

We chose to study the conformational preferences of capped gly-pro and pro-gly sequences in the gas phase as well as the condensed phase. It was earlier established through solution-phase NMR, CD, and IR spectroscopic studies of several dipeptides that the gly-pro sequence has more propensity to form extended conformation whereas the pro-gly sequence prefers the folded β-turn conformation. Similar conformational preferences of the gly-pro and pro-gly sequences have been observed in the proteins. Boc-^{D}Pro-Gly-NHBn-OMe and Boc-Gly-^{D}Pro-NHBn-OMe peptides were used to study the sequence-dependent effect on conformational preferences of peptides in the gas-phase as well as condensed phase. The gas-phase spectroscopic techniques such as R2PI, RIDIR, IR-UV hole burning, and UV-UV hole burning spectroscopy along with DFT calculations at various levels (B3LYP-D3/def2TZVPP, ωB97X-D/def2TZVPP, B97-D3/def2TZVPP, and M062-X/6-311++G(2d,2p) level of theory) were used to predict the conformers observed experimentally in the gas-phase. The gas-phase studies revealed that all three observed conformers of the gly-pro peptide exist in an extended β-strand C5-C7 conformation whereas both the observed conformers of the pro-gly sequence adopt a 2_7-ribbon structure (C7-C7). The solution phase studies through 2-D NMR and FTIR spectroscopy suggest that Boc-Gly-^{D}Pro-NHBn-OMe exists in an extended C5-C7 conformation similar to the conformer observed in the gas phase. On the other hand, Boc-^{D}Pro-Gly-NHBn-OMe exists in a β-turn conformation in the solution phase contrary to the 2_7-ribbon structure observed in the gas phase. The crystal structure of the gly-pro sequence again shows extended C5-C7 conformation where the glycine NH group is involved in the intra-residue C5 hydrogen bond and the NHBn NH group forms a C7 (γ-turn) hydrogen bond. The overall study presents the conformational preferences of pro-gly and gly-pro sequences in the gas phase and

condensed phase. The missing β-turn conformation in the gas phase in the case of pro-gly could be due to the small peptide chain or lack of solvent. The propensity to form β-turn can be further studied in larger peptide sequences such as Pro-Gly-X (where X=Gly, Ala, Ile, and Leu).

The effect of residue X on the stabilization of the β-turn conformation in the Pro-Gly-X sequence has been explored by Brahmachari and co-workers. It has been shown that the X residue has an increasing order of stabilization on β-turn conformation in the Pro-Gly-X sequence as X=Leu > Ala > Gly, Ile > Phe. We have studied the conformational preferences of the Pro-Gly-X (X=Ala) sequence in the condensed phase employing techniques such as X-ray crystallography, DFT calculations, NMR spectroscopy, and FTIR spectroscopy in solution. The X-ray analysis of Boc-^{D}Pro-Gly-Ala-NHBn-OMe reveals that the molecule exists in a double helical C10-C10 structure where the Glycine NH is free and the alanine and NHBn NH groups are involved in type-II' and type I β-turn hydrogen bond, respectively. The 2D NMR and DMSO-d6 titration of the peptide sequence in $CDCl_3$ solvent reveal that it has a similar conformation in the solution phase. The FTIR spectroscopy in $CDCl_3$ further credits the increase of the strength of β-turn conformation as compared to the pro-gly sequence. Various conformers of the molecules generated from MarvinSketch were subjected to DFT calculations at the B97-D/6-31+G(d) level of theory. The calculations performed at the B97-D/6-31+G(d) level of theory show that the global minimum structure has a similar conformation (PGA1-F-C10-C10) as compared to the crystal structure. The PCM calculations at the B97-D/6-31+G(d) level of theory in chloroform solvent ($CHCl_3$) also corroborate the structure of the Boc-^{D}Pro-Gly-Ala-NHBn-OMe peptide obtained from the XRD and solution phase spectroscopy.

6.2. Future perspectives

In chapter 5 we have shown that the peptide Boc-^{D}Pro-Gly-Ala-NHBn-OMe shows a helical double β-turn conformation in solution as well as crystal structure. The DFT calculations at the B97-D level of theory also show that the most stable conformer at 300 K has C10-C10 helical conformation. The gas-phase experiments on this peptide will further reveal whether the helical C10-C10 conformation is stable due to the solvent effect or it is an intrinsic property. Further, we would like to study the effect of the D/L configuration of proline residue by studying the conformational preferences of Boc-^{L}Pro-Gly-Ala-NHBn-OMe in the gas-phase as well as solution-phase. This study will enable us to directly compare the strength of the β-turn formation of peptides by D and L residues of the chiral amino acids.

In future, it will be intriguing to compare the folding motif of the Boc-Gly- ^{D}Pro-Ala-NHBN-OMe peptide in the solution as well as gas phase with that of the Boc- ^{D}Pro-Gly-Ala-NHBN-OMe peptide already reported here in the solution phase. We have already observed that the Gly-Pro sequence in the dipeptide forms an extended structure in the gas phase as well as condensed phase. This study will reveal conformational preferences of Gly-Pro-X where X=Ala, in the gas-phase and condensed phase. Figure 6.1 shows one of the optimized structures of Boc-Gly-^{D}Pro-Ala-NHBn-OMe obtained from the B3LYP/6-31+G(d) level of calculation. The optimized structure shows an extended structure having a C5-C7 hydrogen bond network similar to the preferred conformation of Boc-Gly-^{D}Pro-NHBn-OMe peptide observed from our gas and condensed phase studies.

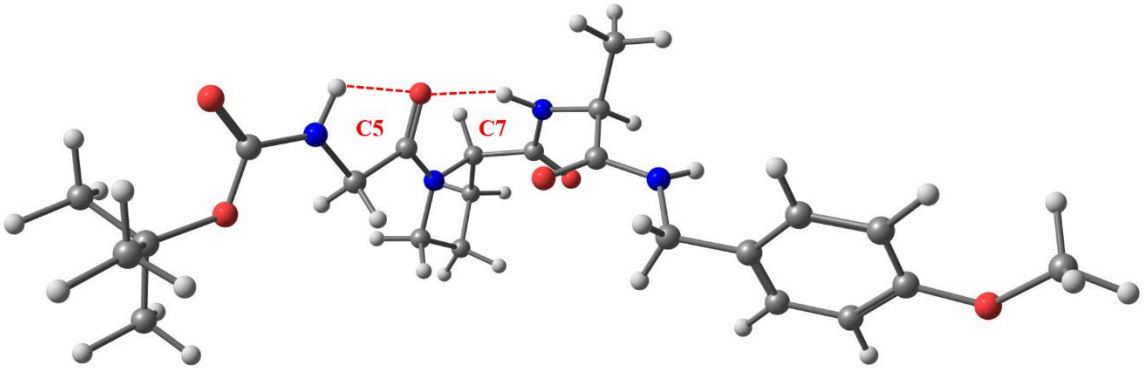

Figure 6.1. The optimized structure of Boc-Gly-^{D}Pro-Ala-NHBn-OMe showing C5-C7 hydrogen bond network after optimization at B3LYp/6-31+G(d) level of theory.

It has been earlier discussed that the strength of the β-turn depends on the residue X in a sequence of Pro-Gly-X, where the order of the stability is according to X=Leu > Ala > Gly, Ile > Phe. We want to further study the conformational preferences of these peptides in the gas phase as well as the condensed phase. Figure 6.2 shows the skeletal structures of the peptides we want to study in the future.

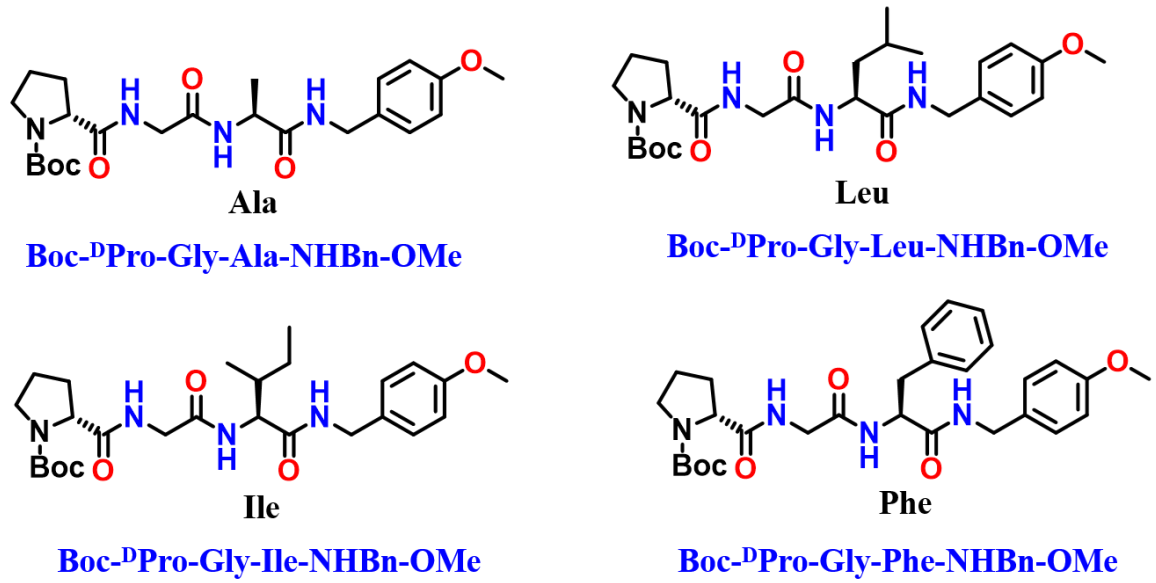

Figure 6.2. Skeletal structures of capped Gly-Pro-Ala, Gly-Pro-Leu, Gly-Pro-Ile and Gly-Pro-Phe sequences.

It will be interesting to explore the strength of the β-turn in all of these capped tripeptides or model tetrapeptides. This study will enable us to understand the role of the neighbouring group

residue or its side chain on the stabilization of the β-turn involving the Pro-Gly sequence. In the case of the Pro-Gly-Ala sequence, the side chain of Ala is small (hydrogen) compared to that of the other neighbouring residues in the sequences such as Pro-Gly-Leu, Pro-Gly-Ile, and Pro-Gly-Phe, where the side chain is much bulkier.

It has been also reported that the Pro-Gly-Val sequence has a very less propensity for the β-turn formation. Although the Valine and Leucine differ very slightly in their side-chain moiety, the effect on the β-turn stabilization by these two residues is completely opposite. It would be interesting to study the effect of the side-chain residue on the stabilization of the β-turn in the gas phase as well as condensed phase.

Incorporation of aromatic amino acids such as tyrosine, phenylalanine and tryptophan at position X in Pro-Gly-X sequence would also enable us to study the conformational preference of these sequences that will be free from any effect from the chromophore group NHBn-OMe, which is used for electronic excitation to obtain both electronic and vibrational spectra in the gas phase.

BIBLIOGRAPHY

References

1. J. C. Dean, E. G. Buchanan and T. S. Zwier, *J. Am. Chem. Soc.*, 2012, **134**, 17186-17201.
2. P. Y. Chou and G. D. Fasman, *Biochemistry*, 1974, **13**, 222-245.
3. F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *Journal of molecular biology*, 1977, **112**, 535-542.
4. K. Takano, J. Funahashi, Y. Yamagata, S. Fujii and K. Yutani, *Journal of molecular biology*, 1997, **274**, 132-142.
5. G. Otting, E. Liepinsh and K. Wuthrich, *Science*, 1991, **254**, 974-980.
6. U. Schieborr, M. Vogtherr, B. Elshorst, M. Betz, S. Grimme, B. Pescatore, T. Langer, K. Saxena and H. Schwalbe, *Chembiochem*, 2005, **6**, 1891-1898.
7. V. R. Mundlapati, Z. Imani, V. C. D'mello, V. Brenner, E. Gloaguen, J. P. Baltaze, S. Robin, M. Mons and D. J. Aitken, *Chem. Sci.*, 2021, **12**, 14826-14832.
8. E. Gloaguen, M. Mons, K. Schwing and M. Gerhards, *Chem. Rev.*, 2020, **120**, 12490-12562.
9. A. M. Rijs and J. Oomens, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, eds. A. M. Rijs and J. Oomens, Springer International Publishing, Cham, Switzerland, 2015, pp. 1-42.
10. W. Chin, J. P. Dognon, C. Canuel, F. Piuzzi, I. Dimicoli, M. Mons, I. Compagnon, G. von Helden and G. Meijer, *J. Chem. Phys.*, 2005, **122**.
11. B. Khatri, P. Majumder, J. Nagesh, A. Penmatsa and J. Chatterjee, *Chem. Sci.*, 2020, **11**, 9480-9487.
12. K. C. Chou, *Anal. Biochem.*, 2000, **286**, 1-16.
13. R. S. Rapaka, V. Renugopalakrishnan, D. Urry and R. S. Bhatnagar, *Biochemistry*, 1978, **17**, 2892-2898.
14. S. K. Brahmachari and V. S. Ananthanarayanan, *Proc. Natl. Acad. Sci.*, 1979, **76**, 5119.
15. K.-C. Chou, *Anal. Biochem.*, 2000, **286**, 1-16.
16. J. K. Myers and C. N. Pace, *Biophys. J.*, 1996, **71**, 2033-2039.
17. R. W. Newberry and R. T. Raines, *Nature Chemical Biology*, 2016, **12**, 1084.
18. R. A. Nyquist, *Spectrochimica Acta*, 1963, **19**, 509-519.
19. S. Kumar, K. K. Mishra, S. K. Singh, K. Borish, S. Dey, B. Sarkar and A. Das, *J. Chem. Phys.*, 2019, **151**, 104309.
20. V. R. Mundlapati, Z. Imani, V. C. D'Mello, V. Brenner, E. Gloaguen, J.-P. Baltaze, S. Robin, M. Mons and D. J. Aitken, *Chem. Sci.*, 2021, **12**, 14826-14832.
21. E. R. Stimson, S. S. Zimmerman and H. A. Scheraga, *Macromolecules*, 1977, **10**, 1049-1060.
22. S. K. Brahmachari, R. S. Rapaka, R. S. Bhatnagar and V. S. Ananthanarayanan, *Biopolymers*, 1982, **21**, 1107-1125.

23. E. Gloaguen and M. Mons, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, eds. A. M. Rijs and J. Oomens, Springer International Publishing, Cham, Switzerland, 2015, pp. 225-270.
24. T. R. Rizzo, Y. D. Park and D. H. Levy, *J. Am. Chem. Soc.*, 1985, **107**, 277-278.
25. M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585-612.
26. B. C. Dian, A. Longarte, S. Mercier, D. A. Evans, D. J. Wales and T. S. Zwier, *J. Chem. Phys.*, 2002, **117**, 10688-10702.
27. S. Habka, W. Y. Sohn, V. Vaquero-Vara, M. Géléoc, B. Tardivel, V. Brenner, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3411-3423.
28. J. Brinckmann, in *Collagen: Primer in Structure, Processing and Assembly*, eds. J. Brinckmann, H. Notbohm and P. K. Müller, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 1-6.
29. M. Alauddin, H. S. Biswal, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2169-2178.
30. L. Pauling, R. B. Corey and H. R. Branson, *Proc. Natl. Acad. Sci.*, 1951, **37**, 205.
31. J. A. Ernst, R. T. Clubb, H. X. Zhou, A. M. Gronenborn and G. M. Clore, *Science*, 1995, **267**, 1813-1817.
32. K. Saxena, B. Elshorst, H. Berk, M. Betz, S. Grimme, T. Langer, B. Pescatore, U. Schieborr, M. Vogtherr and H. Schwalbe, *J. Biomol. NMR*, 2005, **33**, 136-136.
33. H. A. Scheraga, *Carlsberg Research Communications*, 1984, **49**, 1-55.
34. Y. Fu, J. Zhao and Z. Chen, *Computational and Mathematical Methods in Medicine*, 2018, **2018**, 3502514.
35. J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror and D. E. Shaw, *Current Opinion in Structural Biology*, 2009, **19**, 120-127.
36. R. L. Momparler, M. Karon, S. E. Siegel and F. Avila, *Cancer Research*, 1976, **36**, 2891-2895.
37. A. Ben-Ze'ev, S. R. Farmer and S. Penman, *Cell*, 1980, **21**, 365-372.
38. S. Bandyopadhyay, S. Chakraborty, S. Balasubramanian, S. Pal and B. Bagchi, *The Journal of Physical Chemistry B*, 2004, **108**, 12608-12616.
39. Š. Vajda, R. Jimenez, S. J. Rosenthal, V. Fidler, G. R. Fleming and E. W. Castner, *Journal of the Chemical Society, Faraday Transactions*, 1995, **91**, 867-873.
40. G. Wolynes Peter, N. Onuchic Jose and D. Thirumalai, *Science*, 1995, **267**, 1619-1620.
41. J. T. Ngo, J. Marks and M. Karplus, in *The Protein Folding Problem and Tertiary Structure Prediction*, eds. K. M. Merz and S. M. Le Grand, Birkhäuser Boston, Boston, MA, 1994, pp. 433-506.
42. R. Zwanzig, A. Szabo and B. Bagchi, *Proc. Natl. Acad. Sci.*, 1992, **89**, 20-22.
43. M. Karplus, *Nature Chemical Biology*, 2011, **7**, 401-404.
44. B. R. Singh, in *Infrared Analysis of Peptides and Proteins*, American Chemical Society 1999, vol. 750, ch. 1, pp. 2-37.
45. Y. Qian, M. H. Engel, S. A. Macko, S. Carpenter and J. W. Deming, *Geochimica et Cosmochimica Acta*, 1993, **57**, 3281-3293.
46. B. Anfinsen Christian, *Science*, 1973, **181**, 223-230.
47. F. Sanger and H. Tuppy, *Biochem J*, 1951, **49**, 463-481.
48. F. Cárdenas-Bailón, G. Osorio-Revilla and T. Gallardo-Velázquez, *Journal of Microencapsulation*, 2013, **30**, 409-424.
49. J. A. Crapster, I. A. Guzei and H. E. Blackwell, *Angewandte Chemie - International Edition*, 2013, **52**, 5079-5084.
50. S. Hanessian, X. Luo, R. Schaum and S. Michnick, *J. Am. Chem. Soc.*, 1998, **120**, 8569-8570.
51. R. H. Pain and B. Robson, *Nature*, 1970, **227**, 62-&.
52. W. R. Krigbaum and S. P. Knutton, *Proc. Natl. Acad. Sci.*, 1973, **70**, 2809.
53. F. E. Cohen, M. J. E. Sternberg and W. R. Taylor, *Journal of molecular biology*, 1981, **148**, 253-272.
54. J. W. Pflugrath and F. A. Quiocho, *Journal of molecular biology*, 1988, **200**, 163-180.

55. Q.-P. Cao, S. J. Duguay, E. Plisetskaya, D. F. Steiner and S. J. Chan, *Molecular Endocrinology*, 1989, **3**, 2005-2010.
56. K. Chen and L. Kurgan, in *Protein Supersecondary Structures*, ed. A. E. Kister, Humana Press, Totowa, NJ, 2013, pp. 63-86.
57. X. Yu, C. Wang and Y. Li, *BMC Bioinformatics*, 2006, **7**, 187.
58. I. M. Klotz, N. R. Langerman and D. W. Darnall, *Annual review of biochemistry*, 1970, **39**, 25-62.
59. J. Janin, R. P. Bahadur and P. Chakrabarti, *Quarterly Reviews of Biophysics*, 2008, **41**, 133-180.
60. L. Miallau, W. N. Hunter, S. M. McSweeney and G. A. Leonard, *Journal of Biological Chemistry*, 2007, **282**, 19948-19957.
61. T. Svedberg and J. B. Nichols, *J. Am. Chem. Soc.*, 1927, **49**, 2920-2934.
62. C. L. Phillips, B. Ullman, R. G. Brennan and C. P. Hill, *EMBO J*, 1999, **18**, 3533-3545.
63. K. D. Watenpaugh, T. N. Margulis, L. C. Sieker and L. H. Jensen, *Journal of molecular biology*, 1978, **122**, 175-190.
64. C. Toniolo, *Crit. rev. biochem.*, 1980, **9**, 1-44.
65. J. Bella, M. Eaton, B. Brodsky and H. Berman, *Science*, 1994, **266**, 75-81.
66. M. D. Shoulders and R. T. Raines, *Annual review of biochemistry*, 2009, **78**, 929-958.
67. I. Karle, H. N. Gopi and P. Balaram, *Proc. Natl. Acad. Sci.*, 2002, **99**, 5160-5164.
68. K. Kim and C. Frieden, *Protein Science*, 1998, **7**, 1821-1828.
69. P. Y. Chou and G. D. Fasman, *Biophys. J.*, 1979, **26**, 367-383.
70. J. L. Crawford, W. N. Lipscomb and C. G. Schellman, *Proc. Natl. Acad. Sci.*, 1973, **70**, 538.
71. S. Kumar and M. Bansal, *Biophys. J.*, 1998, **75**, 1935-1944.
72. S. Kumar and M. Bansal, *Proteins: Structure, Function, and Bioinformatics*, 1998, **31**, 460-476.
73. C. Chothia, M. Levitt and D. Richardson, *Proc. Natl. Acad. Sci.*, 1977, **74**, 4130-4134.
74. M. Crisma, F. Formaggio, A. Moretto and C. Toniolo, *Peptide Science*, 2006, **84**, 3-12.
75. D. Eisenberg, *P Natl Acad Sci USA*, 2003, **100**, 11207-11210.
76. H. W. Long and R. Tycko, *J. Am. Chem. Soc.*, 1998, **120**, 7039-7048.
77. C. Chothia, *Journal of molecular biology*, 1973, **75**, 295-302.
78. F. Salemme and D. Weatherford, *Journal of molecular biology*, 1981, **146**, 119-141.
79. C. Soto, M. S. Kindy, M. Baumann and B. Frangione, *Biochemical and biophysical research communications*, 1996, **226**, 672-680.
80. O. N. Antzutkin, J. J. Balbach, R. D. Leapman, N. W. Rizzo, J. Reed and R. Tycko, *Proc. Natl. Acad. Sci.*, 2000, **97**, 13045-13050.
81. W. Chin, J. P. Dognon, F. Piuzzi, B. Tardivel, I. Dimicoli and M. Mons, *J. Am. Chem. Soc.*, 2005, **127**, 707-712.
82. J. A. Smith, L. G. Pease and K. D. Kopple, *Critical Reviews in Biochemistry*, 1980, **8**, 315-399.
83. C. M. Venkatachalam, *Biopolymers*, 1968, **6**, 1425-1436.
84. H. Singh, S. Singh and G. P. S. Raghava, *Proteins: Structure, Function, and Bioinformatics*, 2015, **83**, 910-921.
85. J. D. Tyndall, B. Pfeiffer, G. Abbenante and D. P. Fairlie, *Chem. Rev.*, 2005, **105**, 793-826.
86. C. Fang, Y. Shang and D. Xu, *Proteins: Structure, Function, and Bioinformatics*, 2020, **88**, 143-151.
87. M. Avignon, P. V. Huong, J. Lascombe, M. Marraud and J. Neel, *Biopolymers*, 1969, **8**, 69-89.
88. F. R. Maxfield, S. J. Leach, E. R. Stimson, S. P. Powers and H. A. Scheraga, *Biopolymers*, 1979, **18**, 2507-2521.
89. A. Ravi and P. Balaram, *Tetrahedron*, 1984, **40**, 2577-2583.
90. C. P. Rao, P. Balaram and C. N. R. Rao, *Biopolymers*, 1983, **22**, 2091-2104.
91. S. Scheiner, *Molecules*, 2016, **21**, 1426.
92. P. S. Walsh, K. N. Blodgett, C. McBurney, S. H. Gellman and T. S. Zwier, *Angew. Chem. Int. Ed.*, 2016, **55**, 14618-14622.
93. G. N. Ramachandran, *Chemistry of Collagen*, Academic Press, London, 1967.

94. G. N. Ramachandran, *Biopolymers*, 1968, **6**, 1494-+.
95. S. S. Zimmerman, L. L. Shipman and H. A. Scheraga, *Journal of Physical Chemistry*, 1977, **81**, 614-619.
96. S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, 1978, **17**, 1885-1890.
97. S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, 1977, **16**, 811-843.
98. S. Aravinda, V. V. Harini, N. Shamala, C. Das and P. Balaram, *Biochemistry*, 2004, **43**, 1832-1846.
99. Y. K. Kang and H. S. Park, *New J. Chem.*, 2016, **40**, 8565-8578.
100. J. Li, Y. Kuang, Y. Gao, X. Du, J. Shi and B. Xu, *J. Am. Chem. Soc.*, 2013, **135**, 542-545.
101. Z. Feng and B. Xu, *Biomolecular Concepts*, 2016, **7**, 179-187.
102. D. H. Levy, *Annu. Rev. Phys. Chem.*, 1980, **31**, 197-225.
103. J. R. Cable, M. J. Tubergen and D. H. Levy, *J. Am. Chem. Soc.*, 1989, **111**, 9032-9039.
104. Y. D. Park, T. R. Rizzo, L. A. Peteanu and D. H. Levy, *J. Chem. Phys.*, 1986, **84**, 6539-6549.
105. R. H. Page, Y. R. Shen and Y. T. Lee, *J. Chem. Phys.*, 1988, **88**, 5362-5376.
106. R. H. Page, Y. R. Shen and Y. T. Lee, *J. Chem. Phys.*, 1988, **88**, 4621-4636.
107. S. Wiedemann, A. Metsala, D. Nolting and R. Weinkauf, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2641-2649.
108. R. N. Pribble and T. S. Zwier, *Science*, 1994, **265**, 75-79.
109. J. Grotemeyer, U. Boesl, K. Walter and E. W. Schlag, *Organic Mass Spectrometry*, 1986, **21**, 645-653.
110. R. Tembreull and D. M. Lubman, *Analytical Chemistry*, 1987, **59**, 1082-1088.
111. L. C. Snoek, E. G. Robertson, R. T. Kroemer and J. P. Simons, *Chemical Physics Letters*, 2000, **321**, 49-56.
112. W. H. James, E. E. Baquero, S. H. Choi, S. H. Gellman and T. S. Zwier, *J. Phys. Chem. A*, 2010, **114**, 1581-1591.
113. E. E. Baquero, W. H. James, S. H. Choi, S. H. Gellman and T. S. Zwier, *J. Am. Chem. Soc.*, 2008, **130**, 4784-4794.
114. M. Gerhards and C. Unterberg, *Phys. Chem. Chem. Phys.*, 2002, **4**, 1760-1765.
115. J. M. Bakker, I. Compagnon, G. Meijer, G. von Helden, M. Kabelac, P. Hobza and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2810-2815.
116. A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, *P Natl Acad Sci USA*, 2005, **102**, 20-23.
117. A. Abo-Riziq, B. O. Crews, M. P. Callahan, L. Grace and M. S. de Vries, *Angew. Chem. Int. Ed.*, 2006, **45**, 5166-5169.
118. V. Yatsyna, R. Mallat, T. Gorn, M. Schmitt, R. Feifel, A. M. Rijs and V. Zhaunerchyk, *J. Phys. Chem. A*, 2019, **123**, 862-872.
119. S.-i. Ishiuchi, T. Asakawa, H. Mitsuda, M. Miyazaki, S. Chakraborty and M. Fujii, *J. Phys. Chem. A*, 2011, **115**, 10363-10369.
120. S. Chakraborty, K. Yamada, S.-i. Ishiuchi and M. Fujii, *Chemical Physics Letters*, 2012, **531**, 41-45.
121. S. Ishiuchi, K. Yamada, S. Chakraborty, K. Yagi and M. Fujii, *Chemical Physics*, 2013, **419**, 145-152.
122. W. Chin, M. Mons, J.-P. Dognon, F. Piuzzi, B. Tardivel and I. Dimicoli, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2700-2709.
123. W. Chin, I. Compagnon, J. P. Dognon, C. Canuel, F. Piuzzi, I. Dimicoli, G. von Helden, G. Meijer and M. Mons, *J. Am. Chem. Soc.*, 2005, **127**, 1388-1389.
124. K. Okuyama, K. Okuyama, S. Arnott, M. Takayanagi and M. Kakudo, *Journal of molecular biology*, 1981, **152**, 427-443.
125. H. v. Berlepsch, E. Brandenburg, B. Koksch and C. Böttcher, *Langmuir*, 2010, **26**, 11452-11460.
126. J. R. Cable, M. J. Tubergen and D. H. Levy, *J. Am. Chem. Soc.*, 1987, **109**, 6198-6199.
127. D. H. Levy, *Science*, 1981, **214**, 263-269.

128. R. E. Smalley, L. Wharton and D. H. Levy, *Accounts of Chemical Research*, 1977, **10**, 139-145.
129. M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585-612.
130. J. Anderson, R. Andres and J. Fenn, *Molecular beams. Ross J.(ed.)*, 1966, 275.
131. W. C. Wiley and I. H. McLaren, *Review of Scientific Instruments*, 1955, **26**, 1150-1157.
132. L. Frydman, A. Lupulescu and T. Scherf, *J. Am. Chem. Soc.*, 2003, **125**, 9204-9217.
133. Y. Shrot and L. Frydman, *J Magn Reson*, 2011, **209**, 352-358.
134. J. Epp, in *Materials characterization using nondestructive evaluation (NDE) methods*, Elsevier 2016, pp. 81-124.
135. M. J. Frisch, et al., Gaussian 09 (Revision D.01), Gaussian, Inc., Wallingford CT, 2009
136. E. D. Glendening, C. R. Landis and F. Weinhold, *Journal of Computational Chemistry*, 2013, **34**, 1429-1437.
137. J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *Journal of Chemical Theory and Computation*, 2011, **7**, 625-632.
138. E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-Garcia, A. J. Cohen and W. T. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498-6506.
139. C. B. Anfinsen, *Science*, 1973, **181**, 223-230.
140. P. N. Lewis, F. A. Momany and H. A. Scheraga, *Proc. Natl. Acad. Sci.*, 1971, **68**, 2293.
141. I. D. Kuntz, *J. Am. Chem. Soc.*, 1972, **94**, 4009-&.
142. K. A. Dill, *Biochemistry*, 1990, **29**, 7133-7155.
143. D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman and Comapny Ltd, New York, USA, 2008.
144. E. Vass, M. Hollosi, F. Besson and R. Buchet, *Chem. Rev.*, 2003, **103**, 1917-1954.
145. J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry*, W. H. Freeman, New York, 2002.
146. A. W. Burgess and H. A. Scheraga, *Biopolymers*, 1973, **12**, 2177-2183.
147. G. Pohl, A. Perczel, E. Vass, G. Magyarfalvi and G. Tarczay, *Phys. Chem. Chem. Phys.*, 2007, **9**, 4698-4708.
148. I. R. Gould, W. D. Cornell and I. H. Hillier, *J. Am. Chem. Soc.*, 1994, **116**, 9250-9256.
149. M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren and R. A. Friesner, *J. Am. Chem. Soc.*, 1997, **119**, 5908-5920.
150. H. R. Kilgore and R. T. Raines, *J. Am. Chem. Soc.*, 2018, **140**, 17606-17611.
151. S. K. Singh and A. Das, *Phys. Chem. Chem. Phys.*, 2015, **17**, 9596-9612.
152. R. W. Newberry and R. T. Raines, *Accounts of Chemical Research*, 2017, **50**, 1838-1846.
153. G. J. Bartlett, A. Choudhary, R. T. Raines and D. N. Woolfson, *Nature Chemical Biology*, 2010, **6**, 615-620.
154. S. K. Singh, A. Das and G. W. Breton, *J. Phys. Chem. A*, 2016, **120**, 6258-6269.
155. S. K. Singh, K. K. Mishra, N. Sharma and A. Das, *Angewandte Chemie-International Edition*, 2016, **55**, 7801-7805.
156. S. K. Singh, K. K. Mishra, N. Sharma and A. Das, *Angew. Chem. Int. Ed.*, 2016, **DOI: 10.1002/anie.201511925**

157. S. K. Singh, J. K. Vaishnav and A. Das, *J. Chem. Phys.*, 2016, **145**, 104302.
158. S. K. Singh, S. Kumar and A. Das, *Phys. Chem. Chem. Phys.*, 2014, **16**, 8819-8827.
159. J. R. Gord, D. M. Hewett, A. O. Hernandez-Castillo, K. N. Blodgett, M. C. Rotondaro, A. Varuolo, M. A. Kubasik and T. S. Zwier, *Phys. Chem. Chem. Phys.*, 2016, **18**, 25512-25527.
160. P. S. Walsh, J. C. Dean, C. McBurney, H. Kang, S. H. Gellman and T. S. Zwier, *Phys. Chem. Chem. Phys.*, 2016, **18**, 11306-11322.
161. V. Yatsyna, R. Mallat, T. Gorn, M. Schmit, R. Feifel, A. M. Rijs and V. Zhaunerchyk, *J. Phys. Chem. A*, 2019, **123**, 862-872.
162. V. Yatsyna, R. Mallat, T. Gorn, M. Schmitt, R. Feifel, A. M. Rijs and V. Zhaunerchyk, *Phys. Chem. Chem. Phys.*, 2019.

163. C. Cabezas, M. Varela and J. L. Alonso, *Angewandte Chemie-International Edition*, 2017, **56**, 6420-6425.
164. E. G. Buchanan, W. H. James, S. H. Choi, L. Guo, S. H. Gellman, C. W. Muller and T. S. Zwier, *J. Chem. Phys.*, 2012, **137**, 094301.
165. W. Y. Sohn, V. Brenner, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29969-29978.
166. E. Gloaguen, B. de Courcy, J. P. Piquemal, J. Pilme, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier and M. Mons, *J. Am. Chem. Soc.*, 2010, **132**, 11860-11863.
167. Y. Loquais, E. Gloaguen, S. Habka, V. Vaquero-Vara, V. Brenner, B. Tardivel and M. Mons, *J. Phys. Chem. A*, 2015, **119**, 5932-5941.
168. E. Gloaguen, B. Tardivel and M. Mons, *Structural Chemistry*, 2016, **27**, 225-230.
169. S. Blanco, A. Lesarri, J. C. López and J. L. Alonso, *J. Am. Chem. Soc.*, 2004, **126**, 11675-11683.
170. A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899-926.
171. S. Kumar, P. Biswas, I. Kaul and A. Das, *J. Phys. Chem. A*, 2011, **115**, 7461.
172. S. Kumar, I. Kaul, P. Biswas and A. Das, *J. Phys. Chem. A*, 2011, **115**, 10299-10308.
173. S. K. Singh, S. More, S. Kumar, K. K. Mishra, K. N. Ganesh and A. Das, *Phys. Chem. Chem. Phys.*, 2019, **21**, 4755-4762.
174. S. K. Singh, P. R. Joshi, R. A. Shaw, J. G. Hill and A. Das, *Phys. Chem. Chem. Phys.*, 2018, **20**, 18361-18373.
175. K. Y. Hung, P. W. R. Harris and M. A. Brimble, *Synlett*, 2009, 1233-1236.
176. F. Piuzzi, I. Dimicoli, M. Mons, B. Tardivel and Q. Zhao, *Chemical Physics Letters*, 2000, **320**, 282-288.
177. G. Meijer, M. S. de Vries, H. E. Hunziker and H. R. Wendt, *Applied Physics B*, 1990, **51**, 395-403.
178. S. L. Mayo, B. D. Olafson and W. A. Goddard, *The Journal of Physical Chemistry*, 1990, **94**, 8897-8909.
179. G. Imre, G. Veress, A. Volford and Ö. Farkas, *J. Mol. Struct. THEOCHEM*, 2003, **666-667**, 51-59.
180. F. M. J. et.al, M. J. Frisch. et al., Gaussian 16 (Revision C.01), Gaussian, Inc., Wallingford CT, 2016., 2016
181. M. Head-Gordon, J. A. Pople and M. J. Frisch, *Chemical Physics Letters*, 1988, **153**, 503.
182. A. L. Sobolewski and W. Domcke, *Chemical Physics*, 2003, **294**, 73-83.
183. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623-11627.
184. R. E. Stratmann, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 1998, **109**, 8218-8224.
185. I. Usabiaga, J. Gonzalez, P. F. Arnaiz, I. Leon, E. J. Cocinero and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12457-12465.
186. http://www.nist.gov/mm1/csd/informatics_research/thermochemistry_script.cfm.
187. E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498.
188. A. Otero-de-la-Roza, E. R. Johnson and J. Contreras-García, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12165-12172.
189. W. Y. Sohn, J. J. Kim, M. Jeon, T. Aoki, S.-i. Ishiuchi, M. Fujii and H. Kang, *Phys. Chem. Chem. Phys.*, 2018, **20**, 19979-19986.
190. R. J. Plowright, E. Gloaguen and M. Mons, *Chemphyschem*, 2011, **12**, 1889-1899.
191. E. Gloaguen, H. Valdes, F. Pagliarulo, R. Pollet, B. Tardivel, P. Hobza, F. Piuzzi and M. Mons, *J. Phys. Chem. A*, 2010, **114**, 2973-2982.
192. M. Tsuboi, T. Shimanouchi and S.-I. Mizushima, *J. Am. Chem. Soc.*, 1959, **81**, 1406-1411.
193. F. Weinhold and C. R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*, Cambridge University Press, Cambridge, UK, 2005.
194. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics & Modelling*, 1996, **14**, 33-38.
195. H. A. Scheraga, *Pure Appl. Chem.*, 1973, **36**, 1-8.

196. C. B. Anfinsen and H. A. Scheraga, *Adv. Protein Chem.*, 1975, **29**, 205-300.
197. N. J. Zondlo, *Nat. Chem. Biol.*, 2010, **6**, 567-568.
198. K. A. Dill, *Biochemistry*, 1985, **24**, 1501-1509.
199. C. B. Anfinsen, E. Haber, M. Sela and F. H. White, *Proc. Natl. Acad. Sci.*, 1961, **47**, 1309.
200. G. D. Rose, L. M. Glerasch and J. A. Smith, in *Advances in Protein Chemistry*, eds. C. B. Anfinsen, J. T. Edsall and F. M. Richards, Academic Press 1985, vol. 37, pp. 1-109.
201. H. A. Nagarajaram, P. K. C. Paul, K. Ramanarayanan, K. V. Soman and C. Ramakrishnan, *Int. J. Pept. Protein Res.*, 1992, **40**, 383-394.
202. R. H. Pain and B. Robson, *Nature*, 1970, **227**, 62-63.
203. J. P. Schneider and J. W. Kelly, *Chem. Rev.*, 1995, **95**, 2169-2187.
204. W. A. Loughlin, J. D. A. Tyndall, M. P. Glenn and D. P. Fairlie, *Chem. Rev.*, 2004, **104**, 6085-6118.
205. P. N. Lewis, F. A. Momany and H. A. Scheraga, *Biochim. Biophys. Acta*, 1973, **303**, 211-229.
206. I. L. Karle, S. K. Awasthi and P. Balaram, *Proc. Natl. Acad. Sci.*, 1996, **93**, 8189.
207. I. L. Karle, C. Das and P. Balaram, *Proc. Natl. Acad. Sci.*, 2000, **97**, 3034.
208. H. E. Stanger and S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 4236-4237.
209. C. M. Deber, *Macromolecules*, 1974, **7**, 47-51.
210. S. Bakels, M.-P. Gaigeot and A. M. Rijs, *Chem. Rev.*, 2020, **120**, 3233-3260.
211. M. Gerhards, in *Principles of Mass Spectrometry Applied to Biomolecules*, eds. J. Laskin and C. Lifshitz, John Wiley & Sons, New Jersey, 2006, pp. 3-61.
212. E. C. Stanca-Kaposta and J. P. Simons, in *Handbook of High-resolution Spectroscopy*, eds. M. Quack and F. Merkt., John Wiley & Sons 2011.
213. W. Chin, F. Piuzzi, I. Dimicoli and M. Mons, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1033-1048.
214. R. Cohen, B. Brauer, E. Nir, L. Grace and M. S. de Vries, *J. Phys. Chem. A*, 2000, **104**, 6351-6355.
215. D. Toroz and T. van Mourik, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3463-3473.
216. A. Masson, M. Z. Kamrath, M. A. S. Perez, M. S. Glover, U. Rothlisberger, D. E. Clemmer and T. R. Rizzo, *Journal of the American Society for Mass Spectrometry*, 2015, **26**, 1444-1454.
217. S. Hayakawa, M. Hashimoto, H. Matsubara and F. Tureček, *J. Am. Chem. Soc.*, 2007, **129**, 7936-7949.
218. D. Reha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem. Eur. J.*, 2005, **11**, 6803-6817.
219. W. Y. Sohn, S. Habka, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17128-17142.
220. S. Kumar, S. K. Singh, C. Calabrese, A. Maris, S. Melandri and A. Das, *Phys. Chem. Chem. Phys.*, 2014, **16**, 17163-17171.
221. K. K. Mishra, S. K. Singh, S. Kumar, G. Singh, B. Sarkar, M. S. Madhusudhan and A. Das, *J. Phys. Chem. A*, 2019, **123**, 5995-6002.
222. P. Panwaria and A. Das, in *Modern Techniques of Spectroscopy: Basics, Instrumentation, and Application*, eds. K. D. Singh, M. Pradhan and A. Materny, Springer, Singapore, 2021, pp. 57-86.
223. H. Goto and E. Osawa, *J. Chem. Soc., Perkin trans. 2*, 1993, 187-198.
224. E. Gloaguen, F. Pagliarulo, V. Brenner, W. Chin, F. Piuzzi, B. Tardivel and M. Mons, *Phys. Chem. Chem. Phys.*, 2007, **9**, 4491-4497.
225. Y. Umezawa, S. Tsuboyama, H. Takahashi, J. Uzawa and M. Nishio, *Bioorg. Med. Chem.*, 1999, **7**, 2021-2026.
226. F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith and D. G. Watson, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 187-204.
227. Z. Imani, V. R. Mundlapati, G. Goldsztejn, V. Brenner, E. Gloaguen, R. Guillot, J. P. Baltaze, K. Le Barbu-Debus, S. Robin, A. Zehnacker, M. Mons and D. J. Aitken, *Chem. Sci.*, 2020, **11**, 9191-9197.
228. T. E. Bull, *J. Magn. Reson.*, 1988, **80**, 470-481.

229. T. E. Bull, *J. Magn. Reson.*, 1991, **93**, 596-602.
230. B. R. Leeflang and L. M. J. Kroon-Batenburg, *J. Biomol. NMR*, 1992, **2**, 495-518.
231. G. M. Sheldrick and T. R. Schneider, in *Methods in enzymology*, Elsevier 1997, vol. 277, pp. 319-343.
232. F. H. Allen and R. Taylor, *Chemical Society Reviews*, 2004, **33**, 463-475.
233. C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 2016, **72**, 171-179.
234. R. A. Berg and D. J. Prockop, *Biochem Biophys Res Commun*, 1973, **52**, 115-120.
235. M. Bansal, S. K. Brahmachari and V. Sasisekharan, *Macromolecules*, 1979, **12**, 19-23.
236. S. K. Brahmachari, M. Bansal, V. S. Ananthanarayanan and V. Sasisekharan, *Macromolecules*, 1979, **12**, 23-28.
237. M. Nagarajan and V. S. R. Rao, *Current Science*, 1977, **46**, 395-400.
238. J. G. Beeley, *Biochem Biophys Res Commun*, 1977, **76**, 1051-1055.
239. D. Small, P. Y. Chou and G. D. Fasman, *Biochemical and Biophysical Research Communications*, 1977, **79**, 341-346.
240. T. Krieg, U. Feldmann, W. Kessler and P. K. Müller, *Human Genetics*, 1979, **46**, 41-49.
241. J. A. Stearns, C. Seaiby, O. V. Boyarkin and T. R. Rizzo, *Phys. Chem. Chem. Phys.*, 2009, **11**, 125-132.
242. V. Brenner, F. Piuzzi, I. Dimicoli, B. Tardivel and M. Mons, *The journal of physical chemistry. A*, 2007, **111**, 7347-7354.
243. W. Chin, F. Piuzzi, J. P. Dognon, L. Dimicoli, B. Tardivel and M. Mons, *J. Am. Chem. Soc.*, 2005, **127**, 11900-11901.
244. J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct. THEOCHEM*, 1999, **464**, 211-226.
245. S. Miertuš, E. Scrocco and J. Tomasi, *Chemical Physics*, 1981, **55**, 117-129.
246. M. Cossi, N. Rega, G. Scalmani and V. Barone, *Journal of Computational Chemistry*, 2003, **24**, 669-681.